

Why we need responsible AI in times of Corona and how to achieve it

Cateljne Muller, Virginia Dignum, Noah Schöppel

November 24, 2020

This is a publication by ALLAI, a Foundation under Dutch Law, Prinseneiland 23A 1013 LL Amsterdam. This publication is aimed at incepting the ALLAI Project [Responsible AI and Corona](#).

Abstract

In efforts to tackle the Corona-crisis, many public institutions and private organisations are considering to fast-track the deployment of AI applications, both in medical as well as social contexts. There is an increasing concern about the implications of these applications on fundamental rights, ethical principles and societal values. This project aims to address these concerns in several ways. First, it will develop an Observatory to highlight AI applications deployed in the Corona-crisis and their opportunities and challenges. Second, expert sessions will be organised to raise awareness with policymakers, healthcare professionals and the general public on how to deploy AI responsibly in times of crisis. Third, a framework will be developed that contains evaluation criteria to assess the efficacy and the legal, ethical and societal impact of AI-applications in times of Corona, and requirements for their responsible use. This framework will be piloted with a selected group of organisations that are developing or deploying AI to tackle the Corona crisis and culminate in to a quickscan that can be used for self-assessment.

1. Why this Project?

For a few years now, we have been hearing that AI can have significant advantages for society. So now that we are living through the greatest challenge of our lifetime, what better time is there to test, and hopefully validate, that presumption? Can we determine what role can AI play to tackle the Corona-crisis? And, more importantly, how can we make sure that it does so responsibly?

While there is ample research and quite some media buzz around AI tackling this crisis, examples of actual successes are limited. We read that AI applications could contribute to the understanding of the spread of the virus, to the search for a vaccine and to treatment of COVID-19, but caveats are made as to the maturity of these applications, mainly because of limited data about the virus and the disease.¹

We also see AI applications that are meant to guide and underpin policy measures and help gain insight in their long-term impact on society and the economy. While some of them have positive impact, many of them raise serious ethical, legal, societal and even technical concerns. Despite the urgency of this crisis, it is important that AI is developed and deployed in a responsible manner. Robustness, effectiveness, transparency and explainability, but also fundamental rights, inclusion and ethics are vital

to ensure that AI actually helps in tackling the Corona-crisis without causing harm to society along the way.

We hold the conviction that this crisis provides a unique opportunity to accelerate responsible AI. It provides the circumstances and the momentum to step up efforts to find truly innovative and valuable AI applications, and to simultaneously tackle the ethical, legal and societal questions that have up to now been dealt with mostly through high-level principles.

This project will provide insight in current 'AI versus Corona' research and development. It will give an indication of the state-of-the-art and maturity, the level of use and the ethical, legal and societal impact of the applications. It will look at tensions and possible trade-offs between conflicting values, interests and goals, taking into consideration the unique circumstances of this crisis. It will provide guidance on how to make optimal use of AI in this crisis within acceptable boundaries.

2. What kind of AI are we looking at?

For this project we broadly distinguish two areas of AI use: 'medical' and 'societal'. Medical use of AI involves applications that are directly related to the medical aspects of the Corona-crisis: epidemiology, molecular research, clinical practice and provision of care. Societal use of AI

¹ Bullock et al. 2020

involves applications that are related to the public and private aspects of the Corona crisis: policy measures (social distancing, mask wearing, lock-downs, etc.), distance learning, remote working, delivery of information, etc.

We acknowledge that the line between these two areas can blur for example for applications that are both publicly used and medically relevant, such as remote temperature scanning. For the purpose of clarity, we will categorise these applications as 'societal' when they are predominantly used/available outside the medical domain and 'medical' when they are used predominantly within the medical domain.

2.1 AI thrives on the habitual, and this crisis is new

For both categories it is relevant to remember that present day AI needs an abundance of high quality data to perform at an acceptable level. One of the hallmarks of this crisis however, is that this required abundance of data is lacking, both medically and socially.

Medically, we are dealing with an entirely new virus that leads to a disease that we have yet to fully understand. We are still not entirely sure how it spreads, why it affects certain people severely and spares others. There is no complete view of the infection rate and there is no accurate accounting of those that have perished and recovered. This limits the possibilities of AI. It can lead to incorrect or biased outcomes when used to predict the spread of the virus or false positives and false negatives when used for diagnostics.

Socially, there are new phenomena such as reduced traffic and movement as a result of lock downs, an explosion in online purchasing, altered public behaviour, etc. AI has shown to be performing less accurately due to these changed societal dynamics.

2.2 'Medical AI' to combat Corona

Nonetheless, we see ample research in medical AI².

² Bullock et al. 2020

³ Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. Nature. 2020

⁴ Joseph Bullock et al. 2020

⁵ Ibid, 7-8

⁶ Ibid, 13

⁷ Ibid, 10-11

⁸ M.B. Weinstock, MD et al. Chest X-Ray Findings in 636 Ambulatory Patients with COVID-19 Presenting to an Urgent Care Center: A Normal Chest X-Ray Is no Guarantee

⁹ Myura Nagendran et al.

¹⁰ Joseph Bullock et al. 2020

¹¹ <https://news.mit.edu/2020/covid-19-cough-cellphone-detection-1029>

¹² <http://windfall.ai/>

AI & molecular research: We have identified research on AI applications that can be used to estimate the structure of SARS-CoV-2 related proteins³ and to find existing treatments that can be repurposed to fight the disease COVID-19.⁴ There is some research into AI used to help drug discovery and vaccine development.⁵

AI & epidemiology: Epidemiology is a vast research field that has been using classical models such as susceptible-infected-recovered (SIR) and adjusting these to the current situation. Research into the role of machine learning to help model and predict the spread of the Corona-virus has shown complexities of choosing the right model and, more importantly, limitations of machine learning due to the limited availability of data about the virus and how it spreads.⁶

AI & clinical practice: In clinical practice, image recognition of lung CT's and X-rays by AI for diagnostic purposes has been vastly researched and has shown encouraging results. This could help in the event of limited regular testing capacity.⁷ Still it needs to be noted that relying on lung images as a diagnostic tool alone is disputed.⁸ Also, there exist few relevant randomised clinical trials of AI in medical imaging.⁹

Another AI avenue that is being explored is that of disease path prediction. While studies have been limited in scope and data, some propose 'AI triage' by looking at features to predict mortality risk¹⁰ or diagnostics through 'cough detection'¹¹.

AI & provision of care: Less explored is the use of AI to optimise the provision of care, i.e. the distribution of PPE, the testing process, hospital and ICU capacity and vaccine distribution. One example is a tool that predicts the course of infections, hospitalisations and deaths 3 days in advance by using 3 different AI models.¹²

2.2 'Societal AI' in times of Corona

When it comes to AI used in the social and private domain, we see more use of existing AI applications.

AI & policy measures: Policy makers are turning to the use of different types of AI to help decide on, underpin, understand or monitor (the effect of) policy measures.

During the G20 meeting of November, China proposed a global adoption of its COVID QR-code system. A detailed reaction to this proposal has been given by project lead Catelijne Muller and scientific lead Virginia Dignum in their op-ed.¹³

We see examples of 'social distancing AI' that uses image, object or facial recognition to control public adherence to social distancing measures.¹⁴

A project in Spain, run by the National Scientific Research Council (CSIC), uses AI to interpret aggregated and anonymised data from mobile phone records to understand the effects of lock-down measures and help decision makers in the process of strengthening or easing these measures.¹⁵

Several companies have developed 'face mask recognition tools' that they claim are being piloted in 'stealth mode' in several European countries and the US (by public authorities as well as private operators).¹⁶

Other applications we could identify include remote temperature scanning and biometric recognition to supposedly assess fatigue from a person's gait.

AI & remote education and examination: With worldwide school and university closings, we see a number of AI tools being deployed in the educational field. In particular, the UK grading algorithm drew negative attention, because it downgraded students based not on the student's own merits but merely on unrelated characteristics such as overall historic performance of the school.¹⁷

Another example is online proctoring, where students are monitored when taking exams remotely. These types of AI detect several activities such as keystrokes, eye movement and 'copying and pasting' to determine a 'suspicion level of cheating' for each student and flag 'suspicious behaviour'.¹⁸

AI & remote work and worker health: In the workplace and in employer-employee relations we notice an increase in AI-applications to monitor both worker health as well as worker productivity. Remote temperature checks for warehouse workers are being deployed by several companies. Some (mandatory) self-reporting health apps analyse and classify worker's 'risk levels' based on the information they provide.¹⁹

Other systems focus on tracking worker productivity, location and movement, i.e. through tracking of screen time and eye movement, monitoring of key strokes and use of wristbands or camera's to track worker location, social distancing and hand washing.²⁰ All this information can further be processed by AI systems to potentially categorise and classify workers.

AI & information: The World Health Organisation has defined "Infodemic" as a new phenomenon in this pandemic:²¹ '(...) an overabundance of information – some accurate and some not – that occurs during an epidemic. It can lead to confusion and ultimately mistrust in governments and public health response.' Also, the WHO has developed a multi-lingual chatbot that provides trusted pre-coded information on Corona.²²

We see research that employs AI to search social media posts and flag countries or regions where there is a risk of infodemic via the 'infodemic risk index' (IRI).²³ Other AI-research focusses on the dynamics and proliferation of hate speech on social media during this crisis.²⁴

¹³ Oped link

¹⁴ Municipality of Amsterdam '[Anderhalve meter monitor](#)' (1,5 meter monitor)

¹⁵ <https://analytics.ifisc.uib-csic.es/es/respuesta-covid-19/>

¹⁶ <https://api-nationalgeographic-com.cdn.ampproject.org/c/s/api.nationalgeographic.com/distribution/public/amp/science/2020/09/face-mask-recognition-has-arrived-for-coronavirus-better-or-worse-cvd>

¹⁷ <https://theconversation-com.cdn.ampproject.org/c/s/theconversation.com/amp/a-level-results-why-algorithms-get-things-so-wrong-and-what-we-can-do-to-fix-them-142879>

¹⁸ <https://www.vice.com/en/article/n7wxvd/students-are-rebelling-against-eye-tracking-exam-surveillance-tools>

¹⁹ <https://slate.com/technology/2020/05/workplace-surveillance-apps-coronavirus.html>

²⁰ Ibid

²¹ <https://www.who.int/teams/risk-communication/infodemic-management>

²² WHO and Rakuten Viber fight COVID-19 misinformation with interactive chatbot; 2020. <https://www.who.int/news-room/feature-stories/detail/who-and-rakuten-viber-fight-covid-19-misinformation-with-interactive-chatbot>.

²³ Gallotti R, Valle F, Castaldo N, Sacco P, Domenico MD. Assessing the risks of "infodemics" in response to COVID-19 epidemics. arXiv preprint arXiv:200403997. 2020

²⁴ Vel'asquez N, Leahy R, Restrepo NJ, Lupu Y, Sear R, Gabriel N, et al. Hate multiverse spreads malicious COVID-19 content online beyond individual platform control. arXiv preprint arXiv:200400673. 2020; Schild L, Ling C, Blackburn J, Stringhini G, Zhang Y, Zannettou S. "Go eat a bat, Chang!": An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19. arXiv preprint arXiv:200404046. 2020

3. What is Responsible AI in times of Corona?

While all these AI applications are deployed with the best intentions, many of them give rise to ethical, legal, societal and even effectivity concerns. This project looks at these concerns from a holistic and inclusive perspective. It assesses the applications and their impact through an iterative process, using a.o. the Ethics Guidelines for Trustworthy AI of the High Level Expert Group on AI²⁵ and "A socio-technical framework for digital contact tracing" by R. Vinueassa et al.²⁶

3.1 Ethics Guidelines for Trustworthy AI

The EU Ethics Guidelines for Trustworthy AI are the most granular and practical set of AI ethics principles to date. They are grounded in our fundamental rights and based on 3 overarching requirements: (i) socio-technical robustness, (ii) lawfulness and (iii) ethical alignment.

(i) Socio-technical robustness refers to the obligation to take into account the impact on society as a whole, the efficacy to tackle the problem at hand, the impact on human and societal behaviour, the interest of different stakeholders, precedent setting, etc.

(ii) Lawfulness refers to the fact that AI does not operate in a lawless world and that existing regulations are to be complied with. Lawfulness also includes adherence to human rights and respecting democracy and the rule of law.

(iii) Ethical alignment is translated into the '7 requirements for trustworthy AI' that have been endorsed by the European Commission: 1) Human Agency and Oversight; 2) Technical Robustness and Safety; 3) Transparency; 4) Privacy and Data Governance; 5) Inclusion, non-discrimination and Fairness; 6) Societal and Environmental Well-being and 7) Accountability.²⁷

If one takes a closer look the examples of AI deployed in this crisis, one could call into question the adherence of many of them with the above requirements. The UK grading algorithm conflicted with the ethical principles of transparency and fairness for example. Biometric recognition to assess mask wearing, social distancing or fatigue comes into conflict with the human right to a private life and the ethical requirement of human agency. For AI used in the medical field, in particular in diagnostics, there are questions of technical robustness

such as accuracy and reproducibility, in particular because of the lack of sufficient data on the virus and the disease.

Even AI applications that aim to combat the Infodemic, which could advance the freedom of expression, have shown to cause concerns when for example deployed with potential disregard for privacy.²⁸

3.2 Tensions, trade-offs and conditions in times of crisis

Under normal circumstances, tensions can rise between these requirements, that necessitate a careful balancing of interests and sometimes lead to the trade-off of one requirement against the other. The Ethics Guidelines provide guidance for such a 'balancing exercise':

- The benefits of AI must outweigh the individual and collective risks;
- Particular attention should be paid to vulnerable groups such as children;
- The fact that AI carries risks that are difficult to predict should be recognised;
- Adequate and proportional measures should be taken to prevent or reduce those risks;
- There may be situations where there is no ethically acceptable trade-off and the application should not be used.

The current extraordinary circumstances may lead to different tensions, new or different interest, a different balancing of those interests and thus to trade-offs that would perhaps not be acceptable under normal circumstances. This project aims to provide guidance on how to perform this 'balancing exercise' under these extreme circumstances.

4. Project instruments

An AI & Corona Observatory will provide insight in the types of AI that are being developed and/or deployed to tackle the various challenges of this crisis. It will highlight responsible, useful, reliable and safe AI, while at the same time draw attention to unsafe, irresponsible or undesirable AI. It will also attempt to provide insight in the scope of use, the impact domains and the (groups of) people impacted by it. Our experts from various fields will reflect on highlighted applications from their specific area of expertise. We will raise awareness through expert sessions for policy makers and health care professionals and through (social)media activities.

²⁵ <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

²⁶ R. Vinueasa, A. Theodorou, M. Battaglinie, V. Dignum, 2020

²⁷ Ibid

²⁸ <https://www.nrc.nl/nieuws/2020/11/15/leger-verzamelde-data-in-nederland-a4020180>

We will develop an evaluation framework with a number of criteria for Responsible AI in times of Corona. This framework assess AI from the technological perspective, the societal perspective and the governance perspective. These criteria will help organisations perform a 'quicksan' of the AI-application they want to develop, procure or deploy to tackle a certain challenge in this crisis. The quickscan is a self-assessment tool to quickly identify the relevant elements of responsible AI and the level of adherence to these elements. It will help determine the level of impact of the AI-application, and provide options to balance different tensions and interests.

It should be noted that these criteria will not be meant to be used as a 'checklist'. Not in the least because the categories are interlinked and can serve as corresponding vessels. For example, a very effective application could also be highly impactful, being reason not to use the application. Under the extreme circumstances of this crisis however, the right governance measures (e.g. voluntariness, a sunset clause, a controlled environment) might tip the balance in favour of exceptional use of the application. In contrast, an application with a low level of

efficacy could nevertheless be interesting to deploy for research purposes, if it poses no ethical, legal and societal risks.

The quickscan framework will be piloted with a small number of parties that deploy or intend to deploy AI to tackle a certain Corona related challenge. The piloting phase will result in a definitive framework for quickscans that can also serve as a basis for quickscans for AI used under normal circumstances.

Project details

Project webpage: www.allai.nl/corona

Project Lead: Catelijne Muller, LLM.

Scientific Lead: Professor Virginia Dignum

Contact: welkom@allai.nl

Acknowledgements

ALLAI received funding from SIDNFonds for the Project Responsible AI & Corona.