ARTIST: Automated Text Simplification for Task Guidance in Augmented Reality

Guande Wu guandewu@nyu.edu New York University Brooklyn, New York, USA

Shaoyu Chen sc6439@nyu.edu New York University Brooklyn, New York, USA Jing Qian jq2267@nyu.edu New York University Brooklyn, New York, USA

Joao Rulff jlrulff@nyu.edu New York University Brooklyn, New York, USA Sonia Castelo s.castelo@nyu.edu New York University Brooklyn, New York, USA

Claudio Silva csilva@nyu.edu New York University Brooklyn, New York, USA



Figure 1: We propose ARTiST, a text simplification system that is designed for augmented reality (AR) head-mounted display (HMD) environments. Our system combines the findings from a formative study with a novel few-shot prompting to integrate four established text simplification techniques for AR-specific contexts. The example text shown in the bottom-right corner of the figure has been simplified using our approach. The red text indicates removals whereas the green highlights the addition of spatial information. The resulting simplified text is displayed directly in the HMD.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). ACM ISBN 979-8-4007-0330-0/24/05.

https://doi.org/10.1145/3613904.3642669

ABSTRACT

Text presented in augmented reality provides in-situ, real-time information for users. However, this content can be challenging to apprehend quickly when engaging in cognitively demanding AR tasks, especially when it is presented on a head-mounted display. We propose ARTiST, an automatic text simplification system that uses a few-shot prompt and GPT-3 models to specifically optimize the text length and semantic content for augmented reality. Developed out of a formative study that included seven users and

three experts, our system combines a customized error calibration model with a few-shot prompt to integrate the syntactic, lexical, elaborative, and content simplification techniques, and generate simplified AR text for head-worn displays. Results from a 16-user empirical study showed that ARTiST lightens the cognitive load and improves performance significantly over both unmodified text and text modified via traditional methods. Our work constitutes a step towards automating the optimization of batch text data for readability and performance in augmented reality.

CCS CONCEPTS

• Human-centered computing \rightarrow Mixed / augmented reality.

KEYWORDS

augmented reality, text simplification, large language model

ACM Reference Format:

Guande Wu, Jing Qian, Sonia Castelo, Shaoyu Chen, Joao Rulff, and Claudio Silva. 2024. ARTiST: Automated Text Simplification for Task Guidance in Augmented Reality. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3613904.3642669

1 INTRODUCTION

Augmented reality (AR) has evolved into a transformative technology with far-reaching applications across multiple domains, including education [2, 45, 111], entertainment [60, 72, 73, 77], collaborative work [12, 13, 29], and professional training [9, 13, 98, 122]. Notably, AR superimposes digital content onto the physical world in real-time to facilitate more efficient task execution [75, 112]. As a result, AR applications have been increasingly employed for task guidance in manufacturing [71, 125], education [8], and surgery [72]. AR devices have been widely adopted in the manufacturing industry, for example, to reduce reliance on guidance materials or other devices outside of the immediate work environment [54].

A head-mounted display (HMD) is a type of AR device that allows for multi-modal interactions while the user maintains focus on the immediate work environment [31, 54]. Given their hands-free nature, HMDs are frequently used for text-based task guidance. However, compared to desktop displays, HMDs have a relatively small field of view (FoV) limiting the amount of text they can display; longer instructions may occlude a user's view resulting in lower-productivity and higher cognitive load [18], as well as safety risks [72]. As a result, AR text-based instructions require optimization for better utility.

Text simplification offers one potential solution. This process has historically been used to reduce the complexity or length of text for users [94] in non-AR settings, making it more readily understandable. To the best of our knowledge, however, there are presently no established methods for adapting and simplifying text for better utility in the specific context of AR. Furthermore, applying existing methods to AR raises several concerns. Firstly, traditional text simplification methods typically work to facilitate comprehension for individuals with limited reading skills [23, 88], an audience that may not overlap with the AR userbase. Secondly, these methods have not been designed or fine-tuned to accommodate AR-specific constraints, such as the small FoV, a restricted

display area, or the necessity of users performing physical tasks concurrent with reading [72]. Finally, text simplification presents an opportunity to use spatial information AR content by describing a physical object's color, location, or direction. Textually indicating the location of a physical object can, for example, assist users in AR task execution [72, 125].

Accordingly, we aim to implement a text simplification system for AR by tailoring the existing methods to the AR context, with the goal of reducing cognitive load on users and improving their task performance. To do so, we build on insights from prior work [25, 88, 93, 94] and our own formative study to understand the specific challenges of AR text interfaces as well as their limitations and potentials. The formative study contains three parts: a literature survey, an open-ended exploration, and an expert interview. We found that both participants and experts addressed issues related to long-text-induced reading challenges (e.g., cognitive load) and comprehension. Interviews with participants and experts further elicited three design guidelines that helped build ARTiST, an automated text simplification system with few-shot prompting. This system leverages the multi-task capabilities of the large language model (LLM) GPT-3 in combination with our newly formed simplification techniques, eliminating the need for extensively annotated data [79]. We crafted prompts based on chain-of-thought principles, considering text simplification in AR [107]. Specifically, ARTiST introduces two novel simplification methods: the "plan-of-technique" prompt and "error-aware" model calibration, enhancing the effectiveness and reliability of text simplification.

We tested ARTiST via two studies that entailied an empirical evaluation with 16 participants. The first study included task guidance to make pour-over coffee and set up a meeting room according to specific criteria. The second study asked participants to perform video editing on an iPad using AR instructions given through HMD. These studies explore how our proposed system can better benefit participants over the unmodified text and existing methods by assessing related performance metrics, cognitive load, and subjective ratings. The results indicate that ARTiST significantly improved task guidance performance by reducing the number of errors participants made, increasing the number of steps they correctly memorized, and reducing their cognitive load.

In summary, we contribute:

- (1) The ARTiST, a novel system for text simplification in AR using few-shot prompts and customized GPT-3. This system incorporates chain-of-thought, plan-of-technique, and erroraware calibration to tailor text simplification for AR.
- (2) Results and design guidelines from a formative study that includes a literature review, an open-ended exploration with seven participants, and an expert interview with three field experts for text simplification in AR.
- (3) A 16-participant empirical evaluation of ARTiST against baseline and existing methods, which shows that ARTiST significantly reduces errors and overall cognitive load with similarly higher subjective ratings on text readability, memorability, guidance, and trust among users.

To further support the development of the field, we open-source our implementation 1 .

 $^{^{1}}Code\ is\ available\ at\ https://github.com/VIDA-NYU/artist$

2 RELATED WORK

Our research draws inspiration from recent studies in task guidance, AR text display, and text simplification. As text simplification establishes the base for our formative study, we discuss it in detail in the following section akin to [48, 120].

2.1 Task Guidance in AR

HMDs have been increasingly used to guide users in procedural tasks, such as cooking [24, 116], surgery [11], and maintenance [71, 125]. They allow the seamless and interactive display of task elements and steps in the physical working space [9, 57, 99]. The concept of a task guidance system was initially proposed by Ockerman et al., who envisioned it as a reference and guide for procedural tasks [68] such as inspection and assembly. For example, AR-based maintenance task guidance and manuals have been developed to display technical instructions interactively in the workspace. Experiments in engine and factory maintenance demonstrate that such manuals can, in comparison to traditional paper manuals, reduce cognitive load on users and enhance task performance [32, 101]. Further studies confirm that AR task guidance can lead to a reduced error rate and increased user satisfaction [97, 123]. Such benefits were also observed in the military [40] and medical applications [11]. Furthermore, the visual aspect of textual information can be integrated to increase user engagement and highlight important information [51, 65, 87]. However, despite the widespread application of AR task guidance, most existing approaches directly transfer paper-based manuals to AR devices without further customization [32, 51] or simplification, neglecting the reading challenges posed by AR's FoV. Our approach seeks to bridge this gap by first comprehending the unique challenges of AR as identified in our formative study and then implementing an automatic text simplification method based on the guidelines derived from it.

2.2 Text Presentation in AR

Text presentation is an essential function of AR devices; it features in applications that require simple text display through to more robust task guidance. AR text display is especially useful when the text aligns with and supplements the physical environment and the task at hand [9, 76]. However, AR text display poses challenges due to the potential for occlusion of the physical environment. Hardware limitations, such as limited FoV, require that AR developers strategically position text within a constrained display are. To address potential occlusion, solutions that make use of depth information [43, 108], inter-frame motion [56] and the 2D collision detection algorithms [15, 100] have been proposed. Though occlusion and collision are not entirely avoidable due to small display areas and the complexity of physical environments [21], we aim to mitigate them by simplifying sentence structure and length. Moreover, text presentation can be positively influenced by font selection, placement, and style [63, 69]. Orlosky [69] proposed an automated text placement algorithm adaptive to physical and virtual backgrounds. Rzayev et al.further conducted an empirical study on text display type and position in sitting and walking scenarios [85], concluding that top-right placement may not be optimal due

to increased subjective load and reduced comprehensibility. Matsuura et al. later investigated the readability and legibility of six different Japanese fonts used on HMDs for users who were walking and found that fonts with very thin horizontal lines, or thin horizontal and vertical lines, should not be used on HMDs due to decreased text readability caused by vertical shocks [63]. Text coordinates also contribute to the legibility of text presentation in AR applications. Three different coordinates are widely used in AR: body-locked, world-locked, and head-locked [12]. Body-locked displays are based on the user's body position and adapt to the user's movements while walking, making them most suitable for scenarios where users need to walk to perform tasks.

Existing methods for text presentation in AR thus focus on font style and layout, but largely overlook the optimization of the text content itself. Our research addresses this present oversight by utilizing text simplification to save display space but preserve essential content.

3 FORMATIVE STUDY

This study involves three parts: a literature review, an open-ended exploration, and expert interviews to understand the needs around text simplification in AR. We wish to explore the following aspects of text simplification:

- [RQ1] Which text simplification methods from the field of natural language processing (NLP) can be effectively applied in an AR context?
- [RQ2] Would text simplification improve comprehension, per its benefit for low-literacy readers in traditional platforms?
- [RQ3] Can text simplification lead to increased user satisfaction, and hence to a more positive AR experience overall?

3.1 Part I: Survey of Text Simplification

To address 3, we initiated a comprehensive review of existing literature on traditional text simplification; our goal was to identify NLP techniques that might be useful for AR applications. Our review commenced with an in-depth examination of three seminal survey papers [3, 88, 93]. We extended our scope by traversing both the references cited in these papers and consequent citations of them to gain an encompassing understanding of current methodological approaches. Subsequently, we identified four NLP techniques pertinent to our inquiry: content reduction (A1), syntactic simplification (A2), lexical simplification (A3), and elaborative simplification (A4). We describe the four techniques below.

- 3.1.1 A1: Content reduction. Content reduction in text simplification aims to achieve clarity and conciseness by eliminating or restructuring non-essential elements without altering the core message [66]. Strategies include removing non-essential information, shortening sentences, and eliminating repetition. This technique is particularly beneficial in constrained display environments like those of AR, where succinct, clear text enhances user interaction and comprehension [14].
- 3.1.2 A2: Syntactic simplification. Syntactic simplification involves rephrasing complex grammatical structures into simpler ones while still retaining the original meaning [86, 93]. Existing methodologies

often target specific complex linguistic features such as coordination, subordination, relative clauses, passive constructions, and extended sentence lengths [3, 93].

- 3.1.3 A3: Lexical simplification. Lexical complexity often arises from the use of intricate words and phrases. To mitigate this, one widely employed strategy is to replace complex lexical items with simpler synonyms. This form of lexical simplification has seen extensive application in the context of second-language learning, primarily because it aids in comprehension and vocabulary acquisition for learners who may not be familiar with advanced or specialized terminology [70].
- 3.1.4 A4: Elaborative simplification. Elaborative simplification entails providing explanations of complex concepts. This technique is prevalent in professional textbooks, which frequently encompass specialized or technical subject matters [49]. In AR, spatial information becomes increasingly critical for user comprehension and task performance. Consequently, elaborative simplification can be especially beneficial for clarifying spatial metrics and locations. Spatial metrics refer to numerical measurements, such as distances or sizes denoted in units like inches or centimeters. These metrics often need to be elaborated to provide context or improve comprehension. Similarly, spatial locations, which may involve GPS coordinates or relational positioning (e.g., "next to," "above," "beneath"), can be clarified through elaborative simplification to facilitate user orientation and task execution in AR environments [44].
- 3.1.5 Target application and users. Traditional text simplification techniques are normally targeted at non-native speakers or people with cognitive or literacy limitations, e.g., autism [26, 30, 113], aphasia [20, 22, 67], dyslexia [35, 41, 46, 82–84], hearing impairment [4–7, 102] and language learners [59]. The associated benefits are largely attributed to simplified grammar structures and the use of common words, which can significantly reduce information processing time.

In a similar vein, AR users may encounter reduced reading capability due to the challenges associated with the AR setting. Studies have demonstrated that AR users experience reduced reading speed [81], lower comprehension [14, 33, 34], and increased cognitive load [28]. For instance, Rau et al. report that readers' response time in AR is longer than that associated with desktop reading [81]. Hardware limitations comprise a major contributing factor, impacting refresh rate, resolution, and FoV, and ultimately impeding text display due to registration errors [42, 61], extra latency, and visual artifacts. Moreover, users' rapid movements and the surrounding open environment can result in unstable text displays. Prolonged use of AR devices may also lead to eye strain and fatigue due to constant accommodation and vergence adjustments, making reading more challenging than on traditional displays. Finally, AR displays often overlay digital information on real-world information, affecting reading comprehension and focus.

Yet, the main reason for the reduced legibility of AR text is the user's elevated cognitive load in the immersive environment; evidence shows that AR users can experience high pressure and increased cognitive load [28].

Accordingly, drawing inspiration from the research elaborated above, we aim to investigate whether text simplification techniques

can benefit users in AR environments and mitigate the challenges described.

3.2 Part II: Open-Ended Exploration

To explore the effectiveness of the four previously identified text simplification techniques (A1-4) in AR and further investigate 3, we conducted an open-ended exploration with seven participants. According to the literature, the simplification techniques (A1-4) can improve comprehension in paper-based reading. However, these techniques may need to be modified for AR-specific challenges and their benefits in AR require further investigation. Therefore, the open-ended exploration aimed to assess how text simplification techniques might be revised for this context.

- 3.2.1 Participants. Seven participants (four male and three female) were recruited from a school mailing list to experience text simplification in AR. Three out of the seven are native English speakers. All participants have some prior experience with AR (2/7 are frequent users, 4/7 are occasional users, 1/7 is VR-only).
- 3.2.2 Tasks. The open-ended exploration involves two tasks: cooking and gem-hunting. These tasks were selected for being common and applicable to AR scenarios [53, 62, 116]. In the cooking task, participants used the AR system to make a pinwheel sandwich. The AR interface showed step-by-step instructions for ingredient preparation, assembly, and cooking [14]. These instructions are adapted from a wikiHow article on how to make a pinwheel. ² For the gem-hunting task, participants followed clues displayed on the AR device to find a gem hidden in a room. Clues included puzzles, patterns, and spatial information. The task manual is derived from a party game website.
- 3.2.3 Method. Since each task contains multiple steps, the original and simplified text for each step were displayed side-by-side to participants. . Text simplification was manually performed based on the principles of the four existing techniques taken from the literature (i.e., content reduction, syntactic simplification, lexical simplification, and elaborative simplification), with each simplification technique being used an equal number of times. Since this is an exploratory study, quantitative data is not collected. Participants are asked to think aloud while performing their AR tasks. A semi-structured interview collects participants' thoughts on text simplification in AR, its potential, and its challenges.
- 3.2.4 Procedure. Initially, participants completed both tasks using the original text instructions. They shared any challenges they faced in understanding the AR interface. Next, simplified versions of the text were presented. Participants compared and evaluated readability and comprehension. On average, the exploration lasted about one hour. We coded our interview notes and think-aloud notes and summarized participant feedback on the four text simplification techniques. The open-ended study was supervised by the university-approved IRB, and participants were compensated at an hourly rate of \$20.
- 3.2.5 Results. During the study, we found text content and semantics affect the reading experience in AR.

 $^{^2} https://www.wikihow.com/Make-a-Pinwheel \\$

Text length in AR. Text in an AR environment introduces unique challenges that are not present in traditional display mediums. Users can scroll or zoom to manage lengthy texts in conventional formats; these interactions are more challenging to execute in the AR setting [19, 64]. Occlusion and visual clutter are some of the issues pointed out by our participants (P2), who mentioned, "The displayed text takes up too much space and occludes the table." Lengthy text segments also distract users' attention away from physical tasks. P3 found it challenging to focus on the task of sliding floss under the tortilla, perpendicular to the length of the roll, due to the distracting nature of the extended text. These distractions sometimes pose safety risks: P7 was at risk of cutting their finger while engrossed in reading. Furthermore, text length negatively impacts how well information is retained as processing time increases with longer text segments [36, 50]. This was evident in the gem-finding task, where P4 and P5 forgot a crucial step that they had been given earlier after reading a lengthy sentence. Text length thus requires careful design in AR.

Feedback on AR text simplification techniques All participants agreed that *content reduction* is beneficial in AR. For instance, they found the sentence, Roll the tortilla into a log shape, more effective than the original text: Roll the tortilla from one end to the other into a log shape. Most participants mentioned that adding a clause to further explain text may not be necessary (preferring syntactic simplification). When asked about replacing complex words with simpler ones (lexical simplification), most participants (6/7) did not indicate word complexity as an issue. For example, the word perpendicular was not found to be more opaque than at the right angle too, and most of the participants (5/7) preferred perpendicular because it was shorter (4/7) and more precise (3/7). In addition, most participants (6/7) expressed that added details (elaborative simplification) were unnecessary. P6 said that "the 'which' clause is verbose and takes up too much space" (referring to the instruction use the keys to unlock the first drawer below the desk, which should be located to your right). For spatial elaboration, most (6/7) found it helpful when the reference object was present in the scene. P6 remarked that indicating, "'finger size' helps me make a quick estimate of the size." P3 commented that indicating a spatial location in the text is helpful, and a majority of participants (4/7) said that spatial information can complement AR spatial indicators such as bounding boxes or virtual arrows in the scene.

3.3 Part III: Expert Interview

To verify the initial insights gained from the literature review and the open-ended exploration, we further conducted a semi-structured interview with three experts from the industry. All interviewed experts possess extensive experience with AR task guidance systems. Our objective in these interviews was to address 3 by eliciting their insights on text simplification for AR and exploring potential usage scenarios.

- *3.3.1 Expert background.* Each of the three experts (E1-E3) interviewed has over three years of professional experience in AR interface development.
 - E1 is an AR interface designer at a research and development (R&D) company that is currently working on a HoloLens application to support field surgery. E1's users are primarily

- skilled professionals such as teachers and emergency medical technicians (EMTs) who use AR devices to instruct them as they identify and treat injuries such as gunshot wounds.
- E2 is an interface developer at a document solution corporation. E2 collaborated with engine mechanics to develop a HoloLens-based instructional application for displaying engine maintenance documents.
- E3 is an AR/VR researcher with top-tier publications and extensive experience in HoloLens application development.
 E3 has developed AR applications for everyday tasks such as cooking for non-professional users.
- 3.3.2 Method. The interview addressed the experts' backgrounds and experiences, the challenges of AR text interface design, their assessment of the need for text simplification, and the potential benefits and drawbacks associated with it. Additionally, we presented the four commonly used text simplification methods and solicited their opinions on them.

3.3.3 Results. We describe the results in the following subsections. Benefits of text simplification in AR All experts recognized the need to simplify text in AR. They believed this would reduce user impatience and the likelihood of mistakes. E2 said that mechanics, for instance, might be habituated to how they perform a specific task and so rush through it without noticing updates to the process. When the related instructional text is simplified, however, they are more likely to read the instructions. E2 explained: "One of the things that happens is the procedure changes. Users can easily go on a routine and assume they know how to do it without actually reading the instructions." E1 and E3 also mention that the simplified text could reduce the cognitive load and mitigate user anxiety, another set of benefits. For example, E3 indicated that: "Reading the long text may make the users anxious," but this may not be the case for shorter pieces of text.

All experts indicated that simplified text reduces the chance of visual occlusion. Object occlusion (virtual objects being blocked visually by physical objects [100]) is one such instance of this. This leads to users being unable or only partially able to read the AR text, causing frustration and diminished performance. E2 mentioned that: "(Sometimes in engine maintenance) we're gonna have a wall full of the tools, (and sometimes) we are gonna have an engine in front of you, and (so) finding someplace in the visual display is gonna be a challenge." E3 also mentioned that shortening and simplifying text could reduce occlusion.

Both E2 and E3 mentioned that shorter text facilitates the AR reading experience since zooming or scaling long sections of text while reading on an HMD is challenging. E2 emphasized that: "None of the users liked pinching and zooming," highlighting the need for methods that do not require additional interactions.

Finally, the experts mentioned the tremendous opportunity to use text simplification as a way to help automate the conversion of text from traditional digital media (e.g., PDF) to AR. All experts conveyed that the process of creating text instructions for AR is still sub-optimal and requires extra labor. E2 stated: "All the documents we work with start as PDF or Word documents. We basically output them to AR (devices)". In contrast, E1 and E3 mentioned the need to make modifications to the text displayed in AR. For example,

E1 attempts to shrink text or split long sections of text into multiple steps to make them shorter, saying: "We try to keep the words as quick, punchy, and actionable as possible." E3 also mentioned adjusting font sizes and colors to improve legibility in the AR environment. Although full-text automation involves fitting text to the AR scene with different formats, styles, or colors, E2 pointed out that automated text simplification would still be useful as the current manual approach requires expertise that novice workers may not possess. In addition, it is not feasible to manually revise all text when new sections are added regularly.

Challenges in text simplification in AR Current AR applications lack automated solutions and established practices for text simplification (E1-3). All experts concur that manual text revision is impractical due to the constant influx of new text and the absence of a standard framework for AR text readability. This drives home the need for automated methods to adapt existing documents for presentation in AR. However, text simplification for AR poses the following challenges, and current methods are not directly applicable (E1-3).

- All experts raised concern over avoiding accidental changes
 to meaning during text simplification. This concern is unique
 to AR because users perform physical actions live from textual instructions. E2 elaborated: "When working with mechanical systems in real-world scenarios, failure to follow instructions accurately could lead to catastrophic consequences,"
 highlighting the importance of retaining the integrity of the
 original text's meaning.
- Removing duplicated content is crucial in AR given that such redundancies could increase cognitive load for AR users who are already tasked with interpreting and acting upon visual overlays. All experts agreed that elaborative simplification should weigh toward eliminating redundancies rather than adding explanatory details, which is traditional in conventional text simplification.
- Traditional text simplification techniques must be re-adapted for AR (E1-3) as they are primarily geared toward enhancing readability for low-literacy individuals and do not address the attention constraints, high cognitive load, and FoV issues typical in AR. Therefore, the development of an AR-specific text simplification tool presents a challenging yet vital task, as it must harmonize these design goals to suit the unique demands of AR settings.

ID	Technique	E1	E2	E3
A1	Content reduction		√	
A2	Syntactic simplification	✓	√	✓
A3	Lexical simplification	✓		
A4	Elaborative simplification	√	√	√

Table 1: Expert (E1-3) feedback on simplification techniques (A1-4). A check indicates that the expert assesses that the given technique would be useful for AR.

3.3.4 Feedback on existing text simplification techniques. The table 1 summarizes the traditional simplification techniques our experts use in their everyday work. All experts do manual *content*

reduction when creating AR instructions. E3 employs lexical simplification with the aim of retaining the text's original meaning. All experts agree that simplifying syntax, length, and grammar is beneficial for AR interfaces. However, the use of elaborative simplification needs more scrutiny in AR settings as "the subtle balance between the content and text length must be considered" (E1, E2, E3). For example, E1 mentioned that engine maintenance manuals often include explanations of different engine parts that may be unfamiliar to users, and such explanations should not be removed. E2 brought up that both object and numeric elaboration can be beneficial when users need to quickly identify numerous targets in AR. Elaborating AR text to describe objects in the scene is one potential application. E2 explained that using a reference object that is similar in size to the dimensions given in the text (when the object is visible) would facilitate spatial awareness. For example, E2 said that a phrase like Move the handle to seven inches to left can be elaborated as Move the gear to seven inches left, or the length of a screwdriver. Again, as the experts mention, consideration needs to be given to balancing text length against the need for additional content in AR.

3.4 Design Guidelines and Updated Simplification Techniques

3.4.1 Design guidelines. Through summarizing the literature survey, the open-ended exploration, and insights shared by our AR experts, we derived design guidelines and updated the four selected simplification techniques for AR task guidance.

- [DG1] Meaning preservation is paramount in text simplification. Preserving the original text meaning [10, 94]) is the main objective when applying text simplification techniques. This finding is in line with both our interview sessions and observations. Since almost all simplification techniques may compromise original meaning [66, 93], it is essential that any substituted words convey the same meaning as their original counterparts [27].
- [DG2] Text simplification must consider both AR-specific challenges, such as issues with FoV and cognitive load, while exploring AR-specific opportunities. Traditional text simplification techniques (e.g., syntactic simplification, lexical simplification, etc.) do not address challenges associated with AR devices, such as reading the overlayed text while doing a physical task, the constraints of a small FoV, and users' increased cognitive load while completing a task. Minimizing the display space required to render text reduces the chance of visual occlusion while optimizing syntactic structures reduces cognitive load.
- [DG3] Text simplification in AR should give priority to text length over grammatical correctness Traditional text simplification techniques usually prioritize grammatical correctness [3, 93]. However, we find that priority should instead be given to text length and clarity in AR. This was gleaned from the open-ended exploration, where participants expressed the need to minimize occlusion caused by text length and indicated that less strict grammar did not notably affect their comprehension if meaning was preserved. Further expert interviews supported the assessment that AR users tend to

skim lengthy texts, not paying strict attention to grammatical correctness.

3.4.2 Updated simplification techniques (A1-4). Based on our findings, we update the four simplification techniques to fit users' needs in the AR context. We discuss the benefits and address discrepancies within the experts' feedback below.

A1: Content reduction We found that content reduction is beneficial in AR, as both the literature review and experts suggest. However, removed content may contain important task instructions, and its absence may alter the original meaning 3.4.1. Furthermore, as suggested by 3.4.1 and observation of the open-ended exploration, prepositions and pronouns can be cut for more concise.

A2: Syntactic simplification The results from the formative study support syntactic simplification as beneficial in AR contexts, given that complex grammatical structures can consume user attention. However, as with content reduction, syntactic simplification may alter the original meaning [93], necessitating adherence to 3.4.1. Furthermore, simplified grammatical structure can result in overall longer text, contradicting 3.4.1. To mitigate this, syntactic simplification should be applied only when it does not increase the number of lines of the displayed text, as addressed by E3.

A3: Lexical simplification Lexically simplified phrases may deviate from original meanings and lengthen the text, conflicting with 3.4.1 and 3.4.1. To address this, we propose two constraints for lexical simplification: Firstly, it should not alter task-related terms, and, secondly, it should not increase the number of lines of text.

A4: Elaborative simplification Elaborative simplification elicited nuanced opinions from the experts. In the NLP literature, elaborative simplification is described as benefiting second-language learners by elucidating abstract terms. However, as noted by E1-3, explaining terms may not benefit AR users and will likely lead to increased text length (contrary to 3.4.1). Therefore, common-sense explanations and explanations of background knowledge should be excluded from elaborative simplification to support concision. However, E1-3's feedback indicates that elaboration of the spatial context and numerical measures offers greater utility within the AR context. For instance, the user can benefit from spatial positional information such as the cup on your left when multiple cups are present. Additionally, when conveying numeric measures (e.g., seven inches), experts advised elaborating by referencing the size of objects already present within the scene, such as the diameter of a plate. By incorporating spatial context and numerical measure, elaborative simplification can be adapted within AR to adhere to 3.4.1 and enhance task performance.

4 ARTIST SYSTEM

In this section, we describe the design of ARTiST, which has been developed using the updated simplification techniques (Table 1) and design guidelines derived from the formative study.

ARTiST employs three novel methods, shown in Figure 2, to customize the GPT-3 model to stably output the desired simplification results. These include utilizing the chain-of-thought method to enhance GPT-3's reasoning capabilities and the plan-of-technique method for selecting the most appropriate techniques from **A1-4** (3.4.1 and 3.4.1). Additionally, ARTiST implements error-aware

calibration to ensure the preservation of the original text's meaning (3.4.1).

4.1 Plan-of-Technique Prompting

The plan-of-technique method is designed to structure the simplification process through a plan of different simplification techniques (A1-4). These techniques guide the GPT-3 model in executing the simplification as intended. This planning-and-execution model has been widely adopted in code generation [117], open-world agents [106], and robotics [95] for controllable and stable outputs. Figure 3 shows how input texts and the spatial context are fed into GPT-3 to generate the simplification plan.

Step-by-step execution ensures that all necessary simplification techniques can be applied. Our preliminary experiments reveal that GPT-3 sometimes forgets the techniques and design guidelines. One explanation for this is that LLMs like GPT-3 are typically trained on generic corpora without access to specialized design guidelines. Our plan-of-technique thus decomposes the simplification process into different simplification steps, mitigating forgetfulness. Moreover, such a structured pipeline can elicit the GPT-3's multi-hop reasoning capability shown in other NLP tasks [1, 90, 96].

In text simplification, multiple simplification techniques can sometimes conflict with each other and require multi-hop reasoning to resolve. For example, elaborative simplification (A4) may conflict with content reduction (A1). The plan-of-technique guides GPT-3 to consider the different techniques before executing the actual simplification actions, thereby reducing potential conflicts.

4.2 Chain-of-Thought Prompting

Chain-of-thought prompting is used to further enhance GPT-3's multi-hop reasoning capabilities and resolve potential technique conflicts. In few-shot prompting, a series of exemplars are created to instruct GPT-3 on how to generate the desired output based on the input text. Chain-of-thought augments the exemplars with intermediate reasoning steps, leading to the final output [107]. Drawing upon the proven efficacy of chain-of-thought's applications in diverse fields [52, 91, 104], we incorporate chain-of-thought into both the planning and execution phases of the plan-of-technique method. This decision aligns closely with 3.4.1 and 3.4.1, which stress the importance of adaptively applying traditional text simplification techniques (A1-4) to cater to AR-specific needs.

We use an example to show how the chain-of-thought method supports the plan generation in the plan-of-technique method. To simplify the sentence, *Grab a pair of 10 to 12 lb (4.5 to 5.4 kg) dumb-bells and lie on your back with your arms behind you and your legs extended and raised to a 45-degree angle,* we prompt GPT-3 to generate the thoughts about the input text's applicability to AR context. GPT-3 identifies the sentence as overly lengthy, containing more than three phrases, and thus includes syntactic simplification in its simplification plan. The plan involves three steps of syntactic simplification: (1) splitting the sentence at the first *and* because the length of the two joined clauses is too long; (2) splitting the sentence at the second *and* for the same reason. (3) Adjusting the passive voice in *your legs extended and raised* for better readability. After generating the plan, we continue to prompt GPT-3 to apply

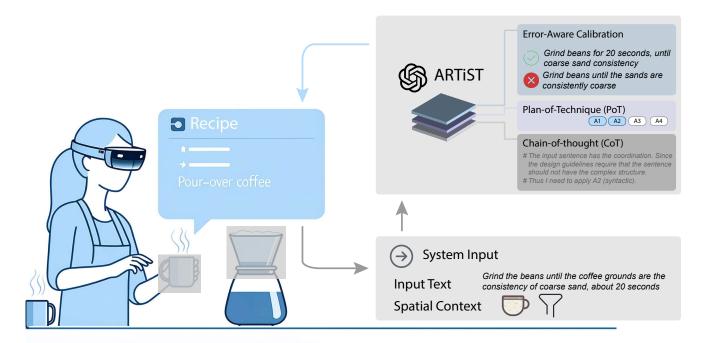


Figure 2: Method overview: ARTiST uses OpenAl's GPT-3 model, prompted with chain-of-thought and technique-as-plan methods, to generate simplified text candidates. The candidates are calibrated to reduce the likelihood of potential errors. The resulting simplified text is then displayed within a HoloLens 2 application. The spatial context is captured by detecting the objects in the scene to support the elaborative simplification.

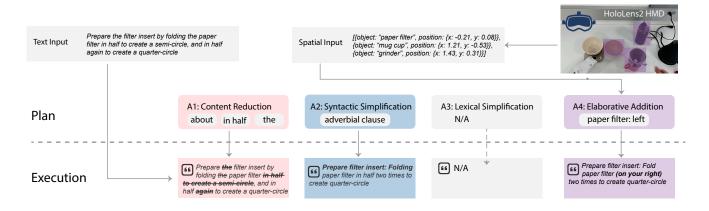


Figure 3: Plan-of-technique: The input text and the spatial context are fed into the LLM, which first generates a plan of the simplification techniques. The techniques will be sequentially applied to the input text to generate the final simplified text.

the simplification techniques outlined in the plan, yielding the result: Grab a pair of 10 to 12 lb (4.5 to 5.4 kg) dumbbells. Lie on your back with your arms behind you. Extend your legs and raise them to a 45-degree angle.

4.3 Error-Aware Model Calibration

To align with 3.4.1 and prioritize meaning preservation in text simplification [10, 94], we propose an error-aware calibration method. Outputs from LLMs are often unstable and exhibit a bias toward certain answers due to the intrinsic bias of the LLMs and the influence of the prompt text, especially when applied to new tasks. Text simplification in AR has requirements that differ significantly from

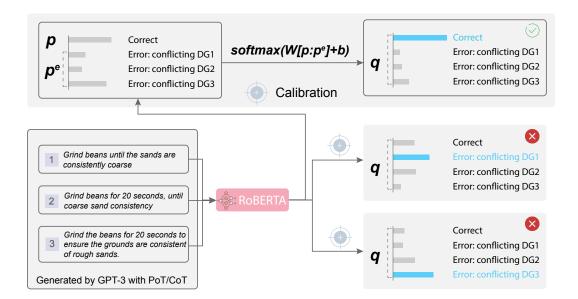


Figure 4: Error-aware model calibration: ARTiST prompts GPT-3 to generate a set of candidate results, which are subsequently analyzed by a RoBERTa-based error classification model (depicted in pink block) to detect any violations of design guidelines. The predicted scores of errors are calibrated with the affine matrix. Scores are adjusted using an affine matrix to ensure that the final selection is the output with the highest probability of correctness.

those of traditional NLP tasks, potentially exacerbating the impact of such intrinsic bias on LLM inference. For instance, LLM outputs may disproportionately reflect the influence of the last example in the prompt text, and within the context of our text simplification, the simplification techniques chosen may also be biased by the simplification techniques used in this last example [119]. To mitigate these issues, our error-aware calibration mechanism adjusts the output probabilities by applying an affine matrix [16], which is learned from a set of annotated datasets. This transformation does not directly rely on the prompt text and can alleviate LLM bias [92, 114, 115, 118]. Moreover, to strengthen LLM against common errors in AR, we enhance the annotated dataset with negative samples that violate 3.4.1 and risk altering the original meaning.

Shown in Figure 4, we use model calibration to stabilize language models in text generation [114, 115]. The affine transformation is defined as:

$$q = softmax(Wp + b), \tag{1}$$

where p refers to the probability of the generated simplified text, q is the calibrated probability, and W and b are learned parameters. We simplify this computation by treating W as a diagonal matrix following [37, 119]. The calibrated errors are identified from our open-ended exploration and expert interview. We then use the RoBERTA model to predict these errors by comparing the simplified text T^* and the original text T as, $p^e = p_1^e, p_2^e, ..., p_m^e = f(T, T^*)$, where p_i^e is the probability of an error and m is the total number of errors. The errors include altering the meaning 3.4.1, producing text that is syntactically complex 3.4.1 and or too long 3.4.1. Since access GPT-3's weights are not publicly accessible, we use RoBERTa instead to predict the error label p^e [58]. Therefore, we modify

Equation 1 by incorporating p^e ,

$$q = softmax(W[p; p^e] + b), \tag{2}$$

For each original text sample, we generate n=5 simplified text samples and calibrate them. The final output is determined by the calibrated probability. The parameter values of W and b are learned from a small set of manually crafted data samples[115]. We first craft a set of gold-standard text simplification samples (64) $D = \{(T_1, Y_1, \hat{q}_1), (T_2, Y_2, \hat{q}_2),$

 \cdots , (T_k, Y_k, \hat{q}_k) } where $(T, Y, \cap e)$ refers to the input text T, the simplified result Y, and whether the erroneous indicator \hat{q} . \hat{q} indicates whether Y is correctly simplified from T and, if not, labels the error in Y. Since W and b have limited dimensions, we can learn the values of W and b through the gradient descent with the logistic loss function $|q \cap e|^2$.

$$\mathcal{L} = -\hat{q}log(q) - (1 - \hat{q})log(1 - q)$$

$$= -\hat{q}log(softmax(Wp + b))$$

$$- (1 - \hat{q})log(1 - softmax(Wp + b)),$$
(3)

where \mathcal{L} is the loss function used to learn W and b.

4.4 Elaborative Simplification with Spatial Information

Following 3.4.1 and implementing elaborative simplification, we enrich the AR text by generating information on the spatial location of objects and object dimensions if they are presented in the original text. The object is detected and located with the Detic model, which runs on the backend server and provides the spatial information to LLM [124]. As shown in the expert interview, many objects may exist in the working environment, while only a subset of them

can be useful. To align with 3.4.1, we require LLMs to select the objects that are mentioned for the first time in the text. Identified object locations are used to signify a spatial relationship to the user, adding a layer of contextual understanding that goes beyond identification. For example, the text Then place the coffee mug with the dripper can be elaborated with the coffee mug's detected position on your right to form the result Then place the coffee mug on your right with the dripper. The spatial location is determined before the user clicks next step and the elaborated content does not change during the execution of the step to avoid distracting the user. The Detic model may incur prediction errors and mismatches in object location due to user movement and latency. For instance, the user's movement can alter the object's location relative to the user, and the Detic model's results may continue to indicate the location of the object before the user's movement. We mitigate this issue by predicting only the spatial relationships between the object and the user (e.g., the object is to the right of the user). Therefore, the minor errors and latency in the Detic model do not significantly impact the final result. Furthermore, in our open-ended exploration, we observed that during the step transition, users typically do not engage in significant movement, thereby reducing the likelihood of potential mismatches. When displayed text includes a numerical measurement and an object with comparable dimensions is identified in the AR environment, the system automatically substitutes the numerical value with a description of the detected object.

4.5 System Implementation Details

We implement ARTiST's functionality using OpenAI's GPT-3 APIs and build a text simplification server with Flask. We run the Detic model on the server and incorporate its object detection result into the prompt text for GPT-3. The interface is developed using the PTGCTL architecture [103], with the HoloLens 2 component implemented in Unity.

5 EVALUATION

The formative study elicited that text-based AR guidance often creates a high cognitive load and that following AR guidance can be challenging due to the HMD's small display, low readability, and user error. Although our system attempted to address these limitations by integrating text simplification into AR, the actual effects on users' cognitive load, performance, and sense of usability required further exploration. To better understand these effects, we conducted two empirical studies. The first (Study 1) focuses on the overall cognitive cost of our system and its effect on performance over unmodified text, and the second (Study 2) focuses on a comparison against other AR text simplification methods and what can be learned from them. Tasks in both studies are everyday tasks that could benefit from AR task guidance [78]. Both studies comprise within-subject designs. Although subtasks in Study 1 have a between-subject component, our primary focus and point of investigation is the text condition. We investigate the following research questions:

- (1) In what ways does our proposed method impact cognitive load in AR (3)?
- (2) In what ways does our proposed method affect task performance with text in AR (3)?

(3) How does our proposed method compare to other text simplification methods in AR (3)?

While the first study focuses on 3 and 3, the second explores 3. We pre-determined the study order so that half of the participants start with Study 1 and the other half with Study 2. Regardless of the order in which they engage the studies, participants are asked to review the study procedures and can only continue after giving their consent on the IRB-approved consent form.

5.1 Participants

Both studies involve 16 participants (average age 25, nine male and seven female). Half have previous experience using head-mounted AR and were recruited through electronic flyers and emails using snowball sampling.

5.2 Study 1: The Effect of Text Simplification on Guidance Tasks

We conducted an empirical study to evaluate the effect of textual simplification on users' cognitive load, performance, and other subjective ratings. We select two common physical tasks that benefit from AR guidance and collect data from real users. To avoid the learning effect while keeping task difficulty levels similar, both subtasks are physical activities that are performed in the same room (See Figures 5 and 6), have instructions of similar lengths, and do not require prior knowledge.

5.2.1 Experiment setup. We present the involved tasks and conditions in Study 1.

Task. The task contains two similar subtasks that have sequential instructions to guide users. In both subtasks, we display the AR text in a dark grey box to ensure visibility. We also adjust the font size to 9pt and have participants confirm that all text is legible. No single instruction is long enough to be cut off by the display. In terms of subtask assignment, we alternate the order of subtasks for each participant to balance the order effect.

- Task 1.1: Pour-over Coffee. This subtask contains nine step-bystep instructions that guide participants to make a pour-over coffee (Figure 5). The instructions are taken from an online tutorial on how to make pour-over coffee. ³ Participants need to read the text to complete the task.
- Task 1.2: Meeting room preparation. This subtask requires that participants follow AR instructions to arrange objects in a meeting room based on a seven-step office menu (Figure 6). The instructions are digitized from an online manual.

Conditions. This study has two conditions: a baseline condition that uses the original imported text and a simplified condition using ARTiST. Each participant will perform one subtask (either Task 1.1 or 1.2) with the baseline condition and the other subtask with the simplified condition. We use a pre-generated table to alternate the order of all trials so that each participant will perform tasks in different orders under both conditions. In total, all conditions and subtasks are evaluated an equal number of times. Samples from the simplified and baseline condition can be found in Table 2.

Apparatus. Participants wear a Microsoft HoloLens 2 and use hand gestures and voice commands to interact with the AR menu.

 $^{^3}$ https://www.wikihow.com/Make-Pour-Over-Coffee

Task	Step	Original	Simplified
Task 1.1	1	To create a coffee, first please carefully place	Place dripper (on your left) on coffee mug.
1ask 1.1		the pour-over dripper over the coffee mug.	
	7	Transfer the coffee grounds to the filter cone.	Move grounds to filter cone. Set coffee mug
		Then place the coffee mug with the dripper	with dripper on scale, zero it.
		on a digital scale and set it to zero.	
Task 1.2	2	Once the desk is clear, bring the power strip	Put power strip on desk, connect phone
		on the desk and connect the Charger to the	charger to it.
		power strip so the meeting attendants can	
		use.	
	5	Next, place cups of water and papers on	Place water, paper onto desk in front of
		each chair. Each person should have one	chairs.
		cup of water and paper;	

Table 2: Four example system outputs in Study 1. The original, unmodified text (baseline) is in the third column; the last column shows the simplified condition with text output from ARTiST showing on the last column.









Figure 5: Task 1.1: Sample frames from user recording. The task requires participants to make pour-over coffee based on a nine-step online tutorial. The frames were sampled from steps 2, 3, 5, and 8.









Figure 6: Task 1.2: Sample frames from user recording. The task requires the participants to arrange objects in a meeting room based on a seven-step office menu. The frames were sampled from steps 3, 4, and 6.

These interactions are native to HoloLens 2, and the AR interactions comprise standard button tapping, translating, and spatial movement. Video and audio recording devices are set up to collect participants' feedback and qualitative data.

5.2.2 Procedure. The experimenters welcome the participants in a physical room; the physical tools necessary for task performance (e.g., coffee machine and ingredients) are present. To maintain ethical standards and comply with the IRB guidelines, each participant is given an informed consent form before the evaluation. Upon signing, each participant is paid an hourly rate of \$20 and is fitted with the Microsoft HoloLens 2 headset. An ill-fitting HoloLens 2 can be detrimental to the AR experience, causing blurry text. A series of initial calibrations are performed to ensure interface functionalities 1.

After all participants successfully interact with the AR interface, including its menus and buttons, using hand gestures, and indicate they can see the AR text clearly on the HMD, experimenters then

explain the two subtasks and ask participants to practice thinking aloud. Meanwhile, video and audio recordings were set up before the trial began. Participants begin the study by air-tapping the AR button marked *Start* at the center of the HMD's screen.

Once the task starts, step-by-step text instructions are automatically displayed in AR. Participants are not informed which condition they are using and are asked to think aloud while we observe and record the trials. Any anomalies or potential safety issues are continuously monitored by the experimenter. At the end of each subtask, we collect the participant's subjective ratings on text readability, comprehensibility, guidance performance, trust, and cognitive load using a NASA TLX form. A semi-structured interview is conducted to better understand their experience.

5.2.3 Data collection. We collect quantitative data to measure performance. We specifically record the number of errors and the number of steps participants recall (i.e., memorability); in addition, we explore self-evaluated performance via subjective ratings. The

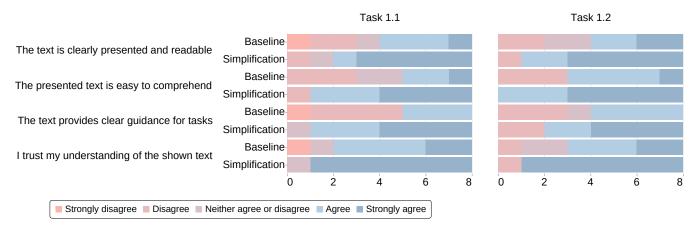


Figure 7: Study 1 results on a five-point Likert scale. Ratings are collected on a scale from "strongly disagree" to "strongly agree" in response to four questions assessing the readability, ease of comprehension, guidance, and trust in both simplified and baseline text versions. The horizontal bar graphs above visually represent the distribution of these ratings. The distribution reveals more positive responses for the simplified text across all questions and tasks.

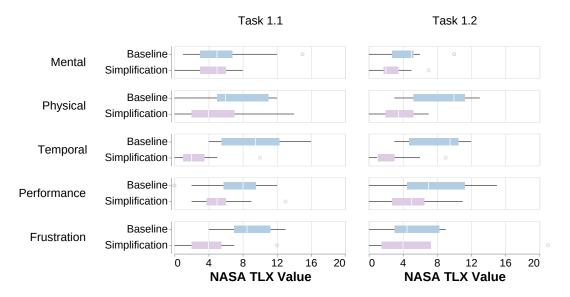


Figure 8: Study 1 results on NASA Task Load Index (TLX) values. The y-axis represents the different aspects of the NASA TLX, while the x-axis shows the TLX values. The simplified text significantly surpasses the baseline in reducing temporal demands and in enhancing performance and reducing frustration, demonstrating its advantages for overall task load.

experimenter counts the number of errors during participants' trials. Memorability is measured because one major challenge in AR guidance is that users only recall limited AR information during physical tasks; remembering steps reduces the need to split attention between AR and the task. Subjective ratings are inspired by the System Usability Scale (SUS), and we collect five-point Likert ratings on AR text readability, comprehensibility, guidance, and trust. We explain that trust reflects how confident the user is with their task performance.

Cognitive load is a primary user performance limitation in AR guidance tasks, and we use a NASA TLX 8(a)(b) form to measure

it. Raw TLX scores are used and summative results are analyzed based on Hart's recommendations [38].

Experimenters also collect qualitative data via video and audio recordings of the study. Interview notes, think-aloud notes, and observations are also collected for later analysis. The sampled frames for the study can be found in Figures 5 and 6.

5.3 Study 1: Results and Discussion

5.3.1 Quantitative results. Using Mann-Whitney's U test, we assess differences among TLX scores, recall, and error data and use the Friedman test to assess differences among subjective ratings. These tests were chosen because the data are non-parametric. The

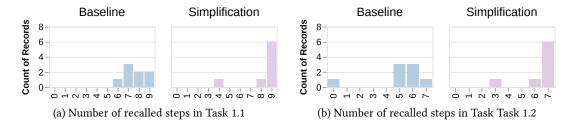


Figure 9: Study 1 results on the number of errors participants made while performing Tasks 1.1 and 1.2. The x-axis indicates the number of steps successfully recalled, while the y-axis shows the count of participants who recall that number of steps.

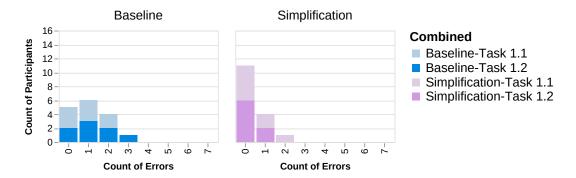


Figure 10: Study 1 results on the number of errors made in Tasks 1.1 and 1.2. The bar graph compares the error count between the baseline and simplified conditions, with the x-axis recording the number of errors and the y-axis depicting the number of participants who make those errors. The data illustrates that participants commit fewer errors when following the simplified text.

TLX analysis shows that the simplified condition significantly reduces the overall cognitive load for both subtasks ($U=52,\ z=2.84,\ p<0.01$). Further evaluation of recall and error found no significant improvement in recall for the simplified condition over the baseline ($U=85,\ z=1.63,\ p=0.10$), but showed the simplified condition significantly reduced the number of errors counted by users over the baseline ($U=72.5,\ z=-2.09,\ p=0.024$), see Figure 9. Test on subjective ratings indicated significant differences in all categories: readability ($\chi^2=10.71,\ p=0.013$), comprehensibility ($\chi^2=15.00,\ p=0.002$), guide ($\chi^2=10.71,\ p=0.013$), and trust ($\chi^2=18.00,\ p=0.001$).

5.3.2 Qualitative results. We coded the transcribed video and audio data along with notes from thinking aloud and observation. Codes sharing similarities were then grouped into themes to summarize analogous findings.

Spatial information can assist users. Participants acknowledged the benefits of spatial information in reducing cognitive load. Elaboration on objects' spatial location eliminates the need to search for them, reducing user effort. P11 reported feeling nervous when presented with multiple objects and new mentions of objects in the AR interface. P11 mentioned, "Sometimes it is overwhelming to face many objects, and the location word (on your left) helps (you) find the object." The reduced effort and pressure were also confirmed by P1, who stated, "Even though it won't save much time, the elaboration on the object eases my (sense of) pressure."

Text length and structural complexity affect participants' performance. Most participants report that shorter text is beneficial. Some participants reported that shorter text takes less time to process (P2, P5, P7) and felt it was "easier to understand" each step during a subtask when the text was shorter (P6-7). This is reflected in the TLX scores, as simplified conditions yielded better cognitive load scores than the original texts. Participants further report that using shorter sentences leads to better comprehension and confidence (P2, P10-12, P14). Multiple participants pointed out that they naturally "skim" text in AR, and stated that complex sentence structures lead to skipping important information and misunderstandings. More than half of the participants further stated that the simplified text improved their trust. When asked to explain their reasons for skimming text, screen resolution, screen size, and the urgency of completing physical tasks (impatience) while wearing a headset were identified. These observations reflect what experts from the formative study indicate.

Simplified text improves task guidance. Participants respond positively to breaking longer sentences into shorter ones (syntactic simplification). ARTiST divides long sentences into shorter ones by adding verbs (elaborative simplification). P5 said "Shorter sentences with clear actionable directions make it easy to know what to do," while P6 said, "It is more convenient to follow smaller step instructions." The participants' positive feedback reflects the benefits introduced by the design guidelines proposed earlier.

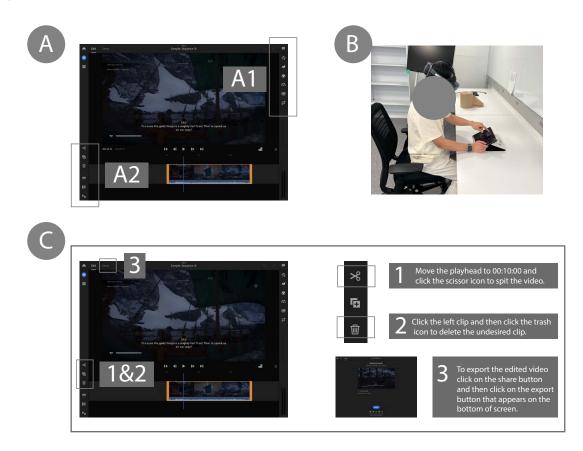


Figure 11: (A) Adobe Premiere Rush Interface: The interface showcases a video player positioned at the center of the screen, accompanied by a timeline below. The top-right section (A1) features buttons for graphics, effects, color, speed, audio, and transform functionalities. The bottom-left section (A2) contains buttons for editing tools and the project panel. (B) User Record: This section captures the user's interactions and activities during the Study 2 session. (C) Example Task Description: An illustration of a sample task description used in the study, providing users with instructions for completing a specific editing task.

5.3.3 Discussion. Our results verify that ARTiST significantly improves cognitive load (3) and reduces task performance errors, generating significantly higher subjective ratings (3). Shortened sentences and syntactic simplification contributed to decreased cognitive load, as participants indicate that the simplified text is easier to read. Additionally, shorter sentences enable participants to quickly skim the text to grasp core concepts, which may also play a part in reducing cognitive load. As we mentioned earlier, reducing cognitive load could help to improve the usability of AR guidance and have a positive effect on users' safety.

ARTiST significantly improves subjective ratings on all four metrics. However, the system yields no significant change in memorability. Both the baseline and simplified conditions reached a fairly high recall count. A possible explanation for this is the fact that all tasks are physical tasks, and participants may rely on their performance more than the text for recall. Yet the simplified text resulted in fewer user errors than the unmodified text, suggesting that the system successfully retains critical information for tasks. Overall, participants felt better guided by the simplified instructions

and more confident (i.e., trust), signaling a positive effect on their overall performances.

5.4 Study 2: Comparing Text Simplification Methods

In the previous study, we evaluated ARTiST against unmodified AR text. The goal of this study is to further understand how ARTiST's process compares with other methods for text simplification. However, almost all currently used methods are not tailored for AR. As such, we selectively integrated these methods into the AR context while keeping their traditional functionalities. This study is a within-subject study that includes five different methods with one task. The study recruited the same set of participants (N=16). In addition to the HoloLens 2 used in Study 1, this study makes use of an iPad as an additional apparatus for task performance.

5.4.1 Experiment setup. We present the tasks and conditions of Study 2 in this section.

Task. The task is designed to have participants wear a HoloLens 2 while also using an iPad. AR instructions for video editing are

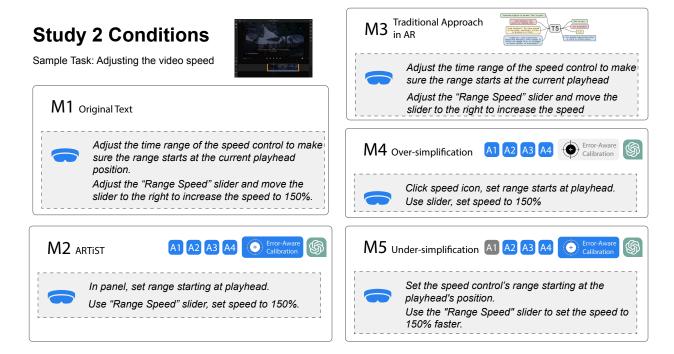


Figure 12: All five conditions for Study 2. M1 is the original text. M2 is the ARTIST condition. M3 is the state-of-the-art T-5 model applied to AR. M4 is ARTIST without engaging error-aware calibration (over-simplification). M5 is ARTIST without engaging content reduction(under-simplification). The text in the method's grey box represents the text after simplification. A1, A2, A3, A4, and error-aware calibration legends denote whether any of these components are used for the condition.

displayed on the HoloLens 2. To minimize the learning effect and for repeated trials, we employ subtasks with similar interactions and difficulty levels but different content. We adopt a series of video editing jobs from Adobe Premiere Rush's official tutorial to AR⁴ to test each method. They involve interaction primitives such as selection, pan, and translation. The task contains five subtasks, including video clipping (S1), speed control (S2), graphics overlay (S3), video filter application (S4), and aspect ratio adjustment (S5). These subtasks are chosen because they have similar interaction difficulty but require diverse types of interactions (e.g., tap, drag, and pinch). Each subtask has three steps and takes about three minutes to complete based on our preliminary testing. Adobe Premiere is installed on the iPad participants use to perform the video editing. An example can be found in Figure 11.

Conditions. To understand how our system differs from other simplification methods and investigate the implications for user performance, we explore five conditions. Beyond making a comparison to the unmodified text (i.e., baseline), we also compare against the state-of-the-art T-5 text simplification model [80]. Further, because our formative study revealed that sentence complexity and grammar structure (text length) have a foremost effect on text reading in AR (which was also indicated by experts in the formative study), we also explore an over-simplification and an under-simplification condition in this study. These two simplification methods represent different levels of length modification relative to the original

sentence. Over-simplification is achieved by removing the erroraware calibration step for maximum simplification at the cost of factual information. Under-simplification is achieved by removing the content reduction technique in ARTiST. We describe the five different conditions below:

- M1: Original text
- M2: ARTiST's approach;
- M3: Traditional state-of-the-art text simplification with T-5 [80] fine-tuned on WikiAuto dataset [47]
- M4: Over-simplification without error correction (i.e., does not force factuality, see 3.4 for details)
- M5: Under-simplification without content reduction (A1)

We use a pre-generated table to balance the learning and ordering effect for the five conditions across the five subtasks. We assign one condition to one of the subtasks to form pairs, and each pair includes three steps (i.e., three trials). For example, the pair M2-S1 stands for the M2 condition used in subtask 1. The pre-generated table ensures that each participant performs these pairs in a unique order. Each participant performs five condition-subtask pairs or 15 trials, for a total of 240 trials. Overall, all subtasks and conditions are evaluated an equal number of times.

5.4.2 Procedure. In the beginning of this study, participants are asked to wear the HoloLens while sitting and holding an iPad. Experimenters explain that their task entails editing videos that are displayed on the iPad. The videos are 30-second clips of stock footage, and participants are told they will use the onboard video

 $^{^4} A dobe \ Premiere \ Rush. \ https://helpx.adobe.com/premiere-rush/tutorials.html$

editing tool on the iPad to seek, crop, change filter, and change the aspect ratio. After a short warm-up period to familiarize participants with iPad functionality and fit the HoloLens, we confirmed that participants can read the AR text clearly, similar to what we did in Study 1. During task performance, condition-specific AR text is displayed to the participants; they are asked to follow the text to perform the editing task. Experimenters count the number of errors made during the trials, and participants are asked to think aloud as they engage in their tasks. At the end of each condition, the experimenter collects recall data, subjective ratings, and TLX scores. A semi-structured interview is conducted at the end of the study to understand the participants' overall impression of each of the five conditions. We collect both quantitative and qualitative data in a similar way to Study 1.

5.5 Study 2: Results and Discussion

5.5.1 Quantitative results. For non-parametric data, we used the independent-sample Kruskal-Wallis' test with repeated measures for performance metrics (error, memory recall, and subjective rating) and for cognitive load with Dunn's Test as post-hoc analysis with Bonferroni correction for multiple tests. For error analysis, we found an overall significant effect (H(4) = 17.189, p = 0.001), with post-hoc analysis showing that M2 (p = 0.014), M3 (p = 0.014), M4 (p = 0.014), and M5 (p = 0.014) reduced errors significantly compared to the baseline. We found that there is an overall difference across the conditions (H(4) = 12.572, p = 0.014) in terms of participants' ability to recall the instruction steps. Post-hoc analysis revealed a significant difference between the original text M1 and ARTiST (p = 0.025). No differences are found between the original text M1 and M3 (p = 0.179), M4 (p = 0.319), and M5 (p > 1.00). No differences are found between the four simplified conditions (M1-4). As for TLX scores, we found that M2 significantly reduced overall cognitive load compared to M1 (p = 0.043): for a detailed breakdown refer to Figure 14. There are no significant differences among the five conditions in readability (H(4) = 0.934, p = 0.934), comprehensibility (H(4) = 0.389, p = 0.983), guidance (H(4) =2.444, p = 0.655) and trust (H(4) = 1.530, p = 0.821); See Figure 13 for details.

5.5.2 Qualitative results. We identify a series of qualitative findings based on the quantitative metrics and the coded recordings.

The level of simplification has a mixed effect on error rates and subjective ratings. While participants reported that they could understand any of the simplified texts (M2-5) better than the unmodified text (M1), we noticed that there is no uniform effect on the level of simplification relative to task errors (P11-12, P3, P5). While several participants experienced increased errors due to over-simplified text omitting important information (P3), others made mistakes due to verbose text that was not simplified enough (P11-12, P5), causing them to overlook important information. This also aligns with the fact that the M1 condition yielded the most user errors. One of the behaviors observed is that the longer the sentences, the less patience a participant appears to have and the faster they skim. Often these behaviors lead to missing details while carrying out the task, such as when P11 and P12 adjusted the wrong button during task performance. Similarly, both over- and

under-simplification methods affected participants' sense of readability and memorability in different ways. When asked about their experience with the M5 condition, P6 reported that "the simplified one uses the more understandable words," but P10 mentioned that the texts are "not simplified enough and can be thrown away." P7 also reported that M5 increased the number of previous steps they could recall, as M5 presents "clear and memorable instructions."

The effect of different text simplification methods on cognitive load. Participants reported during the interview that simplified texts (M2-5) have in lower cognitive load. While reading unmodified text became tedious during task guidance (P2, P9-10, P14), the simplified text could be less so (P2). However, participants reported that over-simplified text (M4) increases cognitive load, as important information is often removed resulting in extra processing time needed (P1, P10). (P1, P10). Moreover, participants indicate that they believe they perform better with the ARTiST condition: As P7 mentioned, "I am pretty sure I successfully completed the steps," while P11 said, "I feel it increased my performance."

5.5.3 Discussion. In exploring 3, we found ARTiST impacted TLX ratings as it was the only condition that significantly reduced cognitive load for participants. Figure 14 shows that TLX variance is much lower for performance with ARTiST, which is in line with our observation that most participants show stable performance with the ARTiST condition.

All four simplified text conditions (M2-5) significantly reduce error rate, but do not necessarily increase recall. This finding reflects our 3.4.1, which addresses the importance of text length in AR. Regardless of the level of simplification, all four conditions shortened the text in some way. The results indicate that only ARTiST significantly improved recall while reducing error rate. This indicates that ARTiST helps users to improve performance in short-span tasks like video editing. In addition, the current state-of-the-art text simplification (M3) does not reduce high cognitive load nor improve memory for AR readers, while ARTiST improved on both. This suggests that the direct application of text simplification to AR might not be optimal and is in line with the results from the formative study.

ARTiST is also the only condition that significantly reduced the TLX scores (3). Sentence length and structure may play an important part in reduced cognitive load as participants noted reduced processing time and more ready comprehension. This reduction addresses the concerns (high cognitive load) brought up by experts during our formative study in Sec. 3.3.

Finally, both over- and under-simplification conditions received mixed feedback from participants. This could be linked to their personal reading habits when wearing an HMD. We observed that participants who comment positively on the over-simplified condition are typically impatient readers when they have the HoloLens 2 on. Others, however, complained that the over-simplified condition does not provide enough detail or is missing critical information, creating obstacles to task completion. Our qualitative results showed that participants who took extra effort going after missing details scored higher in their TLX ratings. These findings reflect the results from the formative study that both text length and meaning preservation are important.

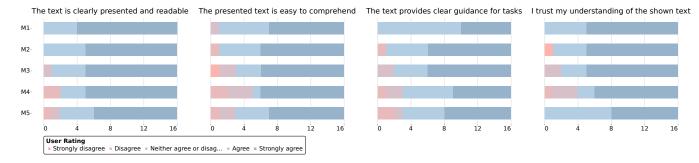


Figure 13: Study 2 results on the subjective Likert scale. Ratings were collected on a scale from "strongly disagree" to "strongly agree" in response to four questions assessing the readability, ease of comprehension, guidance, and trust in both simplified and baseline text versions. The horizontal bar graphs represent the distribution of these ratings, and the results for the four questions are laid out horizontally.

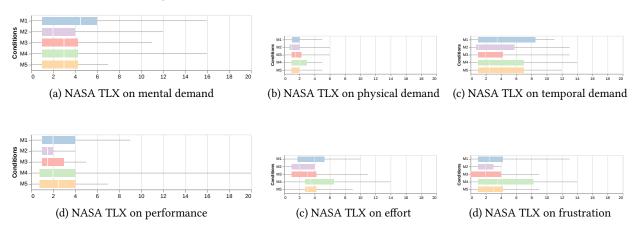


Figure 14: Study 2 results on NASA Task Load Index (TLX) values. The y-axis represents the different aspects of the NASA TLX, while the x-axis shows the TLX values. The results indicate that condition M2 significantly outperforms other conditions in terms of the user's effort.

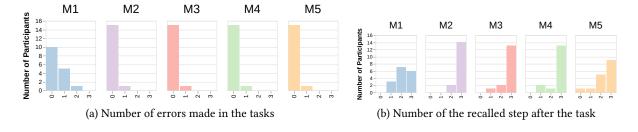


Figure 15: Study 2 results for error count and memorability evaluation. In panel (a), the x-axis represents the number of errors made by participants, while the y-axis shows the count of participants corresponding to each error count. The results indicate that methods M2, M3, M4, and M5 are effective at reducing the number of errors, underscoring the advantages of text simplification in enhancing task success. In panel (b), the x-axis displays the number of task steps correctly recalled after the task, and the y-axis shows the number of participants. This panel demonstrates the impact of the simplification process on the participants' abilities to recall information, with conditions M2, M3, and M4 showing an improved number of steps recalled over M1.

6 DISCUSSION

6.1 The Scope of ARTiST

While our evaluation showed improved performance and reduced cognitive load for participants using ARTiST, the scope in which our system works could be affected by task types (stationary vs mobile), task difficulty, and users' expertise. ARTiST shows promising results for relatively stationary tasks, such as making coffee or preparing meeting rooms, where users can pause between subtasks to read instructions. As a result, we envision that tasks related to medical diagnostics and work productivity can benefit from text simplification. Text-based guidance may perform poorly for tasks where users are in constant motion (e.g., riding a bike or running outdoors), as reading could divert attention and pose safety risks.

Task difficulty could further affect the applicability of text-based instruction. The pure text might not be the best medium to guide highly complex tasks such as surgery or repairing mechanical devices, which may benefit from multimedia guidance comprised of visual or visual-auditory media alongside text. However, the concepts of textual simplification may extend to multimedia guidance and future work could be done to explore the simplification of additional variables (e.g., layout, typography, and visual elements) in the context of more complicated AR tasks.

Finally, the user's own level of expertise also affects the experience of using ARTiST. Experienced users tend to skip instructions and may desire very minimal guidance is a known behavior. A beginner, on the other hand, might benefit from more elaborative text guidance. This creates an opportunity to personalize text simplification for users with different levels of expertise and different guidance preferences.

6.2 Implications of AR Text Simplification

Spatial information benefits AR users. A command like *the cup on your right*, facilitates object identification. Participants give positive feedback about spatial elaboration as this method helps to contextualize content in the original text. Future work could explore how we can further contextualize text via AR animation or AR sound.

Object detection error rarely impairs task performance in AR context We use object detection (i.e., Detic [124]) to gather the spatial information. We acknowledge that the Detic model may not always be stable, and its errors could potentially impair task performance. In our study, we did not observe any significant issues with the Detic model. There are two possible failures: false positives and false negatives. We observed no false positives but some false negatives, where objects were not detected by Detic. Interviewing users who encountered false negatives found that these false negatives were unnoticeable by participants because text was not always simplified; this did not lead to users generating more errors. Although no false positives were encountered, we hypothesized situations where incorrect object locations might be presented. Participants responded that such issues would not significantly impair their performance, as they felt they would identify the error when the object was not found in the indicated location.

Integrating empirical findings from AR into LLMs may open up new opportunities for human-AI collaboration.

Emerging human-AI collaboration tasks often require knowledge about users' needs. This knowledge often remains undeveloped until empirical studies are conducted [89, 105, 110, 121]. The traditional hurdle in applying the resulting knowledge to human-AI collaboration lacks the annotated data [55, 109]. LLMs present a potential solution by using zero-shot or few-shot learning for rapid prototyping with newly discovered knowledge [17, 79]. Our work exemplifies this paradigm by integrating design guidelines into LLMs for AR-specific text simplification.

6.3 Limitations and Future Work

Our system uses OpenAI's GPT-3 API, which has a latency of approximately 2 seconds. This latency can result in object positioning inaccuracies if the user moves or turns around. It can also pose challenges in time-sensitive scenarios, such as emergency medical support (E1). Future work could look into device-side post-processing for faster performance. In terms of AR devices, we primarily tested the HMD format, thus it is not yet clear if our system will work for other AR devices [39, 74].

7 CONCLUSION

In conclusion, this paper presents ARTiST, an automated text simplification system tailored for head-mounted AR devices. We first identify the challenges in AR text presentation via a formative study that includes a survey of the literature, an open-ended exploration with seven participants, and interviews with three experts. The findings lead to design guidelines that help form the ARTiST system. The system leverages OpenAI's GPT-3 models through few-shot learning for automated text simplification. Using chain-of-thought prompting, we present two novel techniques tailored for AR text simplification: a plan-of-technique and error-aware calibration to ensure meaning preservation. We validate our system via a 16participant empirical study, resulting in significant improvements in users' performance, reduced cognitive load, and better subjective ratings when compared to unmodified text, the state-of-the-art T-5 language model, and other methods. These findings underscore the efficacy of our system in enhancing text readability and mitigating cognitive load during task guidance in AR environments.

ACKNOWLEDGMENTS

This work was supported by the DARPA PTG program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. We thank Xiaoan Liu, Jeff Huang, James Tompkin, Leslie Welch, and Ashley Girty for their intellectual contribution and insightful feedback.

REFERENCES

- [1] Leonard Adolphs, Kurt Shuster, Jack Urbanek, Arthur Szlam, and Jason Weston. 2022. Reason first, then respond: Modular Generation for Knowledge-infused Dialogue. In Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 7112–7132. https://aclanthology.org/2022.findings-emnlp.527
- [2] Murat Akçayır and Gökçe Akçayır. 2017. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational research review 20 (2017), 1–11.
- [3] Suha S Al-Thanyyan and Aqil M Azmi. 2021. Automated text simplification: a survey. ACM Computing Surveys (CSUR) 54, 2 (2021), 1–36.

- [4] Oliver Alonzo. 2022. The use of automatic text simplification to provide reading assistance to deaf and hard-of-hearing individuals in computing fields. ACM SIGACCESS Accessibility and Computing 132, Article 3 (mar 2022), 1 pages. https://doi.org/10.1145/3523265.3523268
- [5] Oliver Alonzo, Sooyeon Lee, Mounica Maddela, Wei Xu, and Matt Huenerfauth. 2022. A Dataset of Word-Complexity Judgements from Deaf and Hard-of-Hearing Adults for Text Simplification. In Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Virtual), 119–124.
- [6] Oliver Alonzo, Matthew Seita, Abraham Glasser, and Matt Huenerfauth. 2020. Automatic text simplification tools for deaf and hard of hearing adults: Benefits of lexical simplification and providing users with autonomy. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–13.
- [7] Oliver Alonzo, Jessica Trussell, Matthew Watkins, Sooyeon Lee, and Matt Huenerfauth. 2022. Methods for evaluating the fluency of automatically simplified texts with deaf and hard-of-hearing adults at various literacy levels. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–10.
- [8] KM Augestad, H Han, J Paige, T Ponsky, CM Schlachta, B Dunkin, and J Mellinger. 2017. Educational implications for surgical telementoring: a current review with recommendations for future practice, policy, and research. *Surgical endoscopy* 31 (2017), 3836–3846.
- [9] Ronald T Azuma. 1997. A survey of augmented reality. Presence: teleoperators & virtual environments 6, 4 (1997), 355–385.
- [10] Beata Beigman Klebanov, Kevin Knight, and Daniel Marcu. 2004. Text simplification for information-seeking applications. In On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 735-747.
- [11] Christoph Bichlmeier, Sandro Michael Heining, Mohammad Rustaee, and Nassir Navab. 2007. Laparoscopic virtual mirror for understanding vessel structure evaluation study by twelve surgeons. In 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE, IEEE, 125–128.
- [12] Mark Billinghurst, Jerry Bowskill, Nick Dyer, and Jason Morphett. 1998. An evaluation of wearable information spaces. In Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180). IEEE, IEEE, 20–27.
- [13] Mark Billinghurst and Andreas Duenser. 2012. Augmented reality in the classroom. Computer 45, 7 (2012), 56–63.
- [14] Matt Bower, Cathie Howe, Nerida McCredie, Austin Robinson, and David Grover. 2014. Augmented Reality in education–cases, places and potentials. *Educational Media International* 51, 1 (2014), 1–15.
- [15] David E. Breen, Ross T. Whitaker, Eric Rose, and Mihran Tuceryan. 1996. Interactive Occlusion and Automatic Object Placement for Augmented Reality. Computer Graphics Forum 15, 3 (1996), 11–22. https://doi.org/10.1111/1467-8659.1530011 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.1530011
- [16] Glenn W Brier et al. 1950. Verification of forecasts expressed in terms of probability. Monthly weather review 78, 1 (1950), 1–3.
- [17] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems 33 (2020), 1877–1901.
- [18] Josef Buchner, Katja Buntins, and Michael Kerres. 2022. The impact of augmented reality on cognitive load and performance: A systematic review. J. Comput. Assist. Learn. 38, 1 (2022), 285–303. https://doi.org/10.1111/JCAL.12617
- [19] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating smartphone-based pan and zoom in 3D data spaces in augmented reality. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages.
- [20] Yvonne Canning, John Tait, Jackie Archibald, and Ros Crawley. 2000. Cohesive generation of syntactically simplified newspaper text. In *Text, Speech and Dialogue*. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 145–150.
- [21] Julie Carmigniani and Borko Furht. 2011. Augmented reality: an overview. Springer, New York, NY, 3–46.
- [22] John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of English newspaper text to assist aphasic readers. In Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology. Association for the Advancement of Artificial Intelligence, 7–10.
- [23] John Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999. Simplifying Text for Language-Impaired Readers. In Ninth Conference of the European Chapter of the Association for Computational Linguistics, Henry S. Thompson and Alex Lascarides (Eds.). Association for Computational Linguistics, Bergen, Norway, 269–270. https://aclanthology.org/E99-1042
- [24] Sonia Castelo, Joao Rulff, Erin McGowan, Bea Steers, Guande Wu, Shaoyu Chen, Iran Roman, Roque Lopez, Ethan Brewer, Chen Zhao, Jing Qian, Kyunghyun

- Cho, He He, Qi Sun, Huy Vo, Juan Bello, Michael Krone, and Claudio Silva. 2024. ARGUS: Visualization of Al-Assisted Task Guidance in AR. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 1313–1323. https://doi.org/10.1109/TVCG.2023.3327396
- [25] Raman Chandrasekar, Christine Doran, and Srinivas Bangalore. 1996. Motivations and methods for text simplification. In COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics.
- [26] Antonina Dattolo and Flaminia L Luccio. 2017. Accessible and usable websites and mobile applications for people with autism spectrum disorders: a comparative study. EAI Endorsed Transactions on Ambient Systems 4, 13 (2017).
- [27] Siobhan Devlin. 1998. The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic databases* (1998).
- [28] Matt Dunleavy and Chris Dede. 2014. Augmented reality teaching and learning. Handbook of research on educational communications and technology (2014), 735–745.
- [29] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98.
- [30] Richard Evans, Constantin Orăsan, and Iustin Dornescu. 2014. An evaluation of syntactic simplification rules for people with autism. In Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR), Sandra Williams, Advaith Siddharthan, and Ani Nenkova (Eds.). Association for Computational Linguistics, Gothenburg, Sweden, 131–140. https://doi.org/10.3115/v1/W14-1215
- [31] Catarina G Fidalgo, Yukang Yan, Hyunsung Cho, Maurício Sousa, David Lindl-bauer, and Joaquim Jorge. 2023. A Survey on Remote Assistance and Training in Mixed Reality Environments. IEEE Transactions on Visualization and Computer Graphics 29, 5 (2023), 2291–2303.
- [32] Michele Fiorentino, Antonio E. Uva, Michele Gattullo, Saverio Debernardis, and Giuseppe Monno. 2014. Augmented reality on large screen for interactive maintenance instructions. *Comput. Ind.* 65, 2 (2014), 270–278. https://doi.org/ 10.1016/I.COMPIND.2013.11.004
- [33] Joseph L Gabbard, J Edward Swan, and Deborah Hix. 2006. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence* 15, 1 (2006), 16–32.
- [34] Joseph L Gabbard, J Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. 2007. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In 2007 IEEE Virtual Reality Conference. IEEE, IEEE, 35–42.
- [35] Núria Gala and Johannes Ziegler. 2016. Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. In Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC). The COLING 2016 Organizing Committee, Osaka, Japan, 59–66.
- [36] Dmitriy Genzel and Eugene Charniak. 2002. Entropy Rate Constancy in Text. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Pierre Isabelle, Eugene Charniak, and Dekang Lin (Eds.). Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 199–206. https://doi.org/10.3115/1073083.1073117
- [37] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On Calibration of Modern Neural Networks. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017 (Proceedings of Machine Learning Research, Vol. 70), Doina Precup and Yee Whye Teh (Eds.). PMLR, 1321–1330. http://proceedings.mlr.press/v70/guo17a.html
- [38] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
- [39] Jeremy Hartmann, Yen-Ting Yeh, and Daniel Vogel. 2020. AAR: Augmenting a Wearable Augmented Reality Display with an Actuated Head-Mounted Projector. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '20). Association for Computing Machinery, New York, NY, USA, 445–458. https://doi.org/10.1145/3379337. 3415849
- [40] Steven J Henderson and Steven Feiner. 2009. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In 2009 8th IEEE International Symposium on Mixed and Augmented Reality. IEEE, IEEE, 135–144.
- [41] Firas Hmida, Mokhtar B. Billami, Thomas François, and Núria Gala. 2018. Assisted Lexical Simplification for French Native Children with Reading Difficulties. In Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA), Arne Jönsson, Evelina Rennes, Horacio Saggion, Sanja Stajner, and Victoria Yaneva (Eds.). Association for Computational Linguistics, Tilburg, the Netherlands, 21–28. https://doi.org/10.18653/v1/W18-7004
- [42] Richard L Holloway. 1997. Registration error analysis for augmented reality. Presence: Teleoperators & Virtual Environments 6, 4 (1997), 413–432.
- [43] Aleksander Holynski and Johannes Kopf. 2018. Fast depth densification for occlusion-aware augmented reality. ACM Transactions on Graphics (ToG) 37, 6 (2018), 1–11.

- [44] Michael Hornacek, Hans Küffner-McCauley, Majesa Trimmel, Patrick Rupprecht, and Sebastian Schlund. 2022. A spatial AR system for wide-area axis-aligned metric augmentation of planar scenes. CIRP Journal of Manufacturing Science and Technology 37 (2022), 219–226.
- [45] María-Blanca Ibáñez and Carlos Delgado-Kloos. 2018. Augmented reality for STEM learning: A systematic review. Computers & Education 123 (2018), 109– 123.
- [46] Tatyana Ivanova Ivanova. 2017. Ontology-Based Text Simplification for Dyslexics. Science and Technology 3, 10 (2017), 34–47.
- [47] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Model for Sentence Alignment in Text Simplification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, Online, 7943–7960. https://doi.org/10.18653/v1/2020.acl-main.709
- [48] Matthew Kay, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. When (ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, May 7-12, 2016, Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade (Eds.). ACM, 5092-5103. https://doi.org/10.1145/2858036.2858558
- [49] Julian Keil, Annika Korte, Anna Ratmer, Dennis Edler, and Frank Dickmann. 2020. Augmented reality (AR) and spatial cognition: effects of holographic grids on distance estimation and location memory in a 3D indoor scenario. PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88, 2 (2020), 165-172.
- [50] Frank Keller. 2004. The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Barcelona, Spain, 317–324. https://aclanthology.org/W04-3241
- [51] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F Welch. 2018. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, IEEE, 105–114.
- [52] Seungone Kim, Se June Joo, Yul Jang, Hyungjoo Chae, and Jinyoung Yeo. 2023. CoTEVer: Chain of Thought Prompting Annotation Toolkit for Explanation Verification. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. EACL 2023 - System Demonstrations, Dubrovnik, Croatia, May 2-4, 2023, Danilo Croce and Luca Soldaini (Eds.). Association for Computational Linguistics, 195–208. https://aclanthology.org/2023.eacldemo.23
- [53] Radha Kumaran, You-Jin Kim, Anne E. Milner, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2023. The Impact of Navigation Aids on Search Performance and Object Recall in Wide-Area Augmented Reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, 710:1-710:17. https://doi.org/10.1145/3544548. 3581413
- [54] Jean-François Lapointe, Heather Molyneaux, and Mohand Saïd Allili. 2020. A literature review of AR-based remote guidance tasks with user studies. In Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications. Springer International Publishing, Cham, 111–120.
- [55] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. https://doi.org/10.1145/3491102. 3502030
- [56] Vincent Lepetit and M-O Berger. 2000. A semi-automatic method for resolving occlusion in augmented reality. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), Vol. 2. IEEE, IEEE, 225–230.
- [57] Georgianna Lin, Tanmoy Panigrahi, Jon Womack, Devansh Jatin Ponda, Pramod Kotipalli, and Thad Starner. 2021. Comparing order picking guidance with Microsoft hololens, magic leap, google glass xe and paper. In Proceedings of the 22nd international workshop on mobile computing systems and applications. 133–139.
- [58] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. CoRR abs/1907.11692 (2019). arXiv:1907.11692 http://arxiv.org/abs/1907.11692
- [59] Heather Lotherington-Woloszyn. 1993. Do Simplified Texts Simplify Language Comprehension for ESL Learners?. English for Specific Purposes 68 (1993), 31–46.

- [60] Michael R Lyu, Irwin King, TT Wong, Edward Yau, and PW Chan. 2005. Arcade: Augmented reality computing arena for digital entertainment. In 2005 IEEE Aerospace Conference. IEEE, 1–9.
- [61] Blair MacIntyre, Enylton Machado Coelho, and Simon J Julier. 2002. Estimating and adapting to registration errors in augmented reality systems. In *Proceedings IEEE Virtual Reality 2002*. IEEE, IEEE, 73–80.
- [62] Isaias Majil, Mau-Tsuen Yang, and Sophia Yang. 2022. Augmented Reality Based Interactive Cooking Guide. Sensors 22, 21 (2022), 8290. https://doi.org/10.3390/ S22218290
- [63] Yuki Matsuura, Tsutomu Terada, Tomohiro Aoki, Susumu Sonoda, Naoya Isoyama, and Masahiko Tsukamoto. 2019. Readability and legibility of fonts considering shakiness of head mounted displays. In Proceedings of the 2019 ACM International Symposium on Wearable Computers. Association for Computing Machinery, New York, NY, USA, 150–159.
- [64] Alessandro Mulloni, Andreas Dünser, and Dieter Schmalstieg. 2010. Zooming interfaces for augmented reality browsers. In Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York, NY, USA, 161–170.
- [65] Anton Nijholt. 2022. Towards Social Companions in Augmented Reality: Vision and Challenges. In Distributed, Ambient and Pervasive Interactions. Smart Living, Learning, Well-Being and Health, Art and Creativity: 10th International Conference, DAPI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings, Part II. Springer, Springer-Verlag, Berlin, Heidelberg, 304–319.
- [66] Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P Dinu. 2017. Exploring neural text simplification models. In Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers). Association for Computational Linguistics, Vancouver, Canada, 85–91.
- [67] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. 2021. Designing AACs for People with Aphasia Dining in Restaurants. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (, Yokohama, Japan.) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 496, 14 pages. https://doi.org/10.1145/3411764.3445280
- [68] Jennifer Ockerman and Amy Pritchett. 2000. A review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics* 4, 3 (2000), 191–212.
- [69] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing mobile text in head mounted displays: studies on visual preference and text placement. ACM SIGMOBILE Mobile Computing and Communications Review 18, 2 (2014), 20–31.
- [70] Gustavo H Paetzold and Lucia Specia. 2017. A survey on lexical simplification. Journal of Artificial Intelligence Research 60 (2017), 549–593.
- [71] Riccardo Palmarini, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torab-mostaedi. 2018. A systematic review of augmented reality applications in maintenance. Robotics and Computer-Integrated Manufacturing 49 (2018), 215–229.
- [72] Pranav Parekh, Shireen Patel, Nivedita Patel, and Manan Shah. 2020. Systematic review and meta-analysis of augmented reality in medicine, retail, and games. Visual computing for industry, biomedicine, and art 3 (2020), 1–20.
- [73] Wayne Piekarski and Bruce Thomas. 2002. ARQuake: the outdoor augmented reality gaming system. Commun. ACM 45, 1 (2002), 36–38.
- [74] Jing Qian, Jiaju Ma, Xiangyu Li, Benjamin Attal, Haoming Lai, James Tompkin, John F. Hughes, and Jeff Huang. 2019. Portal-ble: Intuitive Free-hand Manipulation in Unbounded Smartphone-based Augmented Reality. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (New Orleans, LA, USA) (UIST '19). Association for Computing Machinery, New York, NY, USA, 133–145. https://doi.org/10.1145/3332165.3347904
- [75] Jing Qian, David A. Shamma, Daniel Avrahami, and Jacob Biehl. 2020. Modality and Depth in Touchless Smartphone Augmented Reality Interactions. In Proceedings of the 2020 ACM International Conference on Interactive Media Experiences (Cornella, Barcelona, Spain) (IMX '20). Association for Computing Machinery, New York, NY, USA, 74–81. https://doi.org/10.1145/3391614.3393648
- [76] Jing Qian, Qi Sun, Curtis Wigington, Han L. Han, Tong Sun, Jennifer Healey, James Tompkin, and Jeff Huang. 2022. Dually Noted: Layout-Aware Annotations with Smartphone Augmented Reality. In CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 5 May 2022, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 552:1–552:15. https://doi.org/10.1145/3491102.3502026
- [77] Jing Qian, Tongyu Zhou, Meredith Young-Ng, Jiaju Ma, Angel Cheung, Xiangyu Li, Ian Gonsher, and Jeff Huang. 2021. Portalware: Exploring Free-Hand AR Drawing with a Dual-Display Smartphone-Wearable Paradigm. In Proceedings of the 2021 ACM Designing Interactive Systems Conference (Virtual Event, USA) (DIS '21). Association for Computing Machinery, New York, NY, USA, 205–219. https://doi.org/10.1145/3461778.3462098
- [78] Juan Carlos Quiroz, Elena Geangu, and Min Hooi Yong. 2018. Emotion recognition using smart watch sensor data: Mixed-design study. JMIR mental health 5, 3 (2018), e10153.

- [79] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. OpenAI blog 1, 8 (2019), 9.
- [80] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J. Mach. Learn. Res. 21 (2020), 140:1–140:67. http://jmlr.org/papers/v21/20-074.html
- [81] Pei-Luen Patrick Rau, Jian Zheng, Zhi Guo, and Jiaqi Li. 2018. Speed reading on virtual reality and augmented reality. Computers & Education 125 (2018), 240-245.
- [82] Luz Rello and Ricardo Baeza-Yates. 2017. How to present more readable text for people with dyslexia. Universal Access in the Information Society 16 (2017), 29–49.
- [83] Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013. Simplify or help? Text simplification strategies for people with dyslexia. In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility (Rio de Janeiro, Brazil). Association for Computing Machinery, New York, NY, USA, Article 15, 10 pages.
- [84] Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. Frequent words improve readability and short words improve understandability for people with dyslexia. In *Human-Computer Interaction – INTERACT* 2013. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 203–219.
- [85] Rufat Rzayev, Paweł W Woźniak, Tilman Dingler, and Niels Henze. 2018. Reading on smart glasses: The effect of text position, presentation type and walking. In Proceedings of the 2018 CHI conference on human factors in computing systems. 1-9.
- [86] Carolina Scarton, Alessio Palmero Aprosio, Sara Tonelli, Tamara Martín-Wanton, and Lucia Specia. 2017. MUSST: A Multilingual Syntactic Simplification Tool. In Proceedings of the IJCNLP 2017, System Demonstrations, Seong-Bae Park and Thepchai Supnithi (Eds.). Association for Computational Linguistics, Tapei, Taiwan, 25–28. https://aclanthology.org/117-3007/
- [87] Andreas Schmeil and Wolfgang Broll. 2007. Mara-a mobile augmented reality-based virtual assistant. In 2007 IEEE Virtual Reality Conference. IEEE, IEEE, 267–270.
- [88] Matthew Shardlow. 2014. A survey of automated text simplification. International Journal of Advanced Computer Science and Applications 4, 1 (2014), 58-70.
- [89] Chuhan Shi, Yicheng Hu, Shenan Wang, Shuai Ma, Chengbo Zheng, Xiaojuan Ma, and Qiong Luo. 2023. RetroLens: A Human-AI Collaborative System for Multi-step Retrosynthetic Route Planning. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (cconf-loc>, city>Hamburg
 city>, ccountry>Germany
 (country>, </conf-loc>) (CHI '23).
 Association for Computing Machinery, New York, NY, USA, Article 770, 20 pages. https://doi.org/10.1145/3544548.3581469
- [90] Kurt Shuster, Mojtaba Komeili, Leonard Adolphs, Stephen Roller, Arthur Szlam, and Jason Weston. 2022. Language Models that Seek for Knowledge: Modular Search & Generation for Dialogue and Prompt Completion. In Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022. Association for Computational Linguistics, 373–393. https://aclanthology.org/2022.findings-emnlp.27
- [91] Chenglei Si, Weijia Shi, Chen Zhao, Luke Zettlemoyer, and Jordan Boyd-Graber. 2023. Getting MoRE out of Mixture of Language Model Reasoning Experts. In Findings of the Association for Computational Linguistics: EMNLP 2023, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 8234–8249. https://doi.org/10.18653/v1/2023.findingsemnlp.552
- [92] Chenglei Si, Chen Zhao, Sewon Min, and Jordan Boyd-Graber. 2022. Re-Examining Calibration: The Case of Question Answering. In Findings of the Association for Computational Linguistics: EMNLP 2022, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2814–2829. https://doi.org/10.18653/v1/2022. findings-emnlp.204
- [93] Advaith Siddharthan. 2006. Syntactic simplification and text cohesion. Research on Language and Computation 4 (2006), 77–109.
- [94] Advaith Siddharthan. 2014. A survey of research on text simplification. ITL-International Journal of Applied Linguistics 165, 2 (2014), 259–298.
- [95] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2022. Prog-Prompt: Generating Situated Robot Task Plans using Large Language Models. CoRR abs/2209.11302 (2022). https://doi.org/10.48550/arXiv.2209.11302 arXiv.2209.11302
- [96] Weiwei Sun, Pengjie Ren, and Zhaochun Ren. 2023. Generative Knowledge Selection for Knowledge-Grounded Dialogues. In Findings of the Association for Computational Linguistics: EACL 2023, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 2077– 2088. https://doi.org/10.18653/v1/2023.findings-eacl.155
- [97] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the*

- SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 73–80.
- [98] Lamma Tatwany and Henda Chorfi Ouertani. 2017. A review on using augmented reality in text translation. In 2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA). IEEE, 1–6.
- [99] Bruce H Thomas, Gregory F Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, Neven AM ElSayed, Ross T Smith, et al. 2018. Situated Analytics. *Immersive analytics* 11190 (2018), 185–220.
- [100] Yuan Tian, Yuxin Ma, Shuxue Quan, and Yi Xu. 2019. Occlusion and collision aware smartphone AR using time-of-flight camera. In Advances in Visual Computing. Springer, Springer International Publishing, Cham, 141–153.
- [101] Antonio E Uva, Michele Gattullo, Vito M Manghisi, Daniele Spagnulo, Giuseppe L Cascella, and Michele Fiorentino. 2018. Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. The International Journal of Advanced Manufacturing Technology 94 (2018), 509–521.
- [102] Chiara Vettori and Ornella Mich. 2011. Supporting deaf children's reading skills: the many challenges of text simplification. In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (Dundee, Scotland, UK) (ASSETS '11). Association for Computing Machinery, New York, NY, USA, 283–284.
- [103] VIDA-NYU. 2024. ptgctl: A Python Library and Command Line Tool for the PTG API. https://github.com/VIDA-NYU/ptgctl. Available online: https://github.com/VIDA-NYU/ptgctl.
- [104] Boshi Wang, Xiang Deng, and Huan Sun. 2022. Iteratively Prompt Pre-trained Language Models for Chain of Thought. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022. Association for Computational Linguistics, 2714–2730. https://aclanthology.org/2022.emnlp-main.174
- [105] Dakuo Wang, Elizabeth F. Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI 2020, Honolulu, HI, USA, April 25-30, 2020, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, 1-6. https://doi.org/10.1145/3334480.3381069
- [106] Zihao Wang, Shaofei Cai, Anji Liu, Xiaojian Ma, and Yitao Liang. 2023. Describe, explain, plan and select: interactive planning with LLMs enables open-world multi-task agents. In Thirty-seventh Conference on Neural Information Processing Systems.
- [107] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In Advances in Neural Information Processing Systems, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf
- [108] Matthias M Wloka and Brian G Anderson. 1995. Resolving occlusion in augmented reality. In Proceedings of the 1995 Symposium on Interactive 3D Graphics (Monterey, California, USA). Association for Computing Machinery, New York, NY, USA, 5–12.
- [109] Guande Wu, Shunan Guo, Jane Hoffswell, Gromit Yeuk-Yin Chan, Ryan A. Rossi, and Eunyee Koh. 2024. Socrates: Data Story Generation via Adaptive Machine-Guided Elicitation of User Feedback. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 131–141. https://doi.org/10.1109/TVCG.2023. 3327363
- [110] Guande Wu, Jianzhe Lin, and Cláudio T. Silva. 2022. IntentVizor: Towards Generic Query Guided Interactive Video Summarization. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. IEEE, 10493–10502. https://doi.org/10.1109/CVPR52688.2022. 01025
- [111] Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. 2013. Current status, opportunities and challenges of augmented reality in education. Computers & education 62 (2013), 41–49.
- [112] Jianghao Xiong, En-Lin Hsiang, Ziqian He, Tao Zhan, and Shin-Tson Wu. 2021. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications* 10, 1 (2021), 216. https://doi.org/10.1038/s41377-021-00658-8
- [113] Victoria Yaneva. 2015. Easy-read documents as a gold standard for evaluation of text simplification output. In *Proceedings of the Student Research Workshop*. INCOMA Ltd. Shoumen, BULGARIA, Hissar, Bulgaria, 30–36.
- [114] Xi Ye and Greg Durrett. 2022. Can Explanations Be Useful for Calibrating Black Box Models?. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022. Association for Computational Linguistics, 6199-6212. https:

- //doi.org/10.18653/v1/2022.acl-long.429
- [115] Xi Ye and Greg Durrett. 2022. The Unreliability of Explanations in Few-shot Prompting for Textual Reasoning. In Advances in Neural Information Processing Systems, Vol. 35. Curran Associates, Inc., 30378–30392. http://papers.nips.cc/paper_files/paper/2022/hash/ c402501846f9fe03e2cac015b3f0e6b1-Abstract-Conference.html
- [116] Keyu Zhai, Yiming Cao, Wenjun Hou, and Xueming Li. 2020. Interactive Mixed Reality Cooking Assistant for Unskilled Operating Scenario. In Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications - 12th International Conference, VAMR 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 12191), Jessie Y. C. Chen and Gino Fragomeni (Eds.). Springer, 178–195. https://doi.org/10.1007/978-3-030-49698-2_13
- [117] Shun Zhang, Zhenfang Chen, Yikang Shen, Mingyu Ding, Joshua B. Tenenbaum, and Chuang Gan. 2023. Planning with Large Language Models for Code Generation. In *The Eleventh International Conference on Learning Representations*. OpenReview.net. https://openreview.net/pdf?id=Lr8cOOtYbfL
- [118] Shujian Zhang, Chengyue Gong, and Eunsol Choi. 2021. Knowing More About Questions Can Help: Improving Calibration in Question Answering. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (Findings of ACL, Vol. ACL/IJCNLP 2021), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 1958–1970. https://doi.org/10.18653/v1/2021.findings-acl.172
- [119] Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. 2021. Calibrate Before Use: Improving Few-shot Performance of Language Models. In Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139), Marina Meila and Tong Zhang (Eds.). PMLR, 12697–12706. http://proceedings.mlr.press/v139/zhao21c.html
- [120] Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling Stories from Computational Notebooks: AI-Assisted Presentation Slides Creation for Presenting Data Science Work. In CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May

- 2022, Simone D. J. Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Mark Drucker, Julie R. Williamson, and Koji Yatani (Eds.). ACM, 53:1–53:20. https://doi.org/10.1145/3491102.3517615
- [121] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, 351:1–351:19. https://doi.org/10.1145/3544548.3581131
- [122] Ting Zheng, Marco Ardolino, Andrea Bacchetti, and Marco Perona. 2021. The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. *International Journal of Production Research* 59, 6 (2021), 1922–1954.
- [123] Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 2125–2134.
- [124] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. 2022. Detecting twenty-thousand classes using image-level supervision. In European Conference on Computer Vision. Springer, Springer Nature Switzerland, Cham, 350–368.
- [125] Jon Zubizarreta, Iker Aguinaga, and Aiert Amundarain. 2019. A framework for augmented reality guidance in industry. The International Journal of Advanced Manufacturing Technology 102 (2019), 4095–4108.

A EXAMPLE TEXT SIMPLIFICATION RESULTS

Here, we provide the sample results recorded during our experiments. The Study 1 results are listed in Table 3 (Task 1) and Table 4 (Task 2).

Step	Original	Simplification
1	To create a coffee, first please carefully place	Place dripper (on your left) on coffee mug.
	the pour-over dripper over the coffee mug.	
2	Prepare the filter insert by folding the paper	Fold paper filter in half, then half again. Put
	filter in half to create a semi-circle, and in half	filter in dripper, form cone shape.
	again to create a quarter-circle. Place the paper	
	filter in the dripper and spread open to create a	
	cone.	
3	Rinse the filter. Pour enough hot water into the	Wet filter with water to rinse away residue.
	filter to wet it. The entire paper filter should be	
	moist. Rinsing the filter will remove any papery	
	residue so your coffee doesn't have a woodsy	
	taste.	
4	Lift up the dripper and pour out the water. Then	Remove dripper, pour out water, and return drip-
	set the dripper with the wet filter back on the	per to coffee mug.
	coffee mug.	
5	Get out a digital scale and measure out 3 table-	Measure 30g coffee beans on a digital scale
	spoons (about 30 g) of coffee beans. Measure	(right side), place in grinder (right side).
	out 30 g of whole beans and place them in your	
	grinder.	
6	Grind the beans until the coffee grounds are the	Grind beans for 20 seconds, until coarse sand
	consistency of coarse sand, about 20 seconds.	consistency.
7	Transfer the coffee grounds to the filter cone.	Move grounds to filter cone. Set coffee mug with
	Then place the coffee mug with the dripper on	dripper on scale, zero it.
	a digital scale and set it to zero.	
8	Slowly pour the water over the grounds in a	Slowly pour water in circles over grounds, stop-
	circular motion. Do not overfill beyond the top	ping at 100g on scale.
	of the paper filter. Your scale should read 100	
	g once you've poured enough water into the	
	dripper.	
9	Let the coffee drain completely into the mug	Drain coffee into mug and wait for 30 seconds
	and wait for 30 seconds and you can complete	to end.
	the task;	

Table 3: Study 1 Task 1 sample result. The original and simplified versions of the text are listed. The sample is collected from P10's experiment session and the simplified results for other participants may vary slightly due to spatial context.

Step	Original	Simplification
1	Before arranging the meeting room, take a mo-	Tidy desk, move the unnecessary items to other
	ment to tidy up the desk and move anything	desks.
	that's not necessary to other desks;	
2	Once the desk is clear, bring the power strip on	Put power strip on desk, connect phone charger
	the desk and connect the Charger to the power	to it.
	strip so the meeting attendants can use.	
3	Connect the camera's charger to the power strip	Connect camera to strip, facing opposite of TV.
	and position the camera at the opposite end of	
	the desk from the TV.	
4	Arrange the chairs in the meeting room. Make	Arrange chairs on two sides. Leave space of
	sure that there's enough space between each	roughly two A4 papers' length apart. Window
	chair - roughly 1.5 feet should suffice. Position	side: 1 chair. Other side: 5 chairs.
	one chair on the window side, and place five	
	chairs on the other side.	
5	Next, place cups of water and papers on each	Place water, paper onto desk in front of chairs.
	chair. Each person should have one cup of water	
	and paper;	
6	Put up the desk nameplates on on each chair.	Place nameplates: Window side: Alice (win-
	When Alice is on the side of the window, other	dow); sequence (left to right) on other side: Bob,
	desk nameplates should be put on the other side.	Amy, Andy, Dave, Luis.
	The sequence is Bob, Amy, Andy, Dave and Luis.	
7	Since Alice is the VIP in the meeting, place make	Place remote controller at Alice's position on
	it clearly by putting the remote controller to	desk.
	Alice's position.	

Table 4: Study 1 Task 2 sample result. The original and simplified versions of the text are listed. The sample is collected from P10's experiment session and the simplified results for other participants may vary slightly due to spatial context.