Visualization and Interaction Techniques for Single-Text Digital Reading: A Survey

Lei Han[a], Jiandan Song[a], Shuai Chen[a,*] and Zhaoman Zhong[a]

[a]*School of Computer Engineering, Jiangsu Ocean University, Cangwu Road #59, Lianyungang 222000, Jiangsu, China*

ARTICLE INFO

Keywords:
Single-Text Reading
Reading Assistance
Visualization Technologies
Interaction Techniques

ABSTRACT

With the development of information technology, digital reading has become an important way to acquire knowledge. As text resources continue to grow, readers have an increasing need to efficiently understand key information from a single text. To address this challenge, visualization technologies are becoming useful tools for reading assistance. They help present text clearly, highlight important content, and improve reading efficiency. This paper reviews and summarizes recent representative studies on visualization and interaction techniques in single-text reading, and classifies existing methods from two core dimensions. First, by data type: (1) structural information, such as chapters and arguments; (2) content elements, such as data and charts; (3) user interaction data, including highlighting and annotation. Second, by technical approach: (1) Text Presentation Enhancement; (2) Information Content Enhancement; (3) Layout Optimization; (4) Interaction Enhancement. These techniques improve text display in different ways and support better understanding and memory. Based on this classification, the paper reviews the current development of relevant technologies, explores their application potential in academic, educational, and journalistic settings, and summarizes key functions and design concepts of typical reading assistance systems to provide references for future research and system design.

1. Introduction

Print reading is being replaced by the more interactive and processable digital reading. This is especially true in high-frequency and fragmented learning contexts. Digital reading offers greater accessibility. Its good computability and interactivity have also spurred various reading assistance tools. These tools leverage clear text structures and computational advancements. They provide enhanced functions like content navigation, term explanation, summary extraction, and information visualization. This effectively improves reading efficiency and comprehension depth. Meanwhile, the variety and volume of digital texts are continuously growing. These include news reports, academic papers, argumentative articles, textbooks, and emerging digital content. These texts differ significantly in structure and information presentation. This places higher adaptability demands on reading assistance technologies. Recently, researchers have proposed innovative methods using visualization and interaction design to meet this challenge. These methods help readers more effectively perceive, understand, and utilize text information. Notably, the widespread application of Large Language Models (LLM), such as ChatGPT, has further driven the intelligent evolution of reading assistance technology. LLM not only enhance natural language processing capabilities. They also provide new technical support for tasks like text summarization, entity recognition, and logical structure extraction. This expands the auxiliary methods and interaction possibilities in digital reading scenarios.

This survey focuses on techniques of reading in knowledge-intensive, deep-reading scenarios. Specifically, “single-text reading assistance”, as defined in this survey, refers to the technological support provided to a user engaging with a single, self-contained document—such as an academic paper, a news article, or a textbook chapter—to aid in content comprehension, structural navigation, and interactive exploration. Different from tasks like multi-document summarization, cross-textual analysis, or corpus-level search, single-text reading emphasizes in-depth understanding and information extraction within the confines of the document, a common and critical activity in educational, academic, and professional contexts. The challenges in this single-text context include understanding complex academic arguments, navigating structured textbooks, and analyzing detailed news reports. These require specialized visualization and interaction techniques that emphasize semantic enrichment, structural navigation, and

*Corresponding author
2024210905@jou.edu.cn (L. Han); 2023122184@jou.edu.cn (J. Song); shuai.chen@jou.edu.cn (S. Chen); zmzhong@jou.edu.cn (Z. Zhong)

cognitive support. With this as our core object of study, we systematically survey and analyze the related visualization and interaction techniques.

This paper systematically reviews visualization and interaction techniques for single-text reading in knowledge-intensive domains such as news reporting and academic research. Adopting a function-oriented perspective, we aim to summarize core methods, key challenges, and future directions. Our main contributions are as follows:

- **Constructing a Dual-Dimension Analytical Framework:** We propose a novel framework that includes a detailed taxonomy of text-related data types (Section 1.2), encompassing structural information, content elements, and user interaction data. This is complemented by a systematic taxonomy of visualization and interaction techniques (Section 1.3), covering Text Presentation Enhancement (TP ●), Information Content Enhancement (IC ●), Layout Optimization (LO ●), and Interaction Enhancement (IE ●). The specific methods under this taxonomy are elaborated in Section 2.

- **Analyzing Applications in Specific Domains:** Based on this framework, we systematically analyze how representative reading assistance systems are applied in key knowledge-intensive domains, including news reporting, academic papers, and textbooks (Section 3). This analysis reveals how different techniques are combined to solve domain-specific reading challenges.

- **Synthesizing Progress and Discussing Challenges:** We present a synthesis of current research progress (Section 4), summarizing key trends and providing an in-depth discussion of core challenges, with a particular focus on user evaluation. Furthermore, we identify promising directions for future research and system design to foster the development of more effective, user-centric text visualization and interaction techniques.

1.1. Related Work

In 2012, Alencar et al. [1] summarized visualization techniques supporting text analysis tasks from recent years. They categorized these techniques into methods for a single text and document collections. These were further subdivided into categories such as displaying content, emphasizing relationships, highlighting temporal evolution, and assisting search engine result processing. Their research covered a range from simple word clouds (e.g., TagClouds and Wordle) to more complex document maps, time-flow visualizations (e.g., ThemeRiver [2] and TextFlow [3]), and network analysis-based document relationship visualizations (e.g., Action Science Explorer [4] and FacetAtlas [5]). Subsequently, in 2015, Jänicke et al. [6] provided a comprehensive review of text visualization techniques in digital humanities over the past decade. They proposed a systematic classification, dividing these techniques into "close reading", "distant reading", and combined approaches. Close reading techniques emphasize in-depth text analysis, preserve text structure, and support word-by-word reading. Distant reading techniques offer macroscopic views, display global text features, and facilitate rapid analysis of large text volumes. Interactive visualization methods combining close and distant reading allow switching between macro and micro perspectives, supporting multi-level text analysis. In a large-scale, task-driven survey, Liu et al. [7] systematically bridged the fields of text visualization and text mining. Their work established a classification framework linked by analysis tasks, with a primary perspective on exploring and understanding large document collections. Taking a meta-analytical approach, Alharbi and Laramee [8] conducted a "Survey of Surveys" (SoS TextVis) to organize and compare the existing survey literature itself. They classified thirteen representative surveys into five categories based on their primary focus: document-centered, user task analysis, multi-faceted, cross-disciplinary, and satellite-themed, thereby integrating different research traditions at a thematic level.

Building upon these and other related research findings, this review aims to provide a more detailed, cutting-edge, and uniquely focused survey specifically for digital reading assistance of a single text. Compared to previous surveys, this review expands and deepens in several key aspects:

1. **More Focused Research Scope**: While Alencar et al.'s review covered a broad range of text analysis tasks, Jänicke et al. focused on digital humanities, and the work by Liu et al. centers on "distant reading" of large collections, this review has a more focused core. We concentrate on digital reading support for a single text, targeting the everyday reading needs of general readers. This positioning allows us to deeply explore key technologies that directly promote efficient understanding and information acquisition, rather than generally discussing text analysis or specific disciplinary research methods.
2. **More Diverse Text Types Discussed**: Although this review focuses on a single text, in terms of text types, we not only draw from digital humanities research but also consider a wider range of texts commonly encountered by general readers. These include academic papers, news reports, and textbooks. The purpose of this expansion

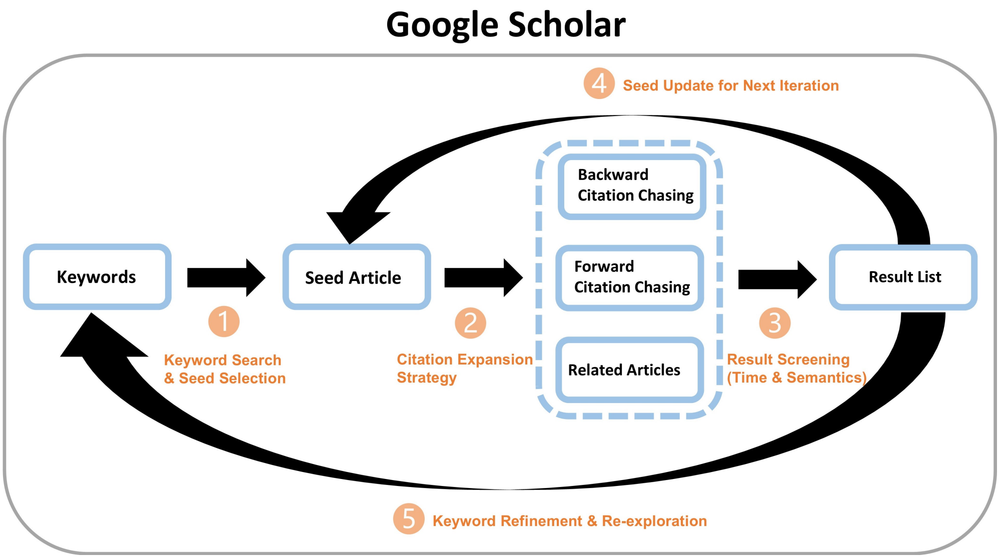

Fig. 1: An iterative literature search and selection process.

is to seek more universally applicable technical solutions within the framework of reading assistance for a single text, not simply to broaden to all text analysis domains.

3. **More Innovative and Practical Classification Perspective**: Alencar et al.'s classification was based on object scale and analysis focus, while Jänicke et al. used the "close/distant reading" paradigm and other surveys approached from task-driven or meta-analytical perspectives. This review proposes a dual-dimension classification system more focused on the functional implementation of reading assistance. First, it is based on data types (subdivided into structural information, content elements, and user interaction data), revealing the information sources for assistive technologies. Second, it is based on visualization and interaction techniques (subdivided into TP ●, IC ●, LO ●, and IE ●), clarifying the specific implementation paths for assistive functions. This classification system aims to more precisely characterize the features, mechanisms, and direct impact of reading assistance technologies on the user's reading process.
4. **More Cutting-Edge Technologies Discussed**: Given the rapid development of information technology, especially after the publication of all these foundational surveys, Artificial Intelligence (AI) technology, particularly Natural Language Processing (NLP) and LLM, has made significant progress in reading assistance. This review specifically focuses on the latest applications of these emerging technologies in single-text reading assistance, such as AI-driven text summary generation, and intelligent term explanation and contextual linking. These aspects were not, or could not be, fully reflected in earlier reviews but are crucial drivers of current reading assistance technology development.

Therefore, this review does not simply reiterate the classification of text visualization techniques. Instead, from the specific application perspective of assisting single-text reading, it systematically reviews and critiques how relevant visualization and interaction techniques help readers improve reading efficiency, enhance the reading experience, and efficiently acquire core content from a single text. We focus not only on visual presentation but also equally on the role of interaction design in the reading assistance process. This unique approach aims to address the shortcomings of existing reviews in specifically tackling the challenges of single-text reading and to provide more targeted references and insights for the future design and development of reading assistance systems.

To comprehensively and systematically collect the literature for this review, we employed an iterative literature exploration strategy centered on Google Scholar. We selected Google Scholar as our primary search engine for two main reasons. First, its broad, cross-disciplinary coverage ensures we capture relevant work from diverse fields such as HCI, NLP, and Visualization. Second, and critically for our methodology, its powerful citation tracking features directly support our Citation Expansion Strategy, specifically the "Cited by" (for forward chasing) and "Related articles" (for semantic expansion) functionalities. This strategy, as illustrated in Fig. 1, combined keyword searches, citation chasing, and recommendations of semantically related literature. The entire process is as follows:

- **Keyword Search & Seed Selection**: In the initial research phase, we used a series of keywords and their combinations on Google Scholar to identify a core set of foundational "seed papers". Our strategy was not to exhaustively count all results, but to ensure comprehensive coverage of key concepts. The search terms were designed to capture multiple facets of the topic, including:
  - *Core Topic*: "single-text reading", "digital reading", "reading assistance".
  - *Key Technology*: "text visualization", "interactive reading", "NLP for reading", "LLM for reading".
  - *Combined Queries*: We systematically combined these terms, for example, ("single-text reading" AND "visualization") or ("reading assistance" AND "interaction techniques").

  This initial phase yielded a robust collection of highly relevant publications that served as the starting point for the next stage.

- **Citation Expansion Strategy**: Focusing on the seed papers, we employed three literature chasing methods to expand the scope of literature:
  - *Backward Citation Chasing*: We reviewed the reference lists of these papers to trace their theoretical foundations and earlier research.
  - *Forward Citation Chasing*: We searched for subsequent studies that cited these papers to grasp the latest developments in that direction.
  - *Related Articles*: We utilized Google Scholar's "Related articles" feature to discover research with potential semantic or thematic connections, compensating for the limitations of keyword matching.

- **Result Screening: Time & Semantic Filtering**: We conducted a multi-pass screening process on the literature obtained from our search and expansion phases. This process was guided by a set of explicit inclusion and exclusion criteria, applied as follows:
  - *Inclusion Criteria*: To be included in this survey, a publication must be a publicly accessible scholarly work that meets the following key criteria:
    - **[I1]** The study must be related to text reading and propose, implement, or evaluate a specific assistance technique or system.
    - **[I2]** The study must be a formal academic publication (e.g., conference paper, journal article, academic thesis, or highly relevant preprint).
    - **[I3]** The publication must be written in English to ensure analysis and comparison within a consistent linguistic context.
  - *Exclusion Criteria*: We explicitly excluded the following types of publications:
    - **[E1]** Works whose core contributions are techniques designed for corpus-level analysis or cross-document comparison. For example, SciDaSynth by Wang et al. [9] is an interactive system that extracts and synthesizes a structured knowledge base from scientific literature at scale. Its primary goal is cross-document analysis, rather than assisting a user in reading and comprehending a single document.
    - **[E2]** Studies that only described algorithms or theoretical concepts without providing a user-accessible technique or interface. For example, QANet by Yu et al. [10] proposes a neural network architecture for machine reading comprehension. While the underlying technology is relevant to reading assistance, it lacks a user-facing tool or interactive interface for human readers, and thus is outside our survey's scope.
    - **[E3]** Publications from before the year 2000, to maintain a focus on contemporary digital reading technologies. For example, Graham's "The Reader's Helper" [11], a 1999 system introducing concepts such as the "Thumbar™" for visual navigation and automatic in-document annotation, is excluded to emphasize more recent advancements.

- **Seed Update for Next Iteration**: From the currently screened results, we selected representative or insightful studies as seed papers for the next round. This initiated a new cycle of citation expansion and screening, iteratively deepening the coverage of the research field.

- **Keyword Refinement & Re-exploration**: During the iterative process, as our understanding of the field deepened and new concepts emerged, we also dynamically adjusted the set of keywords. We then restarted the literature search process to avoid missing potentially important research, continuing until the overall search approached saturation.

Through this method, this paper ultimately selected and included 70 relevant publications. These publications span from 2005 to 2025. The primary focus is on research advancements in single-text digital reading assistance over the past decade. Particular attention has been given to explorations in IE ● and visualization support. Additionally, recent applications of artificial intelligence (AI), natural language processing (NLP), and LLM in this field have also been considered (Table 1).

Table 1
An overview of representative literature on single-text reading assistance, categorized by the subcategories of our proposed taxonomy. A colored cell indicates that a publication's contribution falls within the corresponding subcategory. The color scheme visually groups the four main categories: TP ● and its subcategories (VE ▲, IG ▲, LE ▲) are shaded in red; IC ● and its subcategories (S&O ▲, IS ▲, C&O ▲) are in green; LO ● is in blue; and IE ● and its subcategories (IA ▲, IL ▲, C&N ▲) are in purple. Each publication listed is discussed and cited within the main text of this survey.

Publication	Time	TP			IC			LO	IE		
		VE	IG	LE	S&O	IS	C&O		IA	IL	C&N
Chi et al. [12]	2005	■				■			■		■
Wu et al. [13]	2008					■			■		■
Strobelt et al. [14]	2009	■	■	■	■	■		■	■		■
Oelke et al. [15]	2011	■	■	■	■	■		■	■	■	■
Golovchinksy et al. [16]	2011							■	■		■
Tashman and Edwards [17]	2011							■	■		■
Stoffel et al. [18]	2012	■	■	■	■	■		■		■	
Hinckley et al. [19]	2012				■			■	■		■
Pearson et al. [20]	2012								■		■
Chaudhri et al. [21]	2013	■	■		■	■		■	■	■	■
Singh [22]	2013	■	■						■		■
Kwon et al. [23]	2014	■	■	■	■	■		■	■	■	■
Stoffel et al. [24]	2014	■	■	■		■			■		■
Koch et al. [25]	2014	■	■			■	■	■			■
Gao et al. [26]	2014		■								
Wecker et al. [27]	2014	■				■					
Kim et al. [28]	2015	■	■	■		■		■	■	■	■
Gold et al. [29]	2015	■			■						■
El-Assady et al. [30]	2016	■	■	■	■	■		■	■	■	■
Abekawa and Aizawa [31]	2016			■	■			■	■		■
Nicolas [32]	2016	■							■	■	
Wang et al. [33]	2016		■					■			■
Mehta et al. [34]	2017	■	■		■	■		■	■		■
Yang et al. [35]	2017	■									
Badam et al. [36]	2018	■	■	■							
Kim et al. [37]	2018			■					■		
Metoyer et al. [38]	2018		■								
Sperrle et al. [39]	2019	■		■	■	■				■	■
Kiesel et al. [40]	2020		■								■
Subramonyam et al. [41]	2020		■						■		■
Sevastjanova et al. [42]	2021	■	■	■	■	■		■	■	■	■
Head et al. [43]	2021	■		■		■		■			■
Sultanum et al. [44]	2021			■				■			■
Wang et al. [45]	2021				■		■				■
Fok et al. [46]	2022	■			■			■			■
Han et al. [47]	2022				■			■			■

Publication	Time	TP			IC			LO	IE		
		VE	IG	LE	S&O	IS	C&O		IA	IL	C&N
Head et al. [48]	2022			■							
Kang et al. [49]	2022					■	■		■		
Peng et al. [50]	2022									■	
Rachatasumrit et al. [51]	2022					■	■				■
August et al. [52]	2023	■	■							■	■
Cai et al. [53]	2023		■								
Chee Chang et al. [54]	2023	■			■	■	■				■
Chen et al. [55]	2023	■			■				■	■	■
Chulpongsatorn et al. [56]	2023		■	■				■		■	
Fok et al. [57]	2023	■	■		■					■	
Kang et al. [58]	2023	■	■		■		■				■
Kim et al. [59]	2023	■				■		■			
Masson et al. [60]	2023		■					■			■
Milbauer et al. [61]	2023	■				■					■
Palani et al. [62]	2023	■			■	■	■				■
Park et al. [63]	2023	■			■	■	■				■
Richards Maldonado et al. [64]	2023				■					■	■
Wana and Kim [65]	2023					■					
Zhang et al. [66]	2023	■	■	■	■			■	■		■
Zyska et al. [67]	2023								■		■
Fu et al. [68]	2024		■								■
Kambhamettu et al. [69]	2024	■		■	■						■
Kim et al. [70]	2024	■		■	■		■	■			■
Liu et al. [71]	2024	■			■	■					■
Melin-Higgins [72]	2024				■	■					
Ng et al. [73]	2024	■									
Pitre and Luther [74]	2024		■		■	■				■	
Singh et al. [75]	2024	■				■				■	
Zhang et al. [76]	2024	■	■	■		■			■	■	■
Gunturu et al. [77]	2024		■		■	■		■		■	■
Gunturu et al. [78]	2024		■	■				■		■	
Zou et al. [79]	2025		■	■				■			■
Shao et al. [80]	2025		■	■		■					
Dück et al. [81]	2025	■			■	■		■	■	■	■

1.2. Taxonomy of Data Types

There are many types of texts. From a functional perspective, Reiss [82] broadly classified text types into three categories: (1) Informative texts, which aim to convey facts such as information, knowledge, and opinions. The focus of these texts lies in their content and subject matter; (2) Expressive texts, which emphasize the creative construction of the text and the aesthetic aspects of language, highlighting both the author and the text itself; (3) Vocative texts, which seek to persuade the reader or recipient to take a specific action, prompting behavioral responses. Such texts are often dialogic in nature and focus on appeal.

Texts contain a wide range of data, and the type and structure of data vary significantly across different text types. In this review, we focus on informative texts. In this subsection, we synthesize, summarize, and categorize the types of text-related data emphasized in the surveyed literature on reading assistance. We broadly classify text data into three categories: structural information, content elements, and user interaction data (see Fig. 2).

1.2.1. Structural Information

Structural information refers to the inherent, explicit or implicit structural cues within a text. These cues reflect its content organization and logical framework. While not usually directly involved in expressing semantic content, structural information plays a crucial role. It helps readers understand the article's outline, grasp key content, and navigate effectively. Common structural information includes chapter hierarchies, paragraph divisions, headings,

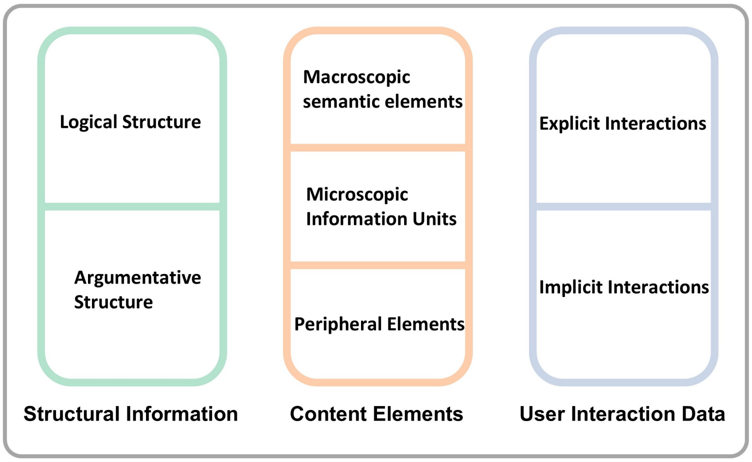

Fig. 2: Taxonomy of Text-Related Data Types in Reading Assistance

summaries, introductions and conclusions, and argument-evidence relationships. Depending on the text type and expressive purpose, structural information can be broadly divided into logical structure and argumentative structure.

Logical structure mainly reflects the hierarchical organization and narrative sequence of the content. It is common in informative texts such as academic papers, expository essays, and news reports. It is usually explicitly marked by chapter titles, paragraph summaries, etc. This type of structure is often used to create document thumbnails or hierarchical tables of contents. These help readers quickly understand content distribution and navigate the entire text. To enhance users' grasp of high-level text structure, researchers have explored using automatic summarization techniques. These extract paragraph gist and generate structured thumbnails [76]. This helps in quickly assessing text readability and value. Another method involves dynamically generating hierarchical views based on paragraphs or chapters. This allows users to flexibly switch between overview and detail, thereby improving reading efficiency for long texts [25].

Argumentative structure is more common in opinion-expressive texts like argumentative essays and review articles. It describes how an author supports or refutes a central claim through a series of arguments and evidence. This structure typically has clear hierarchies and causal logic. To help readers more clearly identify the article's argumentation path and logical relationships, researchers have proposed using structured tree diagrams. These display the relationship between arguments and evidence, classifying them into "Major Claim", "Claim", and "Premise". These are further subdivided into supporting and refuting types. The tree diagram's structure can clearly show the logical connections between various argumentative components, thus enhancing the comprehensibility of the argumentative structure [40]. In news texts, argumentative structure often revolves around a single argument. To enhance multi-angle understanding of content, researchers have tried to introduce external resources to automatically obtain opposing viewpoints [52]. They also use LLM to generate counter-arguments, thereby constructing a more comprehensive view of the argumentative structure [74].

In terms of visual presentation, structural information is usually displayed as document thumbnails, content index trees, interactive tables of contents, etc. This aims to improve the user's grasp of the document's overall structure and logical flow and navigation efficiency. The visualization of logical structure focuses on clear hierarchy and quick positioning. The visualization of argumentative structure emphasizes the clear presentation of logical relationships between components.

1.2.2. Content Elements

Content elements refer to the specific semantic information embedded within a text's structure. They are central to understanding what an article is about in the context of reading assistance. In contrast to structural information, which reflects how content is organized, content elements represent the concrete realization of that content within the structure. These include semantic components such as the article's theme, viewpoints, arguments, methodologies, and conclusions. Extracting and visualizing content elements typically requires contextual understanding, domain knowledge, or external tools, and is a key approach to enhancing reading comprehension efficiency. Depending on the text genre and reading objectives, content elements can be broadly categorized into three types: macroscopic semantic elements, which capture the core meaning at the paragraph or document level; microscopic information units, which refer to fine-grained yet critical elements like terms, formulas, and data; and peripheral elements, which include supplementary materials such as references, charts, and appendices.

Macroscopic semantic elements denote core information present at the article or paragraph level. For instance, in academic papers, these may include the research background, methods, and conclusions; in news articles, core events, timestamps, locations, and participants; and in argumentative essays, main viewpoints and supporting arguments. As this type of information is often scattered across the structure, techniques such as automatic summarization and text clustering are commonly used for extraction. In recent years, generative approaches based on LLM have gained prominence. These methods can produce concise summaries while preserving semantic integrity, thereby allowing readers to quickly grasp the core content of a text [72]. To further improve information acquisition efficiency, some systems also support visual summarization that integrates text with graphical elements, linking semantic content to charts, paragraphs, and images for enhanced presentation [76].

Microscopic information units refer to fine-grained elements that carry limited standalone information but are highly valuable for comprehension. Examples include terms, formulas, symbols, numerical data, and temporal expressions. These elements often exhibit high domain specificity or strong contextual dependence, posing challenges for general readers. To reduce cognitive load, researchers have proposed embedding assistive functionalities within the reading process, such as term definitions, formula-text alignment, and explanatory support for numerical data. For example, Abekawa and Aizawa [31] proposed presenting term annotations alongside the main text, while Head et al. [48] investigated the use of highlighting to align mathematical formulas with their corresponding natural language descriptions, thereby enhancing comprehension efficiency. These approaches contribute to the explicit representation of microscopic elements and improve the interpretability of the text.

Beyond the primary semantic content, reading assistance systems also emphasize peripheral elements that convey implicit or supplementary information, such as references, charts, and appendices. These components often contain rich contextual cues and external evidence. Researchers have explored structuring and visualizing these elements to help readers build knowledge association graphs and better understand auxiliary materials [59]. To support multimodal comprehension, some systems incorporate images, videos, or database links, thereby enhancing the intuitiveness and cross-validation capabilities of textual content [28].

1.2.3. User Interaction Data

User interaction data refers to behavioral traces generated as readers engage with a text. This data not only reflects the reader's focus and comprehension path, but also reveals potential difficulties and cognitive strategies encountered during reading. While the text itself conveys "what the article is about", interaction data primarily reflects how readers use and interpret the content. It serves as a critical foundation for enabling intelligent and personalized support in reading assistance systems. Based on how the behavior is generated, user interaction data can be broadly categorized into explicit and implicit interactions.

Explicit interaction involves deliberate actions taken by users to achieve specific reading goals. This type of data directly reflects user intent and needs. For example, users may highlight text, add annotations or voice comments to facilitate understanding or record thoughts [67, 16]. They may also sketch structural diagrams or organize visual notes through drag-and-drop interactions [41]. Other common forms include selecting and excerpting content—such as copying text snippets, dragging them into a visual workspace [19], creating structured research threads [49], or establishing bidirectional links between excerpts and source text [17]. Explicit interaction also encompasses user-initiated operations such as navigating via tables of contents or thumbnails, switching information views [25, 76], manipulating chart components [60], retrieving term definitions [31, 43], highlighting descriptions of mathematical expressions, viewing embedded videos [59], requesting chapter or full-text summaries [72], and interacting with built-in Q&A systems to obtain answers [76, 21] or explore opposing arguments [74, 52]. All these behaviors constitute explicit interaction data.

It is worth noting that some navigation behaviors may be classified as either explicit or implicit, depending on context. For instance, clicking a table of contents to locate a section reflects clear intent and thus counts as explicit interaction. In contrast, frequent backtracking or switching between paragraphs may indicate confusion, review, or exploration. In such cases, the underlying intent is not immediately clear, and these behaviors are better considered implicit.

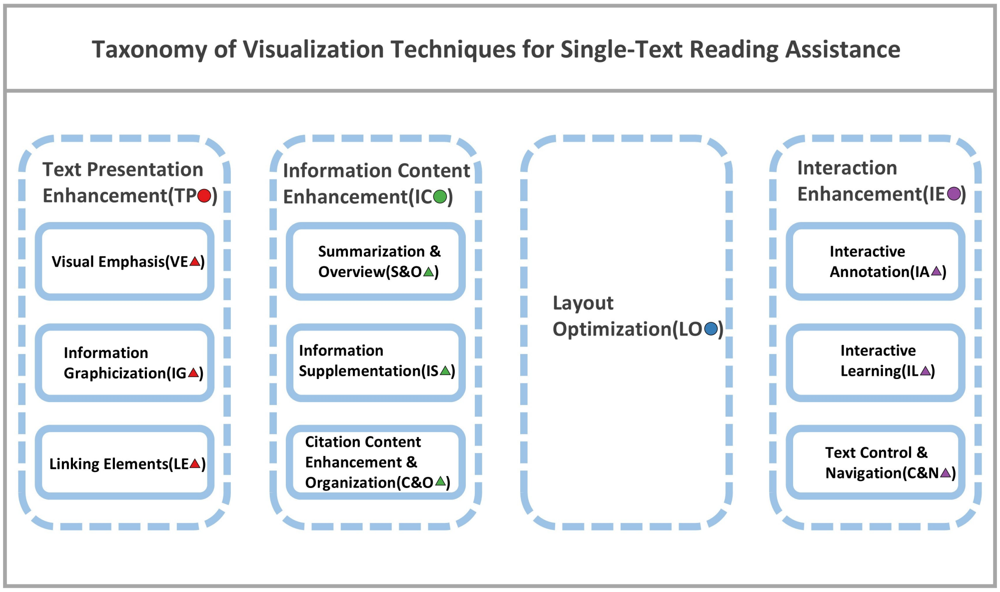

Fig. 3: Classification framework of visualization techniques for single-text reading assistance. It includes four major categories and their corresponding subcategories: (1) *TP* ●, including subcategories such as VE ▲, IG ▲, and LE ▲; (2) *IC* ●, including S&O ▲, IS ▲, and C&O ▲; (3) *LO* ●; (4) *IE* ●, including IA ▲, IL ▲, and C&N ▲. Detailed discussions of each subcategory can be found in Section 2.

Implicit interaction refers to passive behavioral traces generated during natural reading, from which user intent must be inferred. These include reading trajectories and durations (e.g., scrolling behavior, dwell time, frequency of revisits), which can be used to assess attention or generate personalized summaries [55]. Navigation paths, such as hyperlink click sequences, citation browsing orders [54], view-switching patterns, and backtracking trails [43], are also informative. Additional indicators include mouse movements, eye-tracking data, and in-document or search engine queries, which can reflect attention distribution and cognitive load.

Together, explicit and implicit interaction data provide a comprehensive basis for understanding reader behavior and personalized needs. By integratively analyzing these multidimensional signals, reading assistance systems can more accurately infer users' reading intentions, knowledge levels, and cognitive states. This enables adaptive support strategies such as dynamic content reorganization, background knowledge recommendation, and targeted interactive guidance—ultimately enhancing reading efficiency and user experience.

1.3. Taxonomy of Visualization Techniques

In the context of digital reading from a single text, researchers have developed a variety of visualization and interaction techniques to provide reading assistance. Based on a synthesis of existing work, we classify these techniques into four main categories: TP ●, IC ●, LO ●, and IE ● (see Fig. 3). These four categories optimize the reading experience from different dimensions, aiming to improve the efficiency of content acquisition, depth of understanding, and cognitive effectiveness for readers.

- **TP** ●. This category focuses on improving the visual presentation and interaction mechanisms of textual content. It enhances readability, interpretability, and information delivery efficiency by using visual design strategies such as font adjustment, color coding, layout design, highlighting, and graphic representation. The goal is to reduce users' cognitive load and help them quickly and accurately grasp the content and uncover deeper meaning. For example, using color to emphasize key information [73] or adjusting font styles to improve legibility [53] are typical approaches. TP ● can be further divided into three subtypes: Visual Emphasis (VE ▲) (highlighting key content through visual features), Information Graphicization (IG ▲) (transforming data in the text into visual forms), and Linking Elements (LE ▲) (explicitly revealing connections between different text elements).

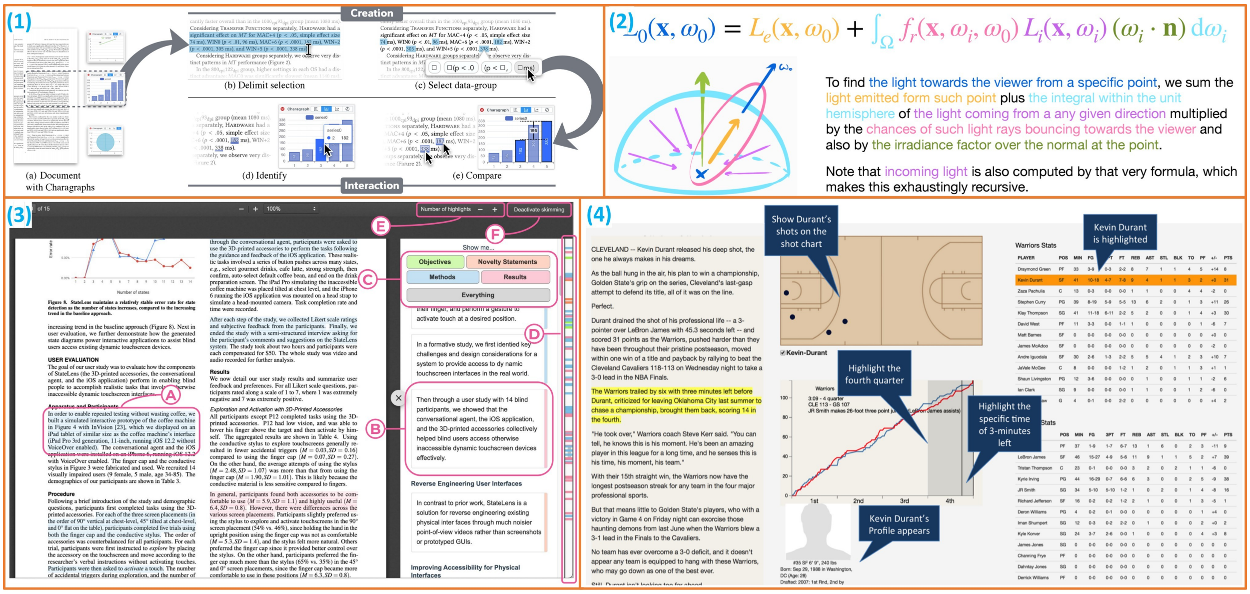

Fig. 4: Visual techniques for TP ● in single-text reading. (1) CHARAGRAPH visualizes numerical data in situ within the text [60]. (2) A method that uses color to highlight key parts of mathematical expressions, helping readers intuitively understand them and relate them to associated charts and textual descriptions [48]. (3) The user interface of Scim, which classifies and marks sentences with four colors based on their importance [46]. (4) A system that visualizes narrative elements in text [38].

- **IC** ●. This category aims to improve the expressiveness and organization of content elements by supplementing, extracting, interpreting, or restructuring the original text. It helps users identify key points, resolve comprehension difficulties, and build a clearer knowledge structure, facilitating the transition from surface reading to deep understanding. Key techniques include Summarization and Overview (S&O ▲) (e.g., automatic generation of article gist [66]), Information Supplementation (IS ▲) (e.g., explanations of technical terms, integration of background knowledge or multimedia), and Citation Content Enhancement and Organization (C&O ▲) (e.g., improving the accessibility and comprehensibility of cited references).

- **LO** ●. These techniques aim to improve the readability and cognitive efficiency of documents by optimizing visual layout and structural organization. Key strategies include combining text with graphics, integrating related information, and supporting structured navigation to reduce cognitive interruption and enhance information exploration. For instance, arranging text and charts in embedded or parallel formats [60], or dynamically generating hierarchical views to support flexible switching between overview and detail [25], are effective LO ● practices.

- **IE** ●. This category enhances users' engagement, comprehension, and control over the reading process by strengthening user interaction with the text and the reading system. It supports personalized reading strategies, enabling users to adjust information presentation, actively construct understanding, and explore content. Common techniques include Interactive Annotation (IA ▲) (e.g., highlighting, graphic notes), Interactive Learning (IL ▲) (e.g., intelligent Q&A, guided exploration), and Text Control and Navigation (C&N ▲) (e.g., gesture-based text folding [17], focus management, and flexible view switching).

These four categories each emphasize a distinct aspect: TP ● focuses on optimizing visual display; IC ● addresses enriching and clarifying content elements; LO ● emphasizes structural organization; and IE ● highlights users' active engagement and control. Together, they form the core framework of visualization and interaction techniques in reading assistance systems. In the following Section 2, we discuss in more detail the specific methods and subtypes under each category.

2. Visualization Technologies

2.1. TP ●

TP ● refers to using various visual techniques and design strategies to optimize the way text information is presented. This aims to improve its readability, comprehensibility, and information transfer efficiency. Its core goal is to reduce user cognitive load through intuitive and efficient visual means. This enables users to grasp text content more quickly and accurately, gain insight into deeper meanings, and enhance overall reading effectiveness. This section will discuss research progress in TP ● technology in the following three directions: (1) VE ▲, which highlights key content of the text through visual means such as highlighting and color-coding; (2) IG ▲, which graphically presents data, structure, and semantic relationships embedded in the text; (3) LE ▲, which explicitly reveals semantic associations between different text elements (such as main text, charts, and formulas).

2.1.1. VE ▲

VE ▲ primarily highlights important parts of the text through visual means, such as highlighting and color coding. It often combines importance or relevance-based grading to guide readers in quickly locating key information or perceiving specific text features, thereby improving reading efficiency. In reading assistance research, common practices include highlighting key paragraphs or sentences and using different colors or saturation levels to indicate relative importance.

As early as 2005, Chi et al. [12] proposed the ScentHighlights system, which applied word co-occurrence and spreading activation to identify concept terms related to the user's search keywords and highlighted relevant sentences. The system not only marked directly matched sentences but also used gray tones to indicate semantically related content, helping users locate information more efficiently in digital reading. Later, Yang et al. [35] introduced the HiText method, which added a "dynamic marking" mechanism to previous techniques. When the user moved the cursor, the system dynamically highlighted sentences highly relevant to the pointer's position, using gradient colors to indicate different levels of relevance. HiText was designed for HTML documents; it first extracted the main content from the DOM tree and then applied deep learning models to evaluate sentence importance, visually enhancing content hierarchy. Additionally, the system could automatically annotate keyphrase terms based on topical relevance, further supporting information location. Fok et al. [46] developed the Scim system for scientific papers. It used pretrained language models and semantic similarity analysis to identify important sentences, which were classified into four semantic roles: research goal (green), novelty (orange), method (blue), and result (red). To improve navigation, Scim introduced the Highlight Browser panel, which listed all highlighted sentences in document order. Users could click to jump to the original context (see Fig. 4 (3)). Unlike HiText, Scim supported PDF documents rather than HTML pages, showing its adaptability to different document formats. VE ▲ can also be used to reveal macroscopic textual structures. The Lexical Episode Plots [29] are a representative method for this. It algorithmically detects and highlights regions of unusually high keyword density within a text using visual markers, such as colored bars. This approach externalizes the thematic rhythm and topic shifts of the text into a visual pattern, guiding readers to quickly locate core segments in a long document.

Beyond emphasizing objective importance or relevance, VE ▲ has also been used to enhance the perception of specific text attributes. For instance, the Semantize system by Wecker et al. [27] analyzed sentiment polarity, emotional terms, and subjective sentences. It applied different colors and font styles to highlight these features, helping readers more intuitively perceive the strength and orientation of emotional expression. VE ▲ is also applied to readability analysis to support text revision. The VisRA system [15] calculates multi-dimensional readability metrics and uses color-coding to directly highlight difficult-to-understand paragraphs and sentences, providing intuitive guidance for the author's refinement process.

VE ▲ techniques have evolved from keyword matching to semantic analysis, from static highlighting to dynamic grading, and from simple text to complex document formats. Their application has also expanded from highlighting objective key information to enhancing perception of specific attributes such as sentiment.

2.1.2. IG ▲

IG ▲ aims to apply diverse visual design and technical methods to transform and optimize the abstract information embedded in the text. This includes numerical data, logical structures, narrative flows, and the visual presentation of the text itself. The main goal is to help readers better understand complex content, identify deeper relationships, and enhance the overall reading experience and cognitive efficiency through intuitive graphical representation, clear structural display, and optimized visual styles.

In data visualization, researchers aim to convert textual numerical information (e.g., time, amount, proportion) into visual charts to enhance clarity and comparability. Charagraph [60] allows users to select numbers during reading and automatically generates editable, interactive charts embedded in the textual context, integrating visualization into the reading flow (see Fig. 4 (1)). Extending this idea, GistVis [79] enables word-scale visualizations by using LLM to automatically extract data insights (e.g., comparison, trend, extremes) and render micro charts (e.g., mini bar charts, sparklines) directly next to relevant words. Compared to Charagraph's paragraph-level embedding, GistVis offers finer-grained, seamless integration, improving immediate comprehension and reducing cognitive load. Gao et al. [26] proposed the NewsViews system, which automatically identifies geographic entities in news text and generates geovisualization maps to reveal spatial trends and comparisons, enriching readers' understanding of geographic context.

For narrative element visualization, Metoyer et al. [38] used natural language processing techniques to automatically extract key components such as events, characters, and time. These elements are then linked with visual components, including timelines and event flow diagrams. In sports journalism, for example, the system can align important game data along a timeline, helping readers intuitively understand the rhythm and turning points of the event (see Fig. 4 (4)). Narrative Player by Shao et al. [80] transforms paragraph-level data narratives from text and tables into dynamic videos. When the narrative is co-constructed by multiple participants, as in a transcribed discourse, graphicization also shifts to revealing its internal interactional relationships and argumentative structures. For example, the VisArgue framework [83] abstracts co-occurrence relationships between participating entities into a node-link diagram and encodes argumentative features as glyphs; whereas VisInReport by Sevastjanova et al. [42] presents the dynamic process of argumentative exchanges and topic evolution through animated speaker activity views and Topic Bars. Taking a more metaphorical approach, the ConToVi system [30] maps topics onto a circular "discussion floor" and uses animated trajectories to represent how speakers move between them over time.

In terms of structural visualization, Kiesel et al. [40] proposed using tree diagrams to display the hierarchical relationships between arguments and evidence. This method supports the understanding of complex logical reasoning. The VarifocalReader system [25] provides multi-level navigation through nested views of chapters and paragraphs, enabling users to switch flexibly between overview and detail. Zhang et al. [76] further introduced a thumbnail-based structural overview that serves as a fast entry point for navigation and improves the reader's sense of control. However, while traditional thumbnails are effective for navigation, their textual content is often rendered illegible due to their small scale. To address this issue, Stoffel et al. [18] proposed a document thumbnail technique with variable text scaling. It enlarges keywords and their context while preserving the overall page layout, transforming the thumbnail into a "focus+context" overview that is suitable for both navigation and the inspection of key information.

Visual style optimization focuses on enhancing readability and visual comfort. The THERIF method proposed by Cai et al. [53] uses a readability-aware theme generation mechanism to automatically adjust parameters such as font, font size, line spacing, and color contrast. This helps meet the personalized visual needs of different users, including those with visual impairments. Enhancing visual style not only improves aesthetic quality but also reduces reading fatigue.

In summary, IG ▲ either transforms implicit data, structure, and narrative content into visual representations or improves the visual presentation of the text. These techniques aim to lower cognitive load and improve readers' comprehension and overall comfort during the reading process.

2.1.3. LE ▲

LE ▲ aims to explicitly establish and strengthen the semantic connections between different parts of a text. This is achieved through techniques such as interactive highlighting, synchronized display, and on-demand visualizations. The core goal is to reduce information silos and support efficient cross-element integration and deeper understanding, thereby lowering cognitive load and improving reading efficiency.

Many studies in reading assistance focus on enhancing the presentation of relationships among different elements in the text. One common direction is linking the main text with data tables. Kim et al. [37] proposed an interactive document reader that automatically links the main text with relevant table cells. When users select a text segment, the associated table cells are highlighted. Conversely, selecting a cell also highlights the related text, enabling direct comparison between textual and tabular information. The Elastic Documents approach by Badam et al. [36] also strengthens the linkage between text content and tabular data. Unlike Kim et al., this method goes beyond static

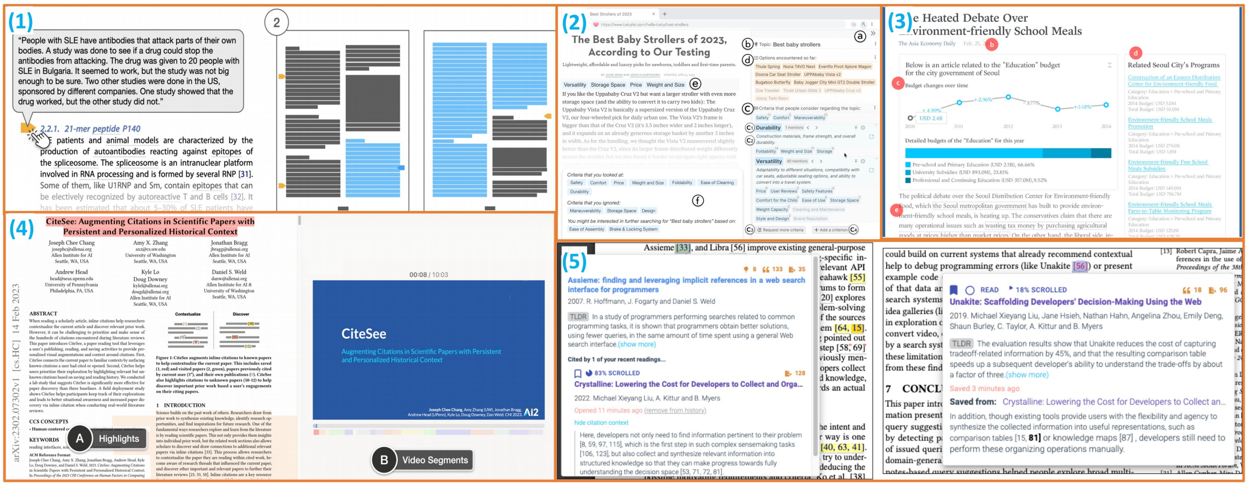

Fig. 5: IC ● Approaches in Single-Text Reading. (1) Paper Plain: Provides concise section summaries via clickable tabs for cross-page content overview [52]. (2) Selenite: Offers global and local topic overviews with dynamic recommendations for webpages [71]. (3) Factful: Verifies budget info in news using public data, supplying contextual background [28]. (4) Papeos: Links PDF paragraphs to video segments with synchronized, color-coded transcript-video views [59]. (5) CiteSee: Uses color coding to track citation exploration and reveals contextual info on click [63].

highlighting. It generates visualizations based on user interactions and focus. Specifically, it first extracts text and tables into separate views, builds links using keyword matching, and then creates charts such as line or bar graphs in a dedicated panel. These charts respond to user focus, allowing interactions such as text selection to highlight related data and filtering of table attributes. This on-demand visualization provides a more intuitive way to understand the context of textual data. However, its effectiveness depends heavily on the accuracy of keyword extraction and the current support for chart types remains limited. For linking formulas, text, and visual charts, the work by Head et al. [48] provides a notable example. Their method highlights specific parts of mathematical formulas and matches them with corresponding textual descriptions using the same color. It also highlights related components in associated charts with the same color coding (see Fig. 4 (2)). This multimodal synchronized highlighting enhances the semantic coherence among formulas, textual explanations, and visual representations, helping readers better understand their internal connections.

These studies demonstrate the potential of LE ▲ to enhance relationships among content elements. From bidirectional highlighting to on-demand visualizations and multimodal synchronization, these methods aim to address the problem of fragmented and implicit relationships in textual information. While challenges remain in link accuracy and interaction design, such approaches have shown clear value in supporting the understanding of complex documents.

In summary, TP ● serves as a foundational layer in reading assistance. Its core objective is to reduce the user's cognitive load by optimizing the visual presentation of information. It achieves this through three main approaches. VE ▲ guides the reader's attention to core content using highlighting and color-coding. IG ▲ enhances comprehension by transforming abstract data, structures, or narratives into intuitive charts. LE ▲ reduces information fragmentation by explicitly connecting different parts, such as text, charts, and formulas. Together, these techniques significantly improve readability, information acquisition efficiency, and the overall reading experience.

2.2. IC ●

In digital reading environments, improving the clarity and organization of text content has become an important approach to support user comprehension and cognitive depth. IC ● refers to the supplementation, refinement, explanation, or restructuring of the original text. This process helps create a more coherent and semantically clear reading experience. It allows users to locate key points more efficiently, overcome comprehension barriers, and construct clear knowledge structures. Compared to TP ●, which emphasizes visual presentation, IC ● focuses more on supporting deep understanding and knowledge organization. It aims to facilitate the transition from surface-level reading to deeper comprehension.

This section reviews major directions in the research on IC ●: (1) S&O ▲, which helps users grasp the main points of a text and build an overall understanding through automatic summarization and structured overviews; (2) IS ▲, which reduces comprehension difficulties and expands contextual understanding by integrating term explanations and external information; (3) C&O ▲, which improves the accessibility and organization of citation-related content to support academic context tracing and analysis. The following subsections introduce representative approaches and systems in each of these areas.

2.2.1. *S&O* ▲

In academic reading, summarization is one of the most common forms of IC ●. By condensing the original text and removing redundant information, a summary highlights the purpose,methods, results, and conclusions of a study. This helps readers quickly grasp the main points and decide whether to read the full text. S&O ▲ are particularly useful for long or information-dense articles, as they serve as a means of knowledge compression and significantly improve reading efficiency.

In recent years, many reading assistance systems have integrated automatic summarization and structured overview features. For example, the Paper Plain system [52] provides a "Section Gists" function, allowing users to click tags next to section titles to view concise summaries of each section (see Fig. 5 (1)). These summaries are generated by GPT-3 [84], using a combination of paragraph lead sentences to represent content across sections. This enables readers to quickly assess whether a section is worth reading in detail and helps them focus on relevant parts to improve overall efficiency. However, automatically generated summaries may contain hallucinated or biased information, which can lead to misinterpretation. To address this issue, the Traceable Text system [69] generates summaries alongside traceable links to their corresponding original paragraphs. Users can view the original text and the summary side by side, with key sentences highlighted for comparison. This design supports quick validation and fosters critical understanding of the summarized content, enhancing both accuracy and transparency. Going further, the Selenite system [71] introduces a more detailed mechanism for S&O ▲ (see Fig. 5 (2)). It provides a "global overview" that identifies the main topics of a web page and common standard options, helping readers form a basic understanding. Meanwhile, the "local overview" offers paragraph-level information related to specific standards or details, assisting users in deciding whether to continue reading. The system also includes a dynamic recommendation feature that suggests related content based on user interaction data, reducing the risk of missing important information. This personalized navigation approach is especially useful for users unfamiliar with a domain, offering more targeted exploration and decision-making.

To provide a more intuitive overview, some research also explores multimodal summaries that combine text and images. The "Document Cards" [14] are an example of this, automatically generating a compact, card-like summary for a single document that integrates both core key terms and representative images. This multimodal condensation offers an effective way for users to quickly assess a document's content.

S&O ▲ play a crucial role in supporting comprehension, saving time, and improving navigation efficiency. From compressed expression to content traceability, and then to structured guidance, these systems progressively extend the functional boundaries of IC ●. They improve both the usability and reliability of complex information access and validation.

2.2.2. *IS* ▲

To help readers better understand texts that contain technical terms or complex content, researchers have proposed various IS ▲ strategies. These approaches reduce the reading threshold by explaining terms and providing contextual background. At the same time, they expand the informational scope of the text by integrating external resources such as multimedia content, external datasets, and multi-source texts.

Term explanation is a key method for improving the readability of professional texts. When only a few technical terms appear in a sentence, readers can often infer their meaning from the context. However, when multiple specialized terms appear together, especially in highly technical texts such as medical papers, readers without the relevant background knowledge may find it difficult to understand. Looking up terms one by one can interrupt the reading process and reduce efficiency. To address this, systems such as Paper Plain [52] and SideNoter [31] integrate external resources like Wiktionary or Wikipedia to provide standardized and authoritative definitions with textual and visual explanations. Paper Plain presents definitions through tooltip boxes on hover, while SideNoter directly links terms to Wikipedia entries. These approaches are particularly suitable for general domain-specific terms and help readers understand them within a broader context. For locally defined terms, which are introduced and defined only within a specific paper, external resources usually offer little support. To address this issue, the ScholarPhi system [43] introduces a context-based method for extracting in-text definitions. The system automatically links each term to its first occurrence and highlights the sentence containing the definition. When the user clicks on a term, the interface navigates to its definition in the text. Compared to external general definitions, these in-context explanations more accurately reflect the specific meaning of the term in the current paper. This method is especially effective for custom concepts or mathematical formulas defined by the author.

In addition to term explanation, many systems enhance the expression of the text by incorporating diverse external information content. Some systems introduce multimedia content to improve clarity and engagement. For example, the Papeos system [59] (see Fig. 5 (4)) links segments of a paper to related conference video clips, supporting reader comprehension through both visual and auditory channels. The WikiTUI system [13] projects relevant images and videos onto printed books to enable interactive presentation of digital content. Other systems focus on fact verification and data supplementation. Factful [28] verifies budget-related statements in news articles using open government data and provides background context (see Fig. 5 (3)). Aletheia [68] structures data claims found in text and compares them with authoritative datasets. It then presents the verification results through IG ▲, helping readers assess their credibility.

Regarding chart interpretability, the system proposed by Wang and Kim [65] can automatically extract "data facts" such as trends, values, and rankings from charts. It then generates customized textual descriptions based on user preferences, which supports clearer and more personalized data understanding. This is particularly useful for data-intensive academic articles.

Some studies also aim to support viewpoint expansion and comparison. For example, NEWSSENSE [61] compares how different news outlets report the same events, revealing consensus and disagreement among sources. ArguMentor [74] uses large language models to generate counterarguments and neutral information for a given claim, encouraging readers to consider multiple perspectives. Similarly, to provide a broader context, the "Enhanced News-reading" system by Stoffel et al. [24] allows users to select keywords while reading a news article to instantly retrieve related social media discussions. Taking this a step further, other systems provide evidential support by retrieving real-world claims from existing corpora. An example is the NEEDLE system by Dück et al. [81], which utilizes Natural Language Inference (NLI) to automatically find sentences that support or contradict a user-selected claim. This approach deepens information supplementation from viewpoint expansion to active reasoning assistance for hypothesis validation.

These systems integrate video, image, data, and multi-source content to significantly expand the explanatory range and cognitive depth of the original text. They further promote deep comprehension and critical thinking during reading.

2.2.3. C&O ▲

In academic papers, inline citations help readers understand the research background and related literature, offering theoretical support for the work. However, citations are often numerous and briefly mentioned, making it difficult for readers to quickly identify the most relevant sources for their own research. Therefore, many studies focus on C&O ▲ to improve the accessibility, interpretability, and contextualization of citation information.

The CiteSee system by Chee Chang et al. [54] and the QuickRef system by Park et al. [63] offer different features to support reading assistance for citations. CiteSee uses colors and markers to distinguish citations that are read, saved, cited, or already published, helping readers quickly locate relevant references during the reading process (see Fig. 5 (5)). When a citation is clicked, CiteSee displays a personalized "Paper Card" that shows metadata and user interaction history with the cited paper. The system also evaluates the relevance of citations based on citation counts and user interactions, using VE ▲ such as highlighting to reveal connections to the current paper. QuickRef focuses on presenting detailed metadata of cited works, such as titles, authors, and publication venues. These elements help readers quickly understand the background of a citation. In addition, QuickRef supports S&O ▲ by providing structured abstracts and charts that highlight the key content of cited papers. It also uses sentence similarity matching to identify the most relevant citation snippets, reducing the time and effort needed to consult source documents. Taking a different approach to providing on-demand context, CiteRead [51] integrates localized citation contexts from subsequent papers directly into the margins of the reference paper, allowing readers to see commentary about a specific passage from follow-on work.

Other systems aim to extract key information from citations and organize them systematically to help readers construct a coherent research context. For example, the Threddy system proposed by Kang et al. [49] allows users to highlight text segments of interest while reading (such as research methods, results, conclusions, and related citations). The system automatically extracts citation information from these segments and stores it in a "holding tank" for later management. Users can save these snippets as new research threads or merge them with existing ones,

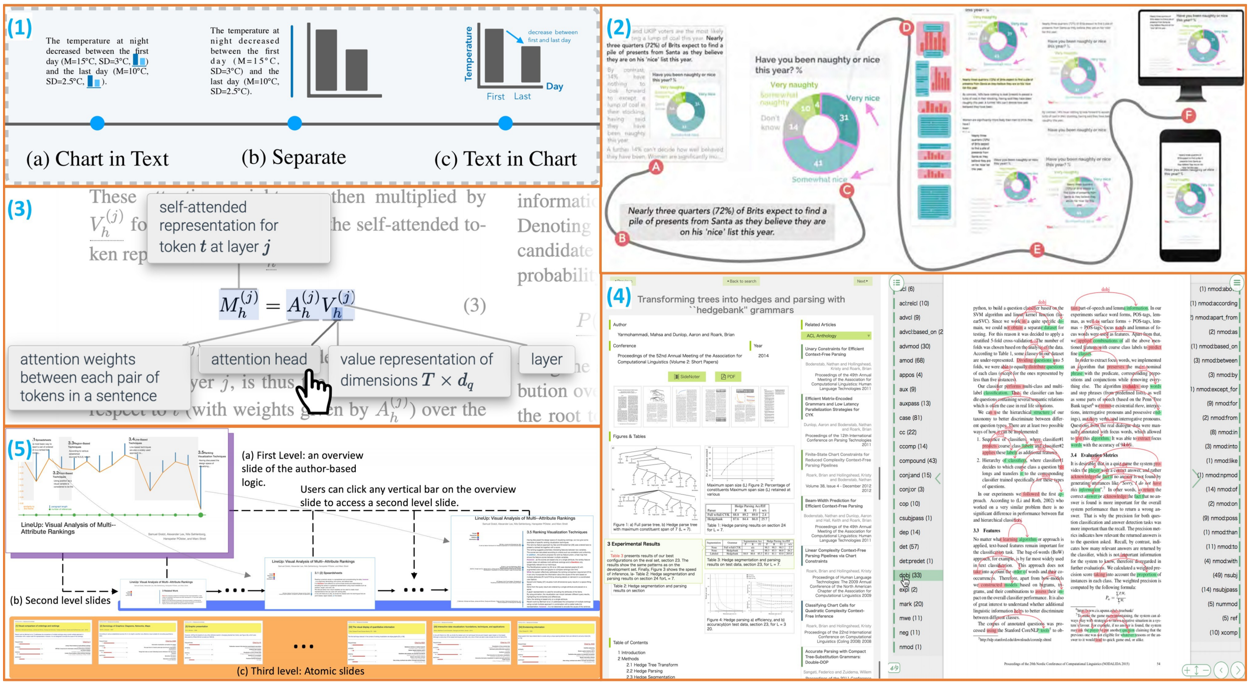

Fig. 6: LO ● Techniques in Single-Text Digital Reading. (1) Charagraphs: Proposes three layout modes to arrange text and charts: embedded, side-by-side, and chart-embedded text [60]. (2) VizFlow: Adds dynamic linking and presentation to static data articles via interactive highlights and text-chart associations [44]. (3) ScholarPhi: Aggregates distributed equation definitions into a unified diagram view for better comprehension [43]. (4) SideNoter: Offers a dual-pane view combining structural outlines and logical sentence dependencies [31]. (5) Narrative Visualization: Structures literature reviews into three hierarchical levels with enhanced citation content organization [33].

and then use drag-and-drop interactions to build a hierarchical structure that reveals relationships between themes. Threddy also recommends related studies based on citation coverage and dynamically updates threads across multiple documents, providing continuous research information management support. Building on this, the Relatedly system by Palani et al. [62] further enhances cross-document theme exploration. It integrates the "related work" sections from multiple papers and helps users review content more efficiently through dynamic re-ranking, highlighting of unexplored dissimilar information, and automatic generation of descriptive paragraph headings. The system's "explore similar paragraphs" feature allows for deep dives into specific topics, and it dynamically adjusts recommendations based on reading history and citation information to ensure a comprehensive and targeted review. Going a step further, the Synergi system by Kang et al. [58] combines user-provided research threads with citation graphs and LLM to automatically generate a structured hierarchy of research themes. Using a Loopy Belief Propagation algorithm with a novel message-weighting scheme, the system retrieves key papers from a local citation graph and recursively summarizes them to form a thematic structure. Users can then iterate on and customize the results in an editor to quickly build a high-quality literature review outline.

Unlike TP ●, which focuses on "how" information is presented, IC ● centers on "what" is presented. It aims to enhance reading depth by supplementing, refining, or restructuring the original content. IC ● technology covers three main areas. S&O ▲ helps users quickly grasp the main ideas of a text. IS ▲ removes comprehension barriers by providing term definitions, background knowledge, and multimedia materials. C&O ▲ focuses on academic contexts, optimizing the retrieval and management of citations. The goal of IC ● is to guide readers from surface-level reading to deep understanding, making it key to achieving in-depth cognition and knowledge construction.

2.3. LO ●

LO ● aims to improve the readability of content and users' cognitive efficiency by refining the visual layout and information organization of documents. Unlike traditional static formatting, this approach focuses on enhancing understanding, association, and exploration through techniques such as image-text integration, information aggregation, and structured navigation.

In terms of image-text integration, researchers explore how to strengthen the coordination between textual content and visual elements such as charts. The Charagraphs system [60] introduces three image-text combination modes: embedding charts within text, placing charts side by side with text, and embedding text within charts (see Fig. 6 (1)). Embedded layouts reinforce logical connections and VE ▲ by closely aligning text and visuals. Building upon this, the VizFlow system [44] implements dynamic mechanisms such as "text fragments on charts", allowing synchronized IE ● between highlighted text and chart content (see Fig. 6 (2)). These methods improve communication efficiency by optimizing spatial arrangement and interaction of visual elements.

The effective integration of contextual information is another key strategy in LO ●. To reduce interruptions in the reading flow, the ScholarPhi system [43] uses an "Equation Diagram" feature that aggregates definitions of formula terms (see Fig. 6 (3)). This reduces cognitive load caused by frequent lookup actions. Similarly, the SideNoter system [31] overlays annotations and related information directly on the document image by parsing PDF structures. It employs highlighting and guiding lines to help users quickly locate key content (see Fig. 6 (4)). Both systems provide IS ▲ by integrating essential contextual elements alongside or within the primary reading path, thereby enhancing coherence and depth of understanding.

In the aspect of structured layering and navigation, the VarifocalReader system [25] introduces multi-level views and dynamic switching, enabling users to navigate freely between a macro-level structural information overview and micro-level content elements. This enhances the understanding of a document's overall logic. The Passages system [47] offers flexible information organization tools. Users can compare key content segments using structured tables or visually arrange and explore relationships on a free-form canvas. Wang et al. [33] propose a narrative visualization system with a three-level interactive slide layout, progressively revealing paragraph structure, internal logic, and citation details of a literature review (see Fig. 6 (5)). When the reading objective deepens to specialized structural analysis, layout strategies adapt accordingly. To support the complex task of "argumentation annotation", the VIANA system [39] employs a task-driven layered interface. It decomposes the analytical workflow into different interface layers, using semantic transitions for switching between them. This workflow-oriented layout is designed to assist expert users in deconstructing a text and building a model of its internal argumentative structure. These systems improve the organization and presentation of complex information by enhancing hierarchical navigation and layout flexibility.

LO ● focuses on the macro-level organization and spatial layout of a document. Compared to traditional linear documents, LO ● techniques break the fixed page flow. They use methods such as text-graphic integration, information aggregation, and structured navigation to create more efficient cognitive pathways. Whether placing a chart next to text, aggregating scattered definitions beside a formula, or providing a multi-level navigation view, the goal of LO ● is to reduce cognitive interruptions from eye-scanning and context switching. This, in turn, improves the efficiency of information exploration and relational understanding.

2.4. IE ●

In modern digital reading, user interaction with text is no longer a one-way process but a dynamic cognitive activity. By actively engaging in interpreting and understanding the text, readers can deepen comprehension, improve memory, and approach content from multiple perspectives to support personalized learning needs. As reading becomes increasingly digital, reading assistance systems have adopted diverse forms of IE ●. These range from basic operations such as text highlighting, clicking, and dragging to more advanced functions including annotation and sketching. These interactions aim to optimize user experience, support flexible reading strategies, and allow users to adjust TP ● according to their needs. Such designs can significantly improve comprehension and retention while helping users construct richer knowledge networks.

This section reviews major directions in the research on IE ●: (1) IA ▲, which supports users in deepening their understanding of content elements through marking, commenting, and visual explanation; (2) IL ▲, which encourages users to actively explore information and expand cognitive boundaries through intelligent question answering, exploratory navigation, and simulation-based experiences; (3) C&N ▲, which enhances users' control over reading content by enabling flexible text manipulation and efficient navigation tools to support information retrieval and comprehension. The following subsections introduce representative approaches and systems in each of these areas.

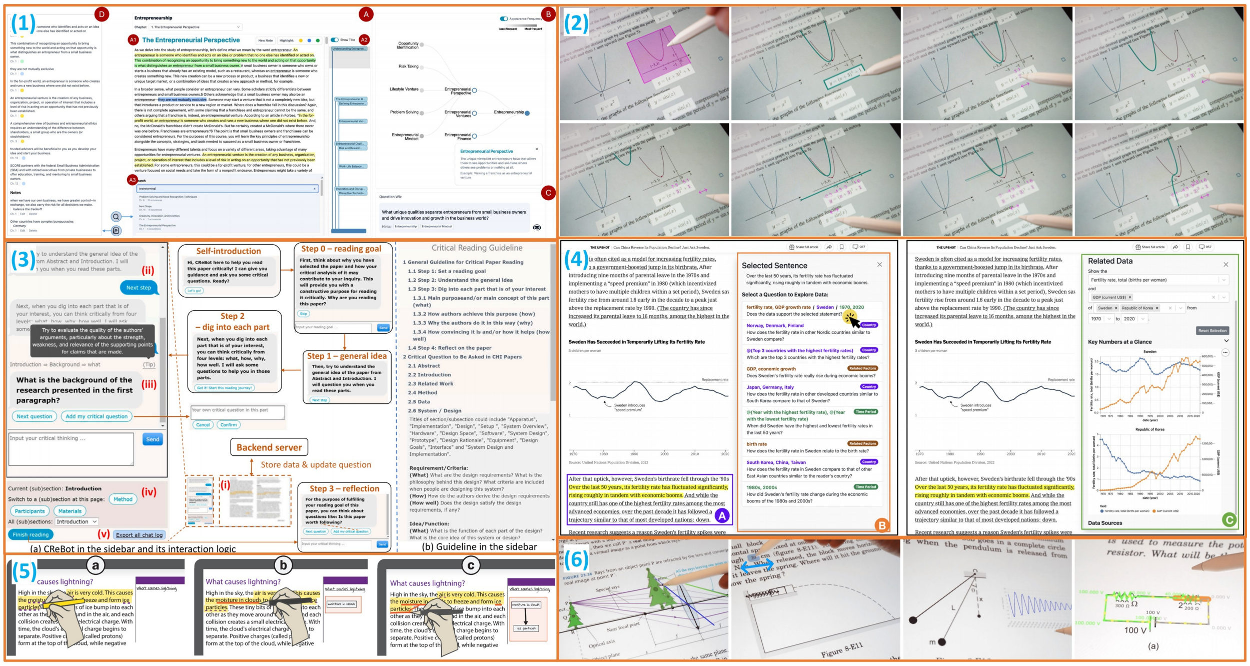

Fig. 7: IE ● Techniques in Active Reading and Learning. (1) IRead: Combines a logical mini-map, concept visualizations, and question prompts to support active textbook reading [76]. (2) Augmented Math: Enables interactive manipulation of equations and graphs extracted from scanned documents [56]. (3) CReBot: Offers real-time prompts and questions to support critical reading of scientific papers [50]. (4) DataDive: Generates questions and visualizations from statistical statements to enhance data understanding [70]. (5) TexSketch: Supports highlighting and visualizing causal relations by linking key terms in a graphical view [41]. (6) Augmented Physics: Converts static diagrams into interactive simulations across various physics topics [77].

2.4.1. IA ▲

As a key form of interaction for deepening understanding, annotation mechanisms directly influence users' cognitive processing and meaning construction during digital reading. Researchers have expanded this concept far beyond simple textual marking, exploring innovations in both the expressive capacity and the interaction modality of annotations. For example, some have investigated more expressive forms, such as the "playful annotations" proposed by Nicolas [32], which use symbols and affective reactions to deepen engagement in interactive textbooks. Concurrently, others have focused on the interaction itself. The GestAnnot system by Singh [22], for instance, replaced traditional toolbars with multi-touch gestures for actions like highlighting and commenting, aiming to make the annotation process more fluid and seamless.

As annotation mechanisms evolve, their applications can be considered at different scales of interaction. At the individual level, various systems support enriched annotation practices. The CARE platform by Zyska et al. [67] enables users to highlight, comment, tag, and add metadata within PDFs. It integrates NLP models to analyze and respond to comment content, offering personalized support. texSketch, proposed by Subramonyam et al. [41], allows users to transform their understanding into causal diagrams using pen-based input, which facilitates the organization of complex content (see Fig. 7 (5)). Annotation forms have also expanded beyond text to include audio. For example, the ARA system by Golovchinksy et al. [16] supports voice-based annotations. The MetaTation system [34] can automatically generate supporting information such as word meanings and usage based on annotations, enhancing language comprehension. Systems such as LiquidText [17] and GatherReader [19] treat annotation and highlighting as core functions of active reading, emphasizing structured information extraction and cross-paragraph linking. At the group level, annotation serves as a key communication and sharing mechanism in collaborative reading. The BuddyBooks system by Pearson et al. [20] introduces tools like "Point-out" and the "Look-at-this" queue to help team members quickly locate key content. It supports real-time, multi-user annotation with color differentiation to enhance communication and reading coordination.The WikiTUI system by [13] bridges digital content and physical books through touch-based interactions and offers the potential to support collaborative annotation. Another key direction for annotation is supporting knowledge synthesis across texts. The "Canvas" feature in the NEEDLE system [81] provides an interactive paradigm for this. It allows users to drag and drop knowledge fragments (i.e., sentences) distilled from different single texts onto a canvas for freeform spatial organization and visual linking. This interaction transforms "annotation" from a reading comprehension aid into a powerful knowledge synthesis tool, shifting the focus from helping users "consume" information to empowering them to "create" new knowledge structures.

These studies expand the space of interaction between users and texts through diverse annotation methods and collaborative mechanisms. They enhance the potential for individual meaning construction and collective knowledge co-creation.

2.4.2. IL ▲

As a key mechanism for transforming reading from passive information consumption to active meaning-making, IL ▲ guides users toward deeper understanding and active knowledge engagement by fostering questioning and exploration. Researchers have explored multiple implementation paths, from systems that proactively generate guiding questions to those that leverage LLM to build more dynamic and responsive interactive environments. For example, the CReBot system [50] (see Fig. 7 (3)) and ReaderQuizzer [64] demonstrate the potential of automatically generated questions for fostering critical thinking. Qlarify [57] introduces the novel paradigm of "recursively expandable abstracts", allowing users to progressively unfold content to retrieve details. Meanwhile, RealitySummary [78] extends Q&A capabilities into the Mixed Reality (MR) environment, enabling on-demand interaction with physical documents. These works can generally be categorized into two technical paths: question-answering based exploration and simulation-based exploration.

In question-answering based exploration, systems support critical thinking and deep information retrieval through dialogue mechanisms. The Inquire Biology system developed by Chaudhri et al. [21] organizes biology knowledge into a structured form and supports natural language questions with reasoning-based answers. The DataDive system by Kim et al. [70] focuses on statistical content, automatically generating exploratory questions when a user clicks on data elements (see Fig. 7 (4)). The IRead system [76] (see Fig. 7 (1)) combines LLM, a concept tree, and an intelligent Q&A assistant to support personalized cognitive path construction. Some systems also focus on training specific skills, such as Argumentor [74] for argumentation skills.

In contrast, simulation-based exploration aims to deepen understanding through the direct manipulation of conceptual representations. Chulpongsatorn et al. [56] enhanced mathematics textbooks with machine learning, allowing learners to clarify abstract concepts through interactive numerical operations and graphical manipulations (see Fig. 7 (2)). In physics education, the Augmented Physics system [77] combines augmented reality (AR) with interactive simulations to provide immersive experiences for exploring physical phenomena (see Fig. 7 (6)). In the context of data narratives, VisJockey [23] allows authors to link pre-defined visualization interactions to specific text segments; readers can then click the text to trigger a "playback" of the intended visual narrative in the chart, guiding their comprehension of the data.

Overall, IL ▲ systems, through the complementary mechanisms of question-answering interaction and conceptual manipulation, expand the user-content relationship into a multi-dimensional interactive space that transcends linear reading. This not only enhances the depth of understanding but also empowers readers to take a more active role in their own knowledge construction.

2.4.3. C&N ▲

In the domain of IE ●, supporting users in managing the reading process, dynamically adjusting the text structure, and efficiently locating and navigating information is a fundamental goal of C&N ▲. These techniques focus not only on helping users manage reading focus, but also on flexible manipulation of text presentation, and seamless navigation within a document or across information content elements.

First, in managing reading focus, Chen et al. [55] proposed the Marvista tool, which dims non-current paragraphs to help users concentrate on specific content. This reduces distraction and enables better control over reading pace, especially when dealing with large volumes of information. Second, for dynamic adjustment of structural information and nonlinear navigation, the LiquidText system by Tashman and Edwards [17] supports multi-touch gestures that allow users to fold or expand sections of the text as needed. Its "dog-earing" feature enables users to set navigation anchors. In combination with excerpting and LE ▲, the system supports efficient context-switching between original

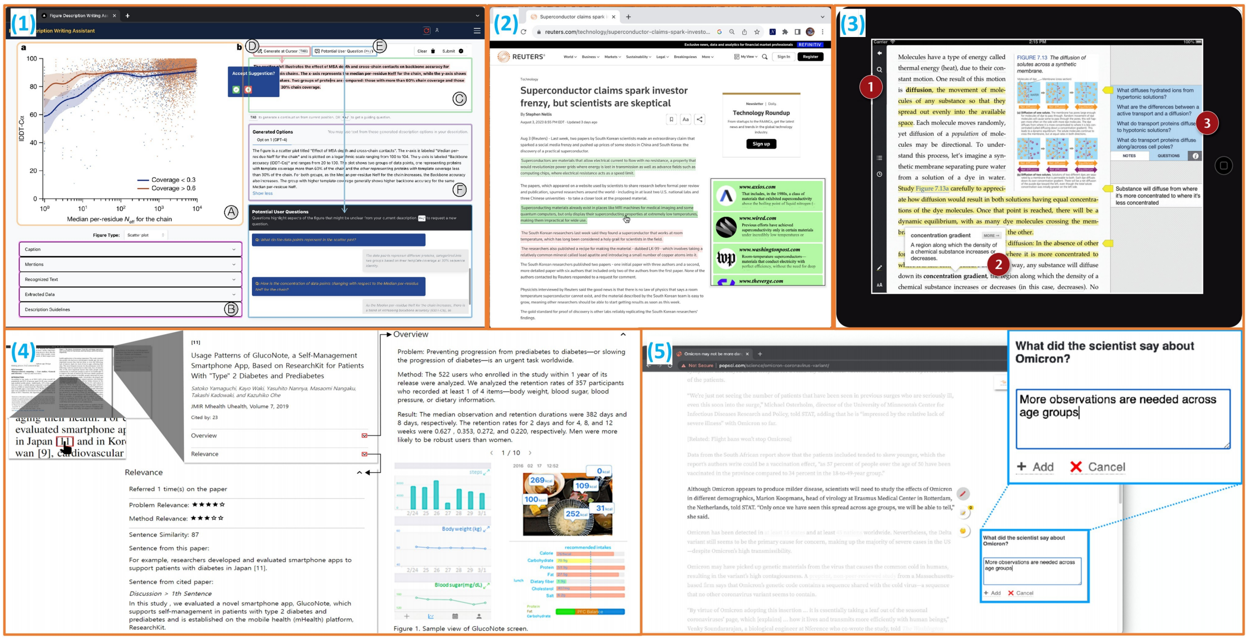

Fig. 8: Reading Assistance Systems in Different Domains. (1) FigurA11y: For visually impaired users, enhances chart access in scientific documents by generating editable alt text [75]. (2) NEWSSENSE: For news reading, highlights conflicting/supportive claims with external evidence overlays [61]. (3) Inquire Biology: For textbook learning, assists science education through definitions and system-generated questions [21]. (4) QuickRef: For academic paper reading, displays citation metadata, summaries, and relevance inline [63]. (5) Marvista: For news reading, supports time-constrained readers with key sentence highlighting and reflective prompts [55].

content and notes, thus enabling more flexible, exploratory reading beyond traditional linear workflows. Lastly, in collecting and navigating information content, The GatherReader system [19] introduced a Frame-and-Drag-to-Pocket technique. It allows users to quickly collect snippets or full pages into a visual clipboard. This flexible organization improves navigation among fragments and supports automatic bidirectional cross-referencing between notes and the source text, enhancing the management and retrieval of key reading information. These systems improve users' control over reading content through refined text manipulation and navigation mechanisms. They also enhance the organization, review, and exploration of information, providing important support for in-depth reading and knowledge construction.

Building upon manual and semi-manual navigation, search functionality further enhances text control and exploration capabilities. Modern search has transcended keyword location to become an interactive exploration of the document. For instance, VE ▲ techniques, as in ScentHighlights [12], not only highlight exact matches but also mark semantically related content; LO ● is often employed through sidebars (e.g., in Scim [46]) to present all search results for efficient, non-linear navigation. More importantly, the goal of search has shifted from "locating" to "understanding". IC ● techniques, such as in ScholarPhi [43], go beyond marking a term's position to directly provide its definition, shifting the search behavior from finding a location to understanding content.

In summary, IE ● shifts the focus from the "text" to the "reader". Its core objective is to empower users with active control, transforming reading from a one-way reception of information into a two-way process of knowledge exploration and construction. It encompasses three key dimensions: IA ▲ supports users in recording personal insights through highlighting and notes, thereby building a personalized knowledge network; IL ▲ motivates critical thinking and deep inquiry through intelligent Q&A, guided exploration, and simulated experiences; and C&N ▲ provides flexible tools for view management and navigation, allowing users to command the information space efficiently. Together, IE ● techniques transform the reader from a passive consumer of information into an active participant, significantly enhancing the personalization, engagement, and cognitive depth of the reading experience.

3. Systems and Applications

This section reviews representative systems and applications, demonstrating how the visualization and interaction techniques discussed in our taxonomy are applied to address the unique challenges of different domains. While most systems employ a combination of techniques, each domain emphasizes certain categories based on its core user needs, as the following examples will illustrate.

3.1. News

Digital news reading presents unique challenges centered on information veracity and viewpoint diversity. Consequently, reading assistance systems in this domain heavily leverage IC ● to support fact-checking and compare multiple perspectives, and IE ● to allow users to actively explore different stances.

For example, the Factful system by Kim et al. [28] (see Fig. 5 (3)) is a web-based reading interface that applies the IS ▲ technique within the IC ● category. It integrates open government data to verify budget-related statements, providing crucial background context that helps users assess credibility. Similarly, the NEWSSENSE system by Milbauer et al. [61] (see Fig. 8 (2)) leverages a combination of techniques. It applies the IS ▲ technique, a part of IC ●, by extracting and comparing related viewpoints from different news sources. It then uses the VE ▲ technique from TP ● to highlight supportive or opposing perspectives.

Furthermore, the Marvista system by Chen et al. [55] (see Fig. 8 (5)) is a comprehensive tool integrating multiple techniques. It applies the S&O ▲ technique within the IC ● category to intelligently highlight content based on a user's set reading time. During reading, its "focus mode" reduces distraction by visually dimming surrounding paragraphs, leaving only the current paragraph fully illuminated. This dynamic alteration of the page's visual hierarchy to guide attention is a form of LO ●. The system also embodies the IL ▲ technique within the IE ● category by dynamically adapting to user interaction and generating reflective questions. By combining IC ●, LO ●, and IE ●, Marvista provides an efficient, personalized news reading experience.

In summary, reading assistance tools in the news domain primarily leverage IC ● to verify claims and present diverse perspectives. These systems aim to equip readers with the tools needed to critically evaluate information, thereby helping to combat bias and misinformation and foster a more informed understanding of complex events.

3.2. Academic Paper

The rapid growth of scientific literature presents a significant information overload challenge. Reading assistance in this domain focuses on improving efficiency and deep comprehension, thus heavily relying on IC ● and IE ●.

For example, Scim by Fok et al. [46] (see Fig. 4 (3)) is a typical academic reading tool that combines techniques from TP ● and IE ●. It applies the VE ▲ technique, which belongs to the TP ● category, by using AI to automatically highlight sentences with four different rhetorical roles. It also implements the C&N ▲ technique from the IE ● category through its "highlight browser" and scrollbar markers. By integrating these techniques, Scim provides powerful support for researchers to quickly screen and understand literature.

To support multi-modal understanding, Papeos by Kim et al. [59] (see Fig. 5 (4)) applies the IS ▲ technique within the IC ● category by combining academic papers with presentation videos. It also utilizes principles of LO ● by structuring the interface to present both text and video concurrently.

To manage citation overload, systems like CiteSee [54] (see Fig. 5 (5)) and QuickRef [63] (see Fig. 8 (4)) focus on the C&O ▲ sub-category of IC ●. CiteSee also employs the VE ▲ technique from TP ● to distinguish citation status, while QuickRef uses the S&O ▲ technique (part of IC ●) to provide structured abstracts. Beyond managing citations, effectively organizing and synthesizing the paper's content itself is equally crucial for knowledge construction. To this end, Threddy by Kang et al. [49] is a prime example of the IA ▲ technique within the IE ● category. It allows users to highlight text fragments and organize them into hierarchical research threads, thereby facilitating systematic information synthesis. Finally, to foster deeper engagement, the CReBot system by Peng et al. [50] (see Fig. 7 (3)) exemplifies the IL ▲ technique, also part of IE ●, by generating critical questions to guide readers.

In summary, the systems designed for academic papers demonstrate a sophisticated integration of IC ●, IE ●, and TP ●. They address the dual needs of researchers: efficient information filtering for rapid literature assessment and in-depth comprehension tools for detailed analysis. Ultimately, these systems aim to streamline the entire research lifecycle, from literature discovery to knowledge synthesis.

3.3. Textbook

In the educational domain, the goal is to transform static textbooks into interactive learning experiences. Systems in this area prioritize IE ●, especially the IL ▲ technique, and use TP ● and LO ● to make abstract concepts more understandable.

The Inquire Biology system by Chaudhri et al. [21] (see Fig. 8 (3)) deeply embodies this approach. Its core feature, supporting natural language Q&A, is an advanced application of the IL ▲ technique within the IE ● category. At the same time, its concept summary pages represent a combination of techniques from the IC ● category, namely S&O ▲ and IS ▲, to help students build structured knowledge networks.

To make abstract concepts tangible, tools like Augmented Math by Chulpongsatorn et al. [56] (see Fig. 7 (2)) leverage a tight integration of techniques. They apply IG ▲ and VE ▲, both part of TP ●, to convert static formulas into dynamic animations. They also embody the IL ▲ technique from the IE ● category by allowing users to directly manipulate values and observe changes. The IRead system by Zhang et al. [76] (see Fig. 7 (1)) combines LO ● with the IG ▲ technique (from TP ●) to visualize concept hierarchies. Its core strength is its Q&A assistant, a key application of the IL ▲ technique (from IE ●), while its note-taking feature is an example of the IA ▲ technique, also within IE ●.

In essence, reading assistance for textbooks is defined by its strong emphasis on IE ●, particularly IL ▲. By converting static text and diagrams into explorable and responsive environments, these tools transform the textbook from a repository of information into an interactive learning companion. This shift fosters deeper conceptual understanding and supports personalized learning paths.

3.4. Special User Groups

Reading assistance for special user groups, such as those with visual impairments, focuses on accessibility and providing alternative means to access information. Consequently, systems in this domain often feature a unique blend of IC ●, used to translate visual content into alternative modalities, and LO ●, which ensures structural compatibility with assistive technologies like screen readers.

For example, the FigurA11y system by Singh et al. [75] (see Fig. 8 (1)) applies the IS ▲ technique within the IC ● category. It automatically generates alternative text for figures, translating visual content into a textual format that assistive technologies can process. Similarly, the Infosonics system by Holloway et al. [85] also employs the IS ▲ technique from the IC ● category. It transforms a data visualization chart into an interactive auditory experience, where users can hear the rise and fall of data, accompanied by spoken annotations that explain key points.

In contrast, the model by Dzhurynskyi et al. [86] leverages the IG ▲ technique, which belongs to the TP ● category. It automatically produces a touch-optimized tactile graphic directly from a text description, converting a textual narrative into a physical representation.

Finally, the SciA11y system by Wang et al. [45] primarily focuses on LO ● to address the inaccessibility of scientific PDFs. The system converts visually formatted but semantically unstructured PDF files into clearly hierarchical HTML. This means the text is imbued with semantic tags for headings, paragraphs, and lists, allowing screen readers to understand the document's structure. As a result, users with visual impairments can navigate efficiently by section, rather than just listening to an undifferentiated stream of text. Furthermore, the system applies the LE ▲ technique from the TP ● category. It creates bidirectional links between inline citations and their references, allowing users to look up a source and seamlessly return.

In conclusion, the approaches for special user groups highlight the critical role of information transformation. They leverage a diverse set of techniques: IC ● (via IS ▲) to translate visual content into alternative modalities like text or sound; TP ● (via IG ▲) to create physical graphics from text; and LO ● to ensure structural compatibility with assistive technologies. These systems are not just about assistance; they are about enabling equitable access and fostering independent, richer reading experiences for all users.

4. Discussion

In this study, we have conducted a comprehensive classification and analysis of the auxiliary reading visualization techniques for text data. Through in-depth discussions on data types (including structural information, content elements, and user interaction data, as detailed in Section 1.2) and visualization technologies (TP ●, IC ●, LO ●, and IE ●), we have summarized the progress of current research and put forward prospects for future research directions.

4.1. Characteristics and Challenges of data types in Reading Assistance

The richness and diversity of text-related data types provide great potential for reading assistance, while also posing considerable challenges. The complexity of structural information (Section 1.2.1), the depth and breadth of content elements (Section 1.2.2), and the personalized nature of user interaction data (Section 1.2.3) all demand flexible and adaptive assistance systems. For example, the argumentative structure of an essay and the logical flow of a scientific paper require different visualization strategies to support effective comprehension. Similarly, the extraction and interpretation of fine-grained content elements—such as technical terms, symbols, and numerical data—often rely on advances in NLP and Artificial Intelligence (AI). In addition, analyzing user interaction data involves not only individual reading behaviors, but also, in some cases, collaborative activities and sharing patterns.

The complexity of structural information is reflected in how different genres organize their content. News articles typically revolve around a central event, argumentative texts often follow patterns of claims, reasons, and evidence, while academic papers emphasize a logical progression from background to methodology and conclusions. Each structure benefits from customized visualization techniques that help readers quickly grasp core ideas and logical organization. The depth and breadth of content elements are evident in the abundance of embedded information, including domain-specific terminology, symbolic expressions, and quantitative data. Accurately extracting and clarifying these elements is essential for comprehension and increasingly supported by NLP and AI techniques. For instance, defining technical terms or linking mathematical symbols to their contextual descriptions can greatly enhance reader understanding. Lastly, user interaction data—such as highlighting, annotations, and search history—reveals personalized reading trajectories and comprehension patterns. This information is key to delivering tailored reading experiences. For example, analyzing how long a user spends on different paragraphs may help infer their focus and interest, enabling the generation of personalized summaries or targeted guidance. Addressing these characteristics and overcoming the challenges of processing such diverse data types are critical to the development of intelligent and effective reading assistance tools.

4.2. Trends in Visualization Techniques

Based on the taxonomy of visualization techniques proposed in this paper and the analysis of surveyed works summarized in Table 1, several notable trends emerge in the development of single-text digital reading assistance systems. These trends reflect a shared trajectory toward intelligent, personalized, and interactive support, with each category exhibiting distinct patterns and emphases.

TP ● and IE ● are the most widely adopted foundational techniques. Nearly all reviewed systems incorporate them to some extent, as shown by the dense color fills in the corresponding columns of Table 1. This highlights a strong consensus that optimizing text presentation and enabling user interaction are central to effective reading assistance. Within TP ●, VE ▲ and LE ▲ are especially prevalent, indicating ongoing efforts to guide user attention and enhance semantic integration across heterogeneous content elements. These methods help reduce information fragmentation and improve reading efficiency through coordinated multimodal presentation. Within IE ●, IA ▲ and C&N ▲ are the most frequently used subcategories, reflecting the increasing importance of supporting active engagement and efficient navigation.

IC ● also shows significant prominence in Table 1. Techniques such as S&O ▲ assist readers in quickly grasping main ideas, while IS ▲ provides relevant background knowledge and contextual details that support comprehension. This demonstrates that current research not only focuses on how information is visually presented, but also on what content is made accessible and how its value can be enriched. C&O ▲, although less frequently applied, plays an important role in academic reading scenarios where managing references and understanding scholarly relationships are essential.

From Table 1, it is evident that compared to other categories, LO ● receives relatively less research attention as an independent focus. This may suggest that layout design is often integrated implicitly into overall system design rather than treated as a standalone technique. However, some works show growing interest in adaptive layout strategies that break from traditional linear structures—integrating text and visuals dynamically to improve cognitive efficiency and support diverse reading tasks.

A clear trend is the combination of multiple enhancement strategies. Most systems integrate techniques across TP ●, IC ●, and IE ● rather than relying on a single category. For example, a system may use color-coded Visual Emphasis (TP ●, as in VE ▲), provide contextual summaries (IC ●, as in S&O ▲), and enable user annotations and navigation (IE ●, as in IA ▲ and C&N ▲), forming a comprehensive multi-layered support framework. This integrative approach reflects a holistic understanding of single-text reading assistance needs. Additionally, the adoption of certain subcategories shows clear dependence on the specific context or application scenario. For instance, IL ▲ within IE ● is more common in educational or tutorial-oriented systems (as discussed in Textbooks), where AI-driven feedback and personalized guidance play a more significant role. In contrast, IL ▲ is less prevalent in general-purpose reading tools.

In summary, the landscape of visualization techniques for single-text digital reading is characterized by intelligent perception, dynamic interaction, and adaptive support. Future systems will likely further integrate context-aware VE ▲ , real-time Information Graphicization, and semantically coherent LE ▲ within TP ●; leverage NLP and large language models for deeper semantic enrichment in IC ●; explore flexible, cognitively efficient layouts in LO ●; and enable increasingly personalized, collaborative experiences in IE ●. Collectively, these developments mark a shift from passive content display toward active, user-centered cognitive assistance.

4.3. Cultural Adaptability and Technological Inclusivity

To enhance the global applicability of reading assistance systems, the needs of different languages and writing directions must be fully considered. However, the existing literature reveals that current systems are predominantly designed with a Left-to-Right (LTR) default, overlooking the specific requirements of Right-to-Left (RTL) and vertical text. This design assumption may limit the cross-cultural application of these systems and introduces complex engineering and design challenges across all categories of reading assistance.

In LO ●, the impact is most direct. Systems like VarifocalReader [25] assume an LTR arrangement for chapters and paragraphs. Layout algorithms trained on LTR data are likely to produce structures that are cognitively dissonant in an RTL context, thereby defeating the purpose of optimization. In TP ●, while some techniques like simple highlighting or sentence marking remain applicable in RTL, directional visual aids, such as color-coding to connect formulas with descriptions [48], must be reversed to align with the reading order. The placement of micro-charts, as seen in GistVis [79], should also follow the text flow, for instance, by being placed to the right in RTL text. In IE ●, user interaction habits often require adaptation. Gesture controls, such as the "swipe left to turn the page" in LiquidText [17], should be changed to "swipe right" in RTL contexts to match the natural reading direction and reduce cognitive friction. Navigation panels, as in Scim [46], should be placed on the right side in RTL interfaces. This is because the user's gaze scans text from right to left, making the right side the visual starting point. Placing controls on the right reduces the burden of eye movement and context switching, while maintaining consistency between the interface and the text flow. Even if the core NLP models for IC ●, such as in Paper Plain [52], can process text from different writing directions to generate summaries or extract key information, the presentation of this content must still align with the text flow. For instance, just-in-time definition pop-ups [43] must be displayed according to the language's directionality and interface conventions to ensure a natural reading experience.

Therefore, the lack of cross-cultural validation severely limits the inclusivity of these technologies. Future research must move beyond simple mirroring and fundamentally rethink layout algorithms, interaction paradigms, and assistive device mechanisms to create truly equitable and globally accessible reading tools.

4.4. Deepening User Modeling through Interaction Histories

As systematically elaborated in subsection 2.4, this survey covers diverse IE ● techniques that empower users to engage in in-depth interactions with digital text. However, most existing research focuses on the functionality of the interaction itself. The deep value of the continuous data streams generated by these behaviors, known as interaction histories, remains underexplored. Shifting the research perspective from isolated events to the analysis of complete processes and patterns, i.e., interaction provenance, can open up new avenues for user modeling and personalized reading support.

Some systems have already begun to leverage interaction data for preliminary modeling. For example, Marvista [55], as discussed earlier, infers attention distribution by analyzing user dwell time on paragraphs to generate personalized summaries. This approach moves beyond immediate feedback by assessing content importance based on behavioral history. However, the ambiguity of dwell time presents a challenge. It can signify either high engagement or comprehension difficulties, and current models struggle to distinguish between these starkly different cognitive states. This limitation reduces the effectiveness of precise interventions. Similarly, the CReBot system provides guided questions at structural nodes [50]. While this is an effective IL ▲ strategy, the timing of its interventions is not well-aligned with the user's real-time cognitive state.

Research has shown that even basic explicit interaction logs, such as the frequency of highlighting or accessing notes, can effectively differentiate learner types and predict academic performance [87]. In the BookRoll system, such logs have been used for learner classification and performance prediction. This approach points to a future direction for tools like LiquidText and Threddy, introduced in subsection 2.4. These tools currently serve as manual organization platforms, yet the complete operational history they record is itself a detailed map of the user's cognitive process. Future versions could analyze this map, transforming the tools from passive recorders into proactive diagnostic partners that offer predictive support driven by IC ● .

To address the "signal ambiguity" in such models, a more advanced path is to integrate multimodal interaction histories, particularly implicit data like eye movements. This first requires specialized software to process and visualize these complex data streams. Systems like EyeMap [88] provide a foundational toolkit for researchers, offering various visualizations such as heatmaps and scan paths to make raw eye-tracking data interpretable. Building on this, the work by Špakov et al. [89] demonstrates how such visualizations can be applied in educational settings to facilitate user modeling. They developed dynamic visualizations like "Gaze replay" and "Word replay" that allow teachers to observe a student's reading process. In this context, the visualizations serve as a cognitive tool for the teacher, who performs the user modeling based on the provided visual evidence. While this human-in-the-loop approach is powerful, an even more advanced direction is to automate the modeling process entirely. The study by Santhosh et al. [90] exemplifies this by combining eye-tracking and physiological signals with deep learning models to achieve real-time assessment of user engagement and affective states. This technology lays the foundation for building truly adaptive reading systems. One can envision an intelligent reading environment that perceives a user's state of confusion in real time and instantly triggers a highly relevant IE ● . Such an enhancement could be a ScholarPhi-style [43] term definition or a dynamically generated structured thumbnail as proposed by Zhang et al. [76], helping the user overcome cognitive barriers in a timely manner.

Notably, the paradigm of analyzing interaction histories can extend beyond traditional text reading to the exploration and reflection upon one's own data. The work by Choe et al. [91] demonstrates that when users explore their personal health data through visualizations, their interaction behaviors (e.g., selection, filtering, comparison) are intricately linked to their cognitive processes of recall, questioning, and gaining personal insights. This reveals a form of "self-modeling" aimed at self-reflection, further broadening the applicability of interaction history analysis for personalized support.

In conclusion, the in-depth utilization of interaction histories and provenance marks an evolution in digital reading assistance from being function-driven to model-driven. Future interactive systems will not just passively respond to user commands. Instead, they will proactively and adaptively provide support by building dynamic user models based on complete interaction histories. This model will become the core engine for advanced assistive functions like personalized summarization, adaptive questioning, and real-time intervention. Ultimately, this will advance the digital reading experience to a new stage that is truly user-centric, personalized, and efficient.

4.5. User Evaluation

In text visualization research, task-based laboratory user studies remain a core method for evaluating system efficacy. Researchers typically design explicit reading or information processing tasks for participants and record their performance in a controlled environment. For instance, the study of HiText by Yang et al. [35] employed TOEFL-style reading comprehension tests to assess the system's impact on skimming efficiency, while Kim et al. [37] utilized timed information-matching tasks to quantify the value of their text–table linking feature. Such experiments often rely on objective behavioral metrics, such as task completion time or accuracy, and are complemented by subjective ratings from standardized scales like the System Usability Scale (SUS) and the NASA-TLX for cognitive load. The research on Charagraph [60] and CReBot [50] both adopted this combined approach to measure quantitative performance alongside user perception.

Nevertheless, user evaluation in text visualization faces several challenges. First, reading tasks are often open-ended and subjective, lacking a single "correct answer", which complicates quantitative measurement. The evaluation of ConceptEVA [66], which addresses the highly subjective task of summary generation, innovatively used participants' own manually created summaries as a baseline, measuring "satisfaction" as the core metric. Second, individual differences are significant. The study of CReBot [50] revealed that an interactive bot effective for "routine readers" could act as a cognitive distractor for "novices", highlighting the importance of personalizing both design and evaluation. Furthermore, reading involves complex cognitive processes that are difficult to capture with existing metrics. For example, the FigurA11y study [75] found that its interactive features, while not reducing cognitive load, prompted users to make substantially more revisions to the AI-generated draft, reflecting deeper engagement that traditional efficiency metrics would miss.

Despite these challenges, user evaluation remains the core pathway to understanding the relationship between system effectiveness and user needs. On one hand, user feedback can reveal unforeseen design issues, as demonstrated in the ScholarPhi study [43], which found that an incomplete AI system could mislead users more than no AI at all. On the other hand, combining behavioral data with subjective ratings helps to model user states at both cognitive and operational levels, providing a basis for adaptive system optimization. The study of IRead [76], for instance, used observations of users with different academic backgrounds to inform strategies for personalized guidance.

In addition to quantitative experiments, qualitative methods are equally crucial. Semi-structured interviews and the think-aloud protocol, through open-ended questions or real-time verbalization, capture users' thought processes and immediate feelings. Formative studies for FigurA11y [75] and DataDive [70], for instance, used these methods to deeply understand users' reasoning paths and design feedback. Because they do not rely on pre-set answers, these methods uncover deep experiences that quantitative metrics cannot reflect, thereby complementing the limitations of lab data.

At the same time, interaction log analysis provides a non-intrusive means of recording users' actual operational traces. FigurA11y [75] revealed human-AI collaboration patterns through editing sequence analysis, while CARE [67] utilized page dwell times to analyze users' attention distribution. These behavioral data, combined with experimental results and qualitative interviews, enable researchers to achieve a more comprehensive understanding of user behaviors and cognitive states.

In practice, these methods are often combined into a mixed-methods evaluation strategy to gain multi-dimensional user insights. For example, CiteSee [54] first validated its core hypothesis in a lab setting, then collected real-world usage data through a long-term field deployment. This approach preserves the control of an experiment while embracing the complexity of real-world contexts, representing a robust trajectory for user evaluation in text visualization and pointing toward more user-centered directions for future research.

4.6. AI-Driven Reading Assistance: Challenges and Future Directions

As demonstrated in this survey, the recent and rapid development of LLM has significantly advanced the intelligent evolution of reading assistance technologies. They show great potential, particularly in areas like IC ● and IE ●. However, while embracing the benefits of these technologies, we must critically examine their inherent limitations. As Haghighatkhah et al. [92] noted, uncertainty is a challenge that pervades the entire text analysis pipeline. The inherent bias and hallucination phenomena in LLM are prominent manifestations of this uncertainty under the current technological paradigm. In the context of single-text reading assistance, these issues pose unique and significant challenges, directly impacting the reliability of these tools for knowledge acquisition.

The core objective of a reading assistance system is to help users accurately and deeply understand a text. Hallucination, where a model generates information that is plausible-sounding but factually incorrect or absent from the source text, fundamentally undermines this objective. Empirical studies have shown that such "confidently incorrect" outputs severely erode user trust, and once trust is damaged, its recovery is often extremely slow [93]. In S&O ▲ scenarios, systems like Paper Plain [52] may hallucinate experimental conclusions not present in the original work, leading readers to misinterpret the author's core arguments. In IS ▲ tasks, if an AI assistant (e.g., an advanced implementation of ScholarPhi [43]) generates a flawed definition for a key technical term, it can disrupt the learning process and lead users to form incorrect mental models.

Bias, which often stems from the large-scale corpora used to train LLM, can subtly distort a user's interpretation of a text. For instance, in VE ▲ highlighting (as in Scim [46]) and summary generation, a model may selectively emphasize sentences that align with mainstream viewpoints while downplaying or ignoring content that reflects minority, critical, or non-mainstream scientific perspectives. This can result in an incomplete or even distorted understanding. In the context of argumentative texts (as with ArguMentor [74]), model-generated counterarguments may reflect common stereotypes or partisan arguments rather than rigorous logical refutations. This risks reinforcing societal biases instead of fostering critical thinking.

Addressing these challenges requires a multi-faceted strategic approach in future research and system design. First, traceability should be a core design principle. All AI-generated content, from a summary sentence to a term definition, must be clearly linked to its source in the original text to allow for rapid user verification. The Traceable Text system [69], which provides sentence-level provenance in its summaries, serves as a valuable reference. Second, the design paradigm must shift from fully automated "authoritative" systems to collaborative "assistant" models. As advocated by Beauxis-Aussalet et al., interactive visualization plays a central role in fostering human agency and dynamic trust calibration [94]. Interfaces should therefore present AI outputs as assistive information and provide intuitive mechanisms for users to review, provide feedback, and make corrections, ensuring that users retain critical agency.

However, this path is not without its own challenges. As demonstrated by Klein et al. in their large-scale benchmark of Explainable AI (XAI) methods, there are significant inconsistencies among different explanation techniques (e.g., saliency maps), and even the metrics used to evaluate their "faithfulness" lack consistent reliability [95]. This implies that even when a system provides a "traceable" explanation, the explanation itself may be a distorted or unreliable representation, introducing a new layer of potential misinformation.

Therefore, future research must move beyond metrics of efficiency and satisfaction to prioritize the robustness and safety of these tools. This requires answering several critical questions: To what extent do users uncritically accept AI-generated information? Can they identify subtle biases or factual errors in the output? Does long-term use of such tools ultimately enhance or diminish a user's own critical thinking abilities? A thorough investigation of these questions is essential for guiding the responsible development of the next generation of reading assistance technologies.

4.7. A Design Philosophy for Cognitive Scaffolding

Synthesizing the systematic findings of this survey, we contend that effective reading assistance systems should be guided by a design philosophy centered on cognitive scaffolding. A first principle is to preserve cognitive flow by minimizing disruption. Systems should prioritize strategies such as in-context embedding (e.g., ScholarPhi [43]), view synchronization (e.g., formula-text linking [48]), and information aggregation (e.g., equation diagrams in ScholarPhi [43]) to avoid operations that break a reader's flow or force context switching. Equally important is to empower active knowledge construction rather than passive text consumption. This requires flexible tools for extraction and organization (e.g., LiquidText [17], Threddy [49]), mechanisms that support the creation of deep connections across information (e.g., TexSketch's chart generation [41]), and features that transform annotations into structured personal knowledge assets—turning reading into a creative and constructive activity.

When incorporating Artificial Intelligence, systems must also safeguard user agency and critical thinking. This entails making AI-generated content traceable and verifiable (e.g., Traceable Text [69]), clearly distinguishing it from original text, and positioning the AI as a "collaborator" rather than an "authoritative interpreter" (e.g., ArguMentor [74], CReBot [50]), thereby encouraging scrutiny and judgment. Finally, systems should deliver personalized and adaptive support for diverse reader goals. This involves providing adjustable levels of detail, dynamically adapting presentation to readers' implicit behaviors (e.g., Marvista [55]), and offering specialized view modes for different reading tasks (e.g., VarifocalReader [25]), ultimately creating a user-centered adaptive experience.

Taken together, this design philosophy highlights a single overarching objective: evolving reading assistance systems from static information browsers into dynamic cognitive partners that actively think and learn alongside the reader.

4.8. Future Research Directions and Challenges

Although reading assistance technologies have made significant progress, several key challenges remain to be addressed. At the core technical level, the depth and breadth of semantic understanding still require improvement. This is especially true when dealing with structurally complex and terminology-intensive texts, such as professional literature and legal documents. Systems need to develop stronger capabilities in context modeling, precise terminology analysis, and inference of implicit semantics. These challenges primarily involve core issues in natural language processing such as word sense disambiguation, coreference resolution, and discourse structure recognition—that is, the system's ability to understand word meanings, pronoun references, and the logical relationships between sentences and paragraphs. These capabilities represent a key bottleneck in advancing the intelligence of reading assistance.

In terms of personalization and adaptation, current systems are still limited in the granularity and responsiveness of user modeling. Future work may incorporate artificial intelligence–driven multi-dimensional user modeling methods. By leveraging machine learning techniques, systems can dynamically analyze user interaction data (as defined in Section 1.2.3) such as reading behaviors, knowledge levels, and cognitive preferences, thereby constructing more accurate and evolving user profiles. Meanwhile, adaptive interfaces should utilize intelligent decision-making mechanisms to support fine-grained dynamic adjustments. This would enable real-time optimization of content presentation and interaction logic based on user feedback.

Multi-modal integration is also a critical direction for enhancing the digital reading experience, particularly for techniques that explicitly link heterogeneous content, such as LE▲ and IG▲. However, it faces technical challenges due to the heterogeneity of data. Information forms such as text, charts, audio, and video differ in semantic, structural, and temporal dimensions. Cross-modal learning and representation alignment techniques are needed to achieve semantic-level integration, ensuring consistency and completeness in information expression. This is particularly important in high-cognitive-load scenarios such as education and scientific research, where coherence among multi-modal content is essential for improving user comprehension and system credibility.

Finally, with the rapid development of artificial intelligence, especially deep learning and large language models, reading assistance systems are evolving toward higher levels of intelligence. Effectively applying these technologies to improve functions such as S&O ▲, content generation, and user behavior modeling and prediction will be a key focus of future research.

4.9. Limitations

While we have strived for a comprehensive and systematic review, this survey has several inherent limitations that we acknowledge here to help readers accurately understand its scope and conclusions.

First, our review is subject to limitations in its literature selection scope and potential biases. Our literature search was primarily conducted on Google Scholar using a curated set of English keywords. This strategy may have led us to overlook relevant research indexed in other databases, published in languages other than English, or not captured by our keyword strategy. Furthermore, our survey has focused mainly on published academic papers and has not systematically covered commercial systems or patents from industry, which may also represent significant technological applications in this field.

Second, our proposed classification framework has its own inherent limitations. The dual-dimension analytical framework is intended to provide a clear, function-oriented perspective, but it is neither the only nor an absolute method of categorization. Many advanced reading assistance systems are inherently multi-functional and complex, and forcing them into a single category is sometimes a simplification made for the sake of clarity. For instance, some systems feature a deep integration of multiple techniques; in classifying them, we had to make a judgment based on their most central innovation, which may not fully capture the entire complexity of their design.

Finally, the boundary of this survey's core scope—the "single-text" context—is becoming increasingly blurred. This review aims to distinguish itself from "text collection analysis", which deals with large-scale document sets. However, in contemporary digital reading practices, this boundary is fluid. For example, an advanced academic paper reader might automatically link to cited references or fetch term definitions from external knowledge bases while assisting with a single paper. Technically, this already involves multi-document information. While our survey covers such features (e.g., under Information Supplementation), it does not delve deeply into the complex interaction patterns that support cross-document knowledge synthesis. Defining and analyzing this emerging scenario of "single-text-centric yet highly interconnected" reading remains a challenge not fully addressed by this survey.

We acknowledge these limitations and hope that future work will complement and extend this survey by considering a broader range of literature, adopting more diverse classification perspectives, and continuously tracking emerging reading scenarios.

5. Conclusion

This paper presents a structured survey of visualization and interaction techniques for enhancing single-text digital reading. We propose an analytical framework based on two dimensions: a taxonomy of data types (structural information, content elements, and user interaction data) and a taxonomy of visualization techniques (TP ●, IC ●, LO ●, and IE ●). Based on this framework, we have compared the strengths and weaknesses of representative systems and identified several remaining challenges, particularly in handling complex documents, supporting diverse reading goals, and effectively leveraging natural language processing techniques.

The contribution of this paper, however, extends beyond this technical review and classification. More importantly, it reveals a profound shift in the field's design philosophy: the research focus is gradually transitioning from pure "Information Presentation" to providing "Cognitive Scaffolding", that is, designing to directly support the user's process of comprehension, analysis, and knowledge construction. Guided by this philosophy, we have further distilled a systematic design philosophy that provides actionable principles for preserving cognitive flow, empowering active knowledge construction, and safeguarding user agency in the age of Artificial Intelligence. It advocates for the evolution of reading assistance tools from static information browsers into dynamic cognitive partners that think alongside the reader.

Realizing this vision requires breakthroughs in several key directions. First, deepening semantic understanding is necessary, moving beyond surface-level analysis to accurately capture complex discourse structures and implicit meanings. Second, achieving true adaptive support requires fine-grained user modeling, enabling systems to dynamically interpret user interaction data and tailor their assistance strategies in real-time. Finally, as documents become increasingly multi-modal, ensuring semantic coherence across heterogeneous content such as text, charts, and videos will be critical for delivering a holistic and effective reading experience.

In summary, single-text reading assistance is evolving from a collection of function-driven tools into a research frontier centered on the user's cognitive process, deeply intersecting with the domains of information visualization, natural language processing, and human-computer interaction. We hope that the analytical framework and the proposed design philosophy will serve as a systematic guide and inspiration for future research and practice in this field.

Acknowledgments

The authors thank anonymous reviewers for their valuable comments. This work is supported by the General Program for Natural Science Research of Basic Disciplines in Universities of Jiangsu Province (24KJB520004), the Scientific Research Start-up Fund of Jiangsu Ocean University (KQ24051), the National Science Foundation of China (NSFC) (72174079), Key R&D Program of Lianyungang (CG2323).

References

[1] A. B. Alencar, M. C. F. de Oliveira, F. V. Paulovich, Seeing beyond reading: A survey on visual text analytics, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2 (6) (2012) 476–492.

[2] S. Havre, E. Hetzler, P. Whitney, L. Nowell, Themeriver: Visualizing thematic changes in large document collections, IEEE Transactions on Visualization and Computer Graphics 8 (1) (2002) 9–20.

[3] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, X. Tong, Textflow: Towards better understanding of evolving topics in text, IEEE Transactions on Visualization and Computer Graphics 17 (12) (2011) 2412–2421.

[4] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, B. Dorr, Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization, Journal of the American Society for Information Science and Technology 63 (12) (2012) 2351–2369.

[5] N. Cao, J. Sun, Y.-R. Lin, D. Gotz, S. Liu, H. Qu, Facetatlas: Multifaceted visualization for rich text corpora, IEEE Transactions on Visualization and Computer Graphics 16 (6) (2010) 1172–1181.

[6] S. Jänicke, G. Franzini, M. F. Cheema, G. Scheuermann, On close and distant reading in digital humanities: A survey and future challenges, in: Proceedings of the Eurographics Conference on Visualization (EuroVis 2015) – State of the Art Reports, 2015, pp. 83–103.

[7] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, D. A. Keim, Bridging text visualization and mining: A task-driven survey, IEEE Transactions on Visualization and Computer Graphics 25 (7) (2018) 2482–2504.

[8] M. Alharbi, R. S. Laramee, SoS TextVis: A survey of surveys on text visualization, in: Proceedings of the Conference on Computer Graphics & Visual Computing, 2018, p. 143–152.

[9] X. Wang, S. L. Huey, R. Sheng, S. Mehta, F. Wang, Scidasynth: Interactive structured knowledge extraction and synthesis from scientific literature with large language model, CoRR abs/2404.13765 (2024).

[10] A. W. Yu, D. Dohan, M.-T. Luong, R. Zhao, K. Chen, M. Norouzi, Q. V. Le, Qanet: Combining local convolution with global self-attention for reading comprehension, in: Proceedings of the 6th International Conference on Learning Representations (ICLR), 2018.

[11] J. Graham, The reader's helper: A personalized document reading environment, in: Proceedings of the Conference on Human Factors in Computing Systems (CHI '99), 1999, pp. 481–488.

[12] E. H. Chi, L. Hong, M. Gumbrecht, S. K. Card, Scenthighlights: Highlighting conceptually-related sentences during reading, in: Proceedings of the 10th International Conference on Intelligent User Interfaces, 2005, pp. 272–274.

[13] C.-S. Wu, S. J. Robinson, A. Mazalek, Turning a page on the digital annotation of physical books, in: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, 2008, pp. 109–116.

[14] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, O. Deussen, Document cards: A top trumps visualization for documents, IEEE Transactions on Visualization and Computer Graphics 15 (6) (2009) 1145–1152.

[15] D. Oelke, D. Spretke, A. Stoffel, D. A. Keim, Visual readability analysis: How to make your writings easier to read, IEEE Transactions on Visualization and Computer Graphics 18 (5) (2011) 662–674.

[16] G. Golovchinksy, S. Carter, A. Dunnigan, Ara: The active reading application, in: Proceedings of the 19th ACM International Conference on Multimedia, 2011, pp. 799–800.

[17] C. S. Tashman, W. K. Edwards, Liquidtext: A flexible, multitouch environment to support active reading, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2011, pp. 3285–3294.

[18] A. Stoffel, H. Strobelt, O. Deussen, D. A. Keim, Document thumbnails with variable text scaling, Computer Graphics Forum 31 (3) (2012) 1165–1173.

[19] K. Hinckley, X. Bi, M. Pahud, B. Buxton, Informal information gathering techniques for active reading, in: Proceedings of the 2012 ACM Conference on Human Factors in Computing Systems (CHI '12), 2012, pp. 1893–1896.

[20] J. Pearson, T. Owen, H. Thimbleby, G. R. Buchanan, Co-reading: Investigating collaborative group reading, in: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, 2012, pp. 325–334.

[21] V. K. Chaudhri, B. Cheng, A. Overtholtzer, J. Roschelle, A. Spaulding, P. Clark, M. Greaves, D. Gunning, Inquire biology: A textbook that answers questions, AI Magazine 34 (3) (2013) 55–72.

[22] V. Singh, Gestannot: A paper annotation tool for tablet, Ph.D. thesis, Dalhousie University (2013).

[23] B. C. Kwon, F. Stoffel, D. Jäckle, B. Lee, D. Keim, Visjockey: Enriching data stories through orchestrated interactive visualization, in: Proceedings of the Computation+ Journalism Symposium, Vol. 3, 2014, p. 3.

[24] F. Stoffel, D. Jäckle, D. A. Keim, Enhanced news-reading: Interactive and visual integration of social media information, in: Proceedings of the VisLR Workshop: Visualization as Added Value in the Development, Use and Evaluation of Language Resources at the Ninth International Conference on Language Resources and Evaluation (LREC 2014), 2014, pp. 21–28.

[25] S. Koch, M. John, M. Wörner, A. Müller, T. Ertl, Varifocalreader—in-depth visual analysis of large text documents, IEEE Transactions on Visualization and Computer Graphics 20 (12) (2014) 1723–1732.

[26] T. Gao, J. R. Hullman, E. Adar, B. Hecht, N. Diakopoulos, Newsviews: An automated pipeline for creating custom geovisualizations for news, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2014, pp. 3005–3014.

[27] A. J. Wecker, J. Lanir, O. Mokryn, E. Minkov, T. Kuflik, Semantize: Visualizing the sentiment of individual document, in: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, 2014, pp. 385–386.

[28] J. Kim, E.-Y. Ko, J. Jung, C. W. Lee, N. W. Kim, J. Kim, Factful: Engaging taxpayers in the public discussion of a government budget, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 2843–2852.

[29] V. Gold, C. Rohrdantz, M. El-Assady, Exploratory text analysis using lexical episode plots, in: Proceedings of the 17th Eurographics Conference on Visualization (EuroVis 2015) - Short Papers, 2015, pp. 85–89.

[30] M. El-Assady, V. Gold, C. Acevedo, C. Collins, D. A. Keim, Contovi: Multi-party conversation exploration using topic-space views, Computer Graphics Forum 35 (3) (2016) 431–440.

[31] T. Abekawa, A. Aizawa, Sidenoter: Scholarly paper browsing system based on pdf restructuring and text annotation, in: Proceedings of the 26th International Conference on Computational Linguistics: System Demonstrations, 2016, pp. 136–140.

[32] N. Nicolas, Exploring playful annotations in interactive textbooks: Engaging the teacher and the learner in an active learning process, Master's thesis, Malmö högskola, Faculty of Culture and Society (KS) (2016).

[33] Y. Wang, D. Liu, H. Qu, Q. Luo, X. Ma, A guided tour of literature review: Facilitating academic paper reading with narrative visualization, in: Proceedings of the 9th International Symposium on Visual Information Communication and Interaction, 2016, pp. 17–24.

[34] H. Mehta, A. Bradley, M. Hancock, C. Collins, Metatation: Annotation as implicit interaction to bridge close and distant reading, ACM Transactions on Computer-Human Interaction (TOCHI) 24 (5) (2017) 1–41.

[35] Q. Yang, G. de Melo, Y. Cheng, S. Wang, Hitext: Text reading with dynamic salience marking, in: Proceedings of the 26th International Conference on World Wide Web Companion, 2017, pp. 311–319.

[36] S. K. Badam, Z. Liu, N. Elmqvist, Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading, IEEE Transactions on Visualization and Computer Graphics 25 (1) (2018) 661–671.

[37] D. H. Kim, E. Hoque, J. Kim, M. Agrawala, Facilitating document reading by linking text and tables, in: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 2018, pp. 423–434.

[38] R. Metoyer, Q. Zhi, B. Janczuk, W. Scheirer, Coupling story to visualization: Using textual analysis as a bridge between data and interpretation, in: Proceedings of the 23rd International Conference on Intelligent User Interfaces, 2018, pp. 503–507.

[39] F. Sperrle, R. Sevastjanova, R. Kehlbeck, M. El-Assady, Viana: Visual interactive annotation of argumentation, in: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, 2019, pp. 11–22.

[40] D. Kiesel, P. Riehmann, H. Wachsmuth, B. Stein, B. Froehlich, Visual analysis of argumentation in essays, IEEE Transactions on Visualization and Computer Graphics 27 (2) (2020) 1139–1148.

[41] H. Subramonyam, C. M. Seifert, P. Shah, E. Adar, Texsketch: Active diagramming through pen-and-ink annotations, in: Proceedings of the Conference on Human Factors in Computing System, 2020, pp. 1–13.

[42] R. Sevastjanova, M. El-Assady, A. Bradley, C. Collins, M. Butt, D. Keim, Visinreport: Complementing visual discourse analytics through personalized insight reports, IEEE Transactions on Visualization and Computer Graphics 28 (12) (2021) 4757–4769.

[43] A. Head, K. Lo, D. Kang, R. Fok, S. Skjonsberg, D. S. Weld, M. A. Hearst, Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–18.

[44] N. Sultanum, F. Chevalier, Z. Bylinskii, Z. Liu, Leveraging text-chart links to support authoring of data-driven articles with vizflow, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–17.

[45] L. L. Wang, I. Cachola, J. Bragg, E. Y.-Y. Cheng, C. Haupt, M. Latzke, B. Kuehl, M. N. van Zuylen, L. Wagner, D. Weld, Scia11y: Converting scientific papers to accessible html, in: Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, 2021, pp. 1–4.

[46] R. Fok, A. Head, J. Bragg, K. Lo, M. A. Hearst, D. S. Weld, Scim: Intelligent faceted highlights for interactive, multi-pass skimming of scientific papers, arXiv: 2205.04561 (2022).

[47] H. L. Han, J. Yu, R. Bournet, A. Ciorascu, W. E. Mackay, M. Beaudouin-Lafon, Passages: Interacting with text across documents, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–17.

[48] A. Head, A. Xie, M. A. Hearst, Math augmentation: How authors enhance the readability of formulas using novel visual design practices, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–18.

[49] H. Kang, J. C. Chang, Y. Kim, A. Kittur, Threddy: An interactive system for personalized thread-based exploration and organization of scientific literature, in: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, 2022, pp. 1–15.

[50] Z. Peng, Y. Liu, H. Zhou, Z. Xu, X. Ma, Crebot: Exploring interactive question prompts for critical paper reading, International Journal of Human-Computer Studies 167 (2022) 102898.

[51] N. Rachatasumrit, J. Bragg, A. X. Zhang, D. S. Weld, Citeread: Integrating localized citation contexts into scientific paper reading, in: Proceedings of the 27th International Conference on Intelligent User Interfaces, 2022, pp. 707–719.

[52] T. August, L. L. Wang, J. Bragg, M. A. Hearst, A. Head, K. Lo, Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing, ACM Transactions on Computer-Human Interaction 30 (5) (2023) 1–38.

[53] T. Cai, A. G. Niklaus, M. Kraley, B. Kerr, Z. Bylinskii, Therif: A pipeline for generating themes for readability with iterative feedback, arXiv preprint arXiv:2303.04221 (2023).

[54] J. C. Chang, A. X. Zhang, J. Bragg, A. Head, K. Lo, D. Downey, D. S. Weld, Citesee: Augmenting citations in scientific papers with persistent and personalized historical context, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2023, pp. 737:1–737:15.

[55] X. A. Chen, C.-S. Wu, L. Murakhovs'ka, P. Laban, T. Niu, W. Liu, C. Xiong, Marvista: Exploring the design of a human-AI collaborative news reading tool, ACM Transactions on Computer-Human Interaction 30 (6) (2023) 1–27.

[56] N. Chulpongsatorn, M. S. Lunding, N. Soni, R. Suzuki, Augmented math: Authoring ar-based explorable explanations by augmenting static math textbooks, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023, pp. 1–15.

[57] R. Fok, J. C. Chang, T. August, A. X. Zhang, D. S. Weld, Qlarify: Recursively expandable abstracts for directed information retrieval over scientific papers, arXiv preprint arXiv:2310.07581 (2023).

[58] H. B. Kang, T. Wu, J. C. Chang, A. Kittur, Synergi: A mixed-initiative system for scholarly synthesis and sensemaking, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023, pp. 1–19.

[59] T. S. Kim, M. Latzke, J. Bragg, A. X. Zhang, J. C. Chang, Papeos: Augmenting research papers with talk videos, in: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023, pp. 1–19.

[60] D. Masson, S. Malacria, G. Casiez, D. Vogel, Charagraph: Interactive generation of charts for realtime annotation of data-rich paragraphs, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–18.

[61] J. Milbauer, Z. Ding, Z. Wu, T. Wu, Newssense: Reference-free verification via cross-document comparison, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2023, pp. 422–430.

[62] S. Palani, A. Naik, D. Downey, A. X. Zhang, J. Bragg, J. C. Chang, Relatedly: Scaffolding literature reviews with existing related work sections, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–20.

[63] S. Park, C. Lee, J. Han, U. Lee, Quickref: Should I read cited papers for understanding this paper?, in: Proceedings of Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–7.

[64] L. Richards Maldonado, A. Abouzied, N. W. Gleason, Readerquizzer: Augmenting research papers with just-in-time learning questions to facilitate deeper understanding, in: Proceedings of Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing, 2023, pp. 391–394.

[65] Y. Wang, Y.-S. Kim, Making data-driven articles more accessible: An active preference learning approach to data fact personalization, in: Proceedings of the ACM Designing Interactive Systems Conference, 2023, pp. 1353–1366.

[66] X. Zhang, J. Li, P.-W. Chi, S. Chandrasegaran, K.-L. Ma, Concepteva: Concept-based interactive exploration and customization of document summaries, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–16.

[67] D. Zyska, N. Dycke, J. Buchmann, I. Kuznetsov, I. Gurevych, Care: Collaborative ai-assisted reading environment, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL 2023), 2023, pp. 291–303.

[68] Y. Fu, S. Guo, J. Hoffswell, V. S. Bursztyn, R. Rossi, J. Stasko, "The Data Says Otherwise"—Towards automated fact-checking and communication of data claims, in: Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, 2024, pp. 1–20.

[69] H. Kambhamettu, J. Flores, A. Head, Traceable text: Deepening reading of ai-generated summaries with phrase-level provenance links, CoRR abs/2409.13099 (2024).

[70] H. Kim, K. D. Le, G. Lim, D. H. Kim, Y. J. Hong, J. Kim, Datadive: Supporting readers' contextualization of statistical statements with data exploration, in: Proceedings of the 29th International Conference on Intelligent User Interfaces, 2024, pp. 623–639.

[71] M. X. Liu, T. Wu, T. Chen, F. M. Li, A. Kittur, B. A. Myers, Selenite: Scaffolding online sensemaking with comprehensive overviews elicited from large language models, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–26.

[72] L. Melin-Higgins, Ai in document interaction: An interaction design approach to enhancing reading practices, Bachelor's Thesis, Malmö University (2024).

[73] H. Y. Ng, Z. He, T.-H. Huang, What color scheme is more effective in assisting readers to locate information in a color-coded article?, arXiv preprint arXiv:2408.06494 (2024).

[74] P. Pitre, K. Luther, Argumentor: Augmenting user experiences with counter-perspectives, CoRR abs/2406.02795 (2024).

[75] N. Singh, L. L. Wang, J. Bragg, Figura11y: Ai assistance for writing scientific alt text, in: Proceedings of the 29th International Conference on Intelligent User Interfaces, 2024, pp. 886–906.

[76] X. Zhang, V. Dörig, P. Cui, V. Zouhar, T. H. Netland, M. Sachan, Harmonizing assistance: Moderating visual and textual aids in ai-enhanced textbook reading with iread, Preprint at Research Square (2024).

[77] A. Gunturu, Y. Wen, N. Zhang, J. Thundathil, R. H. Kazi, R. Suzuki, Augmented physics: Creating interactive and embedded physics simulations from static textbook diagrams, in: Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology, 2024, pp. 1–12.

[78] A. Gunturu, S. Jadon, N. Zhang, M. Faraji, J. Thundathil, T. Ahmad, W. Willett, R. Suzuki, Realitysummary: Exploring on-demand mixed reality text summarization and question answering using large language models, arXiv preprint arXiv:2405.18620 (2024).

[79] R. Zou, Y. Tang, J. Chen, S. Lu, Y. Lu, Y. Yang, C. Ye, Gistvis: Automatic generation of word-scale visualizations from data-rich documents, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), Yokohama, Japan, 2025, pp. 679:1–679:18.

[80] Z. Shao, L. Shen, H. Li, Y. Shan, H. Qu, Y. Wang, S. Chen, Narrative player: Reviving data narratives with visuals, IEEE Trans. Vis. Comput. Graph. 31 (10) (2025) 6781–6795.

[81] M. Dück, S. Holter, R. S. M. Chan, R. Sevastjanova, M. El-Assady, Finding Needles in Document Haystacks: Augmenting serendipitous claim retrieval workflows, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, 2025, pp. 1003:1–1003:17.

[82] K. Reiss, E. F. Rhodes, Translation Criticism-Potentials and Limitations: Categories and Criteria for Translation Quality Assessment, Routledge, 2014.

[83] M. El-Assady, A. Hautli-Janisz, V. Gold, M. Butt, K. Holzinger, D. A. Keim, Interactive visual analysis of transcribed multi-party discourse, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, System Demonstrations, 2017, pp. 49–54.

[84] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, Language models are few-shot learners, Advances in Neural Information Processing Systems 33 (2020) 1877–1901.

[85] L. M. Holloway, C. Goncu, A. Ilsar, M. Butler, K. Marriott, Infosonics: Accessible infographics for people who are blind using sonification and voice, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–13.

[86] Y. Dzhurynskyi, V. Mayik, L. Mayik, Enhancing accessibility: Automated tactile graphics generation for individuals with visual impairments, Computation 12 (12) (2024) 251.

[87] A. A. M. H. Al-Ahdal, Ebook interaction logs as a tool in predicting learner performance in reading, Asiatic: IIUM Journal of English Language and Literature 14 (1) (2020) 174–188.

[88] S. Tang, R. G. Reilly, C. Vorstius, Eyemap: A software system for visualizing and analyzing eye movement data in reading, Behavior Research Methods 44 (2) (2012) 420–438.

[89] O. Špakov, H. Siirtola, H. Istance, R. Kari-Jouko, Visualizing the reading activity of people learning to read, Journal of Eye Movement Research 10 (5) (2017) 10–16910.

[90] J. Santhosh, A. P. Pai, S. Ishimaru, Toward an interactive reading experience: Deep learning insights and visual narratives of engagement and emotion, IEEE Access 12 (2024) 6001–6016.

[91] E. K. Choe, B. Lee, H. Zhu, N. H. Riche, D. Baur, Understanding self-reflection: How people reflect on personal data through visual data exploration, in: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, 2017, pp. 173–182.

[92] P. Haghighatkhah, M. El-Assady, J.-D. Fekete, N. Mahyar, C. Paradis, V. Simaki, B. Speckmann, Characterizing uncertainty in the visual text analysis pipeline, in: Proceedings of the IEEE 7th Workshop on Visualization for the Digital Humanities (VIS4DH), 2022, pp. 25–30.

[93] S. Dhuliawala, V. Zouhar, M. El-Assady, M. Sachan, A diachronic perspective on user trust in ai under uncertainty, arXiv preprint arXiv:2310.13544 (2023).

[94] E. Beauxis-Aussalet, M. Behrisch, R. Borgo, D. H. Chau, C. Collins, D. Ebert, M. El-Assady, A. Endert, D. A. Keim, J. Kohlhammer, The role of interactive visualization in fostering trust in AI, IEEE Computer Graphics and Applications 41 (6) (2021) 7–12.

[95] L. Klein, C. Lüth, U. Schlegel, T. Bungert, M. El-Assady, P. Jäger, Navigating the maze of explainable AI: A systematic approach to evaluating methods and metrics, Advances in Neural Information Processing Systems 37 (2024) 67106–67146.