FIFTEEN RECOMMENDATIONS: FIRST STEPS TOWARDS A GLOBAL ARTIFICIAL INTELLIGENCE CHARTER

THOMAS METZINGER
JOHANNES GUTENBERG UNIVERSITY MAINZ

Introduction

In what follows, I will present a condensed and non-exclusive list of the five most important problem domains in the development and implementation of Artificial Intelligence (AI), each with practical recommendations.

The first problem domain to be examined is the one which, in my view, is constituted by those issues with the smallest chances of being resolved. It should therefore be approached in a multi-layered process, beginning in the European Union (EU) itself.[1]

The "race-to-the-bottom" problem

We need to develop and implement world-wide safety standards for AI research. A Global Charter for AI is necessary, because such safety standards can only be effective if they involve a binding commitment to certain rules by all countries participating and investing in the relevant type of research and development. Given the current competitive economic and military context, the safety of AI research will very likely be reduced in favour of more rapid progress and reduced cost, namely by moving it to countries with low safety standards and low political transparency.

- Recommendation 1
  The EU should immediately develop a European AI Charter.
- Recommendation 2
  In parallel, the EU should initiate a political process and lead the development of a Global AI Charter.
- Recommendation 3
  The EU should invest resources into systematically strengthening international cooperation and coordination. Strategic mistrust should be minimised, and commonalities can be defined via maximally negative scenarios.

The second problem domain to be examined is arguably constituted by the most urgent set of issues, and these also have a rather small chance of being resolved to a sufficient degree.

[1] For a slightly longer treatment, see the following open access publication: Metzinger (2018).

Prevention of an AI arms race

- Recommendation 4
  The EU should ban all research on offensive autonomous weapons within its borders and seek international agreements.
- Recommendation 5
  For purely defensive military applications, the EU should fund research into the maximum degree of autonomy for intelligent systems that appears to be acceptable from an ethical and legal perspective.
- Recommendation 6
  On an international level, the EU should start a major initiative to prevent the emergence of an AI arms race, using all diplomatic and political instruments available.

The third problem domain to be examined is one for which the predictive horizon is probably still quite distant, but where epistemic uncertainty is high and potential damage could be extremely large.

A moratorium on synthetic phenomenology

It is important that all politicians understand the difference between artificial intelligence and artificial consciousness. The unintended or even intentional creation of artificial consciousness is highly problematic from an ethical perspective, because it may lead to artificial suffering and a consciously experienced sense of self in autonomous, intelligent systems. Therefore, it may also lead to artificial subjects or a historically new category of legal persons. Such systems would have to be treated as bearers of rights, because they confer an intrinsic value on themselves by desiring their own, self-conscious existence as an end in itself.

- Recommendation 7
  The EU should ban all research that risks or directly aims at the creation of synthetic phenomenology within its boundaries and seek international agreements. [2]
- Recommendation 8
  Given the current level of uncertainty and disagreement within the na-

[2] This includes approaches that aim at a confluence of neuroscience and AI with the specific aim of fostering the development of machine consciousness. For recent examples see Dehaene, Lau, Kouider (2017), Graziano (2017), Kanai (2017).

scent field of machine consciousness, there is a pressing need to pr promote, fund and coordinate relevant interdisciplinary research projects: evidence-based conceptual, neurobiological and computational models of conscious experience, self-awareness and suffering.

- Recommendation 9
  On the level of foundational research there is a need to promote, fund and coordinate systematic research into the applied ethics of non-biological systems that are capable of conscious experience, self-awareness and subjectively experienced suffering.

The next general problem domain to be examined is the most complex one and likely contains the largest number of unexpected problems and "unknown unknowns".

Threats to social cohesion

- Recommendation 10
  Within the EU, AI-related productivity gains must be distributed in a socially just manner. Obviously, past practice and global trends clearly point in the opposite direction: We have (almost) never done this in the past, and existing financial incentives directly counteract this recommendation.
- Recommendation 11
  The EU should carefully research the potential for an unconditional basic income or a negative income tax on its territory.
- Recommendation 12
  Research programmes are needed to assess the feasibility of accurately timed retraining initiatives for threatened population strata. These initiatives should aim to develop creative and social skills.

The next problem domain is difficult to tackle, because most of the cutting-edge research in AI has already moved out of publicly funded universities and research institutions.

Research ethics

- Recommendation 13
  Any AI Global Charter or its European precursor should always be complemented by a concrete Code of Ethical Conduct guiding researchers

in their practical day-to-day work.

- Recommendation 14
  A new generation of applied ethicists specialised in problems of AI technology, autonomous systems and related fields must be trained.
- Recommendation 15
  The EU should invest in researching and developing new governance structures that dramatically increase the speed at which established political institutions can respond to unexpected problems and actually enforce new regulations.

References

Dehaene, Stanislas; Lau, Hakwan; Kouider, Sid (2017). What is consciousness, and could machines have it? Science (New York, N.Y.), Vol 358 (6362), pp. 486–492.

Graziano, Michael S. A. (2017). The Attention Schema Theory. A Foundation for Engineering Artificial Consciousness. Frontiers in Robotics and AI 4, p. 61.

Kanai, Ryota (2017). We Need Conscious Robots. How introspection and imagination make robots better. Nautilus (47). Available at: http://nautil.us/issue/47/consciousness/we-need-conscious-robots [Accessed 22.11.2018].

Metzinger, Thomas (2018). Towards a Global Artificial Intelligence Charter. In European Parliament (ed.), Should we fear artificial intelligence?, Brussels: European Union PE 614.547, http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf [Accessed 22.11.2018].