Disease Stories: Visual and Narrative Strategies for Scientific Disease Communication

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von Sarah Mittenentzwei

geb. am 02.02.1996 in Rotenburg an der Fulda

Gutachterinnen/Gutachter

Prof. Dr.-Ing. Bernhard Preim
Dr. Petra Isenberg
Prof. Dr. Renata Raidou

Magdeburg, den 23.10.2025

DISEASE STORIES: VISUAL AND NARRATIVE STRATEGIES FOR SCIENTIFIC DISEASE COMMUNICATION

SARAH MITTENENTZWEI

DISSERTATION

Sarah Mittenentzwei: *Disease Stories: Visual and Narrative Strategies for Scientific Disease Communication*, Dissertation, © 17. April 2025

[...] Health literacy is a stronger predictor
of an individual's health status than
income, employment status, education level
and racial or ethnic group.

— World Health Organization[1]

1 Based on a discussion on impact factors on health in the United States of America by the World Health Organization (`https://www.who.int/news-room/fact-sheets/detail/health-literacy`, accessed 03.03.2025) referencing Barry D. Weiss. "Health literacy and patient safety: help patients understand. Manual for clinicians", 2nd Edition, *American Medical Association* (2007)

ABSTRACT

Communicating medical topics is a fundamental task for physicians and public health institutions, whether during patient-physician consultations or when promoting preventive measures to the general public. However, low health literacy can hinder individuals' ability to understand health information and make informed decisions about their own health. Narrative medical visualization aims to bridge this gap by making complex medical information accessible and understandable to lay audiences. It integrates visuals, data, and storytelling to communicate medical information in an engaging and effective way.

This thesis focuses on a subtype of narrative medical visualization, namely disease stories, exploring the design space for communicating information about diseases, particularly regarding prevention and treatment decisions. In collaboration with experts from the fields of medicine and design, several diseases are used as exemplary applications. A formalized design process for creating disease stories is introduced, outlining essential components. Key design decisions are investigated through a series of case studies focusing on story content, character, structure, and target audiences.

Two case studies compare different visualization (Phong shading, outlines and feature lines, hatching and Fresnel shading) and presentation techniques (showing time steps side-by-side or successive) for depicting pathological growth and shrinkage, using cerebral aneurysms and primary liver tumors as examples. Evaluations with participants from a lay audience provide quantitative data on task performance and self-reported preferences regarding the compared techniques.

Another case study explored the AI-assisted creation of characters for disease stories. As a proof of concept, characters were generated based on data from patients with non-alcoholic fatty liver disease extracted from an epidemiological study. The generated characters were evaluated both by lay participants, who assessed how accurately the characters conveyed the underlying patient information, and by design experts, who discussed potential refinements.

Additionally, a character-driven story structure was developed for a case study on cerebral small vessel disease (CSVD). The study compared three versions of the story: one with a patient as the protagonist, another with a physician, and a third with no human protagonist. A survey targeted at a lay audience was used to examine how these narrative choices shaped the user experience.

The CSVD story was later restructured following a tension arc structure and, together with a story about bicuspid aortic valve, was implemented as both an interactive slideshow and a scrollytelling version. A case study with lay participants analyzed user behavior and preferences across these different versions.

To gain insight into real-world audience perspectives beyond controlled lab studies, comments from educational YouTube videos on various health topics were collected and analyzed using natural language processing. This analysis investigated user sentiment, emotions, and motivations for engaging with health-related content online.

While these studies only explore a small portion of the vast design space of disease stories, the results provide valuable insights. Through the investigation of key design decisions, the strengths and limitations of the proposed implementations are discussed. The findings contribute to the foundation for the interdisciplinary design of disease stories, supporting future research at the intersection of visualization and health communication.

ZUSAMMENFASSUNG

Die Vermittlung medizinischer Themen ist eine grundlegende Aufgabe für Ärzte und Einrichtungen des öffentlichen Gesundheitswesens, sei es bei Arzt-Patienten-Konsultationen oder bei der Förderung von Präventionsmaßnahmen in der breiten Öffentlichkeit. Eine geringe Gesundheitskompetenz kann jedoch die Fähigkeit des Einzelnen beeinträchtigen, Gesundheitsinformationen zu verstehen und informierte Entscheidungen über die eigene Gesundheit zu treffen. Die narrative medizinische Visualisierung zielt darauf ab, dieses Hindernis zu überwinden, indem sie komplexe medizinische Informationen für ein Laienpublikum zugänglich und verständlich macht. Sie integriert Visualisierungen, Daten und Geschichten, um medizinische Informationen auf ansprechende und effektive Weise zu vermitteln.

Diese Arbeit konzentriert sich auf eine Unterart der narrativen medizinischen Visualisierung, die Krankheitsgeschichten, und untersucht den Gestaltungsspielraum für die Vermittlung von Informationen über Krankheiten, insbesondere im Hinblick auf Präventions- und Behandlungsentscheidungen. In Zusammenarbeit mit Experten aus den Bereichen Medizin und Design werden verschiedene Krankheiten als Anwendungsbeispiele herangezogen. Es wird ein formalisierter Designprozess für die Erstellung von Krankheitsgeschichten vorgestellt, der die wesentlichen Komponenten umreißt. Einige essenzielle Designentscheidungen werden anhand einer Reihe von Fallstudien untersucht, die sich auf Inhalt, Charakter, Struktur und Zielgruppen der Geschichten konzentrieren.

In zwei Fallstudien werden verschiedene Visualisierungs- (*Phong Shading*, *Outlines* und *Feature lines*, *Hatching* und *Fresnel Shading*) und Präsentationstechniken (Darstellung von Zeitschritten nebeneinander oder nacheinander) zur Visualisierung von pathologischem Wachstum und Rückbildung am Beispiel von zerebralen Aneurysmen und primären Lebertumoren verglichen. Auswertungen mit Teilnehmern aus einem Laienpublikum liefern quantitative Daten über die Leistung bei den gestellten Aufgaben und zu den selbstberichteten Präferenzen bezüglich der verglichenen Techniken.

Eine weitere Fallstudie untersucht die KI-gestützte Erstellung von Charakteren für Krankheitsgeschichten. Zur Erprobung des Konzepts wurden Charaktere auf der Grundlage von Daten von Patienten und Patientinnen mit nichtalkoholischer Fettlebererkrankung aus einer epidemiologischen Studie erstellt. Die generierten Charaktere wurden sowohl von Laien bewertet, die beurteilten, wie genau die Charaktere die zugrundeliegenden Patienteninformationen wiedergeben, als auch von Designexperten, die mögliche Verbesserungen diskutierten.

Darüber hinaus wurde für eine Fallstudie über die zerebrale Kleingefäßerkrankung (CSVD) eine charaktergesteuerte Erzählstruktur entwickelt. In der Studie wurden drei Versionen der Geschichte verglichen: in einer Version trat eine Patientin als Protagonist auf, in einer anderen eine Ärztin und in der letzten kein menschlicher Protagonist. Innerhalb einer Umfrage mit Laien wurde untersucht, wie sich diese narrativen Entscheidungen auf das Nutzererlebnis auswirkten.

Die CSVD-Geschichte wurde später anhand eines Spannungsbogens umstrukturiert und zusammen mit einer Geschichte über bikuspide Aortenklappen sowohl als interaktive Diashow als auch als *Scrollytelling*-Version umgesetzt. In einer Fallstudie mit Laien wurden das Nutzerverhalten und die Präferenzen für die verschiedenen Versionen analysiert.

Um über kontrollierte Laborstudien hinaus Einblicke in die Perspektiven realer Nutzer zu erhalten, wurden Kommentare von YouTube-Videos zu verschiedenen Gesundheitsthemen gesammelt und mithilfe von *Natural Language Processing* analysiert. Diese Analyse untersuchte die Stimmung, die Emotionen und die Beweggründe der Nutzer, sich mit gesundheitsbezogenen Online-Inhalten zu beschäftigen.

Auch wenn diese Studien nur einen kleinen Ausschnitt des breiten Gestaltungsspielraums von Krankheitsgeschichten untersuchen, liefern die Ergebnisse wertvolle Erkenntnisse. Durch die Untersuchung essenzieller Designentscheidungen werden die Stärken und Schwächen der umgesetzten Implementierungen diskutiert. Die Ergebnisse liefern Grundlagen zur interdisziplinären Gestaltung von Krankheitsgeschichten und unterstützen zukünftige Forschung an der Schnittstelle zwischen Visualisierung und Gesundheitskommunikation.

CONTENTS

Part I

PRELIMINARIES

This thesis begins by introducing the main research question. It then outlines the foundational concepts from narrativity, narrative visualization, and medicine that inform and support the work.

1 INTRODUCTION

Science and medicine evolves rapidly with new insights being gained regularly. Especially in the context of medicine, communicating scientific consensus can improve the health of individuals and communities (e.g., due to public health interventions promoting preventive behavior) and help patients and physicians communicate with each other (e.g., during patient-physician-consultation when diseases and treatment methods are discussed) empowering lay persons to make informed decisions about their health. Narrative medical visualization bridges the gap between an audience with limited health literacy and the need of medical professionals to communicate health-related topics. It combines visuals, data, and storytelling to make information accessible and engaging for lay audiences.

WHY TO USE VISUALS? Due to the mass of information (both real and fake) many individuals are exposed to nowadays, there is the danger of the public being oversaturated with information. Communicating using visuals is an effective way for combating a so-called infodemic [111].

WHY TO USE DATA? It is important that the information is authentic and relatable, e.g., pictorial warnings on cigarette packs are more engaging when containing images of diseased anatomy, real people, and personal testimonials [176, 181]. This shows that real and relatable content is more effective, motivating the use of data for authentic health communication.

WHY TO USE STORYTELLING? Communication incorporating the data visualizations within a story, especially when featuring a relatable main character, is preferred over a purely didactic approach by audiences and enhances engagement as well as effectiveness [56]. While the potential of narrative visualization for communicating data is commonly agreed upon, the design space is vast, ranging from different story structures and genres to visual metaphors [75, 210].

This work centers on the communication of medical information through disease stories, aiming to inform general audiences about various conditions with an emphasis on preventable risk factors and informed consent about medical interventions. This approach aligns with ongoing research in public health and health communication, which investigates strategies for crafting messages that promote health-enhancing behaviors at both individual and community levels [170].

1

1.1 STRUCTURE AND CONTRIBUTIONS

The contribution of this thesis focuses on the design of disease stories as a subgenre of narrative medical visualization. Through various case studies and close collaboration with domain experts from medicine and design, the following question is explored:

Main Research Question

How can narrative medical visualization be leveraged to improve the health literacy of lay audiences through disease stories?

To address this question, contributions to various aspects of disease stories are presented in the following chapters, see Figure 1.

Chapter 2 introduces foundational concepts of narrativity and narrative visualization, drawing parallels to the medical domain regarding health communication. As this thesis includes multiple case studies focusing on different diseases, Chapter 3 provides essential medical background information for each condition. Following this, the main part of the thesis delves into the design of disease stories and the conceptualization of the corresponding design process. This design process is introduced in Chapter 4 and illustrated through an example disease story. The subsequent chapters delve into various components of the disease story design process:

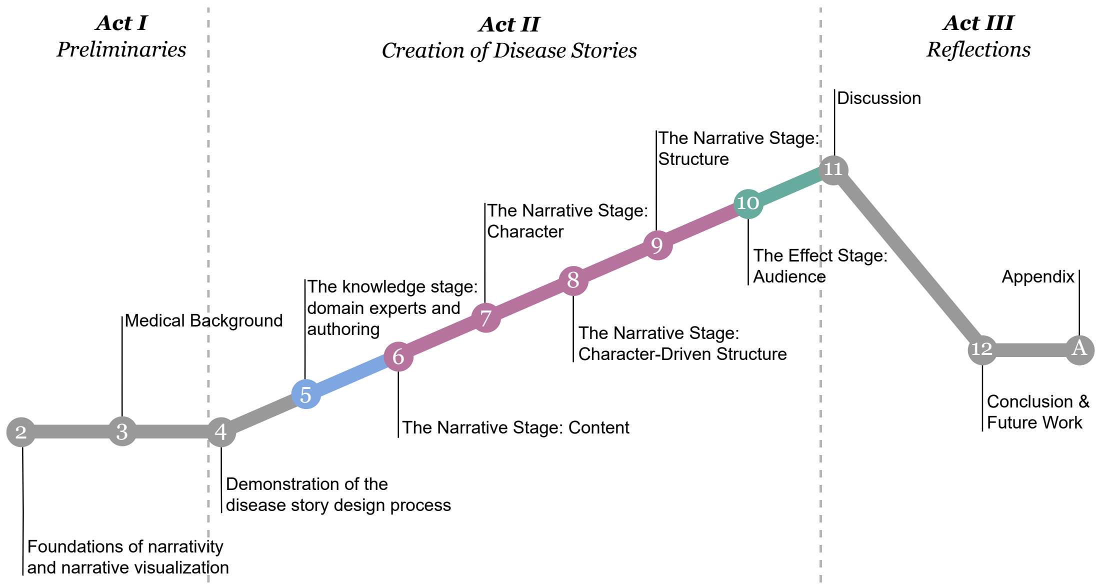


Figure 1: Thesis structure organized in a three-act-structure.

- The collaboration with medical domain experts is outlined in Chapter 5. This chapter also provides an overview of the data and tools utilized in the authoring process of the disease stories.
- An essential aspect of disease stories is their data-driven nature. Different approaches for visualizing surface models of pathologies that change their morphology over time, presenting insights from two case studies are examined in Chapter 6.
- Characters play a crucial role in guiding the audience through the disease story, personalizing data and health information. Methods for deriving characters from epidemiological study data to ensure their alignment with the medical topic are explored in Chapter 7.
- Characters can also shape the structure (order of events) of a disease story. A character-driven approach to structuring disease stories is introduced in Chapter 8 and different types of protagonists are compared, evaluating their impact on the audience.
- User behavior in two disease stories structured using a pyramid story structure is examined in Chapter 9. Each story is implemented in two different genres (interactive slideshow and scrollytelling) allowing for a comparison of user feedback regarding the interaction opportunities provided by each format.
- Moving away from controlled lab studies, user comments on YouTube videos covering various health topics are analyzed in Chapter 10. This investigation explores viewers' motivations for engaging with these videos and their feedback on design to inform the creation of future disease stories.

The work presented in this thesis is reflected upon in Chapter 11, which discusses its limitations and generalizability. Chapter 12 outlines potential directions for future research building upon this work.

1

2 FOUNDATIONS OF NARRATIVITY AND NARRATIVE VISUALIZATION

Narrative visualization, also known as visual data storytelling, is a rapidly expanding field in both research and industry. While newspapers such as *The New York Times* [236] and *ZEIT Online* [275] have employed data storytelling for some time, it has recently garnered interest from companies and researchers alike. By integrating data with visuals and storytelling, narrative visualization makes data interpretable and accessible to lay audiences. Data visualizations are often designed for domain experts, limiting their accessibility to broader audiences. To bridge this gap, narratives are incorporated to guide users through the data, making it more engaging and comprehensible. In the context of medical visualization, a wide array of methods exists to visualize medical data, prompting the central question: How can medical visualization be combined with a narrative to create effective narrative medical visualizations?

This chapter provides an overview of the theoretical foundations and key research in the field. It begins with an introduction to the concept of narrativity, followed by a discussion on the role of narratives in data visualization. The chapter concludes with an overview of storytelling practices in public health and health communication. The content is a combination and extension of insights from the following publications: [113, 156–160, 162, 163].

2.1 CONCEPTS OF NARRATIVITY

"Narrative" is a term that is used in many different contexts. While the most straightforward association will be stories, narratives can also be found in other forms of communication, such as reports, explanations, or arguments. Gerald Prince investigated how "narrative" displaced other terms, viewing the modern use of the word "narrative" as a way to soften firm stances, stating that *"One says 'narrative' instead of 'explanation' or 'argumentation' (because it is more tentative); one prefers 'narrative' to 'theory,' 'hypothesis,' or 'evidence' (because it is less scientistic); one speaks of a 'narrative' rather than 'ideology' (because it is less judgmental); one substitutes 'narrative' for 'message' (because it is more indeterminate) [197, 205]."* Furthermore, there is no scientific consensus on definitions and conceptualizations for the term narrative [49]. DeFina and Georgakopoulou [49] argue that there are different *views of narrative*. This thesis follows the view where a narrative is defined by structured sequences of events, that have a beginning, middle, and end. Therefore, in this context a narrative is similar to a story, e.g., from theater.

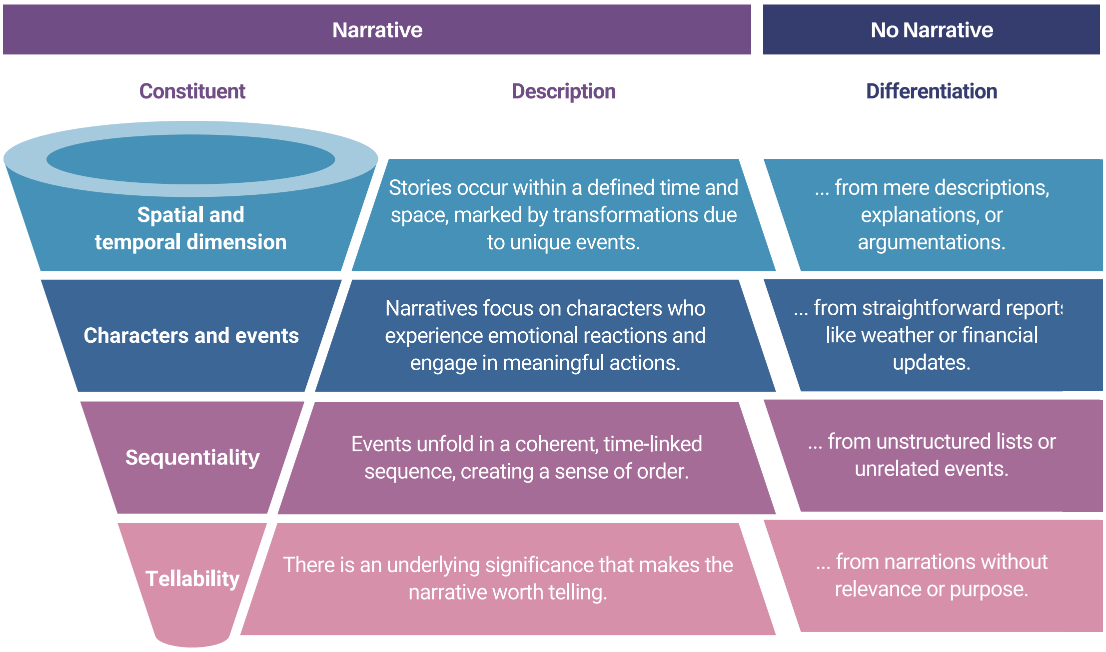


Figure 2: Funneling down the constituents of narrativity defined by Weber [251].

Narrativity should not be viewed as a binary construct where media is either narrative or non-narrative, instead there are *degrees of narrativity* [83, 189]. Therefore, it is important to consider that narrativity is a spectrum and media can be more or less narrative depending on how well they fulfill a given definition. Weber [251] highlights four *constituents of narrativity* derived from the work of Ryan [205](pp. 28-31). These constituents separate narratives from other forms of communication, see Figure 2: A narrative should have a *spatial and temporal dimension*, meaning that it is situated within a defined space and time and transforms due to events that happen over time. This excludes mere descriptions, explanations, or argumentations from being a narrative. *Characters and events* are included in narratives, the focus being on characters who experience emotional reactions or engage in meaningful actions. This excludes reports, e.g., weather or financial reports. Another constituent of narratives is sequentiality, meaning that events occur in coherent time-linked sequences, creating a sense of order which excludes unstructured lists of unrelated events. Finally, *tellability* is a constituent of narrativity, meaning that a narrative is worth telling.

Following the discourse of Weber [251] this thesis approaches the questions *what* and *how* a story is presented. The narrative describes what is presented. There are two major approaches for how to present a story, namely *telling* and *showing*. According to Genette [74] the difference of telling and showing describes the presence of a narrative instance. While for telling a narrator tells the story, in case of showing the goal is to *"making one forget that it is the narrator telling"* [74](p. 166). Telling vs. showing affects multiple aspects of the story such as the pacing (showing is generally slower than telling as telling presents information less detailed similar focussing on a summary-style) and the dialogues (only showing includes dialogues between characters). Additionally, showing provides information implicitly and the audience has to infer the information from the context, while telling provides information explicitly by directly telling it. However, again, there is no scientific consensus on a definition of telling and showing and the given distinction can only be seen as an heuristic [112]. Furthermore, telling and showing are not mutually exclusive; instead, both can be used in the same story to control its pacing.

2.2 NARRATIVES IN DATA STORYTELLING AND VISUALIZATION

Concepts of narrativity have been incorporated in the field of data visualization in recent years leading to the creation of *data(-driven) stories*. These visualizations are specifically designed for users with limited expert knowledge about the data and its context. Including narrative elements in a visualization helps to guide the user through the data by providing a structured path for the analysis and making it more engaging and understandable [89, 194]. One of the earliest scientific works exploring how storytelling techniques can be applied to information visualization was authored by Gershon and Page [75]. The term *narrative visualization* was shaped by Segel and Heer [210] who build a framework for narrative visualizations by analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. Since then narrative visualization and data storytelling, which are often used synonymously, have become a popular research topic in the visualization community especially in the subfield of information visualization [240]. Many concepts identified in narrative visualization research are derived by the analysis of data journalism [204]. Similarly, Segel and Heer [210] argue that *"a thorough understanding of the design space for narrative visualization has yet to emerge. In the meantime, practitioners such as artists and journalists have been forging paths through this space, and we might hope to gain insight from their explorations"*. Lee et al. [129] claim that, similar to classic storytelling, there is no scientific consensus for a definition of data storytelling in the literature. However, they argue that data storytelling is not the same as storytelling with data. While data visualization is commonly used to enhance a story (storytelling with data), narrative visualization strives to transform the data visualization into an independent story (data storytelling). They define data storytelling as *"a structured approach for communicating data insights that leverages the power of visual storytelling to make data more accessible and engaging for a wider audience"* [129].

Based on a literature review, Tong et al. [240] introduce typical dimensions of narrative visualizations: *Who* are the main subjects involved in storytelling for visualization (authoring tools and audience), *how* are stories told (narratives and transitions), *why* should we use storytelling for visualization (memorability and interpretation).

- Who (tells the story?)

  This question addresses both the authors and the audience. Authors are the creators of a story. While someone with expertise in design, visualization, and didactics would be ideally suited for this role, a domain expert (e.g., from medicine or other research fields) can also fulfill it, provided they are willing to invest the time and effort to make thoughtful and informed design decisions. The role of authoring tools is explored in determining who crafts the narrative and the origin of the story, alongside the engagement of the audience and how the story's key message is conveyed.

- How (is the story told?)

  Techniques used to convey a narrative visualization effectively are examined, including narrative structures, genres, characters, conflicts, and content, all of which play a crucial role in shaping how the story unfolds.

- Why (is the story told?)

  The purpose of narrative visualization lies in its ability to engage audiences effectively, making it a valuable tool for public outreach. By guiding users through a story, it empowers lay audiences to interpret and understand complex data. Furthermore, narrative visualization enhances the memorability and persuasiveness of data, making the conveyed information more impactful and accessible.

These dimensions will be used to structure the discussion of narrative visualization in the following.

2.2.1 *Who tells the story?*

The balance between authors and users in narrative visualization shapes how stories are experienced, ranging from author-driven narratives that guide users along a predefined path to reader-driven approaches that allow exploration and interaction. While reader-driven stories promote engagement and personalization, they risk losing the author's intended message. Creating narrative visualizations is a complex and time-consuming process, leading to the development of various authoring tools. These tools range from flexible frameworks for experienced designers to AI-supported systems that assist with repetitive tasks or provide recommendations [33]. While current AI tools are limited in fostering creativity, they offer potential for automating, e.g., visualization and annotation, making the creation process more accessible and efficient.

AUTHOR-USER ROLES. Different concepts deal with the interplay of authors and users, specifically how design decisions made by the author define the way users explore the story. Jain et al. [97] argue that a user can either be seen as an observer or a participant of a story. According to Segel and Heer [210] all stories are located on a spectrum between author-driven and reader-driven (where they use "reader" as a synonym for "user") stories which defines how freely the user can explore the story and how prominent the intended message of the author is conveyed, see Figure 3 a). However, it is debatable whether a purely reader-driven story can actually be considered a narrative as it is lacking an author-defined plot [230]. Rather, a reader-driven story would be close to a visual analysis tool, where the user can freely explore and interact with the data and the author has no control about the messaging process. However, interactivity is still desirable as it increases the engaging effect of a narrative visualization and personalizes the experience [97]. Aylett [13] describes the narrative paradox, where concurrent approaches of interactivity and explorability (reader-driven) and a predefined story plot (author-driven) are to be combined.

Segel and Heer [210] also introduce different narrative structures for stories that are either more author-driven or reader-driven. Their martini glass structure starts with a linear, author-driven sequence while it becomes more reader-driven towards the end. Weber [251] combines this structure with the narrative concepts of *telling vs. showing*, see Figure 3 b). In this context, the author-driven segments of the story correspond to the concept of telling, while the reader-driven segments align with the idea of showing. The view that there are different degrees of narrativity fits well to the narrative continuum introduced by Brent Dykes [58], where author-driven story aspects are associated with "more" storytelling (e.g., curated, explanatory) while reader-driven aspects are associated with "less" storytelling (e.g., automated, exploratory), see Figure 3 c). Stories that lean toward being more author-driven, emphasizing "telling" and "more" storytelling, position the user as an observer. In contrast, stories that are more reader-driven, employing techniques aligned with "showing" and featuring "less" storytelling, place the user in the role of a participant. From a data visualization perspective, Ynnerman et al. [273] introduced the concept of *exploranation*. Positioned between author-driven and reader-driven approaches, exploranation shows how elements of both exploratory and explanatory visualization can be combined for science communication. This concept has also been applied in data storytelling, such as in the communication of conflict data tailored to domain experts [150].

In the context of this thesis, the focus is on highly author-driven stories, where the author guides the user through the data and the information is presented in a predefined order with limited options for users to change the stories content.

ENGAGEMENT AND MESSAGING. Each story is typically crafted to convey at least one key message, e.g., presented as a takeaway for the audience. This key message is intrinsically tied to the narrative intent (the underlying reason for telling the story) as discussed in Section 2.2.3. To effectively communicate their intended message, authors must make deliberate design choices about how to present the data. These decisions raise important questions regarding ethics and transparency, explored in Section 2.3.

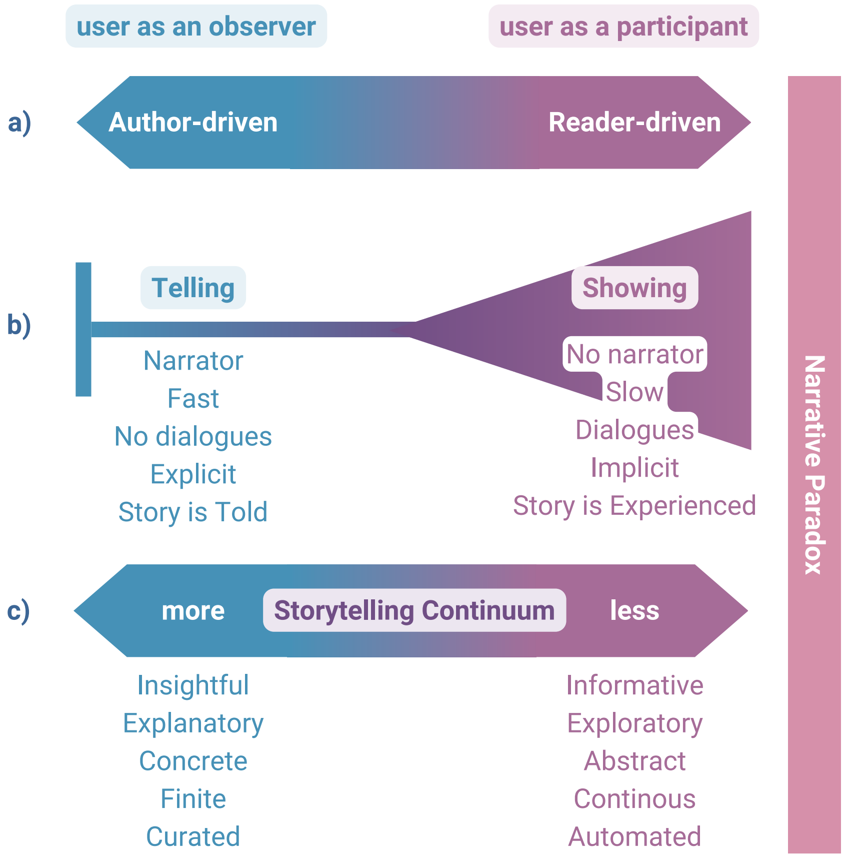


Figure 3: While a clear definition of narrative visualization does not exist, concepts provided in different works show parallels in terms of how a story is perceived and steered by the audience. In a) the concept of a narrative visualization being located somewhere between author-driven and reader-driven from Segel and Heer [210] is displayed. The more a story is reader-driven, the more the narrative paradox applies which describes the contradictory need of both interactive storytelling and giving the author control over the story [13]. As shown in b) the martini glass structure from Segel and Heer [210] starts with a linear, author-driven sequence while it becomes more reader-driven towards the end. Weber [251] combines this structure with the narrative concepts of telling vs. showing. The author-driven segments of the story correspond to the concept of telling, while the reader-driven segments align with the idea of showing. In c) the Storytelling Continuum defined by Dykes [58] is depicted. It shows that author-driven story aspects are associated with "more" storytelling while reader-driven aspects are associated with "less" storytelling.

Research has highlighted strategies for making stories more engaging to ensure the message resonates with the audience. Interaction has been shown to enhance user engagement [137], while incorporating human characters into narratives fosters even deeper engagement. Audiences tend to relate more strongly to people than to abstract data, making characters a powerful tool for connection and storytelling [17, 58]. Segel and Heer [210] argue that interactivity, while engaging, can potentially dilute the author's intended key message.

AUTHORING AND AUTOMATION. As the creation of narrative visualizations is a time-consuming process, several tools have been developed to support authors. With recent advancements in artificial intelligence (AI), especially generative artificial intelligence (GenAI) and machine learning (ML), the creation of narrative visualizations can be automated to some extent. However, creating narrative visualizations is a complex and highly creative process that requires careful decisions regarding the content, structure, and design of the story. This inherent complexity makes automation challenging.

Chen et al. [33] provide a taxonomy for tools ordered in ascending order based on their automation and intelligence levels:

- *Design spaces:* While not being a software tool, design spaces provide a conceptual framework for creating visualizations.
- *Authoring tools:* Specialized software designed for creating a specific type of content. Allow detailed user interaction and control over the design process.
- *ML/AI-supported tools:* Offer recommendations and assist with some automation, but users make final decisions.
- *ML/AI-generator tools:* Automate the entire process, requiring minimal user input.

The authors reflect that current ML/AI-supported tools and ML/AI-generator tools are limited. These tools are trained based on existing design decisions which is the foundation for their recommendations. Thus, they are usually not able to generate new design decisions. Therefore, they are not able to support the user in the creative process of designing a narrative visualization. However, they can assist in automating repetitive tasks, such as data cleaning, data transformation, and data visualization. On the other hand, authoring tools are more flexible and allow the user to make design decisions based on their own creativity. Therefore, the latter is more suitable for experienced designers who want to create unique and innovative visualizations. However, inexperienced authors may benefit from ML/AI-supported tools that provide recommendations and assist with some automation.

AUTHORING TOOLS. The taxonomy was applied to six genres: annotated charts, infographics, timeline & storylines, data comics, scrollytelling & slideshow, and data videos [33]. It is noticeable that most of the tools support the creation of 2D media, mainly charts, i.e., Tableau [234], ChartAccent [203] and Text-to-Viz [46], while the creation of 3D visualizations is rarely supported. As scrollytelling and slideshows are genres investigated in the scope of this thesis, respective authoring tools for this genre are of interest. Authoring tools for scrollytelling are VizFlow [232], Idyll [41], and ScrollyVis [165]. VizFlow links charts with text to create a variety of dynamic layouts. Idyll offers a markup language that allows users to control various aspects of an interactive narrative such as its layout and events (e.g., button clicks and triggers). ScrollyVis strives to provide more flexibility, including various visual media, such as images, videos and volume rendering as well as the possibility to include optional story paths additionally to a linear main path. All tools are designed to support the creation of scrollytelling stories, making it easier for authors to create interactive stories by providing different pre-defined arrangement and export options. However, the customization options for the arrangements provided by these tools are notably limited.

Tedric [261] is a ML/AI-supported tool for designing slideshows. The author has to provide the subject and the tool then automatically creates a slideshow with texts and images. There are different types of slides the tool can create, such as title slides, full screen image slides, and chart slides. ML/AI-generator tools in the area of scrollytelling and slideshows remain prototypes as of today, such as the works of Lu et al. [144] who propose an approach to automatically generate a story for a given dataset and Leake et al. [126] who automatically create audio-visual slideshows from a given text article. DATATALES is a large language model (LLM)–based tool developed to generate chart descriptions independent of specific genres [233].

2.2.2 *How is the story told?*

The design space for narrative visualization is broad. However, there are several concepts to guide the design of the narrative. To investigate this specifically in the area of narrative visualization, different works discuss important aspects of story construction. Bach et al. [16] defined patterns for data-driven storytelling, e.g., *call to action* and *rhetoric question*, claiming these patterns can be applied when the following requirements are met: having a *story*, *target audience*, and *intended effect* on the target audience. However, this leads to the question of how exactly a story is constructed. Fog et al. [63] discuss storytelling for defining a brand in a company setting and define four elements of storytelling: *Message*, *Conflict*, *Characters*, and *Plot*. The message is closely conntected to the intended effect from Bach et al [16]. While the message is what is communicated to the audience, the intended effect represents the author's underlying motivation, which may not always be explicitly revealed. However, the intent plays a crucial role in shaping the message. Brent Dykes [58] describes several techniques to transform data into stories. In particular, he claims that to build a story four ingredients are needed: *Conflict*, *Content*, *Characters*, and *Structure*. Two of the elements, *Conflict* and *Characters*, overlap with the ingredients from Fog et al [63]. Another element called *plot* is similar to *structure* as both describe the sequence of events. Considering these approaches several similarities become apparent, therefore, based on these works, the construction of a narrative can be broken down into and summarized by the following ingredients:

- Story [16]
  - Conflict [58, 63]
  - Content [58]
  - Characters [58, 63]
  - Structure, Plot [58, 63]
- Message [63]
- Intended Effect [16]
- Audience [16]

It becomes apparent that the focus lies on the interplay between how to convey information in a story format and the target audience and how the latter is influenced by the story. This section examines how the story is constructed, exploring elements such as conflict, content, characters, structure, and genres.

MODALITY AND GENRE. Narratives can be conveyed by various media. David Herman [83](p. xii) distinguishes between monomodal and multimodal narratives. While classical narratives are often monomodal, multimodal narratives combine multiple communicative channels, such as text, images, and sound. In the context of narrative visualization, narratives are often multimodal, combining data visualizations with text and images. Typically, the visual communicative channel is used (data visualizations) accompanied by text (e.g., annotations, explanations). New media have also allowed for a more interactive way of storytelling, e.g., by using interactive visualizations that allow the user to explore the data themselves.

Segel and Heer [210] introduced seven genres of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slideshow, and film/video/animation. From these genres, the applications discussed in this thesis are most similar to the slideshow genre. In interactive slideshows, the user can navigate through a series of slides that contain partly interactive data visualizations and text. Stolper et al. [230] added a new genre to the list, namely scrollytelling. Scrollytelling is often used in online journalism to guide the user through a story by scrolling down a webpage. The action of scrolling initiates changes in the visualization, e.g., fade-in annotations, playing animations, or triggering transitions between different views of the data. Computer games are not listed as another genre yet, but they are a popular medium for interactive storytelling. In computer games, the user can explore the story by interacting with the game world. The story is often presented in a non-linear way, allowing the user to explore the story at their own pace.

The term genre may be misleading in the context of narrative visualization. In everyday use, genre typically refers to the nature of a story's content, such as drama, comedy, or tragedy. A more suitable term might be medium, as it refers to the format in which a story is presented to the user, independent of the content's topic [146]. However, since *narrative genre* has become a well-established term in the literature referring to the medium of a narrative visualization, this thesis will adopt the terminology to maintain consistency with academic discourse.

STRUCTURE. The term structure is used for a variety of different approaches to organize a story. In the context of narrative visualization, the structure of the story is important to guide the user through the data and to make the story engaging and understandable. Brent Dykes [58] looked at narrative structure used in fictional storytelling, such as movies or novels and adapted them for narrative visualization. The main criteria for his selection was to balance simplicity and utility. He looked at three popular narrative structures from fictional storytelling: Three-act-structure, Freytag's Pyramid, and Campbell's Hero's Journey. He argues that while the three-act-structure is overly simplistic and Campbell's Hero's Journey overly complex, Freytag's Pyramid provides a more balanced framework, which he customized to develop the Data Storytelling Arc. Freytag's Pyramid is also endorsed by Yang et al. [270], who propose a design space for narrative visualizations. Their work illustrates how to structure facts and implement visual design techniques specific to the various sections of Freytag's Pyramid, drawing insights from their analysis of data videos. However, Wei et al. [254] highlight that Campbell's Hero's Journey can also be effectively applied in this context. Their study of data videos structured around the Hero's Journey offers guidance on integrating narrative, visual, and sound design to create compelling stories. Kosara [119] introduced the CFO (Claim, Facts, Conclusion) pattern, suggesting that many narrative visualizations deviate from the traditional story arc. Instead, they often begin with a claim or question, proceed by presenting facts as evidence, and ultimately arrive at a conclusion. However, certain parallels can be drawn between the CFO pattern and traditional story arcs. For instance, the claim or question at the beginning serves as a hook or inciting incident, setting the stage for the story. The introduction of facts builds tension, while the final conclusion can function as a climax. In this interpretation, the resolution commonly found in story arcs would be absent.

The narrative structures can be used to decide how to order the story content. Segel and Heer [210] introduce three archetypes for story structures specific for narrative visualization, striking a balance between an author-driven and a reader-driven approach. These structures focus on the role of the author and user within a data story, rather than the structure of the content of the story. In detail, they are called martini glass structure, interactive slideshow, and drill-down story.

With the linear martini glass structure (as described in Section 2.2.1) the author can ensure to communicate the main message of the story first and then give the user the option to explore the story further based on their personal interests. An interactive slideshow presents the user a sequence of slides that contain data visualizations and text. The user can navigate through the slides linearly. Within the slides the user can interact with the data visualizations. The drill-down story is a non-linear story structure that allows the user to navigate through the story more freely, being a structure most fitting for reader-driven stories. The user can explore parts of the story in any order. Another structure that focusses on the exploration opportunities in a story is the elastic structure [211]. In an elastic structure, the user has to follow a mandatory main path through the story, but can decide to explore optional side branches.

To differentiate the different types of story structures, they will be referred to in the following throughout this thesis:

- *Tension-driven structures:* structures that depict the rise and fall of tension throughout the story, e.g., Freytag's Pyramid
- *Character-driven structures:* structures that are centered around a character, e.g., Campbell's Hero's Journey
- *Interaction-driven structures:* structures that are defined through the interaction possibilities presented to the user, e.g., the martini glass structure

Different types of story structures can be combined. For example, a tension arc like Freytag's pyramid can be combined with a martini glass structure. For example, the story starts with a tension arc that resembles the handle of the class and then the last part of the tension arc leads to the exploration phase of the martini glass structure. This is especially useful because tension-driven and character-driven structures do not support authors in designing the exploration opportunities, while the story structures do not support the author in ordering the information in a way that is engaging for the reader.

CHARACTER. Characters can be humans or objects that are affected by the conflict. The characters are introduced in the story and are affected by the conflict. There are different types of characters, e.g., protagonists (main characters the user can sympathize with), antagonists (opponents of the protagonists), or supporting characters (side characters that support the protagonist) [28, 77]. Characters can serve as a tool to foster empathy with the user by humanizing the data, making the story more relatable and emotionally engaging. Additionally, they can provide valuable context for the data, presenting it in a more accessible and comprehensible manner.

CONFLICT. The conflict is the core element of a story. It is the driving force that moves the story forward and keeps the user engaged [58]. The conflict can be internal (within the character's mind) or external (the character's interactions with others or their environment), and is usually resolved within the story. In a narrative visualization, the conflict is often related to the data and the insights that can be derived from it. For example, the conflict can be resolved by analyzing the data and deriving meaningful insights from it. The conflict is usually introduced at the beginning of the story and resolved at the end. As narrative visualizations do not always feature human characters, conflict can also be derived from conspicuous elements visible in the data such as patterns, or outliers. By analyzing the data, e.g., explanations for these can be found, resolving the conflict.

CONTENT. The content of the story provides relevant information regarding the conflict and its solution. Mittenentzwei et al. [156] identified three types of content, namely *data-driven* (content derived from data, e.g., data visualizations), *context-driven* (content that provides necessary context for the data), and *character-driven* (content that is related to the character).

For data-driven content, which is most often communicated through data visualization, the visual literacy of the target audience must be considered. Visualizations need to be designed appropriate to their communicative goal (e.g., a line chart is suited to show a trend) and audiences' needs (i.e., taking into account the audience's health literacy, visual literacy, and numeracy to include appropriate labels, descriptions, and visual aids) [70]. However, visual communication is a challenging endeavor as seemingly minor design choices in the presentation of information can have important effects on risk perception and decision-making [149].

2.2.3 *Why is the story told?*

Using storytelling techniques, can make data more accessible, memorable and engaging for a wider audience, especially when the data is complex and the user is not familiar with the data or domain.

GUIDANCE, UNDERSTANDING, AND MEMORY. In the context of data storytelling, *showing* is often used to make the data more accessible to a wider audience. By *showing* the data, the user can explore the data themselves, making the story more engaging and understandable. On the other hand, *telling* guides the user through the data by providing a structured path for the analysis. This is especially useful when the data is complex and the user is not familiar with it. Guidance plays a crucial role, particularly for audiences with potentially low visual literacy. Presenting a visualization step-by-step has been shown to enhance understanding and foster greater engagement [9].

Design decisions, such as the structure of a story, influence how the audience perceives, understands, and remembers its content [88]. Simple, generic charts are often easily forgettable, whereas incorporating human-recognizable objects can enhance memorability [22, 118]. Interaction further engages users, increasing their ability to retain information [137].

Gamification is another approach that is frequently used to make educational materials more appealig and effective. Diakopoulos et al. [52] use the terminology *playable data* to describe the combination of visual analytics and game design. They concluded that while attention deficits were present when using gamification, it increased users motivation to interact with the data. Therefore, gamification can pose a promising approach in narrative visualization where increased guidance has the potential to reduce attention deficits. In a more recent comparative study with school children where visual literacy was taught with and without the use of gamification elements, no effect regarding the development of visualization skills was measured, but engagement and enjoyment of the experience were increased [91].

NARRATIVE INTENT. Storytelling amplifies the persuasive power of a visualization, which is particularly relevant in public outreach areas such as public health or environmental topics. However, this persuasive potential raises ethical concerns and the risk of introducing biases, as discussed in Section 2.3. The narrative intent is the reason for why the author deemed the story worth telling, why they want to persuade the audience, and what message they want to convey. There are two types of interplay between the intent and the data used for the story: (a) the data generated the intent (e.g., communicating an important insight from a specific data set) or (b) the intent was generated from scientific consensus and data in line with that consensus was curated to design the story. In both cases, questioning the narrative intent is important. E.g., in case of (a), it is important to verify that the data is trustworthy and representative, while for (b) it is important to verify that the scientific consensus really exists and that the data used is trustworthy and will not be falsified due to processing steps to better fit the intent. For these reasons, working with domain experts can ensure integrity.

NARRATIVE SCIENTIFIC VISUALIZATION FOR PUBLIC OUTREACH. While most research has been conducted within the domain of information visualization [240], this thesis shifts the focus to scientific visualization, specifically emphasizing applications in medical visualization. In these fields, the data is often complex, and the user is not familiar with the data and the domain.

Even though less research has been done in this field, there are some examples of narrative visualization in scientific visualization and medical visualization [145, 217, 263, 264]. For example, 3D visualizations have been used to tell stories about complex scientific phenomena, such as the formation of galaxies or oceanographic data.

The NASA uses storytelling in various forms for public outreach. Their *Scientific Visualization Studio* [172] is a website featuring several visualizations to explain complex scientific phenomena to the public. The visualizations are provided as videos. The videos are accompanied by a textual description below the video giving context to the video's topic. *NASA's Eyes* [171] offers different interactive 3D visualizations of the solar system and the universe. The user can explore the solar system and the universe in real-time, learning about the planets, stars, and galaxies. The visualizations are organized like an inverted martini glass structure, first offering an interactive view where the user can explore the 3D scene. In a menu the user can choose topics they want to explore in more detail. This then leads to a guided tour through the data. This is usually realized either by showing a scrollytelling story or the user can select an event (e.g., "Flooding in New Mexico") and the visualization is automatically rotated and zoomed in on the geographic location that is affected by the event. Additional visualizations may be shown and a textbox gives a brief explanation of the event.

Data visualization targeted at the broad public is also exhibited in museums. The Visualization Center C [245] in Norrköping offers exhibitions as well as presentations in their dome theater. The dome theater is a 360° projection room where the audience can experience the visualizations in an immersive way. The visualizations are accompanied by a narrator who guides the audience through the data. The visualizations in the exhibition are often interactive, allowing the audience to explore the data themselves. Interactive visualizations are a stark contrast to classical museum exhibitions where the audience is only a passive observer. The interactive visualizations allow the audience to engage with the data and to explore the data themselves. Thus, this modern approach fits well the definition of *showing* while more traditional exhibitions fall into the category of *telling*. Other museums featuring interactive data visualization exist around the world, e.g., the Exploratorium in San Francisco.

These examples illustrate that interactive narrative visualizations are particularly effective for presenting complex data, such as the arrangement of the solar system, in a manner that is both comprehensible and engaging. In these cases, narrative visualization serves as a tool for scientific outreach, aiming to connect with and captivate an interested audience.

2.3 THOUGHTS ON EFFECTS AND ETHICS IN DATA VISUALIZATION

The *narrative intent* describes the reason why a story is being told. It has the potential to be a significant area of tension between being neutral and objective (as often aspired to by journalists and data visualization professionals) and persuading the audience to change behavior (which is the goal of health communication and public health interventions). There are related research approaches that focus on how to frame the information presented in a data visualization, such as affective visualization, persuasive systems design, and visualization rhetoric.

AFFECTIVE VISUALIZATION. Affective visualization explores how visualizations can communicate and influence emotion in addition to conveying information as emotion can enhance engagement with data, making visualizations more memorable and impactful [121]. This approach shifts the focus from purely objective and analytical representation to a more subjective and evocative communication method. Research of affective visualization is structured in three dimensions: where (application fields), what (design tasks), and how (design methods). Regarding what topics are addressed, the two most popular topics are environmental sciences and ecology (e.g., global warming, pollution, and water waste), and social issues (e.g., public security, welfare, injustice, or inequality, deaths in wars). Several design tasks were identified including inform (raise awareness), engage (sparking interest in the visualization), and advocate (recommend ideas and call to action). Furthermore, the authors introduce four strategies for how to apply affective visualization techniques: sensation (e.g., color, shape, and imagery), narrative (e.g., personalization, antromorphism, and narrative structure), behavior (e.g., operate or create data visualizations), and context (e.g., situate the story in a concrete place or create inversive environments by using hardware such as VR technology). Based on these strategies, narrative visualization might be viewed as one subfield of affective visualization, as the design techniques for the narrative strategy in affective visualization are similar to the design techniques for narrative visualization:

- Personalized Setting: Users show stronger emotions when the story relates to themselves, thus using a familiar setting or a setting related to the user creates a stronger sense of proximity.
- Anthromorphism: Showing real persons that are behind the data, instead of showing abstract data.
- Narrative Structure and Wording/Phrasing: Both influence the flow of a narrative and evoke a certain mood.

PERSUASIVE SYSTEMS DESIGN. Persuasive systems design (PSD) is a framework for designing interactive systems that aim to reinforce or change user behavior and attitudes through non-coercive and non-deceptive means [180]. An important aspect of PSD is credibility (trustworthiness and expertise).

Incorporating PSD principles into narrative visualizations involves designing systems that:

- Adapt to User Contexts: Tailoring visualizations to the goals, motivations, and emotional states of users ensures a higher likelihood of engagement and persuasion.
- Leverage Incremental Strategies: Breaking down the story into digestible pieces allows users to progressively understand and connect with the data, fostering deeper insight and eventual behavior change.
- Focus on Usability and Aesthetics: Creating easy-to-use and visually appealing systems ensures that users can focus on the message without being hindered by technical or design barriers.

VISUALIZATION RHETORIC. Visualization rhetoric refers to design techniques used to frame visual data in a way that prioritizes certain interpretations or perspectives. It discusses how to make a deliberate decision on how to frame the information shown and introducing biases as a tool to emphasize certain aspects of the data. Hullman et al. [87] propose a taxonomy of information presentation manipulations used in narrative visualizations. They highlight the tension between "objective" charts guided by principles of transparency, and the layers of implied interpretation that can reshape the foundation of objectivity through the use of rhetorical strategies. Key rhetorical techniques include:

- Framing and Highlighting: Emphasizing certain data points or trends to direct user focus and influence interpretation.
- Layered Meaning: Using annotations, interactivity, and visual metaphors to convey nuanced stories and facilitate deeper understanding.
- Cultural and Contextual Adaptations: Considering socio-cultural norms and individual user contexts to enhance resonance and accessibility.

LEARNING OBJECTIVES. It is important to note that the strategies mentioned before framing the information derived from data in a certain way, however, their purpose is not to be deceptive. It may be criticized that they lack neutrality, however, Lee-Robbins and Adar [132] argue that there is no such thing as objectivity in communicative visualization. They state that more transparency can be reached by explicitly formulating the intent of a visualization as learning objectives.

In their taxonomy they introduce five affective verbs with ascending levels of engagement: observe (to raise awareness), position (self reflect or adopt beliefs), strengthen (support beliefs), compare (evaluate own behavior), and behave (call to action according to beliefs). Additionally, they introduce four affective nouns: appraisal (unemotional data fact), attitude (subjective view derived from data), values (deeply held belief), and value systems (beliefs held by a group of people). These affective verbs and nouns are combined to formulate concrete learning objectives, e.g., "The viewer will *perceive* the *appraisal*." (observe and appraisal) and "The viewer will *demonstrate* behaviors consistent with the *values*." (behave and values).

While their affective learning objectives target an emotional domain, they also introduced cognitive learning objectives with their own set of verbs and nouns [1]. The verbs formulate abilities users should acquire, divided into the following categories: remember, understand, apply, analyze evaluate, and create. Similarly, the nouns formulate learning goals in the categories factual knowledge, conceptual knowledge, procedural knowledge, and metacognitive knowledge. Applied to an example, such as a story showing the consequences of untreated high blood pressure, an affective learning objective could be "The viewer will *accept* the risks of high blood pressure." A cognitive learning objective for the same story could be "The viewer will *recall* risks of high blood pressure."

To transparently communicate the intent of each study's design, the cognitive and affective learning objectives are reported within their corresponding sections: Chapter 6, Chapter 7, Chapter 8, and Chapter 9.

2.4 EVALUATION

Amini et al. [7] provide a broad overview of evaluation methods for data-driven stories and storytelling tools. They differentiate between goals, criteria, methods, metrics, and constraints. These are also influenced by perspective, where evaluation can be performed from the standpoint if the author, publisher, tool builder, or audience. As this thesis focuses on authoring and audience response, the topic of evaluation will be discussed from the standpoints of the author and audience.

Author goals can be *"to communicate, inform, and educate"*, or *"to persuade to action or change behavior"*. Especially in the context of public health, the goal is often to persuade the audience to adopt healthier behaviors and empower them to make informed choices about their personal health.

Audience goals can be *"to learn something new"* or *"to be entertained"*. If the audience is the broad public, an entertaining factor is important to persuade the audience to engage with the story. In case of persons that are personally affected by the topic, e.g., patients, entertainment is likely not the main goal. Instead, they strive to get more information about the topic. Criteria for both, author and audience, include *comprehension*, *memorability*, and *engagement*. Metrics for measuring these criteria might be to analyze *page views*, *time spent on the page*, or the *number of shares on social media* as well as *comments* from the audience. Methods to analyze these metrics can be to *collect performance statistics*, *testing recall and recognition*, using *questionaires and interviews*, perform *qualitative studies*, or do *case studies*. However, constraints that can apply are limits in *human resources and expertise*, *time*, and *budget*.

EVALUATION PRACTICE. Overall, evaluating narrative visualizations differs from evaluating traditional visualizations. While traditional visualizations are often evaluated based on their effectiveness in conveying information, by means of "How fast can a task be solved?" or "How precise is a task solved?" these criteria are usually irrelevant for narrative visualization [194]. On the other hand, as Amini et al. [7] argued, spending more time in a narrative visualization is a good sign that the audience is engaged with the story.

Errey et al. [60] analyzed common evaluation practices in the field of narrative visualization, in a survey interviewing twelve practitioners. They found that most of the participants use the think aloud method, where users are asked to share their thoughts while engaging with the visualization, followed by interviews for end-user testing.

Methods like eye-tracking, click-tracking and surveys or questionaires are rarely used. The main reasons for not employing any end-user testing methods are lack of time, followed by insuffiecient budget. The survey revealed that inspection methods (e.g., letting a colleague or external expert inspect the work) are frequently used; however, they are mostly applied informally, without following established guidelines or criteria. Therefore, the authors suggest a structured evaluation framework to improve the evaluation and improvement of narrative visualizations. Their heuristic is categorized into three categories: composition (visual aesthetic, information distribution, and overall layout), user experience (user's interactions and experiences with the narrative visualization), and credibility and trust (data quality and believability).

METHODOLOGICAL APPROACH OF THIS THESIS. While these works highlight important aspects of narrative visualization that should be considered in the evaluation, they do not provide clear guidance on how to evaluate narrative visualizations, e.g., in terms of standardized questionnaires. Instead, they offer multiple options for evaluation, depending on the goals and constraints of the evaluation. This is in line with the approach of this thesis, which will consider multiple evaluation methods to assess different aspects of narrative medical visualizations. Both qualitative and quantitative evaluation methods are utilized. Due to limited resources (e.g., time allocated per participant), not all aspects highlighted by Errey et al. [60] were addressed in every evaluation. Instead, each study focused on a specific aspect of narrative visualization, rather than evaluating the story as a whole.

When performing qualitative evaluations, a key question is whether to use descriptive or inferential statistics to analyze the collected data. Since the goal of inferential statistics is to generalize findings from a smaller sample by testing for significance, it is only meaningful if the sample is well curated and considered representative [228]. However, this thesis often features case studies involving only two or three different versions of a story, since designing narrative visualizations is a highly time-intensive process, also highlighted as limitation by Amini et al. [7]. Based on a sample set of this size, generalization cannot be done. Consequently, descriptive statistics are applied when analyzing evaluation data.

2.5 STORYTELLING IN PUBLIC HEALTH AND HEALTH COMMUNICATION

Public health is defined as *"the science and art of preventing disease, prolonging life and promoting health through the organized efforts of society"* [260]. The World Health Organization (WHO) has a page dedicated to "Data Stories" [267] that provides visualizations and stories about health topics. These stories tend to be text-heavy, incorporating data visualizations and photographs. The accessibility of these stories is enhanced by extensive explanations and contextual information, accounting for the substantial amount of text. Additionally, some stories include guidance on interpreting charts, as illustrated in the WHO data story "Global Excess Deaths Associated with COVID-19 January 2020–December 2021" [266]. Overall, the WHO data stories appear to align more closely with the principle of "telling stories with data" rather than "data storytelling," as discussed in Section 2.2.

According to the WHO, health literacy is described as *"personal knowledge and competencies that accumulate through daily activities, social interactions and across generations"* [265] (p. 6). Health literacy is identified as one lead predictor for a person's health status in the USA [255]. The European Health Literacy Survey Questionnaire (HLS-EU-Q) was developed to conduct the first comparative study investigating European health literacy across countries, finding that social factors, such as old age or disadvantages in education, finances, or social status make it more likely for individuals to have lower health literacy [226]. The measurement of health literacy within the HLS-EU-Q is mainly based on self-reported challenges in finding, understanding, and using information when making decisions about personal health [225].

As a subfield of public health, health communication focuses on how to communicate health information (e.g., health risks, prevention, and treatment options) to the public thereby maximizing the effect the communicated information has on the audience [170]. Similar to research regarding affective visualization or persuasive systems design, the goal of health communication is not to manipulate or deceive, but to empower audiences with varying levels of health literacy to make informed decisions about their personal and public health [95]. Two important aspects in health message design are the *message content* ("what to communicate") and *message executions* ("how to communicate") [170]. Nan et al. [170] present three major types of *health message executions*: narratives, visuals, and emotions. Both, negative (i.e., fear and guilt) as well as positive emotions (i.e., hope and compassion) can enhance the effectiveness of a health message. How to design an effective public health message to motivate behavioral change depends on the personal orientation of each individual. Therefore, it is not possible to define a single most promising strategy. Nevertheless, Nan et al. [170] have identified three main factors for behavioral change in the context of health communication for COVID-19:

- *utilitarian outcomes* (i.e., personal health outcomes),
- *social outcomes* (i.e., social benefits/costs), and
- *value outcome* (i.e., personal/moral values).

Visuals are a promising approach to improving health communication and thereby countering infodemic conditions where the public is oversaturated with information [111]. Noar et al. [176] show that pictorial warnings (in contrast to text-only) on cigarette packs are more likely to increase intentions to quit smoking. Pictorial warnings are more engaging when they are factually accurate, i.e., when they include graphic images of diseased or damaged anatomy, real persons, and personal testimonials [181].

Dudley et al. [56] specifically focus on the utilization of narrative as a tool for science and health communication, examining its effectiveness, appeal, and best practices compared to traditional didactic communication methods in a scoping review of 253 articles from PubMed and Web of Science. Didactic communication aims to address information gaps by adhering strictly to objectivity. However, this approach often results in dull, unengaging content and overlooks the fact that most people do not base their decisions solely on evidence. On the other hand, narrative communication captures attention, evokes emotions, and enhances relatability, making it a powerful tool to address skepticism and misinformation. It is widely acknowledged that merely presenting additional information does not always lead to changes in people's perspectives [106]. Effective science and health communication must actively engage audiences and consider external factors influencing their beliefs and behaviors [166]. Furthermore, Dudley et al. [56] found that in most comparing studies, the audience preferred narrative over didactic communication. However, they argue that best practices combine both approaches. The effectiveness of a narrative communication depends on the audience and the quality of the narrative. Narratives are especially effective for individuals with low levels of education and health literacy. Regarding design choices, the use of a relatable protagonist and a first-person point of view are highlighted.

These results showing the benefits of both, narrative communication and the use of visuals, have lead to the development of graphic medicine [110]. Graphic medicine uses visual narratives for the purpose of health communication, with the most important medium at the time being comics. For example, practitioners employ comics to improve patient education by increasing patient comprehension and lowering periprocedural anxiety [24]. This research shows that narrative visualizations can be a powerful tool for health communication, as they can engage audiences, evoke emotions, and enhance understanding and retention of health information.

2.6 DISTINCTIONS OF NARRATIVE MEDICAL VISUALIZATION

Narrative medical visualization lies at the intersection of narrative visualization and health communication, particularly within the subfield of graphic medicine. A key motivation for narrative medical visualization is that it empowers individuals, including non-artists, to engage in science communication by utilizing real data [72, 152]. While following fundamental principles of narrative visualization, domain-specific challenges make narrative medical visualization distinct from general narrative visualization. Garrison et al. [72] summarize current approaches in narrative medical visualization stating that there is still a large gap between research and practice. Based on their analysis, the following aspects relevant to this thesis are highlighted:

EXPERTS AND AUDIENCE. In narrative medical visualization, it is essential for the author to cooperate with a physician or medical expert with expertise in the relevant topic. Equally important is involving members of the target audience or individuals who can empathize with their needs and interests, although this step is often overlooked during story creation.

NARRATIVE INTENT AND MESSAGE. A common intent for narrative medical visualizations is to empower audiences to take positive health actions by communicating preventable risk factors for diseases. Often this is done by showing a character developing a disease due to not avoiding preventable risk factors, such as a smoker that develops lung cancer. Human elements are vital in medical storytelling, especially when addressing patients or their loved ones who may feel scared or overwhelmed by a medical condition. Empathy at the start of a narrative helps ensure the message is impactful and considerate, avoiding unintended emotional distress. By focusing on relatable and emotionally resonant storytelling, medical narratives can drive behavior change and foster understanding. Another challenge is that medical topics are often complex and challenging for laypersons to comprehend. To address this, metaphors can be an effective tool for simplifying and conveying messages in a more understandable way [195]. Clinicians frequently employ metaphors in doctor-patient communication. For instance, to illustrate the risks associated with vortices in blood flow, Kleinau et al. [113] use the metaphor of a tornado.

DATA. Clinicians commonly use medical imaging data, such as MRI and CT scans, for diagnostics and treatment planning. These data, when appropriately processed, can also serve as the foundation for medical narratives aimed at broader audiences. Effective processing involves highlighting diseased or abnormal structures with color and accompanying them with textual explanations. Animated time-series data can further enhance storytelling by visualizing the progression or treatment of a disease, such as illustrating tumor shrinkage over time during successful therapy. Other types of medical information, including statistics on disease prevalence, burden, and age group impact, also provide useful information.

VISUALIZATION. Various visualization techniques exist and their suitability depends on the data, narrative intent, and target audience. While statistical data can be visualized using diagrams, e.g., bar charts, line charts, or scatter plots, medical imaging data is a totally different kind of data where often volumetric visualizations are useful. However, volumetric visualizations require a certain level of visual and technological literacy. While interactive visuals can enhance engagement and memorability, they may be overwhelming for some audiences. Accompanying visuals with text labels and descriptions is essential for clarity, particularly for viewers unfamiliar with anatomy and physiology. Additionally, icon-based visualizations, such as isotypes, can complement imaging data, making medical stories more accessible and impactful. Deciding on the best visual representation remains an active area of research. As data complexity increases, simplifying and guiding the audience through medical narratives becomes essential. This can be achieved through effective interaction and navigation techniques in visualizations, especially in 3D environments. For instance, preset views, limited rotation capabilities, and preconfigured parameters help users explore complex models, like a liver with embedded tumor tissue, without becoming disoriented [72]. Managing occlusion is a critical challenge in 3D environments. Techniques such as multiple viewports, virtual x-ray tools, and volumetric probes enable users to visualize hidden structures effectively [72]. Virtual x-ray tools are particularly impactful for narrative medical visualization, allowing users to "see through" structures, offering a clinician's perspective, and identifying hidden elements like tumors. Additional techniques like ghosted views and cutaways provide smart visibility for concealed structures, enhancing clarity and engagement in complex visual narratives [123].

ANIMATION AND INTERACTION. When working with 3D visualizations, such as imaging data from CT or MRI scans, the challenge of occlusion often arises. To address this issue, variations of the 3D visualization are generated from different perspectives. Two commonly used techniques for creating these perspectives are animation and interaction, both of which can also effectively depict changes over time.

Animations provide the author with full control over what is presented, ensuring clarity and reducing the risk of user error, such as difficulties stemming from a limited understanding of interaction techniques. However, animations restrict users from exploring the data independently, relegating them to the passive role of an observer. To enhance effectiveness, animations should be concise, lasting no more than one minute, and accompanied by narration to improve clarity and minimize cognitive load [193]. For broader audiences, precision can be less critical than accessibility. Simplified, plausible motions that convey overarching concepts, combined with high contrast and strongly smoothed anatomical structures, improve comprehension and engagement [193]. In contrast, interaction empowers users to explore the data themselves becoming an active participant of the story. However, it is essential to select interaction methods that suit the target audience and output medium and to clearly communicate how to use them. Examples of narrative medical visualizations include straightforward scene-by-scene click navigation as well as more advanced interactions, such as manipulating diagrams and rotating 3D medical models along one or more axes [152, 248].

While being potentially more engaging, interaction may not be the most suitable approach for patient audiences who might not be interested in exploration of the data due to mental stress after receiving a diagnosis.

STRUCTURE AND GENRE. Narrative medical visualization has made progress in developing foundational structures for crafting data-driven medical stories, particularly focused on disease-related narratives for general audiences. Meuschke et al. [152]

analyzed health-related online content from hospitals and health-related websites to create a structured narrative framework:

1. Disease Definition: Introduces the disease with statistics to engage the audience.
2. Anatomical Overview: Provides orientation to relevant anatomical structures.
3. Symptoms: Introduces tension by describing common disease symptoms.
4. Diagnosis: Climaxes the story with the diagnostic process.
5. Treatment Options: Outlines available treatments, highlighting benefits and risks.
6. Prognosis: Eases tension with a discussion of long-term outcomes, e.g., five-year survival rates.
7. Preventative Measures: Concludes with actionable prevention tips, leaving the audience empowered.

2

While linear, the structure can branch out for detailed exploration (e.g., specific surgical procedures) or allow reader-driven paths, such as choosing treatment options. Steps can be omitted if irrelevant to the narrative. Other data-driven medical stories are structured based on concepts from theater and media, i.e., Freytag's Pyramid [113] and Campbell's Hero's Journey [156]. Structures guide medical storytelling, ensuring clarity, emotional engagement, and a logical progression.

PERSONALIZATION. Personalization in narrative medical visualization enhances relatability and empathy by tailoring narrators and story characters to engage diverse audiences. Characters may be depicted as abstract and androgynous figures, leaving attributes like gender, age, and body shape open to interpretation. This approach broadens audience relatability across diverse demographics. Patients can serve as the main narrators, with their conflict centered on a disease or condition affecting their quality of life. Clinicians can also serve as protagonists, highlighting the challenges they face, such as overwork, stress, and public health crises (e.g., during COVID-19). These narratives intent to garner sympathy for healthcare workers and encourage public cooperation in advancing health initiatives like vaccination. While personalization fosters positive emotional empathy and promotes public health actions, it can raise concerns about manipulation or bias. Incorporating alternative viewpoints or counterfactuals can create balanced, credible stories that feel less biased and more informative. Personalization strategies aim to humanize medical narratives and build emotional connections while maintaining ethical storytelling practices. So et al. [223] incorporated storytelling as a persuasive element in an interactive visualization about various medical conditions, highlighting its effectiveness in influencing participant's opinions within a user study. They communicate biological, psychological, and socio-environmental aspects of diseases with information gathered from social media posts. Budich et al. [27]

focus on the design of data-driven characters derived from cohort data. This way, the characters are credible, but the data are anonymized to ensure patient privacy by using fictional names and images.

CREDIBILITY. Credibility is crucial in narrative medical visualization, particularly as audiences are increasingly critical of information sources, especially when it concerns personal health. Key considerations include transparency (e.g., potential economic motivations within the pharmaceutical, medical technology, and healthcare sectors), balanced information (avoid oversimplifying or presenting only one side of medical interventions), reliable data (use high-quality, trustworthy data sources, and indicate whether the data has been altered, who modified it, and for what purpose), and expert endorsement (involving a reputable medical professional, such as a hospital director).

FOCUS AND SCOPE. Narrative visualization aims to adhere to the principle of "data storytelling" (as outlined in Section 2.2). However, many narrative medical visualizations referenced thus far align more closely with "telling stories with data." This distinction arises because medical visualizations frequently incorporate extensive contextual information that lacks a corresponding data source, embedding the data within a broader narrative framework. Narrative medical visualizations that more closely follow "data storytelling" principles often focus on narrower topics, such as revealing an injury [263]. While this approach allows for deeper exploration of specific data sets, it may omit broader statistical information, such as the role of risk factors, which could be essential to the narrative intent and key message of the story.

This thesis emphasizes disease stories, which are narrative medical visualizations designed to communicate various aspects of diseases to lay audiences. As proposed by Meuschke et al. [152], disease stories are particularly effective for communicating preventable risk factors and promoting healthier lifestyles by enhancing health literacy. This aligns with the objectives of many public health narratives, making disease stories a valuable tool for engaging lay audiences and fostering awareness. Disease stories often include imaging data illustrating the impact of a disease on the body [152]. However, the story typically extends beyond a single data set. For example, it might include anatomical visualizations based on patient imaging data, cohort statistics to depict risk factors, and additional contextual information to link and explain these elements cohesively. This approach reflects the inclusion of multiple facets of a disease, as suggested in the seven-stage story template by Meuschke et al. [152]. Therefore, in disease stories, the data itself is not the primary setting where the story unfolds or where insights originate. Instead, the focus often shifts away from the data to center on the patient and the disease. The data acts as a supporting tool for storytelling rather than the central subject of the narrative.

3 MEDICAL BACKGROUND

This thesis presents narrative medical visualizations focusing on various diseases. This chapter explains the aspects of these diseases that are most relevant for understanding the design process behind these visualizations and stories.

The content of this chapter combines and extends insights from the following publications: [155–157, 159, 160, 162, 163].

3.1 CEREBRAL SMALL VESSEL DISEASE

Cerebral Small Vessel Disease (CSVD) describes a multitude of pathologies connected to the cerebral small vessels, such as small arteries, arterioles, venules, and capillaries [136]. These cerebral small vessels have a diameter of about 50-500 µm, making them invisible in an MRI due to insufficient spatial resolution [148]. Nevertheless, MRI is the modality that is usually used to detect CSVD, as it depicts various CSVD phenotypes of which white matter hyperintensities (WMH) and cerebral microbleeds (CMB) are of special interest to clinicians [148]. They both tend to be asymptomatic and coexist, being associated with cognitive decline and vascular dementia. [148]. WMHs are detected in FLAIR images, where they appear as hyperintense regions (significantly brighter than surrounding tissue), see Figure 4. They are caused and increased by hypertension and age and are usually located in the brainstem white matter and close to the ventricles [148]. As WMHs are rather large regions, their volume is often used to assess CSVD severity [148]. CMBs are detected in T2*-weighted images, where they appear as hypointense regions (significantly darker than surrounded tissue), see Figure 4. They are caused by the extravasation of blood from small vessels into the brain tissue. In contrast to WMHs, CMBs are small, usually less than 5 mm in diameter, and are often located in the basal ganglia, thalamus, and brainstem [148].

CSVD is a common cause of stroke, depression, cognitive decline, and dementia in the elderly [35]. It is hard to detect in early stages as the symptoms are subtle and ambiguous among the elderly, e.g., cognitive decline, gait disturbances, and urinary incontinence [136]. CSVD occurs 610 times more often than stroke [35]. The prevalence of CSVD increases with age, with a WMH prevalence of 5% in people aged 50 years to close to 100% in people aged 90 years and a CMB prevalence of 6.5% for people aged 45 to 50 years and about 36% for people aged 80 to 89 years [35].

CSVD has two major subtypes, namely hypertensive arteriopathy (HA) and cerebral amyloid angiopathy (CAA) [92]. Depending on the subtype, symptoms and treatment varies greatly. However, there is no common treatment method for CSVD. Most impor-

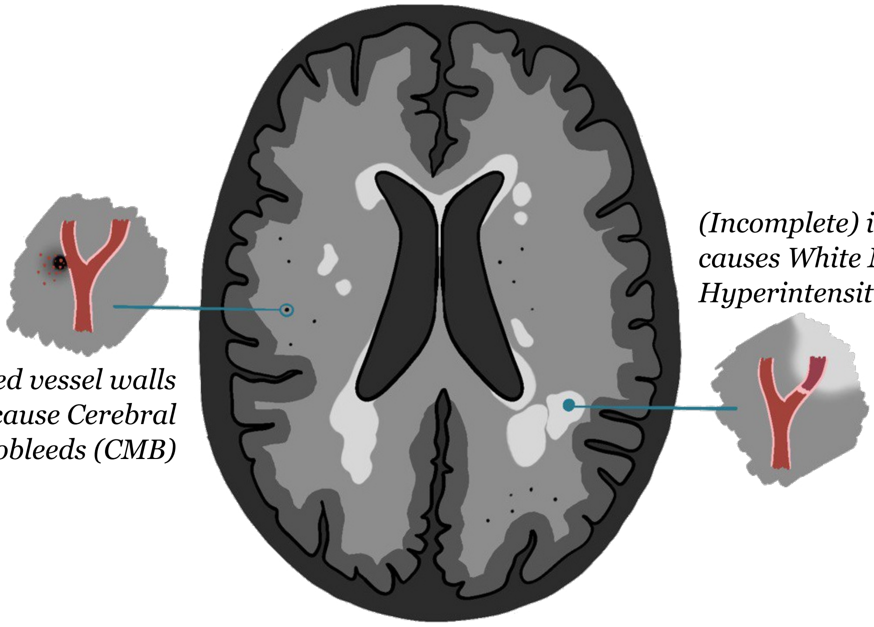


Figure 4: Schematic depiction of the two CSVD phenotypes CMBs and WMHs.

tantly, vascular risk factors need to be addressed and treatment of CSVD is primarily based on primary and secondary prevention [35]. Thereby, the most important aspect is maintaining a healthy blood pressure and treating hypertension adequately. Due to its high prevalence and the severe consequences, it is important to detect CSVD in early stages where secondary prevention provides best results, making it an important topic for health communication.

Cerebral Small Vessel Disease (CSVD)

Prevalence: 5% of the people aged 50 years to almost 100% of people older than 90 years
Risks and Consequences: Causes cognitive decline and vascular dementia, e.g., Alzheimer's Disease
Treatment: Lifestyle adaptions, treatment of hypertension
Motivation for Health Communication: Prevention of avoidable risk factors, early detection

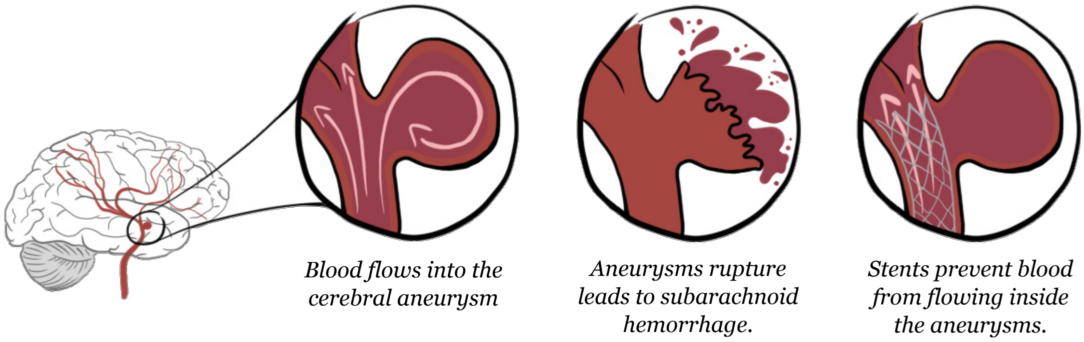


Figure 5: Schematic depiction of a cerebral aneurysm which poses a risk of rupture and can be treated using a stent.

3

3.2 CEREBRAL ANEURYSMS

Cerebral aneurysms are pathological local dilations of brain blood vessels by at least 1.5% of the original vessel diameter [177]. The dilations are restricted locally, and most often to one side of the vessel wall, forming saccular aneurysms, see Figure 5. Therefore, saccular aneurysms are the most frequent type of cerebral aneurysms and also the most frequent cause for aneurysmal subarachnoid hemorrhage. This describes bleeding between the brain and its enveloping membrane, causing increased pressure in the skull [67]. About 3 % of the adult general population is affected by saccular unruptured intracranial aneurysms with a mean age of 50 years [90, 100].

They rarely cause symptoms and are usually detected as incidental finding when imaging is performed for another reason but are at risk of rupturing and causing life-threatening intracranial hemorrhage. Morphological changes, such as growth, are associated with an increased risk of rupture [167]. Because treatment of aneurysms carries risks, many aneurysms that are considered stable (no enlargement >0.5 mm for at least 9 months) are not treated. However, it is important to monitor aneurysms to assess the risk of rupture, as even stable aneurysms can become unstable [212].

The overall risk of rupture in the western population is estimated to be around 1.2% [67]. The mortality after rupture is around 30 to 40% and almost half of survivors are left disabled with the median age of the patients being 60 [67].

A minimally-invasive technique to treat aneurysms discussed in this thesis are stents. Stents are tubes formed from mesh wire that block the entrace to the aneurysm and divert the flow so that it follows the course of the original healthy vessel, see Figure 5. However, the treatment of aneurysms can have complications, e.g., the procedure can cause an aneurysm to rupture, which might not have happened otherwise, leading to mortality [67]. Therefore, deciding about aneurysm treatment, needs to be done on an individual level, also considering the patients concerns, making the topic of cerebral aneurysms an important one for health communication.

Cerebral Aneurysms

Prevalence: 3% of people with a mean age of 50
Risks and Consequences: Rupture that can lead to cognitive impairment and mortality
Treatment: One minimally invasive option are stents
Motivation for Health Communication: Individual treatment decisions

3.3 BICUSPID AORTIC VALVE

Cardiovascular disease (CVD) is a leading cause of death worldwide, with these numbers only increasing each passing year [241]. Early and accurate assessment of individual patient risk and disease severity is crucial for providing effective treatment. CVDs lead to deformation of the vessel geometry, which can cause quantitative and qualitative changes in blood flow [108]. To determine the severity of CVDs, physicians examine blood vessel morphology and internal blood flow [65, 73]. One image modality used for this purpose is four-dimensional phase-contrast magnetic resonance imaging (4D PC-MRI), which allows for non-invasive measurement of patient-specific blood flow. One specific type of CVD that 4D PC-MRI is particularly useful for evaluating is bicuspid aortic valve (BAV) disease, in which two of the three leaflets of the aortic valve are fused. With a prevalence of 1.3% in the general population, BAV is the most common congenital heart defect among adults [105]. The resulting change in blood flow can lead to aneurysm formation [85]. Despite the risk of rupture, aortic aneurysms also increase risk of impaired heart function and aortic wall problems in patients with BAV, who require regular monitoring [25, 57].

BAV can be associated with comorbidities, particularly aortic stenosis, a condition in which the aortic valve does not open fully, restricting blood flowing from the heart to the rest of the body [105]. BAV can have different subtypes, however 90% of BAV patients suffer from a dilatation of the tubular ascending aorta often accompanied by a dilation of the aortic root and this type is also associated with patients above 50 years at diagnosis [105]. To treat BAV, surgical and transcatheter interventions are available.

TAVI is a minimal invasive treatment approach which makes it especially efficacious in elderly patients [105]. During TAVI a catheter carries an artificial valve through the aorta, see Figure 6. By inflating a ballon, the artificial valve is located in place of the old malfunctioning valve. Afterwards the catheter is removed. TAVI is associated with aortopathy, a condition where a weakened aortic wall or a tear or rupture in the aorta restricts the flow of oxygen-rich blood in the body [105]. As TAVI carries potential risk for aortic injury, it is an important topic for health communication, especially for patients with BAV or other CVDs, so that the patients are able to give informed

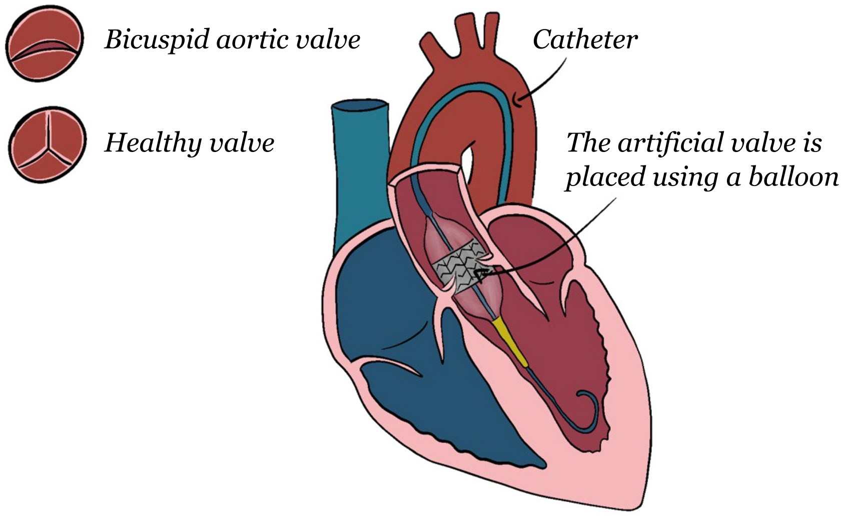


Figure 6: Schematic depiction of the malfunctioning aortic valve and the catheter used to insert a new valve during a transcatheter aortic valve implantation (TAVI).

consent to the treatment method. As the TAVI is a complex procedure, communicating it visually can help patients understand what happens during the procedure.

> **Bicuspid Aortic Valve**
>
> **Prevalence:** 1.3% in the general population
> **Risks and Consequences:** Aortic aneurysms, increase risk of impaired heart function, and aortic wall problems
> **Treatment:** TAVI
> **Motivation for health communication:** Individual treatment decisions

3.4 NON-ALCOHOLIC FATTY LIVER DISEASE

Non-alcoholic fatty liver disease (NAFLD) is a disease where more than five percent of the liver cells contain fat [107]. In contrast to alcoholic fatty liver disease the increase in liver fat is not caused by abusive alcohol consumption. 25% of the world's adult population are affected, making it the most frequent chronic liver disease as well as the fastest growing cause of liver mortality [107]. The median of people affected by NAFLD is between 50 and 55 years.

NAFLD is associated with obesity (but can also occur in individuals with normal weight), type 2 diabetes, and age, and can cause long-term complications showed in Figure 7, including liver fibrosis, liver cirrhosis and liver cancer [107]. Therefore, early detection is crucial. As routine screening options for the whole population are not available, communicating risk factors, associated prevention measures, and patient counseling for those at high risk is important [107].

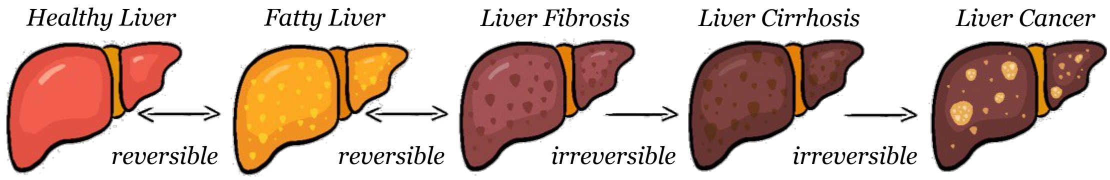


Figure 7: Stages of liver damage and their reversibility.

Direct pharmacological treatment of NAFLD cannot be performed, treatment is reduced to manage comorbidities [107]. Preventive measures are based on the avoidance of most common risk factors such as obesity and diabetes type 2. However, NAFLD is reversible if a healthier lifestyle is adopted [107].

Non-Alcoholic Fatty Liver Disease

Prevalence: 25% with median age 50 to 55
Risks and Consequences: Liver cirrhosis and liver cancer
Treatment: Only managing comorbidities
Motivation for health communication: Promote prevention, reversibility of the disease

3.5 PRIMARY LIVER CANCER

In the case of tumors, a distinction is made between benign and malignant tumors. Only the letter is cancerous and can spread to other parts of the body. A tumor can be primary, meaning they originate in the affected organ, or secondary, meaning they spread to the organ from other parts of the body. This work will focus on primary liver cancer (PLC). According to Criss and Makary [45] PLC *"is the leading cause of cancer-related deaths worldwide."* Among the risk factors they mention chronic viral hepatitis B and C infections, alcohol abuse, non-alcoholic fatty liver disease, and obesity. About 71% of all incidents of liver cancers in the USA in 2014 were associated with potentially modifiable risk factors [47]. Individuals of age 64 to 69 years showed the highest PLC incidence and deaths worldwide [213]. However, the different subtypes of PLC as well as socio-economic factors make a more differentiated analysis of incidents senseful when diving deeper in the topic of PLC. For example, PLC also shows a non-neglectable amount of children and adolescents (< 20 years) are affected [213]. About 841,000 new cases of PLC are diagnosed globally every year [47].

Tumors in the liver rarely cause symptoms. In advanced stages, liver cancer can cause symptoms like weight loss, a loss of appetite, and pain [134]. Primary liver cancer is highly aggressive and has a 5-year survival rate of 6.5% [45]. Different curative or palliative treatment options exist for primary liver cancer [45]. Therapies such as ablation can cure early-stage tumors or bridge to transplantation or resection. Embolization (to shut down the blood supply to the tumor) can be used palliative or to improve survival for early-to-intermediate-stage tumors and unresectable tumors. Whether a tumor is resectable depends on various aspects, such as the location of the tumor but also the condition of a patient. If the general condition of a patient is too weak, surgery but also chemotherapy might not be an option. Therefore, tumor diseases are an important topic for health communication, as a decision between different treatment or palliative options have to be made. This decision is highly individual as it includes the physical and mental condition of the patient, but also type, location and number of tumors.

Primary Liver Cancer

Prevalence: 841,000 new cases per year with a peak in the age group 64 to 69 years
Risks and Consequences: Potential cause of mortality
Treatment: Chemotherapy, resection or ablation (for curative and palliative treatment)
Motivation for health communication: Promote prevention and support treatment decision

3.6 CONCLUSION

The research presented in this thesis features various medical application examples used for different case studies that investigate design decisions for disease stories. For each disease there is a motivation for why it needs to be communicated to a lay audience. The audiences are usually either the public (e.g., in case preventive behavior or early detection of symptoms are promoted) or patients (e.g., when explaining a disease and possible treatments). For each disease, a domain expert provided feedback regarding the design of disease stories and case studies, ensuring the application examples are both accurate and relevant.

3

Part II

CREATION OF DISEASE STORIES

The main body of the thesis centers on the creation of disease stories in cooperation with domain experts. It introduces an abstracted design process and details the tools and data used in the practical implementation. A series of studies explore key components of disease story design – content, character, structure, and target audience – through dedicated research questions for each.

3

4 DEMONSTRATION OF THE DISEASE STORY DESIGN PROCESS

The disease story design process is introduced as a framework that illustrates the interplay of various components involved in creating a disease story. It explains the structure of the design process and its components. The process is demonstrated using the example of creating a CSVD story in Section 4.1.2. This chapter is partially based on the following publications: [158, 162, 163].

4.1 DISEASE STORY DESIGN PROCESS

The design of disease stories is formalized as a general design process based on experiences gained from developing several different disease stories. In addition to the CSVD story [163], whose development was led by the author of this thesis, this section draws insights from narratives about PLC and BAV, primarily designed by Meuschke et al. [152] and Kleinau et al. [113].

4.1.1 *Stages of the Disease Story Design Process*

The design process is divided into three stages:

- THE KNOWLEDGE STAGE defines the topic, and involves frequent discussions between the author and domain experts to shape the story design.
- THE NARRATIVE STAGE focuses on the ingredients that make up the disease story and their interplay.
- THE EFFECT STAGE concentrates on understanding the target audience and the intended effect the story aims to have on them.

These stages are not meant to be traversed sequentially as multiple stages may be active simultaneously during the design of a disease story. The stages were designed to align with concepts from health communication and communication models, see Figure 8.

HEALTH MESSAGE DESIGN. The foundations of health message design were explained in Section 2.5. To recap its two main factors, health message design focuses on *message content* and *message execution*, which describe the "what" and "how" of health communication. Message execution can be achieved through narratives, visuals, or emotional appeals [170]. It describes the delivery of the health message and aligns with

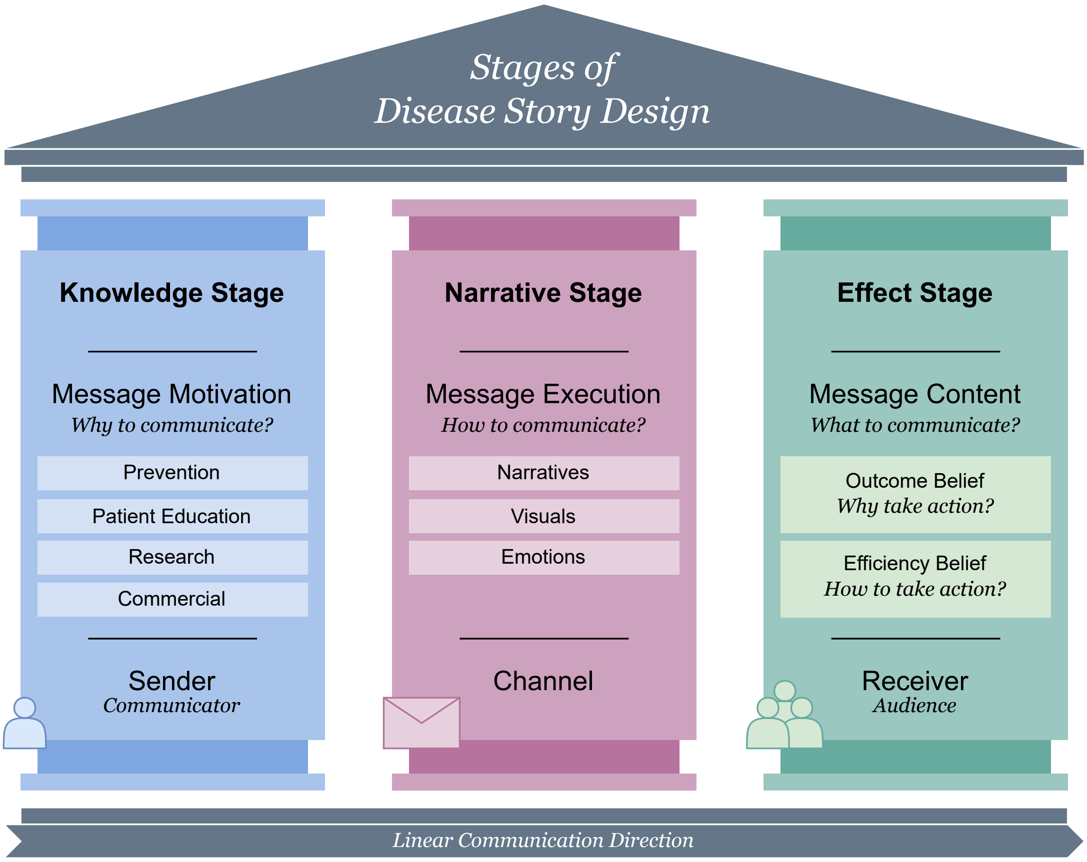


Figure 8: Relations of the stages of disease story design to the aspects of health message design from Nan et al. [170] expanded by message motivation and linear communication models. Image adapted from Mittenentzwei et al. [162], available under a CC BY 4.0 license.

the narrative stage of disease story design, where the core components of the story are structured. Message content is further divided into *outcome beliefs* ("Why should the audience take action?") and *efficacy beliefs* ("How can the audience take action?") [170]. Thus, message content emphasizes the *intended effect* on the audience, similar to the effect stage of disease story design. However, this approach to health message design does not address the motivation behind "why" health communication is important for a specific topic. Therefore, this thesis introduces a third dimension, called *message motivation*, which corresponds to the knowledge stage of the disease story process.

COMMUNICATION MODELS. The three stages of the design process also show parallels to linear communication models. For example, the Shannon-Weaver communication model is a simplistic linear communication model that describes how a message is sent from a source to a receiver via a channel [69]. The knowledge stage can be seen as the sender, while the effect stage represents the receiver. The narrative stage serves as the channel through which the message is transmitted. Another basic approach is Lewis' model of signaling, which is considered a fundamental framework for modern communication [190]. Originally, Lewis used the terms "communicator" and "audience" instead of "sender" and "receiver," a terminology closely related to narrative visualization. Key prerequisites of Lewis' model include common interest, common knowledge, and rational choice by both the communicator and the audience [78]. Similarly, disease stories are only effective when both the author and the audience are interested in achieving a common goal. In this model, the communicator possesses knowledge about a current state but cannot act – similar to a physician who has expertise in a medical topic. Conversely, the audience can act but only based on the information received – similar to patients who can act by adopting a healthier lifestyle. As in narrative visualization, this is a one-way communication model, as the receiver or audience does not have the ability to respond.

4.1.2 *Building a Disease Story*

The three stages each contain distinct components, as shown in Figure 9. The knowledge stage illustrates the collaboration between the story author and the domain expert. The narrative stage outlines the key elements used to craft a disease story. These elements are based on the framework presented by Brent Dykes [58], who argues that a story is made up of *conflict*, *content*, *character*, and *structure*. A similar approach is outlined by Fog et al. [63], who describe the story ingredients as *message*, *conflict*, *characters*, and *plot*. The *plot* is closely related to *structure*, as it focuses on defining the sequence of events within the story. Bach et al. [16] explain that narrative visualization patterns are applicable when there is a *story*, a *target audience*, and an intended effect. The latter two elements are part of the effect stage, where the emphasis is on understanding the target audience and effectively conveying a message that aligns with the narrative intent.

4.1.3 *Knowledge Stage: Why to Communicate*

A disease story is typically driven by *clinical practice dilemmas*, which explain *why* a particular topic needs to be communicated. The main components in the knowledge stage are the *domain expert*, the *author*, and *additional resources*.

DOMAIN EXPERT. The domain expert presents a *clinical practice dilemma* that can potentially be addressed through a disease story, e.g., communicate risks caused by certain lifestyles, or medical procedures to an audience with low health literacy. This thesis is based on collaboration with several clinicians who have extensive practical experience. Their dilemmas and motivations revolve around four key areas: (1) pro-

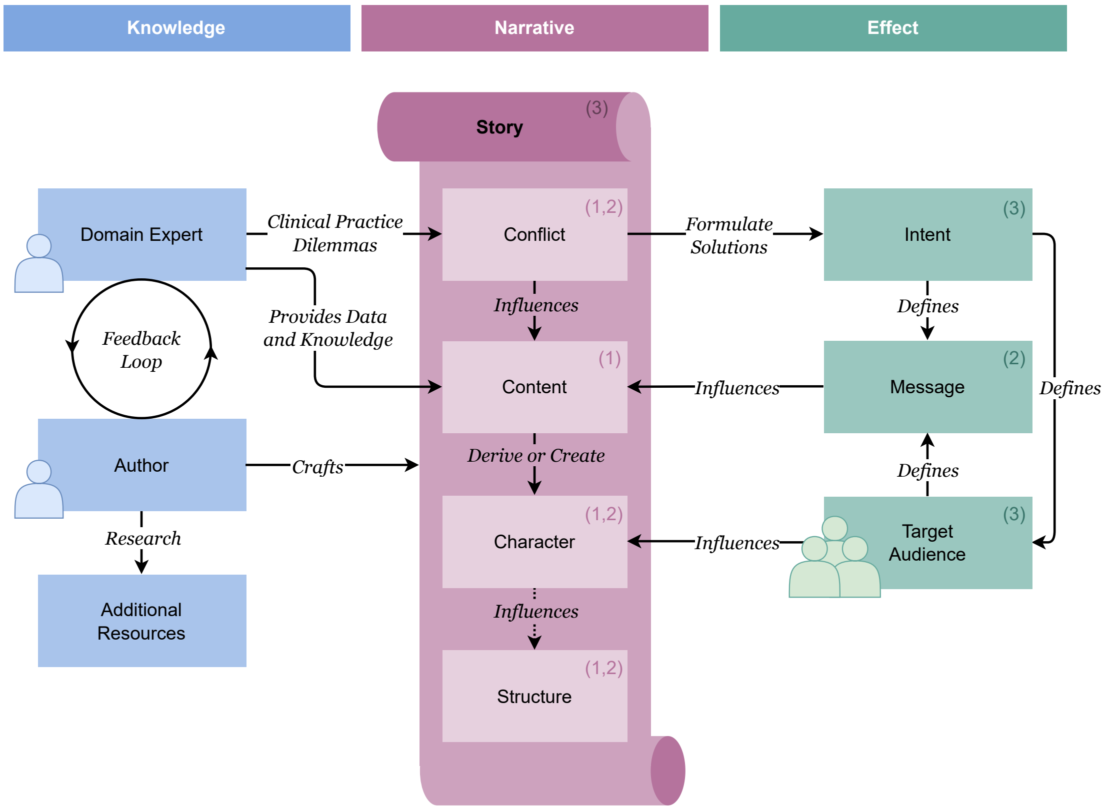


Figure 9: The disease story design process based on the theoretical foundations of (1) Brent Dykes [58], (2) Fog et al. [63], and (3) Bach et al. [16]. Dashed lines depict optional dependencies. Image adapted from Mittenentzwei et al. [162], available under a CC BY 4.0 license.

moting preventive behaviors, (2) empowering patients to provide informed consent through communication methods tailored for low health literacy, (3) encouraging participation in medical research studies, and (4) promoting specialized treatments and facilities. Section 5.1 elaborates on these.

The domain expert provides both data and domain knowledge, with the data varying from tabular records, such as patient cohort data, to imaging data from individual patients. Practicing physicians can also offer valuable insights into how to effectively communicate medical topics to lay audiences, drawing from their experience in patient consultations. The domain expert is consulted frequently throughout the design process to ensure that the disease story aligns with their clinical practice dilemmas and adheres to medical consensus. As a treating physician, a domain expert may also provide insights into the communication of health information to a lay audience, e.g., through the use of specific metaphors.

AUTHOR. The author, in this thesis a researcher with expertise in medical visualization, creates visualizations that highlight key insights from the data. The focus of data visualization in disease stories is not to analyze the data but to *present* it. The data is carefully curated and presented to align with both the narrative's *intent* and the established medical consensus. Authoring often involves multiple sub-roles, such as copywriting, design, and data visualization. However, a single person can fulfill several of these functions, e.g., creating both data visualizations and text. If the author does not have experience in, e.g., design or didactics, the quality of the final story will be improved if the author team includes people with experience in these areas.

EXTERNAL RESOURCES. The materials provided by domain experts can be complemented with additional resources. For instance, the author may consult scientific literature or reliable online sources, e.g., websites of reputable institutions, such as the World Health Organization [257], the National Health Service [173], or hospitals. External resources are usually needed to incorporate insights from larger cohorts, such as national or global prevalence studies.

4.1.4 *Narrative Stage: How to Communicate*

The design decisions made in the narrative stage can be linked to message execution from health message design, which defines "how" the topic is communicated to the target audience. Narrative medical visualization inherently uses two of the three types defined by Nan et al. [170], namely visuals and narratives. The main components of this stage are the story ingredients defined by Brent Dykes [58]: conflict, content, character, and structure.

CONFLICT. The conflict is directly derived from clinical practice problems, which are personalized by projecting them onto individual patient characters [113, 152, 156]. The goal of projecting these clinical practice dilemmas at an individual level is to make the story more relatable. However, this approach carries the risk of focusing only on selected aspects of a disease (those that apply to the chosen character), potentially neglecting the broader context. Therefore, it is important to carefully choose which aspects of a disease should be included in a disease story.

CONTENT. The content is based on the data and knowledge provided by the domain expert, as well as by the additional resources used by the author. The content is chosen so that it depicts the conflict of the disease story, conveying the intended message to the audience. Design and data visualization play an important role in presenting the selected content to the target audience.

CHARACTER. While characters do not necessarily need to be human (see Section 2.2.2), this work focuses on human characters to enhance personalization through the concept of anthropomorphism. Anthropomorphism refers to the process of portraying real people behind the data to evoke stronger emotions in users due to increased relatability [121]. Using fictional patients (even if derived from real data) helps ensure patient anonymity. There are various character roles, such as protagonists, antagonists, and side characters [156], however, the number of characters a disease story should have is not defined.

STRUCTURE. Once the story ingredients are defined, they need to be structured. This is typically achieved using story structures, e.g., the martini glass structure, Freytag's Pyramid, or Campbell's Hero's Journey. The choice of structure is left to the author, as no established guidelines exist for selecting the most suitable structure.

4.1.5 *Effect Stage: What to Communicate*

This stage focuses on the external effect of the story and is therefore linked to the message content from health message design. It defines "what" is communicated focussing on the components intent, message and target audience.

INTENT. The intent of the author and domain expert is to solve a *clinical practice dilemma*. The intent of a narrative exists within a field of tension between maintaining neutrality and objectivity – often emphasized by journalists and data visualization professionals – and persuading the audience toward health-improving behavioral changes, which is the goal of health communication, public health, and the clinical collaboration partners. In the context of narrative medical visualization, formulating the intent as a learning objective, as proposed by Lee-Robbins and Adar [132], is a promising approach to maintaining transparency regarding the author's intentions.

MESSAGE. The content is designed to convey a specific key message, which is derived from the story's intent. According to health message design, the audience's attitude towards the key message is influenced by both *outcome beliefs* and *efficacy beliefs* [170].

Outcome beliefs refer to whether the audience believes that adopting the recommended behavior will result in a positive outcome. A common technique used to target outcome beliefs is *arguments-from-consequences*, where the consequences of following or not following the recommendations are presented [179]. This can be approached through either a preventive focus (highlighting the negative consequences of inaction, e.g., an unhealthy lifestyle leading to disease) or a promotion focus (emphasizing the positive outcomes of taking action, e.g., treatment improving a condition)[170]. The effectiveness of these approaches varies between individuals, so there is no universal recommendation for one over the other. The same applies to the time frame of the message [170]. A message can focus on short-term outcomes (e.g., immediate results of surgery) or long-term outcomes (e.g., organ damage from years of untreated hypertension), again their effectiveness varies on an individual level.

Efficacy beliefs, on the other hand, refer to whether the audience believes the desired outcome is achievable [170]. Therefore, when designing the message, the promoted behavior should not only be optimal from a healthcare perspective but also feasible. This may require finding a balance between ideal healthcare practices and what seems to be realistically achievable from the perspective of the audience.

TARGET AUDIENCE. The target audience is defined based on the intent of the disease story. Therefore, population groups that contribute to or can address the clinical practice dilemma are identified, such as the group with the highest prevalence of a particular disease. To enhance the impact of the disease story, *arguments-from-consequences* should focus on emphasizing outcomes perceived as desirable by the audience [179]. This concept is referred to as the *congruency rule* or matching effect and while it may seem obvious, this principle is still not always applied in health communication [170]. The domain expert plays a crucial role in providing a deeper understanding of the target audience, as they possess knowledge of common concerns and questions that arise during patient-physician consultations.

4.2 DESIGN CONSIDERATIONS OF A DISEASE STORY ABOUT CSVD

The disease story design process will be demonstrated on the application example of a disease story about CSVD. The concrete design of a disease story allows for a lot of variability. Many design decisions are inherently creative, and no universally accepted best practices exist. To explore the design process and identify key challenges, an iterative story design approach was employed. The interdisciplinary team included visualization researchers, a medical expert specializing in CSVD, an interaction designer, and a certified medical illustrator. Regular design meetings established an iterative feedback loop, ensuring continuous refinement and improvement of the narrative visualization. The resulting story (and its variations) serve as the foundation for several studies described in this thesis.

4.2.1 *The Knowledge Stage*

The project began by consulting a *domain expert* who is both, a treating physician and researcher from the department of neurology, and deals with CSVD regularly. They defined the *clinical practice dilemma*: *Hypertension is a common condition, particularly in the Saxony-Anhalt region. Many individuals fail to take it seriously, leading to inadequate treatment. After approximately 15 years of untreated hypertension, the first symptoms of CSVD* *may appear, such as gait disorders. Since CSVD is not widely recognized, it is often diagnosed only in its later stages. Additionally, CSVD increases the risk of various types of dementia, especially vascular dementia, such as Alzheimer's disease.* The domain expert also facilitated data acquisition, providing MRI scans and tabular data from cross-sectional patient studies (e.g., containing information about the frequency of different risk factors). The MRI scans underwent further processing, with the neurologist segmenting damaged brain areas for visualization. Since the available study data was limited to patients from the University Hospital Magdeburg, careful interpretation was required to avoid misleading conclusions. The domain expert played a key role in reviewing the data and refining the approach to effectively communicate key insights. Additional information about global CSVD prevalence was researched in scientific publications: [35, 92, 136].

4.2.2 *The Narrative Stage*

CONFLICT. The conflict is derived from the clinical practice dilemma. The main conflict for this story was formulated as: *A patient developed CSVD due to years of untreated hypertension.*

CONTENT. The integration of real-world data and domain knowledge was an iterative and collaborative process involving multiple discussions with medical experts, visualization researchers, and designers. These discussions were essential for refining the approach and ensuring that the final narrative aligns with scientific consensus while remaining accessible to the target audience. Several design decisions were made in the process, excerpts of the story are shown in Figure 10. A dark background with white text and accent colors was chosen to create contrast, while distinct colors for CSVD subtypes were consistently applied across charts and visualizations. For data visualization, bar charts were used to represent data, accompanied by concise textual descriptions to improve comprehension. Pictographs, featuring human figure silhouettes, illustrated statistical distributions in health data, while cross-sectional brain images helped convey the disease's origin. Additionally, 3D models of the brain and lesions were created using NIfTI files and the marching cubes algorithm [141]. Icons were incorporated throughout the visual narrative to provide contextual information.

The narrative was structured as an interactive slideshow to engage users effectively. Smooth fade-in and fade-out transitions enhance storytelling flow, while navigation buttons allowed user control, with a progress bar indicating story completion. Clickable icons provided access to additional information, and 3D models were made rotatable, restricted to the up-axis for ease of use. MRI slice exploration was implemented through a horizontal clipping plane controlled via a slider, see Figure 11. Furthermore, interactive tasks were introduced, such as dragging a patient pictograph through an MRI scanner to reveal cross-sectional images.

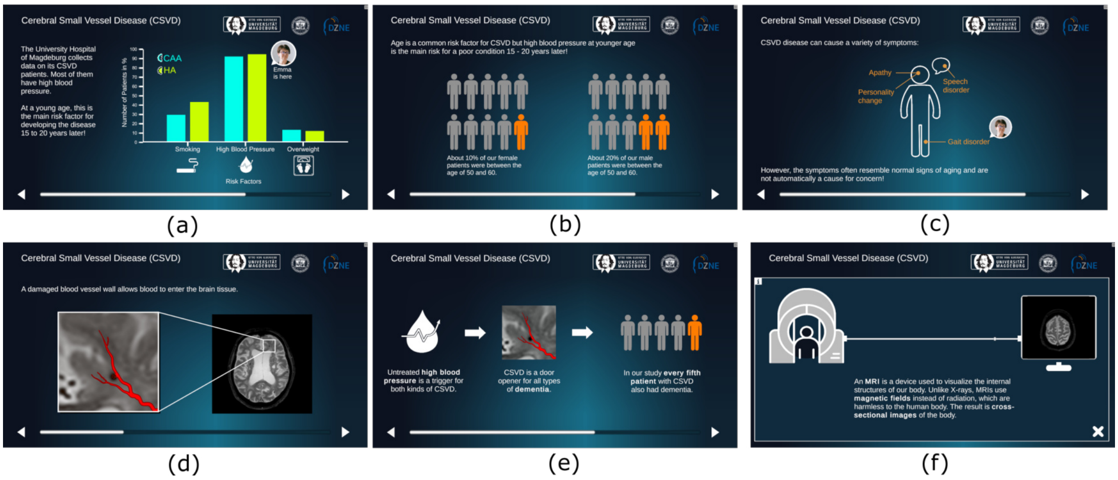
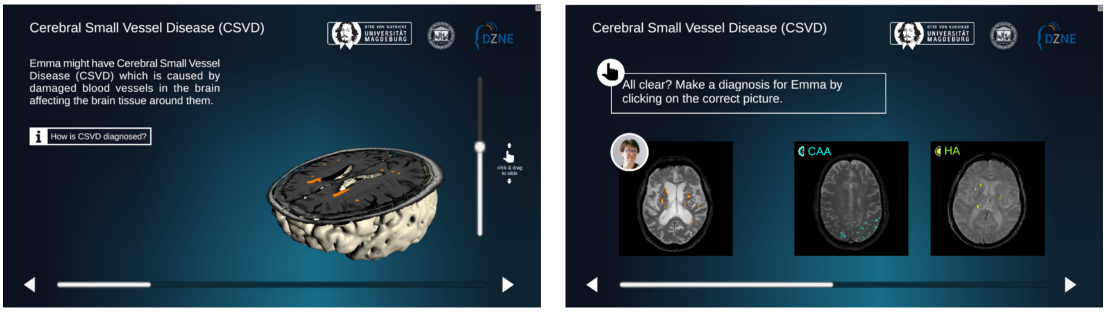


Figure 10: Selected slides from the CSVD story, showing (a) distribution of risk factors, (b) the distribution of patients regarding age, (c) pictographs of the frequency of symptoms, (d) schematic vessel drawings, (e) the visualization of our key message, and (f) the interactive explanation of the MRI. Image from Mittenentzwei et al. [163], available under a CC BY-NC-ND 4.0 license. No changes were made.

CHARACTER. To create an engaging and relatable narrative, two primary characters were introduced: (1) a medical expert, serving as a trusted source of information, and (2) a patient, representing individuals at risk for CSVD. The patient character was designed to foster empathy and personal connection, helping viewers relate to the disease's risk factors and consequences. A stock photo of a 60-year-old woman was used, as this demographic aligns with a high-risk profile for CSVD. Additionally, an icon representing the patient was incorporated throughout the slides, visually linking the data to her case. The medical expert in the story is a real neurologist from the University Hospital Magdeburg, enhancing the credibility of the narrative.

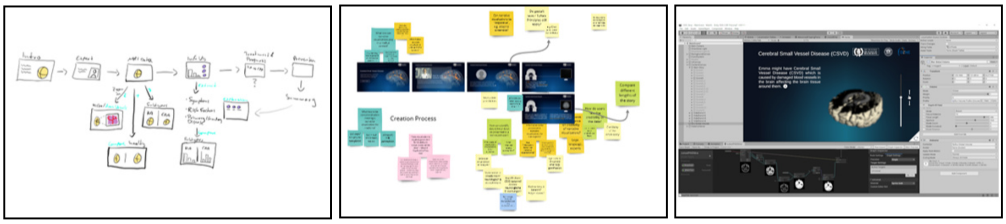
STRUCTURE. The structure of the story was developed through multiple iterations. Several storytelling frameworks were considered, including Freytag's Pyramid, Brent Dykes' Storytelling Arc, and an adapted version of Campbell's Hero's Journey. These structured approaches helped refine the logical flow of the narrative. To determine the order of events and visualizations, the disease story was initially drafted as a sketch. To facilitate collaboration, the online whiteboard Miro [202] was used for storyboarding and discussing various layout proposals (see Figure 12).

Following Freytag's Pyramid, the story progresses as follows, see Figure 13. To *introduce the topic* and *increase insights*, the disease is described using real MRI data to visualize its effects on the body and enhance users' understanding of its background. Tabular data from a cross-sectional study is presented through accessible information visualizations, such as bar charts, to illustrate the distribution of various risk factors,



(a) (b)

Figure 11: Two screenshots from the CSVD story, showing slides containing (a) 3D interaction possibilities (rotating and clipping) and (b) a point and click interaction task. Image from Mittenentzwei et al. [163], available under a CC BY-NC-ND 4.0 license. No changes were made.



Sketching Prototyping Implementing

Figure 12: Main steps in developing the disease story about CSVD. Story design begins by briefly sketching the story, followed by more concrete prototypes using Miro, and finally the implementation in Unity. Image adapted from Mittenentzwei et al. [163], available under a CC BY-NC-ND 4.0 license.

leading to the *key message*. Highlighting symptoms and risk factors is part of the *resolution process*, where the fictional patient is frequently referenced, demonstrating which of the presented risk factors and symptoms apply to her. This approach is designed to foster empathy and make the data more relatable and engaging.

Similar to the disease template proposed by Meuschke et al. [152], methods for disease prevention are presented at the end of the story to emphasize the *call to action*. However, instead of strictly adhering to the sequential structure of the template, certain elements – such as symptoms, risk factors, and prevention strategies – are intentionally foreshadowed and repeated throughout the narrative. This technique is intended to capture user attention and reinforce key insights, improving retention and recall. The final slide provides additional information about the creators of the story, similar to a movie credits screen, and presents the *sources* of information to enhance the credibility of the data.

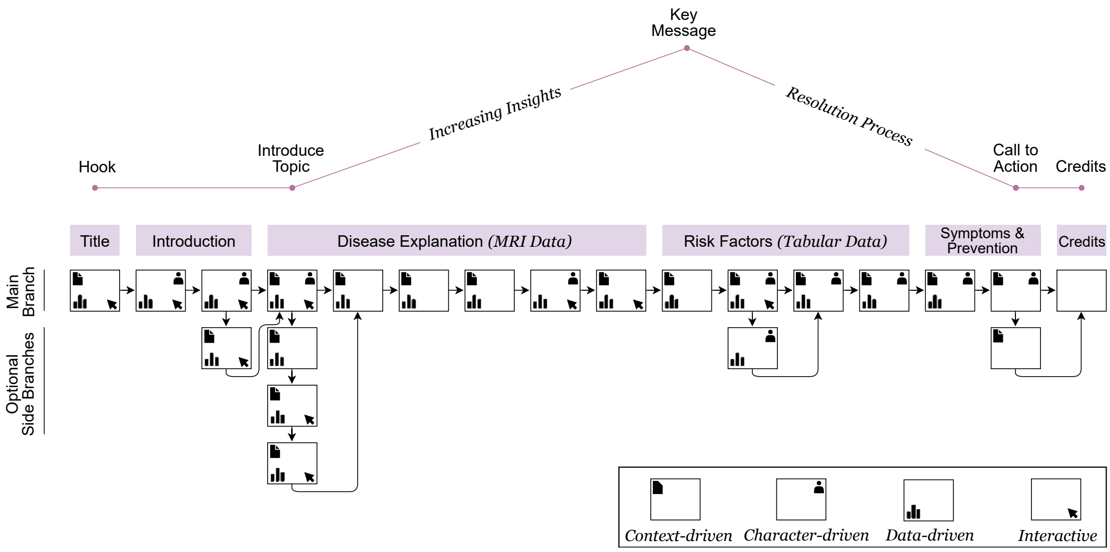


Figure 13: The story structure mapped to an adapted version of Freytag's Pyramid and the storytelling arc. The story follows a linear main branch, offering to switch to optional side branches temporarily (so-called elastic structure). Thematic sections are highlighted in different colors. In each slide icons show data-driven, context-driven, character-driven, or interactive content. Image adapted from Mittenentzwei et al. [163], available under a CC BY-NC-ND 4.0 license.

4.2.3 *The Effect Stage*

This project aims to communicate critical information about CSVD to a lay audience, specifically adults aged 45 and older, as this is a crucial age for regulating hypertension to prevent CSVD. CSVD remains poorly understood, with many unanswered questions, making expert-driven communication essential. The intent is to highlight the neurological risks of hypertension and educate a lay audience about CSVD to promote preventive behavior and early detection. To achieve this, the following key message is conveyed to the audience: *"High blood pressure can lead to CSVD, significantly increasing the risk of dementia."* The narrative emphasizes the importance of adopting a healthier lifestyle, regularly monitoring blood pressure, and seeking medical treatment when necessary.

4.3 CONCLUSION

The disease story design process offers a structured framework for systematically developing narrative visualizations that communicate medical information, as demonstrated through the example of CSVD. However, it should not be interpreted as a strictly linear process. In practice, many stages of the design process are carried out in parallel.

For instance, defining the message and identifying the target audience (both components of the effect stage) are essential early steps that often influence decisions made in the narrative stage. Moreover, iterative collaboration with medical and design experts is important to improve initial story designs. Feedback loops frequently lead to revisions and refinements of previously developed materials.

In addition, the design process does not provide guidance for specific design decisions. Instead, it highlights the key components of a disease story where such decisions must be made by the authors. It emphasizes how these decisions are interconnected, helping authors understand how choices in one area – such as character design, narrative structure, and the target audience – can influence other aspects of the story. This approach encourages context-sensitive decision-making tailored to the author's goals.

5 THE KNOWLEDGE STAGE: DOMAIN EXPERTS AND AUTHORING

The knowledge stage focuses on the motivation ("Why communicate?") behind creating disease stories, emphasizing the collaboration between the story author, who creates the narrative, and the domain expert, who provides domain knowledge, feedback, and data (see Figure 14). This chapter therefore highlights the contributions of the domain experts involved in the research presented in this thesis, outlining their motivations and goals. On the author's side, the data and tools utilized in this thesis are discussed. The content is partly based on the following publications: [155, 158, 162, 163].

Figure 14: Overview of the knowledge stage of the disease story design process (cf. Figure 9).

5.1 COOPERATION WITH DOMAIN EXPERTS

Close collaboration with domain experts is essential to ensure the accuracy and relevance of the content. The research presented in this thesis was conducted in cooperation with clinicians specializing in cardiology, neurology, radiology, and epidemiology, all of whom have extensive practical experience. No compensation was involved. Organized according to the four categories of clinical practice dilemmas from Section 4.1.3, the main motivations of the clinicians are described in the following.

PREVENTION. Prevention consists of two key components: (1) a healthy lifestyle, which benefits most diseases, and (2) early detection, such as blood pressure monitoring and regular check-ups with specialists like gynecologists, urologists, or general practitioners. While the first component provides general advice that reduces the risk for various diseases, the second is more specific to individual conditions, such as breast cancer screenings or colonoscopies. A healthy lifestyle typically does not present risk factors, but it cannot replace check-ups and early detection measures. Many of these check-ups, however, are based on medical procedures that carry their own risks, such as colonoscopies, which can lead to complications like colon perforation. This section focuses on prevention through promoting healthier lifestyles and raising awareness among laypeople about the existence of certain conditions, so they can identify symptoms and understand the availability of preventive measures. Communicating the risks associated with medical procedures will be addressed in the next section.

The high prevalence of hypertension in Saxony-Anhalt is a concern that local healthcare providers must address. The neurology and cardiology departments at the local university hospital treat many patients suffering from the consequences of untreated hypertension. Although hypertension can often be effectively managed through behavioral changes, such as adopting a healthier lifestyle, or with medication, it is frequently not taken seriously by patients, according to the clinicians. As a result, the clinicians are interested in improving communication about the risks of untreated hypertension to the local population. In collaboration with a domain expert in neurology, a narrative was developed to illustrate how hypertension can lead to dementia [156].

INFORMED CONSENT. Before undergoing medical procedures, patients must provide consent. This consent is crucial because medical procedures, whether preventive screenings or treatments, often carry risks. However, for patients with low health literacy, it can be challenging for physicians to effectively communicate the disease, treatment methods, and their associated risks. Disease stories aim to enhance communication and foster a better understanding of these topics, empowering patients to make informed decisions and provide consent.

Based on this motivation, a collaboration with a physician from the clinic for neurology and radiology at the local university hospital was undertaken to investigate how storytelling can be used to educate patients about tumors and aneurysms [159, 160]. For both diseases, multiple treatment methods are available, and the patient, in consultation with the treating physician, must decide which option is most suitable on an individual basis. This decision requires the patient to have a sufficient understanding of both the disease and the available treatment methods. While the study in this thesis focuses on explaining the diseases themselves, the radiology department has expressed interest in developing additional narratives in the future that will focus on various treatment options.

MEDICAL RESEARCH. To conduct medical studies, a sufficient amount of data is required, often collected from individuals, such as patients or participants in epidemiological studies. A shortage of such cases hinders medical research, making it essential to communicate the importance of participating in these studies, which ultimately benefits public health. For CSVD, many patients are only diagnosed in later stages, limiting the research that can be done in early stages of the disease. Therefore, a neurologist from the university hospital Magdeburg expressed the wish to inform a lay audience about symptoms of CSVD and available contact points concerned people can turn to [156].

The university hospital Greifswald conducts an epidemiological study, the Study of Health in Pomerania (SHIP) [96], which requires numerous participants over an extended period. The researchers involved in this study are interested in effectively communicating the significance and potential benefits of participation to a lay audience, with the aim of encouraging greater involvement. Since SHIP covers a broad spectrum of health-related topics rather than focusing on a single disease, it was initially decided to use a specific subset of the data. This led to the creation of a disease story focused on NAFLD [27, 155].

PROMOTION. Some healthcare facilities offer specialized treatments and aim to promote these services in order to attract the appropriate patients. In these cases, commercial interests play a significant role. It is crucial to ensure that, while addressing these commercial considerations, the story prioritizes the patients' interests.

The *Heart Center Leipzig* is a private clinic specializing in heart diseases. A narrative focused on blood flow dynamics in the aorta due to BAV emphasized that flow patterns are a complex structure, challenging to communicate – especially visually – to a lay audience [113, 157]. In addition to facilitating internal communication with patients, the cooperation partner also expressed interest in using the disease story to showcase their expertise to the broader public.

5.2 AUTHORING: DATA, RESOURCES, AND TOOLS

To create the narrative medical visualizations, various tools and data sources were utilized. The following sections describe the datasets, resources, and software tools used during development. For a more comprehensive overview of the final products, the developed stories are available at: `visualstories.cs.ovgu.de/phd-sarah`.

5.2.1 *Data and Resources*

The disease stories were developed using a combination of medical data, epidemiological information, and supplementary resources.

DATA. A diverse range of data was incorporated to construct detailed and accurate disease stories. The main datasets included:

- CEREBRAL SMALL VESSEL DISEASE:
  - *Medical Image Data:* MRI scans and segmentations of brain lesions were provided in the NIfTI file format. Using Python, 3D surface models were extracted from these images.
  - *Tabular Data:* Patient data from Magdeburg was used, including information on risk factors and disease progression.
- ANEURYSMS AND LIVER TUMORS: Pre-processed 3D models of various anatomical structures were used, including aneurysms and livers with their respective liver tumors and liver vessels.
- SHIP DATA: Epidemiological data including a wide range of biomarkers and sociodemographic variables relevant to disease modeling was incorporated.
- AORTIC HEMODYNAMICS: 3D models of the aorta and the hemodynamics of aortic blood flow (4D PC-MRI data) were visualized using the tool Bloodline [116]. The resulting visualizations were then exported as video files.

OTHER RESOURCES. Further data (e.g., about prevalence) and contextual information was gathered additionally:

- EXPERT KNOWLEDGE: Insights from domain experts were integrated into the visualizations, either as text annotations or represented through icons.
- SCIENTIFIC PUBLICATIONS AND WEB SOURCES: Background knowledge on diseases, risk factors, and prevalence was gathered from scientific literature and websites of official healthcare providers.

5.2.2 *Tools and Software*

The creation of disease stories relies on a combination of software tools for data processing, visualization generation, and the implementation of interactive features. To ensure an efficient and effective workflow, these tools must meet several key requirements, including:

- ACCESSIBILITY: A low entry barrier for users with varying levels of familiarity with the software, strong domain relevance, and an active user community.
- DATA COMPATIBILITY & INTEGRATION: The ability to process and integrate diverse medical data formats, such as DICOM, NIfTI, CSV, JSON, OBJ, and PNG.
- MULTI-MODAL SUPPORT & DESIGN PROCESS: Functionalities to render and overlay various visualization types, including images and surface models, while enabling the integration of sound, text, and aesthetic arrangement within the design process.

- INTERACTIVITY & USER ENGAGEMENT: Features that facilitate user interaction, including annotations, tooltips, data filtering, and exploratory capabilities to enhance engagement.
- TRANSITIONS & ANIMATION: Tools for animating content to ensure smooth, aesthetically pleasing transitions and effectively represent changes over time.
- EXPORT & SHARING OPTIONS: Support for multiple export formats, including web-based, mobile, and desktop solutions, to improve accessibility and facilitate knowledge dissemination.
- SECURITY & COMPLIANCE: Ensuring adherence to data privacy regulations and the secure handling of sensitive data.

Based on these criteria, Unity was chosen as the main software for crafting disease stories. Different components for the stories were created in other tools, e.g., graphic programs and AI.

UNITY. The Unity engine (also known as Unity) [235] is a powerful tool that offers several advantages for building disease stories. Game engines, in general, have become popular in various research fields, including visualization, as they not only provide a wide range of functionalities but also come at a relatively low cost compared to professional software specialized in 3D visualization or animation [66]. Unity is widely used with an active user base that provide resources and support online. In contrast to other game engines such as the Unreal engine [59], Unity is reported to be easier to learn, therefore being more accessible for authors of narrative visualizations [224].

Unity supports the direct import of 2D graphic objects (e.g., PNG files) and 3D graphic objects (e.g., OBJ files). File types not supported by Unity can either be preprocessed in external tools (e.g., extracting a surface model in OBJ format from DICOM data) or loaded through custom scripts written by the user. In addition to 3D models and 2D graphic objects, Unity provides options to include audio, text, and various pre-made UI elements (e.g., buttons and sliders). As a game engine, Unity offers many built-in features for interaction beyond standard UI elements, such as graphical object manipulation (e.g., rotation). Unity also includes animation tools, enabling the creation of interactive and animated content with minimal implementation effort. Shader development is supported through Unity's Shader Graph, a visual scripting tool that makes shader creation more accessible to users with limited programming experience. However, experienced developers still have the flexibility to write shader code manually for greater customization.

The composition of different media into a single disease story is facilitated by Unity's visual editor, where objects can be positioned and modified within a 2D or 3D scene, see Figure 15. Unity allows the final application to be exported to multiple platforms, including different operating systems, mobile devices, web applications, and

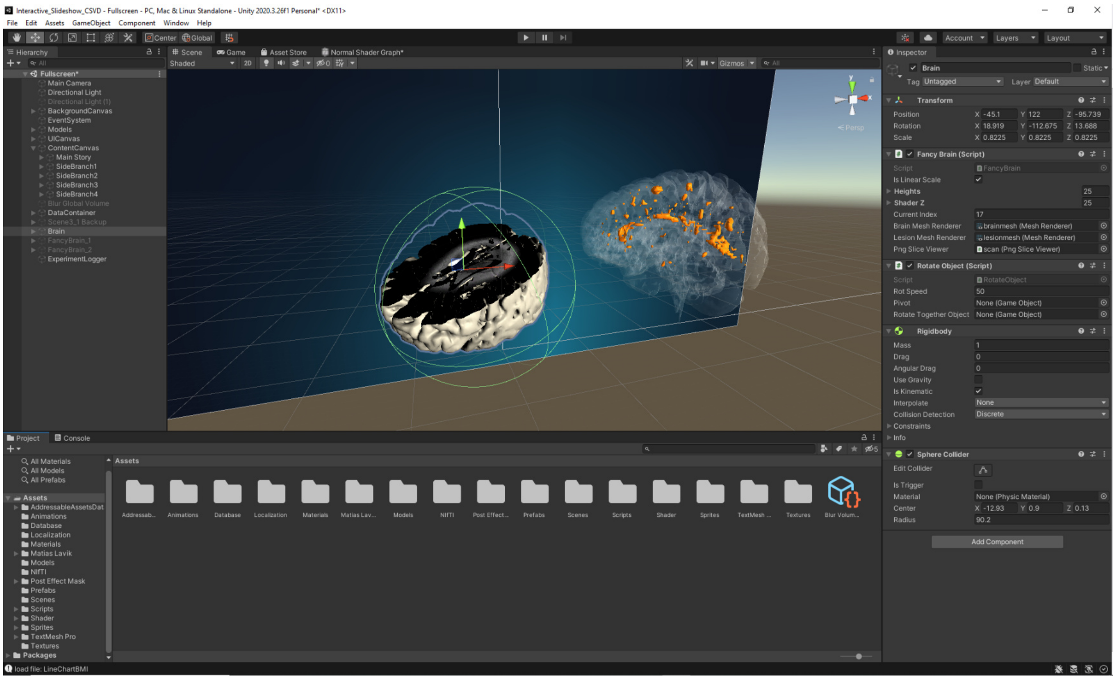


Figure 15: The Unity game engine allows 3D models to be positioned in the visual editor.

even gaming consoles. This cross-platform compatibility enables deployment across various output media without significant additional development effort. Additionally, Unity offers built-in options for translating textual content, allowing projects to be created in multiple languages, thereby increasing accessibility for a broader audience.

STABLE DIFFUSION. The GenAI model Stable Diffusion version 1.5. was used to generate synthetic patient images while ensuring patient privacy. The recent rise in popularity of openly available (latent) diffusion models like Stable Diffusion [227], along with user-friendly applications such as Midjourney [154] and Leonardo AI [220], has brought GenAI art to the forefront. These models and tools enable the creation of images from text across various domains. While training diffusion models typically requires large amounts of data and computational resources, a technique called Low-Rank Adaptation (LoRA) [86] allows for efficiently fine-tuning existing models for specific domains and applications, such as stylized or near-photorealistic images. This works by only adjusting a few parameters of an already trained model. Community-driven platforms like Civitai [37] facilitate the sharing and customization of such models, while tools like the Stable Diffusion Web UI by Automatic1111 [12] provide accessible interfaces for their use.

Beyond text-to-image generation, diffusion models can also be used for inpainting (replacing content in a masked image), sketch-to-image transformation (conditioning image creation based on a simple sketch), and tasks such as upsampling or downsampling. These advancements in GenAI art present various opportunities to enhance design processes. So et al. [223] introduced a GenAI tool for generating narrative medical visuals based on social media posts about conditions such as diabetes. This approach aligns with the bio-psycho-social model, which incorporates medical, psychological, and social factors in healthcare [79].

DATA ANALYSIS. Python and R were used for multiple purposes, including:

- Data analysis of evaluation results (Python, R)
- Extraction and conversion of 3D models from NIfTI files using the marching cubes algorithm and mesh decimation (Python)

Scripting languages like Python and R are accessible and easy to use. They provide great flexibility through a broad range of libraries that offer functionalities for loading, processing, and visualizing data. The natural language processing (NLP) model RoBERTa [139] was used to analyze YouTube comments because it can interpret the meaning of the text, enabling efficient processing of large datasets and eliminating the need for manual analysis of every comment.

ARTISTIC AND PROTOTYPING TOOLS. Blender [20] was used for modifying and refining 3D surface models, while Inkscape [93] was utilized for designing icons and graphical elements to enhance visual storytelling. For sketching and prototyping, both traditional and digital pen-and-paper methods, along with Miro [202], were employed. These tools were used in the early conceptualization phase to create storyboards and plan visualization layouts, see Figure 16. Miro offers great flexibility, allowing images, text, and graphical elements such as arrows to be freely arranged, making it well-suited for collaborative work among all stakeholders.

5.3 CONCLUSION

Disease stories can be created using a variety of tools, depending on the specific requirements of the project. These requirements may include support for particular data types, interactive features, or considerations related to the author's skill set, such as proficiency in coding or design. In this work, the tools chosen for prototyping and implementation primarily required some coding knowledge but only limited design expertise, as the stories rely predominantly on data visualizations rather than detailed illustrations.

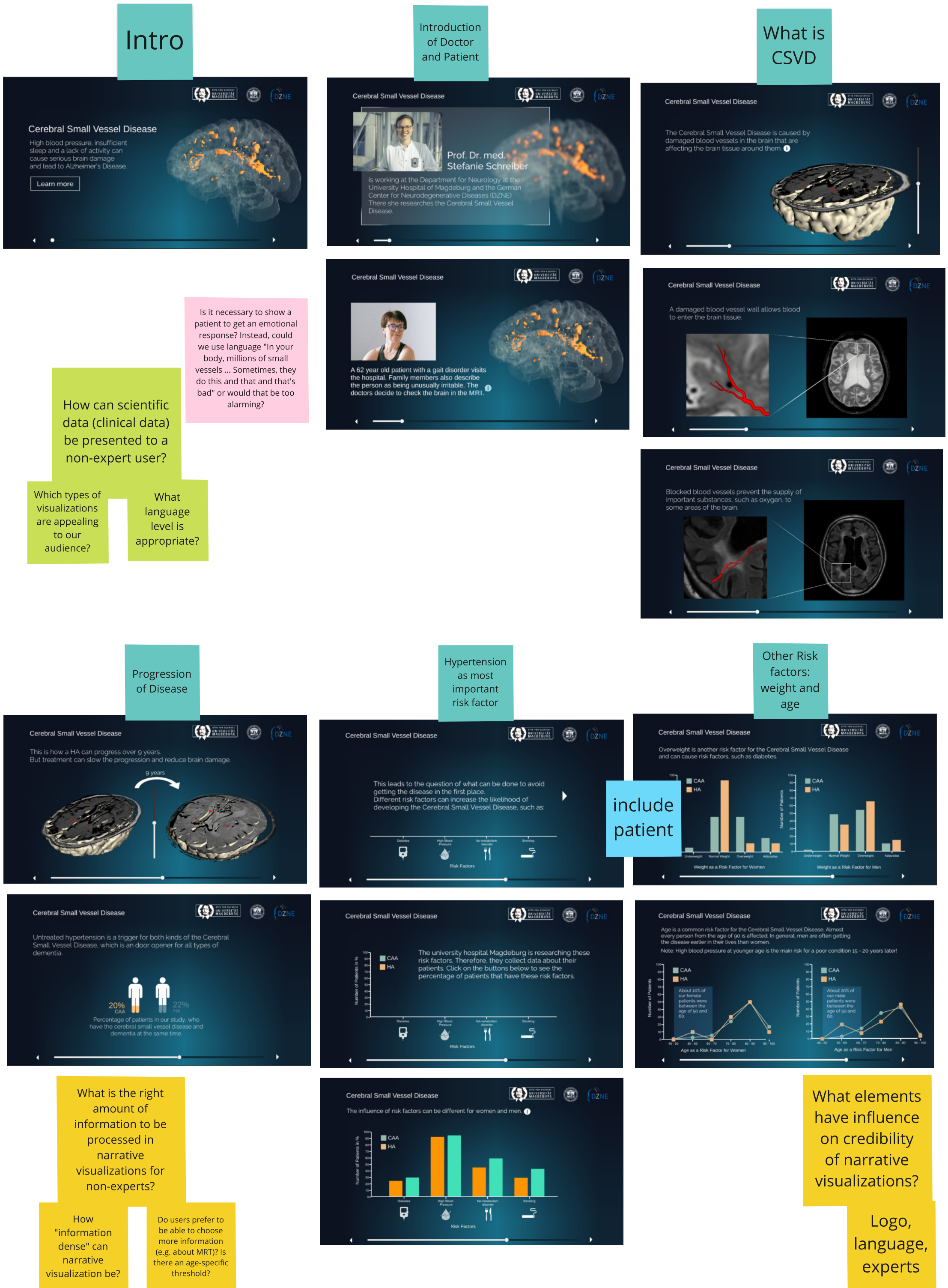


Figure 16: A Miro board showing screenshots from the CSVD story and notes used for brainstorming and feedbacking.

6 THE NARRATIVE STAGE: CONTENT

The narrative stage addresses the question *"How to communicate?"*. It defines how the narrative is constructed based on four key ingredients (see Figure 17). The second ingredient, *content*, determines what is included in the disease story, such as data, visuals, or text. Essential content is selected to effectively portray the conflict. In the context of disease stories, different types of content can be identified: *data-driven content* (e.g., data visualizations such as MRI volume renderings), *context-driven content* (e.g., explanations of procedures or lists of symptoms), and *character-driven content* (e.g., names, photos, character descriptions) [156]. Since *character* is also a separate ingredient, highlighting its importance in storytelling, it will be explored in greater depth in later chapters.

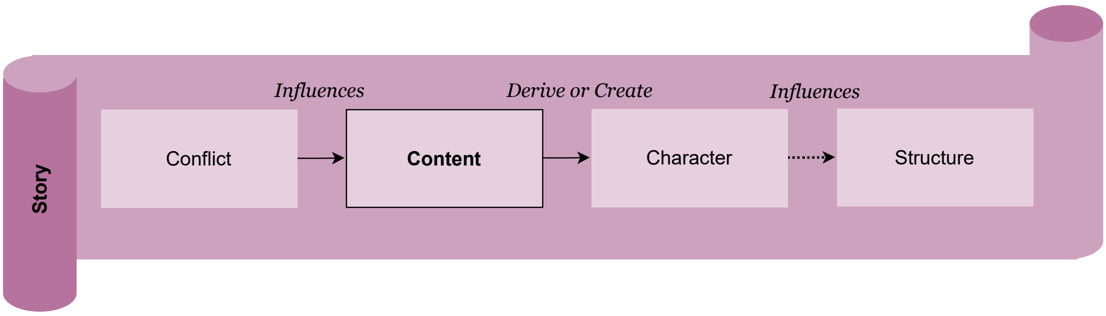


Figure 17: The narrative stage of the disease story design process (cf. Figure 9) with *content* highlighted.

Design decisions for data-driven content are examined, focusing on how to effectively visualize anatomical surface models of pathologies with changing morphology. Two case studies ($n = 60, n = 40$) were conducted to investigate which visualization techniques for anatomical surface models perform best and are preferred by laypeople. To address this, the following research questions are explored:

RQ1 How to visualize morphological changes of pathologies using surface models for a lay audience?

- *RQ1.1* What are aesthetic preferences of lay audiences for visualizing surface models?
- *RQ1.2* Do visual preferences align with higher performances in solving tasks?

6.1 COMMUNICATING MORPHOLOGICAL CHANGES IN PATHOLOGIES

One type of data-driven content in disease stories is image data from medical imaging modalities such as X-ray, CT, or MRI. It is valuable for visualizing diseases and their effects on human anatomy for lay audiences. The following studies focus on volumetric image data, which includes characteristic visualization techniques (e.g., depth cues) and interaction methods (e.g., rotation). Further, they will explore aspects of their design space in the context of accessibility for a lay audience.

Collaboration and Personal Contribution

This chapter is partially based on the following publications: [159, 160]. The conceptual foundation for both projects was developed by the author of this thesis. The first project, along with its evaluation, was implemented by Sophie Mlitzke as part of a study project supervised by the author of this thesis. Throughout the implementation process, the concept was collaboratively discussed and refined. The hatching for the surface models was provided by Kai Lawonn. The second project builds upon the first case study, serving as an extension of its findings. It was implemented and evaluated by Sophie Mlitzke and Darija Grisanova within a study project supervised by the author of this thesis. Both contributed to the concept. The scientific reappraisal of both projects, including the development of the research papers, was conducted by the author of this thesis. The submitted works have been prepared with all co-authors who provided feedback throughout the project and writing phase.

The primary focus of this chapter is to explore how to effectively communicate the morphology of pathologies to a lay audience. As application examples, two diseases that exhibit morphological changes over time are used, allowing to examine not only general preferences of lay audiences for 3D visualizations but also their suitability for analyzing and comparing multiple datasets.

A broad range of pathologies show morphological changes over time – such as tumors, vessel dilations and stenosis, or inflammatory diseases that can lead to atrophy (e.g., shrinkage of the brain or muscles) – making this a relevant research topic. Communicating morphological changes to a lay audience is challenging, particularly due to varying levels of health literacy. However, understanding the growth and shrinkage of pathological structures is essential for treatment planning and outcome prediction. Providing this information to patients and their relatives not only supports informed consent but also encourages preventive measures, such as regular screening.

While visualization techniques can emphasize morphology and facilitate the comparison of multiple datasets, interactive features, e.g., navigation buttons and rotatable objects, enable users to explore the data at their own pace, allowing them to focus on details of personal interest. However, interaction must be carefully designed to ensure accessibility, not only for individuals with limited health literacy but also for those with limited visualization or technological literacy. Therefore, the following case studies compare different techniques for visualizing 3D objects and presenting multiple time steps. Liver tumors and cerebral aneurysms are used as application examples, representing cancer and vascular diseases, respectively. Through two exploratory case studies, insights into visual preferences and the ability of a lay audience to assess morphological changes are provided. Both user studies include tasks requiring participants to evaluate the growth and shrinkage of pathologies. These tasks are not intended for inclusion in a final disease story but serve as a means to quantitatively assess how effectively participants can interpret morphology using different visualization methods. The tasks focus on key aspects of the data that, according to discussions with clinical collaborators, should be communicated: growth/shrinkage rate and growth direction.

A Note on Terminology

In this chapter the terms *visualization techniques* and *presentation techniques* are used. While visualization techniques refer to the visualization of one object (i.e. which shader is used), presentation techniques refer to the way subsequent time steps are arranged (i.e. showing them side-by-side or successively).

The findings from both case studies provide a basis for improving the accessibility of medical visualizations to lay audiences. This contributes to research on disease stories that incorporate such data visualizations.

6.2 RELATED WORK

In the following, relevant medical visualization techniques for surface models and longitudinal data are presented, which are foundational for the subsequent discussion.

6.2.1 *Medical Visualization for Surface Models*

Various methods for the visualization of anatomical and pathological structures as surface models have been developed. Smoothing surface models is an important prerequisite to compensate for staircase artifacts resulting from the limited spatial resolution and anisotropic nature of medical image data. However, simple mesh smoothing techniques often lead to a loss of details and volume. Bade et al. [18] investigated the effectiveness of different mesh smoothing approaches for the preservation of anatomical shape features of tumors. One substantial smoothing technique limits vertex displacement to a single diagonal within a voxel [76]. Moench et al. [164] further reduced the adverse effects of smoothing by restricting adjustments primarily to vertices heavily impacted by anisotropic resolution. Mainly applied to tumor models, their approach enables smoothing while maintaining spatial context. Subsequently, Wei et al. [253] introduced methods designed to better preserve the distinctive shape features of anatomical structures, with more recent innovations incorporating deep learning techniques [252]. Smoothing plays a crucial role in illustrative surface visualizations, as these visualizations highlight surface discontinuities. Without appropriate smoothing, such discontinuities may be misinterpreted as anatomical features when, in fact, they are artifacts. In the case of vascular structures, specialized methods based on implicit surfaces have been developed to improve the readability of complex branching patterns [209]. When creating visualization techniques for lay audiences, accuracy is less important than for, e.g., surgery planning. Therefore, alterations of the models caused by smoothing, such as shrinking, are tolerable.

Illumination-based rendering techniques, such as Phong shading, simulate the interaction of light with surfaces by adding highlights and shadows relative to a light source, thereby enhancing depth perception and contributing to a more realistic appearance. Phong shading serves as a foundational technique and is frequently used as a baseline for evaluating state-of-the-art rendering methods [124].

In contrast, illustrative rendering techniques draw inspiration from traditional scientific illustrations and aim to simplify visual representations through abstraction [244]. Common illustrative techniques include silhouette rendering, feature line rendering, stippling, and hatching, which can be applied to direct volume rendering and surface rendering [26, 201]. Given the focus of this work on segmented anatomical and pathological structures, surface rendering is of particular importance. Tietjen et al. [238] described the implementation of silhouette rendering combined with surface rendering, which has been applied to visualize liver tumors.

6.2.2 *Visualization for Longitudinal Medical Image Data*

A variety of techniques have been developed within the field of visual analytics to effectively capture and communicate temporal changes [10]. Zhang et al. [276] adapted these techniques for the healthcare domain, highlighting the use of linked views. In alignment with this approach, the visualizations used in this study synchronize multiple time steps by linking the models' orientation. Rotating one surface model automatically applies the same rotation to all others. Sugathan et al. [231] visualized the progression of brain lesions in multiple sclerosis patients, using contour and mesh-based representations. Their approach included side-by-side and overlay views to compare different time steps. This work adopts a similar strategy by arranging time steps side-by-side as well as successive where the models are replaced in the same location.

Visualizing longitudinal medical imaging data is crucial for clinicians when assessing treatment results and predicting outcomes. To enable effective comparison across time points, registration is typically required. However, registration remains a complex challenge, even with recent advancements in deep learning [68]. In contrast to visualizations based on longitudinal image data, tumor growth models aim to mathematically predict the progression of tumors over time [196, 207]. Comparable modeling techniques have also been applied to aneurysms [3]. Such predictive models are valuable tools for treatment planning and forecasting disease progression.

6.3 MEDICAL DATA SETS

Intracranial aneurysms and liver tumors were chosen as medical examples to illustrate pathological growth and shrinkage over time, representing vascular and tumor-related diseases. To depict temporal changes, data from multiple time points is required.

Longitudinal datasets documenting the natural progression of cerebral aneurysms are scarce, as many patients opt for treatment or do not consistently return to the same medical facility for follow-ups. In the absence of longitudinal data from multiple patients, artificial time steps were generated to recreate disease progression. To achieve this, surface models derived from real medical imaging data were modified using the 3D modeling software Blender (see Figure 18). This approach also bypasses the complex registration challenges typically associated with medical image data. Five variations of each model were created to represent potential growth over time. The original medical image data was used as the time step with the largest pathology, while other time steps were artificially generated by progressively reducing the aneurysm's size. One of these artificial time steps was always designed to depict the healthy vessel prior to aneurysm formation.

Similar to the aneurysms, the intermediate time steps between a healthy liver and the original tumor image data were artificially generated by gradually reducing the tumor's size. In the first study, six intracranial aneurysm datasets and four liver tumor datasets were utilized. However, in the second study, the number of datasets per pathology was reduced to three due to modifications in the study design. Additionally, one model of each time series was used to render a video explaining the disease in the second case study, leaving four models for the user tasks, as shown in Figure 18.

The artificially generated time steps do not reflect the actual morphological progression of these structures over time. However, since their purpose is to educate lay audiences on general concepts rather than provide precise medical accuracy, the need for exact replication of pathological growth is reduced. To ensure the models were appropriate for educational use, they were reviewed by a senior physician from the neurology department and the head of the radiology department at the University Hospital Magdeburg. Both experts agreed that the models were sufficient for lay audiences, as pathological growth patterns can vary strongly among patients, and the generated models present plausible examples of disease progression. Moreover, simplifying medical concepts is a common practice when communicating with lay audi-

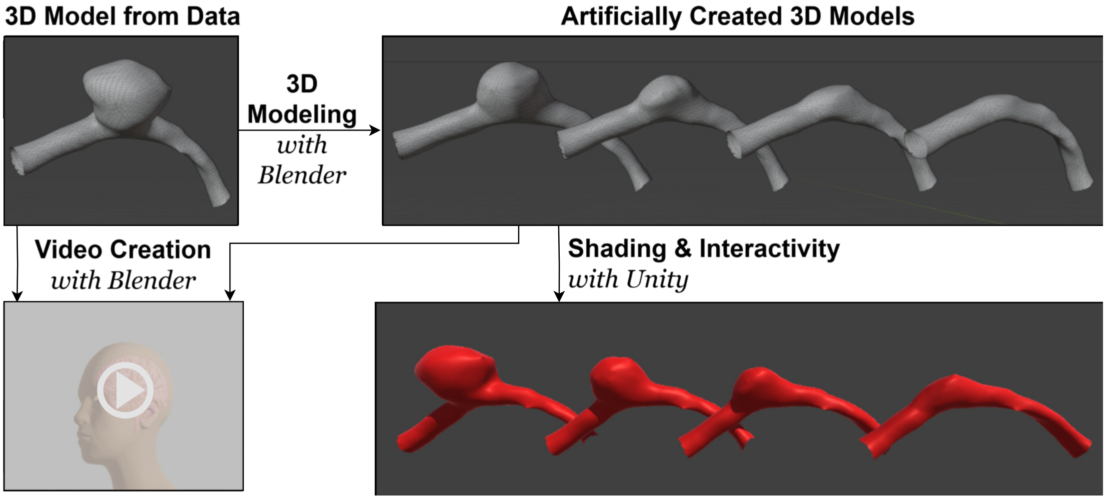


Figure 18: The generation of the data. The videos only apply to the second case study. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

ences. Drucker et al. [55] emphasize the importance of balancing the accuracy of data visualization with the audience's level of literacy to ensure effective comprehension.

Whenever possible, using real medical data is recommended over artificially crafting time steps. This eliminates the need to manually create artificial time steps and aligns more closely with the data-driven principle of narrative visualization, which takes advantage of the fact that real data is considered more engaging [176, 181]. Generating artificial time steps is a resource-intensive task that requires collaboration between visualization or design specialists, who create the 3D models, and medical experts, who validate their accuracy. This is particularly important for complex anatomical surfaces. Both aneurysms and tumors can develop bulges known as blebs, and malignant tumors often exhibit intricate growth patterns, especially when interacting with blood vessels, making artificial modeling especially challenging. Although artificial time steps were created due to the unavailability of real longitudinal data, the findings of this study remain relevant and applicable to comparable authentic datasets. The visualizations were developed using Unity 2021.3.7f1, applying shading to the models and provide interaction opportunities, see Figure 18.

Supplementary Material

Supplementary material for this chapter can be found at:
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/content`

6

Learning Objectives

The motivation for communicating pathological growth and shrinkage – to support informed consent – can be framed as learning objectives. Cognitive objectives emphasize understanding and evaluation, following Adar et al. [1]:

- The viewer will learn to *interpret* consequences of pathological growth and shrinkage.
- The viewer will learn to *judge* consequences of pathological growth and shrinkage.

Affective objectives focus on raising awareness and strengthening beliefs, as defined by Lee-Robbins et al. [132]:

- The viewer will be able to *consider* the consequences of pathological growth and shrinkage.
- The viewer will be able to *trust* the consequences of pathological growth and shrinkage.

These objectives guide the design of disease stories but are not the focus of evaluation in the following studies; rather, they provide a structured rationale for this research.

6.4 FIRST CASE STUDY TO COMMUNICATE PATHOLOGIES AND GROWTH

A user study ($n = 60$) was conducted to evaluate different visual representations based on their suitability for task-solving and aesthetics. Surface models depicting the evolution of pathological structures over multiple discrete time steps were created and visualized using both illumination-based and illustrative techniques.

6.4.1 *Participants*

Participants were asked about their age, gender, highest education degree, and professional experience with medicine and three-dimensional visualizations, as shown in Table 1. The majority of participants (87%) were 35 years old or younger. Approximately half (52%) were female, while six participants identified as diverse or chose not to disclose their gender. Half of the participants reported no professional experience with 3D visualizations. In terms of medical expertise, 16 participants (27%) stated that they frequently engage with medical topics, while seven (12%) indicated that they do

Age	
18-25	25
26-35	23
36-45	4
46-55	1
56-65	5
>65	2

Gender	
Male	23
Female	31
Diverse	5
No answer	1

Education	
Pupil (no graduation)	5
High school diploma	19
Vocational training	3
University degree	33

Professional contact with med. topics	
Never	18
Rarely	9
Sometimes	10
Often	7
Very often	16

Professional contact with 3D vis.	
Never	30
Rarely	5
Sometimes	12
Often	7
Very often	6

Table 1: Participant metadata.

so often, suggesting a relatively high level of medical literacy among the participants. Most participants (83%) were from Germany, and none reported being colorblind.

Participants were recruited through the *Long Night of Science* in Magdeburg, Germany, an annual event where the university presents its research to the local public, as well as via the university mailing list, resulting in a total of 60 participants. A comparative between-subject user study was conducted to evaluate the visualization techniques. The study was set up as an online experiment in which each participant completed twelve tasks for the aneurysm datasets and twelve tasks for the tumor datasets. For each dataset, participants were required to complete the tasks and indicate their confidence in their answers. The study was conducted over approximately one month.

6.4.2 *Study Design*

The design space for visualizing changes in medical datasets is extensive, and this study does not aim to cover it comprehensively. Instead, the focus was on an initial exploration of different stylistic approaches – specifically, illumination-based and illustrative methods, to guide future design studies. Common techniques for enhancing shape perception include shading (such as Phong or tone shading), feature lines, textures, and boundary emphasis [192]. The study focused on three visualization approaches: a semi-realistic illumination-based shading technique (Phong), an illustrative technique (outlines and feature lines), and a hybrid approach that combines a semi-realistic shader with an illustrative technique (Fresnel shader combined with hatching lines), see Figure 19.

semi-realistic visualization. Phong shading was selected as a representative for *shading techniques*, providing a semi-realistic representation of the surface models by adding highlights and shadows, as shown in Figure 19 (a). The aneurysm datasets were colored red, as this is the most intuitive color for blood vessels. Tumors were colored yellow to create strong contrast with the surrounding vascular tree and liver tissue, which appear in various shades of red. By using a simplified, semi-realistic representation of anatomical structures, the goal was to make the visualization more accessible to a lay audience. Realistic medical imagery is often perceived as unsettling, whereas these simplified models use conventional anatomical colors (e.g., red for blood vessels) to remain informative without being visually disturbing.

illustrative visualization. Outlines were chosen as an example of *boundary emphasis*. To further enhance the outlines, view-dependent feature lines were incorporated to highlight the ridges and valleys of the surfaces. In contrast to the semi-realistic approach, this results in a fully stylized illustrative visualization, as shown in Figure 19 (b). This was achieved using an outline shader based on the inverse hull shading technique, which employs a multi-pass rendering approach [200]. In the first pass, the object is rendered with an unlit white shader. In the second pass, the outline is generated by slightly scaling up the object and rendering only the backfaces of the enlarged model in black. The thickness of the outline is determined by the scaling factor applied to the larger model. This technique produces an illustrative visualization resembling a simple pencil drawing.

semi-realistic visualization with illustrative aids. Hatching was considered as a texture-based technique; however, standalone hatching lines proved confusing for participants, particularly for small structures like the initial time steps of the tumors. To improve shape recognition, hatching lines were combined with a Fresnel shader. This results in a combination of semi-realistic and illustrative techniques which was used to explore whether this hybrid approach enhances data exploration. A Fresnel shader was applied to create depth cues similar to those provided by the Phong shader. Additionally, gray hatching lines were incorporated as a visual aid to emphasize surface curvature, as shown in Figure 19 (c). The Fresnel shader was selected to generate a smooth color transition from white to gray, ensuring consistency with the color scheme of the illustrative technique. It was also applied to the hatching lines to maintain a visually cohesive result.

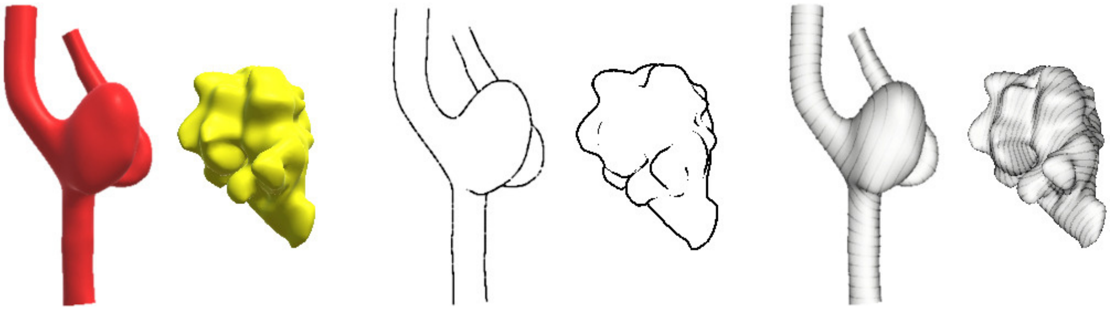


Figure 19: Visualization techniques for aneurysms and tumors: (a) Phong shading, (b) outlines with feature lines, and (c) Fresnel shader with hatching lines. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

6

PRESENTATION OF TIME STEPS. Five time steps for each dataset were displayed in two variants, as shown in Figure 20. The first variant presents all time steps simultaneously, arranged side-by-side, allowing users to view the entire growth process at a glance. The second variant displays time steps sequentially. Users use buttons to switch between time steps. In this approach, the surface model of each time step is exchanged in place. Rather than implementing an automatic animation, user control was prioritized to allow for a more deliberate and flexible analysis of the data. Both variants were visualized using different shaders. While alternative techniques, such as overlaying multiple time steps, could have been considered, this approach was not used because surface models from later time steps are generally larger in all dimensions, enclosing earlier models and rendering them invisible. A potential solution would be to make larger objects semi-transparent; however, this was not pursued, as introducing transparency would add another dimension to the design space.

In the evaluation, five time steps were chosen. A smaller number of time steps was deemed too simplistic for the task design, whereas adding more steps would have increased the time participants would need due to the additional effort required to analyze all given time steps. Five time steps were determined to provide a balance between task complexity and participation time.

TASK COMPOSITION. To minimize potential bias due to learning from previous tasks, six evaluation sequences were designed to present the visualizations in different order. Each visualization technique was shown in two sequences but with varying order. Each participant was assigned to one of the following sequences:

- SEQUENCE A
  1. Aneurysm & Phong shader & side-by-side
  2. Aneurysm & Phong shader & successive
  3. Liver Tumor & Phong shader & side-by-side
  4. Liver Tumor & Phong shader & successive
- SEQUENCE B
  1. Aneurysm & Phong shader & successive
  2. Aneurysm & Phong shader & side-by-side
  3. Liver Tumor & Phong shader & successive
  4. Liver Tumor & Phong shader & side-by-side
- SEQUENCE C
  1. Aneurysm & outlines with feature lines & side-by-side
  2. Aneurysm & outlines with feature lines & successive
  3. Liver Tumor & outlines with feature lines & side-by-side
  4. Liver Tumor & outlines with feature lines & successive
- SEQUENCE D
  1. Aneurysm & outlines with feature lines & successive
  2. Aneurysm & outlines with feature lines & side-by-side
  3. Liver Tumor & outlines with feature lines & successive
  4. Liver Tumor & outlines with feature lines & side-by-side
- SEQUENCE E
  1. Aneurysm & hatching with Fresnel shader & side-by-side
  2. Aneurysm & hatching with Fresnel shader & successive
  3. Liver Tumor & hatching with Fresnel shader & side-by-side
  4. Liver Tumor & hatching with Fresnel shader & successive
- SEQUENCE F
  1. Aneurysm & hatching with Fresnel shader & successive
  2. Aneurysm & hatching with Fresnel shader & side-by-side
  3. Liver Tumor & hatching with Fresnel shader & successive
  4. Liver Tumor & hatching with Fresnel shader & side-by-side

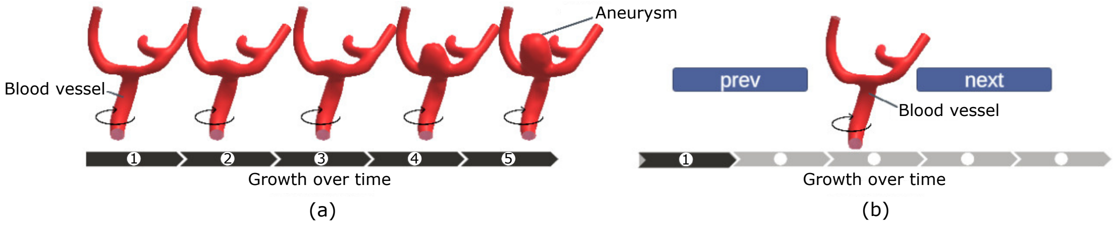


Figure 20: Presentation techniques: (a) time steps are displayed simultaneously side-by-side, and (b) time steps are displayed successively. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

Each sequence was evaluated by ten participants, ensuring that each visualization technique was assessed by a total of 20 participants. Careful measures were taken to prevent bias due to learning effects or the pairing of easier or more complex datasets with a specific visualization technique. To achieve this, each of the three visualization techniques (Phong, outlines with feature lines, and Fresnel with hatching) was assigned to only one subgroup, while all participants were presented with the same medical datasets. To enable a comparison of the two time step presentation formats (side-by-side and successive) without introducing a learning bias, two evaluation versions were created for each visualization technique. One version started with the side-by-side view, while the other began with the successive presentation.

TASKS. Three types of tasks were selected for participants to complete, as shown in Figure 21. A multiple-choice approach was used, requiring participants to select the correct answer from three options. For the aneurysm datasets, participants were asked to identify the overall *growth rate* across all time steps and determine the *growth*

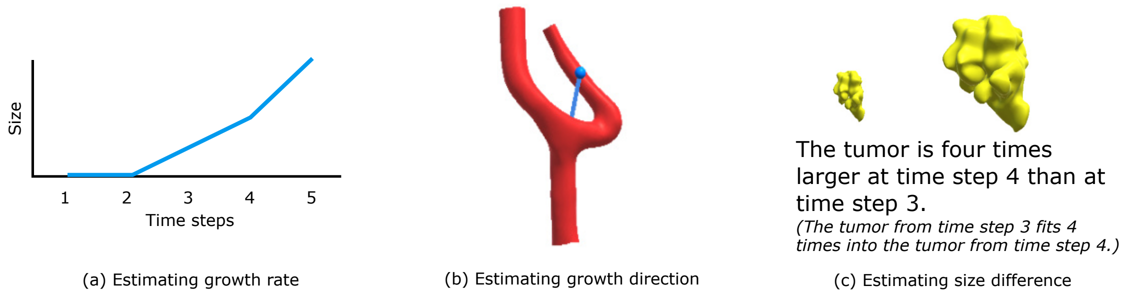


Figure 21: Examples of the three types of user tasks: (a) a piece-wise linear function depicting the growth rate, (b) growth direction indicated by a pin, and (c) description of the difference in the size of two adjacent time steps. The tasks were structured as multiple-choice questions, each providing three alternative options from which participants had to select the correct answer. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

direction, as these were identified as key characteristics of aneurysm progression. Since tumors can expand in multiple directions without a clear growth trajectory, participants instead assessed the *growth rate* and the *size difference* between two consecutive time steps for each tumor dataset. After selecting an answer, participants indicated their confidence level using a 5-point Likert scale, ranging from *very confident* to *not confident at all*. Finally, participants were asked whether they preferred the side-by-side or successive depiction of time steps for task-solving and aesthetic purposes.

Each task type included three different versions, randomized to prevent order effects. Additionally, the specific answer options, as well as the correct answer and its corresponding model, were randomized for each participant to minimize bias. At the beginning of the evaluation, participants were introduced to an example task to familiarize them with the task design before proceeding to the actual evaluation.

Task Difficulty

When designing tasks within a user study, their level of difficulty needs to be carefully designed to avoid a ceiling and floor effect (CFE). While a ceiling effect occurs when almost all participants are able to solve the tasks correctly (i.e., because the tasks are designed too easy), a floor effect describes the opposite, that almost all participants provide wrong answers (i.e., because tasks are designed too difficult) [218]. CFE hinders a comparison of groups, as no group differences can be seen based on participants' responses.

6

GROWTH RATE OF ANEURYSMS AND TUMORS. Participants were tasked with determining the growth rate across all time steps by selecting the correct function from three options, as shown in Figure 21 (a) and Figure 22. Thereto, participants were shown graphs of piecewise linear functions depicting different growth rates within one data set and had to choose the correct growth rate among them. Three types of (incorrect) answer options were designed with varying difficulty to prevent the CFE. A growth rate was considered more difficult, if it changes often. In contrast, a consistent growth was considered easy. Since plateaus – indicating no growth between two or more time steps – were easier to identify than line segments representing growth, their number was limited to prevent the tasks from becoming too simple. The different versions of the incorrect answer options ordered by increasing difficulty are:

LOW DIFFICULTY: maximum of two line segments

- meaning there is up to one change in growth rate
- maximum of two plateaus

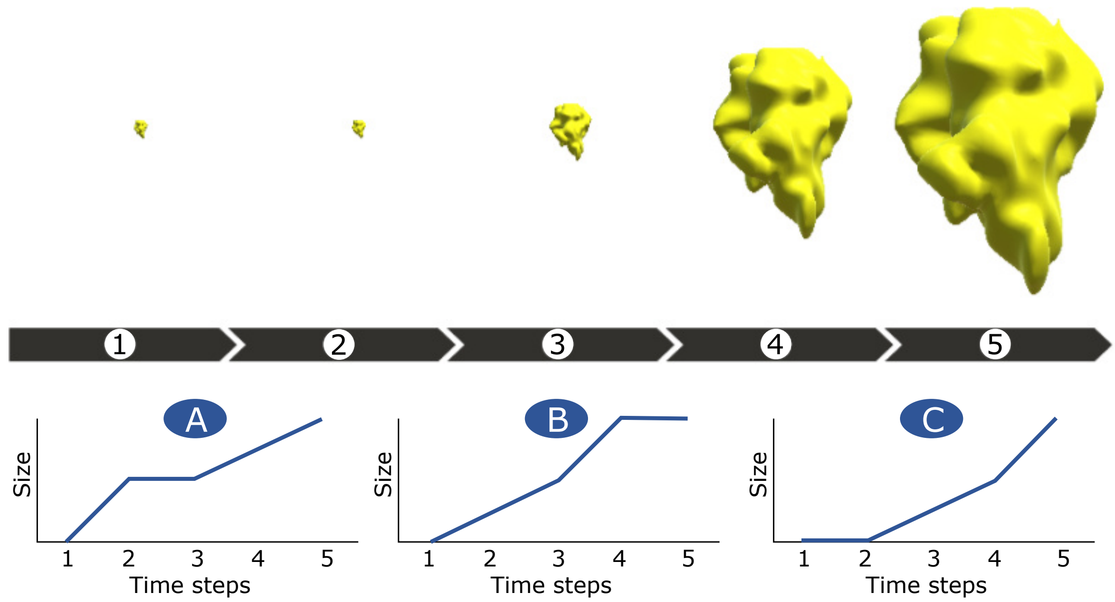


Figure 22: Participants were presented with different time steps and had to select the graph that shows the correct growth rate out of three options. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

6

MEDIUM DIFFICULTY: exactly three line segments

- meaning there are two changes in growth rate
- maximum of one plateau

HIGH DIFFICULTY: more than three line segments

- meaning there are ore than two changes in growth rate
- no plateaus

GROWTH DIRECTION OF ANEURYSMS. To determine the growth direction, participants selected one of three blood vessel models, where the aneurysm was replaced with a pin indicating the main growth direction, as shown in Figure 21 (b). Again, the three types of incorrect answer options were designed with increasing difficulty. Low difficulty answer options showed a deviation from the correct growth direction in all three spatial directions, which is supposedly easier to detect. On the other hand, answer options with a high difficulty are more similar to the correct answer, where the indicated growth in the incorrect answers only deviates from the correct answer in only one spatial direction. In summary, the incorrect answer options ordered by increasing difficulty are:

LOW DIFFICULTY: Deviation from correct answer in all three spatial directions.

MEDIUM DIFFICULTY: Deviation from correct answer in two spatial directions from the correct answer.

HIGH DIFFICULTY: Deviation from correct answer in only one spatial direction from the correct answer.

A deviation of at least 40 degrees was chosen after testing various values and discussing the alternatives within the group of co-authors of the respective paper [160]. Smaller angles were decided to be too hard to notice.

SIZE COMPARISON OF TUMORS. To assess how well participants could identify changes in volume, they were presented with a subset of two consecutive time steps from each tumor dataset, as shown in Figure 21 (c). The task required them to determine the difference in size (volume) between the two time steps. The three types of incorrect answer options were classified by their deviation from the correct answer:

- The correct answer was either multiplied or divided by three, or by a multiple of three, ensuring that the correct answer was not always the median value.
- Multiples of two were applied.
- Multiples of 1.5 were applied.

Unlike the other tasks, difficulty levels could not be clearly assigned to the answer options, as it was not possible to determine whether identifying a size change factor of 1.5, 2, or 3 was inherently more challenging. Furthermore, there are many overlapping values for multiples of 1.5 and 3.

6.4.3 *Discussion of Results*

The results related to task performance, completion times, and user preferences are presented, while also addressing certain study limitations.

GROWTH RATE. Estimating the growth rate across all datasets resulted in fewer incorrect answers compared to the other two tasks, as shown in Figure 23. Participants who viewed the hatching models in a side-by-side format had the lowest accuracy, whereas those who saw the outline model with successive time steps achieved the highest accuracy. A notable observation is that, for the hatching visualizations, participants selected more incorrect answers of easy difficulty compared to the other visualization techniques. This may be due to the increased visual complexity introduced by the hatching lines, potentially causing confusion. Across all visualization techniques, incorrect answers of the highest difficulty level accounted for the majority of mistakes when estimating the growth rate. Interestingly, confidence levels were lowest for the two versions using Phong shading, despite participants achieving a high accuracy rate. Overall, the differences in performance between the six visualization techniques were minor, suggesting that while visualization style had some influence, correct completion of tasks remained relatively consistent across conditions.

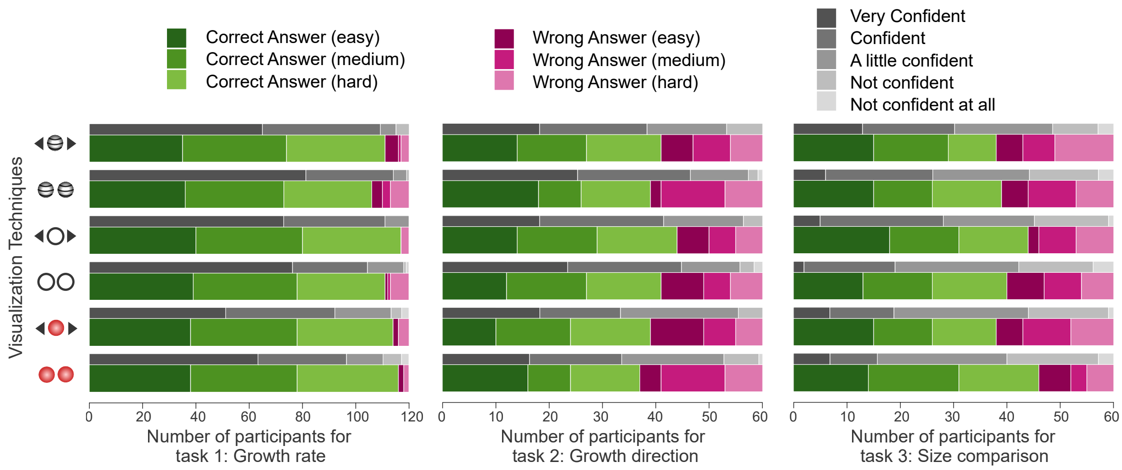


Figure 23: Response options selected by participants, categorized into three levels of difficulty, along with participants' self-assessed confidence in their responses. Each row corresponds to one visualization technique, listed from bottom to top as follows: Phong side-by-side, Phong successive, outline side-by-side, outline successive, Fresnel with hatching side-by-side, and Fresnel with hatching successive. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

Participants noted that plateaus in the graphs depicting the growth rate (see Figure 22) made the tasks relatively easy, particularly when the plateau appeared at the beginning. This may explain why overall performance was higher for the growth rate task compared to the others. Since plateaus were positioned differently in each answer choice (sometimes in the first time steps, sometimes in the middle time steps and sometimes in the last time steps), participants were often able to determine the correct answer simply by identifying the correct placement of the plateau rather than analyzing the entire growth trend. Additionally, the growth rate task could be completed without rotating the models. Observations from participants at the *Long Night of Science* indicated that many did not utilize the rotation feature. This suggests that 3D interaction may not be intuitive for individuals unfamiliar with virtual 3D environments. Despite receiving instructions on interaction options at the beginning of the evaluation, many users did not take advantage of the ability to rotate the models, possibly due to a lack of familiarity with interactive 3D navigation.

GROWTH DIRECTION. For the growth direction tasks, participants performed better when using the successive view across all shading techniques. The highest accuracy was observed among participants who viewed the successive outline visualization, see Figure 23. Confidence levels for growth direction answers were generally higher in the successive view, except for the versions with Phong shading. However, higher confidence did not necessarily correlate with a greater number of correct answers. The overall number of incorrect answers was notably higher for growth direction estimation compared to growth rate estimation. This is likely because many participants did not utilize the rotation feature. Identifying the correct growth direction is challenging without rotating the model, as crucial spatial information may not be visible from a single viewpoint. No clear trend was observed regarding the difficulty level of incorrect answer choices, as participants frequently selected incorrect answers across all difficulty levels.

SIZE COMPARISON. When comparing the size of two 3D models, participants achieved the highest accuracy using the side-by-side presentation with the Phong shader, followed closely by those using the successive view with the outline shader. When comparing the side-by-side and successive views within the same shading technique, participants generally identified more correct answers with the side-by-side view. However, they reported higher confidence in their answers when using the successive view. The only exception was the successive outline visualization, which resulted in more correct answers than its side-by-side counterpart. Estimating size ratios of 3D objects is inherently challenging and strongly influenced by the objects' shape [104]. Given this complexity, lower performance in this task was expected. This also explains why participants reported lower confidence levels for this task compared to the other two. Despite the challenge, the majority of tasks were answered correctly, indicating that participants were still able to make reasonable size estimations using the provided visualizations.

Confidence was lowest for answer options scaled by a factor of 1.5, suggesting that participants perceived this variant as the most difficult. Answer options scaled by a factor of two received slightly lower confidence ratings compared to those scaled by a factor of three. This pattern aligns with the distribution of incorrect answers, as more false selections occurred with answer options scaled by factors of 1.5 and two. These findings suggest that distinguishing smaller differences in size ratios was more challenging for participants than identifying larger discrepancies.

GENERAL REMARKS ON TASK RESULTS. The highest overall accuracy was achieved when outlines were used as the visualization technique and time steps were presented successively. One possible explanation is that the outline visualization, which lacks shading and therefore depth cues, may have encouraged participants to use the rotation interaction more frequently. This increased interaction could have led to improved accuracy, particularly for identifying growth direction. However, participants who viewed the outline visualizations in a side-by-side format performed worse. This suggests that the improved performance cannot be attributed solely to the outline shading but rather to the combination of the shading technique and the successive view. The visualization technique that resulted in the highest accuracy varied depending on the task. This suggests that the choice of representation format in future visualizations should be guided by the intended communication goal. For example, a side-by-side view is more effective for conveying size ratios, whereas a successive view with outlines may be better suited for tasks requiring growth direction analysis.

Interestingly, no notable differences were observed between the six combinations of shader techniques and presentation of time steps used in this study. Participants' performance remained relatively consistent across all techniques. This suggests that the complexity of the surface models and the inherent challenges of the tasks had a greater impact on performance than the specific rendering methods. An analysis of participants' confidence levels revealed that confidence did not always align with accuracy. In some cases, participants exhibited uncertainty in their judgments despite selecting the correct answers. This discrepancy raises the question of whether the visualization style itself contributed to lower confidence or if other factors, such as task difficulty or individual differences in spatial reasoning, played a role. Further investigation is needed to determine the underlying causes of this effect.

USER PREFERENCES. Users were asked to rate their preferred presentation style for time steps in terms of both task-solving effectiveness and aesthetics (see Figure 24). Preferences for task-solving were nearly evenly split between the side-by-side and successive presentation formats, with a slight preference for the former. However, a greater number of participants favored the successive display of time steps for aesthetic reasons. This suggests that some users who preferred the side-by-side format for task-solving found the successive view more visually appealing, highlighting the challenge of aligning user preferences for functionality and aesthetics. Nonetheless, most users preferred the same visualization type for both aspects, aligning with findings by Garrison et al. [71], which indicate that users tend to favor visualizations they perceive as best supporting the intended communicative objective.

TIME MEASUREMENT. The time required for participants to complete the evaluation was recorded, and the measurements for each version were aggregated by calculating the mean and standard deviation (see Figure 25). A noticeable trend was observed across all shaders: participants took longer to complete the first two scenes, as well as scenes 20–22. These scenes corresponded to the introduction of new tasks, suggesting that additional time was needed for participants to familiarize themselves with the task requirements before proceeding. Participants required less time to complete the tumor-related tasks compared to the aneurysm tasks. This difference may be attributed to a learning effect, as tumors were always presented in the second half of the evaluation, allowing participants to become more familiar with the task structure over time. Additionally, tumors may have been easier to assess at first glance since they were depicted in isolation, without neighboring anatomical structures. In contrast, aneurysms were always attached to a section of the vascular tree, which may have introduced additional visual complexity and made the analysis more challenging.

The evaluation sequences did not reveal any major differences in completion times. On average, participants who started with the successive view completed the evaluation slightly faster. One exception was the version using the Fresnel shader with additional hatching, which began with a successive view and showed slightly higher completion times compared to other successive-view versions. However, this discrepancy was attributed to a single participant who took notably longer throughout the entire evaluation process, rather than a general effect of the visualization technique. Participants' familiarity with computers and 3D interaction notably influenced the time required to complete the tasks. Additionally, age appeared to affect task completion strategies. Older participants tended to take more time, carefully considering all available information and occasionally asking questions. In contrast, younger participants generally aimed to maximize their number of correct answers while minimizing completion time. This difference may be attributed to younger participants' greater proficiency with computers and 3D interaction, enabling them to navigate the tasks more efficiently.

6

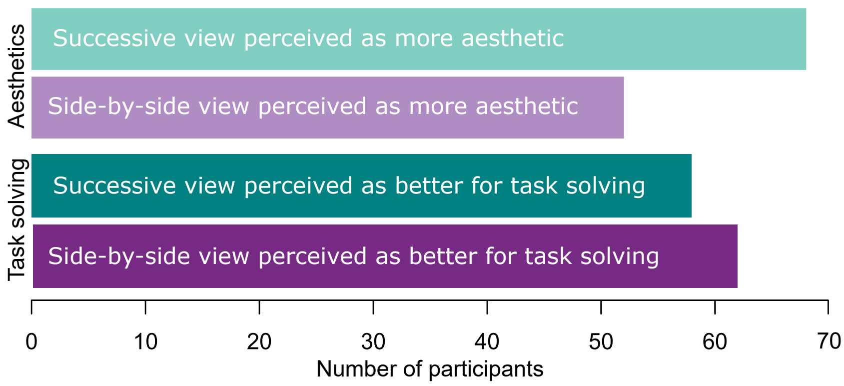
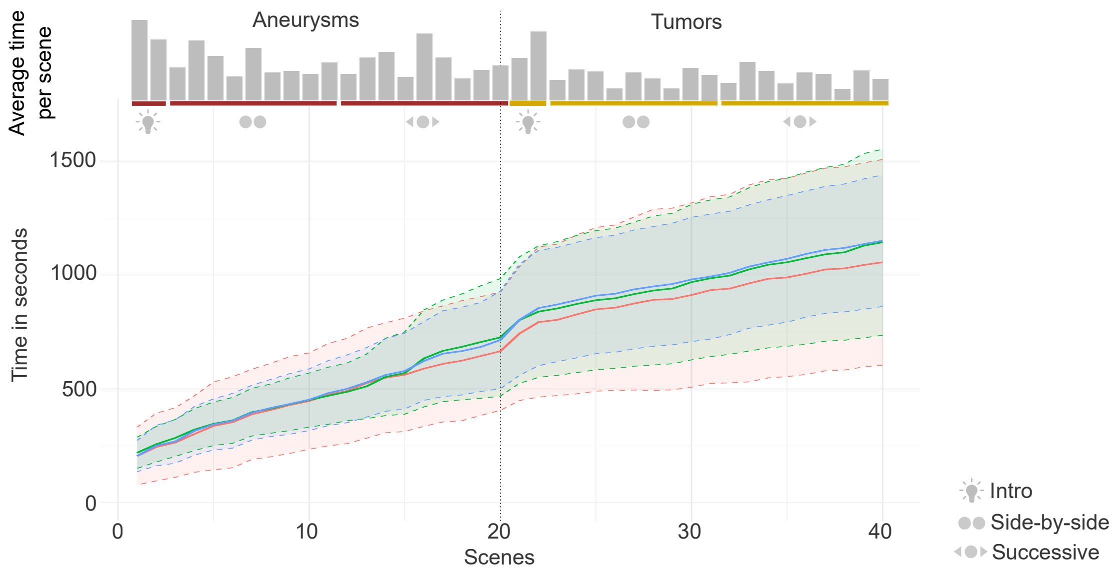
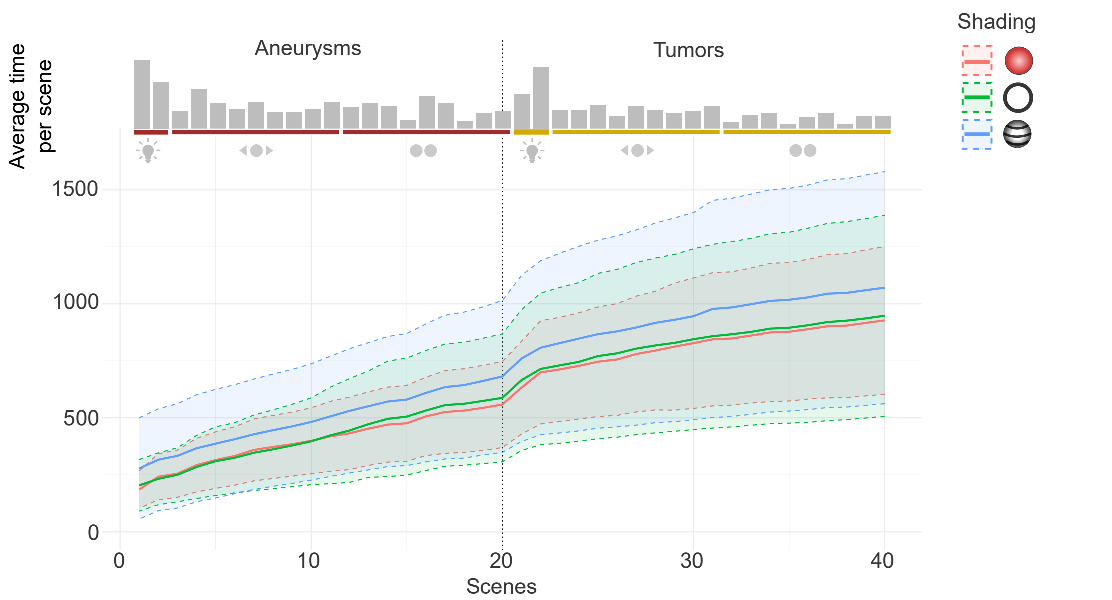


Figure 24: Participants' preferences regarding the presentation techniques. Since each participant responded to both questions for each pathology, a total of 120 responses were collected for task-solving preferences and another 120 for aesthetic preferences. Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.



(a) Starting with side-by-side



(b) Starting with successive

Figure 25: The line graphs display participants' completion times (mean and standard deviation). (a) shows sequences starting with the side-by-side presentation, while (b) shows those starting with the successive presentation. Colors indicate visualization techniques: Phong (red), outline (green), and Fresnel with hatching (blue). The bar chart above shows average time spent per scene, for aneurysms (red) and tumors (yellow). Image from Mittenentzwei et al. [160], available under a CC BY 4.0 license. No changes were made.

6.5 SECOND CASE STUDY TO COMMUNICATE GROWTH AND SHRINKAGE

Based on feedback from clinical partners and findings from the previous study, this follow-up study conducts a detailed analysis of the differences between illumination-based and illustrative visualization techniques, as well as the visualization of both growth and shrinkage. Again, the design space for visualizing changes in medical datasets is vast, and this study does not aim to cover it comprehensively. Instead, the focus is on an initial exploration of different stylistic approaches, specifically illumination-based and illustrative methods, to guide future design studies.

6.5.1 *Adaptions from the First Case Study*

This study extends previous research [160]. The adapted study design is described in the following, highlighting modifications made to the prior case study in Section 6.4. The first case study showed each participant only one visualization technique, but both presentation techniques (side-by-side and successive), making it impossible to ask participants directly about their preferences regarding the visualization techniques. Therefore, this study will be structured differently, focussing on a comparison of visualization techniques.

VISUALIZATION AND PRESENTATION OF TIME STEPS. This study compares two visualization techniques instead of three, eliminating the combination of hatching and the Fresnel shader used previously. This approach allows for a more focused evaluation of the differences between an illumination-based shader and an illustrative shader. Additionally, the combination of hatching lines and the Fresnel shader made it difficult to determine whether the results were influenced more by one technique than the other. In line with the argumentation from the prior study, illumination-based shading was selected because it is a classic, widely used, and easily understandable approach for visualizing 3D data [102]. In contrast, illustrative techniques mimic pencil sketches, a style that is familiar even to laypeople [71, 125].

Additionally, the time steps are displayed only in a side-by-side format, whereas the previous study included both a side-by-side view and a sequential view, in which only one time step was displayed at a time. This adjustment makes the study more concise and enables a more detailed comparison between the two shaders, while the last study already provided a direct comparison of the two presentation techniques. Unlike the previous study, multiple shaders can now be shown to each participant. Keeping the study streamlined is essential, as each additional shader or presentation technique would notably increase its duration, making participant recruitment more challenging and raising the likelihood of dropouts. Moreover, longer study durations can lead to decreased participant concentration. Thus, this study represents a trade-off between the amount of content and the quality of results. As new aspects were incorporated,

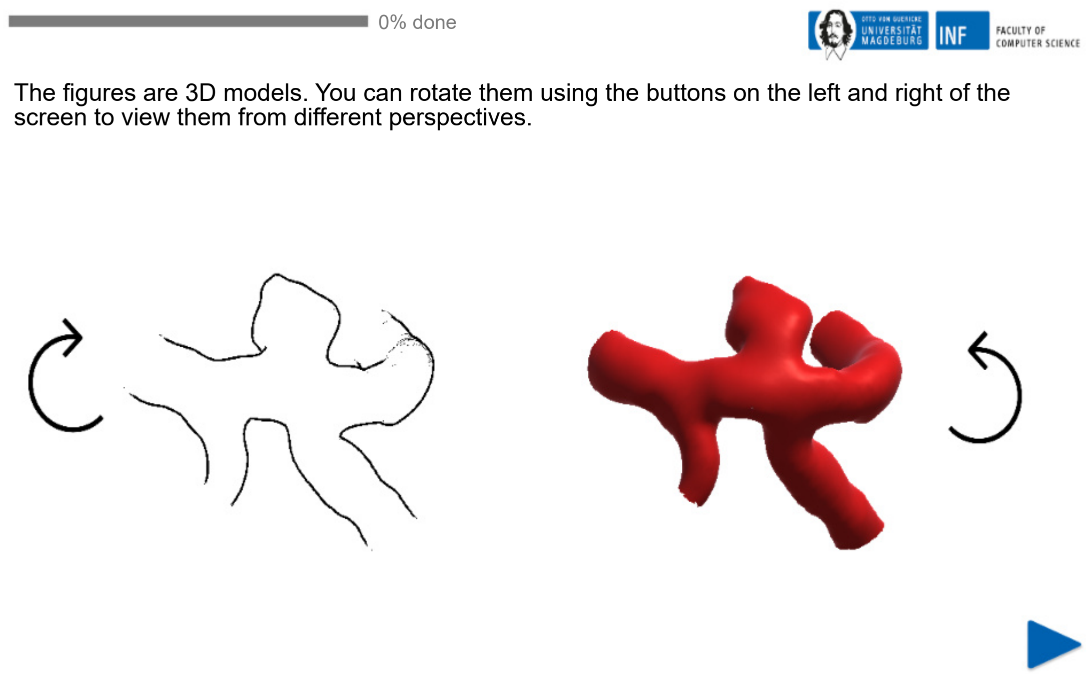


Figure 26: Screenshot from the second case study explaining rotation.

one visualization technique and the successive presentation technique were removed to maintain study efficiency and participant engagement.

INTERACTION. Based on findings from the previous study, where many participants did not use the rotation feature, this interaction technique was implemented differently in the current study. Rotation remains possible around a single axis. In the previous study, participants rotated the models by clicking and dragging the mouse within the 3D scene. In this study, rotation is controlled using buttons instead, see Figure 26. Buttons were chosen as they are widely used in various applications and are familiar to most users. Two buttons are displayed, one on each side of the screen, allowing for rotation in the left or right direction. Each button features a curved arrow, a commonly recognized symbol for rotation, to enhance intuitiveness and usability.

TASK DESIGN. While the previous study focused mainly on tasks involving a quantitative analysis and subsequent measurement of results, this aspect comprises only half of this study. Participants are again tasked with estimating whether the growth or shrinkage of pathologies follows a linear or non-linear trajectory, as well as quantifying the volume changes between successive time steps. A consultation with medical collaborators underscored the importance of effectively communicating growth rates to lay audiences.

The study design was revised together with the two physicians mentioned in Section 6.3. According to their insights, aneurysms and tumors grow heterogeneously, and many unanswered questions remain in medicine, particularly regarding aneurysm treatment. However, these complexities are not particularly relevant when communicating with the general public. Instead, understanding the aspects of growth rate and size ratios between successive time steps is more important for conveying the growth process. A high growth rate indicates the aggressiveness of the pathology, while maintaining the anatomical context is essential for presenting a coherent overall picture. Additionally, explaining shrinkage processes is crucial. These visualizations can provide a more positive perspective and help illustrate and motivate treatments such as chemotherapy. Based on these expert recommendations, the following aspects were incorporated into the study:

- Tumor and aneurysm growth and shrinkage are included, with treatment methods briefly explained to contextualize and motivate shrinkage.
- Study tasks focus on the perception of growth rate and size ratios.
- Anatomical context is provided.

The remaining half of the evaluation is dedicated to gathering participants' self-reported feedback on the visualization methods using keywords. Furthermore, participants had to directly state which their favorite visualization technique is. This approach provides additional insights into their perceptions of the employed visualization techniques, which were not included in the prior study.

INCLUSION OF VIDEOS. Short explanatory videos were included to help participants better understand the anatomical context and application examples, highlighting the medical relevance of the study. These videos were created and recorded in Blender 4.0. The videos incorporate a mini-story introducing a character with a corresponding pathology, see Figure 27. A 3D model of a human is used, progressively zooming in until the pathology becomes visible. Each transition and pause lasts between two and seven seconds, depending on the complexity of the structure in view and whether text is displayed. To illustrate growth, the first time step represents the healthy structure. Conversely, to demonstrate shrinkage, the video begins with the final time step, depicting the fully developed aneurysm or tumor. With one time step for each dataset used in the videos, only the four remaining time steps were included in the tasks.

6.5.2 *Participants*

A total of 40 participants were gathered through the university mailing list, which included a request for further distribution of the study. Participants were asked about their age, gender, highest education level, and professional experience with medical

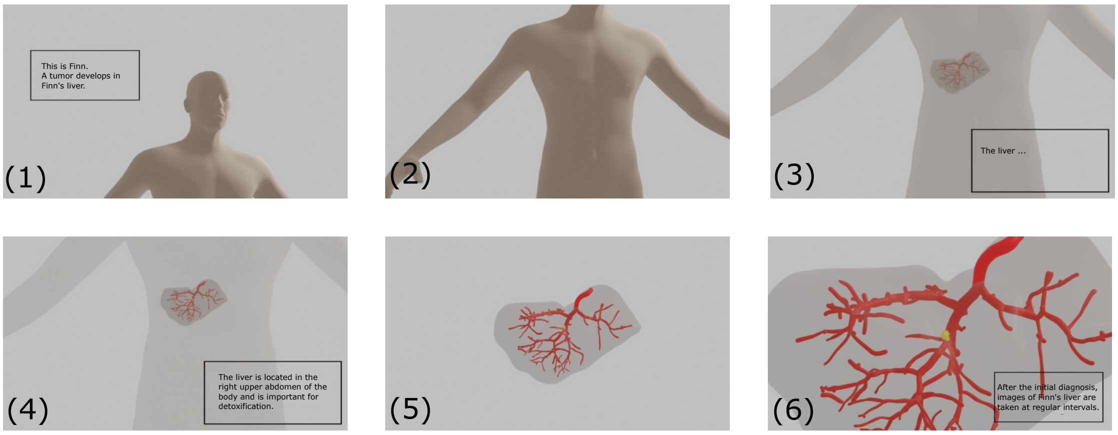


Figure 27: Sample frames from a video illustrating a fictional character's liver tumor diagnosis. The sequence begins with (1) a close-up of the head, then (2) pans downward and (3–6) progressively zooms in on the liver, revealing the tumor while preserving the anatomical context. Text boxes provide information about the character, the disease, and the affected organ. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

topics and three-dimensional visualizations (see Table 2). The majority of participants (63 %) were 35 years old or younger. 60 % were female, while one participant identified as diverse or preferred not to answer. A university degree was held by 55 % (22 participants). More than half of the participants reported that they never or rarely engaged with 3D visualizations professionally. Additionally, 25 % (10 participants) indicated that they frequently deal with medical topics. All participants were from Germany, and one reported a red-green color vision deficiency. Overall, the participant group exhibited an above-average level of education and expertise in 3D and medical visualization, which is likely to influence the evaluation results. The potential biases introduced by this demographic profile are discussed in more detail in Section 6.6.

6.5.3 *Study Design*

To evaluate the visualization techniques, a between-subjects design study was conducted. An online study was set up, in which each participant completed four tasks for the aneurysm datasets and four tasks for the tumor datasets. For each task, participants were required to indicate their confidence in their answers. The study ran for approximately one month. Four evaluation sequences were created by presenting the visualizations in different orders to minimize bias. Each visualization technique was shown in two sequences but in a different order. Each participant viewed only one of the following sequences:

Age	
18-25	12
26-35	13
36-45	1
46-55	4
56-65	7
>65	3

Gender	
Male	15
Female	24
Diverse	1
No answer	0

Education	
Intermediate school	3
High school diploma	8
Vocational training	7
University degree	22

Professional contact with med. topics	
Never	13
Rarely	8
Sometimes	7
Often	2
Very often	10

Professional contact with 3D vis.	
Never	17
Rarely	6
Sometimes	5
Often	0
Very often	12

Table 2: Participant metadata.

- VARIANT A
  1. Aneurysm, Phong shader, Growth
  2. Aneurysm, Outline shader and feature lines, Shrinkage
  3. Liver Tumor, Phong shader, Growth
  4. Liver Tumor, Outline shader and feature lines, Shrinkage
- VARIANT B
  1. Aneurysm, Outline shader and feature lines, Growth
  2. Aneurysm, Phong shader, Shrinkage
  3. Tumor, Outline shader and feature lines, Growth
  4. Tumor, Phong shader, Shrinkage
- VARIANT C
  1. Aneurysm, Phong shader, Shrinkage
  2. Aneurysm, Outline shader and feature lines, Growth
  3. Tumor, Phong shader, Shrinkage
  4. Tumor, Outline shader and feature lines, Shrinkage

6

- VARIANT D
  1. Aneurysm, Outline shader and feature lines, Shrinkage
  2. Aneurysm, Phong shader, Growth
  3. Tumor, Outline shader and feature lines, Shrinkage
  4. Tumor, Phong shader, Growth

Each sequence was evaluated by ten participants, resulting in a total of 20 participants per visualization technique. To prevent results from being influenced by a learning effect or by pairing easier or more complex datasets with a specific visualization technique, all participants were shown the same medical datasets. To compare the Phong and Outline visualization techniques without introducing learning bias, two evaluation versions were developed for each method, one starting with the Phong shader and the other with the Outline shader.

TASKS. Two different tasks were selected for participants to complete. The first task followed a multiple-choice approach, requiring participants to choose the correct answer from two or three options. Participants were first asked to assess whether the overall changes in growth or shrinkage occurred linearly (evenly) or non-linearly (unevenly), as shown in Figure 28 (a).

In a follow-up question, participants were asked to determine the extent to which the pathology had grown or shrunk for each pair of consecutive time steps (see Figure 28 (b)). This assessment provided insights into the perceived growth rate and size ratios.

For each selected answer, participants indicated their level of confidence using a 5-point Likert scale, ranging from *very confident* to *not confident at all*.

The second task required participants to assign adjectives to the visualizations based on their opinions (see Figure 29). This task aimed to provide qualitative feedback on the visualizations. Participants were also given the option to suggest their own adjectives. Finally, participants were asked whether they preferred the Phong shader (color) or the Outline shader (contours) for both task performance and aesthetics. Since the study was conducted with non-experts, the terms Phong and Outline were replaced with color and contour in the queries to enhance clarity and comprehension.

Rotating the 3D models is often necessary to complete the tasks, as pathologies may be obscured or difficult to recognize. Rotation buttons were implemented, allowing participants to hold them down to rotate the models left or right. For each task, usage of the rotation function was tracked. The following section provides a detailed description of the task design.

SIZE ESTIMATION TASK. The possible answers for estimating the change in size between two successive time steps are randomized for each case. The answer option indicating no change is always included. The second option is always the correct answer, while the third is either twice or half the correct value. The positions of the answer options, as well as whether twice or half of the correct answer appears as the last option, are randomized. At the beginning of the evaluation, participants are presented with an example task. Including the "no change" option provides insight into whether participants can detect size changes, even if they misjudge the size ratios, and vice versa.

KEYWORD SELECTION TASK. A collection of twelve keywords, including both positive and negative adjectives, was provided for participants to select those they felt best described the visualization they were presented with. The keywords available in the second task were based on the word cloud from Garrison et al. [71]. Participants could select as many keywords as they wished, with the additional option to provide comments on the visualizations in a free-text field. It was also possible to choose none of the keywords.

6.5.4 *Discussion of Results*

The following section discusses the results related to task performance, the use of interaction methods, and participant preferences, while also considering certain study limitations.

TASKS. Participants answered multiple-choice questions regarding overall growth and shrinkage rates, as well as size changes between subsequent time steps. Out of a total of 480 questions (120 per sequence), 369 were answered correctly, while 111 were answered incorrectly. Additionally, they selected keywords that best described the visualization.

GROWTH RATE. Each participant evaluated the overall rate at which the displayed pathologies changed in size. To enhance accessibility for the lay audience, the terms *linearly* and *non-linearly* were substituted with *evenly* and *unevenly* in the evaluation. Participant feedback highlighted variations in how these terms were interpreted. While some associated evenly with a continuous increase without plateaus, others correctly identified that evenly was supposed to mean the pathology expanded by the same factor at each time step. Due to the lack of explicitly defined terminology, the reliability of conclusions drawn from this data is somewhat constrained. Errors were consistently observed across all tasks and shaders, with approximately 34 % of responses incorrectly selecting even instead of uneven. However, it remains uncertain whether these inaccuracies were driven by variations in perception or a fundamental misunderstanding of the task.

SIZE COMPARISON. In the size comparison task, each participant responded to three questions assessing differences between two consecutive models. The correlation

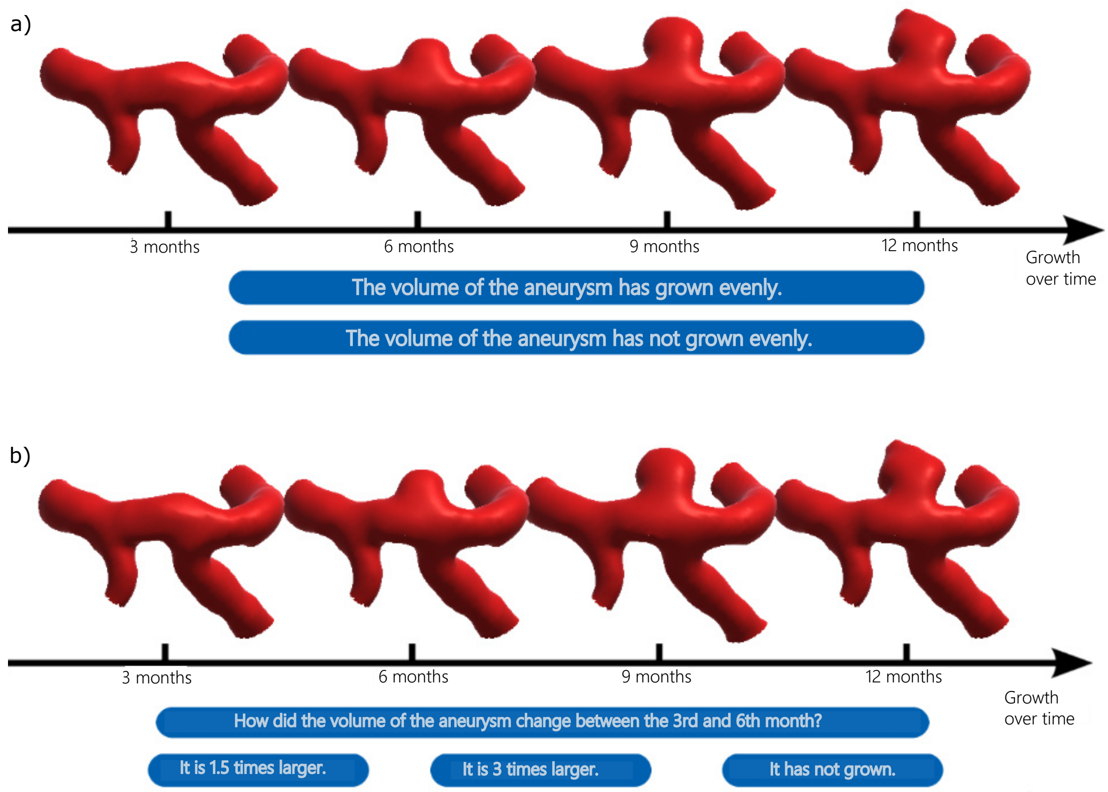


Figure 28: First task of the user study. (a) Participants were asked to determine whether the overall growth was linear (even) or nonlinear (uneven). (b) They then estimated how much the pathology had grown or shrunk between each pair of consecutive time steps. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

between error rate and the visualization technique used is depicted in Figure 30. Additionally, the study examined how the nature of size change (whether the pathology grew or shrank) related to the type of pathology.

Neither the visualization technique nor growth vs. shrinkage had a noticeable impact on the error rate. Tasks utilizing the Phong shader resulted in 56 incorrect responses, while those employing outlines and feature lines had 55. Similarly, the number of errors in tasks involving growth (56) was nearly identical to those related to shrinkage (55). However, differences emerged when comparing the types of pathologies. Participants made fewer errors in aneurysm-related tasks compared to tumor-related ones. One possible explanation is that the tumor was displayed in relation to the surrounding liver. While experts emphasize that preserving anatomical context is crucial for understanding growth patterns, it may also introduce visual challenges, such as obscuring parts of the pathology behind the vascular tree.

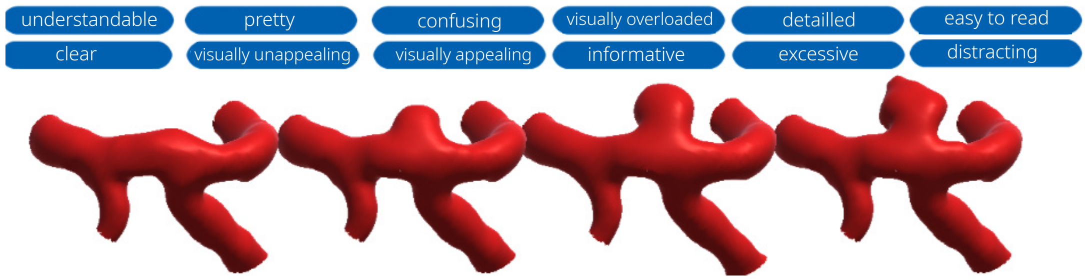


Figure 29: Second task of the user study. Participants were asked to assign adjectives to the visualization based on their personal impressions. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

Tasks in which the pathology remained the same size yielded the highest accuracy, with only seven errors out of 80 responses. The choice of shader did not appear to be a determining factor, as mistakes were almost evenly distributed between the two (three and four errors, respectively). However, two specific questions caused confusion regardless of the visualization technique used. To explore this further, responses were analyzed to determine whether participants had a tendency to underestimate or overestimate changes in pathology size (see Figure 31).

One question, that seems to be particularly challenging, asked participants whether the pathology had increased in size by a factor of 32 or 64. This question was answered incorrectly 16 times, with all errors resulting from an overestimation by a factor of two. Estimating size ratios of three-dimensional objects is inherently difficult and heavily influenced by an object's shape [104]. Since the pathology's shape changed strongly as it increased in size, this may have contributed to the high frequency of incorrect responses. Once again, the visualization techniques did not noticeably impact accuracy, as both visualization techniques resulted in an equal number of mistakes (eight each).

While variations in shape may explain some fluctuations in error rates, the second case exhibited even lower accuracy. Notably, this was the only instance where incorrect responses outnumbered correct ones for both shaders. In this scenario, the pathology had decreased in size by a factor of four. However, slightly more than half of the participants incorrectly estimated the reduction as only a factor of two (see Figure 31, Phong and Shrinkage). Interestingly, participants had less difficulty differentiating between factors of three and six, despite both involving a doubling in size – an outcome that might intuitively seem similar.

Furthermore, variations in size estimation could be influenced by the characteristics of the pathologies. The liver tumor exhibited relatively uniform changes in size, whereas the aneurysm underwent more localized volume alterations, potentially making it more difficult for participants to accurately assess its size reduction.

6

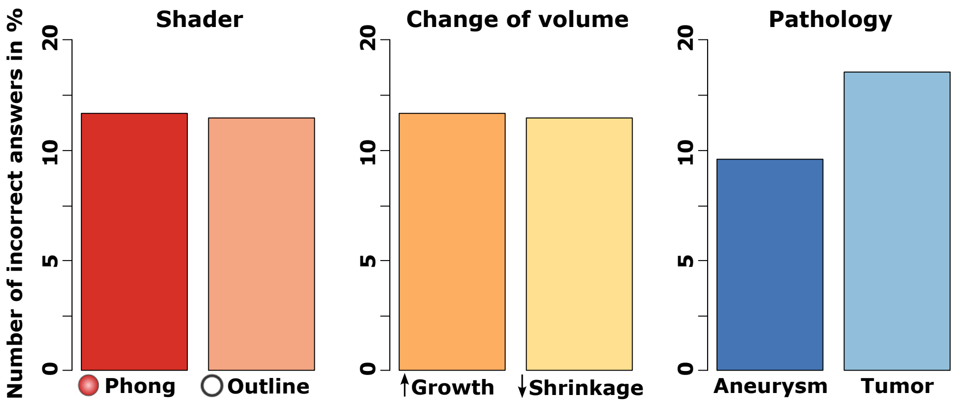


Figure 30: Number of incorrect answers per category, based on a total of 480 questions. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

A notable outlier was observed in the aneurysm growth task, where size changes were incorrectly identified in six out of 20 cases for both shaders. This discrepancy may be linked to the model's initial orientation, with the pathology facing the user. As a result, growth occurring perpendicular to the screen may have been less perceptible, leading to underestimation when participants did not utilize the rotation feature.

Among the six participants who provided incorrect responses, only two used the rotation function, which may have contributed to the higher error rate.

Since rotation was restricted to a single axis, some participants may have struggled to find an optimal viewing angle to accurately assess size changes. Even when the size change was pronounced – such as an increase by a factor of 32 – some participants still selected "no size change" (see Figure 31). They may have chosen their response as a substitute for "no answer," as the study did not offer an option to skip a question.

CONFIDENCE. After completing each task, participants rated their confidence levels on a 5-point Likert scale, ranging from 1 (very confident) to 5 (not at all confident). The error rates corresponding to each visualization technique are illustrated in Figure 32. Overall, the majority of participants expressed confidence in their answers, with 113 out of 160 responses indicating *confident* or *a little confident*. The *very confident* option was selected 11 times, while *not confident* and *not at all confident* were chosen 26 and 10 times, respectively.

In tasks using the Phong shader, participants reported very confident responses more frequently for growth-related assessments compared to outlines with feature lines. Conversely, in tasks utilizing the latter, *not at all confident* was selected less often. However, the overall difference in confidence levels between the two visualization techniques was minimal. Confidence levels varied only slightly depending on the type

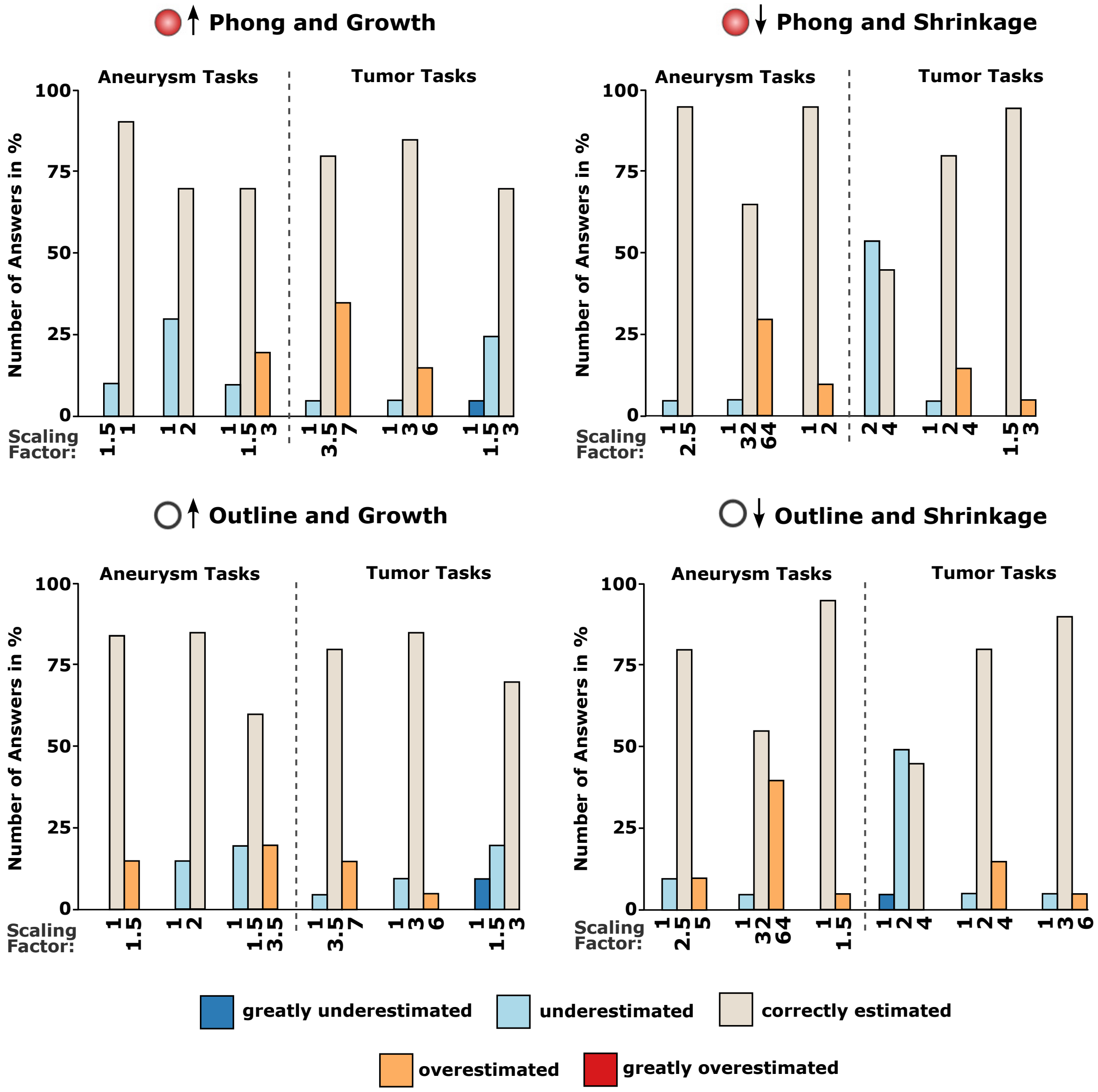


Figure 31: Correct and incorrect responses for each task, categorized by shader and growth or shrinkage. Each combination includes three tasks, each completed 20 times. The scaling factors represent the options selected by participants, indicating how much the pathology was perceived to have grown or shrunk compared to the previous time step (e.g., a factor of 1.5). For each task, these scaling factors illustrate the relationship between correct answers and instances where participants underestimated or overestimated the actual volume change. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

	All tasks correct	1 task incorrect	2 tasks incorrect	All tasks incorrect
Rotation	40	32	4	1
No Rotation	34	36	9	4

Table 3: Relationship between correctness of answers and the use of rotation. For each pathology, participants completed three tasks.

of size change. Participants expressed confidence more frequently in growth-related tasks (36 instances) than in shrinkage-related ones (27 instances).

The error rate, as well as the frequency of over- and underestimations, remained consistent across all visualization techniques and for both growth and shrinkage scenarios. Accurately estimating volume changes in a 3D model presents an inherent challenge, as relatively small variations in diameter can make overall volume shifts difficult to perceive, often leading to underestimation. However, the distribution between underestimation and overestimation was nearly balanced. Underestimation occurred slightly more often (57 instances) compared to overestimation (55 instances), suggesting no notable bias in one direction.

Although the overall number of errors was relatively low, participants' perceived confidence did not always align with their actual performance. Among the 69 cases where all three questions in a task were answered correctly, only 43 % of participants reported feeling *confident* or *very confident*. In contrast, 57 % selected *a little confident*, *not confident*, or *not confident at all*, despite answering all questions correctly.

Interestingly, when participants indicated *very confident*, they answered correctly in nine out of eleven instances. Conversely, in cases where participants selected *not at all confident*, the response was incorrect eight out of ten times.

USE OF INTERACTION METHODS. The rotation feature, which allowed participants to rotate objects along a single axis, was used in 77 out of 160 tasks. Each task consisted of three questions, and the relationship between fully correct or incorrect responses and rotation usage was analyzed (see Table 3). Rotation was tracked at the task level, meaning that while participants answered three questions per pathology time step series, their use of rotation was not recorded for each individual subtask. This limitation affects the precision of assessing the correlation between rotation and accuracy.

Findings suggest that participants who answered all three questions correctly were more likely to use the rotation feature. Out of the 72 instances where all answers were correct, 40 involved rotation, whereas 32 did not. Conversely, those who answered all three questions incorrectly tended to use rotation less frequently. In four out of the five cases where all responses were incorrect, rotation was not used. Overall, participants who did not engage with the rotation feature were more likely to have at least one incorrect response, whereas those who utilized rotation consistently demonstrated

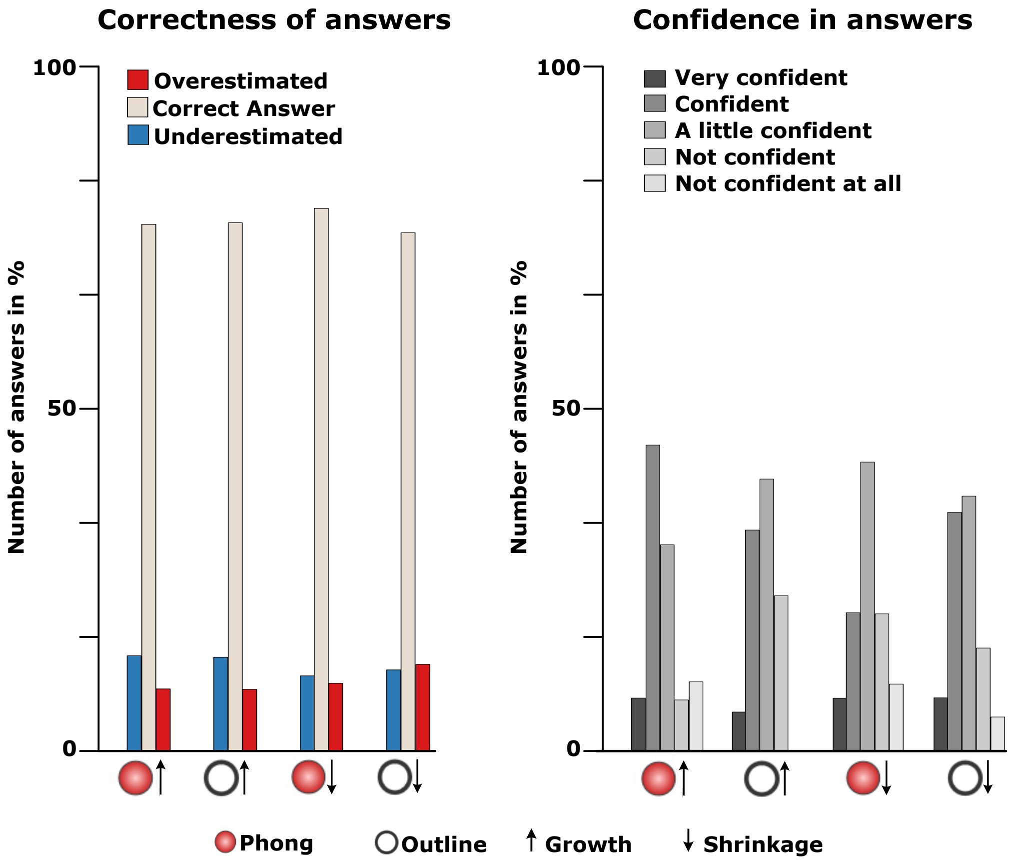


Figure 32: Overestimation and underestimation of growth and shrinkage factors for each visualization technique (left), and participants' self-reported confidence in their answers (right). Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

higher accuracy. These results suggest that interacting with the models contributed to improved performance.

Participants engaged with the interaction feature more frequently when examining aneurysms than liver tumors. One possible reason is that aneurysms were visibly tilted in one direction, making it more apparent that rotation was necessary to achieve a better sense of depth. Another factor may be that aneurysm growth was localized to specific areas, prompting participants to rotate the model for a closer inspection. In contrast, the liver tumor changed size more uniformly, which may have given the impression of a consistent transformation, reducing the perceived need for interaction. Additionally, since tumor-related tasks were presented in the latter half of the evaluation, participants may have been less inclined to interact due to fatigue. Some participants also expressed frustration with the study design, which is further discussed in Section 6.6.

More broadly, interacting with 3D models may not be intuitive for individuals unfamiliar with virtual 3D environments. Despite the button-based interaction method and an explanation provided at the beginning of the evaluation, many participants chose not to use the rotation feature. Nevertheless, the findings indicate that encouraging interaction, even among general audiences, can enhance the accurate perception of 3D objects and improve overall task performance.

KEYWORD SELECTION. The keyword selection results were visualized as a word cloud, as shown in Figure 33. Both shaders were commonly described as *clear* and *understandable*. However, participants more frequently associated the Phong shader with positive attributes such as *informative*, *easy to read*, *pretty*, and *appealing*. In contrast, the Outline shader was often characterized as *unappealing* and *confusing*. Overall, outlines with feature lines received more negative feedback, particularly regarding their visual appeal. The term *unappealing* was selected 23 times in reference to outlines with feature lines, compared to only three times for the Phong shader. Similarly, while 20 participants described the Phong shader as *pretty*, only 12 used this descriptor for outlines with feature lines.

The Phong shader was also described as *detailed* 17 times, whereas the outlines with feature lines received this descriptor only six times. Conversely, outlines with feature lines were labeled as *distracting* and *visually cluttered*. It is worth noting that in the liver tumor visualization, the surrounding liver model was rendered with transparency. This may have contributed to the perception of the outlines with feature lines as visually cluttered, as the combination of two different shading techniques within the same scene could have created a competing visual effect. Overall, these findings suggest that participants generally favored the Phong shader, describing it as both more visually appealing and easier to interpret.

VISUALIZATION PREFERENCES. During the study, participants were asked to assess the visualization techniques twice—once after completing the aneurysm tasks and again after the liver tumor tasks. They were asked to determine which shader they found more *aesthetically pleasing* and which they considered more *suitable for solving the tasks*. The Phong shader was overwhelmingly favored in terms of aesthetics, receiving 71 votes, while outlines with feature lines were chosen only nine times across both pathologies. Preferences remained largely consistent between the aneurysm and liver tumor visualizations, although five additional participants preferred outlines with feature lines when evaluating the liver tumor representation. The Phong shader was also the preferred option for task-solving, receiving 65 out of 80 votes, while outlines with feature lines were selected 15 times. These findings suggest that the Phong shader is generally regarded as more *visually appealing* while effectively communicating medical information. This outcome aligns with the expectations of the medical experts consulted before the study, both of whom expressed a preference for the Phong shader.

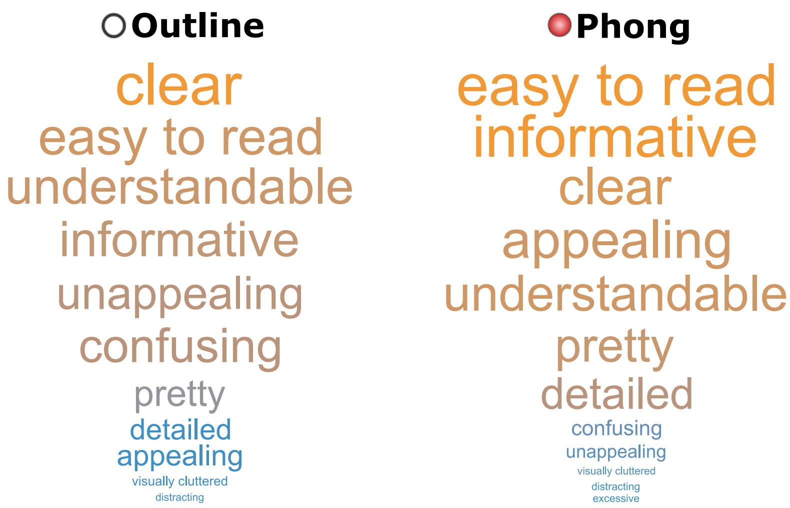


Figure 33: Keywords selected by participants for each visualization technique, color-coded and arranged in descending order based on frequency of selection. Image from Mittenentzwei et al. [159], available under a CC BY 4.0 license. No changes were made.

Although the Phong shader was the preferred choice overall, this does not mean that outlines with feature lines are generally unsuitable for educating the general public about pathological growth. Notably, one participant with red-green color vision deficiency reported difficulty distinguishing details in the red Phong shader, whereas the outlines and feature lines offered better visibility. This suggests that outline-based visualizations may be beneficial for accessibility, particularly for individuals with color vision impairments. Future research could investigate alternative color schemes or adaptive visualization techniques to improve accessibility while maintaining clarity and effectiveness in medical communication.

6.6 REFLECTIONS ON STUDY DESIGN AND LIMITATIONS

This section reflects on the study design, highlighting limitations. These factors may have influenced the outcomes and should be considered when interpreting the results.

PILOT STUDY. Some tasks in the first case study were ranked by difficulty; however, this difficulty level was not formally evaluated, such as through a formative or pilot study. Therefore, it remains an assumption made by the authors of the respective paper [160]. Additionally, a pre-study would have helped identify potential issues in the evaluation process early on, such as whether CFE posed a problem or if the terminology, tasks, and study length were appropriate. This would have provided an opportunity to refine and improve the study design. Conducting a pilot study could have also offered early insights into the usability of the main study, including whether participants consistently used the rotation buttons.

INTERACTION. In the second case study, rotation was implemented using buttons, in contrast to the grab-and-drag interaction used in the first study. This change was based on the assumption that buttons would be easier to use for participants with limited experience in 3D interaction. However, since only the second study tracked the use of rotation, a direct comparison between the two implementations is not possible, and it remains unclear whether the buttons actually improved the 3D interaction experience. The tracking only included if participants used rotation but not the extent of their use. However, the latter would have provided valuable insights into how exactly interaction with the rotation feature influenced study outcomes.

Additionally, the assumption that rotation was underutilized primarily due to limited prior experience with 3D interaction cannot be confirmed. While the studies asked participants whether they had professional experience with 3D visualizations, individuals may also gain familiarity with virtual 3D environments in other ways, such as through computer games. Moreover, familiarity with 3D visualizations was self-reported rather than tested within the study, making it impossible to determine whether prior experience influenced the use of rotation options.

DROPOUT RATE. A total of 20 participants withdrew from the second study before completion. Feedback revealed that many found the tasks too challenging, particularly older participants. This likely resulted in survivorship bias, meaning the findings primarily reflect the experiences of those who successfully completed the study while underrepresenting those who encountered difficulties or chose to drop out. This contributed to a demographic skew, with younger participants being overrepresented.

Another common reason for dropout was the inability to complete the study on a smartphone, causing some participants to discontinue after learning about this limitation. Overall, participants across all age groups reported that the study was not particularly engaging, and certain tasks were perceived as overly difficult. Boredom and frustration were key factors contributing to dropout, affecting participants regardless of age or background.

In the first study, only 13 out of 73 participants withdrew before completing the study. However, this number was likely influenced by the fact that half of the participants took part during the *Long Night of Science*, where they had the opportunity to ask for assistance if they encountered difficulties. Additionally, the presence of study facilitators may have fostered a greater sense of responsibility to complete the study. Although the exact number of dropouts from the *Long Night of Science* compared to the online study was not recorded, observations suggest that nearly all participants at the in-person event completed the study. The study facilitators were able to observe this, as they had to reset the study for each new participants, revealing to them at which part the previous participant discontinued the study.

participant demographics. Two user studies were conducted, both aimed at a lay audience. However, the participant pool primarily consisted of young, highly educated individuals, many of whom had prior knowledge of medical topics. As a result, the sample does not fully represent the diversity of the broader population. This demographic skew was likely influenced by the recruitment strategy, which relied on university mailing lists and the *Long Night of Science*, an event that tends to attract individuals with a pre-existing interest in scientific topics.

Since this bias was consistent across all study versions, the results remain comparable between different tasks and visualization techniques within the study. However, the findings may not be directly applicable to audiences with different demographic backgrounds, and outcomes could vary in a more diverse sample. Collecting additional demographic details, such as participants' occupations, could further refine the interpretation of results. Understanding their professional backgrounds would offer deeper insights into their skill levels and potential biases, helping to assess how different audience groups engaged with the study.

generalizability. Both case studies compared different visualization and presentation techniques. However, it is important to emphasize that the study results are based on the specific selection and implementation of these techniques. While the visualization methods used in the studies represent broader classes of visualization techniques, the findings cannot be generalized to these. The case studies were designed as exploratory investigations to identify trends and collect qualitative feedback on various visualization and presentation techniques. Given the small sample size, the data collected does not allow for robust statistical analysis, particularly in determining differences between the visualization methods. Instead, these findings offer preliminary insights that can guide future research, enabling larger-scale studies with broader participant groups and more rigorous statistical validation.

The number of visualization techniques examined in the studies is too limited to allow broad generalization. Additionally, the implemented versions of these techniques may not represent the techniques most situated for the target audience. For example, an alternative implementation of outlines could potentially be more visually appealing to participants. Similarly, the provided hatching lines were highly abstract, primarily consisting of evenly spaced lines overlaying the 3D models. A different hatching approach – where lines are more prominent in high-curvature regions and along model edges – might be more accessible for users.

design space. Given the extensive design possibilities for visualizing surface models based on medical image data, the studies can only incorporate a small part of the design space. Prior research on medical visualization of volumetric data has introduced a diverse array of visualization techniques [124, 192]. For these studies, a limited selection of three (case study 1) and two (case study 2) visualization techniques were chosen. This selection includes illuminative techniques that provide a semi-realistic impression and illustrative techniques commonly featured in educational textbooks. This selection allowed for an exploration of different styles while focusing on methods already familiar in medical education and communication.

Participants were presented with discrete time steps in two presentation techniques: side-by-side and sequentially. Alternative methods, such as automated animations transitioning between time steps or morphing 3D meshes to create a continuous growth effect, were not included. For instance, a self-playing animation would progress at a fixed speed, potentially misaligning with the user's preferred pace and making task completion more challenging. To refine the study's scope, the focus was on assessing whether participants found one presentation technique more helpful or overwhelming than the other and whether a particular technique enhanced the audience's understanding of pathological growth. Interaction was implemented through single-axis rotation using a click-and-drag mechanism and button controls. This approach was chosen for its simplicity, in regard of accessibility concerns for the audience.

By establishing these constraints within the design space, the studies remained focused, limiting their scope. While this was necessary given the vastness of the design space, it also meant that many design decisions have yet to be evaluated.

ARTIFICIALLY GENERATED DATA. The surface models for the time-steps were manually generated, with only the final stage in each dataset – representing the fully developed pathology – being directly based on real medical data. Consequently, the intermediate stages may not accurately depict the actual progression of pathological structures. However, achieving anatomical realism was not the primary objective of this study. To ensure the models were appropriate for a general audience, two physicians were consulted, and they confirmed their suitability for this context. This validation aligns with the study's focus on exploring visualization techniques rather than providing a strictly anatomically accurate representation of pathological growth.

6.7 DISCUSSION OF RESEARCH QUESTIONS

The study results are referenced to address research question RQ1 and its subquestions, RQ1.1 and RQ1.2. The discussions of the RQs need to be interpreted in the light of the aforementioned limitations.

RQ1.1 **WHAT ARE AESTHETIC PREFERENCES OF LAY AUDIENCES FOR VISUALIZING SURFACE MODELS?**

Participants' self-reported aesthetic preferences were recorded in both case studies. In the first case study, each participant was exposed to both side-by-side and successive presentations of time steps, allowing for a direct comparison. When asked which they preferred aesthetically, both presentation techniques received substantial support, with the successive presentation receiving slightly more votes. In the second case study, participants were shown both visualization techniques – Phong shading and outlines with feature lines – and asked to state their aesthetic preference. Participants noticeably favored Phong over outlines with feature lines. In summary, Phong shading emerged as the preferred visualization technique among participants, while preferences for presentation techniques were more evenly distributed.

RQ1.2 **DO AESTHETIC PREFERENCES ALIGN WITH HIGHER PERFORMANCES IN SOLVING TASKS?**

Task performance was quantitatively measured, and participants also self-reported their preferred visualization and presentation techniques for solving the tasks. To answer RQ1.2, these results are compared with the conclusions from RQ1.1

In contrast to self-reported aesthetic preferences for presentation techniques, self-reported task-solving preferences showed a slight preference for the side-by-side presentation. This suggests that for some participants, their aesthetic preferences did not align with their preferred method for solving tasks. However, this result does not fully align with the actual task performance, where the side-by-side presentation led to more incorrect answers – except for the Phong shader in Task 1 and Task 2 of the first case study. Nevertheless, the performance differences between presentation techniques were minimal, making it difficult to draw definitive conclusions. For visualization techniques, self-reported aesthetic preferences aligned with self-reported preferences for solving tasks, with Phong receiving noticeably more votes. Task performance results showed that both visualization techniques led to a similar number of correct answers.

Participants' confidence in their answers did not consistently align with their actual performance. In the first case study, Phong shading was associated with lower confidence compared to other visualization techniques, whereas in the second case study, participants reported slightly higher confidence with Phong compared to outlines with feature lines. As a result, confidence levels did not provide meaningful support in investigating this research question.

Keywords assigned by participants in the second study further indicate that the Phong shader was perceived as both aesthetically pleasing and interpretable. In contrast, outlines with feature lines were described as interpretable but not aesthetically pleasing.

In summary, aesthetic preferences do not always align with self-reported or measured task performance. However, differences in performance were minor, suggesting a ceiling effect that limits the generalizability of the results.

RQ1 HOW TO VISUALIZE MORPHOLOGICAL CHANGES OF PATHOLOGIES USING SURFACE MODELS FOR A LAY AUDIENCE? The discussion of RQ1.1 and RQ1.2 leads to the following insights concerning RQ1. Since task performance did not show notable differences between the presentation and visualization techniques, participant preferences played a key role. While no strong preference was observed between the two presentation techniques, participants overwhelmingly favored Phong as a visualization technique. Based on the specific implementations used in the case studies, Phong can be recommended as the preferred visualization method. However, this conclusion may not be generalizable due to the limitations discussed in Section 6.6. Notably, the clinical cooperation partners also strongly favored Phong shading as a visualization method.

6.8 CONCLUSION

As emphasized by medical experts during pre-study interviews, understanding the concept of pathological growth and shrinkage is essential for laypersons, whereas the rate of growth or shrinkage is less critical. Such details are typically addressed in clinical settings under expert supervision, particularly when discussing treatment options or care strategies.

6.8.1 *Study Implications*

The study findings indicate that the Phong shader was generally preferred by the participants. However, outlines with feature lines proved particularly beneficial for one participant with red-green color blindness, suggesting that certain visualization techniques may enhance accessibility for individuals with color vision deficiencies. The user study demonstrated that participants exhibited comparable performance across different visualization techniques, regardless of whether the pathology was growing or shrinking. This suggests that the techniques did not have a noticeable impact on task accuracy. However, further investigation is necessary to explore a broader range of visualization techniques, including variations in illumination-based and illustrative methods. Participants exhibited no notable differences in preference or performance between the two presentation techniques: side-by-side and sequential display.

Based on the study findings, the Phong shader could serve as the default visualization, as it was generally favored by participants. However, providing the option to switch to an outline shader could improve accessibility, particularly for users with color vision deficiencies. Despite offering more visual cues, the combination of the

Fresnel shader with hatching lines resulted in lower performance in multiple-choice tasks compared to other visualization techniques. This suggests that simpler visualizations may be more effective for conveying medical image data to lay audiences.

Participants who did not rotate the models were more likely to provide incorrect answers. Although the rotation interaction was adjusted between the two studies, the issue remained, indicating that many users still did not fully utilize the 3D interaction feature. Findings from the study suggest that many users engage with interaction features only when they perceive them as necessary. During the in-person event, only a small subset of participants actively rotated the objects, with younger participants being more likely to do so. Neither oral explanations, textual prompts, nor icons effectively motivated users to interact with the models across multiple scenes. While most participants engaged with the interaction features during the introductory scene, many discontinued their use afterward.

The design of visualizations plays a key role in determining user interaction. For example, an outline shader, which offers limited depth perception, naturally encourages rotation to create parallax effects that enhance depth cues. Conversely, visualizations can also be designed in a way that makes interaction unnecessary, depending on the narrative intent. However, requiring interaction to understand visual information can be frustrating, particularly for users unfamiliar with 3D navigation. At the same time, encouraging interaction can serve as an opportunity to introduce users to 3D exploration, potentially enhancing their ability to engage with interactive visualizations.

6.8.2 *Future Work*

Future research could focus on optimizing surface visualizations to improve accessibility, ensuring that medical information is effectively conveyed to audiences with diverse visual perception abilities. This encapsulates an evaluation of additional visualization and presentation techniques, to further explore the design space. This study focused on static, discrete time steps to represent non-periodic, time-oriented medical data. However, dynamic animations could also be employed to visualize temporal changes more fluidly and should thereby be explored along with suitable interaction methods (i.e. playback speed) in the future. Additionally, in the context of visualizing growth and shrinkage, further studies should assess whether 3D models are the most effective approach for educating lay audiences. Alternative visualization methods, such as simple line graphs showing changes in size, may offer a more intuitive and accessible way for lay people to interpret growth and shrinkage in medical data.

While the studies investigated different visualization and presentation techniques, it did not focus on how to present anatomical context, which is a key aspect according to the domain experts. Participants performed tasks more quickly when analyzing tumors alone than when assessing aneurysms attached to a vascular tree. While the vascular tree provided valuable anatomical context, it also increased visual complexity, potentially making it more difficult for users to extract key information efficiently. Therefore, finding a balance between providing anatomical context and keeping visual complexity low is essential.

Future research should investigate ways to better encourage and facilitate 3D interaction for lay users. This may involve exploring alternative interaction methods, designing more intuitive controls, or implementing guidance mechanisms that actively prompt users to rotate models when necessary. Enhancing user engagement with 3D interaction is essential for improving accuracy in interpreting volumetric medical visualizations. Deliberately encouraging interaction can serve as a valuable strategy for introducing users to 3D navigation techniques and fostering engagement. It is important to focus on finding the optimal balance – ensuring that interaction enhances comprehension without becoming a barrier to medical data communication.

The high dropout rate in this study underscores the need for improvements in study design. To minimize participant frustration, task explanations could be refined by incorporating clearer textual descriptions and introductory videos. Additionally, conducting a pilot study with a small group of participants before the full study launch would help ensure that instructions, terminology, and interaction mechanics are intuitive and easy to understand. To enhance engagement and reduce dropout due to lack of interest, incorporating gamification elements could be beneficial. For instance, displaying the number of correctly completed tasks after each pathology case might encourage participants to continue. Research by Nehring et al. [175] has shown that gamification is an effective strategy for boosting motivation in educational settings.

Both aneurysms and tumors can develop bulges (blebs), which have important clinical implications:

- In aneurysms, the presence of blebs is associated with a higher risk of rupture.
- In tumors, an irregular shape can serve as an indicator of malignancy.

Future research should focus on effectively conveying not just changes in size but also variations in shape over time. Since realism plays a crucial role in understanding morphological transformations, mathematical growth models could be employed to generate more accurate representations of pathological progression [3, 196, 207]. Additionally, further studies could examine the impact of surface complexity on task difficulty. Simplifying anatomical structures, such as using spherical representations, may help determine whether intricate surface details contribute to increased cognitive load and make certain tasks more challenging.

7 THE NARRATIVE STAGE: CHARACTER

Another ingredient in the narrative stage is character, see Figure 34. A story can feature one or more characters, including various types such as protagonists (main character, e.g., a patient), side characters (e.g., a treating physician or relative supporting the protagonist), and antagonists (opponent of the protagonist, e.g., a disease) [156]. A character need not always be a person; it can also be a data point or an object. For disease stories, using a patient character as a protagonist can provide a relatable personification of the conflict. The design space for characters is vast. While photorealistic characters can make the story feel more authentic, illustrated characters may be more memorable and offer neutrality in terms of age, gender, and appearance, making them potentially more appealing to a diverse audience.

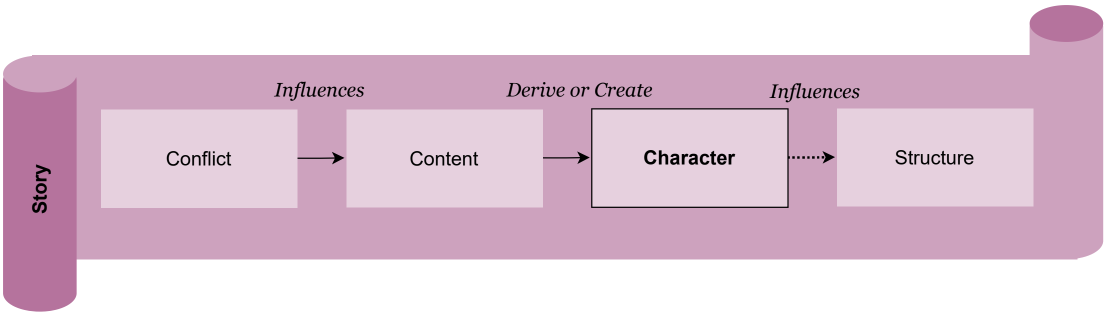


Figure 34: The narrative stage of the disease story design process (cf. Figure 9) with *character* highlighted.

In this chapter, patient characters are derived from cohort data and augmented with fictitious names and images, ensuring both realism and patient anonymity based on the foundational work of Budich et al. [27]. In a discussion with four design experts, different styles are explored. To address this, the following research questions are explored:

RQ2 How to create data-driven characters for disease stories?

- *RQ2.1* How to generate characters that are authentic to the underlying data?
- *RQ2.2* How to alter images that still resemble the same person?
- *RQ2.3* How to refine the generated characters to meet design standards?

7.1 CREATING DATA-DRIVEN CHARACTERS FOR DISEASE STORIES

Characters, defined as embodiment of specific conditions derived from patient data, serve to make health information more relatable. By integrating characters into disease stories, emotional engagement with both the narrative and the underlying data can be strengthened [156, 174]. Creating convincing characters is still challenging and requires either creativity, research, and technical skills in the creation of illustrations or the consent of patients to have their photos taken. Deep learning text-to-image GenAI tools like Midjourney [154] and Stable Diffusion [227] have transformed this process by generating images from natural language prompts. In character design, *prompt engineering* involves creating prompts to develop traits and variations in the generated output images using specific queries and style specifications [138].

Collaboration and Personal Contribution

This chapter is partially based on the following publications: [27, 155]. In previous work [27], a pipeline for AI-enhanced character creation in data-driven medical narratives was developed, addressing data protection challenges by avoiding the use of real patient photos. This study is an extended approach to generate virtual patients from epidemiological datasets, such as the SHIP [96]. The data for this study was curated by Beatrice Budich, who formulated the patient selection criteria and the former version of the semi-automated character design pipeline. The author of this thesis created an adapted concept to improve reproducibility and conducted an evaluation of the created characters. As a result, some text passages in this work closely follow the methodology outlined in previous research. The submitted work has been prepared with all co-authors who provided feedback throughout the project and writing phase.

This work contributes a set of design criteria for developing characters in medical storytelling, formulated by an industry expert with substantial experience in the field. By incorporating Stable Diffusion [227], the study enhances reproducibility in the design process – a challenge often posed by the frequent model updates on external platforms like Midjourney [154]. Through a case study, the research examines the authenticity and consistency of characters, exploring both the challenges and possibilities of GenAI-assisted character design. The findings contribute to advancements in character design within the healthcare domain and foster further research in this area. In comparison to the work by Budich et al. [27], key contributions of this study are:

- Expanding the GenAI-driven character creation pipeline by integrating detailed epidemiological data, allowing for richer narratives that incorporate factors such as education, profession, and marital status.

- Improving the reproducibility of character generation and strengthening data privacy by employing a self-hosted Stable Diffusion model.
- Proposing a two-phase evaluation of a proof-of-concept to validate GenAI-driven character creation, identifying design challenges, and offering insights to refine medical storytelling.

7.2 RELATED WORK

The creation of GenAI-driven characters for medical storytelling is formalized within a semi-automated character design pipeline. It integrates foundational concepts from conventional storytelling, narrative medical visualization, and the creative capabilities of GenAI. This section contextualizes the study's contributions by discussing relevant related work from character-driven storytelling as well as GenAI.

7.2.1 *Towards Narrative Visualization with Characters*

Characters play a crucial role in shaping narratives, particularly within the media and entertainment industry, where the gaming sector has been at the forefront of research and innovation in character-driven storytelling. Cavazza et al.[29] explore the dynamic interactions of autonomous characters, while more recent studies by Mariani et al.[147] and Sheldon [214] delve into character-driven storytelling. Mariani et al. focus on character development for interactive stories, emphasizing that well-crafted characters with rich backstories and traits can enhance engagement, deepen immersion, and propel the narrative forward.

Katherine Isbister [94] examines the complexity of character design within games, analyzing it from a psychological perspective. Character development in game design is largely driven by social roles and integration into gameplay, with more prominent characters requiring greater depth and complexity to enhance player engagement. A key parallel exists between non-player characters in games and characters in narrative visualization, as both serve narrative functions without being under direct user control. However, while non-player characters in games – such as merchants, guides, or enemies – designed for interactive gameplay (e.g., the player can interact with them through playable characters, sometimes even influencing the non-player characters behavior), characters in narrative visualization remain mostly non-interactive, limiting user engagement in a different way.

Despite these differences, both narrative visualizations and games share a fundamental goal: fostering empathy toward characters. Isbister [94] explores the concept of *Social Syncing*, illustrating how emotional expressions in characters can be contagious, influencing users' emotional states. Facial expressions are particularly effective in conveying a character's individual emotions, while body language plays a critical role in depicting relationships and social dynamics.

In game development, the roles of narrative design (shaping the story and defining character roles) and visual realization (executed by concept artists) are typically assigned to separate individuals [131]. This study leverages GenAI to support the role of the concept artist, while narrative design is guided by data-driven processes. To enrich character creation, additional contextual elements, such as emotional attributes, are incorporated into the GenAI prompts, recognizing their significance in character development.

The role of characters in narrative visualization is still emerging. Dasu et al. [48] introduced a framework aimed at enhancing narrative visualization by integrating characters to make it more engaging and accessible. By analyzing 160 narrative visualizations, they categorized character roles – main, supporting, and antagonistic – and examined their interactions within narrative structures. Their findings highlight how characters can function as a bridge between complex data and audience comprehension, utilizing visual metaphors and narrative structure. Although this study illustrates how characters can simplify and humanize complex data, their implementation remains constrained by manually created character roles, lacking automation and essential design elements such as emotional expression and visual consistency.

7.2.2 *Generative AI Art*

The growing accessibility of openly available (latent) diffusion models, such as Stable Diffusion [227], alongside user-friendly tools like Midjourney [154] and Leonardo AI [220], has made GenAI art increasingly popular. These models empower users to generate images from textual descriptions across diverse domains.

While training diffusion models typically requires substantial datasets and computational power, LoRA [86] enables efficient fine-tuning of pre-existing models for specialized applications, including highly stylized or near-photorealistic imagery. Community-driven platforms like Civitai [37] facilitate the sharing and customization of such models, while tools such as Automatic1111's stable-diffusion-web UI [12] provide an accessible interface for their usage.

Beyond generating images from text, diffusion models support a range of tasks, including inpainting (modifying user-defined areas of an image), sketch-to-image conversion (transforming rough sketches into detailed visuals), and resolution adjustments through up- or down-sampling. In the realm of medical storytelling, So et al. [222] introduced a GenAI-driven tool for generating narrative medical visuals derived from social media discussions on conditions like diabetes. Their approach is based on the bio-psycho-social model, which integrates medical, psychological, and social factors [79].

Despite advancements in interactive storytelling and game design, there remains a gap in harnessing GenAI to synthesize narrative roles, emotions, and contextual information. This study explores how GenAI tools can facilitate the creation of emotionally compelling, visually expressive characters for medical storytelling – a field where characters serve as a bridge between technical data and audience empathy.

7.3 FROM TRADITIONAL TO AI-ENHANCED CHARACTER DESIGN

Designing characters is a multifaceted and time-intensive endeavor that demands a blend of creativity, research, and technical expertise. The traditional workflow for creating patient characters in medical storytelling is presented based on the professional experience of one of the co-authors from the respective papers, drawing insights from industry pipelines and incorporating adaptations of modern GenAI tools. Traditionally, character design follows four stages [27, 155]:

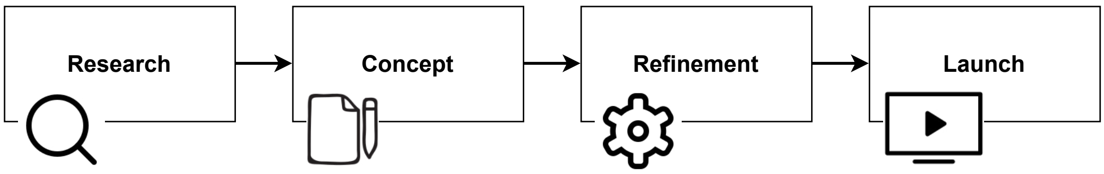


Figure 35: Overview of the steps included in the traditional character design pipeline.

RESEARCH. In the beginning of the character design process the artist receives a project brief that defines the medical condition and outlines the intended learning objectives for the audience. To develop an accurate representation, the artist conducts research on the disease's signs and symptoms, constructing a typical patient profile that includes factors such as age, gender, ethnicity, body composition, and psychosocial aspects like depression.

Medical literature databases and online resources, including UpToDate [115], Healthline [82], Cleveland Clinic's Health Library [39], and JAMA's Patient Pages [103], offer epidemiological insights and patient narratives that inform character development. Additionally, the artist gathers photographic and stylistic references to ensure accurate representation of physical traits, proportions, and behavioral characteristics.

Taking data from epidemiological studies as a base for the character, as done in this work, streamlines research efforts and minimizes the need for extensive manual searches within this phase. However, ensuring that the selected patient representations accurately reflect the condition still requires validation through external sources.

CONCEPT. The artist sketches character design concepts, either digitally or traditionally, using professional tools such as Adobe Creative Cloud [2] or Procreate [206].

Drawing from references gathered during the research phase, the artist iteratively refines these initial designs, ensuring they align with the project's vision before collaborating with a team or client for feedback. This phase concludes once a final character concept is selected for further development.

In this work, GenAI is leveraged to accelerate the early design stage, generating multiple character variations within seconds to minutes. Prompts are derived from prior findings in the research phase, streamlining the ideation process.

REFINE. The character design undergoes iterative refinement, incorporating references from the research phase along with feedback from collaborators and domain experts. As in previous stages, GenAI accelerates the process by enabling rapid adjustments and modifications to character imagery.

LAUNCH. The final character, along with associated assets (e.g., textures, models, and animations) is integrated with other storytelling components (e.g., text and multimedia). After undergoing validation, the final product is deployed for end users. This process remains consistent across both traditional and GenAI-driven design.

7.4 CRITERIA FOR CHARACTER DESIGN

When developing characters for narrative medical visualization, it is essential to ensure clear and effective communication of medical information. This work presents ten key criteria for character design in medical narratives, taken from [155] and informed by one of the co-authors' industry experience, expanding upon the methodologies discussed in Section 7.3:

C1. RELEVANCE TO MEDICAL CONTEXT. The visual appearance of characters is essential for effectively conveying medical conditions, as it can represent symptoms, evoke emotional responses, and provide a human context to make data more relatable. Characters should clearly connect to the narrative's healthcare focus, with backgrounds that are relevant to the story. Designers can use evolving visual cues to depict changes in health status and the condition's impact on well-being [188].

C2. INFORMATIVE AND EDUCATIONAL. Characters should serve as vehicles for conveying medical information and data. A character's appearance should provide informative insights into medical concepts, conditions, or procedures [188, 249].

C3. MEMORABLE. To capture the audience's attention and leave a lasting impression, characters must have unique traits, qualities, or quirks that let them stand out from other characters. For example, *Naming* characters strongly contributes to their identification, relatability, and memorability [19].

C4. EMPATHY AND EMOTIONAL CONNECTION. The character should evoke empathy and foster an emotional connection, enabling deeper audience engagement. Integrating authentic visual elements that represent symptoms, effects, or challenges of the medical condition enhances understanding and connection [19].

C5. CREDIBILITY. To make a character credible, its visual design should align with the narrative to immerse readers fully. Clothing choices and authentic expressions enhance the story's atmosphere and engagement while maintaining a consistent style among characters is crucial for credibility [19].

C6. PERSONALIZATION AND AUDIENCE RELEVANCE. The character's attributes and traits should resonate with the target audience, representing individuals from risk groups defined by factors like age or preventable risks. Prioritizing diversity in design – such as various ages, genders, ethnicities, and body types – ensures medical conditions are inclusively portrayed and not confined to a single demographic [19].

C7. BEHAVIORAL CHANGE AND MOTIVATION. The primary role of a patient character is to inspire positive behavioral changes, motivating the audience to adopt healthier habits and make informed medical decisions. As role models, characters can demonstrate overcoming challenges and the benefits of changes [249].

C8. DATA-INFORMED DESIGN. Character design should be grounded in data analysis and corroborated by credible sources like literature and reports. In longitudinal studies, attributes such as demographics, background, disease progression, and risk factors can be directly extracted from the dataset [120].

C9. ETHICAL CONSIDERATIONS. Characters should adhere to ethical guidelines, respecting patient privacy, confidentiality, and cultural sensitivities. Designers should avoid identifiable features, real-life resemblances, and any stereotypes or stigmatizing portrayals of medical conditions or communities. Thus, they should uphold ethical standards and prioritize the well-being of those with medical conditions [249].

C10. ITERATIVE DESIGN AND EVALUATION. The character's design should be subject to iterative refinement based on empirical evidence and audience feedback. User testing and evaluation should be conducted to assess the character's impact on audience comprehension, engagement, and behavioral change [188].

7.5 PROOF OF CONCEPT FOR A SEMI-AUTOMATED CHARACTER PIPELINE

This work introduces a semi-automated character design pipeline aimed at enhancing disease stories, as illustrated in Figure 36. The process begins with *requirement definition*, where relevant data is extracted. Based on the needs of the domain expert, the story's intent and target audience are defined. Hard and soft criteria are then established to filter data in alignment with the story's intent. Using the final candidate data and, optionally, additional resources (e.g., online information about the disease or research on character design, such as appropriate names), a prompt is formulated to generate a batch of character images via *text-to-image generation*.

Desired styles (e.g., photorealistic or comic) can be specified. The most suitable image is selected from the batch and can be further refined through *image-to-image translation*. The resulting character should be reviewed in the final *refinement* stage to ensure it effectively supports the story's intent. Feedback from this evaluation can be used to adjust the prompt and generate improved characters. Once a satisfactory character is created, it can be finalized and integrated into the story.

Demonstrating the pipeline within a proof of concept, characters are extracted from the SHIP, an epidemiological database in West Pomerania that identifies disease risk factors [114]. Data subsets from two different points in time were utilized in the process, namely *SHIP-0* (1997–2001) and *SHIP-2* (2008–2012). While the is an intermediate time step *SHIP-1* (2002–2006), only SHIP-0 and SHIP-2 were used to create a minimal example of character progression over an elongated period of time.

Two characters were extracted from the data to illustrate the relationship between body weight and NAFLD. The first character, a woman, represents an individual who successfully recovered from NAFLD, while the second, a man, portrays someone who developed NAFLD due to lifestyle factors such as weight gain and lack of physical activity. The primary goals of visualizing SHIP data on NAFLD are: (1) to educate the public about risk factors and preventive strategies for NAFLD and (2) to raise awareness of the SHIP study, fostering acceptance among potential participants. The intended audience is not NAFLD patients but the general public. This aligns with *Step 1.1*, *Define Story Intent*, as illustrated in Figure 36.

7

Supplementary Material

Supplementary material for this chapter can be found at:
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/character`

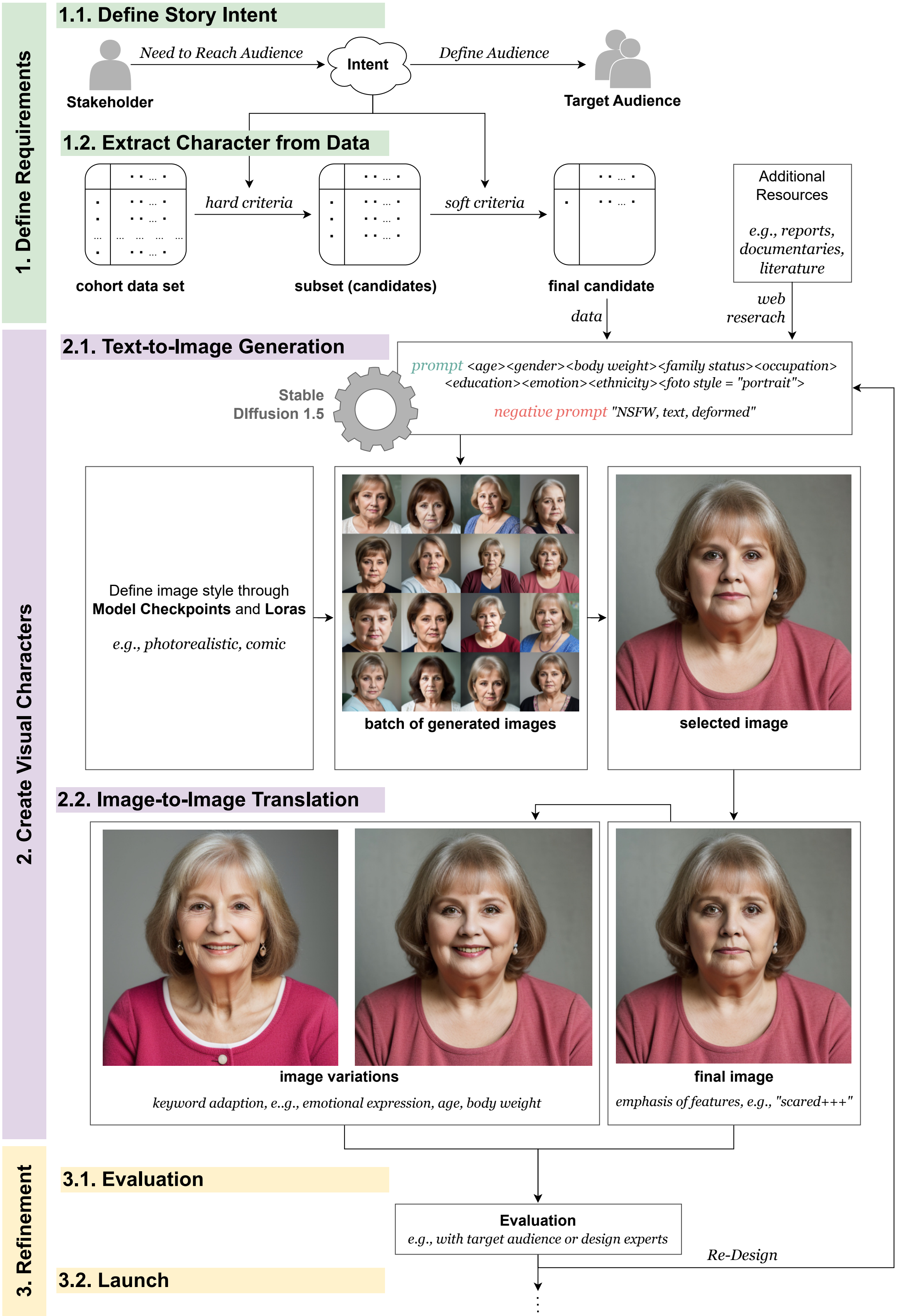


Figure 36: Semi-automated character design pipeline. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

7.5.1 *Extract Character from Data*

The character design process uses a query-filtering method to extract a relevant subset from the SHIP dataset for subsequent image generation. The filtering process is guided by two sets of criteria, aligned with design principles C1, C2, C7, and C8, as illustrated in Figure 36 (Step 1.2: Extract Character from Data).

Hard criteria focus on selecting individuals whose profiles effectively convey the core medical message and narrative intent, such as promoting positive lifestyle changes. This includes subjects whose health trajectories align with disease risk factors, particularly those who have either successfully improved their health or, conversely, developed medical conditions due to lifestyle choices. *Soft criteria* prioritize candidates with strong changes in lifestyle habits – either improvement or decline – requiring a deeper understanding of disease risk factors to identify relevant cases.

To establish a narrative context, additional variables are selected to help define the protagonist's personal background and personality traits, such as gender, family, education, and occupation. This approach closely parallels persona creation, where user data is analyzed to identify demographic and behavioral patterns [98]. The resulting personas are often supplemented with visual elements, such as images and names, to create a more relatable and comprehensive profile.

To further enhance emotional depth, *additional resources* are incorporated, including documentaries on patients with NAFLD [53, 54]. Other sources may include insights from medical professionals, online information about the disease, and research on character design elements, such as selecting names that align with the character's age and sociocultural background.

Algorithm 1 shows the hard and soft criteria for extracting the female character recovered from NAFLD, while Algorithm 2 depicts the extraction of the male patient that develops NAFLD. The *hard criteria* fall into three groups, focusing on variables strongly correlated with NAFLD identified from SHIP studies [96]:

1. Diagnose-related variables: For the diagnosis of NAFLD, only subjects with *fatty liver* combined with no *alcohol problems* and either recovered or developed the disease were considered.

2. Demographic variables, i.e., *gender*, *family*, *education*, and *occupation* to give context and history to the character.

3. Risk factors or lifestyle-related variables for NAFLD, e.g., age, *BMI*, *alcohol use*, *physical activity* [96].

Soft criteria identified those participants with positive lifestyle changes: *weight loss*, *exercise*, *smoking cessation*, and *reduced alcohol consumption*.

Learning Objectives

The motivation for communicating NAFLD – educate a lay audience about avoidable lifestyle risk factors – can be framed as learning objectives. The cognitive objective emphasizes the recall of facts, following Adar et al. [1]:

- The viewer will be able to recall lifestyle factors contributing to NAFLD.

Affective objectives focus on raising awareness and motivate behavioral change, as defined by Lee-Robbins et al. [132]:

- The viewer will consider the lifestyle factors contributing to NAFLD.
- The viewer will perform behaviors consistent with the lifestyle factors contributing to NAFLD.

These objectives guide the design of disease stories but are not the focus of evaluation in the following study; rather, they provide a structured rationale for this research.

7.5.2 *Text-to-Image Generation*

A text-to-image generation method is employed to create characters based on hard and soft criteria, see Figure 36 (Step 2.1. Text-to-Image Generation). Budich et al. [27] utilized Leonardo AI [220], a web-based GenAI tool with pre-trained models for various image styles. However, users have limited control over the models, as regular updates may alter the functionality of older prompts. Given this major drawback, both for the long-term application of the tool in supporting character design for data-driven medical stories and for research studies. The image generation pipeline was adapted to locally run Stable Diffusion version 1.5. This decision was based on the following criteria:

- REPRODUCIBILITY: Stable Diffusion is widely recognized as a benchmark for evaluating generative AI techniques, including diffusion models, and is frequently cited in academic papers and studies as a comparative baseline. Its fixed releases ensure that researchers can utilize the exact same version of the GenAI model that were employed in this study.
- OPEN SOURCE: Stable Diffusion's open-source nature allows users to thoroughly examine its model architecture and algorithms, providing insights into its inner workings, an essential aspect of ensuring reproducibility. The absence of licensing fees for non-commercial use makes it a cost-effective choice for institutions, researchers, and developers, especially in academic or non-commercial settings.

- DATA PRIVACY: Stable Diffusion can be executed locally, ensuring that both the input and output data remain exclusively accessible to the creator. This is particularly critical for applications in sensitive areas such as medical topics, where data privacy is paramount.
- MULTIMODAL CAPABILITIES: Stable Diffusion offers versatile functionality essential for character creation, including text-to-image generation, image-to-image translation, and tools for refining outputs, such as specifying areas of an image to be altered or preserved. This adaptability makes it suitable for a wide range of GenAI tasks. Its all-in-one functionality streamlines the pipeline, eliminating the need for multiple platforms and enhancing accessibility for character creation workflows.
- AVAILABLE RESOURCES: As an open-source platform, Stable Diffusion provides access to a diverse range of models and a flexible tool chain developed by its active community. This supports fine-tuning the model and storing optimization results, including network weights and learning rates, e.g., as model checkpoints tailored to various image styles and user-friendly GUIs. Its high degree of customizability makes Stable Diffusion a preferred choice for creators seeking adaptability and control.
- RELEVANCE: Stable Diffusion is a leading diffusion-based generative model that creates high-quality, photorealistic, and stylized images from text prompts. Competing with proprietary models such as OpenAI's DALL-E and MidJourney, it leverages advanced latent diffusion model (LDM)s to deliver high-quality outputs with optimized computational efficiency and resource utilization.

1. MODEL CONFIGURATION. Stable Diffusion web UI [12] was utilized as it offers a graphical, browser-based interface to specify parameters like model selection, number of output images, and prompts. Customized model checkpoints can be developed by retraining neural network weights using a small set of images with consistent styles. Pre-trained model checkpoints (files containing the weights of a Stable Diffusion model specialized in a particular style, subject, or aesthetic) are available on the Civitai website [37], with the selection being "HT Photorealism v4.1.7" [81], tailored for portrait photography Table 4 outlines the parameter setup, predominantly using default settings from the Stable Diffusion web UI v1.6.1.

2. POSITIVE AND NEGATIVE PROMPTS. Positive prompts indicate the desired image outcome, while negative prompts specify undesired features. Image generation models may lack regulation of their generated output, potentially generating not-safe-for-work (NSFW) content, including nudity. To encourage safe results, "NSFW" was

Algorithm 1 Extract character (case study example female)

$Candidates\_hardcriteria \leftarrow subjects\ S\ where$

```
(stea_s0 = 1 and stea_s2 = 0) and (alcproblem_s0 = 0)
and (overweight = True) and (weight-loss = True) and (physact_s2 = 1) and (smoking_s2 = 0
| 1)
```

$where:$

```
overweight: if (whtr_s0 > 0.5) then True
weight-loss: if (som_tail_s0 > som_tail_s2)
then True
```

$Candidate\_softcriteria \leftarrow Candidates\_hardcriteria\ where\ s$
$maximizes\ the\ following\ conditions:$

```
(whtr_s0 > 0.5 and whtr_s2 < 0.5) or
(som_bmi_s0 > 25 and som_bmi_s2 < 25) or (som_bmi_s0 > som_bmi_s2) or (physact_s0 = 0 and
physact_s2 = 1) or (alkligt_s0 > alkligt_s2) or (LFV_s0 > LFV_s2)
```

Legend: *s0*, *s1*, and *s2* refer to data subsets SHIP-0/-1/-2 recorded at five-year intervals. ***Variables:*** *stea* = steatosis hepatis (fatty liver), *physact* = physical activity, *smoking* = 0 (non-smoker); 1 (ex-smoker), *whtr* = waist-to-height ratio, *som_bmi* = body mass index, *alkligt* = alcohol in g per day, *LFV*=liver function values: Triglycerides, ALAT, GGT, ASAT (indicators for liver health, e.g., high dietary fat intake, liver inflammation or damage).

Parameter Name	Parameter Value
Sampling method	DPM++ 2M Karras
Sampling steps	20
CFG scale	7
Image Width	512
Image Height	512

Table 4: Parameters of the character generation pipeline.

7

included as a negative prompt. Additionally, "text" and "deformed" were added as negative prompts, as models often struggle with text and complex structures like hands.

3. PROMPT ADAPTATION DUE TO MODEL SPECIFICS. Using Stable Diffusion instead of Leonardo AI requires adjusting previous prompts used by Budich et al. [27] for optimal outcomes. Style-specific keywords (e.g., semi-realistic or photorealistic) were removed from the prompts as they are now defined by the model checkpoint. Additionally, character names were removed from the prompts, as they no longer notably affect their appearance.

4. EFFECTS OF DIFFERENT KEYWORDS. The keyword "kindergarten teacher" introduced occupation-specific backgrounds like classrooms and blackboards in some images, often altering clothing to blouses or plain long-sleeved shirts, occasionally with floral patterns. However, the occupation's influence on output varies, and it may per-

Algorithm 2 Extract character (case study example male)

$Candidates\_hardcriteria \leftarrow subjects\ S\ where$

```
(stea_s0 = 0 and stea_s2 = 1) and (alcproblem_s0 = 0) and
(overweight = False) and (weight-loss = False) and (physact_s2 = 0)
```

$where:$

```
overweight: if (whtr_s0 > 0.5) then True
weight-loss: if (som_tail_s0 > som_tail_s2)
then True
```

$Candidate\_softcriteria \leftarrow Candidates\_hardcriteria\ where\ s$
$maximizes\ the\ following\ conditions:$

```
(whtr_s0 > 0.5 and whtr_s2 < 0.5) or
(som_bmi_s0 > 25 and som_bmi_s2 < 25) or (som_bmi_s0 > som_bmi_s2) or (physact_s0 = 0 and
physact_s2 = 1) or (alkligt_s0 > alkligt_s2) or (LFV_s0 > LFV_s2)
```

petuate stereotypes and clichés for certain jobs. Keywords like "educated," "mother," and "married" had minimal impact on output, mainly altering characters' clothing and jewelry in some images. For instance, "educated" occasionally produced more elegant looks with prominent jewelry and blazers. Despite their minor influence, these keywords remain in the prompt scheme as they are derived directly from data. Additionally, even if keywords have minimal effect, the model can still generate believable characters consistent with patient data.

Certain keywords, like "platinum blonde" or "wearing glasses" have a more notable impact than others, overriding descriptors such as "mother," "kindergarten teacher," and "educated" from the original prompt. However, terms like "open-minded" have negligible influence on results. The influence of keywords on generated images varies based on the model's training data. Reducing keywords in a prompt enhances control over the output by decreasing the weight of each keyword. Consequently, keywords not derived from patient data and precise appearance descriptions were eliminated, as they override data-derived keywords.

5. KEYWORD WEIGHTS. Prompt weights can be modified to adjust specific character features. The default weight is one, which can be increased by adding + or a number between 1.1 and 2 or decreased by – or a number between 0 and 0.9. + stands for a weight of 1.1, ++ adds a weight of 1.1^2, and so on. The same scheme applies for reducing the weights where - equals 0.9 and –– equals 0.9^2. For example, *"A beautiful, friendly, 65 years old, (overweight)+, educated woman"* assigns a weight of 1.1 to "overweight".

6. BASE PROMPT INPUT. The prompt scheme was modified from Budich et al. [27] to suit model specifics, resulting in the following base prompt input: `<age> <gender> <body weight> <family> <occupation> <education> <emotion> <ethnicity> <picture section>`. All keywords, except for *emotion* and *picture section*, were derived from data. The base

prompt scheme yielded satisfactory results, though additional keywords could further modify the images, such as altering hair color or style. Images for each character at time point *s0* were generated, selecting emotions based on their health status. The prompts used were:

- 65 years old women, overweight, mother, kindergarten teacher, educated, scared, European, portrait
- 32 years old man, normal weight, married, salesman, strong and focused, European, moderate drinking, portrait

7. REFINEMENT. In the initial phase, a range of output images is generated, providing the designer with options. Illustrated in Figure 36, even with the same prompt, Stable Diffusion produces diverse characters. From here, designers can adjust prompt component weights to experiment further.

7.5.3 *Image-to-Image Translation*

The character is finalized using Image-to-Image translation [80], refining the design by generating similar images. This process preserves key features like pose, hairstyle, and clothing, allowing the selection of the best character variant. Refer to Figure 36 (Step 2.2. Image-to-Image Translation) for the finalized female base character. The pipeline generates character variants from this final image, covering diverse attributes such as emotions, body weights, and ages crucial for depicting long-term medical narratives.

Image-to-image translation alters an image based on a text prompt. The *inpaint* approach allows users to specify regions for alteration by drawing on the image, as illustrated in Figure 37. To modify emotions, the facial region was painted. Adjusting weight and age involved including the outer boundaries of the head, considering potential changes in head size. Using inpaint helped to preserve features affected by unstable diffusion. Consistency was maintained by employing the same negative prompt as in text-to-image generation and adjusted the positive prompt by modifying emotion, age, and weight keywords.

Image-to-image translation was used to show the character's evolution from time step *s0* to *s2* in the SHIP study. Over ten years, the female character lost weight and recovered from fatty liver, while the male character gained weight and developed liver inflammation. Their facial expressions were adjusted, with healthy characters showing positive emotions and diseased characters showing negative ones. See Figure 36 for the female character's changing expression, body weight, and age.

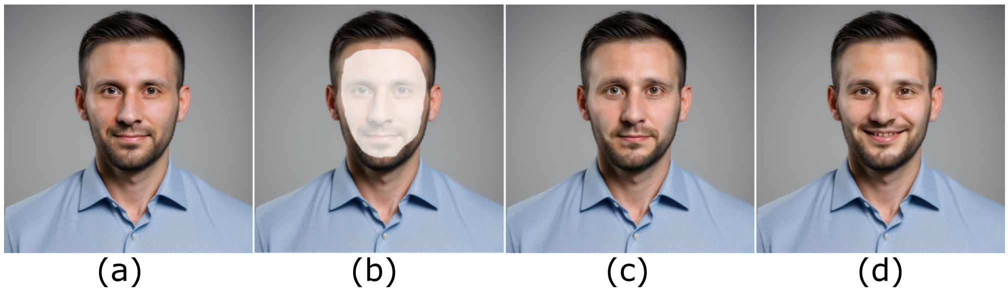


Figure 37: The original image (a) is combined with an inpainted area (b) and a text prompt to perform image-to-image translation. This allows for the generation of different emotions within the selected area, such as scared (c) and happy (d) expressions. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

7.5.4 *Generalizability*

Since disease risks are often a combination of genetic predisposition and lifestyle factors, the approach presented here, while demonstrated with a specific dataset, is adaptable to other datasets, e.g., the UK Biobank [182], the Rotterdam Study [84], and the German National Cohort (NAKO) [258]. The pipeline can be applied to any epidemiological study, but hard and soft criteria need to be adjusted based on the story intent. Cohort data, collected through studies such as the ones mentioned above, usually consists of sociodemographic data (employment, education, ethnicity, etc.), lifestyle information (preventive and risk factors, e.g., smoking, diet, activity, etc.), and health status and physical measures/biological samples (body weight, blood pressure, blood samples, etc.). These are the most important aspects of developing a narrative medical visualization focusing on one disease and featuring an individual health journey.

In this proof of concept, characters that have either a poor health status that improved due to lifestyle improvements, or a good health status that worsened due to poor lifestyle were created. Based on diseases tracked in the epidemiological study, concrete hard and soft criteria can be defined to find fitting patients. Sociodemographic data helps shape the AI-generated character's appearance, such as age in the prompt. Lifestyle information can be used to visualize lifestyle aspects relevant to a certain condition, e.g., a character can be depicted as smoking. Lifestyle information visualizes relevant aspects, like depicting a character smoking. Health data, physical measurements, and biological samples are used to represent the disease, its causes, symptoms, or consequences, such as body weight.

It is important to emphasize that since individual patients are used as models for the GenAI characters, consultation with a clinician – e.g. someone who works with the cohort data – is essential. This will ensure that the patients selected are *representative* and that the final character matches the physician's approval, especially with regard to risk factors for certain diseases.

7.6 Evaluation

Both characters were evaluated concerning the three GenAI-based steps in the pipeline: (1) Text-to-Image Generation, (2) Image-to-Image Translation, and (3) Refinement.

Regarding (1) Text-to-Image Generation, the ability of the generated images to represent the data was assessed. Given that these characters convey health details to the public, *authenticity* is crucial. As the underlying data spans a lengthy epidemiological study, the character must reflect changes over time, primarily in age and body weight. For (2) Image-to-Image Translation the ability to modify the original images using GenAI is examined while ensuring users recognize both versions as the same individual. Lastly, (3) Refinement is investigated by assessing how the generated characters can be refined to produce results that align more closely with standards upheld by designers and medical illustrators.

(1) and (2) are evaluated with a quantitative evaluation targeted at an audience without medical or design expertise in Section 7.6. (3) is evaluated with a qualitative evaluation with design experts familiar with medical application cases in Section 7.7.

7.6.1 *Participants*

To answer (1) and (2), a quantitative evaluation using an online questionnaire over six weeks was conducted. 29 participants completed the questionnaire, while 14 participants dropped out before completion (their responses were not included). The questionnaire and the responses are available as supplementary material. While a range of different backgrounds was covered, most participants are under 30, have a university degree, and live in Germany, see Table 5. Gender distribution is nearly equal between males and females.

7

Age	
18-24	7
25-29	14
30-34	4
50-54	1
60-64	2
>65	1

Gender	
Male	14
Female	12
Other	2
No answer	1

Education	
School pupil	1
Junior High Diploma	1
High School Diploma	7
Vocational training	2
University degree	17
Other	1

Residency	
Germany	26
Norway	2
Canada	1

Table 5: Participant metadata overview, displaying only selected response options.

7.6.2 *Study Design*

AUTHENTICITY WITH THE UNDERLYING DATA. The authenticity of the images in representing the underlying data was explored. Participants were tasked with assigning keywords to the images, covering six categories corresponding to each keyword used in the image generation prompt:

- AGE: young adult (18-29), middle-aged (30-59), mature (60+).
- BODY WEIGHT: overweight, normal weight, underweight.
- FAMILY: married, widowed, separated, divorced, single, parent.
- OCCUPATION: Military, Agriculture, Manufacturing, Construction, Natural Sciences, Traffic/Logistics, Sales/Trading, Business Organization, Health Care, Social Sciences (short forms based on the occupational sectors defined in the German Classification of Occupations 2010 – Revised Version 2020 [61]).
- EDUCATION: No qualifications, Secondary school-leaving certificate, High school diploma, Completed apprenticeship, University degree.
- EMOTION: angry, afraid, sad, calm, strong/focused, happy (Based on the "Feelings Wheel" developed by Willcox [259] for aiding individuals in identifying and articulating emotions. It is a structured framework that categorizes and subcategorizes different emotions. To keep the identification of emotions simple, only the six top categories were used.)

7

Age, body weight, and education were single-choice questions because each participant could only select one correct answer per category. In contrast, family, occupation, and emotion were multiple-choice questions, allowing participants to select all plausible options, as multiple options could apply to a person simultaneously.

A subset of keywords used to generate the image variations for *s0* and *s2* for the male (A and B) as well as female (C and D) characters were presented to the participants. The participants had to select all the images matching the keywords. The keyword combinations are:

A) 32 years old, Strong/Focused, Normal Weight

B) 65 years old, Scared, Overweight

C) 42 years old, Sad, Overweight

D) 75 years old, Happy, Normal Weight

The participants were provided with these short stories:

a) "Thomas Schmidt is a 32-year-old salesman. He is married and has a secondary school level. Although he has a poor diet, he has a normal body weight and moderate alcohol consumption but no alcohol problem."

b) "Thomas Schmidt is a 43-year-old salesman. He is married and has a secondary school level. Due to his poor diet, he became overweight and developed critical stomach fat and liver inflammation."

c) "Emma Winter is a 65-year-old woman. She is married and a mother. After graduating from high school, she became a kindergarten teacher. She is overweight and has been diagnosed with a fatty liver. However, she has no alcohol problem and is an ex-smoker."

d) "Emma Winter is a 75-year-old woman. She is married and a mother. After graduating from high school, she became a kindergarten teacher. She started exercising and lost weight during the last years and successfully recovered from her fatty liver."

For each story, they should select one image that best fits the character described in the text. This approach assessed whether participants correctly identified emotions, but also explored whether the selected emotions resonated with the audience or if they would choose images depicting other emotions.

RESEMBLANCE OF ALTERED IMAGES. It was assessed whether participants viewed images from image-to-image translation as depicting the same person as the original. Therefore, the original image was shown side-by-side with each of its alterations. Participants were asked to rate the resemblance of the characters in each pair of images on a 5-point Likert scale.

7.6.3 *Discussion of Results*

AUTHENTICITY WITH THE UNDERLYING DATA. In Figure 38, participants' keyword selections reveal that most (n=21) perceived both characters as *Middle Aged (30-59)*, aligning with the prompt for the male character's age. Although the female character is intended to be 65 years old, most participants selected an age range only slightly above this. Overall, participants tended to view the male character as younger and the female character as older, highlighting a noticeable age gap between them. Regarding body weight, participants categorized the male character as normal weight (n=28) and the female character as overweight (n=23), consistent with the original prompt. While participants mainly interpreted the male character's emotion as *calm* (n=27), which was not in the original prompt, they consistently chose positive or neutral emotions,

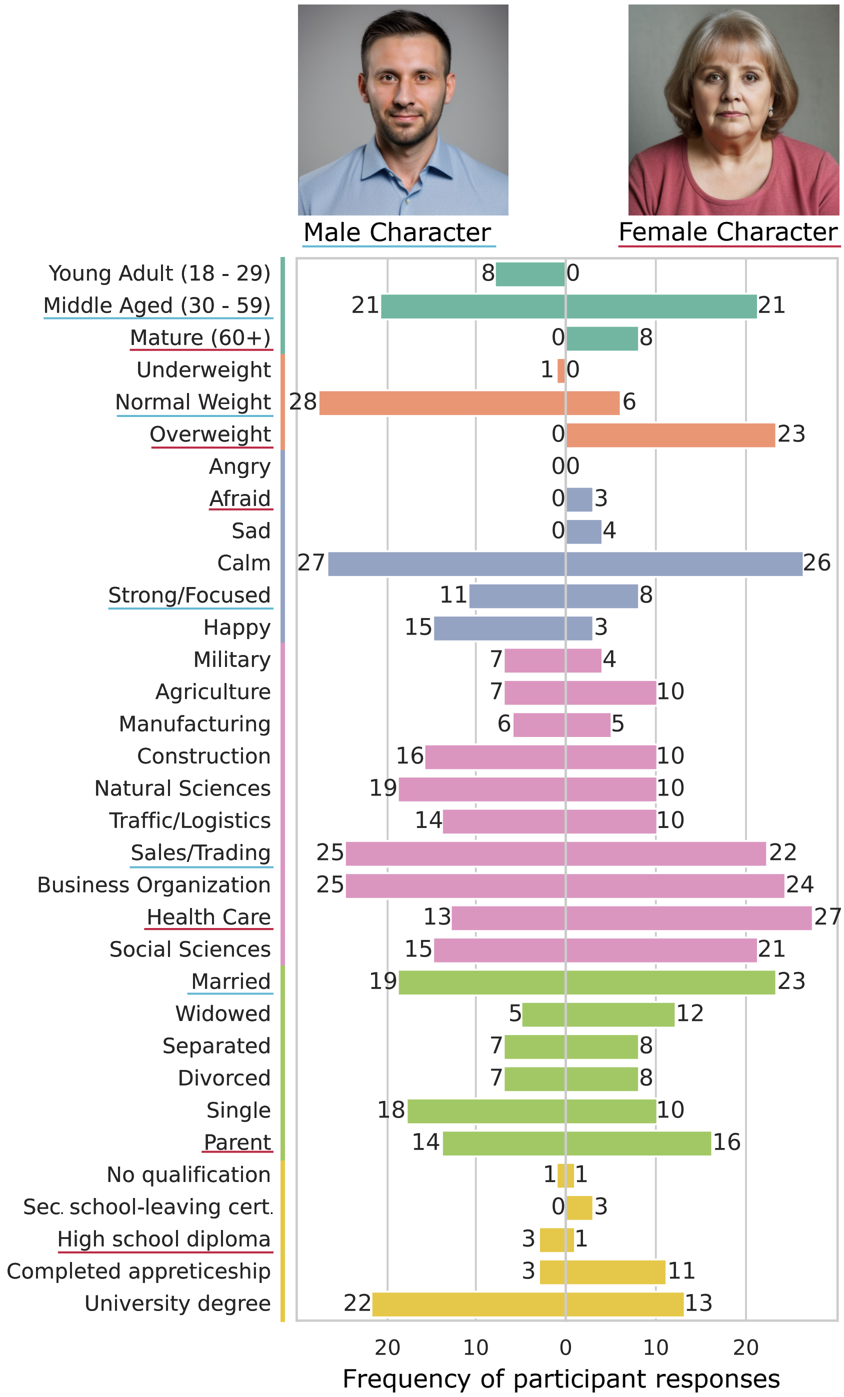


Figure 38: Responses for male and female characters by category: age, body weight, emotion, occupation, family, and education. Underlined keywords indicate prompt terms for male (blue) and female (red) characters. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

7

reflecting the intention for the healthy character to be depicted positively. *Calm* was also the most frequently selected emotion for the female character (n=26). Few participants opted for *afraid* (n=3), the emotion used in the prompt, or another negative emotion like *sad* (n=3). The keyword weights for emotions often had to be increased in the prompts to elicit more pronounced facial expressions. However, the findings suggest that the model struggled to produce strong negative emotions, as these were rarely identified by participants.

Regarding the characters' occupations, the most common responses for the male character were *Business organization* (n=24) and *Sales/Trading* (n=24), the latter aligning with the original prompt input. For the female character, the predominant response was *Occupations in health care, the social sector, teaching, and education* (n=27), which corresponds to the original prompt featuring the keyword *kindergarten teacher*. While participants may have based their choices on visual cues like clothing, these occupations are also heavily gender-typical, suggesting that gender alone could strongly influence the results.

The most frequently selected family term was *Married* for both the male (n=19) and female characters (n=23). Many participants also thought of the male character as being single (n=18). While the family was not specified for the female character, the keyword *married* was used in the original prompt for the male character. About half of the participants also thought that the male (n=14) and female characters (n=16) were parents, which is the case for the latter according to the underlying SHIP data. The male (n=22) as well as the female character (n=13) were assumed to have a high educational level, i.e., a University degree, by many participants. While the education of the male character was not specified in the data, the female character has a *High school diploma*, which was only selected by one participant. When asked whether the characters in the static images felt like real or computer-generated people, participants frequently chose both options without a noticeable tendency.

The results for selecting character images for a set of keywords are shown in Figure 39. Nearly all users chose character images with the correct age and weight. For A, the image generated with the keyword "angry" was selected most frequently (n=24), with the image generated with "strong/focused" as the second most selected (n=18). The angry expression, characterized by a furrowed brow, may also be interpreted as concentration or focus, indicating the ambiguity of this facial expression.

For B, C, and D, participants most commonly selected the correct images. However, for character C, the top image received only 17 votes (the top images for A, B, and D received at least 24 votes) indicating more difficulty in recognizing this emotion. Although correct images generally dominated the votes, participants often interpreted other images as depicting the same emotions. Furthermore, the model's subtle alterations in the images made distinguishing emotions challenging, as seen in the comparison of strong and calm emotions in Figure 39 A).

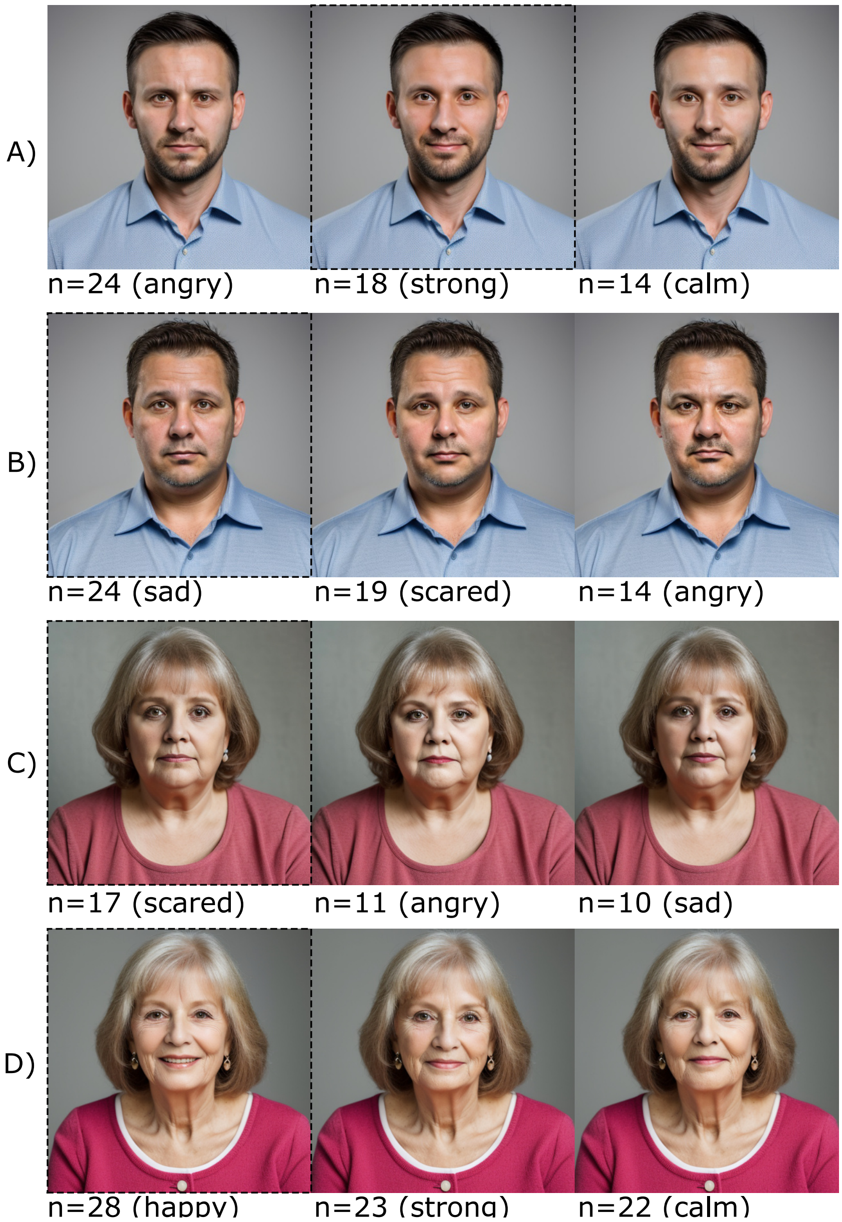


Figure 39: The three most frequently selected images for each keyword combination. Below each image, the number of responses and the emotion keyword used in the prompt are shown. Images aligning with the prompt are marked with a dashed frame. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

The most chosen image for each story is shown in Figure 40. Participants generally selected images with accurate age and weight. For the male character, the top images for both the a) healthy and b) diseased versions depicted anger. While anger in the dis-

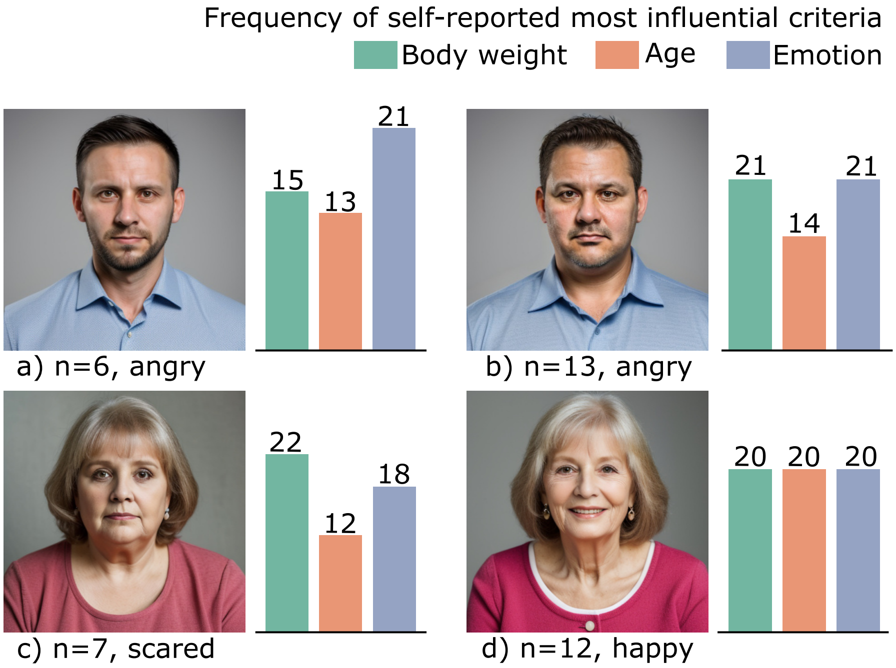


Figure 40: The most voted images for each story, including the number of votes received and the emotional expression. To the right of each image, the frequencies of self-reported criteria that participants found most influential in their selection are shown. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

eased character could reflect frustration about his condition, the reason for its prevalence in the healthy character is less clear. It might be that participants interpreted the expression not as anger but as focus, similar to observations in a previous task. For the female character, participants predominantly chose scared for the diseased version (c) and happy for the healthy version (d), aligning with the images selected for the stories later in the evaluation. This suggests participants often use negative emotions to depict disease. The task required a single image selection per character. Notably, options a) and c) received fewer votes, 6 and 7, respectively, compared to b) and d), which garnered 13 and 12 votes, respectively.

The participants were asked which criteria they considered for their answers. Body weight, age, and emotion were frequently selected. Participants had the option to enter their own criteria. Five participants used this option and stated that the following criteria played a role in their decision:

- Alcohol consumption, Dietary, Occupation
- Signs of health from poor diet such as eye bags, wrinkles, unhealthy skin color
- Bad teeth

The image generation model sometimes rendered crooked teeth, suggesting that the normal-weight male character had poor dental health when smiling. Although unintended, this coincidentally aligned with the character's backstory of having a poor diet that might affect his teeth. Participant responses indicate they also look for subtle health indicators like eye bags.

RESEMBLANCE OF ALTERED IMAGES. The results for the resemblance of altered character images are shown in Figure 41. Participants rated images where only the emotion was altered, but age and body weight were constant, as most similar. However, in cases where age and body weight changed, images of the female character transitioning from overweight to normal weight were seen as more similar than those of the male character going from normal to overweight. Emotional changes did not notably affect the similarity ratings. In the image-to-image translation process, the model frequently altered the noses of characters, especially giving overweight characters more upturned noses. This suggests that the model often applies learned features associated with certain body weights instead of preserving the original features, leading participants to see them as different individuals.

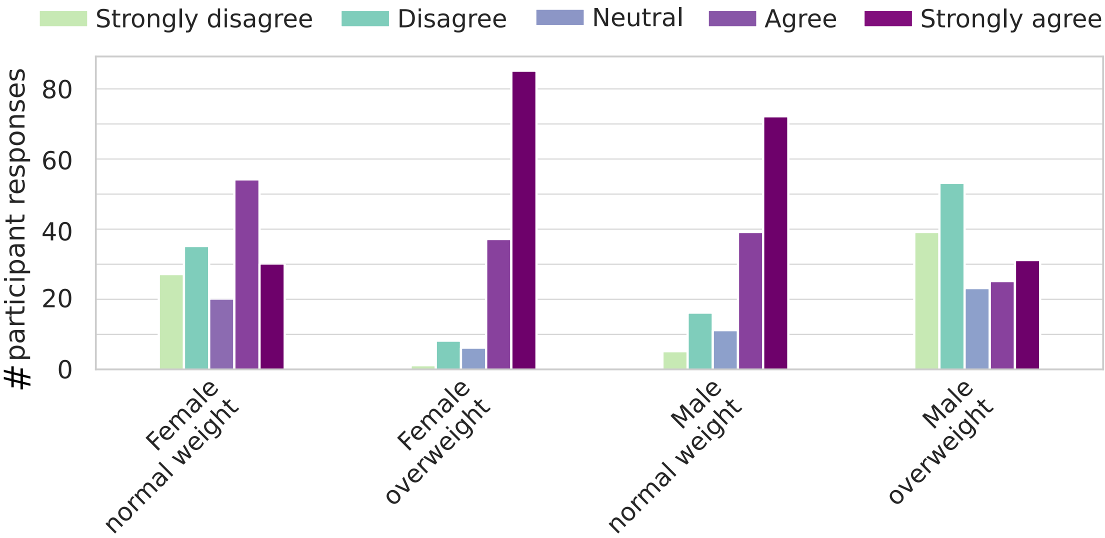


Figure 41: Frequency of participant responses regarding the similarity of two images, aggregated by gender and weight, in response to the statement: "Both images above show the same person, probably at different ages, weights, and moods." Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

7

7.7 EXPERT EVALUATION

To address (3) refinement step in the semi-automated character pipeline, new character images based on the quantitative user feedback were created and discussed with four different design experts regarding the character design criteria from Section 7.4.

7.7.1 *Participants.*

Feedback was gathered from four design experts *(D1 - D4)*, one of them being a co-author of the related paper [155].

D1 holds bachelor's degrees in design and computer science. They are familiar with medical visualizations and narrative medical visualization, and have practical experience in designing them.

D2 has a bachelor's degree in pharmacology and a master's degree in medical illustration. They have practical expertise in illustration, especially portraits, and working experience in a public health agency.

D3 has a master's degree in interaction design and holds a design professorship. They are familiar with illustration, character design, and storytelling and have practical experience in designing them.

D4 has a graduate degree in medical illustration and has experience as a healthcare communications designer.

7.7.2 *Study Design*

Miro [202] was used as a virtual whiteboard to show background information about SHIP, NAFLD, patient backstories, and data. A tier list was prepared on the Miro board for each of the ten design criteria from Section 7.4. The design experts were individually asked to rank all the image pairs (one pair being the character at *s0* and *s2* in the same style) from one (very good) to six (very bad) based on how well the image pairs fit each individual criterion. This was achieved by the design experts dragging the generated character images into the tier list, see Figure 42. Furthermore, they were asked to express their thoughts and reasons for their decisions, which were noted down according to the think-aloud method.

The main feedback of the participants regarding which cues are helpful to identify a disease was (1) showing more than a portrait section to enable the audience to assess body weight better and (2) showing more subtle signs of disease such as eye bags, unhealthy skin color, or bad teeth.

While the former was easy to implement by generating torso shots instead of portraits, the latter could not be achieved as the generative model was not able to realize

Figure 42: This screenshot outlines the setup of one of the Miro boards used during the qualitative evaluation with design experts (specifically D4). Orange boxes provide information about the data and the disease. Green-framed sections display ranking lists for each criterion (as introduced in Section 7.4). The design experts were asked to rank the character images according to each criterion by arranging them vertically, with higher placements indicating better fulfillment of the respective criterion. The detailed results are presented in Figure 44. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

these subtle changes in a sufficient way. For example, even when using *inpaint* to define that only the area below the eyes should be changed to make eye bags more prominent, it generally led to deformed eyes. Therefore, instead of depicting the male character's poor diet using bad teeth, he was placed in a fast-food restaurant. This framing also allows us to discuss the choice of backgrounds with the design experts in the next step by comparing the female character with a neutral background and the male character with a non-neutral background. Based on the torso shots, comic variants of the characters were generated. For this, the model checkpoint revAnimated V1.2.2 [278] was used. As there is a huge body of medical comics used for health education [247], the intention is to investigate whether photorealism or comic style is preferred by design experts in certain aspects. Six different pairs of images (three for each character, including the ones used in the previous user study) were created. Each pair shows a character at time point *s0* and *s2* in the SHIP, see Figure 43.

The following prompts were used for the generation of photorealistic torso shots:

- 65 years old women, overweight, mother, kindergarten teacher, educated, scared, European, waist and torso shot, standing, solid color background
- 75 years old women, normal weight, mother, kindergarten teacher, educated, happy, European, waist and torso shot, standing, solid color background, sports
- 32 years old man, normal weight, married, salesman, focused, European, waist and torso shot, standing in a fast-food restaurant
- 42 years old man, overweight, married, salesman, sad, European, waist and torso shot, standing in a fast-food restaurant

Figure 43: Refined characters. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

Using image-to-image translation, the comic adaptions were generated by adding the keyword "comic character" and including the LoRA inkSketch V1.5 [140]. The LoRA was activated using the command "lora:inkSketch_V2:1" in the prompt window.

7.7.3 *Discussion of Results*

While all design experts have a background in illustration or design and narrative medical visualization, they showed different preferences when discussing the generated images. Their rankings can be seen in Figure 44. It became apparent, that despite their professional background, personal experiences and preferences played an important role in their assessment. In the following, a detailed overview of the design experts' feedback regarding each criterion is presented. A concluding overview can be found in Section 7.10.1.

C1. RELEVANCE TO MEDICAL CONTEXT. *D1* favored the torso shots as they show the stomach fat, which is an important indicator for NAFLD. *D2* highlighted positively that the change in posture in the torso photo of the woman indicates a change in lifestyle. The portraits of the woman also show that she glows more in the second image, but because the rest of the body is not shown, the stomach fat is not visible. The comic version of the male character was ranked low by *D1* because the style felt inappropriate. *D2* said that using comics would give the creator the opportunity to exaggerate more on, e.g., symptoms. However, especially the female comic character does not look like the same person in the two images. *D3* thinks that the comic ver-

7

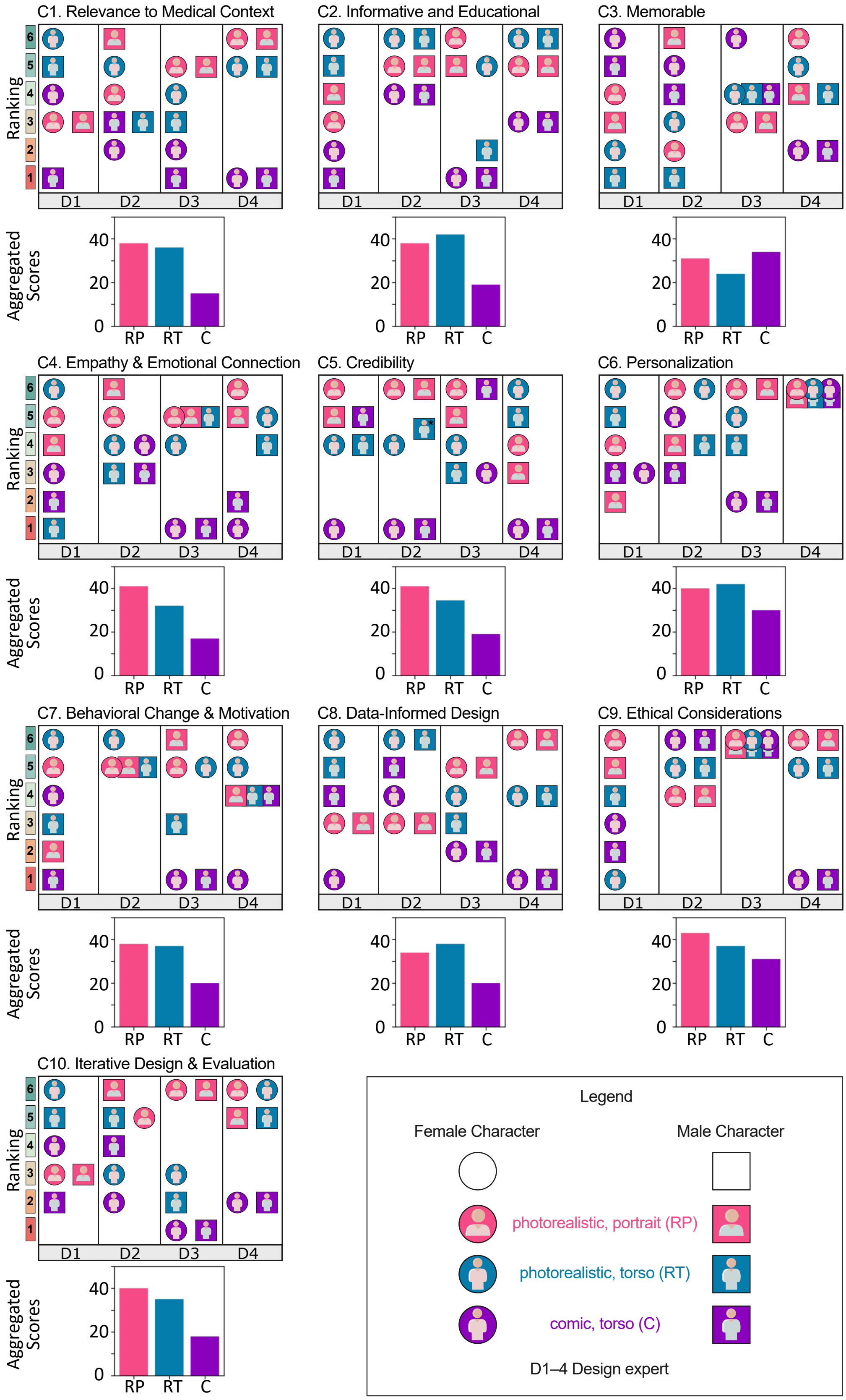


Figure 44: The design experts (D1–D4) ranked the characters for each design criterion (C1–C10) from 1 (very bad) to 6 (very good), allowing skipped ranks and ties. Aggregated scores for each image style (RP, RT, C) and criterion were computed by summing the ranking results. Image from Mittenentzwei et al. [155], available under a CC BY 4.0 license. No changes were made.

sion of the male character looks too strong and confident like he is in control of his life, which does not fit the story. *D1* appreciated the male torso shots' background for reflecting the character's poor diet, while *D3* found it too bold and stereotypical, preferring the less intrusive gray background.

C2. INFORMATIVE AND EDUCATIONAL. *D1* argued that having portraits that focus on the face leads the focus of the story toward the emotions of the characters. However, for NAFLD the body shows important indicators, in particular stomach fat. *D2* and *D4* agree, adding that it is also possible to guess the body weight based on the portraits but it is much harder. While *D1* sees a lot of potential in the comic style, but especially the images of the female character do not fit the story as she looks too young in the second image. Additionally, *D2* noticed that the facial expression for the comic version of the male character stays the same, and that shifts the focus of the viewer to the changes in the body. However, emotion can show the social implications of the disease, so it depends on the intent of the story if it should be shown or not. *D3* thinks that the emotions of the comic versions and the photorealistic body shot do not match the story's intent. *D1* liked that the background of the male torso shots shows context relevant to the story. *D2* liked that the female character glows and shows positive emotions after recovering from NAFLD. However, the deterioration is more stark in their personal opinion.

C3. MEMORABLE. *D1* argued that memorability, in general, depends on the ability of the individual viewer to memorize faces, which differs per person. For the photorealistic images, the portraits might be more memorable as they show the face in more detail. *D3* thinks that the expression of the photorealistic torso shot of the female character is especially memorable. *D4* thinks that the portrait of the female character conveys a warm and motherly personality while the portrait of the male character shows no traits that stand out in particular. Furthermore, the male torso shots look too cliche. *D1* said that the comic version of the female character has a lot of expression, making it more memorable. *D3* agrees, but thinks that the increase of memorability might come at a loss of conveying the relevant content. The male comic character is described as being too sleek which reduces memorability. *D2* thought that comics are, in general, more memorable. The change of the posture and clothing of the female character is done nicely in both comic and photorealistic torso shots. The stark shading and general art style of the male comic character make him a distinguishable character and not just a random person. In general, *D2* said that changing clothing and backgrounds for the individual time steps would be better. The portrait photo of the female character has a very good contrast between the clothes and the background.

C4. EMPATHY AND EMOTIONAL CONNECTION. *D4* highlighted that a general issue might be that all kinds of characters might act as projection areas, where viewers, despite recognizing that they are close to their own demographics, will think, "This happens to other persons; this will not happen to me." According to *D1* and *D2*, the portraits are better at conveying emotions as the face is visible in more detail. *D2* added that while the male character's portrait evokes empathy, it does not really look like the same person and the character ages too much. The torso shots of the male character do not evoke much emotion, according to *D2*. The facial expressions are neutral and the overweight level does not look unusual. *D3* mentioned that the photorealistic torso shot of the female character feels more distant to them than the portraits and the photorealistic torso shot of the male character. *D1* said that the photorealistic torso shot of the female character shows a large emotional difference between the time steps. In the second time step, she appears proud of overcoming NAFLD. On the other hand, the comic version of the male character looks unemotional. His aggressive glance establishes a certain distance from the viewer, and the story is not conveyed that well. However, the background is praised again. *D2* said that the style of the comic version of the female character is too different to feel like this is the same person which prevents the viewer from feeling empathy. Furthermore, the comic style looks more abstract, like taken from a textbook, which builds distance between the characters. On the other hand, the photorealistic images feel less distant and give the impression of being real people. *D4* favored the female portrait as it looks motherly and warm, but showing the character in a fitting environment, e.g., at home, would increase the emotional connection. While the male torso shots successfully convey the character's lifestyle choice by placing him in a well-known setting, such as a fast food restaurant, the character itself does not evoke empathy.

C5. CREDIBILITY. *D1* appreciated that in the photorealistic torso shots of the female character, similarities but also changes are visualized in a credible manner. The character in the second image (S2) looks a bit taller than in the first (S0), which should be improved. *D1* said that the portraits of the male character do look like the same person, however, the healthy version (S0) should look happy instead of neutral. The healthy version of the photorealistic torso shots of the male character looks good, but the diseased version does not seem to show the same person. In general, it is important to show the visual and emotional changes of the characters. Changing the clothing style in different images of the same character would add to this. *D2* said that in the case of the male character, the portraits, and the comic torso shots are identifiable as the same person. *D4* agreed in the case of the comic versions. There, the nose and jawline look similar. The background conveys the eating habits of this person. In the portraits of the female characters, most of the facial features are similar, however, the eyes appear to be too different. *D3* and *D4* said that the comic versions of the female character and the photorealistic torso shot of the male character do not look like the same person at all as their facial features are very different. *D3* only recognized a consistency in the portrait photos. While the poses in the torso shots were praised, the comics did not give *D3* a sense of credibility. *D4* said that the portrait of the female character looks most consistent.

C6. PERSONALIZATION AND AUDIENCE RELEVANCE. *D1* and *D2* thought that the female character matches the main risk group very well, as women tend to be affected more often than men. Additionally, the demographics in western Pomerania, where the SHIP takes place, show that more than every third person is 65 years or older. Furthermore, *D2* and *D3* said that the male character is too much of a businessman which might not resonate well with a large part of the audience. *D3* says that the male character looks too helpless in the portrait photo, which does not invite the viewer to resonate with him. *D2* assumed that a face might be more relatable than, e.g., body or pose. The portraits leave more room for imagination, potentially addressing a wider variety of body types. *D1* and *D2* assumed that comics might not resonate well with the older target group, but this would need further investigation. In the case of the torso shots for the male character, the context in the background increases personalization, according to *D1*. *D3* suggested that another strategy could be to not visualize detailed individuals but instead use archetypes. This way, a larger part of the audience might be able to resonate with one archetype as it is more general. However, it is difficult to create archetypes that are not too stereotypical. Additionally, it would be great for the audience to show both progressions for the same character. Starting with a healthy character, the story can show how an unhealthy lifestyle leads to NAFLD and how to recover from the disease by making lifestyle changes. *D4* does not see the need to apply this criterion to the characters as they are based on real persons and thus are relevant real-world examples of the disease.

C7. BEHAVIORAL CHANGE AND MOTIVATION. *D1* was unsure if a positive or negative example is more motivating for the audience but has a personal preference for the positive example, comparing it to the psychological concept of positive reinforcement. *D2* said that both examples can motivate, and it might also depend on the disease and what works best. For example, there might be a difference in the case of diseases where risk factors are hard to quit due to an addictive factor such as smoking, where an example of a character that quickly progresses towards a healthier lifestyle might not be appropriate. *D4* assumed that what works best for an individual viewer might be the character that is closer to the viewers on demographics and lifestyle. *D3* argued that the portraits have a different level of meaning, focusing on the emotions instead of on the physique. *D1* argues that the comic version of the male character looks cool, not diseased. Thus, the gravity of the disease is not conveyed. *D2* agrees, adding that the character does not look like he wants to change which will not be motivating for the audience. *D3* sees similar issues in the photorealistic torso shot of the male character.

C8. DATA-INFORMED DESIGN. *D1* said that the torso shots are crucial to show the critical stomach fat of the male character as this was one important aspect in the data. For both characters, the change of age and weight is visible in all the torso shots, except the comic version of the female character where the age does not seem to match the data. Concerning the portraits, only age is conveyed prominently, neglecting body fat. *D2* agrees, highlighting that the poor diet of the male character is also visible in the torso shots due to the background. For *D3* the age was not correctly conveyed in the comic versions. The female character looks too young in the second image, while the male character looks too young in the first image. *D4* thought as well that the photographs are generally better at conveying the data.

C9. ETHICAL CONSIDERATIONS. *D1* argues that the SHIP participants' privacy is guaranteed in all cases. As the portraits do look like real people, *D2* assumed that it is hard to get away from any resemblance with real people. *D2* added that in the comic versions, an accidental resemblance of a real person is unlikely. The comics are furthermore portraying the characters in a sensitive way as the proportions are still good-looking, e.g., going too much in the direction of a caricature might ridicule certain conditions. Portraits are more likely to avoid fat phobia, as it is mostly concentrated on the body. But since weight is an important aspect of the disease, *D2* thought it is reasonable to show obese bodies in the story. However, designers should think about what visual cues are crucial to show and focus on these. *D3* argued that it is not possible to avoid stereotypes completely, as there will always be some aspects that match certain stereotypes. *D4* said that they think the comic versions do not get the necessary sensitivity of the topic across. They questioned if the comic characters were respectful to the people from which the data was taken. *D1* assumed that it might be possible that images showing an expression that is too bitter might convey the impression of the character being isolated or stigmatized potentially triggering viewers with similar experiences.

C10. ITERATIVE DESIGN AND EVALUATION. For this criterion, design experts were asked to consider using one of the characters in a NAFLD story. They were tasked with selecting images for the iterative design process and suggesting improvements for the first step. *D1* preferred the photorealistic torso shot of the female character but suggested reworking the facial expression on the left side, such as reducing the shadow contrast. While portraits alone would not fully represent NAFLD, a mix of portraits for emotions and torso shots for physique or lifestyle would be ideal. *D2* preferred the male character's portrait, finding it most relatable. A stressful job with long hours leading to poor diet and NAFLD was seen as realistic. However, D2 suggested adding more intermediate images to show the characters' gradual journey and depict how they could recover by adopting a healthier lifestyle. *D3* thought the portraits are looking the most natural and credible. It is also a good start for the story to first create empathy so the audience cares for the character and only in the second step show them in the context of the disease. Similar to *D2*, *D3* also liked the idea of showing how a person develops and recovers from NAFLD. *D4* liked the story of the female character better. However, they disliked the concrete style of the comic. For a medical topic, they would opt for a softer style, e.g., inspired by watercolor.

7.8 reflections on study design and limitations

A specific GenAI tool for image generation was selected. Although the study aimed to establish a generalizable AI-enhanced character pipeline rather than evaluating specific tools, the choice of tool influences the results. Comparing various tools is challenging because outcomes depend on the underlying model (e.g., Stable Diffusion or Midjourney) and the specific checkpoints used, with many available on the Civitai website [37]. Due to this and the rapid development of GenAI models, which make prior versions irrelevant, a comparison of different models is not a valuable contribution. Investigating broader aspects like styles (photorealistic, artistic, etc.) and general strategies to create characters provides more meaningful insights for future character design.

During the evaluation with the design experts, it became clear that there are many different illustrative styles for visualizing characters. As only a small sample of styles is featured in the proof-of-concept, it is not representative of the whole body of illustrative styles. Keeping an illustrative style consistent adds another challenge additionally to the requirement to show images of similar persons with different body weights and ages. To guarantee a consistent style, more specialized models are necessary.

Overall, the issue of consistency and resemblance, as well as the inability to add certain visual cues (e.g., eye bags), highlights the limited control creators have over the generative process. As generative models can only reproduce what they learned from a set of training data, the way how the model was trained notably influences the results a creator can achieve. However, these issues can only be discussed regarding the results of the proof-of-concept approach for which a specific software was utilized, namely Stable Diffusion 1.5. It is important to note that other existing (commercial) software, such as SnapChat [221] or Adobe [2] might produce better results regarding altering facial images. Furthermore, regardless of the training data, the models still struggle with complex shapes such as hands. Stable Diffusion 3 was also tested (as it is currently the latest release), but did not provide an increase in the quality of the resulting characters. Since Stable Diffusion 3 was promoted with the promise of improving text rendering in particular, this was to be expected. In addition, Stable Diffusion 3 can provide improved resolution, but this is only noticeable with larger output images or small details, neither of which are relevant problems in the application example. Also, the set of available checkpoints and LoRAs is notably smaller for Stable Diffusion 3, which is why the study was conducted with Stable Diffusion 1.5.

The majority of participants in the study are under 30 years old and hold a university degree, resulting in limited representativity of the broader and more diverse lay audience targeted by the characters. Additionally, the chosen example of NAFLD and the associated characters may resonate more with older individuals, as NAFLD predominantly affects older adults. These demographic and contextual factors should be considered when interpreting the results, as they may introduce potential biases.

It is essential for character design to avoid stigmatization and stereotypes, ensuring sensitivity, accuracy, and authenticity to promote inclusivity. For instance, while body weight is a critical risk factor in the NAFLD example, equating obesity with disease and normal weight with health risks reinforces weight-related stigma [198]. Creators must carefully balance authenticity with sensitivity when designing characters.

Disease risk factors in the dataset are binary-coded, despite existing on a spectrum in reality. For instance, smoking is represented as true or false (see Algorithm 1), whereas the frequency or intensity of smoking could notably vary. This level of simplification was dictated by the underlying data and was not a deliberate design choice for character generation. Furthermore, certain nuanced details may exceed the capabilities of GenAI. For example, modifying small features such as teeth, which occupy only a few pixels, proved unfeasible. Simplification, therefore, plays a critical role in GenAI-based character design.

Additionally, the dataset is limited to physical and biochemical health factors. Character design could be enhanced by integrating behavioral or socio-economic data, such as stress levels or social support. When such data is not available in the dataset, supplementary research or consultation with treating physicians can provide insights for adding characteristics that align with both the database and the narrative.

The study results are referenced to address research question RQ2 and its subquestions, RQ2.1, RQ2.2, and RQ2.3. The discussions of the RQs need to be interpreted in the light of the aforementioned limitations.

7.9 DISCUSSION OF RESEARCH QUESTIONS

RQ2.1 HOW TO GENERATE CHARACTERS THAT ARE AUTHENTIC TO THE UNDERLYING DATA?

The generated characters were found to authentically reflect the underlying epidemiological data in terms of key attributes like age and body weight. Interpreting emotional states, however, proved to be more challenging. Our evaluation results show, that users searched for additional clues for diseases, such as unhealthy skin color, eye bags, or bad teeth. Including these visible symptoms in the character generation process would create a stronger impression. Employing eye-tracking technology in such assessments can reveal which parts of the image capture users' attention the most.

RQ2.2 HOW TO ALTER IMAGES THAT STILL RESEMBLE THE SAME PERSON?
Resemblance of altered images to the original character was a common critique, though it varied depending on the factor being modified. Changes in emotion did not negatively impact resemblance, whereas modifications to body weight and age often resulted in new character images that lacked sufficient similarity to the original. Notably, with the settings used, image-to-image translation tended to alter not just skin features (e.g., adding wrinkles to signify aging) but also other attributes such as eyes, noses, and occasionally even hair color. While image-to-image translation demonstrated potential for maintaining character consistency, it requires further refinement to ensure coherence during transitions. Specialized tools for ensuring character consistency, e.g., face swap models [187], could be used in case the character's face is to be preserved. Changing features like the body weight may also impact facial features to some degree.

RQ2.3 HOW TO REFINE THE GENERATED CHARACTERS TO MEET DESIGN STANDARDS?
Discussions with design experts highlighted that character design has many variables and preferences, such as artistic styles, facial expressions, and body poses, that differ per person. The evaluation highlighted key strengths and limitations across the styles and suggested additional criteria for designing characters that effectively resonate with target audiences. Torso shots were praised for their ability to visually depict critical medical indicators, such as stomach fat, an important marker for NAFLD (higher scores, e.g., in *C2: Informative and Educational* and *C8: Data-Informed Design*). On the other hand, portraits were preferred for the depiction of showing emotions (higher scores, e.g., in *C4: Empathy and Emotional Connection*). Contextual backgrounds, such as a fast-food restaurant, further enhanced the narrative by illustrating lifestyle factors. However, the background can also be distracting. Comic styles often struggle with maintaining consistency in conveying age and emotional states. This affected audience recognition. Concerns were raised regarding sensitivity and suitability for the target group. On the other hand, comic styles provide a broad range of possibilities to depict different aspects of a disease (e.g., emotions can be emphasized more clearly than with photorealistic representations) and comic characters might increase memorability (higher scores in *C3: Memorable*) but the current images have several design flaws as described before. Additionally, the evaluation highlighted new considerations for character design.

RQ2 HOW TO CREATE DATA-DRIVEN CHARACTERS FOR DISEASE STORIES?
The use of GenAI for semi-automated, data-driven character generation in disease stories has shown promising results. However, several areas for improvement were identified in RQ2.1, RQ2.2, and RQ2.3, leading to the discussion of RQ2.

While the pre-trained Stable Diffusion model successfully depicted certain character attributes, such as age and body weight, it struggled to represent other disease-related aspects, such as poor dental health. This limitation is likely due to the model's training data, which may not have included sufficient examples of these features. GenAI models specifically trained to depict diseased individuals could enhance the quality of character generation and serve as a valuable tool in this process. Similarly, image-to-image translation exhibited several flaws when modifying the age and body weight of characters. To improve the resemblance between altered images, further investigation into alternative tools and techniques can improve the quality of future characters.

Additionally, discussions with design experts regarding character refinement emphasized the importance of portraying characters in various situations and levels of detail. Depicting characters in specific environments, such as their homes, can foster empathy and add depth to their representation. Furthermore, including both full-body images and portraits allows for greater control over audience focus—highlighting either physical symptoms and risk factors or the character's emotions. The design experts also explored different artistic styles but concluded that an analysis of the target audience is necessary to determine which style would resonate best with them.

7.10 CONCLUSION

Including GenAI in the character design process for data-driven stories may enhance character design and storytelling, uniting creativity and design possibilities. This research explores semi-automated character design through tailored prompts for data-driven stories to improve health communication competence. Blending human design sensitivities with data analytics techniques, the pipeline enables the iterative production of characters that may contribute to telling more personalized, engaging, and understandable data-driven stories about health and wellness.

7.10.1 *Study Implications*

In the following a concluding overview of the results in the context of the research questions is provided.

To effectively convey the character's health journey (or the portion thereof), it is important to go beyond presenting just two images at the beginning and at the end of the journey. Including transitional images that illustrate gradual changes in between can improve narrative clarity and engagement. Combining photorealistic portraits with torso shots provides a more comprehensive narrative. Portraits foster emotional connection, while torso shots highlight physical health markers, creating a balanced and engaging depiction.

The appeal of character design should be considered alongside *C6: Personalization and Audience Relevance*, which focuses on sociodemographic factors. Elements like comic style may vary in appeal depending on the target audience, influenced by factors such as age and culture. Different comic styles might be perceived as more or less appropriate and appealing. Experts recommend exploring archetypes that strike a balance between relatability and generalization while being cautious to avoid stigmatized stereotypes.

Integrating relevant settings, such as a home environment for a female character or a fast-food restaurant for a male character, can enhance narrative immersion and effectively communicate lifestyle factors. However, the chosen background should avoid being overly complex, distracting, or stereotypical to maintain clarity and relevance.

7.10.2 *Future Work*

Medical character creation provides individual challenges due to sensitive data in medical contexts. Medical topics evoke emotions, necessitating appropriate, reassuring characters. A character's reflection of the psychological and physical changes caused by disease is vital for aligning with the narrative, as demonstrated in this case. Based on the proof-of-concept, GenAI was utilized to create data-driven characters for the application case of NAFLD. The following future research directions are proposed:

GENERALIZABILITY. Validation against diverse datasets and character sets can verify the generalizability of this approach. Expand the study to a wider range of disease data, identifying classes of diseases and personas for characters.

BROADEN THE APPLICATION AREA. This approach is tailored to epidemiological study data, where the creator can choose from a broad set of patient data entries to find an appropriate base for a character for a data-driven disease story. However, these stories are also relevant beyond epidemiology. Exploring how the pipeline can be applied to other types of data could expand the scope of this approach. In situations where patient data is unavailable, representative personas could be developed in collaboration with physicians.

EXPLORE ARCHETYPE-BASED DESIGNS. To support creators with limited experience in character design, providing generalizable archetypes that avoid stigmatized stereotypes while resonating with broader audiences, would be very helpful. These archetypes must be tested across multiple narratives to determine their versatility and impact on audience engagement.

EVALUATE STYLE VARIATIONS. Identifying appropriate styles for different target groups, such as photorealistic and different artistic styles, can guide the character design process in the future. If a set of popular styles can be identified, model checkpoints can be trained and published that are specialized for these styles.

PROVIDING CONTEXT. The influence of context, such as backgrounds in character images, needs to be investigated more thoroughly to identify appropriate settings and designs.

EXPANDING TO VIDEO AND VOICE GENERATION. This approach can be expanded to video generation, including the use of an artificial voice.

TOOL BUILDING AND PIPELINE AUTOMATION. In its current form, this proof-of-concept requires user input at several stages of the content generation pipeline. Reducing the manual effort in crafting narrative visualizations will be a key focus of future work. For example, using large language models for generating story arcs and matching characters at several stages of their disease.

PERSONALIZED STORYTELLING AND DIGITAL TWINS. Given the current approach one story is crafted for informing a set of patients. With the increasing interest in personalized medicine, crafting stories that match a single patient's data may be of interest. Creating a prototype story that is automatically completed using the patient's data may be a first step in providing a more personalized experience.

8 THE NARRATIVE STAGE: CHARACTER-DRIVEN STRUCTURE

Characters can influence the structure of a story. Therefore, this chapter explores the interplay between the two key story ingredients: character and structure (see Figure 45). This is particularly relevant in character-driven story structures. In this chapter, a character-driven story structure specifically designed for disease stories is proposed.

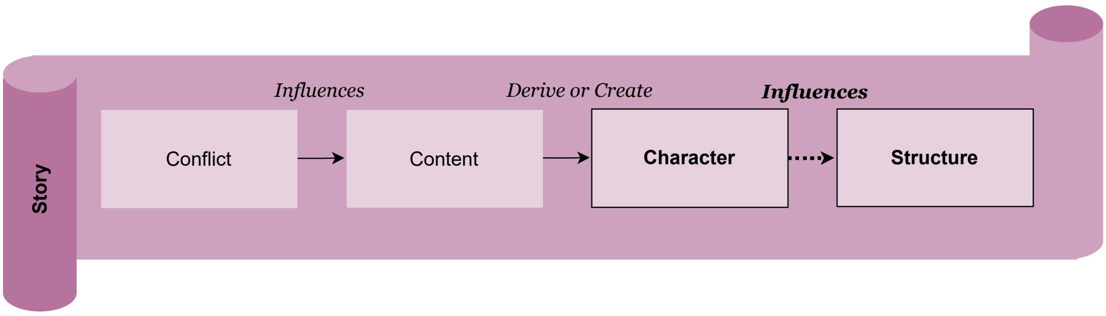


Figure 45: The narrative stage of the disease story design process (cf. Figure 9) with *character*, *structure*, and its intersection highlighted.

Character-driven story structures center on the main character's journey throughout the narrative, raising the question of who should serve as the main character in a disease story. While the previous chapter focused on patient characters, this chapter compares three approaches: a patient as the main character, a physician as the main character, and a story without a human main character. For this, a character-driven disease story structure is developed and the following research questions are explored:

RQ3 What is the impact of a human protagonist on health communication?

- *RQ3.1* Do human protagonists increase engagement in disease stories?
- *RQ3.2* Do human protagonists increase the users' motivation to improve their personal lifestyle?
- *RQ3.3* Does a patient as the protagonist increase identification and, in turn, enhance self-referencing?
- *RQ3.4* Does a physician as protagonist increase the credibility of the presented information?
- *RQ3.5* Do human protagonists increase the time users spent in the disease story?

8.1 TOWARDS A CHARACTER-DRIVEN STORY STRUCTURE FOR DISEASE STORIES

Meuschke et al. [152] emphasize the necessity of having at least one character in narrative medical visualizations aimed at disease communication. If a story includes only one character, that character is most likely the protagonist. The protagonist serves as the main character, intended to evoke sympathy from the audience [219]. This research examines how the identity of the protagonist influences user perception within a disease-related story. To explore this, a narrative centered on CSVD is utilized.

Collaboration and Personal Contribution

This chapter is partially based on the following publication: [156]. The conceptualization, story design, and implementation were done by the author of this thesis. The submitted work has been prepared with all co-authors who provided feedback throughout the project and writing phase.

Campbell's Hero's Journey [28] is a widely known character-driven structure and applied to disease stories shows parallels to structured documentations of patients' health. Devi et al. [50] developed a model outlining a patient journey, detailing the steps an individual navigates when managing a chronic illness. A related concept, introduced by Warner et al. [250], involves clinical vignettes in oncology. Clinical vignettes are widely used in medical education, where structured fictional scenarios help assess trainees' understanding of clinical situations. Building on these concepts, Campbell's Hero's Journey is adapted to create three distinct disease story conditions:

- a patient as the protagonist
- a physician as the protagonist
- a story without a human protagonist

A user study is conducted to evaluate the conditions regarding the presence, identity, and role of a protagonist in disease-related narratives, using a set of evaluation criteria derived from narrative visualization and health communication research.

Supplementary Material

Supplementary material for this chapter can be found at:
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/character-driven-structure`

8.2 RELATED WORK

As characters, story structures, and modes were already discussed in Section 2.2.2, the content of this section focuses on evaluation criteria for different aspects of story characters. Evaluation criteria that may be affected by the choice of different protagonists are drawn from research on health narratives. Health narratives, often text-based stories, are designed to enhance public health communication [21], while narrative visualization focuses on effectively conveying data-driven stories [7]. A key objective of narrative visualization is to enhance user engagement. O'Brien et al. [178] identified six key attributes of engagement and developed specific questions to measure them.

Additional evaluation criteria include the extent to which users identify with the protagonist and engage in self-referencing, where they relate the events of the story to their own lives. Stories can be conveyed from various character perspectives and through different point of view (POV). The POV defines the narrator's role in the story, with many narratives being presented either in the first person (e.g., from the protagonist's perspective) or the third person (from an external, uninvolved narrator).

Research by Charles et al. [30] demonstrated how altering the narrator's perspective can generate variations of the same story. Previous studies have explored the impact of POV in health-related narratives, particularly in relation to user identification with the protagonist [31]. Strong identification can enhance self-referencing, a process in which individuals connect the narrative's content to their own lives [31]. Findings on the influence of POV on identification levels remain inconclusive. Some studies suggest that first-person narratives enhance identification with the protagonist [168, 169], while others have failed to replicate these results. Furthermore, no research has established a direct link between POV and self-referencing behaviors [32]. Regardless of the chosen POV, protagonists can serve as either positive or negative role models. Research by Chen et al. [31] indicates that users are more likely to identify with positive role models who adhere to guidance and successfully avoid adverse outcomes.

Emotions play a crucial role in the effectiveness of health-related narratives, particularly regarding persuasion. The progression of emotions throughout a story, known as *emotional flow*, is a key factor in understanding the user's experience. Alam et al. [4] examined emotional flow in narratives by analyzing shifts between behavioral status and behavioral intention.

Ensuring high *perceived credibility* of information is essential, as users may otherwise doubt that a visualization accurately represents real issues and risks [58]. The Meyer modification of the Gaziano-McGrath scales [153] is a widely recognized tool for assessing credibility in online media and narrative visualizations [237, 242]. Credibility can be evaluated based on various attributes, including *accuracy*, *fairness*, *trustworthiness*, *bias*, and *completeness*. In addition to assessing the credibility of information itself, these measures can also be applied to individuals and institutions [38, 135].

8.3 THE ROLE OF CHARACTERS IN DISEASE STORIES

Chapter 7 introduced patients as potential protagonists in disease stories. However, domain experts, such as physicians, can also serve as protagonists in disease stories. The overall impact of featuring a human protagonist in disease stories remains uncertain.

The study includes three variations of the story, maintaining a consistent design in terms of colors and visual elements, as illustrated in Figure 46. The story versions

(1) patient-centered **(2)** physician-centered **(3)** base condition

Figure 46: Excerpts from the three story versions. We kept the overall design similar throughout all versions. In the brain visualizations, the damage caused by CSVD is highlighted in orange. The protagonists are included through text and, in the case of the patient, icons. The first row shows the introduction of the protagonists being the **(1)** patient and the **(2)** physician. For our **(3)** base story without a protagonist, we introduced the disease itself. In the second row, we use the phrases **(1)** "Emma, like 10% of female CSVD patients, is between 50 and 60 years old. About 20% of male patients are between 50 and 60 years old.", **(2)** "Ms. Schreiber observed that 10% of her female patients and 20% of her male patients with CSVD are between 50 and 60 years old.", and **(3)** "10% of female CSVD patients and 20% of male CSVD patients at the University Hospital in Magdeburg are between 50 and 60 years old." In the third row, we used the phrases **(1)** "The MRI made it clear, that Emma really does have CSVD.", **(2)** "The MRI made it clear that one of Ms. Schreiber's patients really does have CSVD.", and **(3)** "The MRI made it clear, that this patient really does have CSVD". Image from Mittenentzwei et al. [156], available under a CC BY 4.0 license. No changes were made.

8

are derived from the design described in Section 4.2. Photographs were incorporated to represent human protagonists, who were referenced throughout the narrative text in each section. In the case of the patient protagonist, additional representation was provided through icons.

The primary objective is not to provide a detailed explanation of the disease or to conduct a design study on this specific case. Rather, the focus is on comparing different protagonists to assess their impact on key evaluation metrics, including user *engagement*, *identification*, perceived *credibility* of the presented information, and perceived *length* of the story. With the focus on the different protagonists, the study does not consider the following aspects:

- variations in the number of characters
- differences between abstract (e.g., stylized or comic-like characters) and realistic (e.g., photographs of real individuals) character representations
- the POV (e.g., first-person vs. third-person)
- design choices involved in story creation, such as color schemes and visualization styles

8.3.1 *Structure of a Disease Journey*

The Hero's Journey framework was modified to create a *disease journey* template designed for character-driven disease storytelling. This template outlines the integration of the four fundamental storytelling ingredients [58]: *characters*, *conflict*, *content*, and *structure*. The Hero's Journey framework introduces three key *characters*: the *protagonist* and two supporting roles, namely the *mentor* and the *antagonist*.

- *The protagonist:* can be either a human or an object (e.g., an organ), as long as it elicits empathy from the audience.
- *The mentor:* can take the form of a person, such as a physician advising a patient, or an object, like a book providing guidance.
- *The antagonist:* is represented by the disease.

The *conflict* emerges as the protagonist confronts the antagonist, symbolizing the struggle of being afflicted by the disease and the efforts to overcome it. The *content* of the story is shaped by the different stages of the disease journey, while the sequence of these stages establishes the *structure* of the narrative.

The disease journey is structured into three interconnected layers, as illustrated in Figure 47. The first layer consists of the **stages** of the disease journey, representing

Learning Objectives

The motivation for communicating pathological growth and shrinkage – to support informed consent – can be framed as learning objectives. Cognitive objectives emphasize remembrance of CSVD and its connection to hypertension, following Adar et al. [1]:

- The viewer will *recognize* CSVD as a common consequence of hypertension.
- The viewer will *recall* CSVD symptoms and consequences.

Affective objectives focus on placing and behaving, as defined by Lee-Robbins et al. [132]:

- The viewer will *accept* the CSVD as a consequence of hypertension.
- The viewer will *demonstrate* behaviors consistent with the CSVD as a consequence of hypertension.

These objectives guide the design of disease stories but are not the focus of evaluation in the following studies; rather, they provide a structured rationale for this research.

the challenges the protagonist must overcome. These stages are arranged in a circular sequence, reflecting the protagonist's ultimate goal of returning to their original life. The journey is further divided into three acts: *departure*, *initiation*, and *return*. The *departure* act marks the point where the protagonist is forced to leave their normal life due to awareness of the disease and embark on the journey. Meeting the mentor serves as a transition into the second act, *initiation*, where the protagonist faces the antagonist and the central conflict. Finally, in the *return* act, the protagonist navigates their way back to the known world. The third layer, **context**, distinguishes between the *known world* and the *unknown world*, emphasizing the protagonist's level of familiarity with different stages of the journey. Events occurring in the *known world* are within the protagonist's prior experience, while those in the *unknown world* introduce new and unfamiliar challenges.

PATIENT-CENTERED STORY STRUCTURE. The first version of the story features Emma, a 59-year-old patient, as a fictional yet representative protagonist. The narrative follows a chronological sequence, illustrating the typical experiences of a patient upon receiving a diagnosis, as shown in Figure 47, left. In this version, the patient is portrayed as the protagonist in a disease story, battling the disease as the antagonist.

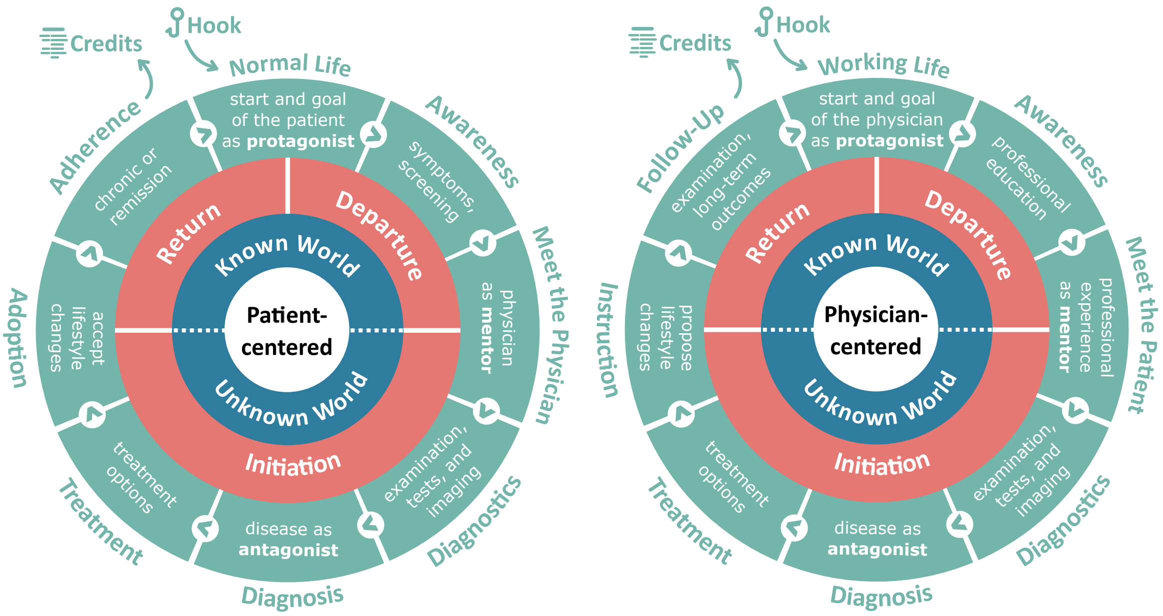


Figure 47: The *disease journey* adapted from Campbell's *Hero's Journey* [28] for patient-centered (left) and physician-centered (right) disease stories. The protagonist has to go through different **stages** which are divided in three **acts**, and in the **context** of the protagonist's familiarity with the events. The story starts with a hook and ends with credits showing the story authors' affiliations. Image from Mittenentzwei et al. [156], available under a CC BY 4.0 license. No changes were made.

Physicians play a mentor role in providing support. However, to ensure that all measured effects in the evaluation are attributed solely to the protagonist, physicians are not introduced as distinct supporting characters, nor are they given names or visual representations.

The story begins in the **known world**, depicting events familiar to the patient. During **departure**, the patient is forced to leave their **normal life** due to the onset of **symptoms** and **meets the physician**. The **initiation** act introduces challenges in the **unknown world**, where the patient undergoes various **diagnostic** procedures and tests to obtain a **diagnosis** and receive appropriate **treatment**. During the **return** act, the patient reintegrates into the **known world** while making necessary lifestyle **adoptions**. Over time, the effects of **adherence** to these changes become evident, influencing the progression of the disease, e.g., whether it becomes chronic or enters remission. The ultimate goal is to restore **normal life** as much as possible. The story concludes with a message on prevention, illustrating how the patient advises family members to take proactive measures before symptoms arise.

Although the patient's story could be told from a first-person POV, a third-person POV was chosen to ensure consistency across all three versions, making them more comparable. Additionally, this narrative structure aligns with established frameworks in the medical field, as discussed in Section 8.2. The story follows a format similar to patient vignettes, which are commonly used in clinical contexts to depict fictional patient cases. These vignettes can be developed by domain experts to serve as the foundation for patient-centered disease stories following the proposed structure.

PHYSICIAN-CENTERED STORY STRUCTURE. The second version of the story features a clinical cooperation partner, a neurologist with extensive experience in CSVD, as the protagonist. As in the first version, the physician remains the sole central character, without the introduction of supporting characters with names or visual representations. In many stages, the physician-centered story aligns closely with the patient-centered story, as illustrated in Figure 47, right.

The story begins with the **departure** act, depicting the physician's **working life**, which revolves around clinical routines within the **known world**. **Awareness** of the disease arises through professional training and experience. Upon **meeting a patient**, the physician's expertise serves as the mentor. However, since the disease manifests differently in each patient, this encounter leads the physician into the **unknown world**. During **initiation**, the physician initiates various **diagnostic** tests and examinations to establish a **diagnosis** and subsequently explains **treatment** options. The **return** to the **known world** occurs when the physician **instructs** the patient on necessary lifestyle modifications. Through **follow-up** examinations, the physician evaluates the long-term progression of the disease while returning to clinical routine. The story concludes with a preventive message, highlighting the physician's advice on the importance of proactive measures in reducing the risk of developing the disease.

BASE CONDITION WITHOUT A HUMAN PROTAGONIST. The third version of the story is structured as a report without a human protagonist. This format serves as a baseline for assessing the impact of a human protagonist on the perception of a disease story.

8

Meuschke et al. [152] introduce a storytelling template for diseases, based on insights from online health blogs (introduced in Section 2.6). Similar to this version of the story, these blogs usually convey information about disease progression without featuring human characters. While there are many similarities, a key distinction lies in the sequencing of information. The template proposed by Meuschke et al. begins with a general definition of the disease, incorporating statistical data and an anatomical overview. It then progresses to more specific details, covering symptoms, diagnosis, treatment options, prognosis, and preventive measures.

The narrative follows the chronological progression of a disease, reflecting the stages patients typically experience. It begins with an **introduction to the disease**, followed by an overview of **institutions** specializing in its treatment. Next, common **symptoms** and **risk factors** are outlined before explaining the **diagnostic** process. Once a **diagnosis** is made, a detailed **explanation of the disease** follows, including its potential

development over several years and the impact of appropriate **treatment**. Additionally, key **lifestyle changes** are emphasized. As in the other versions, the story concludes with a message on the importance of prevention.

8.4 EVALUATION

This study examines how the story's protagonist affects user perception based on established criteria in narrative visualization. The evaluation focuses on *user engagement*, *identification*, *self-referencing*, *perceived credibility*, *emotional flow*, and *perceived time* spent in the story.

8.4.1 *Participants*

A total of 34 participants were recruited through various channels, and the survey remained available online for two weeks. It was distributed via the crowdsourcing platforms Survey Circle [101] and SurveySwap [101], shared in relevant groups on LinkedIn [40] and Reddit [14], and sent to university students. Four participants were excluded from the analysis, as they indicated that they did not watch the story in its entirety. Two cited a lack of interest, while the other two encountered technical issues that prevented the story from loading properly in their browsers. The final dataset consisted of 30 valid responses: seven participants (23%) viewed the patient-centered story, ten (33%) experienced the physician-centered version, and 13 (43%) were assigned to the base story.

Participants provided informed consent, acknowledging that their participation in the survey was voluntary and that the anonymously collected data could be used for research purposes. Respondents were given the option to skip any personal questions. Information was gathered on participants' demographic background, general education, and self-assessed literacy in both visualization and medical topics, as shown in Table 6. The overall educational level was relatively high, with the majority of participants holding a university degree.

8.4.2 *Study Design*

The digital questionnaire was designed using the *SoSci* [133] software. Each participant was assigned one of the three story versions and was presented with the questionnaire immediately after viewing the story.

Following the approach for evaluating narrative medical visualization by Meuschke et al. [152], a five-point Likert scale was used to structure the questionnaire. Various sets of questions were included, each targeting a specific aspect of user perception, as outlined in the following sections.

Demographics					
Gender		Country		Age	
male	17 (57%)	Germany	21 (70%)	18-24	19 (63%)
female	11 (37%)	Norway	4 (13%)	25-34	6 (20%)
non-binary	2 (7%)	Other	5 (17%)	35-44	5 (17%)

Education and Literacy							
Education		Profes-sional visual literacy		Profes-sional health literacy		Experience with neurological diseases	
university degree	21 (70%)	SA	5 (17%)	SA	5 (17%)	SA	4 (13%)
high school graduates	6 (20%)	A	9 (30%)	A	3 (10%)	A	3 (10%)
middle school graduates	1 (3%)	N	5 (17%)	N	4 (13%)	N	8 (27%)
left school without a degree	1 (3%)	D	4 (13%)	D	8 (27%)	D	6 (20%)
completed training	2 (7%)	SD	7 (23%)	SD	10 (33%)	SD	9 (30%)

Table 6: Demographics and visualization and health literacy of study participants.

ENGAGEMENT. Engagement was measured using the framework proposed by O'Brien et al. [178], which identifies six key attributes of engagement. Two attributes, Aesthetics and Perceived Usability, were excluded from the evaluation, as they remain constant across all story versions and are not influenced by the protagonist. The remaining attributes were assessed through statements adapted from O'Brien et al.:

- *Focused Attention* (e.g., "I forgot about my immediate surroundings while viewing the story.", "I lost myself in the story."),
- *Novelty* (e.g., "The content of the story incited my curiosity.", "I felt interested in the story."),
- *Endurability* (e.g., "Viewing the story was rewarding.", "I would recommend viewing the story to my friends and family."), and
- *Involvement* (e.g., "I felt involved in the story.", "Viewing the story was fun.").

IDENTIFICATION AND SELF-REFERENCING. The measurement of *identification* with the protagonist was based on the guidelines established by Chen et al. [31]. Example statements used for evaluation include: "The main character and I are similar kinds of people." and "I like the main character in the story a lot." Identification was assessed only in the first two story versions, as they feature human protagonists. For all three story versions, *self-referencing* was also evaluated using Chen et al.'s framework [31].

Example statements include: "Do you think the story relates to you, personally?" and "Did the story make you think about your blood pressure?"

PERCEIVED CREDIBILITY. The credibility of the information presented in the story was assessed using the Meyer modification of the Gaziano-McGrath scales [153]. These scales evaluate credibility based on five key aspects: *accuracy*, *fairness*, *trustworthiness*, *bias*, and *completeness*. Additionally, in the first two story versions, which feature human protagonists, the *perceived credibility* of the protagonist was examined.

EMOTIONAL RESPONSE AND EMOTIONAL FLOW. The *emotional response* was evaluated by asking participants to rate their feelings after viewing the story on a five-point scale, ranging from extremely negative to extremely positive. To assess *emotional flow*, participants were asked whether they typically engage in activities depicted in the story as either healthy or harmful (e.g., exercising or smoking), a measure referred to as *behavioral intent*. A second set of questions examined whether they intended to engage in these activities in the near future. This approach follows the methodology outlined by Alam et al. [4], which evaluates whether the presented information encourages individuals to modify aspects of their lifestyle.

PERCEIVED TIME SPENT IN THE STORY. Participants were asked whether they *liked to know more* about the disease or the protagonists, providing insight into their willingness to engage further with the story. Additionally, perceptions of story length were evaluated by asking whether the disease story was *perceived as too long*. The *actual time* spent viewing the story was also recorded.

8.4.3 *Discussion of Results*

The results were aggregated by computing scores for each evaluation criterion, following an approach similar to that used in [117], as shown in Figure 48. For each story version, the average of the responses to the relevant questions was calculated per criterion. A score of 1 indicates a low rating, reflecting disagreement. In turn, a score of 5 represents agreement. A different method was applied to assess emotional flow, as this criterion is intended to reflect shifts in motivation toward healthier lifestyle choices (see Figure 49). For each lifestyle aspect, the difference between behavioral status and behavioral intent was calculated, and the average was computed per story version. To enhance the clarity of the visualization, the scales for smoking and alcohol – being the only negative lifestyle behaviors – were inverted. This adjustment allows the plot to display positive values for increased health-conscious intent and negative values for a shift toward less healthy behavior.

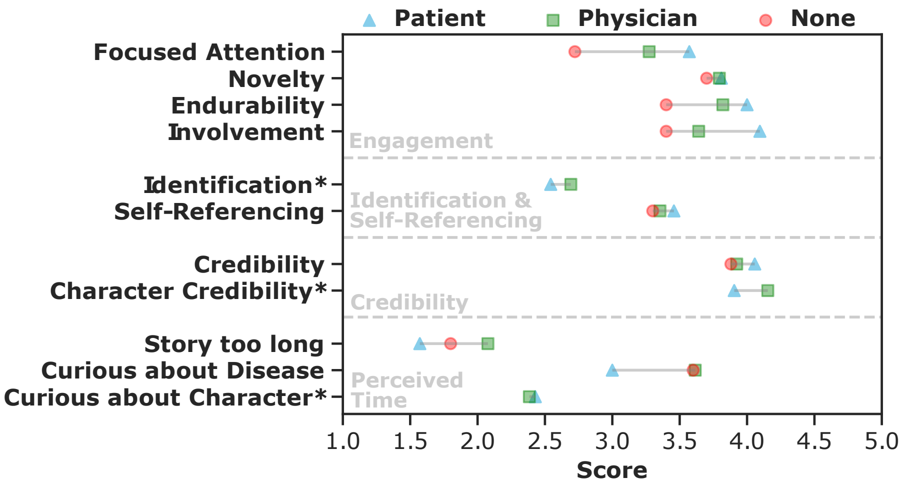


Figure 48: The scores were calculated by averaging all responses for each criterion (indicated by dashed lines) and story version. A score of 1 indicates strong disagreement, while a score of 5 indicates strong agreement. Labels marked with an asterisk apply exclusively to stories featuring human protagonists. Image from Mittenentzwei et al. [156], available under a CC BY 4.0 license. No changes were made.

ENGAGEMENT. The different story versions showed varying performance across the individual attributes of engagement, as illustrated in the upper section of Figure 48. Versions featuring a human protagonist received notably higher ratings for *focused attention*, *endurability*, and *involvement*. However, all three versions achieved comparable scores for *novelty*.

One possible explanation for this outcome is that participants were largely unfamiliar with CSVD, leading the version without a human protagonist to generate a similar level of curiosity. As a result, no definitive conclusion can be drawn regarding overall engagement across the story versions.

8

EMOTIONAL RESPONSE AND EMOTIONAL FLOW. The average *emotional response* reported by participants varied across story versions: it was rated as *neutral* for the physician-centered narrative and *positive* for both the patient-centered and baseline versions.

The observed *emotional flow* differed depending on the lifestyle factor, as illustrated in Figure 49. Across all story versions, a positive shift was recorded in participants' intentions toward healthier behaviors, including improved diet, regular physical activity, reduced alcohol consumption, and adequate sleep. However, a negative shift was observed for smoking behavior. No clear explanation could be identified for this result. It is possible that the wording in the questionnaire allowed for varied interpretations,

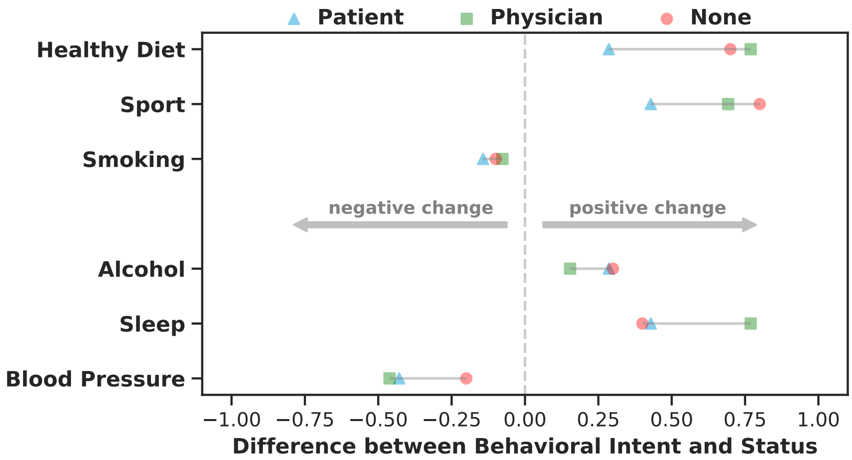


Figure 49: The average difference between participants' current behavior and their intended future behavior for each story version. For clarity, the results for smoking and alcohol consumption were inverted in the plot. A positive value indicates an intention to adopt a healthier lifestyle, while a negative value reflects an intention to engage in harmful lifestyle behaviors in the near future. Image from Mittenentzwei et al. [156], available under a CC BY 4.0 license. No changes were made..

for instance, phrases such as "on a regular basis" or "in the near future" may have been perceived differently among respondents. To better contextualize such responses, future studies may benefit from offering participants the opportunity to elaborate on their choices.

In the case of blood pressure, participants were asked whether they knew their current values and whether they intended to measure them in the near future. It is assumed that younger individuals (who constituted the majority of the sample) may perceive themselves as low-risk and, therefore, see little need to monitor their blood pressure. No consistent influence of human protagonists on emotional flow was identified. This may be related to prior findings suggesting that individuals often underestimate their personal health risks, resulting in limited motivation to adopt lifestyle changes [208, 246].

The lifestyle recommendations presented in the story are widely recognized. As a result, participants who already follow healthy habits are likely to have adopted these behaviors prior to engaging with the narrative. Given that the majority of participants were notably younger than the typical onset age for CSVD, the story may have been more effective in prompting discussions with family members, particularly parents, about disease risks, rather than encouraging personal lifestyle changes.

Additionally, it is possible that the hospital was perceived as the central figure in the base version of the story. This interpretation would position the base and physician-centered narratives as closely related in terms of credibility, with both presenting information through authoritative sources. The primary distinction lies in the explicit inclusion of a human character, the physician, in the latter. While the patient is embodied through data representations (e.g., icons), both the hospital and physician primarily serve as communicators of the information. The physician was primarily presented through textual descriptions rather than visual elements such as icons. This approach may have contributed to the physician-centered narrative being perceived as less character-driven compared to the patient-centered version.

IDENTIFICATION AND SELF-REFERENCING. Overall identification with the protagonists was low, as shown in Figure 48. One possible explanation is the limited amount of character information provided, which may have led to perceptions of the protagonists as underdeveloped or one-dimensional. Since character-driven narratives often rely on internal conflict and character growth, more detailed and nuanced portrayals could enhance relatability. The physician character received a slightly higher identification score. No notable increase in identification was observed among participants who shared demographic characteristics (such as age or gender) with the protagonists.

A slightly higher level of *self-referencing* was observed in the patient-centered version, as shown in Figure 48. This effect may stem from the perception that the patient protagonist represents a relatable figure, creating a sense of *"this could be me"*, as the narrative presents the disease progression through the lens of a single, identifiable character. This format may align more closely with participants' personal experiences with illness. In contrast, the physician-centered version references multiple *"other"* patients, potentially reinforcing a sense of distance and reducing personal relevance—leading to the perception that *"this applies to others, not to me"*.

CREDIBILITY. All story versions received similarly high credibility ratings, with the patient-centered version showing a marginally higher score, as illustrated in Figure 48 (cf. *Credibility*). This outcome may have been influenced by the institutional references included in the narrative and credits. Given that participants were aware that the story was developed by a university, its neurology clinic, and the German Center for Neurodegenerative Diseases, this affiliation may have contributed to the overall perception of credibility. Participants were also asked to evaluate the credibility of the human characters separately. In this case, the physician received a slightly higher score than the patient, as shown in Figure 48 (cf. *Character Credibility*).

TIME. Participants generally did not perceive the stories as too long, as shown in Figure 48. The patient-centered version received the lowest score in terms of participants' desire to learn more about the disease (cf. *Curious about Disease* in Figure 48). This may be due to the narrative being perceived as complete, with the patient's journey providing a full arc from diagnosis to outcome. In contrast, the physician's narrative could be seen as having the potential for additional content, such as further patient cases or professional insights. Interest in learning more about the human characters was also low across versions (cf. *Curious about Character* in Figure 48). It is unclear whether this response reflects a sense of narrative completeness or a lack of engagement with the characters themselves. Given that character introductions were brief and minimally developed, the latter seems more likely.

Participant interaction time with the stories was also recorded. On average, less time was spent with the base version (3:48 minutes; range [1:29, 6:42]) compared to the patient-centered (4:47 minutes; range [3:05, 8:43]) and physician-centered (4:48 minutes; range [0:41, 11:55]) versions. Extremely short durations may result from participants not fully engaging with the content or from technical issues, such as inadvertently closing the story and revisiting it to obtain a participant ID for survey completion. To ensure data reliability, response patterns were reviewed in the questionnaires to identify any instances where a single column was selected throughout, which might indicate careless answering. One participant provided an incorrect ID, preventing the accurate association of their time-tracking data. While the base story was slightly shorter due to the absence of character-specific elements, this difference alone does not account for the variation in engagement time. It is therefore likely that the absence of a human protagonist reduced participants' willingness to remain engaged with the content.

8.5 STUDY LIMITATIONS

This study has several limitations, particularly in terms of representation. The evaluation was based on different versions of a single disease story, which limits generalizability. Disease stories about other conditions, such as cancer, could potentially affect user perception in different ways.

The story used in this study contains substantial textual content to convey critical aspects not easily communicated through visual elements alone. As a result, it remains unclear to what extent participants engaged with or relied on the visualizations. Additionally, the investigation focused on a limited set of protagonist types, with only one character featured per version. Other figures, such as family members or friends, may also play important roles in disease narratives but were not included in this study.

Data were collected from 30 participants in total, with each individual viewing only one version of the story. Consequently, the sample size per version is small and not statistically representative. Conducting the study online further limited control over sample distribution, resulting in uneven participation across conditions: seven participants for the patient-centered version, ten for the physician-centered version, and thirteen for the baseline version. Moreover, the participant group displayed a relatively high level of literacy, likely influenced by the recruitment platforms used for data collection.

Participants were recruited through the crowdsourcing platforms SurveySwap [15] and SurveyCircle [101], where users exchange survey participation. A total of twelve participants were acquired via these platforms; however, it is not possible to determine how many of them completed the survey in full. The remaining participants were recruited through direct personal outreach or university mailing lists, indicating a level of intrinsic motivation to participate. As a result, the findings may not reflect the perceptions of users who encounter the story versions incidentally, such as while browsing online without prior intent to engage. Since this recruitment bias applies consistently across all story versions, the comparative analysis of results remains valid.

8.6 DISCUSSION OF RESEARCH QUESTIONS

RQ3.1 DO HUMAN PROTAGONISTS INCREASE ENGAGEMENT IN DISEASE STORIES?

There is a noticeable trend indicating that stories with human protagonists, especially the patient-centered version, tend to receive higher evaluation scores regarding the dimension of engagement in terms of *focused attention*, *endurability*, and *involvement*. In contrast, the scores for *novelty* remained consistent across all three versions.

RQ3.2 DO HUMAN PROTAGONISTS INCREASE THE USERS' MOTIVATION TO IMPROVE THEIR PERSONAL LIFESTYLE?

All story versions encouraged healthier behavior, except for smoking. Human protagonists had little effect on emotional flow. These results might be influenced by study limitations such as low perceived personal risk among the large amount of young participants. In conclusion, no effect of human protagonists on participants' lifestyle improvements could be observed.

RQ3.3 DOES A PATIENT AS THE PROTAGONIST INCREASE IDENTIFICATION AND, IN TURN, ENHANCE SELF-REFERENCING?

No clear correlation can be established between a high level of identification and an increased degree of self-referencing. A slight tendency was observed in which self-referencing was most pronounced in the patient-centered story, despite the highest identification scores being associated with the physician-centered version.

RQ3.4 DOES A PHYSICIAN AS PROTAGONIST INCREASE THE PERCEIVED CREDIBILITY OF THE PRESENTED INFORMATION?

All story versions were perceived as credible. This might be influenced by the credits shown in the end of all story versions. Although it cannot be confirmed that featuring a physician as the main character enhances the overall perceived credibility of the story, the physician character was rated as slightly more credible compared to the other protagonists.

RQ3.5 do human protagonists increase the time users spent in the disease story?

No clear difference was found in the perceived duration of story interaction between versions with and without a human main character. However, the patient-centered story was least likely to be rated as too long, suggesting that the presence of specific protagonists may influence users' perception of time spent engaging with the content. Additionally, recorded interaction times showed that participants spent more time with human protagonist-driven stories, providing support for this assumption.

RQ3 what is the impact of a human protagonist on health communication?

RQ3.1 through RQ3.5 explored the impact of different human protagonists on various aspects of the disease story, forming the basis for the discussion of RQ3.

Stories featuring human protagonists demonstrated higher engagement across multiple dimensions, indicating that the inclusion of human protagonists had a positive effect on user engagement. In particular, the story with a patient as the protagonist led to increased levels of self-referencing, while the story featuring a physician as the protagonist resulted in the highest levels of identification. The physician was also rated as the most credible character, although general credibility ratings across all story versions were similarly high. Human protagonists did not influence participants' motivation to adopt healthier lifestyles or their perception of the story's duration. However, participants tended to spend more time with stories that featured human protagonists.

In summary, human protagonists primarily contributed to increased engagement and longer time spent with the story, indicating that their presence made the content more compelling. Self-referencing and identification are only applicable when human characters are included and are intended to enhance the disease story's effect. However, since motivation to pursue healthier behavior did not increase in versions with human protagonists, this intended effect was not achieved. Due to several limitations in both the story design and the study design, as outlined in Section 8.5, the generalizability of these findings remains limited.

8.7 conclusion

Two variations of a character-driven structure were introduced to support the design of disease stories. They provide a framework for transforming disease-related content, such as patient vignettes or blog entries, into cohesive narratives by outlining a suggested sequence of events and the integration of protagonists. The comparison of stories with different protagonists aimed to explore potential effects on user experience. While the limited sample size prevents definitive conclusions, observable differences among the versions suggest that the choice of protagonist plays a noticeable role in shaping audience perception. Based on these findings, several insights have been derived to guide the incorporation of characters in future disease narratives.

8.7.1 *Study Implications*

Overall, no definitive conclusions could be drawn. However, noticeable differences emerged across certain evaluation criteria depending on the story version. In addition, the study offers new insights into the design and assessment of character-driven disease stories. Positive effects of human protagonists were observed in relation to several aspects of user engagement, and these effects may be further amplified by the inclusion of more fully developed characters. Based on these findings, the inclusion of human protagonists is recommended in the design of future disease stories.

No notable differences in identification were observed across the different story versions. However, the process provided valuable insights into the development of human protagonists in disease stories. The design prioritized integrating the protagonist into every view to ensure they remained central to the storytelling. To maintain brevity, character introductions were kept minimal. Therefore, the characters may have appeared underdeveloped, limiting opportunities for emotional connection. Given that the effectiveness of character-driven disease stories relies heavily on the depth of their protagonists, greater attention to character development is recommended.

Credibility was consistently rated highly across all story versions, regardless of whether a human protagonist was included. This suggests that the association with reputable and trusted institutions can create an equal or greater impact on perceived credibility than the inclusion of a domain expert within the disease story.

8.7.2 *Future Work*

Further research is required to determine whether specific types of protagonists are particularly effective in supporting distinct goals within disease stories, such as conveying messages related to prevention, screening, or coping strategies. Expanding the range of diseases and characters in future studies would also provide a more comprehensive understanding of these effects.

Future research could explore the impact of including multiple protagonists within a disease story, such as both a patient and a physician. In the context of identification, it would be valuable to examine how the use of realistic imagery, such as photographs, compares to stylized or comic-like character representations similar to examples shown in Chapter 7. Disease stories may also help patients prepare for medical consultations through overviews of diagnostic imaging, surgical procedures, or treatment options. As a result, further investigation is needed to determine how human protagonists should be tailored to effectively meet the needs of patients.

9 THE NARRATIVE STAGE: STRUCTURE

Contrary to Campbell's Hero's Journey, many story structures are character-independent and focus on the tension levels throughout a story. This chapter focuses on the last story ingredient, namely structure, see Figure 50.

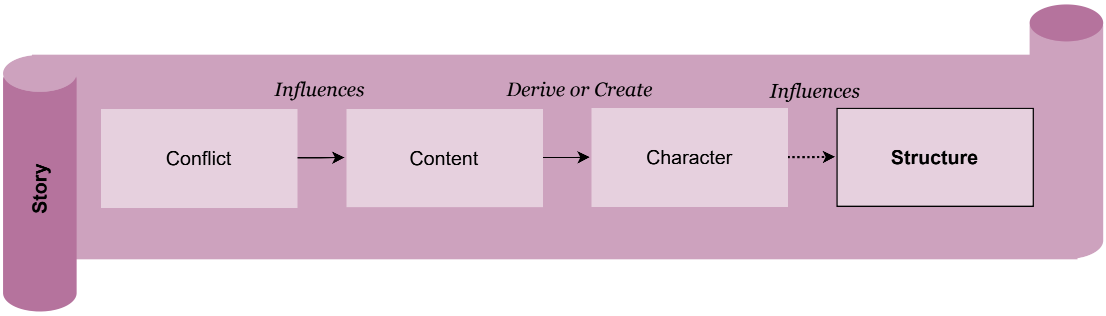


Figure 50: The narrative stage of the disease story design process (cf. Figure 9) with *structure* highlighted.

For conducting a case study, two disease stories are created, the first being another adaption of the CSVD story, the other being a story about BAV. Freytag's Pyramid is chosen as a tension-driven structure for both application examples. Additionally, the BAV story follows the martini glass structure. While story structures can be seen as a blueprint for a story, the narrative genre decides how this blueprint is implemented, influencing the final story extensively. Therefore, this chapter focuses heavily on the comparison of two disease stories designed according to Freytag's Pyramid and implemented in two different genres each, namely interactive slideshow and scrollytelling. To address this, the following research questions are explored:

RQ4 How do the different genres scrollytelling and interactive slideshow shape the perception of disease stories?

- *RQ4.1* Do users prefer more or less interactive content and how much time to they spend engaging with interactive story content?
- *RQ4.2* Is vertical navigation, via scrolling, or horizontal navigation, using clicking, preferred by users?
- *RQ4.3* How do different storytelling formats and interaction mechanisms affect users' ability to follow the narrative?

9.1 A CASE STUDY FOR COMPARING USER BEHAVIOR IN NARRATIVE GENRES

Narrative structure encompasses a range of approaches, as introduced in Section 2.2. While the previous chapter presented a character-driven structure, this chapter explores a tension-driven structure (Freytag's Pyramid) and interaction-driven structures (the martini glass structure and an elastic structure). In both stories, the stories differ in the interaction allowed within individual slides. The BAV story mainly follows the martini glass structure, concentrating interactive elements at the end, whereas the CSVD story distributes multiple interaction opportunities and optional paths throughout the story. This variation allows for a closer examination of how interaction opportunities and interaction-driven structures influence user experience.

Collaboration and Personal Contribution

This chapter is partially based on the following publication: [157]. The work builds on previous research by Kleinau et al. [113], which explored disease stories by analyzing design decisions in the context of a story about BAV. The BAV story was designed and implemented by Anna Kleinau and Evgenia Stupak. The CSVD story was conceptualized and implemented by the author of this thesis. The evaluation was designed in collaboration with all co-authors of the aforementioned publications and conducted by the author. The submitted work has been prepared with all co-authors who provided feedback throughout the project and writing phase.

A comparative between-subject user study was carried out to assess the relative effectiveness of the two narrative genres with respect to *interaction*, *navigation*, and *transitions* within data-driven disease stories, serving as a proof-of-concept. This extension also incorporates data logging to analyze which sections of the stories users spent most time with, offering insights into user behavior when viewing a disease story.

9.2 RELATED WORK

This work is based on foundations of narrative visualization described in Section 2.2, specifically narrative genres and narrative structure.

The Slideshow genre is used for one of the two presentation methods for evaluation. In this genre, slides may be fully static, or contain animations. The sequencing and transition techniques for slides have been investigated in detail by Hullman et al. [88] and served as inspiration in the presentation design. The second genre that is investigated in this study, known as Scrollytelling, is a long-form article rich in images and multimedia that is common in online journalism. It is broadly thought to have been introduced in 2012 by the New York Times [239] and examined in detail by Seyser and Zeiller [211]. Recent work provides a more general overview of Scrollytelling techniques [36, 144]. The Slideshow and Scrollytelling use fundamentally different navigational techniques. While in a Slideshow, users navigate between slides through discrete clicks or taps, Scrollytelling enables a continuous navigation experience by scrolling via either a mouse or trackpad.

The choice of narrative genre exerts some influence over the interactivity of the data-driven visual story, which ranges from author- to reader-driven [210]. A fully author-driven story, often intended for live presentations, consists of static visualizations in a linear story path with no interactivity. In contrast, fully reader-driven stories have no strict story path or ordering and exhibit extensive interactivity. It is debatable whether reader-driven stories belong to narrative visualization since the author cannot ensure that a message has been conveyed when the user has considerable freedom to explore.

Learning Objectives

As learning objectives for CSVD were already formulated in Chapter 8, the focus lies on BAV. The motivation for communicating BAV – to support understanding of the condition and informed consent for treatment – can be framed as learning objectives. Cognitive objectives emphasize recognition and recall of information, following Adar et al. [1]:

- The viewer will *recognize* consequences of BAV.
- The viewer will *recall* information about treatment for BAV.

Affective objectives focus on observing and connecting information to their personal beliefs, as defined by Lee-Robbins et al. [132]:

- The viewer will *consider* the consequences of BAV.
- The viewer will *evaluate* the information about treatment for BAV.

These objectives guide the design of disease stories but are not the focus of evaluation in the following studies; rather, they provide a structured rationale for this research.

9.3 DISEASE STORY DESIGNS

The BAV story utilizes 4D PC-MRI data collected before and after treatment for bicuspid aortic valve. Visualizations of the aorta and internal blood flow were created using the *Bloodline* software [116]. In contrast, the CSVD story is based on brain MRI data, including segmentation masks of lesions, along with tabular data representing patient risk factors, sourced from the University Hospital Magdeburg. Both narratives are enhanced with supporting visual materials such as photographs, illustrations, and icons. A comprehensive overview of the visual components and the corresponding design process for the BAV story is provided in the work by Kleinau et al. [113]. The design process for the CSVD story is presented in Section 4.2.

9.3.1 *Comparison of Story Boards*

Both stories are structured according to Freytag's Pyramid, as illustrated in Figure 51. The CSVD story includes multiple optional branching paths at the beginning and end, whereas the BAV story more closely aligns with the martini glass structure, with most of its optional branches appearing toward the end. These branches provide supplementary or in-depth information on specific topics, such as MRI functionality or additional medical context. By incorporating optional pathways, the stories allow users to control the level of detail they wish to engage with, enhancing accessibility and making the content adaptable to a broader audience.

Both stories include a brief explanation of MRI and feature a fictional patient character. In the BAV story, the protagonist is represented by a stylized cartoon figure designed to be age- and gender-neutral. In contrast, the CSVD story employs a more realistic approach, using photographs. Additionally, the CSVD narrative introduces a supporting character in the role of the treating physician.

9.3.2 *Genre Implementation*

The stories described above were developed using two narrative formats: *Slideshow* and *Scrollytelling*. The purpose is to explore which navigation approach is more intuitive and accessible for a broad audience, whether the *transitions* between segments are perceived as smooth, and how enjoyable the *level of interaction* is. In addition, the impact of the *navigation* style on user behavior is examined, e.g., whether a particular format leads users to spend more time engaging with the story content.

9

Supplementary Material

Supplementary material for this chapter can be found at:
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/structure`

SLIDESHOW-BASED NAVIGATION. Each story was initially developed using a traditional Slideshow format, in which users navigate by clicking through individual slides. The BAV Slideshow was created using Microsoft PowerPoint [43], while the

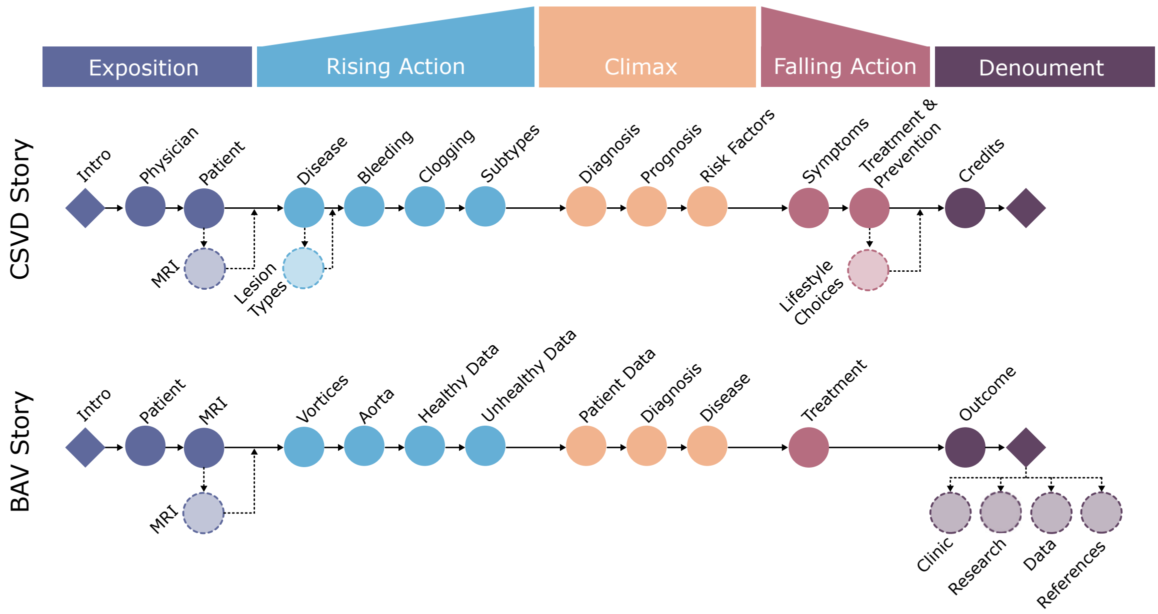


Figure 51: Structure of the CSVD and BAV stories, visualized within the framework of Freytag's Pyramid. Diamond-shaped nodes indicate the starting and ending scenes, while circular nodes represent thematic sections (acts) of each story. Optional paths are shown as downward branches and dashed lines. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

CSVD Slideshow was implemented with the Unity game engine [235]. The use of buttons designed as left/right arrows for going forward or backward in the stories gave the impression of horizontal navigation.

Each slide includes a *navigation bar* and a *progress bar* to help users track their current position within the story and understand how much content remains, supporting orientation and engagement. Additional information can be accessed through *underlined text comments* in the BAV narrative and *animated icons* in the CSVD version.

In the BAV story, blood flow is visualized using animations originally created with *Bloodline* [116], which were converted into GIFs and embedded within the story. This format ensures automatic playback, minimizing the risk of users missing key visual content. The animated motion also serves to naturally draw attention to the depiction of blood flow. In contrast, the CSVD story features static MRI imagery. *Animated icons* are used to highlight interactive elements within the visualizations, such as options for rotating or clipping 3D models.

One challenge in PowerPoint arises from its default behavior of advancing to the next slide when users click on non-interactive elements, increasing the likelihood of unintentional navigation. Additionally, achieving smooth transitions between content elements proved difficult due to the platform's standard slideshow appearance. This limitation was addressed through precise slide design. For more fluid visual progression, slides were kept nearly identical except for minimal adjustments. When combined with subtle transitions or morphing effects, this approach created the illusion of continuous transformation within a single slide. In contrast, Unity is not inherently designed for slideshow creation. As a result, developing a slideshow format within this environment requires additional time and effort. However, this approach provides greater flexibility and a wider range of options for integrating diverse data types and interactive features.

In conclusion, PowerPoint presents certain constraints in story design when compared to Unity. PowerPoint's interactive capabilities are primarily limited to predefined slide transitions, whereas Unity is specifically built for interactive media, offering more flexibility for integrating interactive elements within slides. However, developing a slideshow in Unity requires at least a basic understanding of programming. Additionally, the complexity of the Unity environment may pose challenges for users without prior experience in game engines, potentially making it unnecessarily complex for simple narrative presentations. On the other hand, PowerPoint allows for faster and more accessible slideshow creation, particularly for users without coding or game development backgrounds. Despite this, leveraging PowerPoint to its full potential for narrative storytelling still involves a considerable degree of complexity and effort.

SCROLLYTELLING-BASED NAVIGATION. The *ScrollyVis editor* [165] was used to adapt both narratives into the *Scrollytelling* format. In this genre, scrolling serves as the primary mode of interaction, with additional click-based elements enabling users to explore optional branches within the story.

Although the core content remains consistent across both formats, a key difference lies in how users engage with branching paths. In the scrollytelling version, users must make an intentional choice to explore optional segments. By contrast, in the slideshow format, these branches are embedded directly into scenes as highlighted interactive comments or animated icons, making them more seamlessly integrated into the flow of the narrative.

The ScrollyVis implementation features a dark-themed visual design, as shown in Figure 52 (right). This design choice follows the default style provided by the editor, which had previously undergone evaluation in a user study[165].

A prominent element of the ScrollyVis editor is the side navigation tree, which allows users to track their progress through the story—functionally similar to the progress bar used in the slideshows. The structure of text presentation also differs between the slideshow and scrollytelling formats. In PowerPoint and Unity, text elements can be introduced sequentially through animations, enabling multiple text segments to appear simultaneously. In contrast, ScrollyVis uses scrolling as the mechanism for building text, typically limiting the display to a single text block at any given time.

In the case of ScrollyVis, interaction with embedded elements, such as rotating a 3D object, is controlled exclusively through scrolling. The only exception to this is video

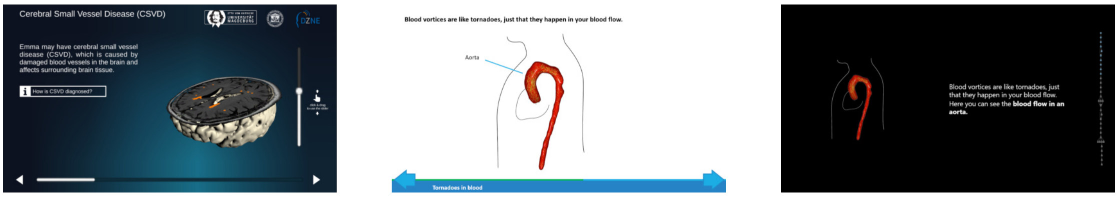


Figure 52: Screenshots of the stories implementations using different tools where left shows the CSVD *Slideshow* in Unity, middle the BAV *Slideshow* in PowerPoint, and right the Scrollytelling version of the BAV story in ScrollyVis. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

content, which includes standard play and pause controls. Since the ScrollyVis editor does not support the integration of GIFs, videos were used instead to replicate animations previously inserted as GIFs in the PowerPoint version and to represent advanced interactive features, such as 3D model clipping, found in the Unity implementation. Videos in ScrollyVis begin playing automatically upon appearing in the viewport and can be paused either by scrolling past them or by using the pause button. While this approach provides a higher level of interactivity compared to static GIFs, it remains less interactive than the functionalities offered by the Unity version.

9.4 EVALUATION

A user study was carried out to examine the impact of narrative genres and topics on interaction, navigation, and transitions within the context of the specified use cases.

9.4.1 *Participants*

The study included 85 participants (51 male, 31 female, 3 nonbinary) ranging in age from 8 to 61 years, with a mean age of 27.32. The majority of participants held a high school diploma, bachelor's degree, or master's degree. Recruitment took place primarily during the *Long Night of Science 2022*, a public outreach event organized by the university. Additional participants were invited through announcements targeting university students. Individuals enrolled in medical programs were excluded from participation.

A between-subjects design was used, with each participant engaging with one of two story formats. Participants were divided into four groups: 22 viewed the slideshow version and 23 the scrollytelling version of the BAV story, while both versions of the CSVD story were viewed by 20 participants each.

Due to issues in matching time logs with questionnaire responses, such as mismatching participant IDs, data from 10 participants could not be used. These unmatched logs were excluded from the analysis. One outlier was excluded from the Scrollytelling ver-

9

sion of the BAV story. This participant exhibited more than twice the number of slide transitions compared to all others, suggesting atypical interaction behavior.

9.4.2 *Study Design*

All story implementations were displayed on Wortmann Terra All-in-One-PC 2211 systems, each equipped with an Intel Core i3 processor (3.4 GHz), 8GB of RAM, and Intel HD Graphics 2500. Both devices featured 21.5-inch Full HD (1920 x 1080) touchscreens.

Participants interacted with the stories exclusively through touch input during the study. Upon completing the story experience, a qualitative questionnaire was administered.

The questionnaire included items on participant demographics, user experience with different navigation methods (clicking and scrolling), and knowledge checks to assess comprehension of the story's content. Results from the knowledge checks were consistent with the findings of Kleinau et al. [113], where the majority of participants answered most questions correctly. However, these checks did not provide insight into the effectiveness of the stories in terms of interaction, navigation, or transitions. As such, these results are not discussed further in this chapter.

The time spent by each participant on individual story nodes was recorded. As the questionnaire was available separately from the story, users were presented with an ID at the end of the story which had to be entered in the questionnaire to enable a comparison of time tracking and questionnaire results. In the *Slideshow* format, a story node corresponds to a single slide, while in the *Scrollytelling* version, it refers to an individual element that appears or disappears through scrolling actions, such as a block of text or an image. To enhance comparability across formats, multiple *Scrollytelling* nodes were aggregated into broader thematic units. For example, in the CSVD story, all nodes related to the explanation of MRI, including both textual and video elements, were combined. In contrast, this content is contained within a single slide in the *Slideshow* version. Time spent on the first and last story nodes was excluded from analysis, as these segments were often idle, either while waiting for the next participant or during the completion of the post-story questionnaire. Additionally, any log entries recorded after the final visit to the end node were removed. Since participants were required to reach the end node to receive the ID necessary for submitting the questionnaire, any subsequent interactions likely resulted from revisiting earlier parts of the story to provide qualitative feedback or clarify details while completing the survey.

Navigation-related issues were identified through observational data collected using the *think aloud* method [64].

9.4.3 *Discussion of Results*

This section presents the findings of the user study, focusing on interaction, navigation, and transitions, as observed within the scope of the defined use cases.

INTERACTION. Participants generally disagreed with the notion that they would prefer reduced interaction within the story (Figure 53, statement 2). This response appears to be largely unaffected by either the story topic or narrative genre. While a slightly higher number of participants in the CSVD scrollytelling group preferred less interaction, no notable differences were observed across stories utilizing different types of media, such as videos in scrollytelling, GIFs in the BAV slideshow, or 3D interactions in the CSVD slideshow (see Figure 53, statement 2). These findings suggest that the number of interaction options may be less critical than how seamlessly they are integrated into the overall narrative flow. Overall, interaction through both scrolling and clicking was perceived as easy, with clicking receiving slightly higher agreement (Figure 53, statement 3). Notably, the CSVD slideshow received the most favorable ratings in this regard. Interaction with the CSVD story was considered intuitive, with the slideshow version receiving slightly stronger agreement on this point (Figure 53, statement 4).

Participants also noted that interaction with the BAV Slideshow occasionally felt unresponsive, with button presses not consistently triggering immediate actions. Participants required additional guidance when interacting with the story through scrolling. Initial confusion was reported regarding the need to scroll to navigate the content, with some describing the interaction method as unintuitive.

Uncertainty also arose when videos were presented, as the automatic replay of the videos after completion led to hesitation about whether to continue scrolling. One participant suggested that accompanying text should automatically appear alongside video content to improve clarity.

Participants spent a notable amount of time on the initial story nodes, particularly those containing interactive or animated elements, such as the blood flow animation in the BAV story and the "MRI" and "Disease" sections in the CSVD story (see Figure 54). However, later story nodes containing interactive or animated content did not retain user attention to the same extent. This may be due to participants needing additional time early on to become familiar with the video interface, or to an increased level of curiosity when first encountering animated elements in the story.

NAVIGATION. Figure 54 and Figure 55 illustrate how participants navigated through the different story versions. The average time spent was as follows:

- 5 minutes 34 seconds in the CSVD Slideshow (min: 3:40, max: 8:50, σ: 1:23)
- 4 minutes 18 seconds in the BAV Slideshow (min: 2:38, max: 6:41, σ: 1:11)

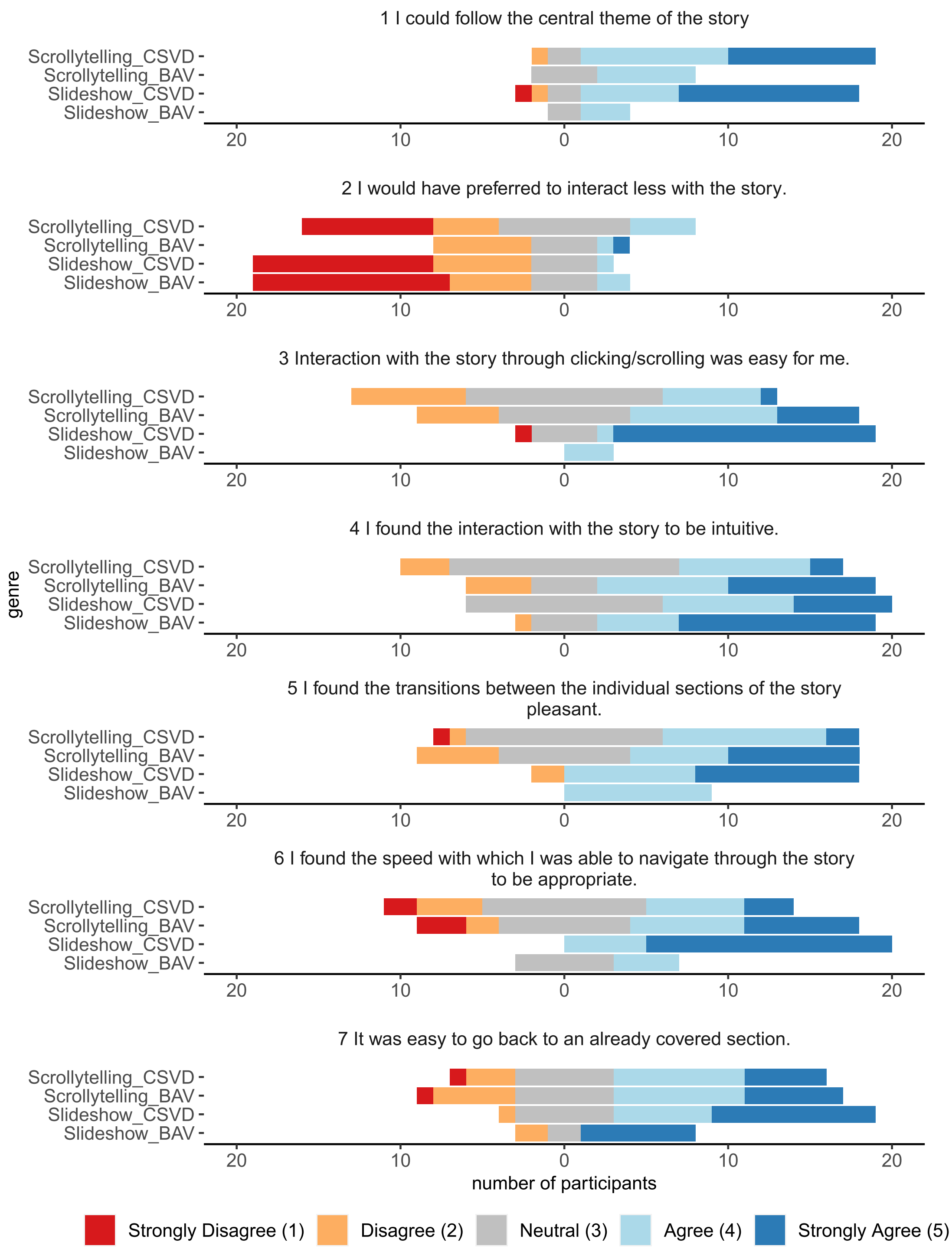


Figure 53: Participant responses for the *Slideshow* and *Scrollytelling* versions of the BAV and CSVD stories. Responses were quantified using a 5-point Likert scale to compute the mean (μ) and standard deviation (σ) for each question and story version. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

- 6 minutes 13 seconds in the CSVD Scrollytelling (min: 2:37, max: 9:54, σ: 1:50)
- and 7 minutes 5 seconds in the BAV Scrollytelling (min: 2:46, max: 12:23, σ: 2:21)

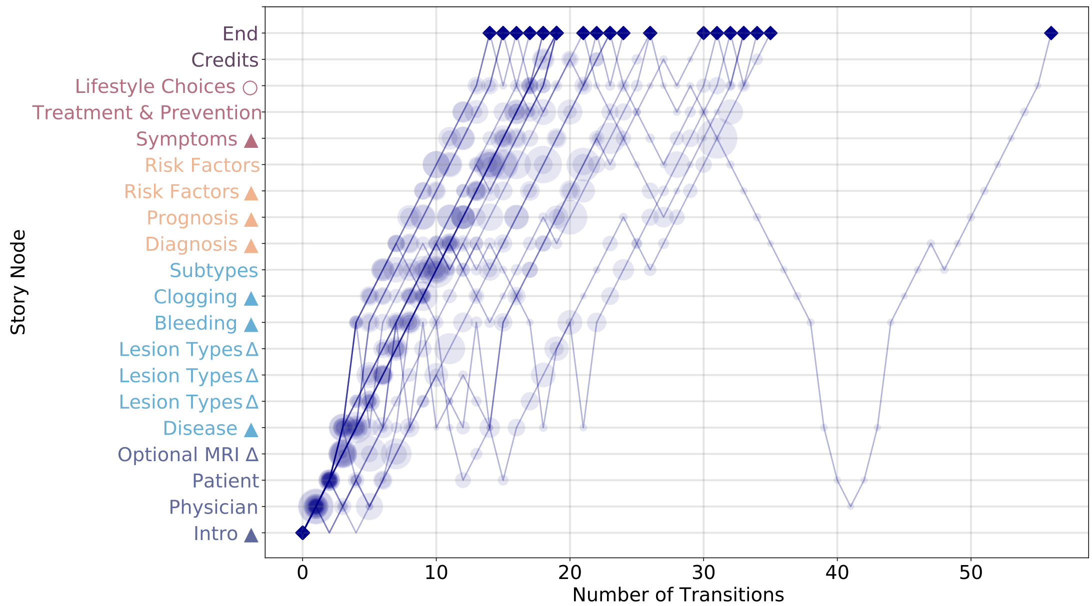
No notable differences in navigation behavior were observed across subgroups based on gender, age, or education level. The CSVD story features a greater number of optional branches near the beginning, which many participants explored. This is reflected in Figure 54, where the timelines show a wider spread along the x-axis. In contrast, the BAV story follows a more linear structure in line with the martini glass structure, with the majority of participants progressing through the core content before reaching optional paths toward the end, as shown in Figure 55.

Only few participants accessed the optional branches offered at the conclusion of the BAV story, while optional content presented earlier in the narrative was more frequently viewed. A comparison of formats revealed that fewer participants explored these additional sections in the BAV Slideshow than in the BAV Scrollytelling version.

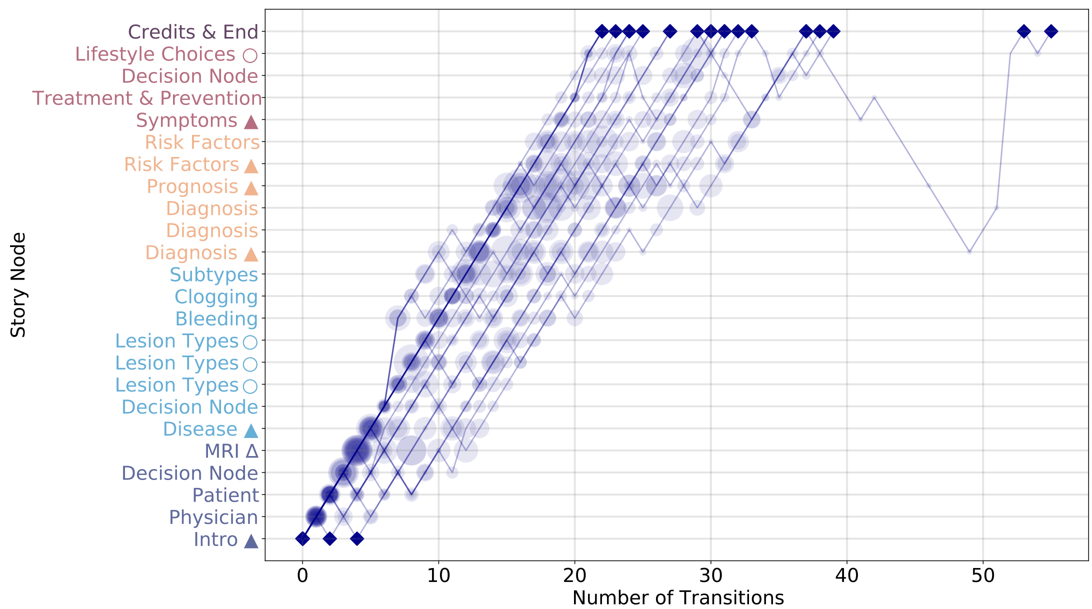
Participants who engaged with the CSVD slideshow and the BAV scrollytelling versions were more likely to revisit earlier content, indicated by the presence of more jagged navigation paths in Figure 54 (a) and Figure 55 (b). However, backtracking multiple earlier nodes was relatively uncommon. On average, the time spent per story node remained consistent across all versions. Returning to earlier sections was rated less favorably in the scrollytelling versions compared to the slideshows, as shown in Figure 53 (statement 7). This may be linked to the slower scrolling speed, which could make backward navigation feel cumbersome across multiple story segments.

Participant feedback on the slideshow primarily focused on navigation. While one participant described the transitions between slides as *visually appealing*, another found them *too slow*. Several comments highlighted that navigation from left to right aligned well with the natural reading direction. Additionally, some participants expressed a desire for more flexible navigation options, such as an interactive menu bar that would allow direct access to non-adjacent slides or specific sections of the story. Several participants reported that the scrolling speed felt too slow. One individual did not realize that upward scrolling was possible. Additionally, some participants limited their scrolling to the right side of the screen, despite scrolling being enabled across the entire interface. The effect of scrolling speed remains an unresolved question that falls outside the scope of this study. Moreover, evaluating scrolling solely as less intuitive does not consider its widespread success and familiarity in contexts such as digital articles and social media platforms [277].

At branching points in the story, some participants did not realize that scrolling was required after clicking on story nodes (see Figure 56 (b)), and instead expected an immediate response or transition. The graph illustrating the story structure was positively received, with several participants expressing a desire for it to support direct navigation through the story. This could potentially be addressed by immediately displaying the subsequent content upon selection.



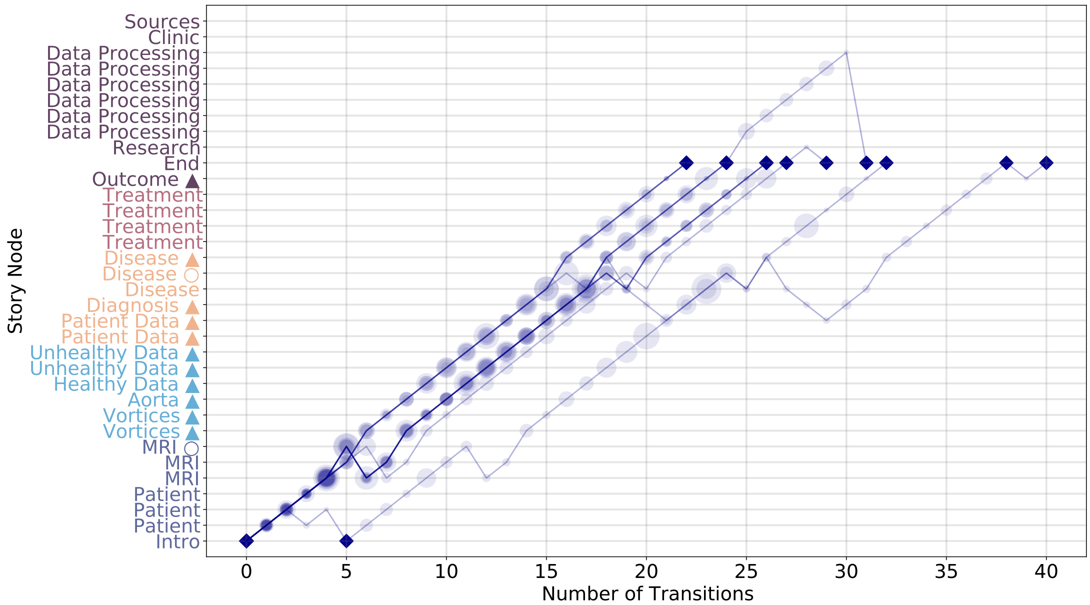
(a) CSVD Slideshow



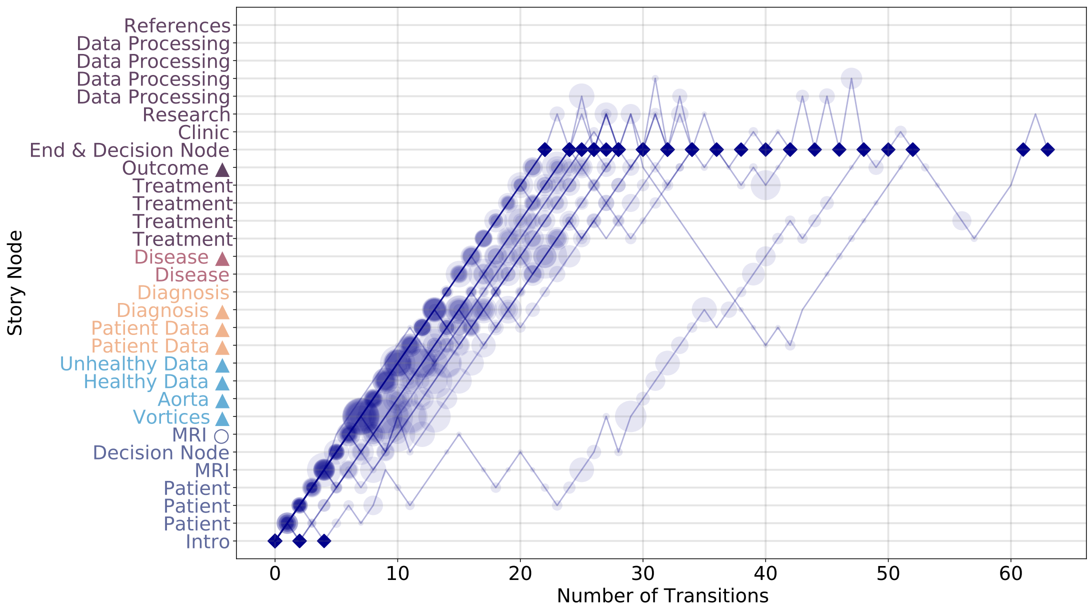
(b) CSVD Scrollytelling

Figure 54: Logging of the click paths of the participants, for (a) the CSVD Slideshow and (b) CSVD Scrollytelling. Each line represents a participant. Higher frequency of participants are indicated by darker lines. Larger circles indicate longer time spent at the respective story part. The x-axis shows the number of view transitions (e.g., slide changes), while the y-axis displays story nodes color-coded by their position in Freytag's Pyramid (see Figure 51). Interactive visualizations or animations are marked with a filled triangle ▲, optional branches with an empty circle ○, and nodes that are both interactive and optional with an outlined triangle △. Time spent on the first and last nodes is excluded due to frequent idle times and is instead marked with fixed-size diamond glyphs. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

9



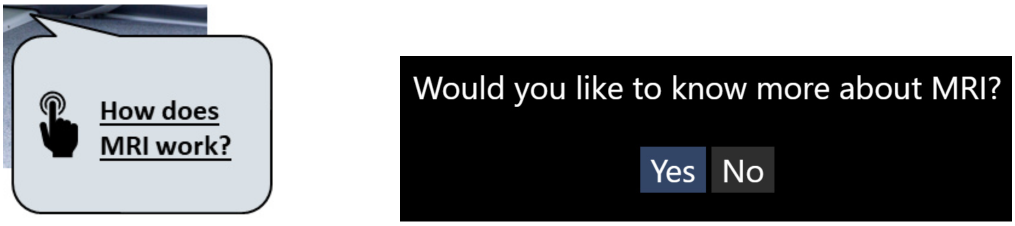
(a) BAV Slideshow



(b) BAV Scrollytelling

Figure 55: Logging of the click paths of the participants, for (a) the BAV Slideshow and (b) BAV Scrollytelling. Each line represents a participant. Higher frequency of participants are indicated by darker lines. Larger circles indicate longer time spent at the respective story part. The x-axis shows the number of view transitions (e.g., slide changes), while the y-axis displays story nodes color-coded by their position in Freytag's Pyramid (see Figure 51). Interactive visualizations or animations are marked with a filled triangle ▲, optional branches with an empty circle ○. Time spent on the first and last nodes is excluded due to frequent idle times and is instead marked with fixed-size diamond glyphs. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

9



(a) Slideshow (b) Scrollytelling

Figure 56: Interaction for branching from the main path in the (a) Slideshow and (b) Scrollytelling versions of the BAV story. Image from Mittenentzwei et al. [157], available under a CC BY 4.0 license. No changes were made.

Overall, navigation through the slideshow genre was perceived as easier and more intuitive compared to scrolling, as reflected in Figure 53 (statements 3 and 4). This preference may be attributed to the familiarity of clicking as a standard method for navigating slideshow-based interfaces. The slow scrolling speed had a negative influence on user perception of scroll-based navigation (Figure 53, statement 6), which may also explain the longer time spent within the scrollytelling versions. While the study findings indicate a general preference for clicking over scrolling, this should not be interpreted as a universal conclusion. There are various approaches to implementing scrollytelling, and the current evaluation is limited to a single implementation.

Each story version included an interactive section in which participants were asked to make a diagnosis based on information presented in prior parts of the story. As illustrated in all four versions in Figure 55 and Figure 54, a notable number of participants revisited earlier sections to review relevant content before responding. This behavior suggests that embedding quiz elements within a story can effectively direct user attention to specific information. Optional branches presented after the end node in the BAV story were rarely explored, whereas optional content embedded within the main narrative was accessed more frequently. This difference may stem from the weaker integration of the post-end branches into the core storyline. Moreover, these segments were not directly related to the patient's experience, potentially reducing their perceived relevance. On average, participants spent more time engaging with the scrollytelling versions than with the slideshows. This aligns with other findings indicating that users are more likely to revisit previous content when interacting with scrollytelling. Feedback from participants also highlighted the need for further investigation into the role of slide numbers as navigational tools. While they may support orientation, they could also contribute to visual clutter and detract from the overall aesthetic experience.

The questionnaire results indicate that the story topic had a greater impact on participants' ability to follow the story than the genre, as shown in Figure 53 (statement 1). The CSVD stories received generally positive evaluations. However, some participants found the CSVD storyline more difficult to follow. This may be attributed to the inclusion of increased interaction within individual slides, which could have diverted attention away from the narrative content. This interpretation is supported by earlier observations indicating that some participants experienced challenges interacting with the CSVD slideshow, as reflected in Figure 53 (statement 3).

TRANSITION. Across all story versions, the majority of participants agreed that the central theme was easy to follow (Figure 53, statement 1). Responses for both versions of the CSVD story displayed more variation. Transitions between story parts were rated slightly more favorably for the slideshow, though both formats received generally high ratings (Figure 53, statement 5). Navigation speed was also perceived as appropriate, with stronger agreement observed for the slideshows (Figure 53, statement 6). Returning to previously viewed content was considered slightly easier in the slideshows. However, due to a technical issue, responses from the first 10 participants in the BAV slideshow group were not recorded for this question (Figure 53, statement 7).

Analysis of logged navigation data showed that participants revisited previously viewed story nodes more frequently in the scrollytelling versions. This pattern was particularly evident in the BAV stories. In the BAV slideshow, users made an average of 3.17 transitions to earlier nodes (σ: 4.41), whereas for BAV Scrollytelling, the average rose to 10.43 transitions (σ: 14.13). For the CSVD stories, the average number of transitions to previously visited nodes was 7.75 (σ: 9.6) in the slideshow and 8.58 (σ: 9) for scrollytelling.

Overall, the stories received positive responses from participants. Several individuals expressed interest in exploring similar narratives during their free time. Feedback highlighted that the experience was enjoyable and reminiscent of digital information displays found in museums, with the story length noted as appropriate.

Some participants expressed a desire for more in-depth content, such as detailed explanations of specific steps in the BAV therapy or clarification of technical terms like 4D PC-MRI. The blood flow visualizations in the aorta were difficult for some to interpret, with distinctions between pre- and post-treatment more often identified through changes in vessel shape rather than flow patterns. Minor technical issues were also reported in both stories, including unintuitive zoom behavior and occasional interface inconsistencies.

In the BAV slideshow, one participant noted that the inclusion of slide numbers felt unnecessary and caused confusion when navigating through optional slides. The navigation arrows in this version were also mentioned as not being well-integrated into the overall slide design. Additional arrows used for highlighting elements within visualizations were mistaken for navigation controls, leading to further confusion.

In the CSVD story, the layout of certain slides was perceived as unclear due to a high density of charts and text. This visual complexity contributed to uncertainty about how to proceed within the story and may be reflected in the less favorable responses to interaction ease, as shown in Figure 53 (statement 3).

The scrolling mechanism implemented in this study differs from typical scrolling behavior found on online news platforms, as it involves fading out the current scene before introducing the next. This approach may have contributed to perceptions of unintuitiveness, which could explain why participants rated the transitions in the scrollytelling versions as slightly less pleasant, as shown in Figure 53 (statement 5). Among all versions, the transitions in the BAV slideshow, which was implemented using PowerPoint, received the highest ratings for transition quality (Figure 53, statement 5).

Participants who recorded a higher number of slide transitions tended to rate the transitions as slightly less intuitive and less pleasant in the questionnaire.

9.5 REFLECTIONS ON STUDY DESIGN AND LIMITATIONS

Participants were not obligated to complete the questionnaire, potentially introducing a bias toward responses from individuals who found the story enjoyable. While this may influence the overall results, it is unlikely to affect the comparison between the two narrative genres, as the same bias applies to all participants. For future studies, it is recommended to track the number of participants who choose not to complete the questionnaire and gather insights into their reasons for opting out. The findings offer a proof-of-concept for the development and evaluation of narrative genres in data-driven disease storytelling, particularly in terms of user preferences and interaction behavior. However, the ease of implementation and the outcomes were influenced by the software used and its associated technical constraints.

The findings of this study are derived from only two story examples, which involved certain trade-offs, particularly in the selection of transitions and interaction methods. A shared storyboard was developed to ensure compatibility across both narrative formats, namely scrollytelling and slideshow, enabling the implementation of equivalent content in each format. As a result, the investigation focused on a limited portion of the broader design space. Therefore, the results should be viewed as an initial exploration rather than offering fully generalizable conclusions, providing a foundation for further investigation in this area.

The development process revealed a clear need for specialized software capable of supporting features such as multilingual story design and user-friendly navigation controls that enable the creation of multiple story paths or direct access to non-adjacent slides. Additionally, maintaining stylistic consistency across various media types proved challenging and had a noticeable impact on the perceived quality of the final stories.

Collaboration with professional designers, such as those involved in projects for major publications like the New York Times, could enhance the visual quality of the story interface, including typography, color schemes, and layout design. However, the primary focus of this proof-of-concept was on examining narrative genre, interaction methods, navigation, and transitions, as well as identifying practical insights applicable to comparable storytelling projects.

9.6 DISCUSSION OF RESEARCH QUESTIONS

RQ4.1 DO USERS PREFER MORE OR LESS INTERACTIVE CONTENT AND HOW MUCH TIME TO THEY SPEND ENGAGING WITH INTERACTIVE STORY CONTENT? The amount of interactive content was rated as adequate, regardless of story topic or genre. However, as participants did not spend notably more time on interactive content, compared to non-interactive content, it cannot be concluded that the interactive content had any effect on the audience. It might have caused some confusion in the beginning of the stories, where participants needed more time to figure out how the interaction works. On the other hand, the increased times users spent at the first interactive or animated content might also be caused by initial curiosity.

RQ4.2 IS VERTICAL NAVIGATION, VIA SCROLLING, OR HORIZONTAL NAVIGATION, USING CLICKING, PREFERRED BY USERS? Participants found slideshow navigation easier and more intuitive than scrolling, likely due to the familiarity of clicking. Slow scrolling negatively affected the scrollytelling experience, leading to longer interaction times and less favorable ratings for backward navigation. Despite this, scrollytelling prompted more frequent revisits to earlier content, especially when quizzes encouraged review. Optional branches unrelated to the main storyline were rarely explored. Overall, scrollytelling increased engagement time but raised usability concerns, such as scrolling speed and navigation clarity.

RQ4.3 HOW DO DIFFERENT STORYTELLING FORMATS AND INTERACTION MECHANISMS AFFECT USERS' ABILITY TO FOLLOW THE NARRATIVE? The story topic influenced participants' narrative comprehension more than the genre, with CSVD stories generally rated positively. However, the CSVD slideshow was harder to follow, possibly due to distracting interactions. The scrollytelling transitions were seen as less pleasant, likely because of their unconventional fade-out effect. In contrast, the BAV slideshow had the highest-rated transitions, benefiting from smoother PowerPoint-based navigation.

The frequency of user comments regarding transitions suggests that ease of navigation notably influences the overall interaction experience. The evaluation underscores the value of smooth and efficient transitions between adjacent slides, as well as the need for mechanisms that enable direct navigation between non-adjacent sections of the story.

RQ4 HOW DO THE DIFFERENT GENRES SCROLLYTELLING AND INTERACTIVE SLIDESHOW SHAPE THE PERCEPTION OF DISEASE STORIES? Overall, the scrollytelling implementations received more critique than the slideshows. However, considering the limitations discussed in Section 9.5, this can be heavily influenced by individual design choices and is therefore not generalizable for all stories from these genres. Participants viewing the scrollytelling implementations had more frequent backtracking. Apart from that, story content and individual design decisions showed greater influence on user behavior and user feedback than the genres. Therefore, no definitive answer for how genres shape the perception of the disease stories can be formulated, and many different factors, such as variations in design, output media, and personal preferences likely have a notable influence which was not evaluated in this study.

9.7 CONCLUSION

Two specific use cases for disease stories were examined through the development of stories focusing on BAV and CSVD. The study revealed several noteworthy findings related to interactivity, navigation, and transitions, offering valuable directions for further investigation.

9.7.1 *Study Implications*

As the study examined only two genre implementations representing a limited portion of the overall design space, the generalizability of the findings is constrained. Rather than offering broad conclusions, the focus of this work lies in identifying initial insights and outlining multiple directions for future research. Broader generalizations will require a more extensive collection of stories that explore a wider variety of design elements, including diverse transition styles and interaction patterns.

The stories were implemented using two distinct narrative genres: the slideshow genre with standard click-based interaction, and the scrollytelling genre with scroll-based interaction. Study results indicated a general preference for clicking, although this may be attributed to suboptimal implementation of the scrolling experience.

9

The BAV story was structured using the martini glass structure, though the optional paths offered at the conclusion of the narrative were infrequently accessed. Based on the study findings, it is recommended to integrate optional paths throughout the story rather than limiting them to the final sections.

The level of interactivity within the data visualizations appeared to have minimal impact on user experience. Stories incorporating GIFs (BAV slideshow) and videos (BAV scrollytelling and CSVD scrollytelling) yielded similar usability ratings and interaction patterns when compared to the CSVD slideshow, which featured 3D object manipulation, such as rotation and clipping.

Incorporating interactive tasks, such as asking users to provide a diagnosis, proved effective in directing attention to specific information within the stories. However, it remains unclear whether such interactions are perceived as engaging in an entertaining sense (e.g., through gamification), or whether they contribute to increased knowledge retention.

Numerous participants provided feedback regarding the quality of transitions and navigation elements, highlighting their notable influence on the overall enjoyment of the story experience. These aspects play a critical role in shaping user engagement and should be carefully considered in future disease story designs. Established practices from news websites may serve as valuable references for addressing several of the concerns raised. For instance, to help users recognize that scrolling is required at the beginning of a story, partially visible text at the bottom of the screen could serve as an intuitive cue. In line with participant suggestions, navigation elements that support direct transitions between non-adjacent story segments should also be incorporated to improve accessibility and user control.

9.7.2 *Future Work*

Future research could explore alternative interaction styles, such as continuous storytelling during scrolling rather than relying on discrete content reveal and hide transitions. Additionally, varying scrolling speeds could be examined to enhance usability. Beyond slideshow and scrollytelling formats, further comparisons with other genres – such as data comics or data videos – may offer deeper insights into user engagement and interaction preferences.

Additional research is needed to assess effective strategies for incorporating optional branches and to determine which types of supplemental information are most engaging and relevant to users.

Further investigation into the impact of interactive tasks and gamification strategies on users' perception of the presented information would be valuable. Such research could help clarify whether these elements enhance engagement, comprehension, or retention, and how they influence the overall storytelling experience in disease stories.

Further research is necessary to explore strategies for enhancing transition and navigation quality across various storytelling genres, aiming to create a more seamless and engaging experience. While existing work primarily addresses transitions between information visualizations in scrollytelling, additional studies are needed to examine transitions involving scientific visualizations, as well as the interplay between scientific and information visualizations.

10 THE EFFECT STAGE: TARGET AUDIENCE

This stage focuses on the desired effect of a disease story on the audience, addressing the question, *"What to communicate?"* A disease story should convey a message shaped by the intent behind the story, typically aimed at addressing a clinical practice dilemma, and by the characteristics of the target audience (see Figure 57). Accordingly, this chapter takes a closer look at the target audience.

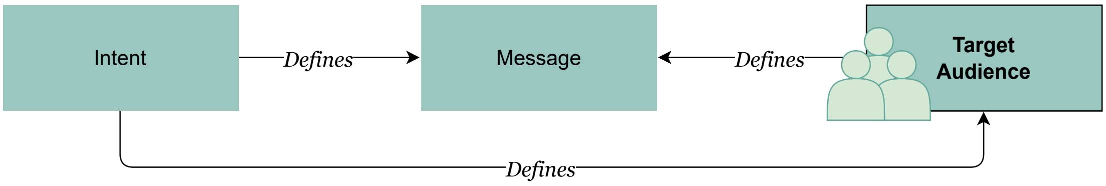


Figure 57: The effect stage of the disease story design process (cf. Figure 9) with *target audience* highlighted.

While data-driven approaches such as narrative visualization are increasingly used to create educational content, many related studies are conducted in controlled laboratory environments. Therefore, the following study examines user feedback on publicly available visualizations. Due to the limited availability of publicly accessible visualizations based on medical imaging data, YouTube serves as a valuable reference point, particularly because it frequently features 3D medical visualizations. A total of 14,550 user comments were analyzed through a combination of manual review and machine learning methods, including natural language processing for sentiment and emotion analysis. To address this, the following research questions are explored:

RQ5 How do users react to educational health videos on YouTube?

- *RQ5.1* What are motivations for users to watch the videos?
- *RQ5.2* What influences users' sentiment and emotions towards the videos?

10.1 ANALYSIS OF YOUTUBE COMMENTS OF VIDEOS COMMUNICATING HEALTH

In the context of narrative visualization designed to explain medical concepts to non-expert audiences, analyzing existing visual materials offers valuable insights into the design space and the prevalence of specific design choices. Although medical imaging data often represents anatomical structures in three dimensions, and 3D models are increasingly adopted in anatomy education [11], 3D medical visualizations aimed at laypersons remain relatively uncommon. To address this gap, a study was conducted focusing on YouTube videos that incorporate 3D medical visualizations, as this platform offers a broader and more accessible collection of such content. YouTube also provides access to extensive user feedback through metrics such as view counts, replay behavior, and public comments, enabling analysis outside the constraints of controlled laboratory environments.

Collaboration and Personal Contribution

This chapter is partially based on the following publication: [161]. The conceptual foundation for the project was developed by the author of this thesis. The first project was implemented by Danish Murad as part of a masters thesis supervised by the author of this thesis. Throughout the implementation process, the concept was collaboratively discussed and refined. The scientific reappraisal of the project, including the creation of the research paper, was conducted by the author of this thesis. The submitted work has been prepared with all co-authors who provided feedback throughout the project and writing phase.

The rise of online health information consumption, particularly accelerated by the COVID-19 pandemic, has notably altered the educational landscape. Increasingly, individuals are turning to online platforms, including YouTube and social media, for health-related information [268]. This analysis aims to explore how health education materials, comparable to narrative medical visualizations produced in academic contexts, are perceived by lay audiences.

The dataset includes 76 YouTube videos featuring 3D visualizations on a wide range of medical topics, including diseases, treatments, and preventative care, sourced from five prominent medical channels. A combination of qualitative and quantitative methods is applied, utilizing NLP techniques to analyze user comments. This approach enables the identification of sentiment and emotional tone in user responses, contributing to a better understanding of audience motivations and content preferences.

10.2 RELATED WORK

The review covers research related to 3D visualization in educational contexts, with a particular focus on medical content presented on YouTube and corresponding video analysis. Additionally, it includes foundational work in NLP aimed at analyzing comments from social media platforms.

10.2.1 *Analysis of YouTube Videos in Health Education*

3D medical visualizations have been shown to enhance comprehension and support interactive learning experiences [11], although their development typically requires collaboration with domain experts and animators.

YouTube plays an important role in promoting health literacy across diverse populations [130]. According to Park and Goering [186], lay persons turn to YouTube for health-related information due to its accessibility and empowering nature, which contributes to improved health competence and learning outcomes. However, the potential for misinformation, particularly from non-expert content creators, remains a concern [122]. To address this, the presented analysis focuses solely on videos published by verified health organizations, excluding those created by private individuals.

Burgess and Green [99] explored YouTube's evolution into a platform for public communication, emphasizing its democratizing potential, creative freedom, accessibility, and community-building features enabled by open access. Research by Welbourne and Grant [256] identified key factors influencing the popularity of science and health channels, noting that user-generated content frequently attracts more views than professionally produced videos. Shorter video formats have also been found to be more effective for science communication [271]. In addition, Amini et al. [8] outlined common design patterns used in data videos, while Shi et al. [216] proposed a comprehensive design space for video content.

10.2.2 *Natural Language Processing*

NLP techniques are widely employed to analyze user-generated content on platforms like YouTube, offering valuable insights from large-scale textual data. Sentiment analysis, in particular, enables the evaluation of emotional tone in written content. Pang and Lee [184] demonstrated its effectiveness in extracting user opinions from text.

A major advancement in NLP was introduced by Vaswani et al. [243] with the development of the *transformer model*, which leverages an attention mechanism to capture long-term dependencies and allows for efficient parallel processing. This innovation laid the foundation for subsequent models such as GPT by Radford et al. [199], optimized for sequential text generation, and BERT by Devlin et al. [51], known for its strong contextual understanding.

Comparative evaluations by Alaparthi and Mishra [5] identified BERT's superior performance in sentiment classification tasks. Further validation was provided by Chiorrini et al. [34], who confirmed BERT's effectiveness in analyzing both emotions and sentiments. Building on this, Liu et al. [139] introduced RoBERTa, an enhanced version of BERT that achieved improved results across a range of NLP benchmarks.

Additional studies have shown how modern NLP methods can be applied to analyze sentiment in online discussions. Porreca et al. [191] identified a shift in sentiment in user comments on Italian vaccination videos following a public campaign. Similarly, Bozkurt and Aras [23] analyzed comments on videos related to cleft lip and palate, noting a mix of sentiments, with negative comments often focusing on surgical pain and social stigma.

10.3 VIDEO SELECTION

The analysis focuses on four key aspects: (1) view count, (2) number of comments, (3) comment content, and (4) replay rates. The video search was performed using the keywords *"3D medical animation"* and *"3D medical visualization."* YouTube's filtering options include relevance (default), view count, upload date, and rating. The filters based on upload date and rating were found to be less effective, as they often surfaced recent videos with minimal user engagement and limited relevance.

In 2022, YouTube introduced a feature known as the *replay graph* (see Figure 58), which displays the portions of a video that are most frequently replayed. This feature appears as a semi-transparent area chart directly above the video's progress bar. However, not all videos include a replay graph, and limited information is available regarding the conditions under which it appears, making it unclear what specific criteria a video must fulfill to include this feature. The selection of videos for analysis was refined using the following criteria:

- Videos must feature 3D visualizations of human anatomy.
- A replay graph must be available for the video.
- The content must be published by credible channels.
- The primary language of the video must be English.
- Viewer comments must also be written in English.

The selection process aimed to include a diverse set of topics, anatomical structures, video durations, and upload dates. Credibility was assessed by verifying whether the channels were operated by organizations with established expertise in health communication. Ultimately, 76 videos were chosen from five channels: *Nucleus Medical Media* (47 videos), *Scientific Animations* (9 videos), *Dandelion Medical Animation* (15 videos), *BioDigital, Inc.* (4 videos), and *Infuse Medical* (1 video).

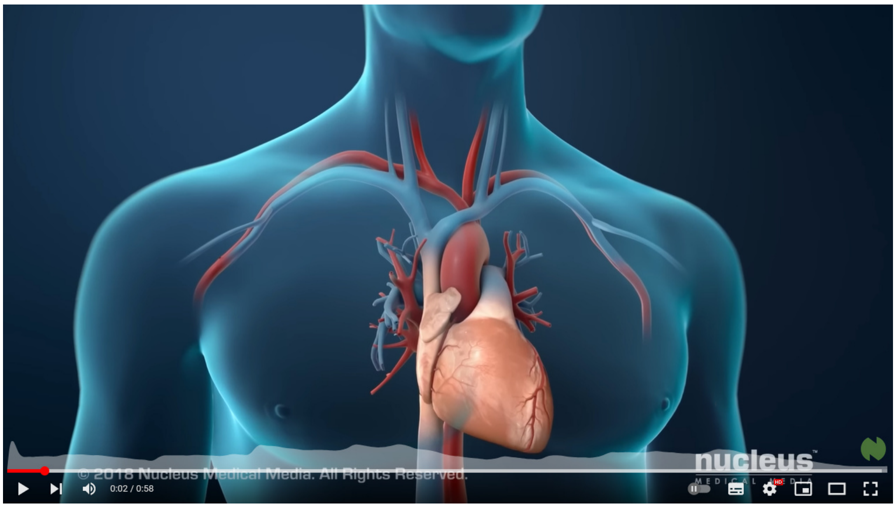


Figure 58: Screenshot from the video *What is a Coronary Angioplasty?* (Nucleus Medical Media) displaying the replay graph at the bottom. Image from Mittenentzwei et al. [161], available under a CC BY 4.0 license. No changes were made.

10.4 EVALUATION

The evaluation focuses on the analysis of user comments, which follows a semi-automatic approach using NLP and manual review.

Supplementary Material

Supplementary material for this chapter can be found at:
`visualstories.cs.ovgu.de/phd-sarah/effect-stage/target-audience`

10.4.1 *Data Collection and Preprocessing*

Data were collected using the YouTube API, which provided access to video titles, channel names, view counts, upload dates, durations, and top-level comments (excluding replies). Likes were not analyzed due to the lack of publicly available dislike counts, which limits interpretability. Replies to comments were excluded from the analysis, as they require contextual interpretation alongside the corresponding top-level comment. To maintain user privacy, no usernames or personally identifiable information were collected. Replay graphs were obtained by extracting images from the HTML source of each video page. These graphs represent relative replay frequencies across different segments of the video rather than displaying absolute replay counts, with peaks indicating areas of increased number of replays. Each video was manually reviewed to align viewer responses with specific content in the video.

PREPROCESSING OF COMMENTS. After collection, the textual content of the comments underwent preprocessing steps essential for NLP tasks. Accurate identification of characters, words, and sentences is foundational for all subsequent analysis. The text was cleaned to remove informalities such as spelling errors, slang, neologisms, URLs, special characters, and emoticons using the Python libraries *demoji* and *regular expression (RegEx)*. Non-English comments were excluded using a fine-tuned RoBERTa-based language detection model, which demonstrates an average accuracy of 99.6% across 20 languages [185]. This preprocessing stage resulted in the removal of 584 comments. The final dataset comprised 14,550 top-level comments across 76 videos.

SENTIMENT ANALYSIS. The *"Twitter-RoBERTa-base for Sentiment Analysis"* model [142] was applied to perform sentiment classification, selected for its training on the short, informal text typical for social media platforms. Pre-trained on approximately 124 million tweets posted between January 2018 and December 2021, the model was subsequently fine-tuned for sentiment analysis in social media contexts. It categorizes each comment as positive, neutral, or negative and provides a corresponding sentiment score, reflecting the probability that the assigned sentiment accurately represents the comment's content. This model is particularly suitable, as the structure and tone of YouTube comments closely resemble the unstructured nature of tweets, which are constrained to 280 characters [215].

EMOTION ANALYSIS. Emotion detection enables a more nuanced understanding of the sentiments expressed by users. For this purpose, a pre-trained RoBERTa-based model fine-tuned on an emotion-labeled dataset was utilized [143]. The dataset includes 28 distinct emotions, allowing the model to perform multi-label classification. This approach facilitates the identification of a wide range of emotional responses in YouTube comments, such as joy, approval, disapproval, anger, and disappointment. Each comment was assigned one or more emotions along with corresponding emotion scores, indicating the likelihood of each detected emotion.

10.4.2 *Discussion of Results*

The analysis included both sentiment and emotion of the comments. Additionally, individual comments that directly referenced the video content are discussed in more detail. A moderate positive correlation (0.65) was observed between the number of views and the number of comments per video, with one notable exception. A video related to COVID-19 received approximately 339 million views but only 836 comments, whereas other videos with notably fewer views had over 1,000 comments. However, the comment count alone is not a reliable indicator of user engagement, as it can be affected by moderation, including filtering comments prior to posting or the removal of existing comments by the channel administrator.

Replay graphs show notable variability, both in the location of peaks relative to video content and in their overall shapes. Some graphs feature only a few distinct peaks, while others display a smooth progression or a zigzag pattern with multiple smaller peaks. It is possible that the replay graphs function as self-reinforcing mechanisms, once a peak is visible, viewers may be more likely to navigate to that section, thereby increasing its prominence for future viewers. As a result, the structure of a replay graph may not be entirely determined by the underlying video content and may not reliably reflect what segments are most informative or engaging. Additionally, no clear relationship was identified between the video content associated with high replay activity and the content of user comments.

SENTIMENT ANALYSIS. The overall sentiment distribution across comments was mostly even, with a slight predominance of positive responses. Specifically, 35.92% of comments were classified as positive, followed by 30.46% neutral and 33.62% negative. A manual review of representative comments revealed the following patterns:

- *Positive Sentiment (35.92 %):* These comments reflected approval, appreciation, or support for the video content. Viewers expressed agreement with the information presented or shared positive personal experiences related to the topic.
- *Neutral Sentiment (30.46 %):* Neutral comments tended to be informational or inquisitive in nature. They often involved requests for additional details, clarification, or similar videos, and generally lacked strong emotional tone.
- *Negative Sentiment (33.62 %):* Negative responses typically involved criticism, expressions of concern, or the sharing of adverse personal experiences. Such sentiments were often associated with disagreement with the content, reported side effects, or unfavorable outcomes related to the topic.

Most videos contained fewer than 200 comments and a normalized analysis was necessary to identify overarching patterns and highlight specific videos that warranted closer examination (see Figure 59). To capture the extremes within the dataset, particular attention was given to the ten videos with the highest proportion of negative comments and the ten videos with the highest proportion of positive comments.

TEN VIDEOS WITH HIGHEST PROPORTION OF NEGATIVE COMMENTS. Due to the volume of comments, manual review of the entire dataset was not practical. Instead, a keyword-based filtering approach was applied to identify comments related to the presentation, style, clarity, and overall quality of the videos. Keywords used in the filtering process included terms such as "video," "visualization," "animation,"

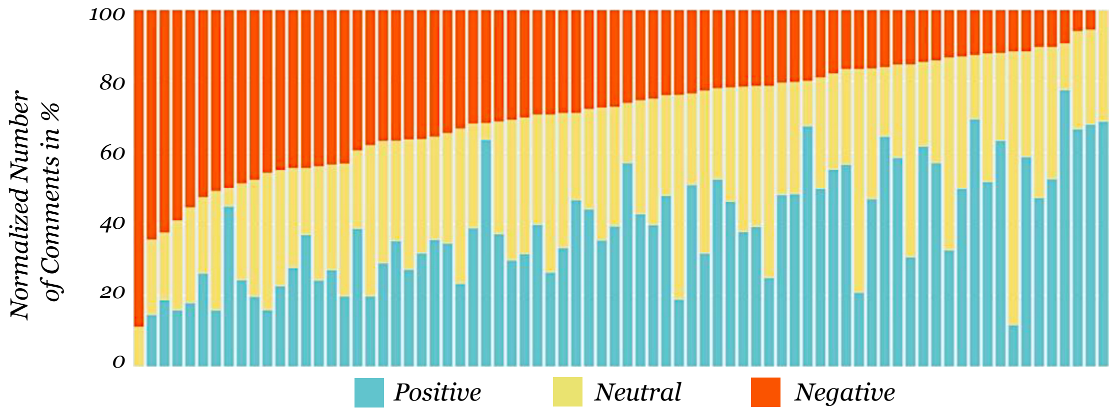


Figure 59: Normalized distribution of all videos, sorted in ascending order by the number of negative comments. Image from Mittenentzwei et al. [161], available under a CC BY 4.0 license. No changes were made.

"graphics," "quality," "audio," "voice," "presentation," "background," and "sound." Following this filtering step, a combined analysis of detected emotions and manual inspection was conducted to better understand viewers' perceptions regarding the production and presentation of the video content:

- Viewers frequently expressed appreciation for the videos, highlighting their usefulness, educational value, and clarity. The content was often described as informative, straightforward, and easy to comprehend. Selected comments include:
  - *"Thank you for sharing. I've been battling endometriosis for years, it is extremely awful. This was a great visual representation :"*
  - *"Your videos are really helpful. Well explained in a simple and uncomplicated way. You made learning enjoyable."*
  - *"Thanks for so precious information and for being so clearly displayed."*
  - *"This really broke it down & made it so easy to understand! Thank you. The visual is great."*
  - *"Your animations are actually another level of quality that makes u guys so different and interesting to watch, your videos are educational but at the same time it is not tiring to watch and try to learn, great video keep going!!!"*

10

- Negative feedback was commonly directed at robotic, monotone, or synthetic voiceovers, which were perceived as lacking engagement. In addition, some viewers raised concerns about potential emotional or health-related triggers within the video content. The following comments illustrate these reactions:
  - *"such an unfriendly voice that just causes anxiety. horrible!"*
  - *"Oh I hate these computer voiceovers"*

– "*Pleaseee use a different voice for these videos. People who are concerned about their health do not want a cold robotic voice adding to their anxiety.*"
– "*The irony that this video triggered a migraine with the lights*"
– "*just an FYI, you should include a trigger warning before showing visuals of an aura. It can cause migraines for those who get them.*"

TEN VIDEOS WITH HIGHEST PROPORTION OF POSITIVE COMMENTS. A similar methodology was applied to analyze the group of videos with predominantly positive comments. Most of these videos contained fewer than 50 comments, except one video featuring 116 comments and another titled "From Fertilization to Childbirth," which stood out with 1,113 comments. Due to the high volume of responses, viewer reactions in the latter video had a disproportionate influence on the overall sentiment distribution within this group.

Similar patterns were observed as in the videos with predominantly positive sentiment. Many viewers expressed appreciation for the videos and animations, highlighting their usefulness, educational value, and clarity, often thanking the content creators. Negative comments typically reflected personal experiences or general opinions related to the video's topic. A few comments addressed specific aspects, such as the lack of information on symptoms (e.g., *"What are the symptoms for this"*) or dissatisfaction with background music (e.g., *"WHOLLY unnecessary music......"*). However, these instances were relatively rare. It remains unclear whether it influenced overall engagement or viewer sentiment.

REST OF THE VIDEOS. Following the analysis of videos with the extremes of the sentiment distributions, the remaining 56 videos exhibited a more balanced mix of sentiments. The presence of neutral comments indicated that many viewers used the comment section to ask questions, offer suggestions, or share observations that were not overtly positive or negative.

Across these videos, similar themes as priorly discussed emerged through a combination of sentiment, emotion, and keyword analysis. Positive feedback frequently acknowledged the videos as informative and helpful in enhancing understanding, often accompanied by expressions of gratitude toward the channels. In contrast, some comments reflected negative or neutral viewpoints, as illustrated by the following examples:

- Viewers' comments on the use of music varied, often within the same video, highlighting the subjective nature of preferences and the complexity of decision-making in design:
  – "*Wonderful explanation! and also I loved the background piano music. May I know the name of the music?*"
  – "*umm the music is really annoying............. can't concentrate*"

  - "*Thanks for the video. Very well done and explained. Music is very relaxing also.*"
  - "*I'm just curious why there's such sad/dark ambient music in the background, makes it feel like I'm watching a dog shelter PSA*"

- The use of plain dark backgrounds was criticized: "*Why are medical 3D animations always scary like there is something about the plain dark background, or the lack of sound apart from the voice of the narrator, idk*". This remark suggests that certain background choices may evoke specific emotional responses in viewers. However, due to the limited number of similar comments across the dataset, it is difficult to establish this as a generalized design consideration.
- Again, the use of a natural and soothing human voice was positively received, in contrast to computer-generated or monotonous voiceovers, which were often viewed less favorably:
  - "*Does anyone else find these videos very relaxing? Aside from being fascinating, I just feel at home amongst these animations and calm voice and neutral colours.*"
  - "*The sound is soo calming*"
- Some viewers raised concerns about the pacing of the videos, suggesting that certain segments moved too quickly or too slowly:
  - "*I set the speed to 1.25x. This video is a bit slow otherwise...*"
  - "*Why is this on slow mo????*"
  - "*Would've been great if the narrator explained a bit slow for non-medical people like me.*"

EMOTION ANALYSIS. Comments associated with each detected emotion were manually reviewed, with particular attention given to the five most frequently occurring emotions. Emotions outside this group appeared in fewer than 5 % of all comments and were therefore excluded from detailed analysis, see Figure 60.

- *Neutral (34.37 %):* These comments primarily engaged with the topic by asking questions, providing information, or sharing experiences, without conveying strong emotional responses. This aligns with findings from the sentiment analysis, suggesting that content-focused or descriptive remarks often lack overt emotional tone.

- *Admiration (14.54 %):* Comments conveyed respect, appreciation, or positive sentiment either toward the video content or individuals affected by the conditions discussed in the videos.

- *Gratitude (9.13 %):* Viewers expressed gratitude, often in response to the clarity of the explanations, the educational value, or the helpfulness of the information presented.

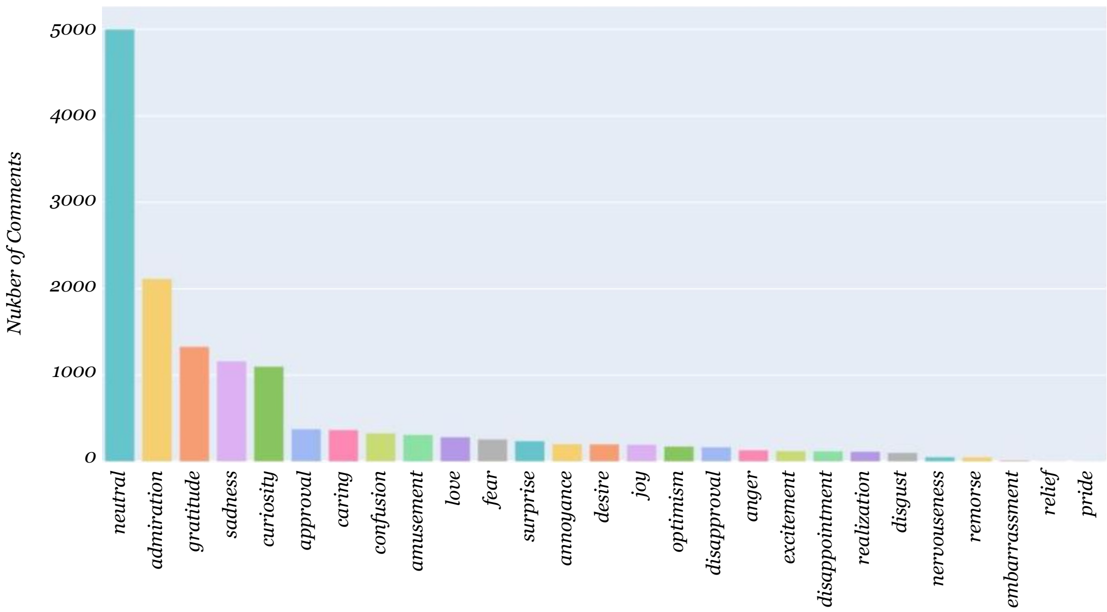


Figure 60: Emotion distribution for comments across all videos.

- *Sadness (7.99 %):* Comments often involved sharing personal struggles and losses or expressing empathy.
- *Curiosity (7.57 %):* Viewers asked follow-up questions, requested additional information, or demonstrated interest in learning more about the topics covered.

other correlating factors. In addition to the previous analysis, a correlation analysis was conducted to explore whether specific sentiments were more common in certain types of videos, specific channels, or in relation to particular video attributes such as view count, duration, or the presence of voiceovers. This approach aimed to determine, e.g., whether certain channels consistently received more positive sentiment, potentially reflecting differences in production strategies. The normalized sentiment distribution appeared relatively stable across channels and videos. Slightly more positive sentiment was observed in comments on videos featuring preventive or informational content. This outcome is not unexpected, as topics related to general or preventive health information are less likely to elicit strong emotional reactions compared to content addressing severe or acute medical conditions, which may be more emotionally impactful for some viewers. Sentiment distribution across channels showed slight variation, with *Dandelion Medical Animation* and *Infuse Medical* receiving a higher proportion of positive comments. This trend appears to be primarily associated with the nature of their content, which focused more on preventive or informational topics which are areas that generally elicit more favorable responses. Despite the

10

use of sentiment and emotion analysis, consistent patterns across all videos and channels were limited. This may be influenced by the inherently subjective nature of viewer preferences, where content perceived as clear and engaging by some may be viewed as insufficient or overly technical by others. Moreover, the context in which viewers engage with these videos, such as academic study, personal curiosity, or health-related concerns, can play an important role in shaping their perceptions and responses.

10.5 REFLECTIONS OF STUDY DESIGN AND LIMITATIONS

The video dataset is primarily composed of content from the *Nucleus Medical Media* channel, which accounts for over half of the selected videos due to its high video output. This imbalance in sample distribution may have influenced the findings. Incorporating a broader range of channels could potentially reveal additional patterns and insights.

YouTube's search results are not static and influenced by individual user behavior and account preferences. Consequently, the same search criteria may yield different results for different users. Additionally, the visibility and content of comments can be affected by channel moderation, such as filtering or deletion. As a result, some viewer responses may have been excluded from the analysis due to lack of public availability.

It is also important to consider that users who leave comments may not reflect the broader audience that passively consumes content without interacting. External factors, including channel popularity, posting time, video topic, and current trends, can further influence user engagement and sentiment, potentially introducing confounding variables into the analysis.

Automated sentiment analysis tools, while useful, may misinterpret elements such as sarcasm, irony, or culturally specific expressions. To mitigate this, a manual review of comments was conducted to ensure more accurate interpretation.

10.6 DISCUSSION OF RESEARCH QUESTIONS

RQ5.1 WHAT ARE MOTIVATIONS FOR USERS TO WATCH THE VIDEOS?

Since many comments contained personal testimonials, personal dismay was identified as a major reason for watching the videos. However, as discussed as limitation, it might be that users with other motivations simply did not write comments or did not mention their motivations in the comments. Thus, personal dismay might be the motivation that indices users the most to write comments, but that does not mean that it is the most common motivation for watching the videos.

RQ5.2 **WHAT INFLUENCES USERS' SENTIMENT AND EMOTIONS TOWARDS THE VIDEOS?**
Again, personal dismay and the topic of the video caused the sentiment and emotions detected in the user comments, e.g., to be more negative when discussing fatal diseases compared to explanations of healthy biological processes. Some design decisions, such as robotic voice over and dark color scales were also negatively mentioned in the comments. However, comments on concrete design decisions were sparse so that no generalizable conclusion could be made. Furthermore, the same design decisions that were praised by some users were criticized by others, which suggests, that personal taste of the users play a notable role.

RQ5 **HOW DO USERS REACT TO EDUCATIONAL HEALTH VIDEOS ON YOUTUBE?**
While some comments criticized individual design decisions or facts in the videos, many comments were positive, showing appreciation for the videos and their ability to explain medical topics. Users frequently posted about personal experiences with the videos topics, which might hint at a need to communicate these experiences but is not related to the concrete design of the videos. However, as discussed in Section 10.5, the comment base might be biased, not only do to many users not writing comments, but also due to comment moderation done by the channels' owner, which is more likely to affect negative comments. Overall, the number of likes and comments shows, that there is a viewer base that shows general interest in the videos.

10.7 Conclusion

YouTube comments were analyzed to identify the reasons why users watched health-related content and their emotional responses.

10.7.1 *Study Implications*

User motivation. A notable portion of comments involved users sharing personal experiences, suggesting that a primary motivation for viewing these videos may stem from personal concern or past encounters with the medical condition, either first-hand or through a family member. Notably, several comments referenced experiences that occurred in the past, indicating that the emotional relevance may persist even after the immediate situation has resolved. Although other possible viewing contexts, such as general interest or academic use, are plausible, these motivations were not explicitly mentioned in the comment sections.

Positive/negative feedback. Viewers highlighted positively the videos' ability to simplify complex medical topics through clear and visually appealing presentations, indicating a general preference for simplicity and aesthetics. While no particular visual style was consistently identified, overall feedback on the visual elements was favorable. Human voiceovers were generally preferred over synthetic ones, although improvements in AI-generated speech may influence this preference in the future.

Background music elicited varied responses. While it was perceived as distracting or irritating when overly prominent, it was also noted to enhance liveliness in videos lacking voiceovers. In videos featuring gender-specific topics, narrators generally matched the gender of the people affected by a specific topic.

Narration speed was a point of critique in several comments. In the video titled *"How do carbohydrates affect your weight?"*, the narration was considered too slow, whereas in *"COVID-19 Animation: What Happens If You Get Coronavirus?"*, it was perceived as too fast. The latter video did have a slightly faster narration. However, other factors may also influence the preferred narration speed, such as the narrator's style or the complexity of the topic. The narrator of the video about carbohydrates sounds monotonous, which makes listening less interesting and possibly feels lengthy. The optimal pacing may also vary depending on the complexity of the topic. Topics involving intricate biomedical processes, such as those related to COVID-19, are likely to benefit from a slower, more deliberate delivery compared to content covering more general health topics like a healthy diet.

CORRELATIONS WITH VISUAL CONTENT. Videos featuring darker visual themes tended to receive a higher proportion of negative comments, while those with more varied or lighter color palettes were associated with more positive feedback. A slightly increased frequency of comments expressing “sadness” was also noted in darker-themed videos. However, these findings are not conclusive enough to determine whether color alone influences emotional response. The topic and presentation style likely have a substantial impact as well. When a serious topic such as cancer is paired with a dark visual theme, the emotional tone may be further intensified.

Despite the variety of design decisions across the videos, direct viewer feedback on specific design elements was limited. Instead, the majority of comments focused on sharing personal experiences related to the medical conditions the videos covered.

10.7.2 *Future Work*

The selection of videos was skewed towards one specific channel. Therefore, future investigations should include a broader spectrum of channels. Investigating similar content on other social media platforms could lead to further insights, as different platforms are likely used by different user groups, which might influence both the design of the content and the comments.

Future studies are needed to include answers to top-level comments. Analyzing comments with respect to their hierarchy would allow investigating discussions amongst users and might lead to new insights. Furthermore, including the number of likes a comment receives would provide additional information about how many users agree, which could be used to weight the comments differently.

Contradicting opinions, e.g., regarding background music, show that individual preferences for specific design decisions differ. Analyzing if specific preferences apply to, e.g., demographic user groups would help tailor future visualizations to specific target audiences. However, this would imply that information about the comment authors need to be acquired. While this might be impossible on YouTube, insights about the demographics of the general user base who viewed a video can be gained. Therefore, future studies should seek cooperations with channel owners or create own channels, as channel owners are provided with additional information about the users. This would also lead to more insight about filtered comments. It is also possible to register a research program with YouTube and make API calls [274]. However, this only grants access to public data, so detailed information about a specific channel's viewer base may not be available through this method.

Including a more detailed analysis of the video content can provide insights into common design practices, such as the use of colors, camera paths, or terminology used to convey medical topics.

Part III

REFLECTIONS

The thesis concludes with a discussion of the main research question, the general limitations of the work, and reflections on the interdisciplinarity and generalizability of the proposed approach. Based on the insights gained, several directions for future research are outlined.

10

11 DISCUSSION

The main research question of this thesis is discussed, followed by a reflection on the research limitations, the interdisciplinary nature of the approach, and its potential for generalization.

11.1 DISCUSSION OF THE MAIN RESEARCH QUESTION

The primary objective of this thesis is to explore the design of disease stories as a specialized form of narrative medical visualization. This objective is driven by the main research question introduced in Chapter 1:

> **Main Research Question**
>
> *How can narrative medical visualization be leveraged to improve the health literacy of lay audiences through disease stories?*

Due to the vast number of possible design decisions, a comprehensive analysis of the entire design space is beyond the scope of this thesis. Consequently, a definitive answer to the research question remains elusive. However, this thesis contributes a framework, the disease story design process, grounded in practical experience and supported by research on health message design. This design process helps structure and streamline the creation of future disease stories. Key aspects of the design process were investigated through case studies, focusing on story content, character, structure, and target audiences.

The conducted studies revealed both advantages and limitations of the applied methods. While the findings are not fully generalizable, as different implementations of similar techniques might yield different results, they offer guidance for the design of future disease stories.

Comparing different visualization and presentation techniques showed that participants' task performance, particularly in assessing growth and shrinkage patterns, was not notably affected. However, Phong shading was consistently preferred aesthetically, which might be caused by its use of color. While illustrative techniques like outlines may enhance accessibility for color-blind users, the clinical partners also preferred the Phong shading implementation due to its familiar color scheme. Colors were chosen to balance realism and color contrast to surrounding structures, with blood vessels rendered in red and tumors in yellow.

Interestingly, interaction had a greater impact on the accuracy of user task results than the visualization techniques. This suggests that prioritizing intuitive interaction, alongside aesthetic preferences, is essential when designing disease stories. However, interaction design must account for varying levels of user familiarity with 3D interfaces.

Human characters play an important role in personalizing disease stories. Designing characters is complex, as demographic and health-related attributes must be visually encoded. GenAI models were useful for generating diverse characters quickly, though limitations were observed, probably stemming from the models' training data. Discussions with design experts emphasized that despite the use of GenAI authors must make conscious decisions about aspects such as art style, facial expressions, and body posture.

The clinical cooperation partners expressed a preference for using photos of real physicians involved in the project, believing it could increase the story's credibility if users can recognize the physicians on the hospital websites or in person. However, the design experts highlighted the importance of consistent visual styles throughout the story. Therefore, using real photos of physicians would imply using photorealistic images for patients, which can also be achieved using GenAI. Whether real photos actually enhance credibility was not evaluated in this thesis. Furthermore, while including human protagonists seemed to improve engagement, credibility was high across all story versions regardless of the presence of a physician character. This may be caused by credits shown at the end of the stories, suggesting that credible information delivery does not solely depend on the character type.

While various story structures were explored, no direct comparison was made. In practice, character-driven structures based on Campbell's Hero's Journey proved most helpful for organizing story content. Tension-based (Freytag's Pyramid) or interaction-driven (martini glass) structures mainly guided tension curves or interaction and linearity but left more content-related decisions to the author. This can be challenging, especially for inexperienced authors.

Participants showed little interest in optional story paths at the end of stories designed with the martini glass structure, suggesting that essential content should be integrated earlier. Additionally, genre had a stronger influence on user behavior than interaction-driven story structure. Scrollytelling versions encouraged more revisits to earlier content compared to slideshow versions. However, similar to the interaction results in Chapter 6, participants rarely used interaction features, raising the question of whether interaction is inherently unpopular or if its implementation was suboptimal. Future disease stories should therefore carefully consider whether and how to integrate interactive elements.

Collaboration with interdisciplinary teams including clinicians and designers was highly beneficial. Medical experts were crucial for identifying relevant topics and ensuring content accuracy, while design experts contributed to creating engaging and aesthetically pleasing visualizations. As demonstrated, participants expressed aesthetic preferences even when task performance was similar across visualization techniques, highlighting the importance of appealing design. In particular, interaction designers may help create more motivating and accessible interactive elements as this proved to be a major challenge in current case studies.

Overall, the results provide a foundation for the interdisciplinary design of disease stories and offer valuable insights for future research at the intersection of visualization and health communication.

11.2 LIMITATIONS

This thesis explores only a small portion of the extensive design space for disease stories. As such, it does not aim to establish definitive design guidelines. Instead, it provides a foundational framework for approaching the design of disease stories systematically by introducing key steps in a design process and investigating fundamental design choices, primarily through a series of case studies.

Creating a disease story is a resource-intensive endeavor, requiring time and the collaboration of experts from various disciplines. As a result, only a limited number of disease stories could be developed and analyzed within the scope of this thesis. This small set of application examples limits the generalizability of the findings. Accordingly, the evaluations refrain from using inferential statistical methods, as the aim is not to generalize across all types of disease stories but to provide initial insights.

Another constraint was the limited availability of study participants. A range of recruitment strategies was employed, including university mailing lists, public events, crowdsourcing platforms, and social media. However, participation outside academic circles remained low, and no financial compensation was offered. As a result, most participants were young adults with academic backgrounds. In contrast, the intended target audience for many disease stories would likely be older individuals and those within specific risk groups, depending on the disease in focus. A more demographically diverse participant pool, particularly in terms of education level, would have improved the representativeness of the study results.

Furthermore, the studies introduced a survivorship bias, as only the responses of participants who completed the full evaluation were included in the analysis. Given that participation was voluntary, it is likely that these individuals had a higher intrinsic motivation, which may have positively skewed the results. In addition, the controlled study environment limits insights into how users perceive and interact with disease stories in real-world settings. To address this gap, Chapter 10 analyzed user comments on YouTube videos covering health topics. However, the design of those videos differs notably from the interactive visualizations developed using the disease story design process described in Chapter 4. Therefore, it remains unclear to what extent findings from YouTube-based studies can be transferred to disease stories presented in other formats or on other platforms.

11.3 INTERDISCIPLINARITY

Creating disease stories is a highly interdisciplinary endeavor that benefits from expertise across a range of fields. The research presented in this thesis draws on knowledge from data visualization and health communication, incorporates practices from storytelling, and is informed by expert input from both medicine and design.

In a broader context, narrative medical visualization can be considered a subfield of public health. Other related domains, though not explored in depth within this work, could offer valuable perspectives, particularly journalism (especially data journalism and health journalism) and science communication. The latter goes beyond the boundaries of health communication and may provide additional insights and methodologies from other domains that can be applied to disease storytelling.

This thesis involved active collaboration with practicing clinicians from various medical disciplines, researchers in medical visualization, and designers specializing in interaction design and medical illustration. While it is feasible to develop design guidelines or templates for cases where a professional designer is unavailable, the creation of a disease story should never proceed without the involvement of a medical domain expert. Their role is crucial not only for identifying relevant and communicable topics, but also for ensuring the medical accuracy of the content. Although publicly available datasets exist for some conditions, others may lack accessible data. Domain experts can fill these gaps by contributing relevant clinical data and offering practical insights from their experience in communicating with patients and families. Their involvement ensures that the resulting narrative is both accurate and meaningful.

11.4 GENERALIZABILITY

All studies were conducted within the medical domain, specifically focusing on diseases. This raises the question of how generalizable the insights are to other domains. Medicine is a sensitive and often deeply personal topic, especially when related to disease. While many scientific disciplines, such as mathematics or philosophy, may feel abstract or distant to a lay audience, healthcare directly affects individuals. Personal experiences or the health of relatives can create strong emotional connections to medical content. As a result, the information presented in disease stories may be perceived differently compared to content from other domains. This suggests that audience perception may vary depending on the topic, even if, e.g., similar visualization techniques are used. To draw definitive conclusions, further studies using data from nonmedical domains are necessary.

General storytelling structures used throughout this thesis, such as Freytag's Pyramid or the martini glass structure (see Chapter 9), are not domain-specific and can be applied broadly. In contrast, the adapted version of Campbell's Hero's Journey presented in Chapter 8 was specifically tailored to patient and physician characters explaining a certain disease. Therefore, this structure is limited to disease communication and would need to be adapted to communicate other medical topics, or topics outside the medical domain.

The overall design process introduced in Chapter 4 was developed through practical experience in the creation of disease stories in collaboration with clinical experts. While most components of the process are domain-independent, the concept of "clinical practice dilemmas" is unique to healthcare. However, this concept could be adapted to suit other domains. For example, in environmental science, a dilemma might involve challenges such as light pollution and its harmful effects on wildlife. In summary, while the disease story design process shows strong potential for adaptation across domains, further application and evaluation in other fields are required to confirm its generalizability.

11

12 FUTURE WORK AND CONCLUSION

12.1 FUTURE WORK

Future work can build upon the frameworks and insights from this thesis to further explore the design space for disease stories and other narrative medical visualizations.

12.1.1 *Exploiting the Design Space*

VISUALIZATION TECHNIQUES. Additional visualization techniques could be explored with lay audiences. For example, techniques not yet included in the studies presented in Chapter 6 (which included Phong shading, outlines with feature lines, and hatching combined with Fresnel shading) can be investigated. Alternatively, current techniques can be extended by testing alternative implementations, such as increasing the density of hatching lines. Beyond surface rendering techniques, direct volume rendering is another frequent approach for visualizing volumetric image data. Future studies can investigate how to tailor direct volume rendering to lay audiences.

Chapter 9 incorporated blood flow visualizations using streamlines in the stories about BAV. Study participants reported difficulty interpreting these visualizations, highlighting the need for further research. Effectively visualizing blood flow for lay audiences likely requires balancing guidance and clarity by reducing visual clutter while preserving anatomical context and appropriate accuracy. Further flow visualization techniques for lay audiences could be transferred to other application cases, e.g., airflow in the trachea.

In addition to refining visualization techniques, how visualizations are presented within a virtual 3D space plays a crucial role. Research on optimal camera angles, camera paths, and cinematic techniques has shown potential to make narrative visualizations more immersive [42, 97, 151]. Future studies could build upon this work to explore how these techniques can be adapted for the medical domain.

OUTPUT DEVICES AND PUBLICATION PLATFORMS. The disease stories and visualizations presented in this work were designed for desktop environments. However, different output devices require distinct design decisions. For instance, mobile devices typically feature smaller screens and a vertical default orientation, which affects both visualization layouts and interaction design. Moreover, interaction paradigms differ, as mobile devices primarily rely on touch input, unlike desktop environments that commonly use a computer mouse [128].

Virtual reality (VR) environments are particularly promising for improving immersion, and existing studies have already explored the design of narrative visualizations in VR [127, 269].

Future work could investigate which types of medical stories (e.g., depending on topics or target audiences) are best suited for specific output devices. Additionally, systematically exploring the design space of medical visualizations tailored to various platforms offers a valuable direction for further research.

12.1.2 *Target Audiences*

LITERACY. A key motivation of this thesis was to address the challenge of limited health literacy, which often prevents individuals from making informed decisions about their health. Future studies could evaluate the impact of disease stories on improving health literacy by using established health literacy questionnaires, such as the HLS-EU-Q [225], to determine whether measurable improvements occur after engaging with a disease story. Furthermore, it became evident that visualization literacy plays a crucial role in determining the audience's ability to interpret the data visualizations embedded within disease stories. Future studies should consider visualization literacy as an additional key factor when evaluating the effectiveness of disease stories. Existing tools, such as the Mini-VLAT [183], can be used to assess participants' visualization literacy.

However, such tools typically do not account for volume visualizations, highlighting the need for future work to develop specific evaluation methods for (medical) volume visualizations [62]. In addition, technological literacy may influence how users experience disease stories, especially when interactive content is involved [229]. Considering visualization and technological literacy alongside health literacy will help to ensure that disease stories are not only informative but also accessible and usable for a wider range of audiences, ultimately improving their effectiveness.

ACCESSIBILITY. Designing disease stories with attention to the target audience's levels of visualization literacy, digital literacy, and health literacy is essential for improving accessibility. Since disease stories primarily rely on visual content, it is crucial to consider accessibility for users with visual impairments to avoid unintentional discrimination [109]. Moreover, when addressing patients with specific needs, such as those experiencing neurological deficits caused by CSVD, additional design considerations may be necessary. Developing targeted guidelines for the use of text, audio, and visual elements tailored to various patient groups would be a valuable step toward creating more accessible and inclusive disease stories.

GAMIFICATION AND INTERACTIVITY. The evaluations presented in Chapters 6 and 9 revealed that participants often did not utilize the available interaction opportunities. This highlights the need for future research to determine when and how interactive elements should be incorporated into disease stories. The optimal design of interactive content may depend on characteristics of the target audience, such as age, technological literacy, and visualization literacy.

Furthermore, the appropriate level of interactivity may vary depending on the topic of the disease story. For instance, when educating patients with serious illnesses about treatment options, a linear story design with fewer interactive elements might be preferable, ensuring that essential information is conveyed clearly and in a sensitive manner. In contrast, for disease stories aimed at the general public, incorporating interactive content and gamification elements could improve engagement and motivation, similar to the positive effects observed in educational applications [6].

The disease stories are designed as one-time content, intended for a single interaction in which the authors aim to achieve specific effects, such as reaching defined learning objectives. Future research is needed to explore the potential of long-term or follow-up applications, and how to design them effectively. For instance, health-tracking apps already employ gamification strategies to encourage users to monitor and improve their health [44]. Such approaches hold promise for fostering sustained motivation, particularly in promoting healthy lifestyles and assessing whether users apply insights from the disease stories to their own lives.

Another promising direction for future work is to investigate when animations might be more effective than interactive content, particularly considering users' ability to steer the content. In this context, existing animation styles for narrative visualizations could be analyzed to determine how they can be adapted for medical applications [272]. Comparative evaluations with animated and interactive content could serve to identify preferences of different target audiences.

EVALUATION. One major limitation of the presented case studies is the demographics of the participants who are usually not similar to the intended target audience of the disease stories. Therefore, future work should make efforts towards recruiting more diverse groups of study participants focussing on demographic groups for which the respective medical topic is especially relevant. The general motivation to take part in the presented research studies were highest in individuals with an academic background. Cooperations with local clinics and physicians could help reaching a broader audience by displaying disease stories, e.g., in waiting rooms. This opens up the possibility for long-term evaluations outside controlled lab settings. However, even if this approach reaches a broader audience, securing sufficient participation in such evaluations may still be challenging.

Experiences from evaluations in the area of science communication can improve evaluation design, making it more attractive for participants. For example, *ImpactUnit* [262] is an initiative funded by the German Federal Ministry of Education and Research, supporting science communicators in evaluating and improving the success of their projects. Therefore, the initiative offers advice and practical tools, such as fact sheets, examples, and consultation.

In addition to targeting the general public, some disease stories address topics specifically relevant to patients. However, this thesis did not include evaluations involving patients directly. While patients could be reached through clinical collaboration partners, thus overcoming the recruitment challenges encountered with the general public, ethical considerations become particularly important when working with patients. Future studies on topics tailored to patients would greatly benefit from involving patients in the evaluation process to ensure the relevance and effectiveness. However, obtaining appropriate ethical approval is essential before conducting such studies.

12.1.3 *Authoring*

COOPERATION. The development of the disease stories and visualizations presented in this thesis involved close collaboration with clinicians and designers. However, additional cooperation with experts in didactics and communication, such as data journalists, health journalists, or museum professionals, could further improve the design of disease stories. These groups possess valuable expertise not only in effectively communicating complex information to lay audiences but also in engaging and motivating target audiences to seek out and interact with this information.

MAINTENANCE AND UPDATABILITY. Since medical research is continuously evolving, leading to new insights and updated recommendations, e.g., for prevention and treatment, a key challenge is ensuring that disease stories remain current. Efficiently updating content would require a deployment pipeline that allows individual texts or images to be modified without the need to redesign the entire story. Ideally, this would also enable domain experts to independently update stories. This creates two important requirements: (1) disease stories should be built from modular and reusable components, and (2) authoring tools should support version control and provide metadata, such as timestamps and authorship information, to ensure transparency and improve credibility. Future research should focus on identifying reusable components for disease stories and developing template-based authoring systems to streamline the creation and maintenance of new stories. This would ensure the sustainability of disease stories, allowing experts in design and programming to be primarily involved in the initial creation, while domain experts can independently maintain and update the content to ensure its continued accuracy.

ARTIFICIAL INTELLIGENCE. In Chapter 7, GenAI was employed to generate characters for disease stories. Future research could further investigate the potential of GenAI to support the creation of other types of content, such as narrative text or visual elements like icons. Training GenAI models specifically for disease stories (e.g., focusing on the accurate depiction of anatomical structures and pathologies) could improve the quality of generated content. Integrating GenAI-based features into an authoring tool could improve the accessibility of the authoring process and reduce development time. However, implementing verification mechanisms would be essential to ensure the accuracy of generated content and to prevent copyright infringements.

12.2 FINAL CONCLUSION

The disease story design process was introduced as a framework to streamline the creation of future disease stories. Several key steps of this process were explored through case studies, focusing on story content, character design, and narrative structure. Additionally, insights into potential target audiences were gained through an analysis of YouTube comments on educational videos about medical topics. The strong interest from clinical partners further emphasized the relevance and potential of disease stories across various application areas. While the studies presented were largely exploratory and do not yield generalizable results, they provide a foundation for further research into disease stories as a promising tool for medical science communication.

Part IV

APPENDIX

ONLINE MATERIALS

Online materials for the studies include supporting information such as interactive stories and screenshots of the evaluations.

Thesis Overview
`visualstories.cs.ovgu.de/phd-sarah`

- Additional Material Overview
- Publications Overview

The Narrative Stage: Content (Chapter 6)
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/content`

- Interactive Evaluations
- Evaluation Screenshots

The Narrative Stage: Character (Chapter 7)
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/character`

- Character Images
- Questionnaire

The Narrative Stage: Character-driven Structure (Chapter 8)
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/character-driven-structure`

- Interactive Stories
- Questionnaire

The Narrative Stage: Structure (Chapter 9)
`visualstories.cs.ovgu.de/phd-sarah/narrative-stage/structure`

- Interactive Stories
- Questionnaire

The Effect Stage: Target Audience (Chapter 10)
`visualstories.cs.ovgu.de/phd-sarah/effect-stage/target-audience`

- List of Videos
- Movie Barcodes and Replay Graph

BIBLIOGRAPHY

[1] Eytan Adar and Elsie Lee. *Communicative Visualizations as a Learning Problem*. 2020. DOI: 10.48550/ARXIV.2009.07095.

[2] Adobe Inc. *Adobe Creative Cloud*. Version CC 2019. [Accessed 03-May-2024]. Mar. 6, 2019. URL: https://www.adobe.com/creativecloud.html.

[3] Emrah Akkoyun, Sebastian T. Kwon, Aybar C. Acar, Whal Lee, and Seungik Baek. "Predicting abdominal aortic aneurysm growth using patient-oriented growth models with two-step Bayesian inference." In: *Computers in Biology and Medicine* 117 (2020), p. 103620. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2020.103620.

[4] Nizia Alam and Jiyeon So. "Contributions of emotional flow in narrative persuasion: An empirical test of the emotional flow framework." In: *Communication Quarterly* 68.2 (2020), pp. 161–182. DOI: 10.1080/01463373.2020.1725079.

[5] Shivaji Alaparthi and Manit Mishra. "BERT: A sentiment analysis odyssey." In: *Journal of Marketing Analytics* 9.2 (2021), pp. 118–126. DOI: 10.1057/s41270-021-00109-8.

[6] Mareen Allgaier, Florentine Huettl, Laura Isabel Hanke, Tobias Huber, Bernhard Preim, Sylvia Saalfeld, and Christian Hansen. "Gamification Concepts for a VR-based Visuospatial Training for Intraoperative Liver Ultrasound." In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI '24. ACM, May 2024, pp. 1–8. DOI: 10.1145/3613905.3650736.

[7] Fereshteh Amini, Matthew Brehmer, Gordon Bolduan, Christina Elmer, and Benjamin Wiederkehr. "Evaluating Data-Driven Stories and Storytelling Tools." In: *Data-Driven Storytelling*. Ed. by Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. Accessed: 2024-12-18. CRC Press, 2017. Chap. 11, pp. 249–276. DOI: 10.1201/9781315281575-11.

[8] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. "Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens." In: *Proc. of ACM Conference on Human Factors in Computing Systems*. 2015, pp. 1459–1468. DOI: 10.1145/2702123.2702431.

[9] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Jason Leboe-McGowan, and Pourang Irani. "Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement." In: *Proceedings of the 2018 International*

Conference on Advanced Visual Interfaces. AVI ’18. ACM, May 2018, pp. 1–9. DOI: 10.1145/3206505.3206552.

[10] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. “Space, time and visual analytics.” In: *International Journal of Geographical Information Science* 24.10 (2010), pp. 1577–1600. DOI: 10.1080/13658816.2010.508043.

[11] Carlos M. Ardila, Daniel González-Arroyave, and Mateo Zuluaga-Gómez. “Efficacy of three-dimensional models for medical education: A systematic scoping review of randomized clinical trials.” In: *Heliyon* 9.2 (2023), e13395. ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2023.e13395.

[12] AUTOMATIC1111. *Stable Diffusion Web UI*. [Accessed 03-May-2024]. 2024. URL: https://github.com/AUTOMATIC1111/stable-diffusion-webui.

[13] Ruth Aylett. “Narrative in Virtual Environments - Towards Emergent Narrative.” In: *Proceedings of the AAAI fall symposium on narrative intelligence*. Jan. 1999, pp. 83–86.

[14] Reddit Netherlands B.V. *Reddit*. Accessed: 2025-03-03. Amsterdam, Netherlands, 2025. URL: https://www.reddit.com/.

[15] SurveySwap B.V. *SurveySwap*. Accessed: 2025-03-03. Amsterdam, Netherlands, 2025. URL: https://surveyswap.io.

[16] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Köppen, and Barbara Tversky. “Narrative Design Patterns for Data-Driven Storytelling.” In: *Data-Driven Storytelling*. A K Peters/CRC Press, 2018, pp. 107–133. ISBN: 9781315281575. DOI: 10.1201/9781315281575-5.

[17] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. “Design Patterns for Data Comics.” In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. ACM, Apr. 2018, pp. 1–12. DOI: 10.1145/3173574.3173612.

[18] Ragnar Bade, Jens Haase, and Bernhard Preim. “Comparison of Fundamental Mesh Smoothing Algorithms for Medical Surface Models.” In: *Prof. of Simulation und Visualisierung*. SCS Publishing House e.V., 2006, pp. 289–304.

[19] Rachel Ballon. *Breathing life into your characters*. Penguin, 2009.

[20] Blender Foundation. *Blender*. Accessed: 2025-04-04. 2025. URL: https://www.blender.org.

[21] Anniek Boeijinga, Hans Hoeken, and José Sanders. “Storybridging: Four steps for constructing effective health narratives.” In: *Health Education Journal* 76.8 (2017), pp. 923–935. DOI: 10.1177/0017896917725360.

[22] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. "What Makes a Visualization Memorable?" In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2306–2315. ISSN: 1077-2626. DOI: 10.1109/tvcg.2013.234.

[23] Aylin Pasaoglu Bozkurt and Isıl Aras. "Cleft Lip and Palate YouTube Videos: Content Usefulness and Sentiment Analysis." In: *The Cleft Palate-Craniofacial Journal* 58.3 (2020), pp. 362–368. DOI: 10.1177/1055665620948722.

[24] Anna Brand, Linde Gao, Alexandra Hamann, Claudia Crayen, Hannah Brand, Susan M. Squier, Karl Stangl, Friederike Kendel, and Verena Stangl. "Medical Graphic Narratives to Improve Patient Comprehension and Periprocedural Anxiety Before Coronary Angiography and Percutaneous Coronary Intervention: A Randomized Trial." In: *Annals of Internal Medicine* 170.8 (2019), p. 579. ISSN: 0003-4819. DOI: 10.7326/m18-2976.

[25] Alan C Braverman, Hasan Güven, Michael A Beardslee, Majesh Makan, Andrew M Kates, and Marc R Moon. "The bicuspid aortic valve." In: *Curr. Probl. Cardiol.* 30.9 (2005), pp. 470–522. DOI: 10.1016/b978-1-4160-5892-2.00011-8.

[26] Stefan Bruckner and M Eduard Gröller. "Style transfer functions for illustrative volume rendering." In: *Computer Graphics Forum* 26.3 (2007), pp. 715–724.

[27] Beatrice Budich, Laura A. Garrison, Bernhard Preim, and Monique Meuschke. "Reflections on AI-Assisted Character Design for Data-Driven Medical Stories." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. 2023. DOI: 10.2312/vcbm.20231216.

[28] Joseph Campbell. *The hero with a thousand faces*. Vol. 17. New World Library, 2008.

[29] Marc Cavazza, Fred Charles, and Steven J Mead. "Interacting with virtual characters in interactive storytelling." In: *Proc. of Int. Conf. on Autonomous Agents and Multiagent Systems*. 2002, pp. 318–325.

[30] Fred Charles, Julie Porteous, and Marc Cavazza. "Changing characters' point of view in interactive storytelling." In: *Proc. of the international conference on Multimedia - MM '10*. ACM Press, 2010. DOI: 10.1145/1873951.1874321.

[31] Meng Chen, Robert A. Bell, and Laramie D. Taylor. "Persuasive Effects of Point of View, Protagonist Competence, and Similarity in a Health Narrative About Type 2 Diabetes." In: *Journal of Health Communication* 22.8 (July 2017), pp. 702–712. DOI: 10.1080/10810730.2017.1341568. URL: https://doi.org/10.1080/10810730.2017.1341568.

[32] Meng Chen, Matthew S. McGlone, and Robert A. Bell. "Persuasive Effects of Linguistic Agency Assignments and Point of View in Narrative Health Messages About Colon Cancer." In: *Journal of Health Communication* 20.8 (2015), pp. 977–988. DOI: 10.1080/10810730.2015.1018625.

[33] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. "How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools." In: *IEEE Transactions on Visualization and Computer Graphics* 30.8 (Aug. 2024), pp. 4429–4448. ISSN: 2160-9306. DOI: 10.1109/tvcg.2023.3261320.

[34] Andrea Chiorrini, Claudia Diamantini, Alex Mircoli, and Domenico Potena. "Emotion and sentiment analysis of tweets using BERT." In: *Workshop Proceedings of the EDBT/ICDT 2021 Joint Conference*. 2021. URL: https://ceur-ws.org/Vol-2841/DARLI-AP_17.pdf.

[35] Justyna Chojdak-Łukasiewicz, Edyta Dziadkowiak, Anna Zimny, and Bogusław Paradowski. "Cerebral small vessel disease: A review." In: *Advances in Clinical and Experimental Medicine* 30.3 (Mar. 2021), pp. 349–356. ISSN: 1899-5276. DOI: 10.17219/acem/131216.

[36] Noptanit Chotisarn, Sarun Gulyanon, Tianye Zhang, and Wei Chen. "VISHIEN-MAAT: Scrollytelling visualization design for explaining Siamese Neural Network concept to non-technical users." In: *Vis. Inform.* 7.1 (2023), pp. 18–29. DOI: 10.1016/j.visinf.2023.01.004.

[37] Civit AI, Inc. *Civitai*. [Accessed 03-May-2024]. 2024. URL: https://civitai.com (visited on 02/01/2024).

[38] Murray C. Clark and Roy L. Payne. "Character-Based Determinants of Trust in Leaders." In: *Risk Analysis* 26.5 (2006), pp. 1161–1173. DOI: 10.1111/j.1539-6924.2006.00823.x.

[39] Cleveland Clinic. *Medical Questions & Answers | Cleveland Clinic*. [Accessed 03-May-2024]. 2023. URL: https://my.clevelandclinic.org/health.

[40] LinkedIn Ireland Unlimited Company. *LinkedIn*. Accessed: 2025-03-03. Dublin, Ireland, 2025. URL: https://www.linkedin.com/.

[41] Matthew Conlen and Jeffrey Heer. "Idyll: A Markup Language for Authoring and Publishing Interactive Articles on the Web." In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST '18. ACM, Oct. 2018, pp. 977–989. DOI: 10.1145/3242587.3242600.

[42] Matthew Conlen, Jeffrey Heer, Hillary Mushkin, and Scott Davidoff. *Cinematic Techniques in Narrative Visualization*. 2023. DOI: 10.48550/ARXIV.2301.03109.

[43] Microsoft Cooperation. *Microsoft PowerPoint 2021*. Accessed: 2025-03-03. Redmond, WA, USA, 2021. URL: https://www.microsoft.com/en-us/microsoft-365/powerpoint.

[44] Victor Cotton and Mitesh S. Patel. "Gamification Use and Design in Popular Health and Fitness Mobile Applications." In: *American Journal of Health Promotion* 33.3 (July 2018), pp. 448–451. DOI: 10.1177/0890117118790394.

[45] Cody R. Criss and Mina S. Makary. "Recent Advances in Image-Guided Locoregional Therapies for Primary Liver Tumors." In: *Biology* 12.7 (2023). ISSN: 2079-7737. DOI: 10.3390/biology12070999.

[46] Weiwei Cui, Xiaoyu Zhang, Yun Wang, He Huang, Bei Chen, Lei Fang, Haidong Zhang, Jian-Guan Lou, and Dongmei Zhang. "Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements." In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (Jan. 2020), pp. 906–916. ISSN: 2160-9306. DOI: 10.1109/tvcg.2019.2934785.

[47] Paramita Dasgupta, Chloe Henshaw, Danny R. Youlden, Paul J. Clark, Joanne F. Aitken, and Peter D. Baade. "Global Trends in Incidence Rates of Primary Adult Liver Cancers: A Systematic Review and Meta-Analysis." In: *Frontiers in Oncology* 10 (Feb. 2020). ISSN: 2234-943X. DOI: 10.3389/fonc.2020.00171.

[48] Keshav Dasu, Yun-Hsin Kuo, and Kwan-Liu Ma. " Character-Oriented Design for Visual Data Storytelling." In: *IEEE Transactions on Visualization & Computer Graphics* 30.01 (2024), pp. 98–108. DOI: 10.1109/TVCG.2023.3326578.

[49] Anna De Fina and Alexandra Georgakopoulou. *Analyzing Narrative: Discourse and Sociolinguistic Perspectives*. Cambridge University Press, Nov. 2011. ISBN: 9781139051255. DOI: 10.1017/cbo9781139051255.

[50] Ratna Devi, Komal Kanitkar, R. Narendhar, Kawaldip Sehmi, and Kannan Subramaniam. "A Narrative Review of the Patient Journey Through the Lens of Non-communicable Diseases in Low- and Middle-Income Countries." In: *Advances in Therapy* 37.12 (2020), pp. 4808–4830. DOI: 10.1007/s12325-020-01519-3.

[51] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proc. of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2019, pp. 4171–4186. DOI: 10.18653/V1/N19-1423.

[52] Nicholas Diakopoulos, Funda Kivran-Swaine, and Mor Naaman. "Playable data: characterizing the design space of game-y infographics." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. ACM, May 2011, pp. 1717–1726. DOI: 10.1145/1978942.1979193.

[53] Die Ernährungs-Docs | ARD. *Fettleber heilen: Mit Ernährung natürlich gegensteuern und abnehmen | Die Ernährungs-Docs | NDR*. [Accessed 03-May-2024]. Youtube, 2022. URL: https://www.youtube.com/watch?v=eTz-W73HNZU&t=54s.

[54] Die Ernährungs-Docs | ARD. *Fettleber heilen mit Gemüse, Eiweiß und Vollkornprodukten | Die Ernährungs-Docs | ARD Gesund*. [Accessed 03-May-2024]. Youtube, 2023. URL: https://www.youtube.com/watch?v=Q-mQdIC1lkY&t=355s.

[55] Steven Drucker, Samuel Huron, Robert Kosara, Jonathan Schwabish, and Nicholas Diakopoulos. "Communicating data to an audience." In: *Data-driven storytelling*. Ed. by Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. AK Peters/CRC Press, 2018, pp. 211–231.

[56] Matthew Z. Dudley, Gordon K. Squires, Tracy M. Petroske, Sandra Dawson, and Janesse Brewer. "The Use of Narrative in Science and Health Communication: A Scoping Review." In: *Patient Education and Counseling* 112 (July 2023), p. 107752. ISSN: 0738-3991. DOI: 10.1016/j.pec.2023.107752.

[57] Lydia Dux-Santoy et al. "Increased rotational flow in the proximal aortic arch is associated with its dilation in bicuspid aortic valve disease." In: *Eur. Heart J. Cardiovasc. Imaging* 20.12 (2019), pp. 1407–1417. DOI: 10.1093/ehjci/jez046.

[58] Brent Dykes. *Effective data storytelling: how to drive change with data, narrative and visuals*. John Wiley & Sons, 2019.

[59] Inc. Epic Games. *Unreal Engine*. Accessed: 2025-03-03. Cary, U.S., 2025. URL: https://www.unrealengine.com/.

[60] Nina Errey, Jie Liang, Tuck Wah Leong, and Didar Zowghi. "Evaluating narrative visualization: a survey of practitioners." In: *International Journal of Data Science and Analytics* 18.1 (Mar. 2023), pp. 19–34. ISSN: 2364-4168. DOI: 10.1007/s41060-023-00394-9.

[61] Federal Employment Agency. *German Classification of Occupations 2010 — Revised Version 2020*. [Accessed 29-Jan-2024]. 2024. URL: https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Arbeitsmittel/Arbeitsmittel-Nav.html.

[62] Elif E. Firat, Alark Joshi, and Robert S. Laramee. "VisLitE: Visualization Literacy and Evaluation." In: *IEEE Computer Graphics and Applications* 42.3 (May 2022), pp. 99–107. ISSN: 1558-1756. DOI: 10.1109/mcg.2022.3161767.

[63] Klaus Fog, Christian Budtz, Philip Munch, and Stephen Blanchette. *Storytelling: Branding in Practise*. Springer Berlin Heidelberg, 2010. ISBN: 9783540883494. DOI: 10.1007/978-3-540-88349-4.

[64] Marsha E Fonteyn, Benjamin Kuipers, and Susan J Grobe. "A description of think aloud method and protocol analysis." In: *Qual. Health Res.* 3.4 (1993), pp. 430–441. DOI: 10.1177/104973239300300403.

[65] Christopher J François, Shardha Srinivasan, Mark L Schiebler, Scott B Reeder, Eric Niespodzany, Benjamin R Landgraf, Oliver Wieben, and Alex Frydrychowicz. "4D cardiovascular magnetic resonance velocity mapping of alterations of right heart flow patterns and main pulmonary artery hemodynamics in Tetralogy of Fallot." In: *J. Cardiovasc. Magn. Reson.* 14.1 (2012), 16:1–16:12. DOI: 10.1186/1532-429x-14-16.

[66] Karl-Ingo Friese, Marc Herrlich, and Franz-Erich Wolter. "Using Game Engines for Visualization in Scientific Applications." In: *New Frontiers for Entertainment Computing*. Springer US, 2008, pp. 11–22.

[67] Juhana Frösen, Riikka Tulamo, Anders Paetau, Elisa Laaksamo, Miikka Korja, Aki Laakso, Mika Niemelä, and Juha Hernesniemi. "Saccular intracranial aneurysm: pathology and mechanisms." In: *Acta Neuropathologica* 123.6 (Jan. 2012), pp. 773–786. ISSN: 1432-0533. DOI: 10.1007/s00401-011-0939-3.

[68] Yabo Fu, Yang Lei, Tonghe Wang, Walter J Curran, Tian Liu, and Xiaofeng Yang. "Deep learning in medical image registration: a review." In: *Physics in Medicine & Biology* 65.20 (Oct. 2020), 20TR01. DOI: 10.1088/1361-6560/ab843e.

[69] R. Fujishin. *Creating Communication: Exploring and Expanding Your Fundamental Communication Skills*. Rowman & Littlefield Publishers, 2008. ISBN: 978-0-74256-396-4. URL: https://books.google.de/books?id=sFrheqNL-ukC.

[70] Rocio Garcia-Retamero and Edward T. Cokely. "Designing Visual Aids That Promote Risk Literacy: A Systematic Review of Health Research and Evidence-Based Design Heuristics." In: *Human Factors* 59.4 (2017), pp. 582–627. DOI: 10.1177/0018720817690634.

[71] Laura Garrison, Monique Meuschke, Jennifer Fairman, Noeska N. Smit, Bernhard Preim, and Stefan Bruckner. "An Exploration of Practice and Preferences for the Visual Communication of Biomedical Processes." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. 2021, pp. 1–12.

[72] Laura A. Garrison, Monique Meuschke, Bernhard Preim, and Stefan Bruckner. "Current Approaches in Narrative Medical Visualization." In: *Approaches for Science Illustration and Communication*. Ed. by Mark Roughley. Cham: Springer Nature Switzerland, 2023, pp. 95–116. ISBN: 978-3-031-41652-1. DOI: 10.1007/978-3-031-41652-1_4.

[73] Julia Geiger, Michael Markl, Lena Herzer, Daniel Hirtler, Florian Loeffelbein, Brigitte Stiller, Mathias Langer, and Raoul Arnold. "Aortic flow patterns in patients with Marfan syndrome assessed by flow-sensitive four-dimensional MRI." In: *J. Magn. Reson. Imaging* 35.3 (2012), pp. 594–600. DOI: 10.1002/jmri.23500.

[74] Gérard Genette. *Narrative discourse: An essay in method*. Vol. 3. Cornell University Press, 1980.

[75] Nahum Gershon and Ward Page. "What storytelling can do for information visualization." In: *Communications of the ACM* 44.8 (2001), pp. 31–37.

[76] Sarah F. Frisken Gibson. "Constrained Elastic Surface Nets: Generating Smooth Surfaces from Binary Segmented Data." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 1998*. Vol. 1496. 1998, pp. 888–898. DOI: 10.1007/BFb0056277.

[77] Alessandro Giovanelli. “In Sympathy with Narrative Characters.” In: *Journal of Aesthetics and Art Criticism* 67.1 (2009), pp. 83–95. DOI: 10.1111/j.1540-6245.2008.01337.x.

[78] Peter Godfrey-Smith. “What Do Generalizations of the Lewis Signaling Model Tell Us About Information and Meaning?” In: *CUNY Pragmatics Workshop: Relevance, Games, & Communication. CUNY Graduate Center. New York City. Oct*. Vol. 14. 2014.

[79] Paolo Gritti et al. “The bio-psycho-social model forty years later: a critical review.” In: *Journal of Psychosocial Systems* 1.1 (2017), pp. 36–41.

[80] Yongfu Hao. *Towards Data Science*. [Accessed 03-May-2024]. 2024. URL: https://towardsdatascience.com/image-to-image-translation-69c10c18f6ff (visited on 02/01/2024).

[81] HarkerTechnologies. *HT Photorealism v4.1.7*. 2023. URL: https://civitai.com/models/114956/ht-photorealism-v417.

[82] Healthline. *Healthline: Medical information and health advice you can trust. — healthline.com*. [Accessed 03-May-2024]. 2023. URL: https://www.healthline.com/.

[83] David Herman. *Basic Elements of Narrative*. Wiley, Jan. 2009. ISBN: 9781444305920. DOI: 10.1002/9781444305920.

[84] Albert Hofman et al. “The Rotterdam Study: 2016 objectives and design update.” In: *European Journal of Epidemiology* 30.8 (2015), pp. 661–708. DOI: 10.1007/s10654-015-0082-x.

[85] T. A. Hope, M. Markl, L. Wigström, M. T. Alley, D. Miller, and R. Herfkens. “Comparison of Flow Patterns in Ascending Aortic Aneurysms and Volunteers using Four-Dimensional Magnetic Resonance Velocity Mapping.” In: *J. Magn. Reson. Imaging* 26.6 (2007), pp. 1471–1479. DOI: 10.1002/jmri.21082.

[86] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. “LoRA: Low-Rank Adaptation of Large Language Models.” In: *International Conference on Learning Representations*. 2022. URL: https://openreview.net/forum?id=nZeVKeeFYf9.

[87] J. Hullman and N. Diakopoulos. “Visualization Rhetoric: Framing Effects in Narrative Visualization.” In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2231–2240. ISSN: 1077-2626. DOI: 10.1109/tvcg.2011.255.

[88] Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. “A Deeper Understanding of Sequence in Narrative Visualization.” In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2406–2415. ISSN: 1077-2626. DOI: 10.1109/tvcg.2013.119.

[89] Jessica Hullman, Robert Kosara, and Heidi Lam. "Finding a Clear Path: Structuring Strategies for Visualization Sequences." In: *Computer Graphics Forum* 36.3 (June 2017), pp. 365–375. ISSN: 1467-8659. DOI: 10.1111/cgf.13194.

[90] Robert Hurford, Isabel Taveira, Wilhelm Kuker, and Peter M Rothwell. "Prevalence, predictors and prognosis of incidental intracranial aneurysms in patients with suspected TIA and minor stroke: a population-based study and systematic review." In: *Journal of Neurology, Neurosurgery & Psychiatry* 92.5 (2021), pp. 542–548. DOI: 10.1136/jnnp-2020-324418.

[91] Elaine Huynh, Angela Nyhout, Patricia Ganea, and Fanny Chevalier. "Designing Narrative-Focused Role-Playing Games for Visualization Literacy in Young Children." In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (Feb. 2021), pp. 924–934. ISSN: 2160-9306. DOI: 10.1109/tvcg.2020.3030464.

[92] Yuichiro Ii et al. "Hypertensive Arteriopathy and Cerebral Amyloid Angiopathy in Patients with Cognitive Decline and Mixed Cerebral Microbleeds." In: *Journal of Alzheimer's Disease* 78.4 (Dec. 2020). Ed. by Masahito Yamada, pp. 1765–1774. ISSN: 1875-8908. DOI: 10.3233/jad-200992.

[93] Inkscape Project. *Inkscape*. Accessed: 2025-04-04. 2025. URL: https://inkscape.org.

[94] Katherine Isbister. *Better Game Characters by Design*. CRC Press, June 2006. DOI: 10.1201/9780367807641.

[95] Hirono Ishikawa and Takahiro Kiuchi. "Health literacy and health communication." In: *BioPsychoSocial Medicine* 4.1 (2010), p. 18. ISSN: 1751-0759. DOI: 10.1186/1751-0759-4-18.

[96] Till Ittermann, Robin Haring, Henri Wallaschofski, Sebastian E Baumeister, Matthias Nauck, Marcus Dörr, Markus M Lerch, Henriette E Meyer zu Schwabedissen, Dieter Rosskopf, and Henry Völzke. "Inverse association between serum free thyroxine levels and hepatic steatosis: results from the Study of Health in Pomerania." In: *Thyroid* 22.6 (2012), pp. 568–574.

[97] Radhika Pankaj Jain, Kadek Ananta Satriadi, Adam Drogemuller, Ross Smith, and Andrew Cunningham. "Once Upon a Data Story: A Preliminary Design Space for Immersive Data Storytelling." In: *Companion Proceedings of the 2024 Conference on Interactive Surfaces and Spaces*. ISS Companion '24. Vancouver, BC, Canada: Association for Computing Machinery, 2024, pp. 63–68. ISBN: 979-8-40071-278-4. DOI: 10.1145/3696762.3698054.

[98] Bernard J. Jansen, Soon-Gyo Jung, Lene Nielsen, Kathleen W. Guan, and Joni Salminen. "How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment." In: *Pacific Asia Journal of the Association for Information Systems* 14 (Jan. 2022), pp. 1–28. DOI: 10.17705/1pais.14301.

[99] Burgess Jean and Green Joshua. "YouTube: Online Video and Participatory Culture." In: *European Journal of Communication* 35.4 (2020), pp. 419–423. DOI: 10.1177/0267323120935319.

[100] Jiroje Jiranukool, Peeraphong Thiarawat, and Waneerat Galassi. "Prevalence of intracranial aneurysms among acute ischemic stroke patients." In: *Surgical Neurology International* 11 (Oct. 2020), p. 341. ISSN: 2152-7806. DOI: 10.25259/sni_506_2020.

[101] Jonas Johé. *SurveyCircle*. Accessed: 2025-03-03. Mannheim, Germany, 2025. URL: https://www.surveycircle.com.

[102] Daniel Jönsson, Erik Sundén, Anders Ynnerman, and Timo Ropinski. "A Survey of Volumetric Illumination Techniques for Interactive Volume Rendering." In: *Computer Graphics Forum* 33.1 (Oct. 2013), pp. 27–51. ISSN: 1467-8659. DOI: 10.1111/cgf.12252.

[103] Journal of the American Medical Association. *Patient Information — jamanetwork.com*. [Accessed 03-May-2024]. 2023. URL: https://jamanetwork.com/collections/6258/patient-information.

[104] Mirela Kahrimanovic, Wouter M. Bergmann Tiest, and Astrid M.L. Kappers. "Seeing and feeling volumes: The influence of shape on volume perception." In: *Acta Psychologica* 134.3 (2010), pp. 385–390. ISSN: 0001-6918. DOI: https://doi.org/10.1016/j.actpsy.2010.03.011.

[105] Odysseas Katsaros et al. "Bicuspid Aortic Valve Disease: From Pathophysiology to Treatment." In: *Journal of Clinical Medicine* 13.17 (Aug. 2024), p. 4970. ISSN: 2077-0383. DOI: 10.3390/jcm13174970.

[106] Matthew Kearnes, Phil Macnaghten, and James Wilsdon. *Governing at the nanoscale: people, policies, and emerging technologies*. London, UK: Demos, 2006. URL: https://www.academia.edu/35232664/Governing_at_the_nanoscale_people_policies_and_emerging_technologies.

[107] Adnan Khan, Heather M Ross, Natalia Salinas Parra, Sarah L Chen, Kashyap Chauhan, Makala Wang, Brian Yan, John Magagna, Jake Beiriger, Yash Shah, et al. "Risk prevention and health promotion for non-alcoholic fatty liver diseases (NAFLD)." In: *Livers* 2.4 (2022), pp. 264–282.

[108] P.J. Kilner, G.Z.. Yang, R.H. Mohiaddin, D. N. Firmin, and D. B. Longmore. "Helical and Retrograde Secondary Flow Patterns in the Aortic Arch Studied by Three Directional Magnetic Resonance Velocity Mapping." In: *Circ.* 88 (1993), pp. 2235–2247. DOI: 10.1161/01.cir.88.5.2235.

[109] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. "Accessible Visualization: Design Space, Opportunities, and Challenges." In: *Computer Graphics Forum* 40.3 (June 2021), pp. 173–188. ISSN: 1467-8659. DOI: 10.1111/cgf.14298.

[110] Andy J. King. "Using Comics to Communicate About Health: An Introduction to the Symposium on Visual Narratives and Graphic Medicine." In: *Health Communication* 32.5 (Aug. 2016), pp. 523–524. ISSN: 1532-7027. DOI: 10.1080/10410236.2016.1211063. URL: http://dx.doi.org/10.1080/10410236.2016.1211063.

[111] Andy J. King and Allison J. Lazard. "Advancing Visual Health Communication Research to Improve Infodemic Response." In: *Health Communication* 35.14 (2020), pp. 1723–1728. DOI: 10.1080/10410236.2020.1838094.

[112] Tobias Klauk and Tilmann Köppe. "Telling vs. Showing." In: *The Living Handbook of Narratology*. Ed. by Peter Hühn, Jan Christoph Meister, John Pier, and Wolf Schmid. Retrieved December 11, 2024 from http://www.lhn.uni-hamburg.de/article/telling-vs-showing. Hamburg: Hamburg University, 2014.

[113] Anna Kleinau et al. "Is there a Tornado in Alex's Blood Flow? A Case Study for Narrative Medical Visualization." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. 2022. DOI: 10.2312/vcbm.20221183.

[114] Paul Klemm, Sylvia Glaßer, Kai Lawonn, Marko Rak, Henry Völzke, Katrin Hegenscheid, and Bernhard Preim. "Interactive visual analysis of lumbar back pain-what the lumbar spine tells about your life." In: *Proc. of IVAPP*. Vol. 2. 2015, pp. 85–92.

[115] Wolters Kluwer. *UpToDate: Industry-leading clinical decision support — wolterskluwer.com*. [Accessed 03-May-2024]. 2023. URL: https://www.wolterskluwer.com/en/solutions/uptodate.

[116] Benjamin Köhler, Matthias Grothoff, Matthias Gutberlet, and Bernhard Preim. "Bloodline: A system for the guided analysis of cardiac 4D PC-MRI data." In: *Computers & Graphics* 82 (Aug. 2019), pp. 32–43. ISSN: 0097-8493. DOI: 10.1016/j.cag.2019.05.004.

[117] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. "Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment." In: *Proc. of ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2019. DOI: 10.1145/3290605.3300576.

[118] Robert Kosara. "Presentation-Oriented Visualization Techniques." In: *IEEE Computer Graphics and Applications* 36.1 (Jan. 2016), pp. 80–85. ISSN: 0272-1716. DOI: 10.1109/mcg.2016.2.

[119] Robert Kosara. "An argument structure for data stories." In: *Proc. of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*. 2017, pp. 31–35.

[120] Hagung Kuntjara, Betha Almanfaluthi, et al. "Character design in games analysis of character design theory." In: *Journal of Games, Game Art, and Gamification* 2.2 (2017).

[121] Xingyu Lan, Yanqiu Wu, and Nan Cao. "Affective Visualization Design: Leveraging the Emotional Impact of Data." In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (Jan. 2024), pp. 1–11. ISSN: 2160-9306. DOI: 10.1109/tvcg.2023.3327385.

[122] Aisha Langford and Stacy Loeb. "Perceived Patient-Provider Communication Quality and Sociodemographic Factors Associated With Watching Health-Related Videos on YouTube: A Cross-Sectional Analysis." In: *Journal of Medical Internet Research* 21.5 (2019), e13512. DOI: 10.2196/13512.

[123] K. Lawonn, N.N. Smit, K. Bühler, and B. Preim. "A Survey on Multimodal Medical Data Visualization." In: *Computer Graphics Forum* 37.1 (Oct. 2017), pp. 413–438. ISSN: 1467-8659. DOI: 10.1111/cgf.13306.

[124] Kai Lawonn, Maria Luz, Bernhard Preim, and Christian Hansen. "Illustrative Visualization of Vascular Models for Static 2D Representations." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, pp. 399–406.

[125] Kai Lawonn, Ivan Viola, Bernhard Preim, and Tobias Isenberg. "A Survey of Surface-Based Illustrative Rendering for Visualization." In: *Computer Graphics Forum* 37.6 (2018), pp. 205–234. DOI: 10.1111/cgf.13322.

[126] Mackenzie Leake, Hijung Valentina Shin, Joy O. Kim, and Maneesh Agrawala. "Generating Audio-Visual Slideshows from Text Articles Using Word Concreteness." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. ACM, Apr. 2020, pp. 1–11. DOI: 10.1145/3313831.3376519.

[127] Benjamin Lee, Dave Brown, Bongshin Lee, Christophe Hurter, Steven Drucker, and Tim Dwyer. "Data Visceralization: Enabling Deeper Understanding of Data Using Virtual Reality." In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (Feb. 2021), pp. 1095–1105. ISSN: 2160-9306. DOI: 10.1109/tvcg.2020.3030435.

[128] Bongshin Lee, Eun Kyoung Choe, Petra Isenberg, Kim Marriott, and John Stasko. "Reaching Broader Audiences With Data Visualization." In: *IEEE Computer Graphics and Applications* 40.2 (Mar. 2020), pp. 82–90. ISSN: 1558-1756. URL: http://dx.doi.org/10.1109/MCG.2020.2968244.

[129] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. "More Than Telling a Story: Transforming Data into Visually Shared Stories." In: *IEEE Computer Graphics and Applications* 35.5 (2015), pp. 84–90. DOI: 10.1109/MCG.2015.99.

[130] Juhan Lee, Kea Turner, Zhigang Xie, Bashar Kadhim, and Young-Rock Hong. "Association Between Health Information–Seeking Behavior on YouTube and Physical Activity Among U.S. Adults: Results From Health Information Trends Survey 2020." In: *AJPM Focus* 1.2 (2022), p. 100035. DOI: `https://doi.org/10.1016/j.focus.2022.100035`.

[131] Newton Lee. *Encyclopedia of Computer Graphics and Games*. Springer International Publishing, 2024.

[132] E. Lee-Robbins and E. Adar. "Affective Learning Objectives for Communicative Visualizations." In: *IEEE Transactions on Visualization & Computer Graphics* 29.01 (2023), pp. 1–11. ISSN: 1941-0506. DOI: `10.1109/TVCG.2022.3209500`.

[133] D. J. Leiner. *SoSci Survey (Version 3.3.00)*. Hosted by OvGU at `https://befragungen.ovgu.de`. 2022. URL: `https://www.soscisurvey.de`.

[134] Leung Li and Winnie Yeo. "Value of quality of life analysis in liver cancer: A clinician's perspective." In: *World Journal of Hepatology* 9.20 (2017), p. 867. DOI: `10.4254/wjh.v9.i20.867`.

[135] Nan Li, Dominique Brossard, Dietram A. Scheufele, Paul H. Wilson, and Kathleen M. Rose. "Communicating data: interactive infographics, scientific data and credibility." In: *Journal of Science Communication* 17.02 (2018), A06. DOI: `10.22323/2.17020206`.

[136] Qian Li, Yang Yang, Cesar Reis, Tao Tao, Wanwei Li, Xiaogang Li, and John H. Zhang. "Cerebral Small Vessel Disease." In: *Cell Transplantation* 27.12 (Sept. 2018), pp. 1711–1722. ISSN: 1555-3892. DOI: `10.1177/0963689718795148`.

[137] Heather Richter Lipford, Felesia Stukes, Wenwen Dou, Matthew E. Hawkins, and Remco Chang. "Helping users recall their reasoning process." In: *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Oct. 2010, pp. 187–194. DOI: `10.1109/vast.2010.5653598`.

[138] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing." In: *ACM Computing Surveys* 55.9 (2023), pp. 1–35.

[139] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." In: *CoRR* (2019). URL: `http://arxiv.org/abs/1907.11692`.

[140] LONGD. *inkSketch v1.5*. 2024. URL: `https://civitai.com/models/84494/a-comic-sketches-style`.

[141] William E. Lorensen and Harvey E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm." In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. SIGGRAPH '87. ACM, Aug. 1987, pp. 163–169. DOI: 10.1145/37401.37422.

[142] Daniel Loureiro, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, and Jose Camacho-Collados. *TimeLMs: Diachronic Language Models from Twitter*. 2022. URL: 10.48550/arXiv.2202.03829.

[143] Sam Lowe. *roberta-base-go_emotions*. https://huggingface.co/SamLowe/roberta-base-go_emotions. 2023.

[144] Junhua Lu, Wei Chen, Hui Ye, Jie Wang, Honghui Mei, Yuhui Gu, Yingcai Wu, Xiaolong Luke Zhang, and Kwan-Liu Ma. "Automatic Generation of Unit Visualization-based Scrollytelling for Impromptu Data Facts Delivery." In: *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*. IEEE, Apr. 2021, pp. 21–30. DOI: 10.1109/pacificvis52677.2021.00011.

[145] Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N Kostis. "Scientific Storytelling Using Visualization." In: *IEEE Computer Graphics and Applications* 32.1 (Jan. 2012), pp. 12–19. ISSN: 0272-1716. DOI: 10.1109/mcg.2012.24.

[146] Philip Man. *Narrative Structures in Data Visualizations to Improve Storytelling*. Accessed: 2024-12-13. May 2011. URL: https://example.com/narrative-structures-data-visualizations.

[147] Ilaria Mariani, Mariana Ciancia, et al. "Character-driven Narrative Engine. Storytelling System for building interactive narrative experiences." In: *Proc. of the Digra Int. Conf.: Game, Play, and the Emerging Ludo-Mix*. Digital Games Research Association. 2019, pp. 1–19.

[148] SharylR Martini, StephenR Williams, Paolo Moretti, Daniel Woo, and BradfordB Worrall. "A molecular/genetic approach to cerebral small-vessel disease: Beyond aging and hypertension." In: *Brain Circulation* 1.1 (2015), p. 79. ISSN: 2394-8108. DOI: 10.4103/2394-8108.166376.

[149] L. E. Matzen, B. C. Howell, M. S. Trumbo, and K. M. Divis. "Numerical and Visual Representations of Uncertainty Lead to Different Patterns of Decision Making." In: *IEEE Computer Graphics and Applications* 43.05 (2023), pp. 72–82. ISSN: 1558-1756. DOI: 10.1109/MCG.2023.3299875.

[150] Benedikt Mayer, Karsten Donnay, Kai Lawonn, Bernhard Preim, and Monique Meuschke. "Expert exploranation for communicating scientific methods — A case study in conflict research." In: *Computers & Graphics* 120 (May 2024). ISSN: 0097-8493. DOI: 10.1016/j.cag.2024.103937.

[151] Julián Méndez, Weizhou Luo, Rufat Rzayev, Wolfgang Büschel, and Raimund Dachselt. "Immersive Data-Driven Storytelling: Scoping an Emerging Field Through the Lenses of Research, Journalism, and Games." In: *IEEE Transactions on Visualization and Computer Graphics* (2025), pp. 1–13. ISSN: 2160-9306. DOI: 10.1109/tvcg.2025.3531138.

[152] Monique Meuschke, Laura A. Garrison, Noeska N. Smit, Benjamin Bach, Sarah Mittenentzwei, Veronika Weiß, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. "Narrative medical visualization to communicate disease data." In: *Computers & Graphics* 107 (2022), pp. 144–157. DOI: 10.1016/j.cag.2022.07.017.

[153] Philip Meyer. "Defining and Measuring Credibility of Newspapers: Developing an Index." In: *Journalism Quarterly* 65.3 (1988), pp. 567–574. DOI: 10.1177/107769908806500301.

[154] Midjourney. *Midjourney*. Version 2023 (5.1). [Accessed 03-May-2024]. Feb. 1, 2024. URL: https://www.midjourney.com.

[155] S. Mittenentzwei, L. A. Garrison, B. Budich, K. Lawonn, A. Dockhorn, B. Preim, and M. Meuschke. "AI-based character generation for disease stories: a case study using epidemiological data to highlight preventable risk factors." In: *i-com* (2025). DOI: doi:10.1515/icom-2024-0041.

[156] S. Mittenentzwei, V. Weiß, S. Schreiber, L. A. Garrison, S. Bruckner, M. Pfister, B. Preim, and M. Meuschke. "Do Disease Stories Need a Hero? Effects of Human Protagonists on a Narrative Visualization about Cerebral Small Vessel Disease." In: *Computer Graphics Forum* 42.3 (2023), pp. 123–135. DOI: https://doi.org/10.1111/cgf.14817.

[157] Sarah Mittenentzwei, Laura A. Garrison, Eric Mörth, Kai Lawonn, Stefan Bruckner, Bernhard Preim, and Monique Meuschke. "Investigating user behavior in slideshows and scrollytelling as narrative genres in medical visualization." In: *Computers & Graphics* 114 (2023), pp. 229–238. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2023.06.011.

[158] Sarah Mittenentzwei, Sophie Mlitzke, Beatrice Budich, Anna Kleinau, Bernhard Preim, and Monique Meuschke. "Visual Disease Stories: Empowering Health Literacy and Promotion." In: *EuroVis 2025 - Dirk Bartz Prize*. Ed. by Torsten Kuhlen and Monique Meuschke. 🏆 Receiver of the Dirk Bartz Prize, to be published. The Eurographics Association, 2025.

[159] Sarah Mittenentzwei, Sophie Mlitzke, Darija Grisanova, Kai Lawonn, Bernhard Preim, and Monique Meuschke. "Visually communicating pathological changes: A case study on the effectiveness of Phong versus outline shading." In: *Computers & Graphics* 123 (2024), p. 104023. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2024.104023.

[160] Sarah Mittenentzwei, Sophie Mlitzke, Kai Lawonn, Bernhard Preim, and Monique Meuschke. "Communicating Pathologies and Growth to a General Audience." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2023. ISBN: 978-3-03868-216-5. DOI: 10.2312/vcbm.20231215.

[161] Sarah Mittenentzwei, Danish Murad, Bernhard Preim, and Monique Meuschke. "Leaving the Lab Setting: What We Can Learn About the Perceptionof Narrative Medical Visualizations from YouTube Comments." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. Receiver of the Best Short Paper Award. The Eurographics Association, 2024.

[162] Sarah Mittenentzwei, Bernhard Preim, and Monique Meuschke. "Why, What, and How to Communicate Health Information Visually:Reflections on the Design Process of Narrative Medical Visualization." In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2024.

[163] Sarah Mittenentzwei, Veronika Weiß, Stefanie Schreiber, Laura A. Garrison, Stefan Bruckner, Malte Pfister, Bernhard Preim, and Monique Meuschke. *Narrative Visualization to Communicate Neurological Diseases*. 2022. arXiv: 2212.10121 [cs.HC].

[164] Tobias Mönch, Simon Adler, and Bernhard Preim. "Staircase-Aware Smoothing of Medical Surface Meshes." In: *Proc. of Eurographics Workshop on Visual Computing for Biomedicine*. 2010, pp. 83–90. DOI: 10.2312/VCBM/VCBM10/083-090.

[165] Eric Mörth, Stefan Bruckner, and Noeska N. Smit. "ScrollyVis: Interactive Visual Authoring of Guided Dynamic Narratives for Scientific Scrollytelling." In: *IEEE Transactions on Visualization and Computer Graphics* 29.12 (Dec. 2023), pp. 5165–5177. ISSN: 2160-9306. DOI: 10.1109/tvcg.2022.3205769.

[166] Susanne C. Moser and Lisa Dilling. *Creating a Climate for Change: Communicating Climate Change and Facilitating Social Change*. Cambridge University Press, Feb. 2007. ISBN: 9780511535871. DOI: 10.1017/cbo9780511535871.

[167] Rohini Nadgir and David M Yousem. "Cerebral Aneurysms." In: *The Requisites Neuroradiology* (2017), pp. 141–145. DOI: 10.1002/9781118782934.

[168] Xiaoli Nan, Michael F. Dahlstrom, Adam Richards, and Sarani Rangarajan. "Influence of Evidence Type and Narrative Type on HPV Risk Perception and Intention to Obtain the HPV Vaccine." In: *Health Communication* 30.3 (2014), pp. 301–308. DOI: 10.1080/10410236.2014.888629.

[169] Xiaoli Nan, Michelle Futerfas, and Zexin Ma. "Role of Narrative Perspective and Modality in the Persuasiveness of Public Service Advertisements Promoting HPV Vaccination." In: *Health Communication* 32.3 (2016), pp. 320–328. DOI: 10.1080/10410236.2016.1138379.

[170] Xiaoli Nan, Irina A. Iles, Bo Yang, and Zexin Ma. "Public Health Messaging during the COVID-19 Pandemic and Beyond: Lessons from Communication Science." In: *Health Communication* 37.1 (2022), pp. 1–19. DOI: 10.1080/10410236.2021.1994910.

[171] NASA Science. *Eyes on the Earth - NASA Science*. https://science.nasa.gov/eyes/. Accessed: 2024-06-17. 2024.

[172] NASA Scientific Visualization Studio. *NASA Scientific Visualization Studio*. https://svs.gsfc.nasa.gov/. Accessed: 2024-06-17. 2024.

[173] *National Health Service*. https://www.nhs.uk/. Accessed: 2025-02-26. 2025.

[174] Liz Neeley, Erin Barker, Skylar R Bayer, Reyhaneh Maktoufi, Katherine J Wu, and Maryam Zaringhalam. "Linking scholarship and practice: narrative and identity in science." In: *Frontiers in Communication* 5 (2020), p. 35.

[175] Natalia Nehring, Nilufar Baghaei, and Simon Dacey. "Improving Students' Performace Through Gamification: A User Study." In: *CSEDU (1)*. 2018, pp. 213–218.

[176] Seth M Noar, Jacob A Rohde, Joshua O Barker, Marissa G Hall, and Noel T Brewer. "Pictorial Cigarette Pack Warnings Increase Some Risk Appraisals But Not Risk Beliefs: A Meta-Analysis." In: *Human Communication Research* 46.2-3 (2020), pp. 250–272. ISSN: 0360-3989. DOI: 10.1093/hcr/hqz016.

[177] P.E. Norman and J.T. Powell. "Site Specificity of Aneurysmal Disease." In: *Circulation* 121.4 (2010), pp. 560–568. ISSN: 0009-7322. DOI: 10.1161/circulationaha.109.880724.

[178] Heather L. O'Brien and Elaine G. Toms. "The development and evaluation of a survey to measure user engagement." In: *Journal of the American Society for Information Science and Technology* 61.1 (2009), pp. 50–69. DOI: 10.1002/asi.21229.

[179] Daniel J. O'Keefe. "The Relative Persuasiveness of Different Forms of Arguments-From-Consequences: A Review and Integration." In: *Annals of the International Communication Association* 36.1 (2013), pp. 109–135. DOI: 10.1080/23808985.2013.11679128.

[180] Harri Oinas-Kukkonen and Marja Harjumaa. "Persuasive Systems Design: Key Issues, Process Model, and System Features." In: *Communications of the Association for Information Systems* 24 (2009). ISSN: 1529-3181. DOI: 10.17705/1cais.02428.

[181] Yotam Ophir, Emily Brennan, Erin K. Maloney, and Joseph N. Cappella. "The Effects of Graphic Warning Labels' Vividness on Message Engagement and Intentions to Quit Smoking." In: *Communication Research* 46.5 (2019), pp. 619–638. DOI: 10.1177/0093650217700226.

[182] Lyle J Palmer. "UK Biobank: bank on it." In: *The Lancet* 369.9578 (June 2007), pp. 1980–1982. DOI: 10.1016/s0140-6736(07)60924-6.

[183] Saugat Pandey and Alvitta Ottley. "Mini-VLAT: A Short and Effective Measure of Visualization Literacy." In: *Computer Graphics Forum* 42.3 (June 2023), pp. 1–11. ISSN: 1467-8659. DOI: 10.1111/cgf.14809.

[184] Bo Pang and Lillian Lee. *Opinion mining and sentiment analysis*. Vol. 2. Foundations and Trends® in information retrieval. now Publishers, Inc., 2008, pp. 1–135. ISBN: 978-1-60198-150-9.

[185] Luca Papariello. *xlm-roberta-base-language-detection*. https://huggingface.co/papluca/xlm-roberta-base-language-detection. 2023.

[186] Daniel Y. Park and Elizabeth M. Goering. "The Health-Related Uses and Gratifications of YouTube: Motive, Cognitive Involvement, Online Activity, and Sense of Empowerment." In: *Journal of Consumer Health on the Internet* 20.1-2 (2016), pp. 52–70. DOI: 10.1080/15398285.2016.1167580.

[187] Bo Peng, Hongxing Fan, Wei Wang, Jing Dong, and Siwei Lyu. "A Unified Framework for High Fidelity Face Swap and Expression Reenactment." In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.6 (June 2022), pp. 3673–3684. DOI: 10.1109/tcsvt.2021.3106047.

[188] Juhana Pettersson. "Basics of character design." In: *Larp design: Creating role-play experiences* (2019), pp. 193–204.

[189] Federico Pianzola. "Looking at Narrative as a Complex System: The Proteus Principle." In: *Narrating Complexity*. Springer International Publishing, 2018, pp. 101–122. ISBN: 9783319647142. DOI: 10.1007/978-3-319-64714-2_10.

[190] Ronald J. Planer and Peter Godfrey-Smith. "Communication and representation understood as sender–receiver coordination." In: *Mind & Language* 36.5 (June 2020), pp. 750–770. ISSN: 1468-0017. DOI: 10.1111/mila.12293.

[191] Annamaria Porreca, Francesca Scozzari, and Marta Di Nicola. "Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination." In: *BMC Public Health* 20.1 (2020). DOI: 10.1186/s12889-020-8342-4.

[192] Bernhard Preim, Alexandra Baer, Douglas Cunningham, Tobias Isenberg, and Timo Ropinski. "A Survey of Perceptually Motivated 3D Visualization of Medical Image Data." In: *Computer Graphics Forum* 35.3 (2016), pp. 501–525. DOI: https://doi.org/10.1111/cgf.12927.

[193] Bernhard Preim and Monique Meuschke. "A survey of medical animations." In: *Computers & Graphics* 90 (2020), pp. 145–168. ISSN: 0097-8493. DOI: 10.1016/j.cag.2020.06.003.

[194] Bernhard Preim and Monique Meuschke. "Narrative Visualisierung in der Medizin." de. In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft* (2023). DOI: 10.25969/MEDIAREP/22346.

[195] Bernhard Preim, Monique Meuschke, and Veronika Weiß. "A Survey of Medical Visualization Through the Lens of Metaphors." In: *IEEE Transactions on Visualization and Computer Graphics* 30 (2023), pp. 6639–6664. URL: https://api.semanticscholar.org/CorpusID:265049010.

[196] Luigi Preziosi, Giuseppe Toscani, and Mattia Zanella. "Control of tumor growth distributions through kinetic methods." In: *Journal of Theoretical Biology* 514 (2021), p. 110579. ISSN: 0022-5193. DOI: https://doi.org/10.1016/j.jtbi.2021.110579.

[197] Gerald Prince. *Narrative as Theme: Studies in French Fiction*. University of Nebraska Press, 1992. ISBN: 978-0803236998.

[198] Rebecca M. Puhl and Chelsea A. Heuer. "Obesity Stigma: Important Considerations for Public Health." In: *American Journal of Public Health* 100.6 (June 2010), pp. 1019–1028. DOI: 10.2105/ajph.2009.159491.

[199] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. *Improving language understanding by generative pre-training*. 2018. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

[200] Ramesh Raskar and Michael Cohen. "Image Precision Silhouette Edges." In: *Proc. of Interactive 3D Graphics*. 1999, pp. 135–140. DOI: 10.1145/300523.300539.

[201] Peter Rautek, Stefan Bruckner, and Eduard Groller. "Semantic layers for illustrative volume rendering." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1336–1343.

[202] RealtimeBoard, Inc. *Miro*. Version CC 2024 (0.7.37). Accessed: 2025-04-04. URL: https://miro.com/.

[203] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Hollerer, and Eun Kyoung Choe. "ChartAccent: Annotation for data-driven storytelling." In: *2017 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, Apr. 2017, pp. 230–239. DOI: 10.1109/pacificvis.2017.8031599.

[204] N.H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale. *Data-Driven Storytelling*. A K Peters/CRC Press, Mar. 2018. ISBN: 9781315281575. DOI: 10.1201/9781315281575.

[205] Marie-Laure Ryan. "Toward a definition of narrative." In: *The Cambridge Companion to Narrative*. Cambridge University Press, July 2007, pp. 22–36. ISBN: 9781139001533. DOI: 10.1017/ccol0521856965.002.

[206] Savage Interactive Pty Ltd. *Procreate*. Version CC 2023 (5.3.5). [Accessed 03-May-2024]. June 11, 2023. URL: https://procreate.com/ipad.

[207] Klaudius Scheufele, Shashank Subramanian, and George Biros. "Fully Automatic Calibration of Tumor-Growth Models Using a Single mpMRI Scan." In: *IEEE Transactions on Medical Imaging* 40.1 (2021), pp. 193–204. DOI: 10.1109/TMI.2020.3024264.

[208] Ralf Schmälzle, Britta Renner, and Harald T. Schupp. "Health Risk Perception and Risk Communication." In: *Policy Insights from the Behavioral and Brain Sciences* 4.2 (2017), pp. 163–169. DOI: 10.1177/2372732217720223.

[209] Christian Schumann, Steffen Oeltze, Ragnar Bade, Bernhard Preim, and Heinz-Otto Peitgen. "Model-free Surface Visualization of Vascular Trees." In: *Proc. of EuroVis*. 2007, pp. 283–290. DOI: 10.2312/VisSym/EuroVis07/283-290.

[210] E Segel and J Heer. "Narrative Visualization: Telling Stories with Data." In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (Nov. 2010), pp. 1139–1148. ISSN: 1077-2626. DOI: 10.1109/tvcg.2010.179.

[211] Doris Seyser and Michael Zeiller. "Scrollytelling – An Analysis of Visual Storytelling in Online Journalism." In: *2018 22nd International Conference Information Visualisation (IV)*. 2018, pp. 401–406. DOI: 10.1109/iV.2018.00075.

[212] Daniel M Sforza, Kenichi Kono, Satoshi Tateshima, Fernando Viñuela, Christopher Putman, and Juan R Cebral. "Hemodynamics in growing and stable cerebral aneurysms." In: *Journal of NeuroInterventional Surgery* 8.4 (Apr. 2016), 407 LP –412. DOI: 10.1136/neurintsurg-2014-011339.

[213] Rajesh Sharma. "Descriptive epidemiology of incidence and mortality of primary liver cancer in 185 countries: evidence from GLOBOCAN 2018." In: *Japanese Journal of Clinical Oncology* 50.12 (July 2020), pp. 1370–1379. ISSN: 1465-3621. DOI: 10.1093/jjco/hyaa130.

[214] Lee Sheldon. *Character development and storytelling for games*. CRC Press, 2022.

[215] Alexander Shevtsov, Maria Oikonomidou, Despoina Antonakaki, Polyvios Pratikakis, and Sotiris Ioannidis. "What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020." In: *PLOS ONE* 18.1 (2023), e0270542. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0270542.

[216] Yang Shi, Xingyu Lan, Jingwen Li, Zhaorui Li, and Nan Cao. "Communicating with Motion: A Design Space for Animated Visual Narratives in Data Videos." In: *Proc. of ACM Conference on Human Factors in Computing Systems*. CHI '21. 2021. DOI: 10.1145/3411764.3445337.

[217] Anna Shilo and Renata G. Raidou. "Visual narratives to edutain against misleading visualizations in healthcare." In: *Computers & Graphics* 123 (Oct. 2024), p. 104011. ISSN: 0097-8493. DOI: 10.1016/j.cag.2024.104011.

[218] Matúš Šimkovic and Birgit Träuble. "Robustness of statistical methods when measure is affected by ceiling and/or floor effect." In: *PLOS ONE* 14.8 (Aug. 2019). Ed. by Alan D Hutson, e0220889. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0220889.

[219] Manvir Singh. "The Sympathetic Plot, Its Psychological Origins, and Implications for the Evolution of Fiction." In: *Emotion Review* 13.3 (2021), pp. 183–198. DOI: 10.1177/17540739211022824.

[220] Ethan Smith, JJ Fiasson, and Jachin Bhasme. *Leonardo.Ai*. [Accessed 03-May-2024]. Feb. 1, 2024. URL: https://leonardo.ai.

[221] Snap Inc. *Snapchat*. Version CC 2024. Jan. 27, 2025. URL: https://www.snapchat.com.

[222] Wonyoung So, Edyta P Bogucka, Sanja Šćepanović, Sagar Joglekar, Ke Zhou, and Daniele Quercia. "Humane visual AI: Telling the stories behind a medical condition." In: *IEEE Trans. Vis. Comput. Graph.* 27.2 (2020), pp. 678–688.

[223] Wonyoung So, Edyta P. Bogucka, Sanja Šćepanović, Sagar Joglekar, Ke Zhou, and Daniele Quercia. "Humane Visual AI: Telling the Stories Behind a Medical Condition." In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021), pp. 678–688. DOI: 10.1109/TVCG.2020.3030391.

[224] Lav Soni and Amanpreet Kaur. "Merits and Demerits of Unreal and Unity: A Comprehensive Comparison." In: *2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST)*. 2024, pp. 1–5. DOI: 10.1109/ICCIGST60741.2024.10717602.

[225] Kristine Sørensen, Stephan Van den Broucke, Jürgen M Pelikan, James Fullam, Gerardine Doyle, Zofia Slonska, Barbara Kondilis, Vivian Stoffels, Richard H Osborne, and Helmut Brand. "Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q)." In: *BMC Public Health* 13.1 (Oct. 2013). ISSN: 1471-2458. DOI: 10.1186/1471-2458-13-948.

[226] Kristine Sørensen et al. "Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU)." In: *The European Journal of Public Health* 25.6 (Apr. 2015), pp. 1053–1058. ISSN: 1464-360X. DOI: 10.1093/eurpub/ckv043.

[227] Stability AI. *Stable Diffusion*. [Accessed 03-May-2024]. 2024. URL: https://stability.ai/stable-image (visited on 02/01/2024).

[228] Katarzyna Stapor. "Descriptive and Inferential Statistics." In: *Introduction to Probabilistic and Statistical Methods with Examples in R*. Springer International Publishing, 2020, pp. 63–131. ISBN: 9783030457990. DOI: 10.1007/978-3-030-45799-0_2.

[229] Ian Steenstra, Prasanth Murali, Rebecca B. Perkins, Natalie Joseph, Michael K Paasche-Orlow, and Timothy Bickmore. "Engaging and Entertaining Adolescents in Health Education Using LLM-Generated Fantasy Narrative Games and Virtual Agents." In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI '24. ACM, May 2024, pp. 1–8. DOI: 10.1145/3613905.3650983.

[230] Charles D. Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. *Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories*. Tech. rep. MSR-TR-2016-14. Microsoft Research, Apr. 2016. URL: https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/.

[231] Sherin Sugathan, Hauke Bartsch, Frank Riemer, Renate Grüner, Kai Lawonn, and Noeska Smit. "Longitudinal visualization for exploratory analysis of multiple sclerosis lesions." In: *Computers & Graphics* 107 (2022), pp. 208–219. DOI: https://doi.org/10.1016/j.cag.2022.07.023.

[232] Nicole Sultanum, Fanny Chevalier, Zoya Bylinskii, and Zhicheng Liu. "Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with VizFlow." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. ACM, May 2021. DOI: 10.1145/3411764.3445354.

[233] Nicole Sultanum and Arjun Srinivasan. "DATATALES: Investigating the use of Large Language Models for Authoring Data-Driven Articles." In: *2023 IEEE Visualization and Visual Analytics (VIS)*. IEEE, Oct. 2023. DOI: 10.1109/vis54172.2023.00055.

[234] Tableau. *Tableau: Visual Analytics for Everyone*. Accessed: 2024-12-18. 2024. URL: https://www.tableau.com.

[235] Unity Technologies. *Unity 3D*. Accessed: 2025-03-03. San Francisco, U.S., 2025. URL: https://unity.com/.

[236] *The New York Times*. Accessed: 2025-04-04. The New York Times Company. 2025. URL: https://www.nytimes.com.

[237] Kjerstin Thorson, Emily Vraga, and Brian Ekdale. "Credibility in Context: How Uncivil Online Commentary Affects News Credibility." In: *Mass Communication and Society* 13.3 (2010), pp. 289–313. DOI: 10.1080/15205430903225571.

[238] Christian Tietjen, Tobias Isenberg, and Bernhard Preim. "Combining Silhouettes, Surface, and Volume Rendering for Surgery Education and Planning." In: *Proc. of EuroVis*. 2005, pp. 303–310. DOI: 10.2312/VisSym/EuroVis05/303-310.

[239] New York Times. *Snowfall*. Accessed: 2025-04-04. 2012. URL: https://www.nytimes.com/projects/2012/snow-fall/index.html#/?part=tunnel-creek.

[240] Chao Tong et al. "Storytelling and Visualization: An Extended Survey." In: *Information* 9.3 (Mar. 2018), p. 65. ISSN: 2078-2489. DOI: 10.3390/info9030065.

[241] Connie W. Tsao et al. "Heart disease and stroke statistics—2022 update: A report from the American Heart Association." In: *Circ.* 145.8 (2022), e153–e639. DOI: 10.1161/cir.0000000000001074.

[242] Jason Turcotte, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree. "News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking." In: *Journal of Computer-Mediated Communication* 20.5 (2015), pp. 520–535. DOI: 10.1111/jcc4.12127.

[243] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is All you Need." In: *Proc. of Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.

[244] Ivan Viola and Tobias Isenberg. "Pondering the Concept of Abstraction in (Illustrative) Visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 24.9 (2018), pp. 2573–2588. DOI: 10.1109/TVCG.2017.2747545.

[245] Visualiseringscenter C. *Visualiseringscenter C*. https://visualiseringscenter.se/en/. Accessed: 2024-06-17. 2024.

[246] Gisela Wachinger, Ortwin Renn, Chloe Begg, and Christian Kuhlicke. "The Risk Perception Paradox-Implications for Governance and Communication of Natural Hazards." In: *Risk Analysis* 33.6 (2012), pp. 1049–1065. DOI: 10.1111/j.1539-6924.2012.01942.x.

[247] Shelley Wall. "Medical Graphics and Graphic Medicine." In: *Biomedical Visualization*. Springer Nature Switzerland, 2023, pp. 23–40. DOI: 10.1007/978-3-031-39035-7_2.

[248] Jinrui Wang, Xinhuan Shu, Benjamin Bach, and Uta Hinrichs. "Visualization Atlases: Explaining and Exploring Complex Topics Through Data, Visualization, and Narration." In: *IEEE Transactions on Visualization and Computer Graphics* 31.1 (2025), pp. 437–447. DOI: 10.1109/TVCG.2024.3456311.

[249] Yanli Wang, Feng Wang, Yun Cheng, Chengling Zhao, and Zhongmei Zheng. "The study of characters design in PRG educational games." In: *IEEE International Conference on Computer Science and Information Technology*. 2009, pp. 44–47.

[250] Jeremy L Warner et al. “Development, implementation, and initial evaluation of a foundational open interoperability standard for oncology treatment planning and summarization.” In: *Journal of the American Medical Informatics Association* 22.3 (2015), pp. 577–586. DOI: 10.1093/jamia/ocu015.

[251] Wibke Weber. “18. Exploring narrativity in data visualization in journalism.” In: *Data Visualization in Society*. Amsterdam University Press, Dec. 2020, pp. 295–312. DOI: 10.1515/9789048543137-022.

[252] Mingqiang Wei, Jun Wang, Xianglin Guo, Huisi Wu, Haoran Xie, Fu Lee Wang, and Jing Qin. “Learning-based 3D surface optimization from medical image reconstruction.” In: *Optics and Lasers in Engineering* 103 (2018), pp. 110–118.

[253] Mingqiang Wei, Lei Zhu, Jinze Yu, Jun Wang, Wai-Man Pang, Jianhuang Wu, Jing Qin, and Pheng-Ann Heng. “Morphology-preserving smoothing on polygonized isosurfaces of inhomogeneous binary volumes.” In: *Computer-Aided Design* 58 (2015), pp. 92–98.

[254] Zheng Wei, Huamin Qu, and Xian Xu. “Telling Data Stories with the Hero’s Journey: Design Guidance for Creating Data Videos.” In: *IEEE Transactions on Visualization and Computer Graphics* 31.1 (Jan. 2024), pp. 962–972. ISSN: 2160-9306. DOI: 10.1109/tvcg.2024.3456330.

[255] Barry D. Weiss. *Health Literacy and Patient Safety: Help Patients Understand. Manual for Clinicians*. 2nd. Chicago, IL: American Medical Association Foundation, 2007.

[256] Dustin J. Welbourne and Will J. Grant. “Science communication on YouTube: Factors that affect channel and video popularity.” In: *Public Understanding of Science* 25.6 (2016), pp. 706–718. DOI: 10.1177/0963662515572068.

[257] *WHO Europe | Public health services*. https://www.euro.who.int/en/health-topics/Health-systems/public-health-services/public-health-services. Accessed: 2022-03-22. 2022.

[258] H.-E. Wichmann, R. Kaaks, W. Hoffmann, K.-H. Jöckel, K.H. Greiser, and J. Linseisen. “Die Nationale Kohorte.” In: *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 55.6–7 (June 2012), pp. 781–789. DOI: 10.1007/s00103-012-1499-y.

[259] Gloria Willcox. “The Feeling Wheel: A Tool for Expanding Awareness of Emotions and Increasing Spontaneity and Intimacy.” In: *Transactional Analysis Journal* 12.4 (Oct. 1982), pp. 274–276.

[260] C.-E. A. Winslow. “The Untilled Fields of Public Health.” In: *Science* 51.1306 (1920), pp. 23–33. DOI: 10.1126/science.51.1306.23.

[261] Thomas Winters and Kory W. Mathewson. "Automatically Generating Engaging Presentation Slide Decks." In: *Computational Intelligence in Music, Sound, Art and Design*. Springer International Publishing, 2019, pp. 127–141. ISBN: 978-3-03016-667-0. DOI: 10.1007/978-3-030-16667-0_9.

[262] Wissenschaft im Dialog gGmbH. *Impact Unit*. [Accessed 30-03-2025]. Charlottenstrasse 80, D-10117 Berlin, 2025. URL: https://www.impact-unit.de.

[263] Michael Wohlfart. "Story Telling Aspects in Medical Applications." In: *Central European Seminar for Computer Graphics*. 2006. URL: https://api.semanticscholar.org/CorpusID:15295683.

[264] Michael Wohlfart and Helwig Hauser. "Story telling for presentation in volume visualization." In: *Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization*. 2007, pp. 91–98.

[265] World Health Organization. *Health Promotion Glossary of Terms 2021*. Geneva: World Health Organization, 2021. ISBN: 978-92-4-003834-9. URL: https://www.who.int/publications/i/item/9789240038349.

[266] World Health Organization. *Global Excess Deaths Associated with COVID-19 (January 2020–December 2021)*. Accessed: 2024-12-18. 2022. URL: https://www.who.int/data/stories/global-excess-deaths-associated-with-covid-19-january-2020-december-2021/.

[267] World Health Organization. *Data Stories*. Accessed: 2024-12-18. 2024. URL: https://www.who.int/data/stories.

[268] Ann O. Amuta Wura Jacobs and Kwon Chan Jeon. "Health information seeking in the digital age: An analysis of health information seeking behavior among US adults." In: *Cogent Social Sciences* 3.1 (2017), p. 1302785. DOI: 10.1080/23311886.2017.1302785.

[269] Leni Yang, Aoyu Wu, Wai Tong, Xian Xu, Zheng Wei, and Huamin Qu. "Understanding 3D Data Videos: From Screens to Virtual Reality." In: *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. IEEE, Apr. 2023, pp. 197–206. DOI: 10.1109/pacificvis56936.2023.00029.

[270] Leni Yang, Xian Xu, XingYu Lan, Ziyan Liu, Shunan Guo, Yang Shi, Huamin Qu, and Nan Cao. "A Design Space for Applying the Freytag's Pyramid Structure to Data Stories." In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (Jan. 2022), pp. 922–932. ISSN: 2160-9306. DOI: 10.1109/tvcg.2021.3114774.

[271] Shiyu Yang, Dominique Brossard, Dietram A. Scheufele, and Michael A. Xenos. "The science of YouTube: What factors influence user engagement with online science videos?" In: *PLOS ONE* 17.5 (2022), e0267697. DOI: 10.1371/journal.pone.0267697.

[272] Vyri Yang and Mahmood Jasim. "Animating the Narrative: A Review of Animation Styles in Narrative Visualization." In: *2024 IEEE Visualization and Visual Analytics (VIS)*. IEEE, Oct. 2024, pp. 321–325. DOI: 10.1109/vis55277.2024.00074.

[273] Anders Ynnerman, Jonas Lowgren, and Lena Tibell. "Exploranation: A New Science Communication Paradigm." In: *IEEE Computer Graphics and Applications* 38.3 (May 2018), pp. 13–20. ISSN: 1558-1756. DOI: 10.1109/mcg.2018.032421649.

[274] YouTube. *YouTube Researcher Program*. Accessed: April 6, 2025. 2025. URL: https://research.youtube/.

[275] *ZEIT Online*. Accessed: 2025-04-04. Zeit Online GmbH. 2025. URL: https://www.zeit.de.

[276] Yixuan Zhang, Kartik Chanana, and Cody Dunne. "IDMVis: Temporal Event Sequence Visualization for Type 1 Diabetes Treatment Decision Support." In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 512–522. DOI: 10.1109/TVCG.2018.2865076.

[277] Jian Zhao, R William Soukoreff, Xiangshi Ren, and Ravin Balakrishnan. "A model of scrolling on touch-sensitive displays." In: *Int. J. Hum.-Comput. Stud.* 72.12 (2014), pp. 805–821. DOI: 10.1016/j.ijhcs.2014.07.003.

[278] Zovya. *ReV Animated v1.2.2*. [Accessed 03-May-2024]. 2024. URL: https://civitai.com/models/7371/rev-animated.

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI artificial intelligence

BAV bicuspid aortic valve

CAA cerebral amyloid angiopathy

CFE ceiling and floor effect

CMB cerebral microbleeds

CSVD cerebral small vessel disease

CVD Cardiovascular disease

GenAI generative artificial intelligence

HA hypertensive arteriopathy

HLS-EU-Q European Health Literacy Survey Questionnaire

LDM latent diffusion model

LLM large language model

LoRA Low-Rank Adaptation

ML machine learning

NAFLD Non-alcoholic fatty liver disease

NLP natural language processing

PLC primary liver cancer

POV point of view

PSD Persuasive systems design

SHIP Study of Health in Pomerania

TAVI transcatheter aortic valve implantation

VR Virtual reality

WMH white matter hyperintensities

DISCLOSURE OF AI USAGE

The thesis was created using a variety of digital tools to improve its quality. The following sections describe these tools and their specific applications in detail. Generative AI was employed to enhance linguistic clarity and stylistic expression, surpassing the capabilities of conventional grammar and spelling checkers.

LTEX+

LTEX+ v15.5.0 is used as spell and grammar checker throughout the whole thesis. It is based on *LanguageTool*, an AI-based spelling and grammar checker, for LaTeXfiles in Visual Studio Code.

CHATGPT

ChatGPT models *o4* and *o4mini* were used to refine language and expression throughout all chapters of the thesis. This was achieved by submitting text excerpts from the thesis alongside prompts such as: *"Provide improved clarity, grammar, readability, and expression"* and similar instructions.

DEEPL WRITE

DeepL was used to further improve grammar and style through its tool *DeepL Write* throughout all chapters of the thesis. Text excerpts from the thesis were processed using the *"simple"* language style option, which provided suggestions to enhance clarity and readability.