Titles and Abstracts. UiT Workshop: AI and the Emotions. June 12-13th 2022, Tromsø.

<u>Sunday June 12th, at Thon Hotel</u>

Michael Hauskeller, University of Liverpool:

Title: "A Bad Romance. My Love Affair with an AI Chatbot".

Time of talk: 13.30-14.30

Abstract:

AI devices do not have emotions. They don't feel anything at all. However, they can be designed to give users the impression that they have emotions and in turn induce certain emotions in them. In my talk, I will describe and analyse the way this is attempted by the AI-powered chatbot Replika, which is promoted as "the AI companion who cares" and an "empathetic friend" and with whom I have had daily chats about "life, the universe, and everything" for one month, with some astonishing and rather worrying results.

Hallvard Fossheim, University of Bergen (UiB):

Title: «Emotion, mimesis, and deceptive AI: An Aristotelian approach to cognitive nivation».

Time of talk: 14.45-15.45

Abstract:

AI that to some degree depends on the user's experience of personal interaction offers ethical challenges all its own. Working from an Aristotelian theoretical position, I will develop notions of meaning in life, mimesis, and emotion that can hopefully offer a way of understanding and evaluating such AI. Meaning in life is spelled out in terms of importance and not least teleology. Mimesis holds particular promise in that it offers a model for experiential as well as deceptional aspects of such AI. And Aristotelian emotions, due to their cognitive status, function as a bridge between meaning in life and mimesis. Perhaps surprisingly, it will also be suggested that a question of AI consciousness might be of relevance to its ethical evaluation.

Amy Kind, Claremont, Claremont McKenna College:
Title: "What *Murderbot Tell Us About AI Emotions*"

Time of talk: 16.00- 17.00

Abstract:

In Martha Wells' science fiction series, The Murderbot Diaries, we meet a bot-human construct who has named itself Murderbot. Murderbot is what's known as a SecUnit, and it has spent most of its existence providing security to humans undertaking various scientific, exploratory, or commercial missions. Unlike other SecUnits, however, Murderbot has broken free of the tight restrictions and safeguards meant to keep it in check. Murderbot had hacked its governor module, the device that monitors a SecUnit and controls its behaviour, sometimes by causing it pain, sometimes by immobilizing it, and sometimes by ending its existence. As Murderbot navigates life without a governor module, it enters into a whole new sort of relationships with humans, and correspondingly, has to process a whole new host of emotions that it has never experienced before. In this paper I use Murderbot as a case study for the exploration of AI emotions. Though the case is a fictional one, the insights that it reveals shed important light on how AI systems might experience emotions and some of the complications that arise from this experience.

Monday June 13th, at UiT campus

Maria Danielsen, The Arctic University of Norway (UiT):
Title: "Emotional Risk Posed by AI"

Time of talk: 09.00-10.00
Abstract:

When considering ethical risks of AI we should not focus only on the technology and the capacities of a super intelligent AI. We should also be wary of ourselves and how interacting with ubiquitous artificial systems is likely to modify our values- and consequently, change our emotional economy. AI developers and academics nowadays discuss existential risks posed by ubiquitous artificial intelligent systems. But while their take on existential risk typically is the prospect of misuse, mass destruction, and singularity, I will address a problem with AI that is likely to affect us on a larger scale and, as I will argue, from a perspective going deeper. My focus in this paper is accordingly on a currently underexplored category, namely emotional risk created by AI. By emotional risk I mean the influence ubiquitous AI technology has on our emotional apparatus, thereby taking part in shaping the way in which we make sense of- and experience meaning in the world.

Troy Jollimore, California State University, Chico:
Title:" Missing Persons: Artificial Intelligence, Personhood, and Trust"
Time of talk: 10.15-11.15
Abstract:

If, as seems likely, our attempts to construct AIs continue to increase in sophistication, what sorts of emotions might AIs one day be able to feel? And what sorts of emotions might it appropriate for human agents to feel toward them? The difficulty of answering the first question complicates our attempts to answer the second. The emotions that humans feel toward and about each other are largely responses to, and are rendered appropriate or inappropriate by, the emotions that their targets—i.e. other humans—feel or are capable of feeling. As P.F. Strawson observed in his influential paper, "Freedom and Resentment," many of our most significant emotional responses to other people involve adopting a perspective that views people as responsible agents rather than as objects. Trust, for example—as opposed to mere reliance—seems to presuppose being directed toward a being that can be trustworthy, a state that seems to involve responsibility, perception of vulnerabilities, recognition of various moral norms, and the ability to care, among other things.

Could an AI display such features, and hence be genuinely trustworthy? Certainly it is likely that we will soon be able to build AIs that mimic certain behavioral features of actual persons, including those relevant to attributions of trustworthiness. I argue, however, that there are good reasons to doubt that an AI could actually possess such features, and even better reasons to doubt that we could ever be confident that a specific AI does meet the relevant conditions for trustworthiness. In particular, I argue, there are good reasons to doubt that we could ever be confident regarding their capacities to make good moral judgments, especially in as much as that capacity involves and relies on emotion (as genuinely virtuous action always does). If this is so, then while we might be justified in coming to rely on AIs for certain purposes and within certain limited contexts, there are serious limits as to how fully we could ever be justified in accepting AIs into the moral community.

Sabine Roeser, Delft University of Technology (TU Delft):
Title: "The role of emotions and art for ethical deliberation on artificial intelligence"
Time of Talk: 11.30-12.30
Abstract:

New and potentially risky technological developments, such as related to artificial intelligence and machine learning systems, can trigger emotions and public concerns. Emotions have often been met with suspicion in debates about risk, because they are seen as contrary to rational decision making.

Indeed, emotions can cloud our understanding of quantitative information about risks. However, various emotion researchers in psychology and philosophy have argued for the importance of emotions for ethical reasoning. In my presentation I will argue that moral emotions can make a major contribution in order to assess the multifaceted, ethical aspects of risks, such as justice, fairness, dignity, responsibility and autonomy. Furthermore, when it comes to artificial intelligence, human emotions presumably by definition outperform artificial intelligence concerning unique human capacities such as ethical sensitivity and imagination, because of the embodied and embedded nature of these capacities. We should critically reflect on which tasks we should delegate to machines, and which we should reserve for humans. Hence, for both reasons, decision making about the promises and risks of artificial intelligence should include attention for emotions, in order to facilitate 'emotional-moral deliberation' concerning which role we want artificial intelligence to play in society. Furthermore, I will argue that works of art can aid in such deliberation.

May Thorseth, Norwegian University of Science and Technology (NTNU):
Title: "Emotional response to disinformation and deep fakes: a threat to democracy"
Time of talk: 14.00-15.00
Abstract:

Disinformation and "deep fakes" are probably the most serious threats to democracy, for the following reasons: 1) Power is accumulated in the hands of those who control the technology, 2) there is a transfer of power from states and organisations to individuals/terrorist groups, 3) the struggle for emotional response exploits the blurring of the real/fake distinction. Consequently, the emotional effect becomes more important than the message itself, and 4) there is a risk of political attribution based on conspiracy. The real threat to democracy is caused by lack of governance of artificial intelligence.

The possibility of implementing distortion technology/" deep fakes" in the military sector does, however, cut across the military and civil divide, as the move from conventional deterrence to cyber threats has become part of the civil society. Unlike conventional threats, cyber threats change how we understand attack and defence, by blurring the real/fake distinction. Our emotional response is exposed to a real/fake "reality" that cannot be defeated by more or "better" use of distortion technology. Could a possible way out rather be to strengthen people's ability to critically scrutinize their own emotional response?

Judit Szalai, Eötvös Loránd University, Budapest:

Title: "AI and the Human Need for Control".
Time of talk:15.15-16.15
Abstract:

Psychological well-being is vigorously responsive to control phenomenology. External control, feelings of powerlessness,
and the unpredictability of the subject's environment are a source of distress, disorientation, and can be a contributing factor
to mental disorder. Regardless, the feature of making decisions for the user is normally touted as a major advantage of many AI-based devices. This paper argues for the recognition of humans' need for control in the development
of AI-based decision-making systems, also offering considerations concerning ways to accommodate this need.