AI Ethics Lab

Meeting notes 29 Sep 2021

The purpose of the Swedish AI Ethics Lab is to provide guidance and support in implementing ethics in AI development, and the initiative will be piloted during the course of 2021. On the 29th of September 2021, the AI Ethics Lab convened for the second time.

The members that have been appointed as part of this group

Anna Nordell Westling, Sana Labs
Daniel Akenine, Microsoft
Evelina Anttila, Peltarion
Helena Thybell, Save the Children
Katarina Gidlund, Mid Sweden University
Louise Callenberg, PublicInsight
Magnus Boman, KTH Royal Institute of Technology and Karolinska Institutet
Martin Engström, Region Halland
Sara Övreby, Google
Stefan Larsson, Lund University
Theodor Andersson, Agency for Digital Government

Anna Nordell Westling and Sara Övreby were not able to participate in the meeting on the 29th of September.

Content of this document

This document reflects the strategic insights from the group discussion that will help direct future efforts in supporting the implementation of ethical AI. T

The discussion was based on a presentation by Magnus Sahlgren, RISE and AI Sweden, on the challenges and opportunities with applying large language models in Swedish.

Strategic key insights

The following insights were derived from the group discussions, which used the presentation and use case mentioned above as a starting point.

The focus for the AI Ethics Lab discussions is not on how ethical issues are or should be regulated, but on which ethical aspects and additional perspectives that should be considered in the implementation of the AI solution.

Challenges
There are several challenges related to applying large language models, such as

- **Energy use and climate change**
    - Developing and running large language models demand huge resources, and will inevitably impact our energy consumption. Moreover, it begs the questions of who has access to such resources. The current system for research funding incentivizes developing new, state-of-the art and unique language models instead of reusing existing ones, even when this would be a good solution requiring far less energy resources.
- **Representation**
    - There is a built-in optimization of language models that is tilted towards the global West, and white, male-dominated research and developing teams with similar disciplinary backgrounds risk reinforcing this. Although there is a great gain in developing Swedish language model, given that the global trend is focusing on larger languages such as English or Chinese, it is important to note that there will be issues of representation also within this context, not least regarding dialects and minority languages.
- **Responsibility**
    - Given the wide variety of risks associated with implementing large language models, ranging from climate change aspects in running the models, to issues on democracy and fair use of the AI solutions - the importance of the age-old question about responsibility is reinforced.

Need to take potential gains into account when balancing risks
When considering the ethical aspects of an AI solution, risks such as bias or a lack of fairness are often highlighted. These risks are real and should be given appropriate attention. However, it is also important to weigh the risks, and how they can be mitigated, against the potential gains from developing or implementing a certain AI solution. In the case of large language models, we know that the models risk not being representative for example due to a lack of focus on minority languages. Yet there is also a risk connected to *not* implementing the AI solution, given that the potential upside of doing so is great. The potential gains therefore need to be analyzed and clearly stated in order to be taken into account.