ISSN 2186-7437

NII Shonan Meeting Report

No. 214

The Future of Education with AI

Andreas Dengel
Laurence Devillers
Andrew Vargo

March 4–7, 2024

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

The Future of Education with AI

Organizers:
Andreas Dengel (University of Kaiserslautern-Landau / DFKI)
Laurence Devillers (Sorbonne University / CNRS- LISN (LIMSI))
Andrew Vargo (Osaka Metropolitan University)

March 4–7, 2024

Abstract

AI and its effect on education is certainly an important and divisive research topic. From a public standpoint, since the launch of ChatGPT 3.5 in the fall of 2022, the paradigms of teaching, learning, and education have fundamentally shifted. This has led to a great amount of uncertainty, even within research communities. But the reality is that, beyond large language models (LLMs) and generative AI, there are a myriad of technologies that can alter and augment how humans approach education. Implementation of these technologies and their adoption by different parts of society do not happen in a vacuum, creating a challenge for both developing and implementing these technologies so that they do the most good, as well as accurately analyzing their cognitive, physical, perceptual, and societal impacts on humans. This meeting was held in order to discuss AI and Education, primarily through a Human-Computer Interaction (HCI) lens. It also served as a culmination of the Trilateral AI Learning Cyclotron (LeCycl) Project [1] supported by the Japan Science and Technology Agency (JST), France National Research Agency (ANR), and German Research Foundation (DFG).

[1]https://lecyl.org

Background

Artificial Intelligence (AI) can benefit many human endeavors and transform how humans interact and function in many different fields. One such application area where the introduction and entrenchment of AI can be transformative is in education. From primary, secondary, and tertiary schooling to independent adult learners, AI could be interwoven into different ecosystems to help students learn, master, and share knowledge more effectively and efficiently by acting as effective dialog partners and work assistants [8]. However, the implementation of AI in education is not conducted in a controlled or systematic manner but is rather occurring asymmetrically across diverse geographies and organizational structures with different policy initiatives [36]. This leaves researchers and system designers with the conundrum of what it means to successfully implement an AI aided ecosystem for learners.

The purpose of this seminar is to explore, discuss, and shape the future of AI within the field of education and learning with regards to its cognitive, physical, perceptual, and societal impacts on humans. Within the education and learning fields, deployment of augmentation tools and systems will likely have a direct impact on the social structures of humanity. We aim to bring experts from Human-Computer Interaction (HCI), Education, Cognitive Psychology, Public Policy, and Philosophy to explore the development of AI educational ecosystems. In particular, we will explore the following interconnected themes:

Achieving Individualized education through AI aided systems: One of the greatest potential uses of an AI education ecosystem would be the fully individualized systems that can tailor content and strategies to an individual's needs as opposed to a system that optimizes for a majority or plurality of individuals. Learners would then be able to save time and effort when studying by having the optimal routines prepared for them. This could be especially useful for at-risk learners [44]. The path to achieving a system that works well has some difficulties. Privacy and validation remain a challenge. This is especially true when considering in-the-wild usage. The type of training data that is needed to successfully build a system is difficult to collect and doing so may put participants in studies at disadvantages.

Encouraging knowledge transfer platforms: The transfer of knowledge from individuals to other individuals is difficult to manage on a large-scale platform. Typically, there are two types of platforms: those which do not consider information redundancy (the same things can be asked and learned repeatedly) and those which seek to create an information corpus (each exchange should be unique). In the former, the information exchange and creation is inefficient and difficult to search. In the latter, the restrictive nature of the exchange makes it difficult to encourage new entrants to the system. In both cases, the exchanges will almost certainly experience power law, where most content is created and maintained by relatively few users, regardless of the assumed incentive system [41]. An AI mediated system brings many opportunities and challenges for creating knowledge sharing platforms. At the same time, the incentive systems needed to make such systems work are unknown.

Trade-offs from AI use: While AI aided systems promise many benefits to its users, it is also necessary to consider if there are negative externalities where the technology may impact the cognitive or perceptual abilities of humans. An example is how pervasive availability of GPS maps on mobile devices and in automobiles have potentially hindered the wayfinding ability of individuals [17]. While the ability of an individual to quickly find an exact location has been enhanced through technology, the ability to navigate without the device has regressed. In a similar way, technologies that are introduced in education may enhance specific performance but may harm performance outside of the system structure.

Splintering infrastructures in education: Left on its own, and regardless of the intentions of researchers, the asymmetrical implementation of these systems is likely to result in uneven and potentially damaging outcomes for different types of participants [16]. The access to technology and requisite knowledge needed to harvest its benefits are naturally skewed to benefiting wealthy groups, and the systems themselves may have biases baked in which discriminate against different groups. A potential danger is that implementation results in splintering infrastructures, where certain groups continue to use mostly traditional methods of education, while groups with access to technologies experience a beneficial cycle where technologies and systems are continually updated and optimized for their specific needs.

Meeting Schedule

March 3:
15:00 - 19:00: Check-in
19:00 - 20:30: Welcome Banquet

March 4:
7:30 - 8:30: Breakfast
8:30 - 9:10: Introduction of Shonan Seminar
9:10 - 10:00: Seminar Welcome (Andrew Vargo and Koichi Kise)
10:00 - 10:30: PhD Student Lightning Talks
- Ko Watanabe
- Peter Neigel
10:30 - 11:00: Break
11:00 - 12:00: Lightning Talk Session 1
- Tilman Dingler
- Akiko Aizawa
- Yutaka Arakawa
12:00 - 14:00: Lunch and Group Photo
14:00 - 15:30: Keynote I: Andreas Dengel & Christoph Maerz
15:30 - 16:00: Break
16:00 - 16:50: Lightning Talk Session 2
- Ralph Rose
- Yasuyuki Sumi
- Benjamin Tag
16:50 - 18:00: Module 1: AI and Human Behavior
18:00 - 19:30: Dinner

March 5:
7:30 - 9:00: Breakfast
9:00 - 10:30: Keynote II: Laurence Devillers
10:30 - 11:00: Break
11:00 - 12:00: Lightning Talk Session 3
- Justine Cassell
- Eisaku Maeda
- Christoph Maerz
12:00 - 13:30: Lunch
14:00 - 18:15: Excursion to Kamakura
18:15 - 21:00: Main Banquet

March 6:
7:30 - 9:00: Breakfast
9:00 - 10:30: Lightning Talk Session 4
- Anisia Popescu
- Shigeo Matsubara
- Simon Knight
- Nicolas Großmann
- Leo Van Waveren
10:30 - 11:00: Break
11:00 - 12:00: Module 2: Access and AI
12:00 - 13:30: Lunch
13:30 - 15:00: Module 2: Access and AI
15:00 - 15:30: Module 2 Wrap-up
15:30 - 16:00: Break
16:00 - 18:00 Lightning Talk Session 5
- Judy Wang
- Motoi Iwata
- Chris Blakely
- Théo Deschamps-Berger
- Hannah Nolasco
18:00 - 19:30: Dinner

March 7:
7:30 - 9:00: Breakfast
9:00 - 10:30: Closing Session & Overflow Time
10:30 - 11:00: Break
11:00 - 12:00: Seminar Close (Andrew Vargo)
12:00 - 13:30: Lunch and End of Meeting

Overview Keynote Talks

The Pulse of Generative AI for Education

Andreas Dengel, University of Kaiserslautern-Landau / DFKI
Chrsitoph Merz, DFKI

Generative AI is increasingly central to educational technology, providing tools that can potentially transform teaching and learning processes. This keynote addressed the current landscape of generative AI in education, its transformative potential, and anticipated future developments. At the heart of the discussion was the role of large language models (LLMs) and foundational models in creating more dynamic and responsive learning environments. The presentation detailed the initiatives of the Foundation and Large Language Model Group at DFKI which aims to integrate LLMs into existing educational projects, developing a generalized multi-agent system based on LLMs to foster innovative educational solutions.

Several practical applications of generative AI in education were highlighted, ranging from the creation of customized learning materials and interactive tutoring systems to more efficient administration and support for educators through automated content generation and task management. For instance, generative AI can assist in creating texts, managing emails, and even programming, which streamlines workflows and frees up educators to focus on more creative and strategic educational roles.

A central and extensively discussed project during the presentation was the Generative Tutor Assistant. This innovative application embodies a generative avatar that can engage with users in multiple languages across a variety of topics. The queries posed to the avatar are processed by a sophisticated multi-agent system, which then generates a live video response. This capability enables the Generative Tutor Assistant to function effectively in diverse educational and professional set-tings. It can serve as a virtual teacher for pupils and students, providing interactive and personalized learning experiences. Additionally, it has practical applications in corporate environments, such as facilitating employee onboarding or offering general customer support through an accessible and informative avatar. This project exemplifies the versatile and impactful use of generative AI in enhancing interaction and support across different sectors.

The presentation also touched on the ethical considerations and challenges associated with deploying AI in educational settings, including data privacy, the potential for bias, and the importance of maintaining a human-centric approach in education.

From the ensuing discussions and QA session, it was clear that while enthusiasm for the potential of generative AI in education is high, there is also a cautious recognition of the need for careful implementation. Key takeaways included the importance of aligning AI technologies with educational goals to enhance learning outcomes, the necessity for robust privacy protections, and the ongoing need to critically assess the impact of AI on educational equity and accessibility. Looking forward, the integration of generative AI into education promises significant shifts in educational paradigms. However, it will require thoughtful oversight to ensure that these technologies are used to enhance educational outcomes and foster an inclusive and equitable learning environment.

Generative AI, Affective computing and Chatbot/Robot for Education : Ethical Issues

Laurence Devillers, Sorbonne University/ CNRS-LISN (LIMSI)

In this talk, I offered studies and reflections on the ethical issues of Generative AI, affective computing and Chatbot/Robot for Education. Conversational agents and social robots using autonomous learning systems and affective computing will change the game around ethics. We need to build long-term experimentation to survey Human-Machine Co-evolution and to build ethics by design chatbots and robots. In the chair of HUMAAINE (L. Devillers, CNRS), we aim to study the Human-Machine Affective interactions and relationships, in order to audit and measure the potential influence of intelligent and affective systems on human beings, and finally to go towards a conception of "ethical systems", by design or not and to propose evaluation measures. In this purpose, the planned scientific work focuses on the detection of social emotions in human voice, and on the study of audio and spoken language "nudges", intended to induce changes in the behavior of the human interlocutor. This work should be complemented by experimental studies (long- term, human vs. machine influence, etc.) to evaluate ethical aspects and confidence in AI, chatbot and robots, as well as by demystification of these technologies among the general public which naturally tends towards anthropomorphism. The importance of this subject also lies in the variety of its societal applications from care of the elderly and vulnerable people, to the economy, and to education.

Overview Lightning Talks: In Alphabetical Order

Lightning Talk 1

Akiko Aizawa, NII

Since the advent of ChatGPT, natural language processing has become a large-scale, accelerator-like science, even in Japan. The rapid development of large language models (LLMs) is often represented by an evolutionary tree, where one model is derived from another, and then various models that improve it appear, leading to an explosion of diverse models.

Simultaneously, several problems with LLMs have been noted since their emergence. Among them are that they have not yet been fully adapted to the Japanese language and that they need to be adapted to Japanese society. For this reason, an attempt has been made to construct an open and Japanese-competent LLM. As an extension of this effort, NII will establish a new LLM research and development center in April 2024. The goal of this center is to conduct research to ensure the transparency and reliability of LLMs.

Lightning Talk 2

Yutaka Arakawa, Kyushu University

I provided information about my research on Human Activity Recognition and Behavior Change Support Systems.

In addition to the SDGs and environmental considerations, there has recently been a movement to change existing behaviors, such as movement restrictions during the COVID-19 pandemic. For treating lifestyle-related diseases, changing habits is crucial. To prompt such human behavior changes, traditionally awareness-raising and counseling by people have been the mainstream approaches, but recently AI interventions using smartphones and wearable sensors have been expanding, and digital therapeutics for lifestyle diseases have also emerged.

Given this technological background, we discussed the future of AI and education. In the future, AI teachers may accompany and support students, taking into account their individual learning ability, motivation, and learning progress. In other words, personalized learning support that can even transform learners' attitudes is expected to advance.

With a generative AI like ChatGPT, it would answer questions and teach many things without a teacher present. However, it will also reply to wrong things, and students will not be able to acquire the ability to write reports if a report writing feature is available.

So, we need to pay attention to what abilities humans can acquire and, conversely, what they cannot acquire due to such support. Can a human who asks AI for everything be called a capable person who can use AI skillfully, or a helpless person who cannot do anything without AI?

The boundary is ambiguous, and a careful discussion is needed on how far AI should support us.

Lightning Talk 3

Chris Blakely, KCGI

Imparting knowledge and developing the skills of students to be prepared for future career opportunities is often seen as the ideal of higher education institutions. A significant value is often placed on literacy and familiarity with new technologies and cutting-edge applications. Generative AI (gen-AI) is one of the most recent examples of a new technology quickly coming to prominence and changing various domains and industries. With the increasing power and convenience of generative AI tools available, it remains uncertain if these new tools will undermine the need for such skilled, literate talent in the future. Nonetheless, current gen-AI tools are known to hallucinate data, fabricate sources or produce poor-quality results without knowledgeable, corrective human oversight [1, 18]. Despite this, gen-AI tools remain extremely powerful and ubiquitous causing many fields to require less human input or labor. This poses a unique and frightening prospect for the education of future professionals.

The view of these new gen-AI tools across different higher education institutions are inconsistent: Many institutions globally have not developed any clear policy or guidance on the use of AI tools [40]. Furthermore, these is no clear indication whether the use of AI tools should be promoted as part of the educational curriculum or restricted as a new form of academic misconduct. The lack of a universal policy approach across institutions to gen-AI is not the only concern; there is also the concern that different institutions adopting widely different policies and attitudes toward gen-AI can lead to divergence across educational institutions. For some schools, encouraging students to learn these tools and how to get good results from gen-AI is seen as inevitable and necessary for student development. Other schools respond to the use of gen-AI tools as a form of cheating, strictly forbidding students from learning how to use these tools well. Such a divergence could result in inequitable access to career opportunities across student populations. If this situation remains unchanged, there is the danger that only a small minority of students allowed to train and develop their skills with AI would be able to access future career opportunities. Such a divergence scenario could potentially cause an irreversible consolidation of access to knowledge and limit professional advancement to a small, elite minority at the expense of other students.

In pursuit of understanding such a possible threat and its likelihood, we began research into the attitudes and impressions toward gen-AI tools, specifically ChatGPT, by administering an online survey among faculty, students and administrative staff at a Japanese university. This initial survey asks respondents questions about their beliefs and attitudes toward ChatGPT, their awareness of institutional policies or unofficial practices, their AI usage habits and their views on the future and value of such tools in their respective fields. Through this ongoing study, we aspire to build a qualitative data set that can provide a clearer, more accurate picture of the state and impact of institutional policy-making and user practices with regards to gen-AI technologies.

One significant insight gained from this workshop has been the recommendation to incorporate an educational policy framework (i.e., the Digital Competence Framework for Educators) to ground our survey results on policy objectives for social advancement and educational literacy [14] The workshop also introduced additional sources on AI hallucination and the veracity of AI results to strengthen the literature review for this on-going study [45, 22].

Lightning Talk 4

Theo Deschamps-Berger, CRNS

I am a PhD candidate at the Laboratoire Interdisciplinaire des Sciences du Numérique (LISN) at CNRS, focusing on Audio Language Models and Large Language Models (LLM). My research involves developing deep learning models to recognize emotions in speech from multimodal sources, aiming to enhance human-computer interactions.

In my talk at the NII Shonan Meeting 214, I presented my research on speech emotion recognition (SER) for social interactions, specifically from *in-the-wild* emergency call center conversations. The goal is to provide emotional cues to agents, enhancing interactions between callers and agents. I discussed the inherent challenges in speech emotion recognition (SER), including the scarcity of *in-the-wild* emotional corpora for training robust models. Often, existing corpora are collected in controlled lab environments, which may not capture the variability and natural emotional expressions of the real world. This distinction is crucial as it affects the nature of the data and, consequently, the applicability and generalizability of the research results obtained from these corpora. The corpus used in my research is a collection of spoken telephone dialogues from a French emergency call center (CEMO) between callers and agents, and annotated with major and minor labels by two psycho-linguists from a list of 21 fine-grained emotions, [13]. This annotation scheme is supported by other works, such as the *push-pull* model from Scherer [35]. Emotions in social interactions often reflect social-cultural elements, resulting in mixed emotions. The *push-pull* theory, outlined by Scherer and illustrated by the Tripartite Emotion Expression and Perception (TEEP) model by Bänziger [5], distinguishes between distal information (internal state estimated by voice analysis) and proximal information (listeners' perception). The term "push" refers to natural, instinctive expressions, while "pull" covers expressions shaped by social norms and cultural expectations.

The core of my research is to specialize Audio Language Models and Large Language Models to recognize social emotions such as in the CEMO real-world conversations. It includes techniques of transfer learning, to effectively specialized unimodal pre-trained models [10, 9]. Other focuses include integrating dialog context from conversations into Audio Language Models and LLMs, as discussed in [12]. This approach integrates specific dialogue turns in addition to the speech turn to be classified as input to specialized language models. The embeddings produced by the encoders are filtered before classification to ensure they correspond to the speech turn being classified. The intuition is that the attention layers from the encoders weigh the speech turn embeddings considering their context, which includes embeddings from the surrounding speech turns.

I also presented a multimodal fusion technique called symmetric cross-attention, which leverages acoustic and textual information from unimodal specialized encoders. This fusion method proved beneficial for speech emotion recognition (SER) [11, 10]. The technique involves an alignment layer to join temporal information from both acoustic and textual encoders before fusion. Grounded on the multi-head mechanisms from [42], the symmetric cross-attention combines information from both modalities by applying attention mechanisms in parallel to acoustic and text data. This approach allows each modality to be contextualized by the other, effectively weighting the information. The final representations from both modalities are then merged for classification, enhancing the accuracy of speech emotion recognition.

During the Q&A session, several key takeaways emerged. Participants emphasized the importance of handling mixed emotions and suggested exploring multilabel learning strategies. Questions also centered on the integration of multimodal information, with discussions on the potential benefits of combining acoustic and linguistic features for more accurate emotion recognition. These interactions provided valuable feedback and highlighted areas for future research, such as refining the models to better handle overlapping emotional states and improving the interpretability of the predictions.

Lightning Talk 5

Tilman Dingler, TU Delft

Generative AI tools, such as Copilot, have been reported to make people more productive and creative, as well as help them save time. They increase personal productivity and lift up the organisation as a whole. 70% of Copilot users indicated to be more productive with such tools and 68% were under the impression that they improved the quality of their work [2].

Similarly, Noy and Zhang [27] conducted an online experiment with 444 college-educated professionals, in which they allowed half of their study group to use ChatGPT and compared the results to a group without access. They found that ChatGPT raised people's average productivity, decreased time efforts, improved output quality and even reduced inequality between workers, especially benefiting low-ability workers.

However, there is no such thing as a free lunch.

We also know that technologies affect our cognitive abilities, often not for the better. In a study by Storm and colleagues [37], participants using the Internet to retrieve information altered their propensity to use the Internet to retrieve other information, hence technology seems to make us more reliant on online information per sé. Technologies have also been shown to affect our relationships in unintended ways where the mere presence of a mobile phone can negatively impact the quality of our face-to-face interactions [28].

With the rise of large language models (LLM) and their invitation to outsource our thinking and writing to them, the question, therefore, becomes:

What do we loose when we resort to LLMs "to aid" us?

Cai and colleagues [6] presented a tool to inspire designers, yet Wadinambiarachchi et al. [43] showed that the use of generative AI in the design process can lead to 'design fixation' and hence limit the creative bandwidth. The process of creating–be that a design or text–is complex and the process itself a way of thinking. If we resort to LLMs to do that thinking for us, what do we lose when the process itself gets shortcircuited? And what is the meaning of putting

[2] /urlhttps://www.microsoft.com/en-us/worklab/work-trend-index/copilots-earliest-users-teach-us-about-generative-ai-at-work

down such information in the first place? Should we expect people to attach any value to the machine output that we prompted?

In "The Work of Art in the Age of Mechanical Reproduction," Walter Benjamin [4] argues that the authenticity of a work of art is not solely inherent in the piece itself, but also in the context in which it exists. When Benjamin was writing, the world was undergoing significant changes due to inventions like the phonograph, television, and radio. He expressed his concern about how technologies reshapes the way art is experienced. Before these technologies, the only way to experience a concert, speech, or play was to be physically present as it was being performed. Similarly, prior to photography and mass production, most visual art was directly crafted by the artist's hand. Benjamin points out that the immediacy of an experience is important highlighting that there is a profound difference between owning an original painting, where the artist's hand lead the actual brushstrokes, vs owning an no-matter-how perfect reproduction.

Coming back to LLMs and their relentless ability to produce text, images, and multimedia. Information, however, only exists to the degree that people can perceive, process, and understand it. It, therefore, only has meaning if it contributes to a human experience. Information should lead to some sort of learning, which prompts the question of what do we lose if our education entails passing our learning to an AI? When outsourcing parts of our our learning opportunities and creativity to a machine, do we recognise the cost?

LLMs are here to stay and as we grapple with their effective and ethical utilisation, we should also ask ourselves what the effects and unintended consequences of using generative AI tools may be on our, the creators', creativity and thinking, sense of ownership, achievement, skill and mastery, and last but not least, meaning itself.

Lightning Talk 6

Nicolas Großmann, DFKI

One of the biggest advantages of digital learning material in contrast to regular learning material is the ability to adapt to the needs of the learners and their individual skill levels. The adaptation of can be automatically done with the help of Artificial Intelligence but for an AI to fulfill this task the current learning state of the learners must be measured, quantified and analyzed.

In this talk we shared the DFKI approach and vision of quantified learning in the iQL - Immersive Quantified Learning Lab. This classroom size lab uses sensor technologies to measure the cognitive states of the learners to use this data with the help of AI and adapt the learning to the users. The lab mainly uses the eye tracking technology to follow the scan path of the eyes of the users which does not only give information about attention at a certain time but evaluating using the gaze behavior also allows to calculate information like interest, comprehension and self-confidence. Other sensor technologies like pen data or electro dermal activity are used standalone or complementary to give an even better insight into the cognitive states of the users.

A big help for fulfilling the iQL goal has now arrived with the rise of Large Language Models (LLM). These AI models allow the processing of texts and different operations on it. Functions like summarizing a text, searching and highlighting special information, translating and rephrasing entire paragraphs can now be easily done by sending a prompt to a LLM like ChatGPT. This means we can focus on cognitive state recognition and there is no need to program time-consuming functions by ourselves. New functions can also easily be implemented into the existing framework.

Lightning Talk 7

Motoi Iwata, Osaka Metropolitan University

My main research topics are Support systems for learning foreign languages, Comics analysis and processing, and Digital watermarking. For the support system for learning a foreign language, I am mainly engaged in the estimation of learner's cognitive state based on their behavior, where the target cognitive states are confidence, understanding, unknown words, and so on. I use learners' eye gaze, voice, posture, movement, heart rate, and so on for estimating the cognitive states. In this presentation, I will talk about mainly three methods of the support system for learning a foreign language, that is, Multimedia vocabulometer, Unknown words estimation on a smartphone app, and Unknown words estimation when reading comics aloud.

Multimedia vocabulometer is implemented as an app on smartphones. News articles and YouTube videos with English subtitles can be imported into this app as teaching materials. Users read or watch them and record unknown words manually. After that, they can relearn the unknown words with a flashcard with the sentences in the news article or the video including unknown words. Unknown words estimation on smartphone app is also implemented as an app on smartphones. This can generate teaching materials by GPT. Users can obtain the unknown words list just by reading the teaching materials. By introducing this method into Multimedia vocabulometer, we expect that the unknown words list can be automatically obtained just by using the app (this is our future work). Unknown word estimation when reading comics aloud is implemented as an app on a desktop PC. This is because this method requires the stable recording of eye gaze and speech. In Japanese manga, short texts are distributed in speech balloons on a page. Therefore, it is suitable to record the reading behavior of each speech balloon (short text) or between them. Moreover, Japanese manga is popular and helpful in keeping learners motivated.

Lightning Talk 8

Koichi Kise, Osaka Metropolitan University

I gave an overall introduction to the Learning Cyclotron (LeCycl) Project, which is a trilateral (France, Germany, Japan) AI project for 4.5 years. After showing the goal of the project, which is to accelerate the knowledge flow among humans, I introduced the three major steps for the knowledge flow: perceiving, mastering and transferring. Then I showed recent results for each of the steps such as mobile and desktop systems for English learning, desktop learning systems for programming, and nudging strategies as well as fundamental technologies for the above processes.

Lightning Talk 9

Simon Knight, University of Technology Sydney

Learning for ethical (dis)engagement with AI focused on the intersection of learning and AI, and the ethical implications thereof. The talk framed this both in terms of the importance of learning regarding ethics and AI generally (i.e., how we learn regarding ethics and AI), and in the specific context of AI in education. My use of the parenthetical (dis)engagement reflects two concerns:

1. that sometimes ethical practice should involve not engaging with particular technologies/activities (i.e., it isn't enough to assume we'll build a tool, and then consider how to do so ethically, we should question whether the tool should exist at all). 2. that disengagement can also have ethical implications, and there are times we may consider we have obligations to engage (ethically).

I am particularly concerned to with framing ethics and AI in terms of learning, and the kinds of material (learning) resources we draw on in practice. These include ethics guidelines, which have proliferated, but with very little research on their characteristics or structures to support learning. Indeed, in my review many documents noted as guidelines or principles documents in earlier reviews in fact provide very little guidance.

In this context, I framed ethics in three ways:

1. Ethics as an expression of values sees ethics in terms of the expression of principles or values that we act by. This view of ethics helps us because through shared principles and values we develop shared language to consider ethics. However, principles only get us so far, with challenges in navigating predicaments or dilemmas in action where principles come into tension.

2. I then discussed ethics as enacted, and the role that practical cases that apply principles to fine grain issues can help us learn, and become attuned to ethical issues. I noted that many examples of cases are weak in this regard, because they are often examples of clearcut moral harm, and not cases where some of the more nuanced concerns can be drawn out. This matters, if we want to learn about ethics, we actually have to talk about ethics. A particular example of this is the undervaluing of 'merit' (or worth) as an ethical principle - i.e., that it is unethical to implement systems for which there is no evidence, or poor evidence, because they are at best an opportunity cost.

3. I then highlighted that these issues arise in particular contexts, and that in this way we can see ethics as embedded. It is embedded in the material resources we work in, that exist in networks of norms and cultures. I pointed to some examples from our work of where taking this situated approach has helped us understand the particular nuanced concerns at play in our context.

Lightning Talk 10

Peter Neigel, Osaka Metropolitan Unviersity

After a short biographical introduction of the author, the author takes on the topic of stress detection in a university cohort [23]. In this presentation, we showed daily usage data of the Oura Ring by the OMU cohort for the time between August 2021 and November 2023 is presented. The data from this cohort consisting of university students is the basis for a stress analysis aiming to detect cohort-wide stress, i.e. without singling out individual participants. The presentation asks the question whether this is possible given that every participant shows individual baselines of physiological measurements and reacts differently to different stressors. But the homogeneity of the group - all students at Japanese university, most between 20-24 years old, therefore similar stressors – makes this analysis viable. The presentation then shows results for several mixed effects linear models on the group with different pre-defined periods as the fixed effect, the individual participants as a random effect and different physiological measurements as dependent variables. The results show with statistical significance after Bonferroni correction, that there are changes in Heart Rate, Heart Rate Variability and Daytime Maximum Heart Rate for the whole cohort during specific periods like exams or breaks. Since these physiological measurements are shown to be connected to stress in laboratory studies, the authors infer that a cohort-wide analysis in this fashion can be used to detect new group-wide stressors e.g. in cases of pandemics or other macroscopic events. The authors conclude that the use of wearable technology can reveal periods of cohort-wide stress with around 100 participants/devices even if the data is messy, but further investigation is needed to pinpoint the source of the stressful events. This could help in policy making at e.g. universities, where the policed group is relatively homogeneous.

Lightning Talk 11

Hannah Nolasco, Osaka Metropolitan Unviersity

Wearables are becoming increasingly accurate at capturing physiological health markers and are fast approaching a degree of precision comparable to medical grade sensors. The Oura Ring, a sleep monitoring device, is equivalent in fidelity to an electrocardiogram (ECG) in heart rate tracking [19] and can also reliably detect sleep stages [2]. Devices such as this can effectively assist in mitigating the cognitive stress of building and maintaining health-related habits [21] but only if the user appropriately adheres to the technology. 'Adhere' in this context refers to successful usage compliance at a level that enough personal data is collected to deliver insights on the user's behavior and wellbeing and these insights are impactful enough to mobilize them into changing their habits and attitudes.

In our previous research, we found that usage compliance is not actually difficult to achieve with most users: Based on a 1-month study that we conducted with the Oura Ring on a population of 31 Japanese adults between the ages of 20 and 50, device adherence with a satisfactory compliance rate (80 percent of sufficient usage in 4 out of every 5 days of the experiment) is possible regardless of age or gender [24]. However, in a follow-up study to determine if reported changes in lifestyle would reflect in the recorded data of a different cohort that used the ring for a period of 1 year, we discovered that all users did not exhibit any improvement in sleep quality over time or even had a slight deterioration in sleep quality as the months passed [25]. This demonstrates a failing in the device to modify behavior. It also suggests that it may only be providing users the impression that they are improving their habits for the betterment of their wellbeing.

One potential method to address this gap in perception of device benefit versus actual recorded data is to implement self-reflection into the process. Self-reflection is said to be an important and crucial step in personal informatics and wearable interaction as it is a great way to gain useful insight on the self [7]. Despite this, most existing wearables today do not have the necessary tools to facilitate reflection [20]. We tested out a brief 1-minute daily self-reflection prompt on university students who had been using the Oura Ring for some time, giving them a platform to log habits that encourage or discourage good sleep based on a list of behaviors known to impact sleep quality. They were also asked to evaluate the quality of their sleep and the accuracy of Oura's sleep and Readiness Scores. Our results showed that all users found the calculations of Oura to be either moderately or extremely inaccurate or they would express uncertainty ("I don't know") over it. Apart from this, users also did not exhibit any visible attempts to change any behaviors that could negatively impact their sleep quality: When they reported on any harmful activities, such as blue light exposure before bedtime, users would continue to do them for the entire two weeks of the experiment [26].

Self-efficacy—or whether one perceives themselves to be capable of executing a task—is one of the major factors affecting this lack of long-term substantial behavioral change [21]. Users with high agency tend to regard their wearable as a co-creator of meaning as opposed to an instrument that must direct their actions; users with low agency may tend to wait for the right answers. In the next stages of our research, we endeavor to see whether self-reflection is really the right solution to building self-efficacy in users or if other forms of delivering data may be more suitable to making wearables devices truly influential.

Lightning Talk 12

Shigeo Matsubara, Osaka University

I introduced my research on multi-agent systems research, focusing on two studies related to problem-solving through human-machine collaboration in dynamic environments to achieve large-scale collective intelligence. The first study involves collective prediction through human-machine collaboration [1]. We proposed an ensemble method for estimating the expected error of a machine forecast and dynamically determining the optimal number of humans included in the ensemble. We evaluated the proposed method using seven datasets on US inflation and confirmed the superiority of our method. The second study focuses on the realization of human-to-human coordination in the use of limited resources [2]. We proposed a novel distributed user-car matching method based on a contract between users to mitigate the imbalance problem between vehicle distribution and demand in free-floating car sharing. Our method overcomes the drawback of previous regulation methods, assuming that the system operator can obtain accurate data on origin-destination (OD) demand. In these studies, diversity is crucial for high performance in prediction and regulation tasks.

During the workshop, we engaged in a thought-provoking discussion on the transformative potential of AI in education. We explored its implications across various stages of learning, from perception to mastery and transfer. These discussions were enriched by the perspectives of diverse disciplines, including human-computer interaction (HCI), education, and cognitive psychology. While my research does not directly focus on learning/teaching support, I can offer a unique perspective: diversity. With advances in sensing technology, we now have the capability to capture physical behavior and emotions, which is expected to significantly contribute to our understanding of the importance of maintaining diversity of perception. However, many presentations seemed to concentrate on individual learning outcomes. Therefore, establishing mechanisms to recognize and embrace differences from others will be a crucial and inspiring challenge.

Lightning Talk 13

Ralph Rose, Waseda University

Hesitation phenomena (HP) encompass various linguistic and paralinguistic behaviors that can delay the transmission of a speaker's intended message to listeners. These behaviors include silent pauses, filled pauses (FP: e.g., uh/um, e-to/ano), repairs, repeats, and lengthenings. My research focuses on understanding their acoustic features, factors contributing to their occurrence, cross-linguistic variations, perceptual effects on listeners, and their use by second language (L2) learners during speech development. Key findings of mine include the observation that L2 speakers gradually move from native FP forms to L2 target forms [**?**]. Additionally, Japanese speakers are slower to 'fill' pauses than English speakers [34]. Furthermore, people use silent pauses more than filled pauses to parse language during both listening [29] and reading [30]. Acoustic features such as jitter and shimmer may be anticipatory indicators of disfluency [31]. My recent research interests revolve around automating the insertion of hesitation phenomena (HP) into AI-generated speech. One practical application is to direct a language learner's attention to specific elements in an AI tutor's speech, such as vocabulary items, syntactic structures, or pragmatic markers. Achieving this involves addressing several challenges, including identifying text-to-speech sources capable of authentic HP insertion with appropriate prosodic patterns. Additionally, determining the optimal locations or types of HP for maximum learning impact is crucial. An underlying ethical consideration is whether and how such HP-insertion aligns with ethical guidelines (cf., [32]). During the Shonan workshop, there were extensive discussions about AI-related ethical questions, including how educational institutions should address AI usage by faculty, staff, and students. Notably, transparency and accountability emerged as key themes. Through conversations with fellow participants, I found a solution to a longstanding challenge: How can we guide students to understand their role in a meaningful relationship with AI tools? Drawing inspiration from Japanese puppetry (Bunraku), I presented this perspective in-depth at a subsequent conference [33].

Lightning Talk 14

Yasayuki Sumi, Future University of Hokkaido

Yasuyuki Sumi presented several of his own past studies to discuss the impact of IT in education. First, he introduced his own research cases where the measurement and analysis of verbal and non-verbal behaviors in multiparty conversations (e.g., utterances, standing position, gestures, head movements, gaze, etc.) required the use of various special sensors 15 years ago, and how, thanks to the development of computer vision in recent years, similar measurement and analysis are now possible in everyday situations (e.g., tutoring conversations between students). He then introduced related projects such as Face Counter for measuring the amount of daily communication, a system that encourages the circulation of casual conversations in a workshop, and a game that encourages the use of the university library. The possibility of continuing to use such experimental systems in actual educational settings was discussed in terms of motivating participants and protecting their privacy.

Lightning Talk 15

Benjamin Tag, Monash University

Emotions manifest as subjective, physiological, and behavioural responses to everyday life's opportunities and challenges. These responses often arise unexpectedly due to uncontrollable external events. Recent advancements in ubiquitous technologies, Machine Learning, and Artificial Intelligence have introduced novel methods for quantifying, generating, and regulating emotions. These innovations open new avenues for emotion regulation, particularly in educational settings, where AI can provide personalized emotional support, enhancing learning outcomes. Better understanding and the ability to regulate emotions are crucial, especially in educational settings, as they can improve students' focus, motivation, and resilience. This can ultimately lead to better academic performance and overall well-being. However, there are significant concerns regarding the ethical implications, privacy risks, and potential for emotion manipulation associated with these technologies. This lightning talk explores (1) the challenges and opportunities in studying human emotions, (2) the impact of our digital habits on our emotional well-being, (3) the role new AI models may play in emotion regulation before discussing (5) the ethical risks posed by AI in terms of privacy and the potential for emotion manipulation.

The domain of Affective Computing has gained new relevance, driven by recent global events [39] and advances in Human-Computer Interaction (HCI) and Artificial Intelligence (AI) research. New devices and systems now allow us to study human activities and emotions passively, contactless, and in everyday contexts [38]. This has led to a deeper understanding of human behaviour and emotions, yet it has also sparked critical discussions about the efficacy of emotion detection technologies, as well as concerns about privacy and ethics. Despite promising results, many studies highlight the difficulties in using physiological signals to reliably detect emotions. Multimodal sensing approaches can offer a more robust foundation for emotion inference but often overlook the subjective and experiential nature of emotions.

The theory of constructed emotion suggests that the brain continuously generates emotional categories [3], challenging the notion of universal emotions. This theory underscores the complexity of sensing and quantifying emotions accurately. It posits that what we experience as emotion is primarily the felt emotional experience, not merely its physiological or behavioural components. Thus, physiological or behavioural signals may be detected without the presence of an actual emotion, making the subjective experience essential for confirming an emotion's existence. This makes a critical evaluation of the feasibility and accuracy of current emotion-sensing technologies crucial.

To achieve a higher control when studying emotions in naturalistic settings, we propose studying instances of Emotion Regulation, which, according to [15], describe the intentional, i.e., with a certain level of agency, use of strategies to change emotions.

Emotion regulation is a critical component in educational settings, influencing students' ability to learn, adapt, and thrive. Effective emotion regulation can help students manage stress, maintain focus, and stay motivated, which are essential for academic success. AI-driven emotion regulation tools promise to provide real-time feedback and personalized interventions to support students' emotional well-being. For instance, AI can monitor students' emotional states through voice analysis and physiological sensors, offering timely interventions such as mindfulness exercises, motivational messages, or adaptive learning tailored to individual needs.

Moreover, AI can help educators understand the emotional dynamics of their classrooms better, allowing for more empathetic and responsive teaching approaches. Emotionally intelligent AI systems can identify when a student is struggling and suggest appropriate support strategies, fostering a more inclusive and supportive learning environment.

While AI and emotion regulation offer promising opportunities, especially in educational contexts, they also raise significant ethical concerns. The potential for emotion manipulation, privacy breaches, and sensitive emotional data misuse are substantial risks. Emotion detection technologies can be intrusive, and the data collected can be highly personal, requiring strict privacy protections. There is also the risk of AI systems being used to manipulate emotions for commercial or political purposes, leading to ethical dilemmas around consent and autonomy.

It is crucial to establish clear guidelines and ethical frameworks for the development and deployment of affective computing technologies. This includes ensuring transparency in how emotional data is collected and used, obtaining informed consent from users, and implementing robust security measures to protect sensitive information. Balancing the benefits of AI in emotion regulation with the need to safeguard ethical standards and privacy is essential for the responsible advancement of affective computing.

Lightning Talk 16

Leo van Waveren, RPTU Kaiserlautern

Learners may fail to answer questions correctly despite having mastered the corresponding learning goals (slipping), or solve questions by chance (guessing) without attaining the intended learning processes. This means providing feedback to learners should not only rest on the evaluation of single items, but take the overall performance into account. The understanding of feedback is often reduced to a knowledgeable party providing an assessment on performance to a learning party. On the one hand, this view overlooks the internal processes of the feedback provider – i.e., comparing the observed performance to an (arbitrary) standard. On the other hand, the learning party has to compare their own performance to an internal standard. In addition, the outcome may then be used as a point of reference to include external feedback into considerations for adjustments of performance. Regardless of whether the feedback is provided by a human or an AI system, often the conveyed message is similar to "the answer is incorrect" (knowledge of performance) or "The correct answer was 18" (answer of correct result). Both forms are less helpful in fostering learning processes.

And while current Large Language Models may influence how teaching is conducted in the future, I propose that this decision needs to be a deliberate one. The RAT-Approach assumes that digital support measures can change how learning goals are pursued by superseding previous goals (replace), lead to an enhancement (amplify), or may change the overall process (transform). This assumption appears to rest on the assumption that the supporting party is trustworthy enough to attribute meaning to their feedback. For this reason, I argue that learning institutions need to provide a "safe space," not devoid of AI, but with controlled exposure, where learners can take supported steps in their learning journey. Encountering AI content with guardrails and enough context to judge the extent of its merits and drawbacks. To this end, the introduced RAT-Approach may need an adjustment in terms of sections where AI is consciously omitted from the picture and a regular learning process is kept in place (retain), even if tools in this area may exist, so relevant mastery can be acquired before the capabilities of AI are introduced.

Lightning Talk 17

Ko Watanabe, TU Kaiserslautern

Knowledge transfer is significant in discussing future education. It is the transmission of mastered knowledge from one individual to another through communication. This intricate process depends on three critical facets of communication: the appearance and demeanor of the participants, verbal articulation, and nonverbal cues. Several projects worked on will be to analyze the full potential of quantifying and enhancing communication.

In the workshop I mainly shared works on knowledge transfer activity recognition. In the DisCaaS project, we aim to quantify micro-behaviors that happen during meetings using cameras as sensors. In collaboration with a research group in Japan, we meticulously collected a dataset of 295 videos, totaling 21.7 hours, with 40 participants from online and onsite meetings. Remarkably, we achieved F1 scores of 0.812, 0.949, and 0.973 for nodding, talking, and smiling detection, respectively. The EnGauge project has been instrumental in quantifying engagement levels, a critical dimension of internal and cognitive human behavior. We collected data from 30 participants and achieved a result of engagement detection of 0.895 in the F1 score with leave-one-participant-out cross-validation. We also shared the system Metacognition-EnGauge. The system allows self-and-group engagement level feedback in gauge-interface real time. Several challenges exist in the accelerating knowledge transfer, which has been discussed.

After joining "The Future of Education with AI", a shonan-meeting workshop, I gained lots of in-sights especially from the perspective of psychologists and linguists. Our work is done mainly on the interest as a computer science researcher and hence feedback and discussions from different domain experts deepend the direction of the research. Related to that, I gained a chance to meet more researchers from abroad, which was a great opportunity to further collaborate and accelerate the field of future education AI.

Summary of discussions and findings

Throughout the meeting many new technologies and interaction scenarios were introduced and explored by the participants. The discussion sections centered around these technologies and interactions and how they fit into an educational ecosystem. While the breadth of topics was expansive, there were some prominent themes that emerged from the dialogues:

- HCI research needs to be at the forefront of investigating the potential impact on cognition that using AI technologies can have on users. While the adoption and proliferation of the technologies are iterative, patterns of usage will become prevalent. It is necessary to test and observe how these usage patterns affect people in order to inform policy. There is great risk in waiting to see if clearly evident negative effects are visible, thus allowing negative externalities to either proliferate too far or go completely undetected.

- Delivering technologies to education scenarios is difficult and needs support to find target audiences. While generative AI is popular and easily adopted in educational scenarios, it might not be the best or most effective tool. As we found from the discussion, there are numerous technologies that may be more effective in different contexts but never actually become usable by the general population (or specified target populations). Thus, while AI in educational spaces will be prevalent, it may only rely on the dominant marked choices and not the best choice for the task at hand.

- There is a risk of splintering with the use of AI in education based on social and cultural stratification. While pay-to-play models inherently benefit the wealthy when it comes to early adoption of AI technologies, equal access does not guarantee equal outcomes. The use of the same technologies in the same ways may be viewed differently based on social status and context.

- Policy is difficult to discuss because stakeholders are difficult to identify and address. The introduction of new technologies and interactions requires the creation of new guidelines and standards in many situations. In order to reach a working consensus on implementations, the inclusion of stakeholders is a necessity. However, the discussions uncovered how difficult this is due to the disruptive nature of these technologies and the constant shifting needs of stakeholders relevant to their contexts. This highlights the tension between the reactive nature of humans and the need to be proactive.

Identified issues and future directions

While there are many issues that came up in the 4-day workshop, the main problem that kept appearing was the difficulty in developing policy for educational ecosystems. It is difficult to understand the technologies, the technologies are not always accessible, the technologies are disruptive and can cause teachers and administrators extra work, and we are not sure who all the stakeholders are and what they want. There is also the reasonable fear that policy can hinder innovation.

In an ideal world, policy, while iterative, is proactively based on evidence. An HCI prospective can help us understand how humans interact with the technology and incorporate it into their education ecosystems. This requires a deep understanding of features and limitations of specific AI technologies and approaches with real human subjects, as well as a fundamental grasp of social and psychological aspect of interaction.

Based on this perspective, we feel that it is vital to bridge the gap between specific technologies and the theoretical aspects of their implementation. As such, the organizers of this workshop and many of the participants are producing a book on this topic for the Shonan Meeting Book Series.

List of Participants

- Akiko Aizawa, National Institute of Informatics, Japan
- Yutaka Arakawa, Kyushu University, Japan
- Chris Blakely, KCGI, Japan
- Justine Cassel, Carnegie-Mellon University, USA
- Andreas Dengel, RPTU Kaiserslautern / DFKI, Germany
- Theo Deschamps-Berger, University of Paris-Saclay, France
- Laurence Devillers, Sorbonne University/ CNRS-LISN (LIMSI), France
- Tilman Dingler, Delft University of Technology, Netherlands
- Nicolas Großmann, DFKI, Germany
- Motoi Iwata, Osaka Metropolitan University, Japan
- Koichi Kise, Osaka Metropolitan Unviersity, Japan
- Simon Knight, University of Technology Sydney, Australia
- Eisaku Maeda, Tokyo Denki University, Japan
- Christoph Maerz, DFKI, Germany
- Shigeo Matsubara, Osaka University, Japan
- Peter Neigel, Osaka Metropolitan University, Japan
- Hannah Nolasco, Osaka Metropolitan University, Japan
- Anisia Popescu, University of Paris-Sarclay, France
- Ralph Rose, Waseda University, Japan
- Yasuyuki Sumi, Future University Hokkaido, Japan
- Benjamin Tag, Monash University, Australia
- Leo Van Waveren, RPTU Kaiserslautern, Germany
- Andrew Vargo, Osaka Metropolitan University, Japan
- Judy Wang, Waseda University, Japan
- Ko Watanabe, TU Kaiserslautern, Germany

References

[1] Alkaissi, H., and McFarlane, S. I. Artificial hallucinations in chatgpt: implications in scientific writing. *Cureus 15*, 2 (2023).

[2] Altini, M., and Kinnunen, H. The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring. *Sensors 21*, 13 (June 2021), 4302.

[3] Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Social cognitive and affective neuroscience 12* (2017), 1–23.

[4] Benjamin, W. The work of art in the age of mechanical reproduction, 1936. *New York* (1935).

[5] Bänziger, T., Hosoya, G., and Scherer, K. Path Models of Vocal Emotion Communication. *PloS one 10* (2015).

[6] Cai, A., Rick, S. R., Heyman, J. L., Zhang, Y., Filipowicz, A., Hong, M., Klenk, M., and Malone, T. Designaid: Using generative ai and semantic diversity for design inspiration. In *Proceedings of The ACM Collective Intelligence Conference* (2023), pp. 1–11.

[7] Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto Ontario Canada, Apr. 2014), ACM, pp. 1143–1152.

[8] Dengel, A., Devillers, L., and Schaal, L. M. Augmented Human and Human-Machine Co-evolution: Efficiency and Ethics. In *Reflections on Artificial Intelligence for Humanity*, B. Braunschweig and M. Ghallab, Eds., Lecture Notes in Computer Science. Springer International Publishing, Cham, 2021, pp. 203–227.

[9] Deschamps-Berger, T., Lamel, L., and Devillers, L. End-to-End Speech Emotion Recognition: Challenges of Real-Life Emergency Call Centers Data Recordings. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)* (2021).

[10] Deschamps-Berger, T., Lamel, L., and Devillers, L. Investigating Transformer Encoders and Fusion Strategies for Speech Emotion Recognition in Emergency Call Center Conversations. In *24th ACM International Conference on Multimodal Interaction (ICMI)* (2022).

[11] Deschamps-Berger, T., Lamel, L., and Devillers, L. Exploring Attention Mechanisms for Multimodal Emotion Recognition in an Emergency Call Center Corpus. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2023), pp. 1–5.

[12] Deschamps-Berger, T., Lamel, L., and Devillers, L. Multiscale Contextual Learning for Speech Emotion Recognition in Emergency Call Center Conversations. In *International Conference on Multimodal Interaction (ICMI)* (2023), ACM, pp. 337–343.

[13] Devillers, L., Vidrascu, L., and Lamel, L. Challenges in real-life emotion annotation and machine learning based detection. *Neural networks : the official journal of the International Neural Network Society 18* (2005), 407–22.

[14] DigCompEdu. The european framework for digital competence of educators (digcompedu), 2024.

[15] Gross, J. J. Emotion regulation: Current status and future prospects. *Psychological Inquiry 26* (2015), 1–26.

[16] Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., and Koedinger, K. R. Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education 32*, 3 (Sept. 2022), 504–526.

[17] Ishikawa, T., Fujiwara, H., Imai, O., and Okabe, A. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology 28*, 1 (Mar. 2008), 74–82.

[18] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. Survey of hallucination in natural language generation. *ACM Computing Surveys 55*, 12 (2023), 1–38.

[19] Kinnunen, H., Rantanen, A., Konttä, T., and Koskimäki, H. Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. *Physiological Measurement 41*, 4 (Apr. 2020), 04NT01.

[20] Li, I., Dey, A. K., and Forlizzi, J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing* (Beijing China, Sept. 2011), ACM, pp. 405–414.

[21] Lin, F.-r., and Windasari, N. A. Continued use of wearables for wellbeing with a cultural probe. *The Service Industries Journal 39*, 15-16 (Dec. 2019), 1140–1166.

[22] Munn, L., Magee, L., and Arora, V. Truth machines: synthesizing veracity in ai language models. *AI & society* (2023), 1–15.

[23] Neigel, P., Vargo, A., Tag, B., and Kise, K. Using wearables to unobtrusively identify periods of stress in a real university environment. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers* (New York, NY, USA, 2024), ISWC '24, Association for Computing Machinery, p. 17–24.

[24] Nolasco, H. R., Vargo, A., Bohley, N., Brinkhaus, C., and Kise, K. Examining Participant Adherence with Wearables in an In-the-Wild Setting. *Sensors 23*, 14 (July 2023), 6479.

[25] Nolasco, H. R., Vargo, A., Komatsu, Y., Iwata, M., and Kise, K. Perception Versus Reality: How User Self-reflections Compare to Actual Data. In *Human-Computer Interaction – INTERACT 2023*, J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, and M. Winckler, Eds., vol. 14144. Springer Nature Switzerland, Cham, 2023, pp. 665–674. Series Title: Lecture Notes in Computer Science.

[26] Nolasco, H. R., Vargo, A., Moreeuw, M., Hara, T., and Kise, K. Augmenting Sleep Behavior with a Wearable: Can Self-Reflection Help? In *Proceedings of the Augmented Humans International Conference 2024* (Melbourne VIC Australia, Apr. 2024), ACM, pp. 278–281.

[27] Noy, S., and Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science 381*, 6654 (2023), 187–192.

[28] Przybylski, A. K., and Weinstein, N. Can you connect with me now? how the presence of mobile communication technology influences face-to-face conversation quality. *Journal of Social and Personal Relationships 30*, 3 (2013), 237–246.

[29] Rose, R. L. The structural signaling effect of silent and filled pauses. In *The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019)* (Budapest, Hungary, September 2019), R. Rose and R. Eklund, Eds., pp. 19–22. ELTE Eötvös Loránd University; September 12-13, 2019.

[30] Rose, R. L. The structural signalling effect of filled pauses during reading. Poster presentation at CUNY Human Sentence Processing Conference, Mar. 2019.

[31] Rose, R. L. Variation in jitter, shimmer, and intensity of filled pauses and their contexts in native and nonnative speech. In *The 10th Workshop on Disfluency in Spontaneous Speech (DiSS 2021)* (St. Denis, France, August 2021), R. L. Rose and R. Eklund, Eds., pp. 59–64. Université Paris VIII Vincennes; August 25-26, 2021.

[32] Rose, R. L. Ethical considerations in the use of disfluencies in ai-generated speech. *Proceedings of the 15th International Conference on Computer Supported Education - Volume 2* (4 2023), 128–135.

[33] Rose, R. L. Constructing policies on use of ai tools in language education: Considering chatgpt as a collaborator. Oral presentation at 58th RELC International Conference, Mar. 2024.

[34] Rose, R. L., and Watanabe, M. A crosslinguistic corpus study of silent and filled pauses: When do speakers use filled pauses to fill pauses? In *Proceedings of the 19th International Congress of Phonetic Sciences* (Canberra, Australia, 8 2019), S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds., Australasian Speech Science and Technology Association Inc., pp. 2615–2619.

[35] Scherer, K. R. Emotion in Action, Interaction, Music, and Speech. In *Language, Music, and the Brain: A Mysterious Relationship*, M. A. Arbib, Ed. The MIT Press, 2013, p. 0.

[36] Schiff, D. Education for AI, not AI for Education: The Role of Education and Ethics in National AI Policy Strategies. *International Journal of Artificial Intelligence in Education 32*, 3 (Sept. 2022), 527–563.

[37] Storm, B. C., Stone, S. M., and Benjamin, A. S. Using the internet to access information inflates future use of the internet to access other information. *Memory 25*, 6 (2017), 717–723.

[38] Tag, B., Sarsenbayeva, Z., Cox, A. L., Wadley, G., Goncalves, J., and Kostakos, V. Emotion trajectories in smartphone use: Towards recognizing emotion regulation in-the-wild. *International Journal of Human Computer Studies 166* (2022).

[39] Tag, B., van Berkel, N., Vargo, A. W., Sarsenbayeva, Z., Colasante, T., Wadley, G., Webber, S., Smith, W., Koval, P., Hollenstein, T., Goncalves, J., and Kostakos, V. Impact of the global pandemic upon young people's use of technology for emotion regulation. *Computers in Human Behavior Reports 6* (2022), 100192.

[40] UNESCO. Less than 10% of schools and universities have formal guidance on ai, 2023.

[41] Vargo, A. W., and Matsubara, S. Editing Unfit Questions in Q&A. In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (2016), pp. 107–112.

[42] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems 30: Annual conference on neural information processing systems 2017, december 4-9, 2017, long beach, CA, USA* (2017), I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., pp. 5998–6008.

[43] Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., and Velloso, E. The effects of generative ai on design fixation and divergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024), pp. 1–18.

[44] Yang, S. J. H., Ogata, H., Matsui, T., and Chen, N.-S. Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence 2* (Jan. 2021), 100008.

[45] Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., et al. Siren's song in the ai ocean: a survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219* (2023).