THE DAGSTUHL DECLARATION ON THE APPLICATION OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE FOR SOCIAL GOOD

WE the undersigned, practitioners in the areas of Machine Learning, Artificial Intelligence [1], academics from across the world and from non-governmental organizations across various sectors of social development, the initial group of whom met in Dagstuhl, Germany, on 18th - 22nd February 2019, hold and declare as follows:

1. Worldwide, the use and uptake of Artificial Intelligence (AI) and Machine Learning (ML) technologies in everyday life has risen to become an integral part in many everyday tasks and activities.

2. These advances in the fields of AI and ML have the potential to present new frontiers for the transformation of society.

3. Notwithstanding the great strides of AI and ML technology, there still exist systemic imbalances in the equity of all in society, as is envisioned in the United Nations Sustainable Development Goals.

4. We acknowledge the potential for AI and ML, when properly used, to be a direct driver in achieving social good.

5. We acknowledge the existing work by industry, national and international organizations to develop shared principles governing AI [2]. While these high-level guidelines are useful for considering the impact of AI systems broadly, there is a need for specific guidelines that cover AI for Social Good projects that govern partnerships between non-governmental organizations, local non-profits and academics/industry sectors.

6. Our understanding of social good is anchored in the United Nations Sustainable Development Goals which “provide a shared blueprint for peace and prosperity for people and the planet” [3].

In pursuing applications of AI and ML to Social Good, the signatories to this document will commit to the following recommendations:

1) Operating within a just and fair ethical framework that respects human rights standards and law.
2) Striving for AI projects to be built on accessible technology.
3) Striving for AI projects to be deployable across various settings and communities.
4) Encouraging the collection of highest quality data available.
5) Fostering deep collaboration and partnership between the AI and NGO communities.

We believe that the following principles can be beneficial to anyone working on AI for social good projects, which will help us form a community of practice around AI for Social Good.

1) ***Operating within a Just and Fair Ethical Framework***

AI for Social Good projects should take place within a fair and just ethical framework rooted in human rights standards and law, such as ensuring respect for and protection of the liberty, equality, dignity and privacy of all affected individuals and communities [4].

Those working on AI for Social Good projects should strive to have a deep understanding of the discrimination against, and vulnerabilities of, the users, beneficiaries, and of the people potentially affected by such technologies. Such vulnerabilities may exist because of race, age, religion, culture, language, gender, sexual orientation, gender identity, ideology, socio-economic status, disability, belonging to an ethnic minority, or other reasons, as well as at the intersection of two or more of these dimensions.

An awareness of the vulnerabilities of these groups and the potential harmful effects of AI on them is important if practitioners want the technology to have a meaningful impact. In the current climate, the AI community constantly needs to reflect on the implications of the tools developed and how such tools may introduce new or reinforce previous bias and unfairness towards vulnerable groups. As a community interested in AI for Social Good, we need to navigate this domain with even greater care, to ensure that the AI systems we develop do not exacerbate some of the vulnerabilities which we are trying to alleviate.

2) ***Accessible Technology***

In order for AI and ML to have a significant impact for social good, the technology we develop needs to be accessible to all. Accessibility implies the ability of technology to transcend both cultural and linguistic boundaries.

The AI and ML technology also needs to take into account and work fairly for those whom it is intended to benefit. In this spirit, domain experts and AI researchers need to ensure that technologies are designed to be accessible to all. AI and ML technologies should be designed with all users in mind, particularly those who are vulnerable. Accessibility also implies that people across a broad range of mental, physical, and technical abilities can interpret and utilise the technology.

Another aspect of accessibility is affordability. In an attempt to deploy technology and make it accessible and inclusive to multiple sectors of society, it is important that these

technologies are affordable to develop, use and maintain. Some ways of ensuring affordability is through open access of code and technology. It is important that computation requirements are minimized, and such requirements are related to the principles of creating technology which is accessible *by design*.

In order to achieve this, the costs associated with each stakeholder need to be taken into account. The potential overhead costs, such as access to dependent technologies and associated infrastructure such as gas, electricity, mobile phones and internet are also important.

For long-term affordability, machine learning researchers, technologists and NGOs should strive to develop technologies and collaborations that are cost-effective. This should be done with the view of developing more self-sustainable (automated) systems where possible, where the burden of external expertise costs such as specialist statisticians or AI and ML technologists is reduced. Even with partially automated AI and ML systems, NGO and domain experts of course need to remain an integral part of the system to ensure effectiveness and human oversight.

3) ***Deployable across various settings and communities***

AI and ML for social good projects should strive to be deployable across various geographical and cultural settings so as to maximize the benefits of a wide range of people. Scarcity of, and competition for, resources in the social good sector may also translate into scarcity of AI and ML expertise. In order to have maximal impact for social good, AI solutions for problems within the developmental domain need to be transferable to grassroot developers and communities.

Effective deployment also requires continual learning and updating of modelling frameworks. This also facilitates being able to reimplement algorithms into other similar settings and also make the necessary adjustments and fine tuning to algorithms for communities with other pressing needs. Al solutions and their documentations should be open source/access when possible, however with caution to prevent potential malicious usage.

Although in an ideal world, we would want models that generalize broadly, we may also have situations in which we only have good enough data to build models that generalize for a very particular setting (e.g., country/region) and not beyond. In such cases, we should avoid the temptation to apply the models beyond their limitations.

4) ***Good Data Quality***

An essential element for successful development in AI for social good is high quality of data. Data is what powers AI systems, and bad data going in will result in bad AI-driven outputs going out. If the data is of poor quality, the best and fastest algorithms in the world will be rendered useless. Also, it is important to understand the provenance of data to ensure relevance and prevent unintended biases. Hence there is a need for a systematic approach to study design and the necessary supporting infrastructure for collecting such data and documenting the data collection process. It is essential to build data pipelines that have the appropriate degree of complexity for the particular project. Data collection, retention, and use should be done in accordance with accepted professional and legal standards of privacy and data protection.

5) ***Fostering Deep Partnerships***

Tight collaboration and equal partnership between the AI and social good communities are powerful means to achieve AI for social good. Successful AI for Social Good requires tight collaboration and equal partnership between the AI and social good communities in order to create a deep understanding of the problem domain. We, the participants of the Dagstuhl Workshop on "AI for Social Good", have benefited from the close collaboration to reach a common understanding and a common language for thinking about the application and influence of AI for social Good.

However, this workshop cannot remain an isolated event. We rather need to think of ways to achieve greater continuity and sustainability. This could be done through systems that have the capacity of integrating a feedback loop of external factors. Such a feedback loop would include incorporating domain knowledge from experts in the NGO sector. Close collaboration is also important for setting expectations.

AI and ML cannot solve every problem, nor is AI or ML needed to solve every problem. ML practitioners and local communities must work together to properly scope out the problem at hand and set clear expectations with each other for what can be done.

[1] We use the High-Level Expert Group on Artificial Intelligence (AI HLEG) definition of AI found on https://ec.europa.eu/futurium/en/ai-alliance-consultation: “Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).”

[2] For example: Toronto Declaration, Asilomar Principles, OECD Principles, AI HLEG’s principles for Trustworthy AI.

[3] United Nations. (2015) General Assembly Resolution A/RES/70/1. Transforming Our World, the 2030 Agenda for Sustainable Development. [cited 2016 Feb 10]. Available from: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

[4] United Nations. General Assembly. (1949). Universal declaration of human rights (Vol. 3381). Department of State, United States of America.

Drafting committee members:

Danielle C. M. Belgrave, Microsoft Research, Cambridge, UK,
Department of Medicine, Imperial College London, UK
Gerald Abila, BarefootLaw, Kampala, UG
Ruben De Winne, Oxfam Novib, The Hague, NL
Tom Schaul, DeepMind, London, UK
Claudia Clopath, Department of Bioengineering, Imperial College London, UK
Mohammad Emtiyaz Khan, RIKEN Center for AI Project, Tokyo, JP
Julia Proskurnia, Google, Zurich, CH
Yee Whye Teh, DeepMind, London, University of Oxford, Oxford, UK
Julien Cornebise, Element AI, London, UK,
Department of Computer Science, University College London, UK
Nenad Tomašev, DeepMind, London, UK
Frank Hutter, Department of Computer Science, University of Freiburg, Freiburg, DE
Angela Picciariello, Oxfam GB, Oxford, UK
Bec Connelly, RNW Media, Hilversum, NL
Daphne Ezer, University of Warwick, Warwick, UK, Alan Turing Institute, London, UK
Fanny Cachat van der Haert, International Commission of Jurists, Brussels, BE
Frank Mugisha, Chemonics International Inc., Kigali, RW
Hiromi Arai, RIKEN Center for AI Project, Tokyo, JP
Hisham Almiraat, Justice and Peace Netherlands, The Hague, NL
Kyle Snyder, RNW Media, Hilversum, NL
Mihoko Otake, RIKEN Center for AI Project, Tokyo, JP
Mustafa Othman, Shaqodoon Organization, Hargeisa, SO
Shakir Mohamed, DeepMind, London, UK
Tobias Glasmachers, Institute for Neural Computation, Ruhr-University Bochum, DE
Wilfried de Wever, SEMA, Kampala, UG, Humanity Solutions

This declaration was conceived in the Dagstuhl Seminar 19082 - AI for the Social Good. Please note that the declaration was not made by Schloss Dagstuhl - Leibniz-Zentrum für Informatik GmbH, and thus does not represent a statement by Schloss Dagstuhl - Leibniz-Zentrum für Informatik GmbH.

Signatories:
The list of all signatories can be found here.

This Declaration was published on 5th July 2019 on the following website:
https://www.dagstuhl.de/fileadmin/redaktion/Programm/Seminar/19082/Declaration/Declaration.pdf