Research Statement

As an aspiring researcher, I am deeply driven by a passion for understanding the intricate mechanisms underlying human language processing and how artificial intelligence endeavours to emulate such processes. This fascination was sparked by my previous projects where I compared the performance of XGLM (multilingual autoregressive language model) and GPT2 models on multilingual dataset and explored methods to enhance language model performance. Further, I delved into multilingual representation spaces, seeking insights into how these models process and represent language across different languages and layers. Another project focuses on training GloVe embeddings from scratch, developing a sentiment classifier, analyzing and evaluating performance across various domains, offering insights into their robustness and generalizability. Building on my previous experiences in analyzing language model behaviour, I am drawn towards exploring the discrepancies observed between the predictive capabilities of large language models (LLMs) like GPT-3.5 and human behaviour in tasks such as reading times and neural responses.

While the advancements in LLMs have showcased remarkable accuracy in language modelling, there are open questions regarding how these models align with human language processing. One way to assess this aspect is to compare language models to human behaviour e.g. reading times or neural responses such as ERPs. Specifically, while mid-sized models like GPT-2 demonstrate proficiency in predicting reading times, larger models like GPT-3 or GPT-4 exhibit poor performance in this regard. This could be either due to a divergence between language processing in humans vs. artificial language models or how we process reading times. This observation prompts critical questions about the underlying mechanisms of language processing and motivates my proposed research endeavour.

In light of recent findings, such as those presented by Liu et al. (2024), highlighting the challenges posed by multi-token words in predicting reading times accurately, I propose a comprehensive re-analysis of reading times considering the subword tokenization utilized by LLMs. By recalculating reading times based on subword units and possibly exploring alternative tokenization techniques grounded in morphological analysis, I aim to mitigate the artificial complexities introduced by long words in traditional word-level analyses.

This research endeavour will entail leveraging eye-tracking corpora to scrutinize LLMs' predictions of human reading times, thus discerning whether these models align more closely with human processing when assessed on a subword-token level. Moreover, we seek to elucidate whether discrepancies observed in word-level analyses are attributable to inherent artefacts or the inadequacies of current methodologies in capturing the nuances of language processing, particularly concerning long words.

By undertaking this investigation, I aspire to contribute novel insights into the comparative analysis of LLMs and human language processing, potentially paving the way for advancements in both artificial intelligence and cognitive science. Through meticulous experimentation and rigorous analysis, this thesis endeavours to unravel the intricacies of language processing, thereby enriching our understanding of the capabilities and limitations of contemporary language models in mimicking human cognition.