Long-term Research Strategy for

Artificial Intelligence and Ethics at the Police

June 2019

I. AI and Ethics at the Police

1. AI at the Police

AI has many potentially beneficial applications in law enforcement including predictive policing, automated monitoring, (pre-) processing large amounts of data (e.g., image recognition from confiscated digital devices, police reports or digitized cold cases), finding case-relevant information to aid investigation and prosecution, providing more user-friendly services for civilians (e.g. with interactive forms or chatbots), and generally enhancing productivity and paperless workflows. AI can be used to promote core societal values central to police operations (human dignity, freedom, equality, solidarity, democracy, and the rule of law), but values carefully guarded in existing operations and procedures may be challenged by the use of AI.

It is impossible to anticipate all the effects of the use of AI in society, and more specifically in the law enforcement domain. Therefore, it is essential that adoption and use of any application be continuously evaluated, in order for the Dutch police to ensure policing practices in line with the values acknowledged by the Dutch state and the European Union. As identified in the whitepaper *"AI & Ethics at the Police: Towards Responsible use of Artificial Intelligence in the Dutch Police"* (hereafter *Whitepaper*), treating the introduction of new technologies as a social experiment could be helpful in this respect. [1] This gives the police organization the responsibility to consistently reflect and assess every application of AI from the first pilot phase throughout the life-cycle.

2. On the Law and Ethics

Similar to other authorities of the state, the police necessarily operate within a specific legal framework. This framework includes but is not limited to preventing misuse of powers, conflicts of interest and discrimination, and is ensured through active accountability measures. For example, police actions can be challenged and reviewed by the independent judiciary.

The police organization in the Netherlands is committed to protect *fundamental human rights* and to ensure respect for the rule of law. The police are directly obliged to comply with domestic and international legal instruments that specify this commitment. These legal requirements apply to all police work regardless of the means used and thus include the use of AI.

However, the application of AI raises some challenges that are not—or it is unclear if they are—covered by current legal provisions. As identified in the Whitepaper, the opacity of reasoning that is inherent to some AI techniques might decrease transparency and weaken human agency in the police's decision-making, and thereby pose a threat to the legitimacy of and trust in the police. Therefore, for the spaces left open by the law, the police should incorporate ethical considerations through practical measures to ensure responsible use of AI and contribute towards enhancing (rather than limiting) legitimacy of and trust in the police.

Such ethical considerations should stem from the ethical principles and values that statutory law aims to uphold, and these can be found in the fundamental rights framework. In the Whitepaper, we identified six morally salient requirements for the Dutch police:

1. Accountability;
2. Transparency;
3. Privacy & data protection;
4. Fairness & inclusivity;
5. Human autonomy & agency; and
6. (Socio-technical) robustness and safety.

It is important to note, that while AI can sometimes replace human labor we should be careful not to attribute any human characteristics to it—notably we cannot have "responsible AI" but the police must be responsible in their use of AI.

II. Research Strategy Ethics & AI

We emphasize that further research is essential for the police to explore ways to realize their goals of increasing (a) efficacy and efficiency on the one hand, and (b) trust and trustworthiness on the other (to boost public trust and the perception of the legitimacy of the police). With these goals in mind, we identify the following research directions on AI and Ethics at the Police, divided into tracks for (1) impact on humans, (2) organizational embedding, and (3) technical work:

1. Impacts on human beings

1.1. Impacts on human dignity

Human dignity is the inviolable value upon which the human rights framework rests. It illustrates the fundamental belief in the intrinsic worth of a human being, protecting his/her autonomy and self-determination. Belief in human dignity can be understood as the raison d'être for the law the police aims to enforce.

Example RQ: In which ways can the automation of police tasks threaten or support human dignity?

1.2. Public trust

Public perception of the legitimacy of the police and subsequent trust is as important as the legal framework in which the police operate. While automation and prediction to some extent increase efficacy of the police, the study could explore if such increase in potency is desirable from the societal perspective.

Example RQ: How do perceptions of AI usage affect public opinion w.r.t.

competence, fairness, trust and legitimacy of the Dutch police?

2. Impacts on the police organization

2.1. Ethics guidelines and oversight

The police does not operate in isolation, and the use of AI takes place across the entire judicial chain: OM, local government, the Ministry of Justice and Security, judiciary. Responsible use of AI within the Dutch police ideally follows from a robust ethics framework for the entire chain. Such a framework can establish criteria to follow throughout the AI development and application cycle. The criteria should afford evaluating the introduction of applications in the light of the morally salient requirements on a case-by-case basis, irrespective of whether AI is automating existing workflows or providing entirely new functionality. Furthermore, proper oversight on the broader ethical impacts of the development and application of AI should be put in place.

Example RQ1: Which concrete guidelines can the Dutch police organization adopt to satisfy the assessment criteria for trustworthy AI provided by the European Commission's High-Level Expert Group on AI? [2]

Example RQ2: What are necessary and sufficient criteria for the oversight on responsible use of AI within the Dutch Police?

2.2. Impacts on police personnel

AI can be used to support the police organization in achieving its goals of efficiency, traceability, uniformity and integrity. However, the change of operations may come with displacement of employees and changing roles. Research is required to ensure that workers with non-traditional skillsets fit into the police organization in a way that empowers police personnel.

Example RQ: How can AI tools assist personnel in making more uniform judgments without corroding their agency?

3. Technical aspects

3.1. Explainable AI

The aforementioned oversight can only be adequate and meaningful if automated decisions can be explained and justified on the technical level.

Example RQ: What explainable AI solutions (e.g. argumentation [3], LIME [4], counterfactual explanations [5]) make automated decisions adequately contestable, fair and accurate for a particular application (e.g. risk assessment)?

3.2. Justifiable/verifiable AI

Justification provides the reasons behind the results and the choices for particular approaches. Mathematical tools for formal verification make AI systems themselves and their decisions reviewable.

Example RQ: What are useful justifications of, e.g. predictive policing applications?

III. Organizational embedding

While the Dutch National Police are actively working on developing AI that would enhance their capabilities, they also recognize the importance of aligning police work with the ethical considerations and requirements for responsible use of AI.

The National Police Lab AI (AI Lab) is a collaborative initiative of the Dutch Police, with two Dutch universities, that works on developing state-of-the-art AI techniques to maintain safety in the Netherlands in a socially, legally and ethically responsible way. AI use by the Dutch Police is not limited to the applications explored at the AI lab, and neither are the topics in this paper. We advise that the police develop a focal point for dealing with ethical considerations around the application of AI; for example, by connecting all applications of AI to the AI Lab.

Funding for the research described in this report should not necessarily (or even: preferably) come from internal budgets of the police alone. For collaboration with other parties in the judicial chain, such as ministries, budgets could be combined. It is also important to preserve independence of the research by looking for external funding.

Funding opportunities include but are not limited to national funding schemes, new calls for which may be announced in the future: e.g. Digital Society, Dutch National Research Agenda (NWA), Responsible Innovation (MVI) and other future AI research calls by The Dutch Research Council (NWO). Funding opportunities may also open through research initiatives from the Ministries (in particular the Ministry of Justice and Security — SAPAI), other institutions (e.g. Netherlands Standardization Institute — NEN) and/or supra-national organizations or bodies (in particular the European Union).

The Dutch police may also look for research collaborations with other countries (e.g. Sweden, Australia as there are contacts in the police as well as academia). Such collaboration can for example take the form of comparative studies on AI and Ethics in different jurisdictions.

Finally, in the realization of this long-term research strategy, collaborations with external parties are essential, as it provides external validation of police practices when it comes to the complex subject matter.

Dr. Francien Dechesne
Assistant Professor at Leiden University
f.dechesne@law.leidenuniv.nl

Dr. Virginia Dignum
Associate Professor at TU Delft
M.V.Dignum@tudelft.nl

Jordi Bieger, MSc
Researcher at TU Delft
J.E.Bieger@tudelft.nl

Lexo Zardiashvili, LLM
Researcher at Leiden University
a.zardiashvili@law.leidenuniv.nl

This document is a deliverable of research conducted by the Leiden University Center for Law and Digital Technologies (eLaw) and TU Delft Design For Values Institute, commissioned by the Dutch National Police.

References

[1] F. Dechesne, V. Dignum, A. Zardiashvili and J. Bieger, "AI and Ethics at the Police: Towards Responsible Use of Artificial Intelligence at the Dutch Police," 2019.

[2] High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy Artificial Intelligence (AI)," European Commission, Brussels, 2019.

[3] F. J. Bex, Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory, Dordrecht: Springer Netherlands, 2011.

[4] M. T. Ribeiro, S. Singh and C. Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 2016.

[5] S. Wachter, B. Mittelstadt and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and The GDPR," *Harvard Journal of Law & Technology,* vol. 31, no. 2, 2018.