DEVELOPING A HYBRID VECTOR-GRAPH RETRIEVAL SYSTEM FOR ENTITY-PRESERVING AND INSPIRING STORYLINE CREATION OF PRESENTATION SLIDES

Completed Research Paper

Alexander Meier, University of St. Gallen, St Gallen, Switzerland, alexander.meier@unisg.ch
Roman Rietsche, Bern University of Applied Sciences, Bern, Switzerland, roman.rietsche@bfh.ch
Mahei Manhai Li , University of St. Gallen, St Gallen, Switzerland, mahei.li@unisg.ch

Abstract

Effective presentation slide creation is crucial for impactful communication, yet fully automating this task with AI is insufficient. Hybrid human-AI solutions often perform worse than pure AI or human creation due to overreliance on AI. To address this, we develop design principles for configuring human-AI hybrid systems in complex knowledge tasks using a design science research approach. Our prototype, NarrativeNet Weaver, leverages an underutilized corpus of existing presentation slides, applying generative AI advances in hybrid dense embedding and graph-based retrieval techniques. Evaluated through 15 think-aloud sessions and 73 user trials, users with NarrativeNet Weaver exhibit greater engagement and achieve equal or improved slide quality compared to those using a ChatGPT-based chatbot with a vector database. We contribute design knowledge for human-AI systems for complex multimodal content and offer a new approach to retrieving and visualizing existing slides, enhancing the utilization of valuable but underused resources.

Keywords: Overreliance, Presentation Slide Creation, Graph-based Retrieval, Narrative Structuring

1 Introduction

Effective presentation slide creation is crucial for impactful communication, yet fully automating this task with artificial intelligence (AI) yields insufficient results. Despite the advancements of AI tools for fully automated slide generation, they often oversimplify complex content, produce unreliable outputs, or fail to engage slide creators critically, thereby diminishing presentation effectiveness (Guo et al., 2023; Zhai et al., 2024). While hybrid human-AI solutions may appear to offer a promising alternative, they frequently underperform compared to pure AI or human creation due to overreliance on AI, where content creators place undue trust in AI outputs even when unwarranted (Buçinca et al., 2021).

Consultants often reuse existing slides to save time and maintain consistency. However, current methods for accessing and retrieving relevant slide content from large databases are inefficient and fail to support the construction of coherent narratives from disparate slides. The challenge lies not merely in finding seemingly relevant slides but in integrating them into a structured, engaging presentation narrative. As one interviewee noted[1], "You start directly in PowerPoint and realize later that you do not have a proper story, which leads to logical jumps and disjointed diagrams."

To address these challenges, we pose the following research question (RQ):

RQ: *What design principles should guide the development and instantiation of a hybrid vector-graph retrieval system, and how does it perform in enhancing presentation slide creation in consulting contexts compared to conventional vector-database chatbot approaches?*

[1] PS4; we give detailed insights into the conducted interviews in section 5.

To address this RQ the remainder of the paper is structured as follows. We adopt a design science research approach to generate design knowledge and construct artifacts that resolve real world challenges through iterative development and rigorous assessment (Peffers et al., 2007). We present the NarrativeNet Weaver, a novel human AI slide retrieval system that leverages an underutilized corpus of existing presentation slides. The system applies recent advances in generative AI for dense embedding and graph based retrieval to enhance the access, retrieval, and visualization of slides, thereby facilitating the construction of coherent narratives. We summarize our main steps as artifact conceptualization, iterative refinement, and comprehensive evaluation. We evaluate the NarrativeNet Weaver through qualitative think aloud sessions with fifteen participants and quantitative user trials with seventy three participants. Our results reveal that users of the NarrativeNet Weaver engage more deeply with content and produce slide creation quality that equals or exceeds that of a ChatGPT based chatbot using a vector database. In doing so, we contribute design knowledge for human AI systems that create complex multimodal content, offer a new approach to retrieving and visualizing existing slides, and advance the utilization of valuable yet underused resources in consulting contexts.

2 Background

2.1 Presentation Slides and Narratives

Presentation slides are essential tools in the consulting industry for crafting convincing storylines and arguments in both C-level briefings and internal workshops. They organize and display information effectively while communicating complex ideas persuasively. Slides vary in design and structure based on their intended purpose, reflecting the required depth of argumentation (Bourgoin & Muniesa, 2016; Kim et al., 2017). Well-crafted slides showcase consultants' professional expertise, reinforcing essential discursive practices such as collaboration and visual storytelling (Kaplan, 2011).

Kaplan's analysis (2011) reveals two critical discursive practices in presentation slides: a collaborative effort representing knowledge development and meaning negotiation and a strategic visual effort to assert and disseminate interests. Together, these practices imbue slides with characteristics of materiality, mutability, modularity, and digitality, profoundly shaping organizational communication. The modular structure of slides facilitates the integration of disparate content into coherent presentations, aligning diverse interests.

The creation of presentation slides in the consulting industry often necessitates a structured approach (Bourgoin & Muniesa, 2016). While specific methodologies may vary, a common framework for creating presentation slides emphasizes several key steps (Bourgoin & Muniesa, 2016; Fu et al., 2022).

The initial step is *(1)* the *development of a narrative*, which begins with constructing a coherent and compelling storyline that resonates with the intended audience. Subsequently, the *(2) content identification* phase ensues, wherein the essential sections and supporting elements are identified, ensuring the effective conveyance of the vital message. *(3) A structural organization* establishes a logical flow by outlining each section's order, hierarchy, and length. Subsequently, a *(4) content summary* is produced, condensing information into a concise and impactful format, typically in bullet point or summary form. Lastly, a *(5) visual enhancement* is incorporated through the use of relevant visuals, such as charts, graphs, and images, to enhance clarity and engagement (Bourgoin & Muniesa, 2016).

With the NarrativeNet Weaver, we primarily aim to streamline steps (1) and (2), helping users develop coherent narratives and identify essential content while also providing inspirational support for visual enhancements in step (5).

2.2 AI-based Presentation Slide Creation and Overreliance

Before the integration of AI, slide creation predominantly relied on individual cognitive efforts and social collaborations (Alavi & Leidner, 2001). However, the emergence of AI has catalyzed a paradigm shift in knowledge creation, particularly in the consulting environment (B. Chen et al., 2023). Previous AI-based slide creation research has focused on multimodal summarization that integrates text and visuals into concise, structured outputs and on utilizing neural networks to synthesize syntactic, semantic, and contextual sentence features (Fu et al., 2022; Sefid et al., 2019).

Advancements in foundation model architectures, including transformers and diffusion models, have fueled the current transformation by enabling the generation of modalities such as text, images, and audio that closely resemble human-generated content (Han et al., 2024). GenAI models offer a range of functionalities, including text-to-text, text-to-image, image-to-image, and audio-related capabilities (Han et al., 2024). In slide generation, the most common technique involves the automatic creation of bullet points inserted into contextually appropriate templates (Fu et al., 2022; Sefid et al., 2019). Tools like Beautiful.ai and Microsoft's Office 365 Copilot facilitate not only slide creation but also the development of documents and emails. Despite sourcing data broadly from platforms like Common Crawl, these foundational models often lack the specialized knowledge of dedicated experts, impacting the quality and depth of generated content (Huan & Zhou, 2024). This deficiency can lead to oversimplified visual representations that do not adequately convey complex messages. Moreover, the results frequently comprise oversimplified bullet points on unpopulated slides, which may not adequately represent complex content and may simplify it excessively (Tufte, 2003). Furthermore, in human collaboration with AI, scenarios emerge where users trust AI when they should not, a phenomenon known as overreliance on AI (Buçinca et al., 2021). Overreliance can lead users to perform worse on tasks compared to the performance of the user or AI working alone (Bansal et al., 2021). This issue becomes especially acute when users work with new content; users expect AI to maintain its performance on new content but assume its own performance will worsen (Chiang & Yin, 2021). Consequently, users rely more on AI when dealing with unseen content, leading them to trust AI more when its performance is questionable and uncertain. Research has found that cognitive forcing functions, which require more effort to process displayed information, can reduce overreliance on AI (Vasconcelos et al., 2023).

In summary, there are three possible approaches: automating human tasks, hybrid human-AI task solving, or purely human task solving. Regarding performance, purely AI-driven solutions for automating tasks do not yet suffice for most cases of complex knowledge work, such as presentation slide creation (Guo et al., 2023). In hybrid human AI settings, overreliance is a well-overserved phenomenon, resulting in decreased performance (Bansal et al., 2021).

This has implications for the design of human-AI hybrid complex content creation systems, as we aim to enhance quality performance, which has yet been insufficiently considered in this context.

2.3 Vector and Graph-based Retrieval

Although LLMs can learn substantial, in-depth, data-driven knowledge and function effectively as parameterized implicit knowledge bases without external memory, they have significant limitations. They struggle to expand or revise memory, provide transparent insights into their predictions, and exhibit a propensity for generating "hallucinated" information (Marcus, 2020).

RAG has been developed to integrate parametric and non-parametric memory (Hashimoto et al., 2018). In RAG, LLMs serve as the parametric memory, while a vector index (e.g., Wikipedia) represents the non-parametric memory (Lewis et al., 2020). Due to the context size limitations of LLMs, external documents are divided into smaller chunks, converted into vector indices via an embedding model, and stored in a vector database. The retrieval component performs similarity searches to rank and retrieve the most relevant chunks concerning the query. These top-ranked chunks are aggregated and provided as context to the LLM (Sarmah et al., 2024).

KG-based RAG applications have been developed to address hallucinations and missing factuality (Agrawal et al., 2024; Peng et al., 2023). A KG is a graph representation of facts and their semantic relations, initially coined by Google (Ehrlinger & Wöß, 2016). Knowledge is stored as triplets of ⟨subject, predicate, and object⟩, forming the essential elements of KGs, where the predicate is represented by a directed edge, creating a network of interconnected information (Edge et al., 2024; Zhong et al., 2024). KGs enhance search efficiency and accuracy by leveraging a more realistic semantic representation of subject matter (Peng et al., 2023). Due to their formal structure, KG embeddings map

KGs into dense vector spaces, improving RAG performance. Recent KGs employ graph embedding methods to include more features, such as indirect relationships between entities (Peng et al., 2021). We argue that Storypoints—representations of a presentation slide's content and their indirect relationships to other Storypoints within a shared slide deck—are not explicitly considered during conventional RAG. We assume that integrating a KG-based RAG system can leverage this information to find slide candidates that are not only similar in content but also contextualized adequately within the overall storyline. LLMs are employed for KG construction, typically drawing upon semi-structured data such as text, databases, and existing ontologies. The primary objectives are entity recognition, relationship extraction, and co-reference resolution (Sarmah et al., 2024).

The graph structure of KGs is often employed with graph-based databases, providing both visual graph representations and query interfaces, thus improving data manipulation and exploration (Hsuan Yuan et al., 2023). Advanced approaches have shown that visualizing semantic relationships aids in exploring non-obvious relationships between entities, finding relevant information beyond direct connections, and increasing the overall interpretability of KGs (Hsuan Yuan et al., 2023).

We propose an entity-preserving hybrid vector and graph retrieval system that integrates similarity-based retrieval from vector databases with the retrieval of a subgraph directly accessible to the user without relying on the semantic corpus of an LLM. While vector databases are not comprehensible to humans and require LLMs to arrange them in a semantic shell, KGs are directly accessible and support understanding the underlying relationships of complex datasets.

3 Designing a Hybrid Slide Graph Retrieving System

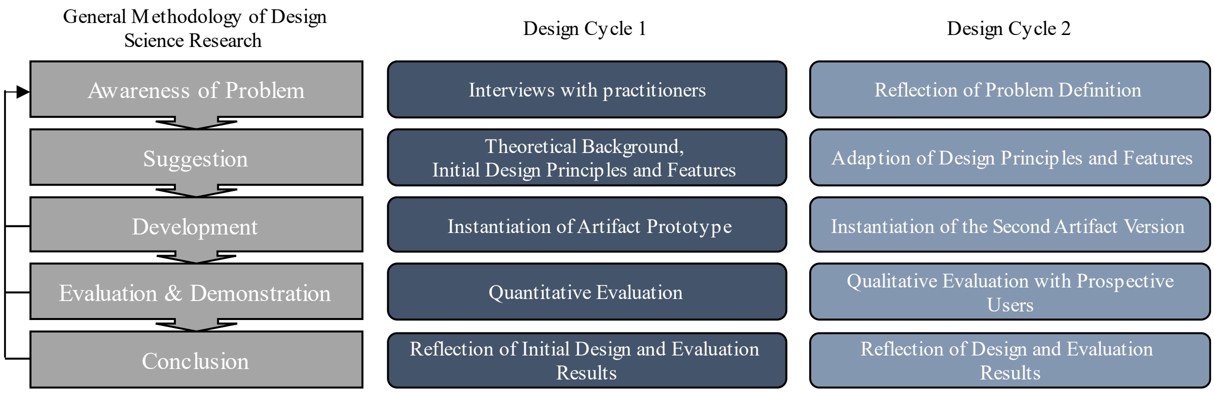


Figure 1. Design Science Research Cycles based on Kuechler and Vaishnavi (2008).

Our study employs a DSR approach to develop a set of DPs. DSR in information systems is widely accepted as a rigorous framework that facilitates the creation of innovations that enhance human capabilities, as discussed in foundational contributions (Hevner et al., 2004; Walls et al., 1992). Our DPs, specifically tailored to address challenges in the problem domain, are demonstrated through implementing the NarrativeNet Weaver. The system integrates genAI's semantic and multimodal capabilities with knowledge graphs and user-centric interfaces. This approach serves a bifurcated role: it not only directs the preliminary design but also provides critical insights for the ongoing development of the artifact, aligning with the insights from Meth et al. (2015).

Regarding the procedural aspect of our research, the DSR literature suggests a cyclical artifact refinement process, where each cycle involves reflection, construction, and revision (Hevner et al., 2004). Following the framework by Peffers et al. (2007), our project proceeded through two iterative cycles. Based on the five phases commonly recognized in DSR studies, as illustrated in Figure 1, we structure and specify our working steps conducted to serve rigor and transparency (Kuechler & Vaishnavi, 2008; Walls et al., 1992).

3.1 Requirements Derived from Expert Interviews

We initiated our study by creating awareness of the problem through two approaches: an informal literature review and expert interviews (Onwuegbuzie & Frels, 2016).

We conducted semi-structured interviews with eight expert coaches specializing in developing workshops for executive management, aiming to collect requirements relevant to the presentation slide creation process systematically. Utilizing the qualitative content analysis methodology outlined by Gläser and Laudel (2010), these interviews affirmed the practical relevance of our research. The interview guide comprised three thematic sections and seven guiding questions to explore current slide creation practices, utilized resources and tools, encountered challenges, time-consuming tasks, and to solicit potential solutions and envisioned interactions with AI systems.

Each 40-minute interview was conducted between November 2023 and May 2024 with experts based in Switzerland and Germany. All of them identified presentation slides as their primary communication medium and have at least 5 year experience in creating slides. They had little or no experience with AI based slide generation yet. All interviews were conducted and transcribed in German, and all quotes below have been translated into English.

First, an integrated storyline framework emerged as essential for guiding users in constructing coherent and logically progressing narratives (DR1). This aligns with the structured approach to slide creation emphasized in prior research, where developing the narrative is the critical initial step (Bourgoin & Muniesa, 2016). Interviewees noted that narrative development is crucial yet remains one of the most time-intensive tasks. One expert noted, "It is difficult to strategically create slides because I never learned how".

To lower the burden of interacting with AI, the system should facilitate communication in natural language, akin to engaging with another human being (DR2). This reduces user resistance and enhances usability (McTear et al., 2016). Inspiration is also critical for facilitating efficient access to relevant information from previous projects (DR3). Users expressed a desire for systems capable of retrieving and suggesting past content; as one stated, "I would like an intelligent knowledge management system that can search through old projects." This resonates with the limitations of foundational AI models, which often lack specialized knowledge (Huan & Zhou, 2024).

An effective search functionality is necessary to swiftly locate and incorporate pertinent information from extensive existing resources (DR4). An interviewee mentioned, "It is necessary to create an application that links AI slide evaluation with my knowledge management tool." Transparency and source disclosure in AI assistance is essential to allow users to adjust the level of AI support and maintain control over the creative process (DR5). One interviewee voiced concerns regarding quality and reliability: "There are risks in terms of quality if AI-generated materials do not disclose their sources, making it hard to evaluate the content's reliability". This concern can also be found in the literature about AI systems' potential oversimplification of content and the need for transparency to ensure content accuracy and trustworthiness (Tufte, 2003).

Finally, maintaining human expertise and control ensures that the system supports rather than supplants human input, preserving the depth and quality of presentations (DR6). This sentiment is captured in the observation, "The greatest value of consultants remains their deep understanding of the company". This also reflects the broader discourse on human-AI collaboration, where AI augments but does not replace human cognitive efforts (B. Chen et al., 2023; Sefid et al., 2019).

3.2 Designing the NarrativeNet Weaver

To conceptualize the DPs, we follow the approach of Meth et al. (2015) with corresponding DFs to tackle the stated DRs. We specify the DPs using the actors, context, mechanism, and rationale comprising the schema of Gregor et al. (2020) (see Figure 2).

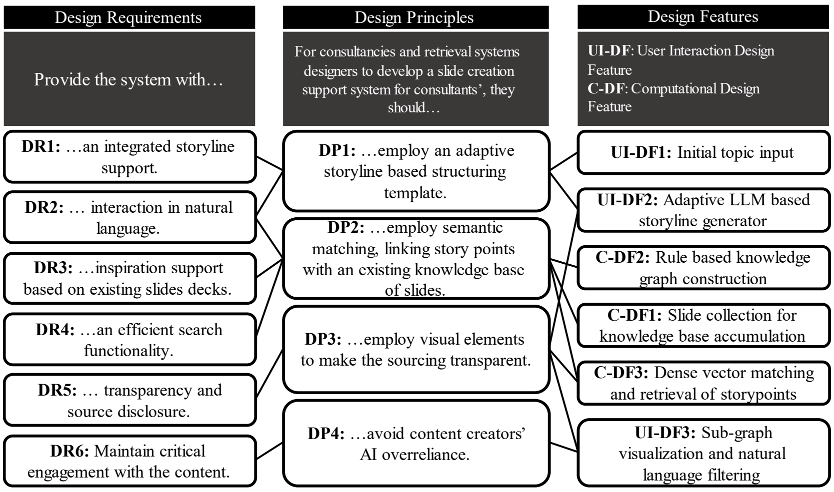


Figure 2. Linking Design Requirements to Design Principles and Design Features according to Gregor et al. (2020).

We developed the NarrativeNet Weaver based on the DRs and corresponding DPs (DP1–4), integrating user interaction modules (UI-DF), leading the interactive narrative curation, and computational modules (C-DF), responsible for graph-based retrieval and customization[2]. Users follow three logical interaction steps to support slide creation: retrieval, access, and visualization (Figure 3).

3.2.1 Interactive Narrative Curation

Initially (UI-DF1), users define the workshop objective and select a relevant topic addressing a skill gap aligned with strategic goals, setting learning goals for participants, and specifying the number of Storypoints and critical milestones to highlight significant learning outcomes (DP1).

Subsequently (UI-DF2), they refine and develop workshop content using AI-generated Storypoints, adapting, adding, or removing points to tailor the presentation into a final storyline that aligns with the workshop's learning goals (DP1). *Making it an adaptive storyline-based structuring template*.

Finally (UI-DF3), users view the retrieved subgraph entailing *visual elements to make the sourcing transparent* (DP3) and apply filters to clarify how the content corresponds to learning goals and Storypoints, exploring relationships within slide decks, individual slides, and Storypoints to ensure comprehensive topic coverage and potentially inspire further narrative development. Identifying gaps or inconsistencies allows for refining learning goals or Storypoints in earlier steps. Natural language-based filtering enables users to input custom queries in plain language, providing precise control over displayed content and facilitating the isolation and review of materials most relevant to the narrative (UI-DF3).

An iterative approach speeds up the slide creation process while still allowing the user to be in control. This addresses the needs expressed by the interviewees, who do not deliver finished slides but rather support specific steps. This constructs an interactive approach, cognitively forcing content creators to engage with the sourced materials. Thus, we avoid content creators' overreliance (DP4).

[2] The implementation of the web application is openly accessible in the online appendix: https://github.com/alexanderHSG/slideGraphRetriever

3.2.2 Graph-Based Retrieval and Customization

We describe three computational features (Figure 3). The NarrativeNet Weaver relies on the structural properties of a large, interconnected knowledge graph (KG) constructed through three core steps. First (C-DF1), utilizing a corpus of existing slides from Speakerdeck.com, which hosts millions of presentations across varied audiences and topics, we employed a crawler to randomly harvest 7,945 unique slide deck URLs, ensuring maximum diversity in topics, languages, and audiences. We retrieved the slide decks in PDF format along with metadata, including titles and unique author identifiers. Second (C-DF2), we constructed the KG from unstructured input data. Using OpenAI's function calling, we obtained structured output from “GPT-4o”. Each slide, sent as an encoded image to the “GPT-4o” with an iterative engineering prompt, was parsed to generate detailed descriptions of visual and textual content. A title was generated for each slide. We constructed the KG using properties from the parsing pipeline and slide collection data: SLIDE nodes connected by IS_FOLLOWED_BY relationships according to their order in the slide deck; each SLIDE assigned to a SLIDE_DECK via CONTAINS; SLIDE_DECKs assigned to authors using CREATED_BY and to categories via BELONGS_TO, as supplied by metadata. A rule-based node and relationship creation approach ensured a deterministic KG, contrasting with the variability of purely LLM-based inference (Edge et al., 2024).

Figure 3. Instantiated NarrativeNet Weaver and Design Features.

We faced challenges in calculating dense embeddings using OpenAI's "text-embedding-3-large" for each SLIDE node due to the high-dimensional (3,072 dimensions) feature space. To mitigate this, we applied Uniform Manifold Approximation and Projection (UMAP) for dimension reduction into a two dimensional space, which is feature-number independent and offers remarkable runtime performance, preparing data for clustering SLIDE nodes using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and thus mitigating the "curse of dimensionality" (Y. Chen et al., 2018; McInnes et al., 2020). DBSCAN, requiring no predefined cluster number and effectively handling noise by classifying dissimilar SLIDE nodes as outliers, was employed with ϵ=0.1 and minimum samples = 2 determined through iterative inspection. Clustered SLIDE nodes form STORYPOINTs, connected via ASSIGNED_TO relationships (Ester et al., 1996); outliers are connected to a STORYPOINT constituted by a single SLIDE node. STORYPOINTs represent core meaningful messages, arranging SLIDE modes. To contextualize STORYPOINTs similar to SLIDES, we sent cumulative descriptions of associated SLIDE nodes to "GPT-4o", obtaining meaningful descriptions, and computed "text-embedding-3-large" embeddings saved as STORYLINE properties.

Third (C-DF3), the computational user interaction system includes three core elements. The storyline supporter (DP1) uses GPT-4o with function calling to generate a deterministic number of Storypoints, guided by an iteratively engineered prompt; users specify the number of Storypoints and the topic. The graph-based slide retriever uses the generated and user-refined Storypoints to calculate "text-embedding-3-large" embeddings for each narrative point; cosine similarity between user-generated Storypoints and all STORYPOINT nodes in the KG is calculated, returning a subgraph of the most similar STORYPOINTs and associated nodes like SLIDE_DECK and SLIDE. In summary, this enables *semantic matching, linking Storypoints with an existing knowledge base of slides* (DP2). The slide customizer performs filtering based on a natural language request, translated into declarative KG query language using "GPT-4o" with the subgraph context and KG schema (UI-DF1).

4 Demonstration and Evaluation: Quantitative Study 1

4.1 Method

To assess the NarrativeNet Weaver system with potential end users, we recruited 73 consultants (43 male and 29 female) through Prolific for an online experiment. We targeted individuals fluent in English and proficient in technology (using it at least twice weekly), specifically those in consulting-related roles such as data scientists, researchers in corporate or management consultancies, and consultants at management consulting firms. Prolific was selected due to its high response quality and sample diversity (Peer et al., 2017). The quantitative evaluation comprised two parts (see Figure 4): a pre-survey and a slide creation task. After the slide creation task, we assessed participants' interaction with the NarrativeNet Weaver and their outputs. Participants completed a pre-survey to control for randomization and demographic information such as years of work experience, age, and gender. The mean age was 29 (SD = 6.69), with an average of 7.99 years of work experience (SD = 10.88) and 6.66 years of experience with presentation slide software (SD = 4.77).

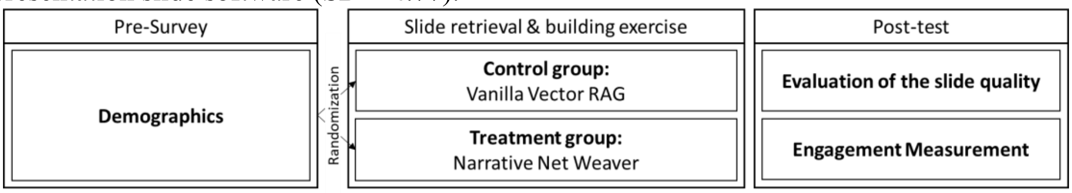


Figure 4. Quantitative Field Experiment Design.

Participants were randomly assigned to either a control group (CG, n = 34) or a treatment group (TG, n = 39). Both groups were tasked with creating exactly five slides using their preferred software (typically PowerPoint or Google Slides) in a scenario where they, as consultants, prepared a workshop aimed at introducing cybersecurity resilience to colleagues. The workshop's goals were twofold: to upskill team members in cybersecurity best practices and to expand the consultancy's offerings in this critical area. Before beginning the task, each group watched an introductory video specific to their version of the system. Participants were informed that a minimum of 30 minutes was required to complete the task, enforced by a countdown timer. They were incentivized with a $5 bonus for the top 10% best-evaluated presentation slides. The TG was given access to a graph and embedding-enhanced version of the NarrativeNet Weaver, while the CG utilized a ChatGPT-4-based chatbot with access to a standard vector database. Both tools were populated with the same repository of presentation slide resources. Participants who did not interact with the provided system at least once were excluded.

4.1.1 Assessment Ratings for Presentation Slide Quality

Information quality, defined as a subjective assessment based on expectations and available options, has been extensively explored in information retrieval but less so in the context of presentation slides (Hilligoss & Rieh, 2008). Kim et al. (2017) proposed a taxonomy for slide evaluation comprising intrinsic, representational, contextual, and reputational quality. Intrinsic quality pertains to the truthfulness and clarity of objectives; representational quality evaluates understandability, visual appeal, and navigability; contextual quality encompasses completeness, informativeness, timeliness, and task suitability; reputational quality reflects the credibility of the source, often associated with the author's reputation (Meier et al., 2024). Wang et al. (2023) introduced five ratings: satisfaction, clarity of structure, ease of understanding of the content, proper organization, and aesthetics. Building upon these dimensions, we developed a set of evaluative ratings (R1–R5) for slide decks: (R1) Accuracy and Relevance—the accuracy and relevance of the content and absence of incorrect or irrelevant material; (R2) Structural Clarity and Organization—the effective layout and organization, including organizational aids like tables of contents and clear delineations between topics (Wang et al., 2023); (R3) Narrative Coherence—the creation of a coherent and engaging narrative that maintains thematic continuity across slides (Kim et al., 2017); (R4) Accessibility and Usability—the ease with which the content can be understood, and the accessibility of language and visuals (Meier et al., 2024; Wang et al., 2023); (R5) Visual Design—the overall aesthetic appeal, consistency in visual design, and effective use of color, graphics, and animation to enhance understanding.

4.1.2 Annotation of Workshop Slides

We recruited two experienced annotators for the study, each with over five years of experience creating presentation slides and at least one year of professional work experience, though not subject matter experts in the specific topics. They were provided detailed annotation guidelines explaining the five evaluation criteria (R1–R5) and associated items, as shown in Table 1. Using a pairwise comparison methodology, annotators were shown one slide deck from the TG and one from the CG for each comparison and asked to select which better reflected each criterion. To mitigate potential biases, the order of criteria and position of slide decks were randomized and anonymized. Annotators were required to take breaks after two hours to maintain focus and accuracy, and two attention checks ensured continued engagement and reliability. The technical display of each slide deck was verified to prevent any issues from affecting evaluations. With 34 submissions from the CG and 39 from the TG, there were 1,326 possible pairwise comparisons, of which 174 pairs were annotated.

4.1.3 Results

Before annotating the 174 slide decks, the two annotators independently annotated 20 identical pairs to measure inter-annotator agreement, calculating Cohen's kappa (κ) (Cohen, 1968). A κ value of 0.5 indicated moderate agreement (Landis & Koch, 1977). The annotators reviewed and discussed the pairs where discrepancies occurred to enhance agreement. After resolving these disagreements, they were randomly assigned new pairs to ensure a sufficient number of annotated slide decks. Across the 174 annotated slide decks, each presentation was annotated at least once.

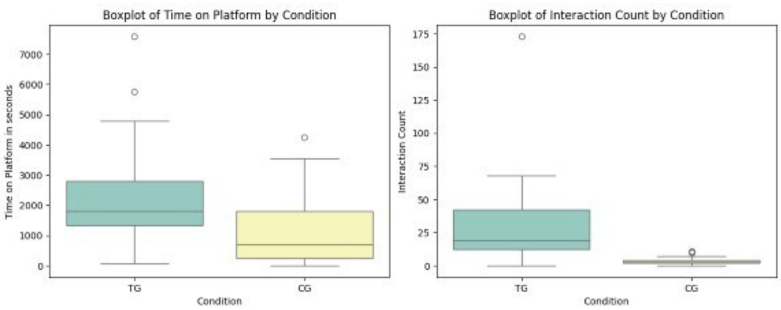
Our alternative hypothesis *(H1) posited that using NarrativeNet Weaver enables consultants to create presentations that achieve higher ratings compared to those supported by state-of-the-art vector-based chatbots.* Given that the pairwise comparison between the two conditions represents a discrete sequence of independent wins without replacement, we used a one-sided Fisher's exact test. For each evaluation criterion, we calculated the p-value and the win rate for the TG using NarrativeNet Weaver (see Table 1). Setting the significance level at $\alpha = 0.05$, we accepted the alternative hypothesis (H1) for two criteria: R1.1 (accuracy and relevance) and R5.1 (aesthetic appeal). This indicates that slides produced with the support of NarrativeNet Weaver were rated significantly higher in terms of subject-specific terminology usage and overall aesthetic appeal compared to those generated with the vector-based chatbot.

Additionally, we measured user engagement with the platform, including the number of interactions and the total time spent in each condition. A paired t-test revealed significant differences in user behavior.

Participants using NarrativeNet Weaver engaged more frequently with the support system and spent more time interacting with it (see Figure 5). Despite this increased engagement, participants in the TG demonstrated superior or equal performance in slide creation, depending on the specific rating criteria.

#Item	Assessment Item	Win Rate (TG)	p-value
R1.1	The information presented in the slide deck uses terminology that is specific to the subject matter.	0.56	0.0207*
R1.2	The content of the slide deck is relevant to the topic at hand.	0.55	0.0538
R1.3	The slide deck includes relevant examples.	0.52	0.2265
R2.1	The slide deck is effectively laid out and organized.	0.52	0.296
R3.1	The slide deck presents a coherent and engaging narrative.	0.51	0.3739
R3.2	There is thematic continuity across the slides in the slide deck.	0.51	0.4573
R3.3	The content effectively conveys the intended narrative or concept.	0.52	0.296
R4.1	The content of the slide deck is easy to understand.	0.49	0.6261
R4.2	The language and visuals in the slide deck are accessible to my fellow consultants.	0.51	0.3739
R5.1	The slide deck has an overall aesthetic appeal.	0.55	0.0341*
R5.2	The visual design of the slide deck is consistent.	0.53	0.1191
R5.3	Color and graphics are used effectively to enhance understanding of the content.	0.51	0.4573

Table 1. Slide annotation pairwise comparison. P-value is for a one-tailed Fishers exact test. p<=0.05.*



Group	Time on platform	Interaction count
Mean (TG)	2151.679	29.412
Mean (CG)	1143.607	3.294
Median (TG)	1722.5	18.0
Median (CG)	707.5	3.0
SD (TG)	1732.15	32.00
SD (CG)	1151.04	2.88
p-value	0.00394 **	0.00004 **
t value	-3.151	-4.725

Figure 5. Time on platform and interaction count. Paired double-tailed t-test. p<=0.05, p<=0.01**.*

5 Demonstration and Evaluation: Qualitative Study 2

5.1 Method

To evaluate the effectiveness of the NarrativeNet Weaver during its development, we conducted a qualitative formative evaluation (Nunamaker Jr. et al., 2015; Venable et al., 2016). We employed semi-structured interviews incorporating "Think-Aloud" sessions, where experts used the tool while verbalizing their thought processes (Nielsen, 1993). Interviewees were selected based on their relevance to the context and experience with presentation slide creation; all participants engaged with presentation slides daily in their professional roles when the interviews occurred in August 2024. Potential participants were identified through direct and indirect referrals and social media.

We coded interviews by stakeholder group and interview number—for example, PS4 refers to the fourth presentation specialist interviewed, an external supplier to consulting firms (see Table 1 for participant details). The think-aloud sessions focused on functionality, usability, and reliability. We analyzed the qualitative data using an inductive content analysis following Mayring and Fenzl (Mayring & Fenzl, 2019).

Stakeholder group	Interview code	Position	Duration
Consultant	C1	Associate Consultant	45 min
	C2	Associate Consultant	30 min
	C3	Senior Associate Consultant	40 min
	C4	Associate Consultant	30 min
	C5	Associate Consultant	35 min
Manager/Partner	M1	Manager, Project Lead	55 min
	M2	Partner (AI Lead)	30 min
	M3	Senior Director	40 min
	M4	Manager, Project Lead	35 min
	M5	Manager, Project Lead	35 min
Graphic Designers/ Presentation Specialist	PS1	Team Lead Design & Graphics	50 min
	PS2	Senior Graphic Designer	35 min
	PS3	Graphic Designer	45 min
	PS4	Executive	50 min
	PS5	Director (AI Lead)	45 min

Table 2. Think-aloud participant overview.

5.2 Results

Functionality emerged as a key aspect, particularly the capability of contextual search, which proved highly effective for navigating complex content within a company. One interviewee noted, "The real value was that it works very well, just as one might imagine beforehand—you had a search engine, you searched in the library" (C1). Participants appreciated using the tool as a starting point to enhance creativity: "We now have a standard slide pool, fixed with recommendations, where old slides are ranked. And yes, I often use this as a starting point and have also put together my own little slide pool" (C2). The NarrativeNet Weaver significantly accelerates the search process, as another user mentioned, "At least from what I have seen, there is superb time-efficiency..." (C1). Interviewees also highlighted the need for a timeliness filter to enhance search functionality: "Ideally, I would have all my proposals that I have ever made in such a context search... The system would then fetch the right slides from, let's say, the 100 proposals I've made in my life, sorting them by recency" (M3). The necessity of integrating existing slides within a storyline context was underscored by challenges in developing presentations directly in tools like PowerPoint: "One limiting factor is when you start directly in PowerPoint, you often realize at the end that you do not have a proper story... you end up with logical jumps" (PS4).

Regarding usability, interviewees mentioned positive aspects related to managing and accessing knowledge, particularly the navigation enabled by the knowledge graph. It allows users to start with a consolidated view and then drill down into details, effectively narrowing the data. A user stated, "...it's important that you can expand and reduce things. That you can focus on an environment from the entire graph" (M1). They acknowledged the superior iterative storyline creation abilities: "Through this graph, which is essentially a linkage to the context... The result should be that I only see relevant topics in my graph environment first... Then... I can adjust my storyline and perhaps do a drag & drop to pull all the information from the graph into the storyline" (M1). Another participant appreciated the simplistic design of the user interface without the need to click through multiple layers: "...it's one page, it's straightforward, it's very clear to operate... it's all about the slides" (C4).

In terms of reliability, the indispensability of contextual search emerged. Interviewees with experience using current knowledge bases found them insufficient: "You had a search engine; you searched in the library. Moreover, if it didn't match exactly, you often didn't find it... a lot of knowledge was actually lost... because buzzwords and synonyms were not recognized" (C1). The NarrativeNet Weaver could mitigate this issue by recognizing synonyms and related terms, thereby improving knowledge retrieval.

6 Discussion

6.1 Theoretical Contributions

Our study has three contributions to the literature on IS: information retrieval, presentation slide creation, and prescriptive design knowledge.

First, our study introduces a novel KG-based RAG design for retrieving multimodal content, whereby visual and textual elements are contextualized within the task of presentation slides. By structuring these elements into Storypoints and aligning them into a coherent storyline, the approach retrieves relevant content while embedding it into a meaningful narrative flow. We implement the NarrativeNet Weaver to combine the visual representation of knowledge graphs with the semantic matching capabilities of vector encoding and thereby enhance the contextual retrieval of relationships and nodes represented by Storypoints and Storylines. We differentiate our retrieval augmented generation application from conventional methods by incorporating application specific structures and using large LLM technologies as an intermediate step to search for and identify relevant retrieval candidates. Our multimodal approach integrates both visual and textual content and now considers visual elements in complex documents such as slides. Finally, we demonstrate that our ability to depict and model relationships among slide elements advances the organization and reuse of digital multimodal content.

Second, the study's results show that the NarrativeNet Weaver participants were more engaged with the content than those using a conventional LLM chatbot with a vector database (naïve RAG) (Li et al., 2024). Users relying on the chatbot tended to directly copy the generated output onto slides without critically engaging with it. This suggests that without structured guidance or visual aids that illustrate the semantic relationships leading to a retrieval candidate, personal interaction with the content becomes minimal and less effective (Moore, 1989). The NarrativeNet Weaver, which uses a KG-based approach, prompted users to interact more and helped them visualize and understand the relationships between content elements. The subgraph visualizations allowed for a deeper examination of the content, discovering connections and insights that might have been overlooked in traditional workflows. This aligns with recent studies' results on semantic visualizations (Hsuan Yuan et al., 2023). This engagement with the content could have significant implications, especially in educational settings, as increased interaction is often associated with improved retention and understanding (Moore, 1989; Thurmond & Wambach, 2004). The NarrativeNet Weaver's ability to encourage such engagement may benefit learner retention and improve slide creation, ultimately impacting knowledge transfer and communication practices.

Third, we provide valuable prescriptive design knowledge to IS researchers, developers of information retrieval, and consultancies, particularly RAG applications, for integrating the advantages of both LLM-based interfaces and KG-based visualizations. This integration leverages the ability of LLMs to find similarities in dense vector spaces while using KGs to represent semantic relationships visually. We have formulated design principles based on insights from literature and user interviews to create systems that support the presentation slide creation process. Furthermore, this design knowledge can be instantiated to support other creative processes, such as writing reports, proposal development, and instructional design in classrooms - where structuring content into coherent narratives is crucial and engagement with the content, thus avoiding overreliance, is desired.

6.2 Practical Contributions

This work offers valuable insights not only for IS researchers and system developers but also for practitioners across various domains where presentation slides are commonly used, such as education, consulting, marketing, corporate training, and project management. The NarrativeNet Weaver helps users overcome procrastination and creative blockades by providing a structured starting point with pre-existing content and a draft storyline, thereby mitigating the "blank page" dilemma (Su et al., 2023).

Novices, particularly those with limited experience in slide creation, benefit from this system as it gives them an initial framework that can be refined and tailored. Moreover, the system promotes self-directed content creation by enabling users to explore existing resources and customize them according to their goals and narrative structure. This support empowers users to take control of the slide-creation process instead of solely relying on AI.

6.3 Limitations and Future Work

It is important to note that our system and user study are not without limitations. The NarrativeNet Weaver's front end still needs improvement. Some participants in the qualitative user study mentioned that it would be helpful to expand and collapse nodes and relationships in the KG to show alternative matching slides by clicking on them. This would support the NarrativeNet Weaver's explorative and interactive properties.

Another important step would be to make the retrieved slides more easily exportable, for example, by enabling users to drag and drop them directly into a presentation slide creation software such as Google Slides or Microsoft PowerPoint.

While data privacy is not a limitation here, as all slide data used is publicly available, companies implementing the artifact should explicitly consider the privacy, security, and compliance implications of their chosen OpenAI endpoint to avoid ethical concerns associated with exposing potentially sensitive information.

Currently, our retriever relies solely on distance metrics calculated by embeddings. However, distance alone does not necessarily indicate relevance. We propose to enhance the retrieval process by incorporating a comprehensive relevance measure. This measure could draw from existing approaches used to assess the quality of presentation slides, which include numerical characteristics such as the number of images, words, and pages, as well as textual characteristics identified holistically by Natural Language Processing (Meier et al., 2024).

Lastly, the quantitative field experiment is based on 174 annotated pairs out of the 1,326 available pairs, which limits the scope of the analysis. Additionally, the annotators were not subject matter experts on the topics for which the slides were created. As a result, the findings reflect preliminary tendencies that should be strengthened with more annotated pairs and the inclusion of domain experts, leading to greater significance in the comparison for slide quality. Also, overreliance is only approximated by engagement with the retrieved content, which is the number of interactions and time spent on the task. An approach to measure is to count the number of incorrect AI suggestions that are included in the user's submission (Vasconcelos et al., 2023). Counting the incorrect adopted number of suggestions in such complex content as slides is not trivial. However, this is necessary in a future study to undoubtedly quantify and compare overreliance.

7 Concluding Remarks

This paper presents NarrativeNet Weaver, a novel interactive system designed to support the creation of presentation slides by leveraging innovative ways for retrieving and visualizing slides embedded in a coherent Storyline while increasing user engagement. The instantiated DPs of our prescriptive study demonstrate that NarrativeNet Weaver promotes more significant interaction with content, improving engagement and facilitating the reuse of slide materials to encourage narrative refinement. Additionally, the system's hybrid approach, which combines the strengths of KGs and vector-based retrieval, offers users precision in retrieving relevant content, fostering a more efficient slide-creation process.

In summary, NarrativeNet Weaver provides a step forward in the design of human-AI collaboration in the problem class of multimodal content creation. It offers a powerful tool for users who seek to improve narrative coherence and streamline the integration of existing content into presentations.

Acknowledgments

We thank Jan Reinhardt for conducting and transcribing a major part of the interviews and all interviewees for their time. This work was supported by Innosuisse and is part of the flagship project SCESC, PFFS-21-29.

References

Agrawal, G., Kumarage, T., Alghamdi, Z., & Liu, H. (2024). *Can Knowledge Graphs Reduce Hallucinations in LLMs? : A Survey* (arXiv:2311.07914). arXiv. https://doi.org/10.48550/arXiv.2311.07914

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445717

Bourgoin, A., & Muniesa, F. (2016). Building a Rock-Solid Slide: Management Consulting, PowerPoint, and the Craft of Signification. *Management Communication Quarterly*, *30*(3), 390–410. https://doi.org/10.1177/0893318916629562

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.*, *5*(CSCW1), 188:1-188:21. https://doi.org/10.1145/3449287

Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023). Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence*, *5*, 100184. https://doi.org/10.1016/j.caeai.2023.100184

Chen, Y., Tang, S., Bouguila, N., Wang, C., Du, J., & Li, H. (2018). A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data. *Pattern Recognition*, *83*, 375–387. https://doi.org/10.1016/j.patcog.2018.05.030

Chiang, C.-W., & Yin, M. (2021). You'd Better Stop! Understanding Human Reliance on Machine Learning Models under Covariate Shift. *Proceedings of the 13th ACM Web Science Conference 2021*, 120–129. https://doi.org/10.1145/3447535.3462487

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., & Larson, J. (2024). *From Local to Global: A Graph RAG Approach to Query-Focused Summarization* (arXiv:2404.16130). arXiv. https://doi.org/10.48550/arXiv.2404.16130

Ehrlinger, L., & Wöß, W. (2016, September 1). *Towards a Definition of Knowledge Graphs*.

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996, August 2). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Knowledge Discovery and Data Mining.

Fu, T.-J., Wang, W. Y., McDuff, D., & Song, Y. (2022). DOC2PPT: Automatic presentation slides generation from scientific documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(1), 634–642. https://ojs.aaai.org/index.php/AAAI/article/view/19943

Gläser, J., & Laudel, G. (2010). Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. *Lehrbuch*.

Gregor, S., Kruse, L. C., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, *21*(6). https://doi.org/10.17705/1jais.00649

Guo, Y., Zhang, Z., Liang, Y., Zhao, D., & Duan, N. (2023). *PPTC Benchmark: Evaluating Large Language Models for PowerPoint Task Completion* (arXiv:2311.01767). arXiv. https://doi.org/10.48550/arXiv.2311.01767

Han, Y., Qiu, Z., Cheng, J., & Lc, R. (2024). When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3613904.3642133

Hashimoto, T. B., Guu, K., Oren, Y., & Liang, P. S. (2018). A Retrieve-and-Edit Framework for Predicting Structured Outputs. *Advances in Neural Information Processing Systems*, *31*.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, *44*(4), 1467–1484. https://doi.org/10.1016/j.ipm.2007.10.001

Hsuan Yuan, C.-W., Yu, T.-W., Pan, J.-Y., & Lin, W.-C. (2023). Interactive Visual Exploration of Knowledge Graphs with Embedding-based Guidance. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3544549.3585596

Huan, X., & Zhou, H. (2024). Integrating Advanced Language Models and Vector Database for Enhanced AI Query Retrieval in Web Development. *International Journal of Advanced Computer Science and Applications*, *15*(6). https://doi.org/10.14569/IJACSA.2024.0150601

Kaplan, S. (2011). Strategy and PowerPoint: An Inquiry into the Epistemic Culture and Machinery of Strategy Making. *Organization Science*, *22*(2), 320–346. https://doi.org/10.1287/orsc.1100.0531

Kim, S., Lee, J.-G., & Yi, M. Y. (2017). Developing information quality assessment framework of presentation slides. *Journal of Information Science*, *43*(6), 742–768. https://doi.org/10.1177/0165551516661917

Kuechler, W., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. *EJIS*, *17*, 489–504.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Li, M. M., Nikishina, I., Sevgili, Ö., & Semmann, M. (2024). *Wiping out the limitations of Large Language Models—A Taxonomy for Retrieval Augmented Generation* (arXiv:2408.02854). arXiv. https://doi.org/10.48550/arXiv.2408.02854

Marcus, G. (2020). *The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence* (arXiv:2002.06177). arXiv. https://doi.org/10.48550/arXiv.2002.06177

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 633–648). Springer Fachmedien. https://doi.org/10.1007/978-3-658-21308-4_42

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3

Meier, A., Rietsche, R., & Blohm, I. (2024). An AI Approach for Predicting Audience Reach of Presentation Slides. *ECIS 2024 Proceedings*. https://aisel.aisnet.org/ecis2024/track04_impactai/track04_impactai/9

Meth, H., Mueller, B., & Maedche, A. (2015). Designing a Requirement Mining System. *Journal of the Association for Information Systems*, *16*(9). https://doi.org/10.17705/1jais.00408

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Nielsen, J. (1993). Chapter 6—Usability Testing. In J. Nielsen (Ed.), *Usability Engineering* (pp. 165–206). Morgan Kaufmann. https://doi.org/10.1016/B978-0-08-052029-2.50009-7

Nunamaker Jr., J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research. *Journal of Management Information Systems*, *32*(3), 10–47. https://doi.org/10.1080/07421222.2015.1094961

Onwuegbuzie, A. J., & Frels, R. (2016). *Seven Steps to a Comprehensive Literature Review: A Multimodal and Cultural Approach*. SAGE.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Peng, C., Vu, D. T., & Jung, J. J. (2021). Knowledge graph-based metaphor representation for literature understanding. *Digital Scholarship in the Humanities*, *36*(3), 698–711. https://doi.org/10.1093/llc/fqaa072

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review*, *56*(11), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

Sarmah, B., Hall, B., Rao, R., Patel, S., Pasquali, S., & Mehta, D. (2024). *HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction* (arXiv:2408.04948). arXiv. http://arxiv.org/abs/2408.04948

Sefid, A., Wu, J., Mitra, P., & Giles, C. L. (2019). *Automatic Slide Generation for Scientific Papers*.

Su, X., Wambsganss, T., Rietsche, R., Neshaei, S. P., & Käser, T. (2023). Reviewriter: AI-Generated Instructions For Peer Review Writing. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 57–71). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.5

Thurmond, V., & Wambach, K. (2004). Understanding interactions in distance education: A review of the literature. *International Journal of Instructional Technology and Distance Learning*, *1*(1), np.

Tufte, E. (2003). The Cognitive Style of PowerPoint. In *Journal of The American Statistical Association—J AMER STATIST ASSN* (Vol. 99).

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M., & Krishna, R. (2023). *Explanations Can Reduce Overreliance on AI Systems During Decision-Making* (arXiv:2212.06823). arXiv. http://arxiv.org/abs/2212.06823

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Walls, J. G., G., J., Widmeyer, R., G., Sawy, O., & A., O. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*, 36. https://doi.org/10.1287/isre.3.1.36

Wang, F., Liu, X., Liu, O., Neshati, A., Ma, T., Zhu, M., & Zhao, J. (2023). Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3544548.3580753

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2024). A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Computing Surveys*, *56*(4), 1–62. https://doi.org/10.1145/3618295