Purpose, Background, and Significance

Within the world of the mind, for most everyone, there are two main representations of information, verbal and nonverbal (Paivio 1969). Within the world of the artificial mind (i.e. the Large-Language Model - LLM), in opposition, there is typically only one representation, semantic sequences: a representation that most closely maps to that of the mind's verbal representation (Sutskever et al. 2014). In recent years developments have been made through latent diffusion (and similar methods) to generate visual representations of information (Rombach et al. 2022). Similarly, methods like CLIP now exist to convert image representations into semantic sequences (Radford et al. 2021). Most commercial LLMs are built upon the transformer framework which directly takes an input or output semantic sequence, creates an embedding, and then performs a series of steps through several neural network layers to generate further sequences or outputs (Vaswani et al. 2017). Notably, these models do not include any nonverbal (i.e. nonsemantic) representations of information.

This project proposes to bridge this gap by developing an integration of the aforementioned technology to allow for models compatible with dual-coding by introducing virtual mental imagery: a modified transformer framework which includes latent diffusion to encode information visually (i.e. through images or their mathematical representations) as well as verbally (i.e. through traditional semantic methods). This is an important step towards human-like artificial intelligence rather than the encyclopedic and unrealistic AI of the present day (Herbold et al. 2023). Notably, very little research has been done regarding the integration of nonsemantic representations of information with state-of-the-art models. There has been a framework proposed using nonsequential generation of language through latent diffusion, but it does not separate the semantic and the nonsemantic representations (Lovelace et al. 2024). Success within this area could lead to vast improvement of models towards human-like behavior.

Objectives and Methods

The key objective of this project is to create an Artificial Intelligence model framework for multi-modal internal representation (i.e. verbal and nonverbal or semantic and nonsemantic). This will involve testing and creating frameworks containing only pre-trained models, models fine-tuned on representing information abstractly, and from-scratch models integrating each of the portions described above. These models will be tested against each other and against similar existing baselines to ensure their compatibility in the existing world. This framework should allow the creation of a LLM which performs as well or better than the transformer in at least BLEU (Papineni et al. 2002), but also the human-like comparison metric, Mauve (Pillutla et al, 2021).

The key objective will be attained through a systematic, hands-on process. First and foremost a direct implementation of Vaswani et al., Sutskever et al., and Lovelace et al. will be created for baseline metrics on a single machine. These will be direct implementations, trained for the same amount of time, on the same machine, and tested through benchmarks as described and referenced in Chapter 2 of Maslej et al. (2024). Additionally these will be compared to the results given within the same chapter on state-of-art models to find a point of comparison between a shorter training period, and a commercial one.

From this point the second of the goals will take place: piecing together the three novel models. The novel models (especially the from-scratch model) will be directly compared

against Vaswani et al.'s benchmarks to show potential improvement, but critically equal performance. With these benchmarks the models will then be tested on more subjective tasks similar to those described in Herbold et al. This methodology of testing will help determine whether the model performance is more human-like than existing models, and if so, to what degree.

Finally, a manuscript will be prepared describing the performance of the framework, its components, as well as its creation process. These three steps will take place over the duration of the award period. The baseline model creation will be allocated 2 weeks to allow for sufficient training time. The creation of the unique framework will be allocated 6 weeks: 1 week for the pre-trained existing version, 2 weeks for the fine-tuned version (to allow for some training time), and 3 weeks for the from-scratch implementation. The final 2 weeks will be dedicated to manuscript preparation and wrap-up of the project.
Additionally, as many of the allocated weeks will involve periods of training, there will be some flexible time between each of the segments where the next segment can begin earlier.

While the key objective is hands-on, it will be critical to continue to stay abreast of the ever evolving research in the area as the project continues. A new state-of-the-art framework could be released during the project, in which case it is critical to test it as well against the project's framework. By modularly adapting to the ever evolving world of artificial intelligence the best possible results will be achieved. This is an important area to dedicate at least half a day each week.

There are many potential challenges related to this project. The largest is that the novel model does not perform as well as the transformer at baseline. One possible way to tend to this challenge is to examine how the information is represented. Weighting the representations between semantic and nonsemantic will be essential as it is impossible to know (even in humans) which is more important. The next possible challenge is that the models cannot be trained in sufficient time (commercial models train for hundreds of thousands of compute hours). The solution to this problem is to compare our results against the baseline of other models, not their prime, state-of-the-art examples. This does, of course, lead to a question of scalability, but if the model can be compared against a baseline, it can also be compared against itself after further training. More challenges will ultimately arise as the project continues so continuing to review the literature for similar scenarios is critical and some solutions may already exist.

This project's resources will be fairly minimal. The Subjectivity Lab has access to a powerful enough GPU for baseline metrics and fine-tuning. Additionally the lab provides the perfect space to work on the project throughout the semester (an assigned desk and a large HD monitor). No other financial burdens are expected of the project beyond wages and paper submission fees. For any expert advice outside of members of the Subjectivity Lab (which will provide mentorship and a lab discussion group), the lab can work with Dr. Peter Bex or Dr. MiYoung Kwon in the Psychology Department's Vision Science group and plans to contact Dr. Paul Hand of Khoury College and the College of Science.
This is not a group project and does not involve human or animal subjects.

Outcomes, Evaluation, and Dissemination

The primary benchmarks used for testing the created framework will be those referenced and described in Maslej et al. 2024 Chapter 2. By taking a direct comparison with the state-of-the-art it will be possible to evaluate the status of the project from the very

beginning. Additionally weekly mentorship meetings will take place to assess progress on the project to create adjustments if necessary (e.g. how to correct course if the framework is not performing to the same level as the transformer baseline). With the combination of direct feedback and explicit benchmarks progress in the project will be quantifiable and therefore correctable in the event things go awry. The expected outcome of the research project is a model framework as well as a comparison report between the framework and the other existing models. This will consist of performance in key benchmarks, the modular composition of the model (i.e. a similar report to Vaswani et al.), and future considerations of improvements or changes that the same or a different research group could improve upon. This outcome will be evaluated by examining to what degree this model improves upon (or stays at the same level) the two main metrics: benchmarks versus baseline, and comparisons versus open source, publically available human data (Maslej et al. 2024, Herbold et al. 2023, Pillutla et al 2002).

The aim of the project is to present a manuscript for publication at the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025 (publication due date of Nov. 14th, 2024). Additionally, presentations of the work at other conferences (beyond CVPR and RISE) in a poster or talk style are being considered depending on dates that align with the end of the project.

About the Learner

The preparation for the project has been extensive. It is the culmination of several years worth of theorizing better ways to approach the current problems (at least in my viewpoint) that exist within the AI industry: namely, the cost and the scalability. The most recent approach towards building better models has been to throw thousands of GPUs at the wall and hope for improved performance. While this has led to marginal improvement in the past year, improvement appears to have slowed down relative to the amount of money entering the space. As such, I have been theorizing different ways to take a new approach towards AI rather than continue down the same path. During my previous co-op in the Subjectivity Lab I was lucky enough to be exposed towards research around aphantasia, a condition associated with the lack of mental imagery. This exposure led me to question the relevance of such a capability in AI models (which upon reflection seems necessary in my own life almost everywhere). I conceived and developed the framework behind this project as a mixture of my own interests in AI and the work being done in the Subjectivity Lab in the Department of Psychology. I see this project as a gateway into graduate studies. This is the culmination of what I have worked on as an undergraduate and I want to take on a challenging project before entering a doctoral program where I will continue to develop and work on more projects like this. For these reasons I will have full responsibility for decision making and project progress. As far as mentorship is concerned, I plan on meeting weekly with my mentor as well as discussing issues with the other members of the Subjectivity lab, with at least one or two presentations during the semester's weekly lab meetings. To me this project is less of the Summit of undergraduate research, but a journey along the ridge towards doctoral studies.