DocHub: Facilitating Comprehension of Documents via Structured Sensemaking with Large Language Models

Zhaoxun "Lorenz" Liu
lorenz@cs.toronto.edu
University of Toronto
Toronto, Ontario, Canada

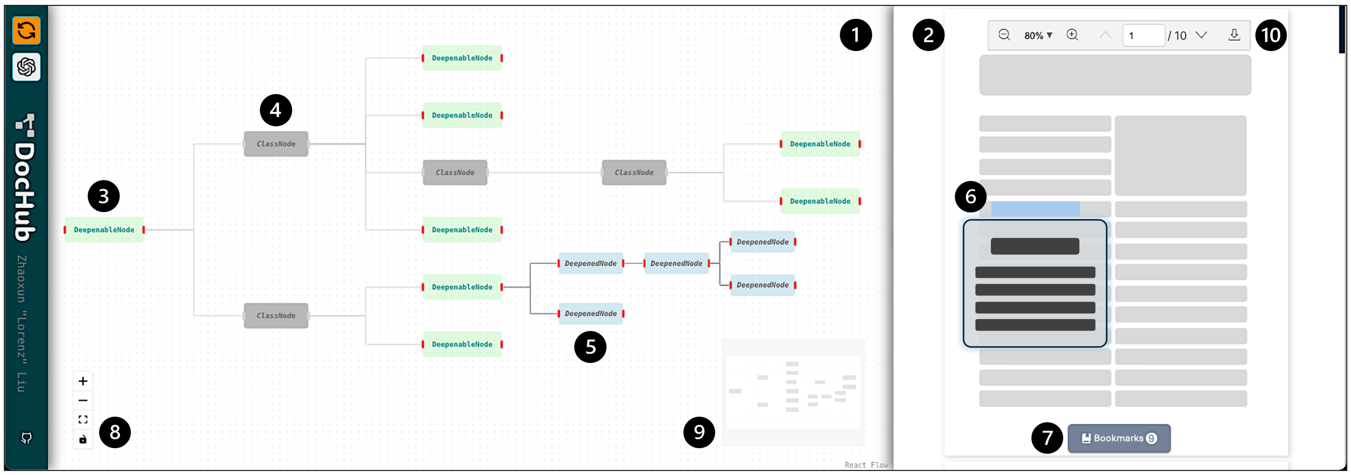

Figure 1: The key user interface components of DocHub (detailed in Section 4.2): 1. Diagram View: holds the visual representation of the document's structure through a node-link diagram; 2. Document View: displays the document, allowing for direct reading and interaction; 3. `DeepenableNode`: stands for a specific part of the document and supports further interaction via creating `DeepenedNodes`; 4. `ClassNode`: categorizes information within the document, aiding in the hierarchical organization; 5. `DeepenedNode`: Created from `DeepenableNodes`, `DeepenedNodes` allow for in-depth exploration on the contextual topics; 6. InstantOp: offers immediate, context-sensitive assistance related to selected text within the document; 7. Bookmark System: enables users to store and easily revisit crucial information; 8. Diagram Toolbar: manipulates the viewport; 9. MiniMap: offers an overview of the document's diagram; 10. Document Toolbar: adjusts zoom level and navigates through the document. The two vertically-arranged buttons in the upper left corner are for restarting and inputting OpenAI API Key.

ABSTRACT

As large language models (LLMs) continue to evolve, they have achieved widespread popularity for generating responses to a broad spectrum of user queries. Furthermore, the integration of capabilities such as file uploading and information retrieval in platforms like ChatGPT, powered by GPT-4, has seen a surge in using LLMs for document comprehension. Despite these advancements and needs, a systematic method to enhance document comprehension through LLMs remains absent, underscoring the urgent need for a structured approach to support users in navigating and understanding complex documents effectively. Therefore, we propose DocHub, a LLM-based interactive system that (1) identifies and visualizes crucial data and their interconnections within documents as node-link diagrams, (2) offers an interactive interface allowing users to modify these visualizations for tailored insights and to pose detailed, context-specific queries for deeper understanding, and (3) features a non-linear abstraction framework to adeptly handle and streamline the complexity of information presented. Our within-subject study demonstrates that DocHub significantly enhances user comprehension, enabling a deeper and more thorough understanding of the documents provided.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Interactive systems and tools**; **Interaction techniques**; *Empirical studies in HCI*.

KEYWORDS

human-AI interaction; structured sensemaking; computer-aided document reading; large language models; visualization

1 INTRODUCTION

Large language models (LLMs) like ChatGPT have become indispensable for a vast user base by facilitating a transition from manual searches to conversational interactions, improving access to and

CSC2524, Winter 2024, University of Toronto

This is an independent project for the CSC2524 Winter 2024 course.

comprehension of information [6]. Despite their widespread use for content creation and information retrieval, current LLM platforms struggle with complex tasks like document comprehension due to limitations in context processing, lack of visual data interpretation, challenges in deep analysis, and linear interaction models, leading to potential information overload [45].

Guided by the outlined challenges, we designed DocHub, a LLM-based interactive system, which is meticulously crafted to address the core issues of document interaction and understanding through a three-fold approach. (1) It adeptly identifies and visualizes essential information and the intricate network of relationships within documents, utilizing node-link diagrams. This visual representation is designed to overcome the inherent limitations of text-based information processing, providing a clear and intuitive overview of document contents and their interconnections. (2) DocHub introduces an interactive platform that empowers users to customize these visualizations according to their specific needs. This functionality not only allows for the adjustment of the graphical presentation for more personalized insights but also enables users to submit detailed, context-specific queries. (3) To manage the inherent complexity of dense and information-rich documents, DocHub incorporates a sophisticated non-linear abstraction framework. This framework organizes the presented information into various levels of granularity, simplifying the user's interaction with complex data sets and enabling a more structured approach to document comprehension. Through these innovative features, DocHub aims to significantly enhance the efficacy and efficiency of document reading and comprehension, providing a comprehensive solution to the limitations currently faced by users of existing LLM platforms.

In summary, we contribute:

- DocHub, a LLM-based interactive system that facilitates document comprehension through non-linear abstraction, streamlining the sensemaking process.
- Externalization of node-link and non-linear sensemaking that enhances the effectiveness of document comprehension;
- A formative study that uncovered limitations in using a conversational interface for complex document comprehension;
- A user study demonstrating that DocHub significantly improves users' efficiency in comprehending documents.

2 RELATED WORK

2.1 Sensemaking

In managing complex information, sensemaking involves noting, curating, and crafting representations like graphs or concept maps to structure information for better understanding and reflection [41]. Externalizing information reduces cognitive load and enhances processing and contemplation of connections at varying abstraction levels [15, 16]. Sensemaking is inherently nonlinear, blending deduction and induction, and requires significant cognitive effort and time for reflecting on information interconnectedness [7, 20, 21]. Efficient sensemaking is crucial to navigate complex information landscapes without distractions from using multiple tools [9, 38].

2.2 Graphical Representation of Information

Graphical representations have long been recognized for their ability to enhance comprehension, memory, and inference, leveraging human aptitude for visual information processing across various domains [1, 4, 27–29, 36, 43, 46]. Significant HCI and visualization research has explored their design, creation, and modification, such as sketching and annotating, making them ideal for sensemaking tasks [2, 8, 34, 40]. The advent of advanced NLP technologies has enabled the automatic generation of diverse graphical content, including visualizations [39], 3D scenes [11], animations [24], and videos [3, 17, 18, 31, 37]. Similarly, the generation of node-link diagrams from texts like video transcripts, documents, and social media enhances the analysis of various data types [12, 22, 25, 33, 47].

The visuo-spatial organization of information addresses the challenges of managing complex data. This method helps reduce cognitive overload, facilitates problem-solving, and supports the manipulation of information through spatial reasoning [15, 23, 26, 30, 32, 35, 44, 48]. Early systems like SemNet [19], the Information Visualizer [10], Workscape [5], and Data Mountain [42] introduced the concept of 3D spatial layouts for documents. Despite potential challenges, such as the risk of clutter and occlusion in 3D interfaces, careful design considerations can mitigate these issues, enhancing usability [13, 14].

3 FORMATIVE STUDY

3.1 Participants and Procedure

We conducted this study with ChatGPT[1] as the LLM platform. We recruited six participants with varying experiences using ChatGPT, including two first-time users, two casual users, and two experienced users who use it daily with advanced prompting techniques and have developed applications using OpenAI's API. The study sessions were conducted via Zoom, lasting one hour each, with participation being voluntary and without financial compensation.

Before starting, participants filled out a pre-task survey to gather demographic data and their experience with ChatGPT. To ensure consistency across sessions, we selected an academic paper as the document for comprehension; academic papers, known for their complexity, structured format, and depth of expert knowledge, served as an ideal document type for this study. Participants first confirmed they had not previously read the chosen paper, then, They were given a set of questions (e.g. *"What research questions is the paper attempting to address?"*) to guide their reading. With 30 minutes allocated for the task, participants were required to understand the paper as thoroughly as possible. Following this, we conducted interviews to reflect on their experiences, particularly focusing on the challenges of using ChatGPT for organizing and comprehending information. Participants were also encouraged to suggest features that could alleviate the difficulties they encountered.

3.2 Findings and Discussion

All participants had successfully completed the formative study and the post hoc interview. We have identified the key **C**hallenges from the interviews:

C1. Verbose and Unstructured Responses: Participants noted that responses from ChatGPT often contained excessive detail, making it challenging to extract pertinent information quickly. The lack

[1] ChatGPT by OpenAI: https://chat.openai.com/

of structured responses further complicated the comprehension process, as users had to sift through dense responses to locate answers or to make sense of different responses, leading to severe information overload.

C2. Hard to Revisit: The oversimplified linear structure of responses by ChatGPT significantly restricted users' ability to engage in or emerge from previous topics.

C3. Difficulty with Deep Analysis: Despite ChatGPT's capacity for generating coherent responses, participants struggled with its limitations in performing deep, analytical tasks. This was particularly evident in the context of academic papers where understanding requires not just comprehension of the text but also critical analysis and synthesis of complex concepts.

C4. Contextual Limitations: ChatGPT, while adept at handling individual queries, sometimes struggles with maintaining context over extended interactions. This can be problematic when attempting to comprehend documents that require an understanding of information presented across multiple queries or sections, as the model might lose track of previous interactions or fail to integrate information cohesively over a longer dialogue.

4 DOCHUB

4.1 Design Rationale

The design rationale of DocHub is guided by the challenges derived from the formative study. In response, DocHub adopts a tripartite design strategy to innovatively address these identified challenges. We visually mark each rationale with the challenges (**C1** - **C4**) they are supposed to address.

Enhanced Visual Representation [C1, C2]: DocHub uses dynamic node-link diagrams to clarify document structures, reducing cognitive load and improving document navigation.

Customizable User Interaction [C3]: DocHub allows users to customize visuals for detailed exploration or broader overview, enhancing personal engagement and understanding.

Non-linear Information Abstraction [C4]: At its core, DocHub employs a non-linear framework for easier data navigation, preventing context loss and facilitating a more intuitive exploration of information.

4.2 User Interface & Features

Dochub consists of two main views: Diagram View (Section 4.2.1) and Document View (Section 4.2.2). Below, we describe each view, its features, and how they help address the challenges (**C1** - **C4**).

4.2.1 Diagram View. DocHub supports document interaction with a diagrammatic interface that organizes texts into node-link diagrams, featuring `ClassNode`, `DeepenableNode`, and `DeepenedNode`. These elements respectively facilitate quick orientation, detailed examination, and in-depth analysis with AI insights. Additionally, a contextual zoom feature dynamically adjusts content visibility to manage information density, enabling seamless navigation between document overviews and focused analyses.

Node-Link Diagram [C1, C2, C4] is the central of DocHub, where documents are transformed into node-link diagrams, featuring three node types: `ClassNode`, `DeepenableNode`, and `DeepenedNode`. Upon initializing a DocHub session, the system automatically generates

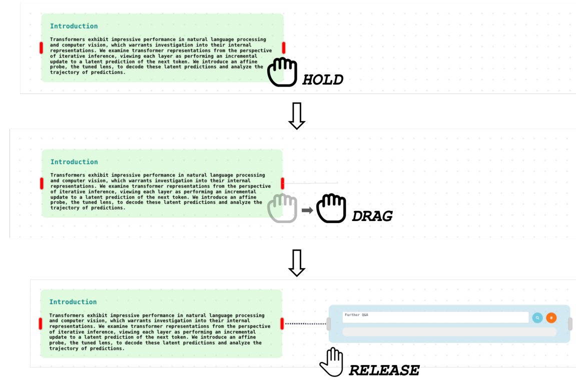

Figure 2: Creating a `DeepenedNode` from a `DeepenableNode`.

this diagram, with `ClassNode` and `DeepenableNode` as fundamental components.

ClassNode [C2] categorizes document content into thematic clusters, providing a clear overview for streamlined navigation.

DeepenableNode [C2, C3] is characterized by their titles and associated content summaries, `DeepenableNodes` delineate the document's finer details. Serving as conduits for in-depth engagement, these nodes enable users to interrogate specific sections, thereby fostering a direct interaction with the document's substance.

DeepenedNode [C3, C4] can be created by users' dragging from `DeepenableNodes` (Figure 2). This type of node hosts more in-depth dialogues or analyses with LLMs.

Contextual Zoom [C1, C2] is to manage information density from LLM responses, smoothly transitioning between thumbnail and detailed views to match user needs. In thumbnail view, only titles of `DeepenableNodes` are shown, with `DeepenedNodes` hidden, facilitating easy navigation between document overviews and in-depth analysis (Figure 3).

4.2.2 Document View. DocHub's Document View features an Interactive Document Viewer for direct engagement with text, an InstantOp popup for real-time AI-assisted insights, and a comprehensive toolbar for visual customization. It also includes a bookmark

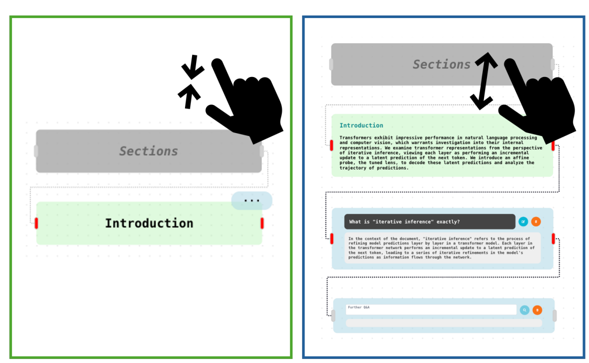

Figure 3: Contextual Zoom: the diagram switches between thumbnail and detailed views based on the zoom level.

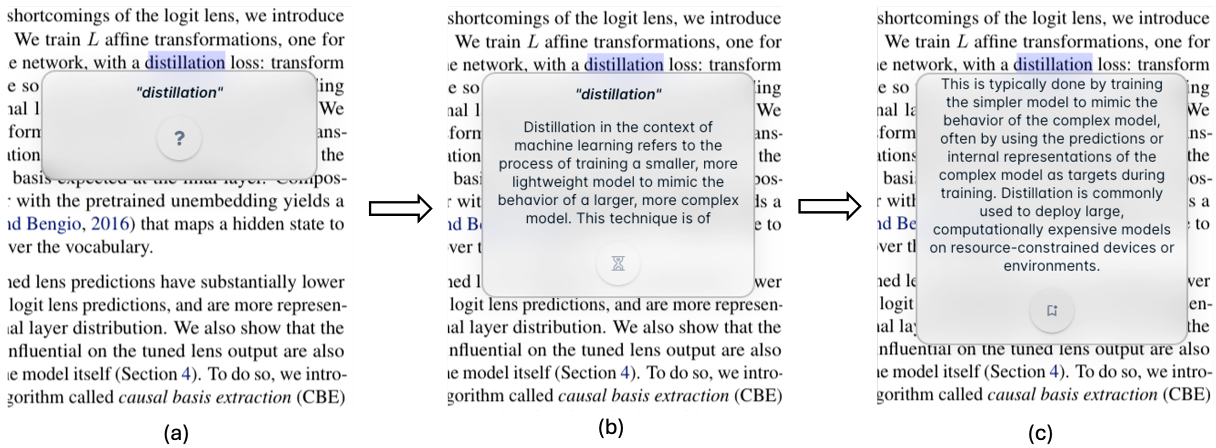

Figure 4: InstantOp usage process: (a) A piece of text is selected; (b) press the button to start in-depth exploration with the LLM; (c) once the LLM finishes responding, users can choose to add a bookmark.

system for annotating and revisiting significant text segments and an adjustable layout for personalized interface arrangement.

Interactive Document Viewer [C2] is central to the Document View functionality, the Interactive Document Viewer displays the document and enables users to engage directly with the text.

InstantOp [C1, C4] is a popup that provides instant LLM-generated explanations for selected text, enhancing document comprehension with real-time, context-specific assistance (Figure 4).

Bookmark System [C2, C3] lets users annotate, save, and search key document excerpts, streamlining navigation with a centralized panel for notes and LLM-generated explanations.

4.2.3 Miscellaneous Features. DocHub is also equipped with a variety of features to make it more user-friendly.

Multiple Node Selection enables users to press the left Shift key and drag the mouse to select and move overlapping nodes together in the diagram.

Diagram Toolbar is positioned at the left bottom corner. Users can zoom in and out, fit view, or lock the viewport via the toolbar.

MiniMap is located at the bottom right corner of the Diagram View, which renders an overview of the diagram and visualizes where the current viewport is in relation to the rest of the flow.

Document Toolbar is positioned at the document's top, offering zoom controls, page navigation, and a download button.

Adjustable Layout is enabled by a draggable splitter, allowing users to customize the display proportions between the document and its diagram, enhancing navigation and visualization according to user preferences.

4.3 Implementation Details

DocHub is developed with the React frontend framework and utilizes React Flow[2], an open-source library, for building interactive node-link diagrams. Our backend leverages two OpenAI models[3]: `gpt-3.5-turbo` and `gpt-4-turbo-preview`. The GPT-3 model is to handle typical language tasks, while the GPT-4 model is reserved for processing token-heavy tasks such as document parsing. The integration with the Retrieval Augmented Generation (RAG) service by LangChain[4] streamlines model interactions, balancing performance with cost efficiency in our system.

5 USER STUDY

To evaluate whether DocHub supports effective comprehension of documents, we conducted a within-subject study. Specifcally, we aimed to answer the following questions:

- **RQ1. Does DocHub facilitate document comprehension?**
- **RQ2. How do people see DocHub being useful in their reading tasks?**

5.1 Baseline & Setup

5.1.1 Baseline. We chose ChatGPT as our baseline due to its widespread use, support of document parsing, and strong capabilities in text generation and comprehension, providing an ideal benchmark for evaluating DocHub's innovations in document interaction and comprehension.

5.1.2 Setup. The study was conducted remotely via Zoom with a new cohort distinct from the Formative Study to eliminate any bias. The study encompassed three types of complex documents: academic papers, technical documentation, and legal documents. Participants were tasked with selecting a designated document from each category that they had not previously read; the complete list of designated documents can be found in Appendix A. The study supervisor was the author of this paper.

5.2 Participants

We recruited 10 participants from a local university, with an average age of 23.3 ($\sigma^2 = 2.7$). The cohort comprised 4 females and 6 males, with a diverse range of academic and professional backgrounds, including physics, mathematics, neuroscience, and engineering. All participants frequently engage with complex document reading,

[2]React Flow https://reactflow.dev/
[3]OpenAI API released models: https://platform.openai.com/docs/models/overview
[4]LangChain Retrieval: https://www.langchain.com/retrievals

such as academic papers and technical reports, in their daily activities. Regarding interactions with LLM platforms like ChatGPT, 6 participants indicated they had extensive experience, 3 had moderate experience, and 1 lacked any experience. This cohort was responsible for all user study tasks, and as their participation was voluntary, no financial compensation was made.

5.3 Metrics

5.3.1 Objective Metrics. Throughout the study, we track two objective metrics: (1) the duration required to comprehend a document fully; (2) the quantity of concepts or information acquired, as indicated by participants in the post-task survey where they list all the concepts or information they learned. To ensure accuracy in measuring the concepts explored, two raters independently evaluated the responses, achieving an inter-rater reliability score of 0.93 on the Intraclass Correlation Coefficient (ICC) scale (2,1), indicating a high level of agreement. It's important to note that switching between ChatGPT and DocHub was not considered as revisiting information in this analysis.

5.3.2 Subjective Metrics. In the post-task survey, two subjective metrics are measured: (1) NASA Task Load Index (NASA-TLX): participants were asked to fill out the NASA-TLX questionnaire for both tools (according to the relevance of the NASA-TLX metrics to this task, we only measure: *Mental Demand, Performance, Effort, Frustration*); (2) overall opinions: This metric assesses participants' overall perspectives regarding whether DocHub helps to address the challenges posed by the baseline platform (see Section 3.2). We used responses to the post-study survey and interview as measures. For example, this included responses to questions asking for their agreement (1: Strongly Disagree; 5: Strongly Agree) with statements, such as *"DocHub successfully addressed the challenge of Verbose and Unstructured Responses."*

5.4 Procedure

5.4.1 Pre-task Survey. Participants completed a pre-task survey on demographic information, their familiarity with document comprehension, and experience with LLM platforms like ChatGPT, ensuring a clear baseline for comparison and analysis. Another crucial purpose is to make sure the participants have no prior reading experience in the designated document.

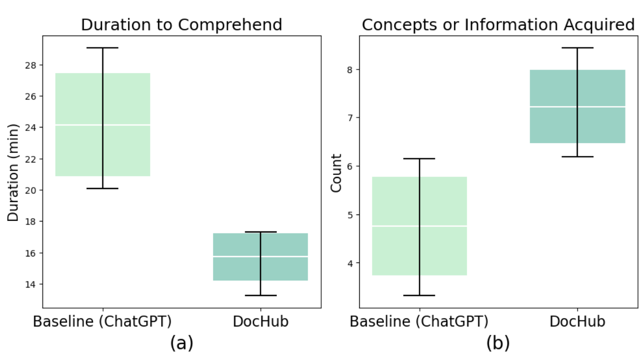

Figure 5: With DocHub, participants simultaneously exhibited (a) significantly reduced durations for document comprehension; (b) a substantial increase in the acquisition of concepts or information.

5.4.2 Pre-task Exercise. Participants started by freely exploring DocHub to learn its interface and features, aiming to prepare them for subsequent tasks. They could also ask for clarifications.

5.4.3 Tasks. The tasks within the user study were designed to probe DocHub's effectiveness across a diverse set of complex documents, including academic papers, technical documentation, and legal documents. This variety aimed to test DocHub's adaptability and efficiency in handling documents of varying complexity and structure. A within-subject study design was employed, facilitating a direct comparison between participants' experiences using both ChatGPT and DocHub for document comprehension tasks. This comparative assessment followed an A-B testing format, where:

(1) *Initial Allocation*: Participants were randomly assigned to begin with either ChatGPT or DocHub, mitigating initial bias toward either platform.
(2) *Main Tasks*: Participants received a document from an unfamiliar field and used the currently designated platform to read and comprehend it at their own pace, without time constraints, until they felt they had sufficiently grasped the content. This approach, including assigning documents outside participants' expertise, aimed to eliminate bias from prior knowledge.
(3) *Platform Switch and Rest Period*: Participants received a flexible rest period of at least 15 minutes between platform switches to mitigate mental fatigue and ensure consistent performance.
(4) *Repeat Tasks with New Documents*: After the rest period, participants switched to the other platform and repeated the *Main Tasks* with a new piece of documents analogous in complexity and type to the initial piece. This step ensured that each participant's experience with both platforms was based on a comparable range of content.
(5) *Study Conclusion*: Upon completion of the *Main Tasks* across all three document categories using both platforms, yielding a total of $3 \times 2 = 6$ *Main Tasks*, participants will have concluded this phase of the study and proceed to the post-task survey.

5.4.4 Post-task Survey. The post-task survey is divided into three sections: (1) participants' self-reporting as many concepts or pieces of information as they have learned; (2) completing the NASA-TLX questionnaire to evaluate their cognitive load using both platforms; (3) responding to a questionnaire asking about their overall opinions regarding DocHub's capability in address the posed challenges.

5.5 Results

5.5.1 RQ1. Does DocHub facilitate document comprehension? We answer this research question from two metrics: the duration required to comprehend a document fully and the quantity of concepts or information acquired.

For the duration metric, participants using ChatGPT demonstrated an average comprehension time of 24.16 minutes (M = 24.16, SD = 3.28), while those utilizing DocHub recorded a reduced average of 15.73 minutes (M = 15.73, SD = 1.51). The t-test comparing these two conditions revealed a statistically significant difference,

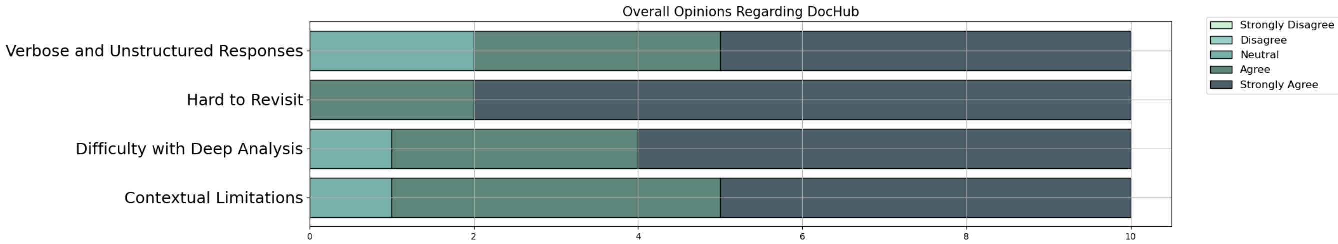

Figure 6: Participants concurred that DocHub significantly alleviates the challenges associated with comprehending documents when compared to using ChatGPT.

$t(18) = 7.00, p = 1.56 \times 10^{-6} < 0.05$, indicating significantly faster comprehension times with DocHub (see Figure 5).

In terms of the quantity of concepts or information acquired, participants using ChatGPT reported learning an average of 4.76 concepts (M = 4.76, SD = 1.02). Conversely, DocHub users learned more, with an average of 7.22 concepts (M = 7.22, SD = 0.76). The difference in the number of concepts learned was statistically significant as indicated by the t-test, $t(18) = 5.82, p = 1.63 \times 10^{-6} < 0.05$, suggesting a more effective comprehension when using DocHub. These findings suggest that DocHub enhances the efficiency and depth of document comprehension.

5.5.2 RQ2. How do people see DocHub being useful in their reading tasks? We answer this research question by analyzing the NASA-TLX and overall opinions reported in the post-task survey.

The participants' overall opinions regarding DocHub is shown in Figure 6, which demonstrates DocHub's capability. In assessing cognitive workload through the NASA-TLX scale, the study revealed distinct differences between ChatGPT and DocHub across various dimensions (see Figure 7). Participants reported a higher Mental Demand using ChatGPT (M = 62.70, SD = 14.16) than DocHub (M = 50.20, SD = 7.03), $t(18) = 2.37, p = 0.029 < 0.05$, suggesting that

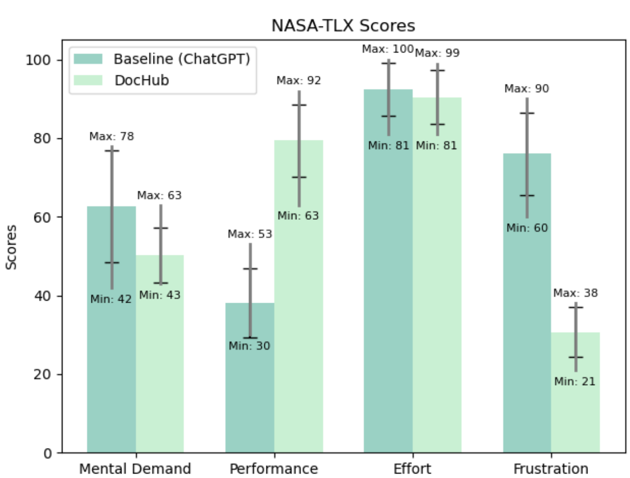

Figure 7: NASA-TLX subjective workload assessments reported by participants, from which we can tell DocHub provides a significantly better overall user experience.

DocHub facilitates a less cognitively taxing experience, although the exact statistical significance remains to be calculated. Conversely, in terms of Performance, participants felt more effective with DocHub (M = 79.40, SD = 9.12), $t(18) = -9.76, p = 1.28 \times 10^{-8} < 0.05$, in contrast to their experience with ChatGPT (M = 38.10, SD = 8.81). Effort levels were comparable across both platforms; however, a marked difference was observed in the Frustration dimension, with users experiencing less frustration with DocHub (M = 30.60, SD = 6.34) than with ChatGPT (M = 76.0, SD = 10.58), $t(18) = 11.03, p = 1.91 \times 10^{-9} < 0.05$. These findings demonstrate DocHub's usefulness in complex document reading.

6 DISCUSSION

The development and evaluation of DocHub highlight its effectiveness in enhancing document comprehension through structured sensemaking with LLMs. Our study demonstrates DocHub's capacity to significantly streamline the comprehension process, reducing cognitive overload and facilitating a deeper understanding of complex documents. These achievements resonate with our initial design rationale, addressing the identified challenges of verbose responses, difficulty in revisiting content, deep analysis, and contextual limitations inherent in existing LLM platforms like ChatGPT. The interactive visualization tools and non-linear abstraction framework employed by DocHub not only aid in navigating the intricacies of documents but also personalize the user experience, catering to varied user preferences and learning styles. Foreseeable future work could include integrating support for multimodal documents like videos, involving the visual and auditory data.

7 CONCLUSION

In this paper, we propose DocHub, a LLM-based interactive system that facilitates document comprehension via structured sensemaking. DocHub forecasts an advancement in utilizing LLMs for document comprehension, marked by its ability to offer a more efficient, user-friendly, and insightful exploration of text. The system's innovative approach, leveraging dynamic visualizations and tailored interactions, underscores the potential of integrating AI with user-centric design principles. As we anticipate future enhancements and broader application scopes, DocHub sets a foundational step towards revolutionizing the way users interact with and understand complex information, fostering a more intuitive and engaging learning environment.

REFERENCES

[1] Maneesh Agrawala, Wilmot Li, and Floraine Berthouzoz. 2011. Design Principles for Visual Communication. *Commun. ACM* (April 2011), 60–69. http://graphics.berkeley.edu/papers/Agrawala-DPV-2011-04/

[2] Brad Asher. 1994. The Professional Vision. *American Journalism* 11 (1994), 304–320. https://api.semanticscholar.org/CorpusID:141946525

[3] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2019. Elastic Documents: Coupling Text and Tables through Contextual Visualizations for Enhanced Document Reading. *IEEE Transactions on Visualization and Computer Graphics* 25 (2019), 661–671. https://api.semanticscholar.org/CorpusID:52042590

[4] Jeff Baker, Donald Jones, and Jim Burkman. 2009. Using Visual Representations of Data to Enhance Sensemaking in Data Exploration Tasks. *J. AIS* 10 (07 2009). https://doi.org/10.17705/1jais.00204

[5] Joseph M. Ballay. 1994. Designing Workscape: an interdisciplinary experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 10–15. https://doi.org/10.1145/191666.191676

[6] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL]

[7] John M. Budd. 2004. Relevance: Language, Semantics, Philosophy. *Libr. Trends* 52 (2004), 447–462. https://api.semanticscholar.org/CorpusID:6302795

[8] Alberto J. Cañas, Roger Carff, Greg Hill, Marco Carvalho, Marco Arguedas, Thomas C. Eskridge, James Lott, and Rodrigo Carvajal. 2005. *Concept Maps: Integrating Knowledge and Information Visualization.* Springer Berlin Heidelberg, Berlin, Heidelberg, 205–219. https://doi.org/10.1007/11510154_11

[9] Robert G. Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. 2010. Tools-at-hand and learning in multi-session, collaborative search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010). https://api.semanticscholar.org/CorpusID:8949187

[10] Stuart K. Card, George G. Robertson, and Jock D. Mackinlay. 1991. The information visualizer, an information workspace. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1991). https://api.semanticscholar.org/CorpusID:14383080

[11] Angel X. Chang, Will Monroe, Manolis Savva, Christopher Potts, and Christopher D. Manning. 2015. Text to 3D Scene Generation with Rich Lexical Grounding. *ArXiv* abs/1505.06289 (2015). https://api.semanticscholar.org/CorpusID:802243

[12] Antoine Clarinval, Isabelle Linden, Anne Wallemacq, and Bruno Dumas. 2018. Evoq: a Visualization Tool to Support Structural Analysis of Text Documents. *Proceedings of the ACM Symposium on Document Engineering 2018* (2018). https://api.semanticscholar.org/CorpusID:52899091

[13] Andy Cockburn and Bruce McKenzie. 2002. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 203–210. https://doi.org/10.1145/503376.503413

[14] Andy Cockburn and Bruce J. McKenzie. 2000. An Evaluation of Cone Trees. In *British Computer Society Conference on Human-Computer Interaction.* https://api.semanticscholar.org/CorpusID:10686243

[15] Anita Crescenzi, Yuan Li, Yinglong Zhang, and Robert G. Capra. 2019. Towards Better Support for Exploratory Search through an Investigation of Notes-to-self and Notes-to-share. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2019). https://api.semanticscholar.org/CorpusID:197928076

[16] Anita Crescenzi, Austin R. Ward, Yuan Li, and Robert G. Capra. 2021. Supporting Metacognition during Exploratory Search with the OrgBox. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2021). https://api.semanticscholar.org/CorpusID:235792408

[17] Weiwei Cui, Xiaoyu Zhang, Yun Wang, He Huang, B. Chen, Lei Fang, Haidong Zhang, Jian-Guang Lou, and Dongmei Zhang. 2019. Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements. *IEEE Transactions on Visualization and Computer Graphics* 26 (2019), 906–916. https://api.semanticscholar.org/CorpusID:198147846

[18] Patrick Esser, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasis Germanidis. 2023. Structure and Content-Guided Video Synthesis with Diffusion Models. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (2023), 7312–7322. https://api.semanticscholar.org/CorpusID:256615582

[19] Kim M. Fairchild, Steven E. Poltrock, and George W. Furnas. 1990. Semnet: three-dimensional graphic representa-tions of large knowledge bases. https://api.semanticscholar.org/CorpusID:60205603

[20] Nigel Ford. 1999. Information retrieval and creativity: towards support for the original thinker. *J. Documentation* 55 (1999), 528–542. https://api.semanticscholar.org/CorpusID:38268837

[21] Allen Foster and Nigel Ford. 2003. Serendipity and information seeking: an empirical study. *J. Documentation* 59 (2003), 321–340. https://api.semanticscholar.org/CorpusID:40143767

[22] Tessai Hayama and Shuma Sato. 2020. Supporting Online Video e-Learning with Semi-automatic Concept-Map Generation. In *Interacción.* https://api.semanticscholar.org/CorpusID:220528397

[23] Sarah Henderson. 2009. Personal document management strategies. In *ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction.* https://api.semanticscholar.org/CorpusID:32614371

[24] Fangzhou Hong, Mingyuan Zhang, Liang Pan, Zhongang Cai, Lei Yang, and Ziwei Liu. 2022. AvatarCLIP: zero-shot text-driven generation and animation of 3D avatars. *ACM Trans. Graph.* 41, 4, Article 161 (jul 2022), 19 pages. https://doi.org/10.1145/3528223.3530094

[25] Mengdie Hu, Krist Wongsuphasawat, and John T. Stasko. 2017. Visualizing Social Media Content with SentenTree. *IEEE Transactions on Visualization and Computer Graphics* 23 (2017), 621–630. https://api.semanticscholar.org/CorpusID:17688893

[26] Janellen Huttenlocher and Clark C. Presson. 1973. Mental rotation and the perspective problem. *Cognitive Psychology* 4 (1973), 277–299. https://api.semanticscholar.org/CorpusID:122229679

[27] Gwo jen Hwang, Po-Han Wu, and Huirong Ke. 2011. An interactive concept map approach to supporting mobile learning activities for natural science courses. *Comput. Educ.* 57 (2011), 2272–2280. https://api.semanticscholar.org/CorpusID:29209287

[28] Peiling Jiang, Jude Rayan, Steven W. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (2023). https://api.semanticscholar.org/CorpusID:258823121

[29] Peiling Jiang, Fuling Sun, and Haijun Xia. 2023. Log-it: Supporting Programming with Interactive, Contextual, Structured, and Visual Logs. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 594, 16 pages. https://doi.org/10.1145/3544548.3581403

[30] William P. Jones and Susan T. Dumais. 1986. The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Trans. Inf. Syst.* 4 (1986), 42–63. https://api.semanticscholar.org/CorpusID:16201222

[31] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating Document Reading by Linking Text and Tables. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (2018). https://api.semanticscholar.org/CorpusID:52905094

[32] David Kirsh. 2001. The Context of Work. *Human–Computer Interaction* 16 (2001), 305 – 322. https://api.semanticscholar.org/CorpusID:28915179

[33] Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative Concept Mapping for Video Learning. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://api.semanticscholar.org/CorpusID:5046960

[34] Michael Xieyang Liu, Andrew Kuznetsov, Yongsung Kim, Joseph Chee Chang, Aniket Kittur, and Brad A. Myers. 2022. Wigglite: Low-cost Information Collection and Triage. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (2022). https://api.semanticscholar.org/CorpusID:251223638

[35] Thomas W. Malone. 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.* 1 (1983), 99–112. https://api.semanticscholar.org/CorpusID:1486850

[36] J.H. McClellan, L.D. Harvel, R. Velmurugan, M. Borkar, and C. Scheibe. 2004. CNT: concept-map based navigation and discovery in a repository of learning content. In *34th Annual Frontiers in Education, 2004. FIE 2004.* F1F–13. https://doi.org/10.1109/FIE.2004.1408581

[37] Ronald A. Metoyer, Qiyu Zhi, Bart Janczuk, and Walter J. Scheirer. 2017. Coupling Story to Visualization: Using Textual Analysis as a Bridge Between Data and Interpretation. *23rd International Conference on Intelligent User Interfaces* (2017). https://api.semanticscholar.org/CorpusID:3795712

[38] André N. Meyer, Thomas Fritz, Gail C. Murphy, and Thomas Zimmermann. 2014. Software developers' perceptions of productivity. *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering* (2014). https://api.semanticscholar.org/CorpusID:2484132

[39] Arpit Narechania, Arjun Srinivasan, and John T. Stasko. 2020. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Transactions on Visualization and Computer Graphics* 27 (2020), 369–379. https://api.semanticscholar.org/CorpusID:221292836

[40] Srishti Palani, Yingyi Zhou, Sheldon Zhu, and Steven W. Dow. 2022. InterWeave: Presenting Search Suggestions in Context Scaffolds Information Search and Synthesis. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (2022). https://api.semanticscholar.org/CorpusID:253205375

[41] Peter Pirolli and Stuart Card. 2015. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. https://api.semanticscholar.org/CorpusID:237007033

[42] George G. Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten van Dantzich. 1998. Data mountain: using spatial memory for document management. In *ACM Symposium on User Interface Software and Technology.* https://api.semanticscholar.org/CorpusID:12723851

[43] Michail Schwab, Hendrik Strobelt, James Tompkin, Colin Fredericks, Connor Huff, Dana Higgins, Anton Strezhnev, Mayya Komisarchik, Gary King, and Hanspeter Pfister. 2017. booc.io: An Education System with Hierarchical Concept Maps. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 571–580. https://ieeexplore.ieee.org/document/7536150
[44] Roger N. Shepard and Jacqueline Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171 (1971), 701 – 703. https://api.semanticscholar.org/CorpusID:16357397
[45] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (2023). https://api.semanticscholar.org/CorpusID:258822925
[46] Tamara Sumner, Faisal Ahmad, Sonal Bhushan, Qianyi Gu, Francis Molina, Stedman Willard, Michael Wright, Lynne Davis, and Greg Janée. 2005. Linking learning goals and educational resources through interactive concept map visualizations. *Int. J. Digit. Libr.* 5, 1 (mar 2005), 18–24. https://doi.org/10.1007/s00799-004-0112-x
[47] Chien-Lin Tang, Jingxian Liao, Hao-Chuan Wang, Ching-Ying Sung, and Wen-Chieh Lin. 2021. ConceptGuide: Supporting Online Video Learning with Concept Map-based Recommendation of Learning Path. *Proceedings of the Web Conference 2021* (2021). https://api.semanticscholar.org/CorpusID:235324809
[48] Barbara Tversky. 2004. Visuospatial Reasoning. https://api.semanticscholar.org/CorpusID:9393679

A LIST OF DESIGNATED DOCUMENTS

All third-party hyperlinks mentioned have been verified as valid at the time of writing this paper; however, they may become invalid in the future due to changes by the third party or for other reasons.

A.1 Academic Papers

- Deuze, M. (2004). *What is multimedia journalism?*. Journalism studies, 5(2), 139-152.
- Franklin, B. (2012). *The future of journalism: Developments and debates*. Journalism studies, 13(5-6), 663-681.
- Purnomo, A., Afia, N., Prasetyo, Y. T., Rosyidah, E., Persada, S. F., & Maulana, F. I. (2022). *Business Model on M-Business: A Systematic Review*. Procedia Computer Science, 215, 955-962.
- Fairlie, R., & Fossen, F. M. (2021). *The early impacts of the COVID-19 pandemic on business sales*. Small Business Economics, 1-12.
- Mathew, A., Amudha, P., & Sivakumari, S. (2021). *Deep learning techniques: an overview*. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 599-608.
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). *Machine learning and deep learning*. Electronic Markets, 31(3), 685-695.
- Thesing, T., Feldmann, C., & Burchardt, M. (2021). *Agile versus waterfall project management: decision model for selecting the appropriate approach to a project*. Procedia Computer Science, 181, 746-756.
- Blaskovics, B., Czifra, J., Klimkó, G., & Szontágh, P. (2023). *Impact of the applied project management methodology on the perceived level of creativity*. Acta Polytechnica Hungarica, 20(3), 101-120.
- Liao, Q. V., Gruen, D., & Miller, S. (2020, April). *Questioning the AI: informing design practices for explainable AI user experiences*. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-15).
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). *What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education*. Smart Learning Environments, 10(1), 15.

A.2 Technical Documentations

- Sphinx: https://www.sphinx-doc.org/en/master/index.html
- PyTorch: https://pytorch.org/docs/stable/index.html
- TensorFlow: https://www.tensorflow.org/api_docs
- React: https://react.dev/
- Go Programming Language: https://go.dev/doc/
- R Programming Language: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
- Swift: https://www.swift.org/documentation/
- Matplotlib: https://matplotlib.org/stable/index.html
- Unity: https://docs.unity3d.com/Manual/index.html
- Unreal Engine: https://dev.epicgames.com/documentation/en-us/unreal-engine/unreal-engine-5-3-documentation

A.3 Legal Documents

- Agreement of Purchase and Sale, Ontario Real Estate Association: https://crm.agentlocator.ca/UserFiles/6779/files/OREA-Form-100%20(2)Agreement%20of%20Purchase%20and%20Sale.pdf
- Residential Tenancy Agreement (Standard Form of Lease), Ottawa Region Landlords Association: http://www.orla.ca/Ontario_Standard_Lease_2021.pdf
- CIBC Mutual Funds Account Agreement, Canadian Imperial Bank of Commerce: https://www.cibc.com/content/dam/personal_banking/investments/pdfs/mutual_funds/reporting_and_governance/8798-cibc-securities-account-agreement-and-disclosures-en.pdf
- Personal Account Agreement, Canadian Imperial Bank of Commerce: https://www.cibc.com/content/dam/personal_banking/bank_accounts/pdfs/personal-account-agreement-en.pdf
- Agreement to Purchase a GIC, Canadian Imperial Bank of Commerce: https://www.cibc.com/ca/pdf/gic-en.pdf
- Service Agreement for Personal Accounts, Royal Bank of Canada: https://www.rbcbank.com/siteassets/Uploads/pdfs/Service-Agreement-for-Personal-Accounts.pdf
- RBC Royal Bank® Business Credit Card Agreement, Royal Bank of Canada: https://www.rbcroyalbank.com/credit-cards/documentation/pdf/rbc-royal-bank-business-credit-card-agreement.pdf
- Scotiabank Privacy Agreement, Scotiabank: https://www.scotiabank.com/content/dam/scotiabank/canada/common/documents/Scotiabank_Privacy_Agreement_EN_CLEAN_FINAL.pdf
- Digital Banking Agreement, TD Canada Trust: https://www.td.com/content/dam/tdct/document/pdf/dba-eng.pdf
- Schedule 28 Insurance Trust Agreement, City of Ottawa: https://documents.ottawa.ca/sites/documents/files/TLPA-Schedule%2028%20-%20Insurance%20Trust%20Agreement%20-%20Execution%20Version%20%28Redacted%29-AODA.pdf