Can AI Bridge the Literacy Gap? Developing a GPT-4 Summarization Tool for Low Literacy

Sichen Guo[1][0009−0005−2163−4218], François Leborgne[1][0009−0000−9639−9519], Jun Hu[1][0000−0003−2714−6264], and Walter Baets[2][0000−0003−1860−2483]

[1] Department of Industrial Design, Eindhoven University of Technology
[2] Eindhoven Engine
s.guo3@tue.nl

Abstract. In the Netherlands, 2.5 million adults grapple with low literacy—a gap between integration and an increasingly information-dense society. To address this, we have developed an innovative intervention using GPT-4, integrated within the user interface built via Streamlit, to summarize and simplify the complexity of official documents and everyday text. This paper outlines the design and anticipated functionality of our tool, which uses the summarization capabilities of GPT-4 to transform complicated language into concise, accessible content with clear action points and connect with related organizations. Six participants were invited to this study. We evaluate the tool's impact on enhancing the societal participation of adults lacking basic skills, proposing that such an application can empower low-literacy individuals and bridge the gap to an informed and equitable society.

Keywords: LLM · AI · Human AI Interaction · low literacy

1 Introduction

On a typical quiet afternoon, you exit your room to find a letter waiting in your mailbox. The envelope's light blue color and prominent logo indicate its importance, originating from the government. Nevertheless, feelings of anxiety and irritation prompt you to delay retrieving the letter. Fears of unfavorable news, such as a fine or additional payment, weigh heavily on your mind. These concerns cause you to leave the letter unopened for six months. Ultimately, you muster the courage to open it. Initially, you experience uncertainty. Soon, this uncertainty gives way to discouragement. The reason lies in the letter's content, which features complex numbers and multicolored hyperlinks, rendering its message unclear.

This vignette illustrates a harsh reality. Lacking basic reading, math, and writing skills leads to many daily challenges. It makes managing money, getting information, dealing with health issues, and relationships difficult [1, 2]. The Survey of Adult Skills (PIAAC) shows adults' ability in three key skills: literacy, numeracy, and problem solving in technology-rich environments. Literacy is the ability to understand and respond to written texts. Numeracy is the ability to use numbers and math. Problem solving in technology-rich environments is the ability to access, interpret, and analyze information in digital environments [3]. Low literacy refers to individuals who struggle with reading and writing without being completely illiterate. It is known to be a significant issue in developed countries, including the Netherlands and the consequences are usually severe, leading to feelings of shame, insecurity, and dependence [4].

In the Netherlands, about 2.5 million people struggle with reading, math, and using digital devices due to low literacy. This includes about 1.3 million residents aged 16-65. Research shows that many adults are hesitant to admit their struggles with literacy and, as a result, rarely seek help [5, 6]. These people often struggle to find and keep jobs. Nearly half of those with poor basic skills are unemployed. Additionally, people with limited reading and writing abilities often face health and communication challenges [4, 7, 8]. Therefore, the Dutch government aims for every municipality to assist individuals with low literacy by 2024 [9].

Nowadays, many researchers focus on early intervention. It targets children's and teenagers' education [10]. However, the Survey of Adult Skills (PIAAC) reveals that, like most participating countries, a large minority of adults in the Netherlands are poor at literacy, numeracy, and problem-solving. Moreover, foreign-language immigrants in the Netherlands exhibit significantly lower Dutch proficiency levels than native-born adults with Dutch as their first language [3]. The National Library's Library and Basic Skills program supports efforts to address this issue by establishing a nationwide library infrastructure to improve basic skills for 45,000 vulnerable adults, including those who are low literate, refugees, migrants, unemployed, elderly, and computer illiterate [5].

In this context, the Emergency Lab from Eindhoven Engine[3] and Eindhoven University of Technology[4] aim to explore a more innovative solution to this problem. This work is part of the Met Mij [11] project from Eindhoven Engine. It investigates how AI can improve interaction and user experience for individuals with low literacy in Dutch using a well-known large language model (GPT-4) designed in an online interface (Streamlit) to tailor to low-literacy people. This study focuses on the development, design, and evaluation of a document summarization tool based on GPT-4. With a user-centered approach, qualitative and quantitative data are collected and analyzed. The goal of this study is to contribute to exploring how Artificial Intelligence (AI) can address the low literacy problem through more accessible design. The findings provide key insights into user interaction, covering information on informed effectiveness and efficiency, personal needs, and trust. These insights give a clearer picture of how low-literacy individuals interact with and obtain information from official documents despite their lack of proficiency in the Dutch language.

[3] https://eindhovenengine.nl/

[4] https://www.tue.nl/en/

2 Related work

2.1 Wicked Problem

In 1973, design theorists Horst Rittel and Melvin Webber introduced the term "wicked problem". They did this to show the complexities and challenges of addressing planning and social policy problems. Wicked problems lack clear aims and solutions. They face real-world constraints that prevent many risk-free attempts to solve them [12]. For these kinds of problems, there are no clear rules to stop complex problems, and there is no standard "right" or "wrong" criterion for solving them, only "good" and "bad" solutions [13].

Mari Suoheimo et al. [14] observed a significant shift in design practice over the last fifteen years, influenced by increased interaction with social and political matters. They suggest that designers should improve at navigating complex contexts. They should also get better at aligning designs with challenging situations. This will help them to effectively address tricky issues. Therefore, designers face complex, unpredictable, wicked problems. They must evolve their understanding and proposals while being open and sensitive.

Low literacy is a wicked problem in Dutch society, affecting a diverse group with various needs. Researcher Inge Hootsmans has identified a big constraint. It's the hidden group of young adults aged 18-40 who struggle with literacy issues [15]. This problem leads to feelings of shame, lack of well-being, and marginalization. As a result, it is crucial to address this long-standing issue in a more creative and innovative way.

2.2 Target Audience

In Eindhoven, 7% of the working population does not have the necessary basic skills, which is approximately 19,000 inhabitants [16]. The number is even higher in the report of the National Audit Office [17]. In recent years, efforts have been made to assist residents with limited basic skills. Although many hope for a decline in these numbers, the expectation is that they will actually rise. This increase is expected because more young people are leaving education without enough language skills. There are also more non-native speakers, and society is becoming more digital [18].

Low literacy individuals are split into different levels, as shown in Table 1. The "Everyone Basic Skills City Plan 2024-2028" outlines the levels in Eindhoven. It categorizes people based on their Dutch proficiency as a first (NT1) or second (NT2) language [18]. The levels are aligned with the Common European Framework of Reference for Languages (CEFR) and other educational standards. Below is a summary of the different literacy levels and the corresponding number of individuals in each category [19, 18]. The NT1 group represents about 54.2 percent of the low literacy population.

The city plan of Eindhoven aims to ensure that every adult in Eindhoven possesses basic language, arithmetic, and digital skills. A significant challenge lies in reaching and motivating NT1 individuals mentioned by other research [11,

Table 1. Language Level Reference Framework for Language and Arithmetic

Dutch as First Language (NT1)	Illiterate	1F	2F	3F	4F
Dutch as Second Language (NT2)	A0	A1-A2	B1	B2	C1-C2
Comparable Education Level		End of primary school	End of VMBO[5] and MBO[6]1,2,3	End of MBO-4 or HAVO[7]	End of pre-university education (vwo), higher professional education (hbo), or university (wo)
	Low literate		Not low literate		
Legal regulations	Integration, Language Requirements, WEB				

source https://www.eindhoven.nl/

[5] Preparatory secondary vocational education

[6] Secondary vocational education

[7] General secondary education

15]. To gain more insights into their behaviors and feedback for improvement, we studied NT1 together with the NT2 group.

According to the city plan, this study will focus on both NT1 and NT2 groups who belong to 1F (Table 1). These individuals, at rudimentary and basic literacy levels, are the target audience for this study.

2.3 Large Language Models(LLMs) and GPT-4

Large language models (LLMs), like GPT-3 and GPT4, have greatly improved natural language processing (NLP). They were trained on huge amounts of text data to generate human-like text and do language tasks accurately [20–22]. Recent advancements in LLMs have demonstrated their potential to achieve a level of intelligence comparable to humans [23, 24].

OpenAI developed the Generative Pre-Trained Transformer (GPT), a language model capable of producing human-like text [25]. The development of GPT follows a two-step process: generative, unsupervised pretraining using unlabeled data, and discriminative, supervised fine-tuning [26, 27]. GPT stands out due to the scale of its training program and the extensive amount of data used. With access to the entire internet, the algorithm is trained on billions of data sources [28]. As a result, GPT can perform a wide range of language-based tasks, such as translation, question answering, and text generation.

The latest large-scale, multimodal model, GPT-4, developed by OpenAI, is a significant area of study due to its ability to process both image and text input and generate text output. This model holds great potential for various applications, such as dialogue systems, text summarization, and machine translation [29].

Currently, there is much research related to how the GPT model can solve or improve financial literacy and health literacy [30–34]. Moreover, there is also research on GPT-3 on simplified mathematics problems in education and people with cognitive impairments, which all contribute to better understandable results for the target audience [35]. However, research also indicates that the current GPT interface is not suitable and accessible for people with low literacy, especially those who have difficulties in reading and writing [36].

Additionally, XuanXin Wu et al. found that GPT-4 makes simpler outputs with fewer errors. It also keeps the original meaning better than the best current model, Control-T5. Their research highlights GPT-4's superior performance in text simplification tasks [37].

In conclusion, the GPT-4 model demonstrates an effective ability to comprehend context, summarize, and simplify text, thereby generating more understandable and accessible content. Due to its superior performance, we have chosen the GPT-4 model for text summarization and simplification for our work.

3 Method

3.1 Interactive Prototype

The interface prototype prioritizes logic and practicality. It has an interactive interface tailored to our research. It also takes into account the significant cultural disparity between users and developers. Dutch experts conducted initial testing and interviews during the first iteration. Figure 1 illustrates the logic underlying the prototype. This prototype includes a scenario and interactive interface built with Streamlit[8]. They help users form immersive during the evaluation study and are easy to interact with.

The prototype uses the prompt test to give clear context and visuals through icons. It categorizes the summarized content into four sections. The comprehensive user summary page is illustrated in Figure 2.

- *Sender Information* - Identification of the party responsible for sending the correspondence
- *Action Points* - Specific tasks the recipient is expected to undertake upon receiving the letter
- *Contact Details* - Means of communication
- *Direct Action* - A direct avenue for recipients to establish contact

[8] https://streamlit.io/

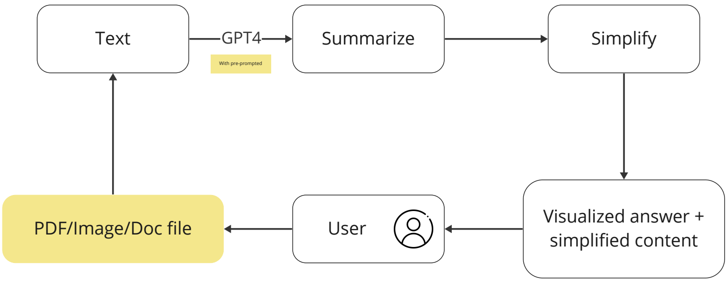
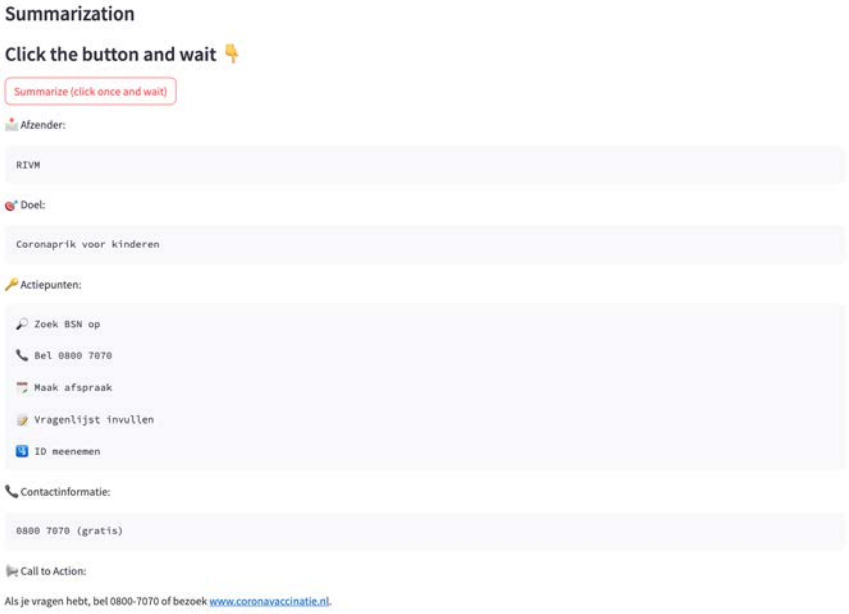


Fig. 1. user information flow.



Summarization

Click the button and wait 👇

📩 Afzender:

```
RIVM
```

🎯 Doel:

```
Coronaprik voor kinderen
```

🔑 Actiepunten:

```
🔎 Zoek BSN op
📞 Bel 0800 7070
📅 Maak afspraak
📝 Vragenlijst invullen
🆔 ID meenemen
```

📞 Contactinformatie:

```
0800 7070 (gratis)
```

📢 Call to Action:

Als je vragen hebt, bel 0800-7070 of bezoek www.coronavaccinatie.nl.

Fig. 2. Summarized content and interface

3.2 Evaluation Study

In the Evaluation study, we first did the whole process test with baseline and the prototype with the NT1 group (P1, P2). An oral assistant and a text content Dutch assistant together with researcher work on the study. Then, we tested with four NT2 group members (P3-P6) with a link they can test on their digital devices. The qualitative and quantitative data are collected. It was based on the six questions about the four categories from the letter. The User Experience Questionnaire (UEQ) combined with Cognitive Load Questionnaire(CLQ) are used in this study [38, 39]. It focuses on effectiveness, efficiency, and cognitive load. Additionally, we did semi-structured interviews for the NT1 and NT2 groups. We asked open-ended questions to learn more about their user experience.

3.3 Participants

In total, 6 users were evaluated and interacted with the tool and did the interview. Hereby, there are 2 people from the NT1 group and four people from the NT2 group. (n=6) We found five themes in the data. They are: (1) user interaction, (2) informing effectively, (3) reading efficiently, (4) understanding context, and (5) personalization and customization needs

Set up The evaluation consists of two sessions, as shown in Figure 3. In the first session, the user chooses the letter and reads the original text. There are four designated topics. However, only two are currently accessible: health and finance. After thoroughly reviewing the correspondence, participants will be prompted to use the evaluation questionnaire. Subsequently, respondents will address six inquiries about content across three distinct categories. A secondary questionnaire focusing on user experience and cognitive load will ensue, followed by an invitation for participants to partake in an interactive session involving our prototype. Upon completing the text overview, participants will once again complete the evaluation questionnaire and the content-related inquiry. Lastly, an in-depth semi-structured interview is reserved for individuals categorized under NT1 to glean additional insights.

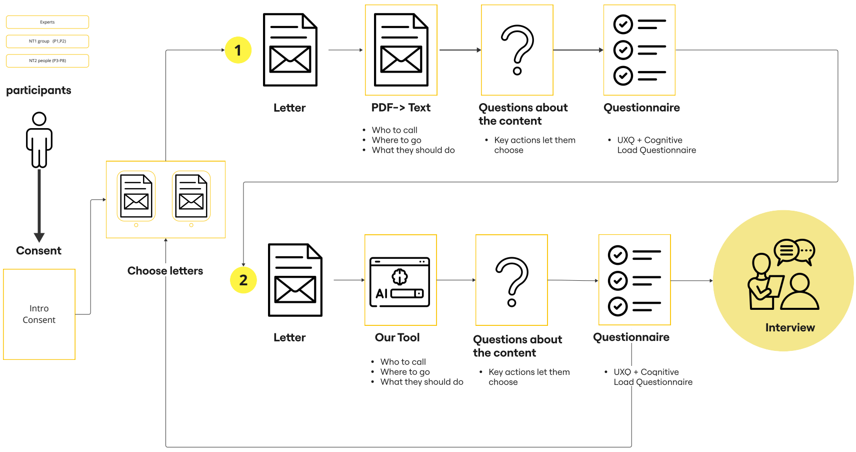


Fig. 3. Evaluation set up.

4 Findings

4.1 Quantitative Data Analysis

The data analysis shows participant performance across two groups (NT1 and NT2) and two letter types (health and financial). It covers before and after an intervention. The metrics evaluated include sum scores (SS) and accuracy ratings (AR). The data is summarized in Table 2

Table 2. Participant Performance Before and After Intervention

Group	**Participant**	**Letter Type**	**Sum (Before)**	**Accuracy (Before)**	**Sum (After)**	**Accuracy (After)**
NT1	P1	health	11	5	44	5
NT1	P2	health	25	2	34	1
NT2	P3	health	29	5	42	5
NT2	P3	financial	24	5	40	5
NT2	P4	health	32	6	45	6
NT2	P4	financial	45	6	45	6
NT2	P5	health	31	5	31	6
NT2	P5	financial	31	6	44	6
NT2	P6	health	34	6	44	6
NT2	P6	financial	34	6	43	5
Average			29.6	5.2	41.2	5.1

The data in Table 2 demonstrate a significant increase in average sum scores (SS) from 29.6 before the intervention to 41.2 after, indicating enhanced participant performance across both health and financial letter types. Despite this improvement, the average accuracy rating (AR) slightly decreased from 5.2 to 5.1, suggesting that overall task performance improved while accuracy did not uniformly benefit. However it is notable for P2 from NT1 group, even though P2 claimed to get a higher score, the accuracy after the tool is actually less.

Figure 4 and Figure 5 show a substantial increase in cognitive load post-intervention, with median and mean values rising significantly. Participants handling financial letters generally maintained or improved their SS and AR, as exemplified by P4. Besides, the bar chart in Figure 5 highlights big increases in cognitive load related to effectiveness and efficiency. Post-intervention scores increased to 4.53 and 4.55, respectively. This suggests that participants found tasks more effective and efficient.

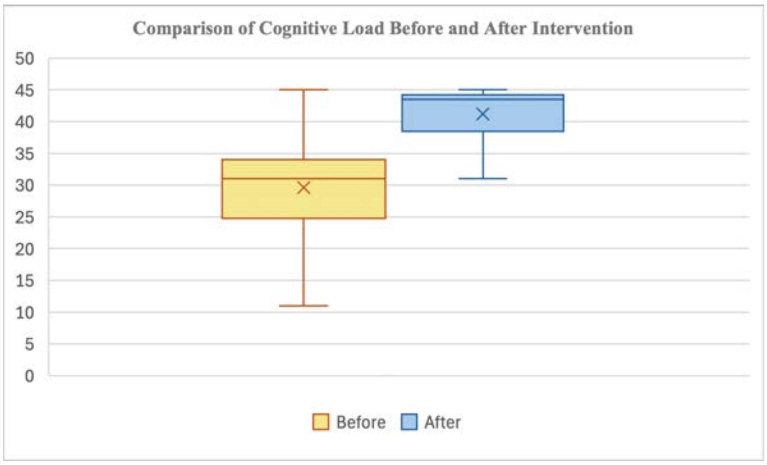


Fig. 4. Comparison of Cognitive Load Before and After Intervention.

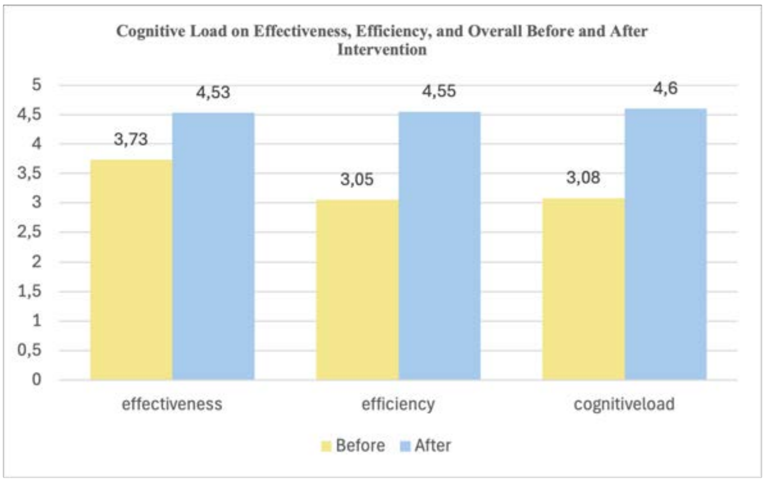


Fig. 5. Cognitive Load on Effectiveness, Efficiency, and Overall Before and After Intervention.

4.2 Qualitative Data Analysis

A significant difference in attitude between NT1 and NT2 groups Aiming to explore the whole user experience further, each participant got a semi-structured interview and provided qualitative data. Researchers found that there were significant behaviors in the NT1 group and the NT2 group. The NT1 and NT2 groups can quickly search for answers when they get the letter. However, NT2 has more confidence in answering due to their better searching skills, as shown in Table 2. When P1 from the NT1 group got the letter, it was noticeable that the participant mentioned, "If it's blue or thick... I don't open it right away... I just put it aside." This reaction highlights the emotional barrier and procrastination associated with dense, official-looking mail." The NT2 group has a learning attitude toward reading letters. P3 (NT2) mentioned, " I can understand most of the content, about 50%." Generally, the NT2 group showed more confidence than the NT1 group when they received official documents. Also, P1 mentioned that after finishing the letters, "I found it took a lot of effort because I found these list things very unclear." This underscores the need for clear and concise communication in official documents. Most of the NT2 group mentioned that they have memorable letter-related to the daily topics .

Preference for Summarized Content There was a clear preference for summarized content over full text. One participant mentioned, "But if something just says like... Immediately what you have to do. Then there is very little room for noise and panic." This preference highlights the importance of direct and clear communication. Furthermore, an NT1 participant mentioned twice that the text was perfect. They said they wanted to use it for other letters. "I always find it very nice because reading a (summarized) letter is exciting. It helps me avoid misunderstanding the information and prevents me from creating fake news in my panic." From the NT2 group, there are three participants who mentioned that it would be good to use it for financial problems or to help them select them of different letters. One participant mentioned specifically that "it would be nice if I could use it for reporting taxes, and I don't need to search for pieces of information online."

Improved Confidence and Efficiency Using the tool especially made NT1 participants feel more confident and efficient in handling official documents. One participant stated, "This app with summarizing is always faster than waiting two weeks to open that letter." This reflects the tool's ability to streamline the process and reduce the cognitive load. Additionally, the NT2 group mentioned that this tool can enhance the time spent reviewing official documents. One mentioned that if the tool can help them know the topic quicker, it can also help them tell if this letter is for them or not.

Accessible user experience design elements. Participants expressed a need for customization, such as larger font sizes and clear visual indicators. One from the NT1 group noted, "I would have made it a bit bigger personally." This feedback points to the importance of adaptable design elements in enhancing user experience. Meanwhile, participants mentioned that it would be nice if there were an audio option or if some people could read the content. NT2 people want more layers; for example, they can adjust how detailed information they can see. All participants mentioned that it would be nice if the tool could scan the document on the mobile.

Trust between technology and users. Trust is the biggest issue mentioned by NT2 participants mostly; one participant noted, “I want to know more explanation on how the tool summarized the key action points. For example, if I click on the summarization, it can turn to the original file automatically”. There is a trust gap between LLMs and users; they want to know more about why and how the information is given. Besides, it depends on the topic. If they know the documents are important, they prefer to use a summarization tool first and then double-check with the original files.

5 Discussion

The digital summarization tool enhances the reading of official documents and streamlines the process of responding to them. NT1 participants, who had previously delayed engaging with such documents due to anxiety and confusion, reported significant improvements in their ability to handle these tasks promptly. The tool effectively presents key information, enabling users to make decisions and act quickly without prolonged hesitation. This efficiency reduces both the time spent and the cognitive burden associated with understanding complex official documents.

Accessibility emerged as a critical factor in the tool’s effectiveness. Participants emphasized the need for customization options, such as larger font sizes, clear visual indicators, and audio support, to cater to their diverse needs. These features are needed for an inclusive user experience. They are especially important for low-literacy people who may struggle with standard text. By adding accessible design elements, the tool can serve a wider range of users, ensuring that individuals with different literacy levels can manage official documents effectively.

The qualitative feedback underscored the emotional challenges faced by NT1 participants when dealing with official documents. The tool’s design addressed these challenges by reducing anxiety and fostering a sense of control over the information. Participants prefer summarized content and they mention it is important to communicate directly and clearly. This helps reduce overwhelm and panic. Researchers analyze the users’ feelings and aims to reduce negative ones. This approach improves user engagement and satisfaction.

The study revealed distinct behavioral differences between NT1 and NT2 groups. NT1 participants often feel anxiety and delay dealing with official documents. In contrast, NT2 participants show more confidence and a proactive attitude, likely due to their better-developed basic skills. This difference shows the importance of tailored interventions that address the specific needs and challenges of each group. The tool enhances NT1 participants. on getting information simpler and clearer. This is especially helpful for overcoming emotional barriers and promoting timely action.

Trust is crucial for adopting digital summarization tools. This is especially true for NT2 participants. Since they want more transparency in how the tool works. Users need explainable AI features to understand how the summarization works. The functions let users compare the summary with the original documents, which will be useful to solve this problem. These features can boost trust and reliability. Building this trust is essential. Users need it to feel confident in using the tool for critical tasks. This trust will increase the tool's overall effectiveness and user satisfaction.

6 Conclusion

In conclusion, the digital summarization tool with GPT-4 has shown great potential. It can improve the efficiency, accessibility, and emotional comfort of low-literacy individuals with official documents. While the tool enhances performance and reduces cognitive load, further refinements in accessible design, emotional considerations, and explainability are necessary to fully meet the needs of both NT1 and NT2 groups. By addressing these areas, the tool can play a more transformative role in empowering low-literacy individuals and bridging the gap to an informed and equitable society.

About Sichen Guo

Sichen Guo is an Engineering Doctorate (EngD) working in Eindhoven University's Technology, Netherlands, Human System Interaction program. She is also a member of the Emergency Lab at Eindhoven Engine, which focuses on designing to bridge the gap between people who lack basic skills and society. She has a bachelor's and master's degree in Industrial Design. During her master's study, she focused on socially inclusive design, user interaction, VR, and gamification. As a designer, engineer, and researcher, she takes a user-centered approach to real-life problems, combining innovative technology and design methodologies.

References

1. B Suresh Lal. The economic and social cost of illiteracy: an overview. *International Journal of Advance Research and Innovative Ideas in Education*, 1(5):663–670, 2015.
2. Nancy D Berkman, Darren A DeWalt, Michael P Pignone, Stacey L Sheridan, Kathleen N Lohr, Linda Lux, Sharon F Sutton, Tameka Swinson, and Arthur J Bonito. Literacy and health outcomes: summary. *AHRQ evidence report summaries*, 2004.
3. Education gps, oecd. http://gpseducation.oecd.org. last accessed: 2024/05/21.

4. Laaggeletterdheid-slo. https://www.slo.nl/thema/vakspecifieke-thema/nederlands/laaggeletterdheid/. last accessed: 2024-05-21.
5. Adriaan Langendonk and Maaike Toonen. Dutch approach to prevent and curate low literacy. 2017.
6. World Literacy Foundation. The economic and social cost of illiteracy: A white paper by the world literacy foundation. *The World Literacy Summit*, 2018.
7. Lezenenschrijven. https://www.lezenenschrijven.nl/. last accessed: 2024-05-21.
8. Basisvaardigheden. https://basisvaardigheden.nl/themas/thema-basis-over-basisvaardigheden. last accessed: 2024-05-21.
9. Rijksoverheid. https://www.rijksoverheid.nl/onderwerpen/laaggeletterdheid/aanpak-laaggeletterdheid. last accessed: 2024-05-21.
10. Sanne L Tamboer, Inge Molenaar, Tibor Bosse, and Mariska Kleemans. Testing an intervention to stimulate early adolescents' news literacy application in the netherlands: A classroom experiment. *Journal of Children and Media*, 18(1):60–79, 2024.
11. Jéssica Messias Goss Dos Santos. 'met mij': Ai-enhanced tool for identifying and connecting people and services, 2024.
12. Randall Bass. What's the problem now? *To Improve the Academy: A Journal of Educational Development*, 39(1), 2020.
13. Horst WJ Rittel and Melvin M Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973.
14. Bo Westerlund and Katarina Wetter-Edman. Dealing with wicked problems, in messy contexts, through prototyping. *The Design Journal*, 20(sup1):S886–S899, 2017.
15. Inge Hootsmans. Navigating through low literacy, 2024.
16. Geletterdheid in zicht. https://geletterdheidinzicht.nl/. last accessed: 2024-05-21.
17. Rekenkamer. https://www.rekenkamer.nl/. last accessed: 2024-05-21.
18. eindhoven.nl.
19. CONSIGLIO D'EUROPA. Common european framework of reference for languages (cefr). *Learning, teaching*, 2018.
20. Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.
21. Elizabeth D Liddy. Document retrieval, automatic. 2005.
22. Mark Burry. A new agenda for ai-based urban design and planning. In *Artificial Intelligence in Urban Planning and Design*, pages 3–20. Elsevier, 2022.
23. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
24. Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
25. Robert Dale. Gpt-3: What's it good for? *Natural Language Engineering*, 27(1):113–118, 2021.
26. Dumitru Erhan, Aaron Courville, Yoshua Bengio, and Pascal Vincent. Why does unsupervised pre-training help deep learning? In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 201–208. JMLR Workshop and Conference Proceedings, 2010.

27. Paweł Budzianowski and Ivan Vulić. Hello, it's gpt-2–how can i help you? towards the use of pretrained language models for task-oriented dialogue systems. *arXiv preprint arXiv:1907.05774*, 2019.
28. Luciano Floridi and Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694, 2020.
29. Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
30. Pragnya Ramjee, Bhuvan Sachdeva, Satvik Golechha, Shreyas Kulkarni, Geeta Fulari, Kaushik Murali, and Mohit Jain. Cataractbot: An llm-powered expert-in-the-loop chatbot for cataract patients. *arXiv preprint arXiv:2402.04620*, 2024.
31. Ziqi Yang, Xuhai Xu, Bingsheng Yao, Ethan Rogers, Shao Zhang, Stephen Intille, Nawar Shara, Guodong Gordon Gao, and Dakuo Wang. Talk2care: An llm-based voice assistant for communication between healthcare providers and older adults. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2):1–35, 2024.
32. Md Mushfiqur Rahman, Mohammad Sabik Irbaz, Kai North, Michelle S Williams, Marcos Zampieri, and Kevin Lybarger. Health text simplification: An annotated corpus for digestive cancer education and novel strategies for reinforcement learning. *arXiv preprint arXiv:2401.15043*, 2024.
33. Hana L Haver, Anuj K Gupta, Emily B Ambinder, Manisha Bahl, Eniola T Oluyemi, Jean Jeudy, and Paul H Yi. Evaluating the use of chatgpt to accurately simplify patient-centered information about breast cancer prevention and screening. *Radiology: Imaging Cancer*, 6(2):e230086, 2024.
34. Julie Ayre, Olivia Mac, Kirsten McCaffery, Brad R McKay, Mingyi Liu, Yi Shi, Atria Rezwan, and Adam G Dunn. New frontiers in health literacy: using chatgpt to simplify health information for people in the community. *Journal of General Internal Medicine*, 39(4):573–577, 2024.
35. Nirmal Patel, Pooja Nagpal, Tirth Shah, Aditya Sharma, Shrey Malvi, and Derek Lomas. Improving mathematics assessment readability: Do large language models help? *Journal of Computer Assisted Learning*, 39(3):804–822, 2023.
36. A Fiora, F Piferi, P Crovari, and F Garzotto. Exploring large language models for the education of individuals with cognitive impairments. In *INTED2024 Proceedings*, pages 4479–4487. IATED, 2024.
37. Xuanxin Wu and Yuki Arase. An in-depth evaluation of gpt-4 in sentence simplification with error-based human assessment. *arXiv preprint arXiv:2403.04963*, 2024.
38. M Schrepp. User experience questionnaire handbook version 8. 2019. *URL: https://wWww. ueq-online. org/Material/Handbook. pdf*, 2021.
39. Gordon Matthew. The effect of adding same-language subtitles to recorded lectures for non-native, english speakers in e-learning environments. *Research in Learning Technology*, 28, 2020.