Report from Dagstuhl Seminar 19082

AI for the Social Good

Edited by

Claudia Clopath[1], Ruben De Winne[2], Mohammad Emtiyaz Khan[3], and Tom Schaul[4]

1 **Imperial College London, GB,** c.clopath@imperial.ac.uk
2 **Oxfam Novib – The Hague, NL,** ruben_de_winne@hotmail.com
3 **RIKEN – Tokyo, JP,** emtiyaz.khan@riken.jp
4 **Google DeepMind – London, GB,** schaul@google.com

Abstract

Artificial intelligence (AI) and machine learning (ML) have made impressive progress in the last few years. Long-standing challenges like Go have fallen and the technology has entered daily use via the vision, speech or translation capabilities in billions of smartphones. The pace of research progress shows no signs of slowing down, and demand for talent is unprecedented. AI for Social Good in general is trying to ensure that the social good does not become an afterthought, but that society benefits as a whole. In this Dagstuhl seminar, we brought together AI and machine learning researchers with non-governmental organisations (NGOs), as they already pursue a social good goal, have rich domain knowledge, and vast networks with (non)-governmental actors in developing countries. Such collaborations benefit both sides: on the one hand, the new techniques can help with prediction, data analysis, modelling, or decision making. On the other hand, the NGOs' domains contain many non-standard conditions, like missing data, side-effects, or multiple competing objectives, all of which are fascinating research challenges in themselves. And of course, publication impact is substantially enhanced when a method has real-world impact.

In this workshop, researchers and practitioners from diverse areas of machine learning joined stakeholders from a range of NGOs to spend a week together. We first pursued an improved understanding of each side's challenges and established a common language, via presentations and discussion groups. We identified ten key challenges for AI for Social Good initiatives. To make matters concrete, we organised a hackathon around some existing technical questions within the NGOs to scope the applicability of AI methods and seed collaborations. Finally, we defined guidelines and next steps for future AI for Social Good initiatives.

Seminar February 17–22, 2019 – http://www.dagstuhl.de/19082
2012 ACM Subject Classification Computing methodologies → Machine learning
Keywords and phrases Machine Learning, Artificial Intelligence, Social Good, NGO, sustainable development goals, Non-governmental organisation
Digital Object Identifier 10.4230/DagRep.9.2.111

1 Executive Summary

Ruben De Winne (Oxfam Novib – The Hague, NL, ruben_de_winne@hotmail.com*)*

The purpose of Dagstuhl Seminar 19082: AI for the Social Good was to bring together researchers in artificial intelligence (AI) and machine learning (ML) with non-governmental organisations (NGOs) to explore if and how AI and ML could benefit the social good. Indeed,

AI and ML have made impressive progress in the last few years. Long-standing challenges like Go have fallen and the technology has entered daily use via the vision, speech or translation capabilities in billions of smartphones. The pace of research progress shows no signs of slowing down, and demand for talent is unprecedented. But as part of a wider AI for Social Good trend, this seminar wanted to contribute to ensuring that the social good does not become an afterthought in the rapid AI and ML evolution, but that society benefits as a whole.

The five-day seminar brought together AI and ML researchers from various universities and industry research labs with representatives from NGOs based in Somalia, Rwanda, Uganda, Belgium, United Kingdom and The Netherlands. These NGOs all pursue various social good goals, such as increasing access to justice for vulnerable people, promoting human rights & protecting human rights defenders, and defeating poverty. On these topics, NGOs have rich domain knowledge, just like they have vast networks with (non-)governmental actors in developing countries. Mostly, NGOs have their finger on the pulse of the challenges that the world & especially its most vulnerable inhabitants are facing today, and will be facing tomorrow. The objective of the seminar was to look at these challenges through an AI and ML lens, to explore if and how these technologies could help NGOs to address these challenges. The motivation was also that collaborations between AI and ML researchers and NGOs could benefit both sides: on the one hand, the new techniques can help with prediction, data analysis, modelling, or decision making. On the other hand, the NGOs' domains contain many non-standard conditions, like missing data, side-effects, or multiple competing objectives, all of which are fascinating research challenges in themselves. And of course, publication impact is substantially enhanced when a method has real-world impact.

The seminar facilitated the exploration of possible collaborations between AI and ML researchers and NGOs through a two-pronged approach. This approach combined high-level talks & discussions on the one hand with a hands-on hackathon on the other hand. High-level talks & discussions focused first on the central concepts and theories in AI and ML and in the NGOs' development work, before diving into specific issues such as privacy & anonymity, data quality, intellectual property, accessibility and ethical issues. These talks and discussions allowed all participants – in a very short timeframe – to reach a sufficient level of understanding of each other's work. This understanding was the basis to then start investigating jointly through a hackathon how AI and ML could help addressing the real-world challenges presented by the NGOs. At the start of the hackathon, an open marketplace-like setting allowed AI and ML researchers and NGOs to find the best match between technological supply and demand. When teams of researchers and NGOs were established, their initial objective was not to start coding, but to define objectives, assess scope and feasibility. Throughout the hackathon, group membership was fluid, as some projects finished early, were deemed out of scope, or needed to wait for data. Some groups managed to build a viable initial prototype, others established the seeds for future collaborations, and a few were proposed as full summer projects within the "Data Science of Social Good summer school". The projects' aims were diverse. They included better seeds for farmers, modelling cognitive age and decline, scalable legal assistance and scalable citizen feedback. As a result of the hackathon, all NGOs could take concrete results home – some to build on further, some as mature solutions.

Finally, a result of the seminar that is relevant for the entire AI for Social Good community are the ten key challenges for AI for Social Good initiatives that participants identified:

1. the importance of deep, long-term partnerships,
2. clear and well-defined goals and use cases,

3. bias towards simpler solutions,
4. data readiness,
5. setting expectations with regards to both impact and the pace at which technology can be applied,
6. ensuring privacy and security of data,
7. inclusivity and ethics of the applications,
8. factoring in the limitations of both communities,
9. challenges in overcoming the barriers to NGOs utilising the potential of AI/ML, and
10. the relative cost of AI/ML for social good.

2 Table of Contents

3 Overview of Talks

3.1 Introduction to the workshop

Claudia Clopath (Imperial College London, GB)

Claudia Clopath introduced the notion of "AI for Social Good" and the motivation for the workshop. In particular, she listed the following goals: 1) Bring NGO participants and AI researchers together, 2) Create awareness, 3) Make sure incentives are aligned, 4) New innovations coming out of the newly formed collaborations.

The potential outcomes of the workshops were: 1) Understanding the link between the two worlds: Machine Learning and NGOs, 2) Understanding the ways to make AI for Social Good initiatives work, 3) Build Prototypes during the hackathon time, 4) Build collaborations, create a network, make friends, 5) and prepare a set of follow-up plans. Finally, she defined a set of rules for the workshop: Connect, Respect, Bottom-Up, Brain-storm, Fun.

3.2 Successful examples of AI for Social Good

Julien Cornebise (Element AI- London, GB)

The movement of "AI for Good" is undergoing a great gain of interest, with the United Nations AI for Good Summit being in its third iteration and several "AI for Good" workshops proposed for NeurIPS 2018. Yet it is still pretty recent, and as with any young domain, this might cause some teething problems. The current hype around Artificial Intelligence and Machine Learning can also induce some measure of "tech saviour syndrome". In the field of "for good" however, resources are often scarce and lives can be at stake, bringing a very high cost to any such early mistakes.

In this talk, we draw parallels with three decades of ICT4D (Information and Communications Technology for Development), to learn some of key lessons and avoid repeating mistakes from the past. We start by suggesting a way to navigate the obvious question of "What is good?" by partnering with actual domain experts. We show some of the key factors of success and explain how to navigate some of the pitfalls that can rise in such projects. We illustrate the point with two concrete successful projects over the last three years, which brought together domain experts from Amnesty International, and machine learners. We finally open the conversation with a wider reflection on how the current specialists in machine learning, mainly employed in the private sector, may be channelled to help with much larger social problems that we are facing as a society and as a species.

3.3 Using Machine Learning to Address Global Challenges

Shakir Mohamed (Google DeepMind – London, GB)

The sustainable development goals (SDG) give a broad framework for addressing global challenges that have been agreed to by the governments of the world, and we now have the opportunity to think of the ways in which machine learning can contribute to addressing this global agenda. To achieve this, we will discuss approaches for doing this using the integrated transformations given by the SDGs, the need for high-frequency and diverse data sources, and the central role of multi-disciplinary teams. We'll use the example of satellite imagery and look at how they provide a data source to look at global challenges in poverty prediction, deforestation, global fishing monitoring, and the deployment of domestic solar panels. We will also explore other data sources and areas in weather and early warning, energy, healthcare and conservation.

3.4 Introduction to Machine Learning I

Mohammad Emtiyaz Khan (RIKEN – Tokyo, JP)

In this talk, I will present an initial introduction to Machine Learning (ML). I will start with the historical context of the relationship of ML with AI, and formulate a general definition to differentiate it from other fields such as statistics, data mining etc. I will summarise some of the recent successes, but also talk about some documented failures. I will conclude by mentioning subfields within machine learning and some of the open problems.

3.5 Introduction to Machine learning II

Yee Whye Teh (University of Oxford, GB)

Yee Whye Teh introduced the difference between unsupervised, supervised and reinforcement learning. He explained the different levels of difficulty and what is required to make them work. He explained–in simple language for the NGO participants–the advantages and disadvantages of the three different methods and illustrated these points with examples.

3.6 Introduction to Reinforcement Learning – chalk talk

Tom Schaul (Google DeepMind – London, GB)

Reinforcement learning (RL) is both one of the most general types of AI and one of the more difficult ones to apply. It permits the automation of complex sequential decision making processes, maximising any well-defined objective. It learns by interacting, trying things out and learning from its successes and mistakes, and it thus needs only minimal knowledge about the task. To work well, however, it requires a lot of direct interaction, thus making many potentially costly/harmful mistakes while learning is still going on. The second drawback for practical applications is the fact that defining the reward/success can be complicated or ambiguous–yet, clarity of objective is a desirable thing in itself, so framing a problem as an RL problem often leads to more clarity.

4 Working groups

Cracking each other's code

We broke up into small sessions and discuss what each participant is going on an everyday basis at work and what are the terms we commonly use. This was a way to get to know each other, understand each other's work and challenges and understand each other's jargon. One important point was to outline what defines success in their work.

Topics for discussion

We dedicated one session to eliciting a broad range of topics of discussion, surrounding AI for Social Good, from our highly diverse and interdisciplinary participants. On the basis of this list (see below) we grouped and prioritised, and those resulted in the break-out groups for the following days–but of course, not all of them could be discussed within the limited time.

- Trust and transparency
- Privacy and confidentiality
- Equality
- Ethics and principles
- Affordability, funding
- Talent, capacity, matchmaking
- Intellectual property
- Access, usability, bandwidth, illiteracy
- Who is the user?
- SDG alignment
- Urgency, relevance
- Maintenance, technology transfer, integration, sustainability
- Robustness, verifiability, open data
- Fake claims, adversarial actors, exploiters, aid shopping
- Text (Docs + informal), NLP (natural language processing), dialogue
- Voice

- Privacy, anonymisation, safety
- Data quality (scarcity, missing data, sample size, confidence in the data)
- Data safety
- Sensors
- Visualisation
- Predictive modeling
- Causality, outliers, classification similarities
- Performance evaluation
- Auto ML
- Transfer + generalisation
- Bias, interpretability
- Continual learning, adaptation
- Measurement models
- Feature design
- On-line versus off-line

Theory of Change

Theories of change are an important technical term in civil society and the NGO world. They capture a problem specific understanding of causal structure. Most government, institutions, and grant making institutions use theories of change to structure interventions. Causality, as studied by machine learners, can be a complementary tool to theories of change.

Urgency, Relevance and Goals

NGOs have a tension between quality and quantity of impact. For example, donors would like to see big numbers of beneficiaries, but deep individual stories are important for private supporters. This leads to a risk of cherry-picking if the strategy is only going for quality, which in turn can be perceived as unfair in the community of the beneficiaries. NGOs need to find an empty niche where they can avoid competing with other NGOs. But importantly, the relevance of the projects needs to be aligned with their mission, needs to reflect the donor's priorities and needs to be centred on the beneficiaries. NGOs have tend to not plan long-term, as funding cycles are short, and many funders avoid taking risks. However, the reality is that long-term investments tend to help in many ways, including the urgent issues (e.g., education is useful for addressing climate change or epidemics). We discussed ways to go around that, namely 1) using for-profit companies, 2) by investing in long-term monitoring like sensors for example, 3) exerting political pressure to change the culture and constraints. Finally, we discussed the problem of "free riders" who don't contribute to solving the problem but take advantage of the solution.

Privacy & anonymity

We discuss the ways to encrypt data both for storing and sharing. We discuss the tools to anonymise data, e.g., removing fields with personally identifiable information (PII) or using tools to obtain k-anonymity. Multi-party computation permits collaborating without sharing data. If we do linear data analysis, adding noise makes sure individuals cannot be identified. Finally, we talked about how to find and remove biases in data and the risk of fake data.

Use of Auto-ML

Auto-ML is a tool to democratise ML so that non-experts can use ML technology. NGOs could start using it, but they at least need a software engineer to begin with. We discussed the fact the the model is only as strong as the data, so if the data is not good, Auto-ML cannot do miracles. We talked about the issue of interpretability: we can actually understand what the machine is doing. This is particularly important if one needs to be accountable for the conclusions the machine is making.

Data quality

We discuss of concept of "Garbage in, garbage out": if the data is not good, the ML won't be good either. We discussed ways to clean the data both on the NGOs side and on the ML side, and acknowledged the fact that data cleaning is typically a large part of the whole project. We discussed the different possible data formats: tabular, speech, text. Finally, we talked about predictive models and that fact that each project separately needs to define what is the minimum viable solution, and what are requirements in terms of fairness and beyond raw performance. Finally, we presented the "Data Science Hierarchy of Needs", where there are 6 levels, from the bottom to the top:

1. Collect data: instrumentation, logging, sensors, external data, user generated content.
2. Move and Store: reliable data flow, infrastructure, pipelines, data storage.
3. Explore and transform: cleaning the data, detect anomalies, etc.
4. Aggregate and label: analytics, metrics, segments, aggregates, features, training data.
5. Learning/optimise: A/B testing, experimentation, simple ML algorithms.
6. AI and Deep learning.

Only when the 5 first levels are established, does it make sense to move on to the 6th level and invest in state-of-the-art AI methods.

Intellectual property

We first discussed the how important it is for NGOs to know their level of maturity in terms of data availability, cleanliness and usability. It is important that intellectual property (IP) is taken into account throughout the project. In particular, NGOs need to understand the value of their data and also use that to pitch it better to their donor: if the data Is better/richer/more abundant, the results will be more interesting.

Accessibility

We first discussed the accessibility to the most vulnerable, namely in terms of language barriers, disabilities, minority groups, access to devices, internet connectivity and infrastructure. We talked about the fact that technology should be "accessible by design". We then talked about usability both from the NGOs side and from the beneficiaries side. These principles should be aligned with the SDGs to ensure inclusiveness. In order to achieve that, we should look at the feedback from users and how we can correct the system to ensure accessibility. We should think of the eligibility criteria for the users and staff, and about training them.

Ethical issues

As this topic is so complex and multi-faceted, we had multiple separate break-out groups on the topic of ethics. Among the issues that were raised were the following:

- Involving the voice of affected people versus not endangering them
- Transparency versus privacy
- Transparency versus solving the problem
- Meaningful informed consent, building trust, buy-in from all stakeholders
- Gradual/iterative deployment versus quick impact with risks
- Job loss risks versus efficiency gains
- Safe exploration of private data
- Framing of risks, perception, awareness and public relations
- Responsibility to identify the most vulnerable
- Who is responsible for the decisions/mistakes the ML system makes?
- Mitigating risks
- Delegating to the machine versus helping humans to make better decisions

On the solution side, we discussed the Toronto framework on responsibly dealing with data, and the Data4Development criteria.

5 Hackathon

Two days of the workshop were dedicated to a hackathon. After project presentations by the NGO participants and tool presentations by the ML participants, we had a couple of match-making rounds to form initial mixed groups. The initial objective for each group was not to start coding, but to define objectives, assess scope and feasibility. Throughout the hackathon, group membership was fluid, as some projects finished early, were deemed out of scope, or needed to wait for data. Some groups managed to build a viable initial prototype, others established the seeds for future collaborations, and a few were proposed as full summer projects within the "Data Science of Social Good summer school". Among the projects were the following:

Better seeds for farmers. We built a clustering algorithm that matches the similarity in score for (standardised) crop varieties and traits, so that users of the application can view which varieties of a specific crop display similar traits based on their own scoring of a certain variety.

Modelling cognitive age and decline. We built a prototype to distinguish between groups of people with different levels of cognitive decline. We used publicly available data.

Scalable legal assistance. Legal advice is typically accessible to only a certain subset of people. Here we tried to improve an existing process of providing legal recommendations to the least privileged beneficiaries. In particular, a recommendation system prototype (based on past responses) was designed to making some of that process less labour-intensive, so that the same number of lawyers can serve more requests.

Competent judges. Another way of improving outcomes in the justice system is to improve the quality and consistency of judgements, by training judges and automatic monitoring. In the workshop, we defined steps towards enabling a such a project and how to gather the necessary data.

Scalable citizen feedback. Two projects were centred around how to effectively handle large amounts of citizen feedback, but they differed in the data modality (voice messages versus tabular). The voice message feedback service service looked promising at first, but we

realised that there was very little clean and annotated data, and there do not currently exist out-of-the-box tools for voice recognition in languages like Somali.

Spaces for trusted discussions. We assessed the feasibility to study the effect of moderation on social media data.

6 Guidelines / Takeaways

The seminar highlighted ten key insights amongst domain experts from NGOs and technical specialists from machine learning, in relation to:

1. the importance of deep, long-term partnerships,
2. clear and well-defined goals and use cases,
3. bias towards simpler solutions,
4. data readiness,
5. setting expectations with regards to both impact and the pace at which technology can be applied,
6. ensuring privacy and security of data,
7. inclusivity and ethics of the applications,
8. factoring in the limitations of both communities,
9. challenges in overcoming the barriers to NGOs utilising the potential of AI/ML, and
10. the relative cost of AI/ML for social good.

7 Follow-up actions

In the last part of the seminar, we discussed ways forward for AI for Social Good initiatives in general, and with the collaborations we started during the week. We discussed the following action plan:

- We plan to write a manuscript with the take-home messages that we have learned during the workshop. The audience targeted are people who are running or want to run AI for Social Good initiatives.
- We plan to write a Global Challenges manuscript that relates examples of AI for Social Good projects to the Sustainable Developmental Goals.
- We plan to write a third document with the key principles of AI for Social Good projects, to serve as a call to action.
- We plan to identify, organise and share the different sources of funding that could be used for AI for Social Good projects.
- We listed possible future meetings and workshops.
- We finally made sure that the collaborations we started at the seminar will continue. To that end, we defined multiple milestones and commitments, we built working groups, etc.

Participants

- Gerald Abila
BarefootLaw – Kampala, UG
- Hisham Almiraat
Justice and Peace Netherlands – The Hague, NL
- Hiromi Arai
RIKEN – Tokyo, JP
- Danielle C. M. Belgrave
Imperial College London, GB
- Fanny Cachat
ICJ – Brussels, BE
- Claudia Clopath
Imperial College London, GB
- Bec Connelly
RNW Media – Hilversum, NL
- Julien Cornebise
Element AI- London, GB
- Wilfried de Wever
SEMA – Kampala, UG
- Ruben De Winne
Oxfam Novib – The Hague, NL
- Daphne Ezer
University of Warwick, GB
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Frank Hutter
Universität Freiburg, DE
- Mohammad Emtiyaz Khan
RIKEN – Tokyo, JP
- Shakir Mohamed
DeepMind, GB
- Frank Mugisha
Chemonics International Inc. – Kigali, RW
- Mihoko Otake
RIKEN – Tokyo, JP
- Mustafa Othman
Shaqodoon Organization – Hargeisa, SO
- Angela Picciariello
Oxfam – Oxford, GB
- Julia Proskurnia
Google Switzerland – Zürich, CH
- Tom Schaul
Google DeepMind – London, GB
- Kyle Snyder
RNW Media – Hilversum, NL
- Yee Whye Teh
University of Oxford, GB
- Nenad Tomasev
Google DeepMind – London, GB