IICCSSS 2024 - Contributed talks

Book of abstracts

Lost in Translation: The Effect of Linguistic Closure on Consciousness and Qualia Studies

Zeynep Beste Coşkun

This paper introduces "linguistic closure," a subset of cognitive closure, suggesting that linguistic constraints significantly contribute to the perceived opacity of subjective experience, leading to the concept of qualia. While recognizing a link between subjective experiences and neural states, I argue that language limitations obstruct our understanding of this relationship, by creating a gap between empirical research and the experience itself which is known as the explanatory gap (Levine, 1983) and leads some to believe subjective experience is unanalyzable or non-relational. By responding to Metzinger's (2003) model of consciousness and qualia, and incorporating perspectives from Constructed Emotion Theory (Barrett, 2016; Lindquist et al., 2015) and cognitive linguistics (Lakoff, 1987), I propose that subjective experiences are fluid, continuous and not structural like language and conceptual apparatus which causes the mismatch. This view suggests that linguistic structures can distort our understanding, affecting empirical research, therapeutic interventions, and public discourse. To solve the problem I suggest studying this limitation deeply. A more comprehensive approach to the study of consciousness that incorporates a brain-language interface could provide a stronger foundation for philosophical inquiries into consciousness as well as its relation to artificial intelligence. Without a solid understanding of consciousness and its components, we risk being misled by potential claims that AI systems possess human-like consciousness. To effectively navigate this future, we must solidify our understanding of consciousness to address potential disinformation and to prepare for the ethical and philosophical challenges posed by advances in humanistic AI

Dogs or gods? Perception of AI risk - a terror-management perspective

Daniel Friedrich

My participants were asked to think about death, opportunity or comfort. Then, I measured their attitudes towards AI existential risk, empathy and implicit philosophy of mind. Thoughts on death and opportunity decreased AI risk concern for men, but increased AI concern for women. Empathy and attribution of mind-like qualities to AI both independently positively correlated with AI concern. People who were reminded of death and people with higher empathy also reported higher concern. I hypothesize these factors can be explained by how vividly people imagined the inner world of an AI and the moral weight of a catastrophe.

Detecting brain state changes during naturalistic free-viewing

Aruna Nyssanbay

Naturalistic viewing (NV) engages a broad range of brain regions during the processing of visual scenes. How the brain reconfigures during NV and how these changes reflect the dynamics of brain network connectivity is not well understood. We hypothesized that brain regions would shift their

central state configuration depending on the visual input. To address this hypothesis, we acquired functional magnetic resonance imaging (fMRI) data on 19 healthy human subjects while we presented 5-minute movie clips containing ego-motion and face scenes. Assessment of brain activation revealed a distinctive pattern for ego-centric motion scenes engaging the dorsal cortical network, with areas active in the frontal eyes fields, lateral intraparietal areas and the parahippocampal place area. Additional regions, including the precuneus and the motor cerebellum, were not seen in previous macaque studies using the same movies. In contrast, face scenes activated the fusiform face area, the superior temporal sulcus, the ventral striatum and the lateral orbitofrontal cortex. Next, we conducted functional connectivity analyses and identified the optimal parameters for constructing static and dynamic connectivity matrices. Our findings and analyses move forward the use of NV paradigms for capturing brain state changes and will contribute to a better understanding of brain network dynamics.

A Platform for Interactive Classification with Decision Trees

Daria Radcenko

State-of-the-art artificial intelligence uses black-box models, which are opaque and do not allow for traceable decision paths. While decision trees underperform on most data unless used in ensembles, using those ensembles lowers the interpretability of the model. Single-tree models do not have that downside so their transparency and interpretability enable users to incorporate their knowledge. This helps to mitigate the model's flaws and help users better understand them. Recent work finds that attention on decision trees in this context has been increasing, but interactively learning them has yet to be explored. We develop and study the first platform for interactively learning decision trees that users can assist and whose classification results and interpretability they can improve through various types of interaction. Thereby, we investigate whether there is a benefit in users assisting the algorithm by using more expressive feature types to improve model interpretability, performance, and size. To evaluate its usefulness, we have users interactively classify weather data and provide feedback regarding its effectiveness via questionnaire. In that context, we also compare the size and performance of the model before and after human intervention. The results indicate a potential to leverage human knowledge for increasing model performance and interpretability.

Enhancing Human-AI Interaction with Probabilistic Behavior Prediction for Situational Autonomy Adaptation

Ainur Raviqoh

This thesis explores the enhancement of human-AI interactions through the integration of probabilistic behavior prediction models to enable situational autonomy adaptation. The primary objective is to develop AI systems capable of anticipating human actions and intentions by analyzing contextual data, historical interactions, and environmental factors. By employing advanced machine learning techniques, the AI can generate probabilistic forecasts of user behavior, allowing it to adjust its level of autonomy dynamically based on the predicted scenarios.

In practical applications such as autonomous vehicles, healthcare, and collaborative robotics, this adaptive approach ensures that AI systems respond appropriately to varying levels of human input and decision-making needs. For instance, an autonomous vehicle might assume full control in routine driving conditions but defer to the driver during unexpected or complex situations, thereby enhancing safety and user trust.

Furthermore, this research addresses the balance between AI autonomy and human control, ensuring that the system remains transparent and respects user preferences. Ethical considerations, including privacy and the mitigation of unintended biases in behavior prediction, are integral to the development process. Ultimately, this thesis aims to create more intuitive, reliable, and user-centric AI systems that foster seamless collaboration and improve overall interaction quality between humans and artificial

intelligence.

Analyzing Speaker Variability in Conversational Data

Ali Gharaee

This talk presents an ongoing study focused on examining speaker variability within conversational datasets. We utilize advanced natural language processing techniques to explore how linguistic behavior differs across individual interactions. By analyzing the variability in responses to various prompts, we aim to uncover patterns that indicate the diversity and dynamics of dialogue. Our methodology includes identifying conversations with the highest and lowest variability, providing insights into both dynamic and stable interaction patterns. The findings from this study can inform applications in dialogue systems, communication strategy development, and understanding social dynamics in conversational contexts. While the project is still in progress, initial results highlight significant variability in speaker responses, which could have important implications for the study of human communication.

Study Project: Automatic Music Transcription

Deniz Gün

Automatic Music Transcription (AMT) is the process of transforming musical audio into notated musical symbols. Research into AMT using signal-processing methods dates back to the 1970s. While progress in music transcription with classical signal-processing methods has seemingly plateaued, the advent of neural networks has sparked renewed interest in the field. Notably, the conceptualization of AMT as a sequence-to-sequence task, an area where transformers have demonstrated exceptional performance, appears promising. However, the field of AMT still faces several open challenges, many of which we believe can be effectively addressed within the framework of cognitive science. In our talk, we will briefly present the challenges we are working to address and discuss future avenues that we find exciting.

Evaluation of Polish-German neural machine translation models with heterogeneous corpus data

Victoria Pierz

This thesis presents a comparative analysis of machine translation models by DeepL and Huggingface using a set of standardized metrics. Machine translation between Polish and German is evaluated, in particular the Hamburg Corpus of Polish in Germany (HamCoPoliG), as well as books for language learners, used as a gold standard. An automatic evaluation was performed, calculating the BLEU score, precision, recall, F1 score and accuracy. As the HamCoPoliG is a monolingual corpus, a proper gold standard was not provided. A forwards and backwards translation was performed, consequently it was decided to conduct a manual evaluation as well. For the manual evaluation, Polish native and heritage speakers were requested to rate the translation quality by grammar, readability and flow, while for the automatic evaluation the translated output was compared to the original transcriptions. Comparing the scores from the automatic evaluations, the scores were similar throughout the different translations, whereas for the manual evaluation major differences were visible. Overall, DeepL showed higher scores and a better ranking. For a more differentiated evaluation of machine translation on semi-spontaneous data, a proper gold standard and a larger dataset need to be acquired