Designing, Developing, and Evaluating AI-driven Text Entry Systems for Augmentative and Alternative Communication Users and Researchers

Boyin Yang
University of Cambridge
Cambridge, UK
by266@cam.ac.uk

Per Ola Kristensson
University of Cambridge
Cambridge, UK
pok21@cam.ac.uk

ABSTRACT

Non-speaking individuals with motor disabilities heavily rely on augmentative and alternative communication (AAC) text entry systems to communicate. However, designing, developing, and evaluating AAC text entry systems for users and researchers gives rise to several challenges, such as the difficulty of eliciting requirements for a highly heterogeneous user population using a wide variety of assistive devices. This research aims to address such challenges and bridge the gaps between AAC stakeholders while exploring the potential of artificial intelligence (AI)-powered AAC systems to cater to diverse AAC needs and wants. To achieve these goals, we propose an *imperfect surrogate user* model for AAC system design and evaluation, develop a dedicated machine learning language model for AAC purposes, and design an AI-driven text entry system *Tinkerable AAC (TAAC)* which is ready for user testing. The objective of these efforts is to ultimately reduce the communication gap between AAC users and their speaking partners.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility design and evaluation methods**; **Accessibility systems and tools**; **Text input**.

KEYWORDS

AI-driven systems, text entry design, predictive text entry, augmentative and alternative communication

ACM Reference Format:
Boyin Yang and Per Ola Kristensson. 2023. Designing, Developing, and Evaluating AI-driven Text Entry Systems for Augmentative and Alternative Communication Users and Researchers. In *25th International Conference on Mobile Human-Computer Interaction (MobileHCI '23 Companion), September 26–29, 2023, Athens, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3565066.3609738

1 INTRODUCTION

Non-speaking individuals with motor disabilities, such as people with motor neuron diseases (MNDs), rely on augmentative and alternative communication (AAC) devices to communicate. Literate users widely adopt keyboard-based AAC text entry systems equipped with information retrieval-based auto-complete and word prediction features that enable typing anything and having it spoken by the system speech synthesizer. For instance, users relying on eye typing are heaily reliant on word prediction [3]. However, AAC users typically exhibit a very low text entry rate on keyboard-based systems, often achieving less than 20 words per minute (WPM) [6], whereas speech rates in common conversations are around 150–200 WPM [2]. This stark contrast makes it challenging for AAC users to communicate freely with able-bodied individuals without extra assistance.

The advent of large language models (LLMs) has demonstrated substantial potential for enhancing AAC systems design [1, 7]. However, designing an effective, efficient and pleasant AI-driven AAC text entry system poses considerable challenges. For example, it is not yet clear how we can effectively exploit advanced AI techniques for accurate text prediction and generation in practical AAC settings. Despite the rapid development of natural language processing (NLP) technologies, interactive accessible controllers, and human-computer interaction theories, AAC technologies are still lagging behind the latest technology developments [8].

This is in part because AAC is a highly interdisciplinary domain, encompassing various stakeholders, including AAC users, AAC researchers, natural language processing (NLP) researchers, and human-computer interaction (HCI) researchers. Each stakeholder group faces unique challenges and perspectives within the AAC domain. The purpose of this research is to address and engage with these diverse perspectives. The research seeks not only to contribute to each stakeholder group individually but to also establish connections and bridge the gaps that exist between these groups, ultimately empowering AAC research as a whole.

2 RESEARCH QUESTIONS, METHODS, AND PROGRESS TO DATE

The research questions are listed in Table 1. RQ1 serves as the primary question in this research, encompassing the following reasons that link to the remaining research questions (with corresponding published works cited as indicators of current progress):

(1) AAC researchers' insights are not effectively translated into advanced AAC systems (RQ2 [10]).
(2) There is a limited availability of advanced AAC systems for both researchers and users (RQ5 [8]).

MobileHCI '23 Companion, September 26–29, 2023, Athens, Greece

ACM ISBN 978-1-4503-9924-1/23/09.
https://doi.org/10.1145/3565066.3609738

Number	Question
RQ 1	Why has the AAC domain witnessed great developments, while AAC technology has progressed slowly in terms of communication rate over the past few decades? What are the gaps in this domain that can be addressed through HCI research?
RQ 2	How can the design of an AAC system effectively engage AAC researchers and users to fulfill their complex requirements?
RQ 3	How can we parameterize and simulate the performance of the AAC system and user prior to development in order to facilitate the design at an early stage in the design process?
RQ 4	How can we develop a dedicated generative language model for AAC purposes to increase the text entry rate for AAC users? What are the positive and negative qualities of generative language models and information retrieval methods?
RQ 5	How do we best design, develop, and evaluate such an AAC system?
RQ 6	What are the key aspects influencing the acceptance of AI-generated sentences by AAC users and able-bodied users for communication purposes? In terms of AI-generated sentences, can experiments with able-bodied users be conducted to alleviate the challenges associated with conducting experiments with AAC users? If so, to what extent can this substitution be made?
RQ7	How do AAC users interact with an AI-driven AAC text entry system to construct sentences and how can they steer the system to generate their desired sentences?

Table 1: Research questions

(3) There are considerable challenges in conducting experiments with AAC users for AAC systems design, development, and evaluation (RQ3 [8], RQ5 [9], RQ6).
(4) There is inadequate integration of state-of-the-art AI technology into AAC (RQ4 [7], RQ6-7).

To answer these research questions, we use the following approaches, referencing the corresponding published works as indicators of current progress:

- RQ1–2: Literature review and engagement in discussions with experienced AAC researchers [10].
- RQ3: Integration of human performance factors into a keystroke-level model (KLM), and utilization of this new model along with a function structure model to conduct envelope analyses [9].
- RQ4: Development of a dedicated generative LLM for AAC purposes [7].
- RQ5: Translation of AAC requirements into system functions, resulting in a fully developed tinkerable AAC text entry system equipped with conventional and state-of-the-art NLP methods [8].
- RQ6: Plan to carry out studies on the impact of AI-generated sentences for communication for both able-bodied individuals and AAC users using a variety of methods, including experiments and semi-structured interviews with both groups using the tinkerable AAC system.
- RQ7: Plan to iteratively improve the tinkerable AAC system design for experiments once RQ6 is investigated and cooperate with an AAC company to develop and test a new system with their users, comparing it with their existing system.

To date, substantial progress has been made in addressing RQ1–5, with the exception of the user evaluation in RQ3. However, RQ6 and RQ7 still remain to be investigated.

3 KEY RESULTS AND REMAINING RESEARCH

3.1 Key Results from RQ1, RQ2, RQ4, and RQ5: Tinkerable AAC

We introduced the concept of tinkerable AAC (TAAC) and developed such a system. The TAAC system allows for modifications during interaction to accommodate various use scenarios and conditions, enabling users and researchers to achieve specific goals, such as gaining higher efficiency and analyzing users' text entry strategies [8, 10]. Figure 1 presents the conceptual framework explaining how the qualities of a TAAC system facilitate AAC research and development. Table 2 illustrates the relevant functions in the TAAC system that are aligned with stakeholders' attributes and requirements. The TAAC system integrates several AI language models, including KWickChat, a multi-turn language model [7] that utilizes conversation history and persona tags for sentence generation, and ChatGPT [5], one of the most powerful LLMs to date.

3.2 Key Results from RQ3: The Imperfect Surrogate User Model and Envelope Analysis

Designing and evaluating AAC systems is notoriously difficult due to the speech and motor challenges of AAC users. To address this, we introduced a conceptual AAC system for sentence generation [9] based on a computational design of a word prediction AAC system [4]. This model simulates the information flow during user interaction. Further, we proposed an imperfect surrogate user model that offers a more realistic simulation of user performance compared to traditional KLM. With these two models, several valuable insights are obtainable using a technique known as envelope analysis. For example, combining word and sentence predictions leads to a higher text entry rate than using only one prediction function, whereas the keystroke savings do not necessarily translate to a positive net entry rate, especially when considering the impact of human performance factors [9]. This work complements conventional qualitative user experience evaluation methods by

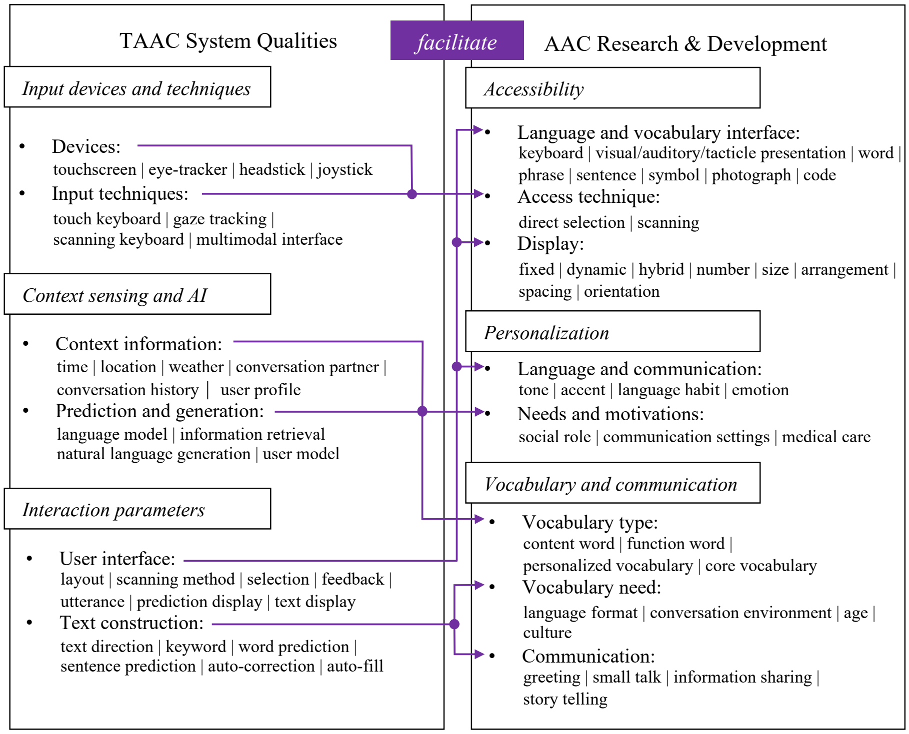

Figure 1: A conceptual framework for explaining how qualities of a TAAC system (left) facilitate AAC research and development (right). The purple arrows indicate pathways linking TAAC system qualities that can be used for investigating related areas in AAC research and development. There are many pathways, but we only show a few in the figure for clarity. [10]

	AAC User	AAC Researcher	Natural Language Processing (NLP) Researcher	Human-Computer Interaction (HCI) Researcher
Stakeholders' attributes and requirements	1. Unique motor capabilities 2. Developing physical conditions 3. Unique communication needs 4. Rely on text prediction functions to type	1. Have access to AAC users 2. Language threapists, psychologists, or other specialists who have deep understanding in the AAC domain and user needs 3. Use existing tools to conduct experiments	1. Investigate state-of-the-art NLP models for various downstream tasks 2. Develop dedicated language models for AAC purposes	Bridge stakeholders to enable the design and development of AAC systems for users, professionals and researchers
Relevant functions in the TAAC system	1. Text entry system with word and sentence prediction functions 2. Tinkerable keyboard layout 3. Tinkerable conversation modes with different interaction approaches	1. Multiple text input methods 2. Different text prediction algorithms for different scenarios	1. Tinkerable parameters for each prediction method. 2. The system is modualized for quick integration of up-to-date NLP models	Quantitative analysis functions for automatically tracking and evaluating human performance

Table 2: The stakeholders' attributes and requirements and the relevant TAAC system's functions. [8]

quantitatively evaluating the emerging impacts of human performance factors from a system at design-time, which is particularly important for AAC research, where it is frequently difficult to carry out conventional user studies with AAC users.

3.3 Remaining Research: RQ3, RQ6, and RQ7

We are currently conducting a pilot test to evaluate the TAAC system, which will be followed by real user tests as part of RQ3. Additionally, we are in the progress of preparing a questionnaire on AI-generated sentences for communication purposes, targeting both able-bodied individuals and AAC users (RQ6). Once the survey results are analyzed, we will proceed with experiments on the current TAAC text entry system and semi-structured interviews involving both user groups (RQ6). Further, the survey results from AAC users will be leveraged to enhance the sentence prediction function of the current TAAC text entry system (RQ7). Simultaneously, we will also collaborate with an AAC technology company to develop an AI-driven AAC system based on a symbol-based AAC system, enabling a more flexible and accurate sentence generation method (RQ7). Moreover, we are developing an LLM-based keyword prediction language model for keyword-based text entry, exploring this new sentence construction approach for AAC users, which we conjecture may dramatically increase text entry rate (RQ4 and RQ7).

REFERENCES

[1] Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Ajit Narayanan, Meredith Morris, and Michael Brenner. 2022. Context-Aware Abbreviation Expansion Using Large Language Models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 1261–1275. https://doi.org/10.18653/v1/2022.naacl-main.91

[2] Ann Copestake. 1997. Augmented and alternative NLP techniques for augmentative and alternative communication. In *Natural Language Processing for Communication Aids*.

[3] Per Ola Kristensson, Morten Mjelde, and Keith Vertanen. 2023. Understanding Adoption Barriers to Dwell-Free Eye-Typing: Design Implications from a Qualitative Deployment Study and Computational Simulations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 607–620. https://doi.org/10.1145/3581641.3584093

[4] Per Ola Kristensson and Thomas Müllners. 2021. Design and Analysis of Intelligent Text Entry Systems with Function Structure Models and Envelope Analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445566

[5] OpenAI. 2022. *Introducing ChatGPT*. OpenAI. Retrieved Aprol 6, 2023 from https://openai.com/blog/chatgpt

[6] Ondrej Polacek, Adam J Sporka, and Pavel Slavik. 2017. Text input for motor-impaired people. *Universal Access in the Information Society* 16 (2017), 51–72.

[7] Junxiao Shen, Boyin Yang, John J Dudley, and Per Ola Kristensson. 2022. KWickChat: A Multi-Turn Dialogue System for AAC Using Context-Aware Sentence Generation by Bag-of-Keywords. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 853–867. https://doi.org/10.1145/3490099.3511145

[8] Boyin Yang and Per Ola Kristensson. 2023. A Demonstration of a Tinkerable Augmentative and Alternative Communication Keyboard. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 138–140. https://doi.org/10.1145/3581754.3584153

[9] Boyin Yang and Per Ola Kristensson. 2023. Imperfect Surrogate Users: Understanding Performance Implications of Augmentative and Alternative Communication Systems through Bounded Rationality, Human Error, and Interruption Modeling. In *Proceedings of the 25th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '23)*. Association for Computing Machinery, New York, NY, USA, 33 pages. https://doi.org/10.1145/3604260

[10] Boyin Yang and Per Ola Kristensson. 2023. Tinkerable Augmentative and Alternative Communication for Users and Researchers. In *Design for Sustainable Inclusion*, Joy Goodman-Deane, Hua Dong, Ann Heylighen, Jonathan Lazar, and John Clarkson (Eds.). Springer International Publishing, Cham, 137–145.