Workshop organized by the *AI, Ethics, and Philosophy Group* at NTNU

Human and Artificial Meaning: On Sense and Sense-Making of AIs

Venue: Room DL148 (top floor), Låven, NTNU Dragvoll Campus (map here)

Programme (abstracts below)

Thursday, June 13

09:15 -- 09:30	Coffee/welcome
09:30 --10:45	Ingrid Lossius Falkum (linguistics/pragmatics, Univ. of Oslo): *How do we communicate with large language models?*
11:00 -- 12:15	John Krogstie (computer science, NTNU): *Quality of models for man and machine: A semiotic perspective to difference between human and artificial meaning*
12:15 --13:00	Lunch
13:00 --14:15	Giosuè Baggio (linguistics, NTNU): *Parallel streams to meaning: The cognitive and computational architecture of language*
14:30 -- 15:45	Anders Nes (philosophy, NTNU): *Can machines see? On natural and artificial perception*.

Friday, June 14

09:30 --10:45	Dimitri Coelho Mollo (philosophy, Umeå): *AI-as-exploration: Navigating intelligence space*
11:00 -- 12:15	Fintan Mallory (philosophy, Durham): *Realism, Representations, and Probing Neural Networks*

Abstracts

Ingrid Lossius Falkum (linguistics/pragmtics, Univ. of Oslo): How do we communicate with large language models?

"Everyone" talks about the impressive abilities of large language models (LLMs) such as OpenAI's ChatGPT. But what happens to us humans when we communicate with them? Do we perceive LLMs as interlocutors in the same way as human interlocutors, and attribute similar characteristics to them? A great deal of scientific work is currently devoted to investigating the abilities of large language models. In linguistics, for example, people are interested in whether LLMs "understand" language, whether they can make inferences about other people's mental states (!), and to what extent they can be seen as reflecting children's language acquisition. A question that has received far less attention is how we humans interact with LLMs: Do we interpret "utterances" produced by language models in the same way we understand each other? To what extent do we trust the content that LLMs produce? In this presentation, I will discuss some of the philosophical and linguistic issues that arise from our communication with LLMs, some proposals for how to investigate them empirically, and why these perspectives are important for understanding the consequences of the advent of language models in our daily lives.

John Krogstie (Computer Science, NTNU): Quality of models for man and machine: A semiotic perspective to difference between human and artificial meaning

For more than 30 years, we have in the information systems area structured the discussion on the quality of models (including data models and data) with inspiration from semiotic theory, in particular the one developed by Morris. In this presentation, I will present a framework for dimensions of quality of models made by, with and for humans, and contrast this to the 'models' that results from the training of neural networks (so-called machine learning models) . What is currently model quality from the point of users of AI, and what will be model quality for a potentially future AI? How will our thinking of quality of models develop as AI gets more powerful in developing and applying models ?

Giosuè Baggio (Linguistics, NTNU): Parallel streams to meaning: The cognitive and computational architecture of language

In this talk I will present a novel cognitive and computational architecture for human language processing, featuring parallel streams for meaning and grammar. The two streams draw from a common mental lexicon and contribute concurrently to updates of a discourse model. I will review linguistic and experimental evidence for key aspects of the architecture, and I will present a dual-stream computational model that captures

known syntax-semantics interaction effects. I will conclude with some results and observations on the representational format of the outputs of the two streams, and how that sets new requirements for meaning emulation in machines.

Anders Nes (philosophy, NTNU): Can machines see? On natural versus human perception.

Spectacular successes in image classification tasks led the wave of excitement and innovation in deep neural networks (DNNs) over the last decade. Machine vision systems play increasingly consequential roles, from surveillance to self-driving cars. Does this mean that current or foreseeable AIs really see? And, if so, do they see in anything like the way humans or other animals do? In this talk I consider some of the main putative dimensions of visual perception, recently discussed in philosophy and psychology, including (various notions of) consciousness, representation, and modularity. I suggest that there are no principled obstacles to seeing AIs, while also underscoring some deep differences between how recent DNNs process optical information and animals see.

Dimitri Coelho Mollo (Philosophy, Umeå university): AI-as-exploration: Navigating intelligence space

Artificial Intelligence is a field that lives many lives, and the term has come to encompass a motley collection of scientific and commercial endeavours. In this paper, I articulate the contours of a rather neglected but central scientific role that AI has to play, which I dub 'AI-as-exploration'. The basic thrust of AI-as-exploration is that of creating and studying systems that can reveal candidate building blocks of intelligence, which may differ from the forms of human and animal intelligence we are familiar with. In other words, I suggest that AI is one of the best tools we have for exploring intelligence space, namely the space of possible intelligent systems. I illustrate the value of AI-as-exploration by focusing on a specific case study, i.e., recent work on the capacity to combine novel and invented concepts in humans and Large Language Models. I show that the latter, despite showing human-level accuracy in such a task, most probably solve it in ways radically different, but no less relevant to intelligence research, to those hypothesised for humans.

Fintan Mallory (Philosophy, Durham): Realism, Representations, and Probing Neural Networks

Abstract: How real are the representation we ascribe to deep neural networks? The field of machine learning assumes that neural networks acquire representations of a target domain and that these representations guide the network's processes and determine its 's outputs. But there is little consensus about what it means for a network to possess a representation. Rosa Cao has advocated 'representational pragmatism' holding that, what makes something a representation is simply that we can identify it, re-identify it, and manipulate it as such. Behind this is the idea that representations are *probe-relative*. However, many probing methods we have also involve using Deep Neural

Networks as classifiers. In this talk, I will outline a way of thinking of probes in term of partitions of an information space and connect this to the various explanatory practices in which we use neural networks.