AI Ethics Lab

Meeting notes 20 May 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 20th of May 2021, the AI Ethics Lab convened for the first 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

Content of this document

This document contains a summary of the first meeting, reflecting both the presentations by external speakers and the learnings and insights from the group discussions.

The document does not include case-specific recommendations or technical learnings. Going into the work with the AI Ethics Lab, the aim was to produce very concrete learnings for specific case studies. After the first meeting is concluded, it is instead assumed that the AI Ethics Lab at this pilot stage will primarily produce strategic insights on how to direct future efforts in supporting the implementation of ethical AI.

Summary of presentations by external speakers

Göran Sundin, the Swedish Tax Agency
Sustainable AI to enhance trust
Trust is a key factor in collecting tax revenue and minimizing the tax deficit. The Swedish Tax Agency has a history of enjoying a high level of trust. Going forward and using AI, maintaining and even enhancing this trust is crucial. Sustainable AI is therefore one of the guiding principles of the AI policy adopted by the Swedish Tax Agency this spring. The

principle is now being concretized in “Guidelines for sustainable AI”, focusing on responsibility, transparency and explicability. The ambition is to appoint an ethical council and to provide the organization with tools and best practices to ensure sustainable AI. The key questions to be discussed center around *what is appropriate to do*, not if it can be done within current legal frameworks, nor if it is feasible on a technical level.

Johan Larsson Hörkén, Recorded Future
The project has created two sentiment models for narrow definitions of fear and violence respectively, both in Swedish. Such sentiment models can be used for instance in opinion analysis, to monitor inappropriate language/behaviors in forums and social media, or used by the police to detect escalating threat levels.

Strategic key insights

The following strategic insights were derived from the group discussions, which used the presentation and use case described 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.

Purpose is fundamental but not everything
The purpose and intended use of the AI solution is a strong guiding factor as to how to decide on ethical issues. Tracking young people’s use of gaming apps might be laudable if the intention is to alert the system to overuse, but not if the end goal is to encourage addiction. However, in reality it is not always feasible to constantly keep the purpose in mind during the development process when multiple smaller, but important, ethical-oriented decisions are made by a larger team.

A multi-disciplinary approach is needed
Diverse teams are often mentioned as a solution to discovering unwanted bias that a more homogenous team might not capture or test for. However, the discussions here showed that although diversity based on personal attributes is important, such as gender balance in AI developing teams, we need to shift our focus to disciplinary diversity. Bringing in for example behavioral social scientists or sociologists in AI development will have a direct effect. Not only for understanding of how algorithms may produce undesirable results, but for a wider understanding of the effect on society as a whole. What are the behavioral shifts triggered by the use of algorithm-based solutions by new users, target groups or in a new context?

Knowledge gap between management and developers
In order to make strategic decisions about how and when to use AI, ethics need to be considered and well understood. Consequently, knowledge and awareness about AI and ethics needs to be present in organizations also on a management level. But this understanding still needs to be grounded in the reality of AI developing teams for it to be effective. A good understanding of AI will be a prerequisite for this and should be encouraged.

Getting the AI developing community on board is key
Ethics need to be a factor when developing AI solutions. However, in reality, pragmatism and beating state of the art often takes precedence. Simply put, there is a tradeoff between performance and ethics. Developing teams should have the support and incentives needed to navigate these tradeoffs. Getting the AI community on board with the importance of ethics will be key for organizations to dare to challenge the current state of often favoring performance.

We need to work on both a short-term and a long-term timeline
AI ethics is a huge field. The legal perspective, policies, and standards aside, there are still many issues to address in order to get one step closer to implementing AI ethics in practice. It is clear from the discussions that we need to entertain several timelines at the same time. Certain topics, such as introducing non-technical competence into AI development teams to diversify the disciplinary knowledge will take time. Still, AI developers are already today confronted with ethical dilemmas on a daily basis and getting access to the right tools and support structures is fundamental and should be made a priority right here, right now. These two perspectives are not mutually exclusive and should both be addressed, although with different timelines and target groups in mind.

What this means going forward

Next step
The AI Ethics Lab will meet two more times during 2021 and will continue to discuss AI ethics with real use cases as a starting point. The intention is to extract cross-sectorial learnings on a level high enough to understand what areas to focus on to best support the implementation of AI ethics in practice. Technical learnings and recommendations for each case will not be feasible. We encourage members of the AI Ethics Lab to contribute with use cases and ideas for how to collaboratively shape the AI Ethics Lab together.

Questions?
Feel free to reach out to Karin Vajta Engström, karin.vajta@ai.se.