Research Statement

I am an ordinary person, so my research story does not start from my childhood, it commenced much later, precisely when I felt it's imperative. I started thinking about my research story that inspires others and secures me a job. Many ideas came and past, and I worked on many, some I got published some I pursued and I got some jobs to pay off my bills. Yes, I am stretching it. Admittedly, I may be stretching it a bit.

Primarily, I worked in machine learning and formally started my research in machine learning with a project on neuromorphic devices. It allowed me to understand artificial neural networks as it was inspired the other way around from earlier strategies for artificial intelligence (AI) where the inspiration was to mimic the working of biological neural networks.

We worked to mimic the human brain by creating a neuromorphic device that keeps on learning. An ideal choice is generative modeling (it was not a buzzword at that time), and we used restricted Boltzmann machines (RBMs). Although it was learning, the memristor properties used in devising the device were not ideal, so the difference in accuracy was significant. But it helped us to achieve primitive speech recognition with competitive performance. However, the concepts of generative modeling stuck with us. It gave us a chance to apply the RBM with foreground and background segmentation in videos.

Meanwhile, another opportunity opened up, and we began working on object detection, particularly license plate detection in videos and images. Keeping our wish for generative modeling alive, we utilized the Region-based Convolutional Neural Network (R-CNN), which is considered an object detection model. However, generating likely bounding boxes for the objects presents a genuine generative modeling problem, and the solution worked much better than conventional license plate detection solutions.

This concept extended into a full-scale autonomous vehicle development project, focusing on perception and planning for autonomous vehicles. The goal was to understand the 3D environment from its 2D representation. Again, this posed a generative modeling problem, and we started labeling 2D images with 3D ground truth construction. The environment presented challenges as we lacked the leverage of a panoptic setup in the open, leading us to utilize Lidar. However, frames captured by the Lidar lacked 3D perspective, so we devised techniques using a sequence of frames. Ultimately, a demonstration was only possible by fusing camera, Lidar, and radar data for vehicle motion planning.

Simultaneously, a privacy challenge emerged during open data collection. Preserving the anonymity of people and cars in the data became critical. Thus, we turned to generative modeling once again for de-identification. We employed RBM to generate faces, and opportunities arose for the use of generative adversarial networks (GAN).

Despite some diversions during the early post-PhD settlement, ideas sprouted and materialized for various projects, publications, and the integration of new techniques in machine learning and AI. Another relevant idea related to generative modeling involved scene understanding in images. This provided us with an opportunity to comprehend and apply transformers as an alternative to sequential learning models like Recurrent Neural Networks (RNN) and their variants such as LSTM, GRU, etc.

After gaining this exposure, we re-examined our self-driving approach, initially reliant on a fusion of extremely expensive sensors, particularly Lidar. We contemplated transitioning to using only camera perception and planning. Our inspiration stemmed from the dominant role of human vision in decision-making, especially in psychomotor learning and development. Commencing with the understanding of the environment, we delved into vehicle localization in scenarios lacking prior location information. This posed yet another generative modeling problem.

Beyond the applied problems lies an inherent concern for understanding machine learning algorithms and their alignment with human objectives, a persistent consideration in my subconscious. The alignment challenge encompasses both technical and social dimensions. For me, the social aspect of alignment is intrinsic to the technical alignment of algorithms. My comprehension stems from observing the autoregressive development process within the ecosystem, reflecting the knowledge's adherence to the principle of evolution. The symbiotic relationship between perception and planning initiates a perpetual cycle of knowledge acquisition and its practical application. Present learning systems excel in one-shot perception but lack in certain areas. Many influential researchers share similar thoughts, ideas that might have embedded themselves in my mind through discussions and exchanges.

The delineation of cognitive processes in Daniel Kahneman's System-1 and System-2 concepts, revered by numerous machine learning researchers, has piqued my interest. I have contemplated incorporating System-2 functionalities into learning processes. While our machine learning algorithms surpass human-level performance in System-1 tasks, they lack the nuanced capabilities of System-2. The emergence of large language models (LLMs) and their understanding of human-generated language harbors an inherent potential to emulate System-2 functions through prompts. Despite LLMs heavily relying on self-supervised learning, the incorporation of prompts reintroduces a supervised-learning paradigm.

The challenge with System-2 in machine learning, compared to humans, lies in the finite interaction between the two and the disconnect in communicating thoughts. Human-machine interaction typically involves brief, task-specific interactions, unlike humans, the inherently connected intelligent beings. An experiment we designed using today's machine learning paradigm attempts to predict hierarchies by expanding and reducing the search space in predictions.

Another related yet slightly divergent challenge involves knowledge transfer—a skill at which humans excel. Numerous discussions highlight how swiftly children effortlessly transfer knowledge across domains. This phenomenon underpins the establishment of foundation models, later fine-tuned for specific tasks through a deliberate process. Mapping System-2 to conscious human behavior seems fitting, given the pivotal role of consciousness in decision-making.

However, does consciousness solely direct learning, or is it an input to the learning process?

I believe exploring the avenue of knowledge transfer within learning context should delve into the unconscious. The unconscious mind might serve as the enabler, providing room for imagination based on prior conscious inputs. Various techniques like prompts and reinforcement learning are already in use, but consciously harnessing the unconscious learning ability in machines is an avenue I eagerly anticipate exploring. I perceive it as a means to comprehend problems concerning artificial general intelligence (AGI) and their solutions. Additionally, integrating System-1 and System-2 with causal reasoning should intertwine conscious deliberation with unconscious input mapping.

I've realized that it's hardly just single science—it's interconnected so it asserts my admiration for polymaths.