D1.2a: KOBAN Framework

Prof. dr. mr. Marc Schuilenburg & Martijn Wessels MSc.

Erasmus University Rotterdam

Date: 25-11-24

Version: 0.1

Abstract

This task (T1.2) has the objective to carry out extensive desk-based research into the current models, frameworks and tools relating to responsible development and deployment of technologies and applications for community policing in EU-countries. A first version of a general KOBAN Framework is developed for assessing societal, legal, cultural, ethical and gender aspects for responsible technological applications for community policing. The KOBAN Framework contains sets of public values that are drawn from a systematic analysis of documents relating to the use of AI tools and technologies by police organizations. An elaborate literature search was conducted on aspects of data protection, accountability, human factor and privacy as integral considerations of the KOBAN project and the to be developed AI-driven solutions.

The KOBAN Framework offers an overview of the ethical, societal, legal, cultural and gender aspects relating to the development and use of data-driven and AI tools in community policing. Three overarching sets of public values are described: driving, underpinning and procedural values.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

Set 1: Driving Values
Public Safety & Security
Effectiveness & Efficiency
Set 2: Underpinning Values
Data Bias & Fairness
Privacy & Mass Surveillance
Cultural & Gender Sensitivity
Digital & Environmental Inclusion
Set 3: Procedural Values
Accountability
Transparency & Explainability
Human Oversight

This deliverable (D1.2a) contains the first version of the KOBAN Framework. Based on the findings of the pilots in the EU-countries and implementing it in all other work packages, the framework will be updated to ensure a careful and responsible way of designing and dealing with AI-led solutions in community policing. The experiences with the KOBAN framework will be compiled and lead to a new revision of the framework (D1.2b).

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

Table of contents

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

List of Tables

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

1. INTRODUCTION

Although the concept of 'community policing' is not applied in the same way in all European countries, its underlying idea is to bring the police – in a physical sense – closer to the population and create support and legitimacy for police-work. Other important objectives of community policing are small scale, problem-orientation and cooperation with public and private partners in tackling public safety and improving the liveability of neighbourhoods.

Several social developments have put pressure on the concept of community policing, from a more diverse population in European Union (EU)-countries and changed views on the role of the police to the global trend of digitalisation. The digitalisation of society, for instance, has widened the playing field of crime from a physical to an online environment. Many times, this does involve an intertwining of the physical and digital worlds. At the same time, digitalisation also provides opportunities for the police to strengthen community policing. The introduction of web care teams and digital agents, an AI chatbot that 'speaks' with people who want to report a crime, is an actual example of this. Another example is the use of Artificial Intelligence (AI) tools to analyse large volumes of body camera footage, provide insights into interactions between police officers and citizens, and contribute to better training and policymaking.

One of the ambitions for community policing is to remain up to date on the latest digital trends and forms for building trust within communities, while also encouraging community engagement and communication with local police forces. As such, an important goal of KOBAN is to develop data-driven solutions in order to build trust and strengthen community engagement for community policing. One of the challenges of the project is to define what role AI can play in support of community policing and which relevant values and considerations must be taken into account. In community policing, relevant goals are not only to improve security and efficiency; but also, to not encroach upon but rather support and strengthen other public goals such as maintaining accountability in policing and ensuring equal treatment of citizens. This KOBAN Framework provides an overview of the public values relating to the use of AI tools and technology for community policing.

1.1 OBJECTIVE TASK 1.2

This task has the objective to carry out extensive desk-based research into the current models, frameworks and tools relating to responsible development and deployment of technologies and applications for community policing in EU-countries. Currently available and upcoming (policy-)documents and guidelines on ethical, legal, and societal aspects for tools and technologies for 'responsible community policing' from around Europe will be reviewed and compared to identify policy trends and gaps.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

1.2 SCOPE

The development of the KOBAN Framework focuses primarily on the ethical, societal, legal, cultural and gender deployment of AI tools for the purpose of community policing. Broadly, three sets of public values are discerned: driving, underpinning and procedural values. These public values – as described in this document – must at all times be weighed in a local social, cultural and legal context and translated into the design and use of AI tools and technology.

In a legal context, there are a number of important categories of legislation surrounding the application of AI tools by police organizations. These include i) the European primary laws drawn up (e.g., Charter of Fundamental Rights), ii) secondary laws which include, for example, the Law Enforcement Directive, and the recently released AI Act, iii) the UN human rights treaties, and iv) legislation of individual member states, relating to national laws implementing the Law Enforcement Directive, as well as legislation on policing powers and criminal procedures.

These instruments provide specific rules relating to the use of AI by Law Enforcement Agencies (LEAs). The Law Enforcement Directive and its national implementations cover the data protection angle. The AI Act contains a significant number of rules, exceptions and specific provisions related to the law enforcement use of certain AI applications. Other applications are not regulated, or are only subject to transparency requirements, and often there are specific exceptions for LEAs. KOBAN will ensure to help educate LEAs on the exact scope of these rules and obligations.

In addition to the AI Act and future AI national legislation, other guidance (e.g., governmental and/or organizational) may impact a LEA-policy on the use of AI. Even when certain rules or guidance are not binding, these will still be important input to develop AI that fits the norms and values of a certain EU-country. See for instance the vision presented by the Dutch government on the use of generative AI (Government of the Netherlands, 2024), or Interpol's 'Report On Artificial Intelligence For Law Enforcement' (Unicri & Interpol, 2020).

This layering of different types of rules, laws and regulations, and guidance, underscores the importance of looking closely at the local context within KOBAN pilots and determining which rules and guidance are relevant. Especially in situations where there is collaboration with other organizations or citizens for community policing, different legislation could become intertwined. Hence, the legal context must be assessed locally to ensure compliance. Subsequently, ethical, societal, cultural and gender considerations must be made on the various dimensions that will be presented in the proposed KOBAN Framework.

Description Deliverable 1.2a: KOBAN Framework

This task intends to develop a general KOBAN Framework for assessing societal, legal, cultural, ethical and gender aspects for responsible technological applications for community policing. The KOBAN Framework is presented in the current deliverable. In line with the scope of the

framework (see also: *scope*), this assessment list does not provide any detailed advice on ensuring national legal compliance, but offers mainly guidance on meeting the second and third components of 'trustworthy AI' ('ethical' and 'robust AI').

An update to the KOBAN framework is expected to be necessary during the project's lifetime. When this framework is used to assess the ethical, societal, legal, cultural and gender aspects of the different KOBAN pilots and in other KOBAN work packages, it is expected that gained insights can lead to a revision of the framework to make the framework more applicable and complete. In this way, the other pilots and the most up-to-date version of the framework can be used for future pilots (see: concluding remarks & future revision, for more details).

1.3 RELATION TO OTHER KOBAN WORK-PACKAGES (WP) AND TASKS (T)

The main function of the KOBAN Framework is to help directing the ethical, legal, societal, cultural and gender deliberations when developing and applying the community policing tools for the different KOBAN pilots. The KOBAN framework aims to contribute to Key Objective 2 of KOBAN: *to deliver adapted co-designed and state-of-the-art tools, methods and CP [community policing] solutions which are tailored to the stakeholders' needs and compliant with societal, legal, cultural, ethical and gender aspects*. Herein, the KOBAN framework is a result of one of the efforts performed in T1.2. Together with the other activities performed within T1.2, the *Working across Cultures workshop* and the *Rapid Ethical Deliberation workshop,* this framework aims to promote the awareness regarding ethical, societal, cultural and gender dimensions within the KOBAN consortium, and help direct the responsible development and use of community policing tools within the different pilot-contexts.

The KOBAN framework can assist in directing with the development of community policing solutions (WP3: Development of solutions for effective future-proof community policing). It can help with the development of the KOBAN app factory (T3.1), KOBAN AI Assistant (T3.2), AI-driven early detection and risk project of security threats (T3.3), Multifaced intelligent dashboard for CP (T3.4), and the KOBAN Knowledge enrichment and management (T3.5) solutions.

Furthermore, the KOBAN framework can also aid in the problem definition of the six KOBAN pilots and in designing the evaluation framework to incorporate ethical, legal, societal, cultural and gender dimensions (WP4). Within the monitoring and reflective assessment of these dimensions, feedback will be provided after each monitoring exercise, including an assessment overview, findings, and recommendations (if needed) (T4.3).

2. METHOD: DESK RESEARCH

The KOBAN Framework and sets of public values drawn from the extensive and systematic analysis of documents relating to the use of AI tools and technologies by police organizations. Between September 2024 and December 2024, an elaborate literature search was conducted on aspects of data protection, accountability, human factor and privacy as integral considerations of the KOBAN project and the to be developed AI-driven solutions. The literature review consisted of four phases.

The first stage sought to retrieve an overview of relevant EU legislation, AI frameworks, and (national) reports on the use of AI. The frameworks were gathered by exploratory queries in Google and Google Scholar (e.g., 'AI ethical framework', 'ethical framework policing'). These were combined with legislation, frameworks and reports the researchers and the KOBAN consortium were familiar with (e.g., frameworks developed in previous Horizon Europe projects). This led to an overview from which public values were distilled and categorized in different groups. Our preliminary conclusion was that there are many different hierarchies and groupings of public values in the frameworks, consisting of different sub-values. Therefore, we decided to make a distinction between three different sets of public values: driving, underpinning and process-based values (WRR, 2011). Driving values support the use of technological tools, whilst underpinning values should be preserved, maintained, and ideally enhanced through the use of the technology. The last set are procedural values that enable and assist the balancing of both sets of values.

Table 1 provides an overview of the frameworks, reports and legal documents used for our initial exploratory analysis. It gives a general rather than an exhaustive overview of key public values in the mentioned frameworks. Where a public value is described in the framework as an independent category, it is included in Table 1.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

Table 1 – Overview reviewed frameworks

			Driving Values		**Under-pinning values**				**Procedural values**		
			Security & Public Safety	Effectiveness & Efficiency	Data Bias & Fairness	Privacy & Mass surveillance	Cultural & Gender Sensitivity	Digital & environmental	Accountability	Transparency & Explainability	Human oversight
Title	**Document type**	**Source type**									
AI and policing: the benefits of artificial intelligence for law enforcement (Europol, 2024)	Report	LEA			x	x			x	x	
Big Data Ethics (Ryan & Inguanzo, 2021)	Paper	Sci.			x	x		x	x	x	
Ethical Sensitivity Tools: Social ReadinessTool and Ethical Role-playing Tool (Francis et al., 2023)	Framework	Sci.			x	x		x	x		x
Legal and Ethical Framework and Risk analysis (Casanovas et al., 2021)	Framework	Sci.			x	x		x	x	x	x
AI Ethics Framework: Enabling principles (Victoria Police, 2024)	Framework	LEA			x	x			x	x	x
Ethics, Transparency and Accountability Framework for Automated Decision-Making (UK Government, 2021)	Framework	Gov.			x				x	x	x
A guide to using artificial intelligence in the public sector (UK Government, 2020)	Framework	Gov.		x	x				x	x	
Data Ethics Framework (Victoria Police, 2024)	Framework	Gov.			x			x	x	x	
D2.4:Ethical frameworks for the use of LEAs	Framework	Sci.		x	x	x		x	x	x	x
D4.7: An Ethical framework for the development and use of AI androbotics technologies (Brey et al., 2020)	Framework	Sci.			x	x		x	x	x	
A guide on assessing unintended societal impacts of different CM functions - Version 2 (Driver+, 2019)	Framework	Sci.				x	x	x	x	x	
Ethics guidelines for trustworthy AI (HLEG, 2019)	Framework	Gov.			x	x		x	x	x	
The Artificial Intelligence Act	Legal	Gov.		x	x	x	x	x	x	x	x

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

Current and emerging trends in the use of AI for community surveillance (Salgado, 2024)	Paper	Sci.				x		x		x	
The Ethics of AI Ethics: An Evaluation of Guidelines	Paper	Sci.			x	x	x	x	x	x	x
Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance	Paper	Sci.			x	x	x	x	x	x	x
AI & Ethics at the Police (Dechesne et al., 2019)	Paper	Sci.			x	x		x	x	x	x
Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (United States White House, 2023)	Legal	Gov.	x	x	x	x	x	x	x	x	x

The second stage of the desk research was to gain an overview of the state of the art of academic literature. An advance search query was performed in Scopus and Web of Science by entering keywords, including: AI, public values, data justice, algorithms, big data, ELSA and community policing. These keywords were used in various combinations in order to gain more insight in the available literature on public values in the context of AI tools and technologies for community policing.

The third stage consisted of a selection of relevant documents for the KOBAN Framework. A publication was only selected if it mentioned terms as 'public values' or affiliated terms, as well as one of the other keywords (e.g., 'data justice', 'ELSA', 'community policing', 'public safety', 'Law Enforcement Agencies') in the title, abstract or keywords. This was necessary to limit the literature search to publications that are directly relevant to the ethical, societal, cultural and gender challenges of developing and implementing AI tools and technologies for community policing. We analysed these documents qualitatively to identify how the publications conceptualize the issue of public values in the context of AI-tools and technologies for community policing.

To gain a deeper understanding of the different public values, it was necessary to work with new search terms – the fourth and final stage of the literature review. In the fourth stage, we were solely interested in publications that mention one or more of the public values within the three overarching sets. In Google Scholar, the following combinations of keywords were used during this phase, including: 'security', 'effectiveness', 'bias', 'fairness', 'privacy', 'human factor', 'accountability', and 'transparency'. Through bibliographic snowballing, the additional search terms led to another potentially relevant set of publications. Subsequently, a full text analysis was conducted in order to gain a better understanding of the identified public values in relation to AI tools and technologies for community policing and the – to be developed – AI-led solutions in the six projects in the EU-countries.

3. KOBAN FRAMEWORK

The KOBAN Framework offers an overview of the ethical, societal, cultural and gender aspects relating to the development and use of AI in community policing. Three overarching sets of public values are described: driving, underpinning and procedural values. In what follows, we identify these key values and unpack their relevance and implications for community policing. Importantly, each value should be applied to KOBAN – even if it's just a statement of how the project will take care to not let the six pilots in the EU-countries breach these public values.

Table 2 - Overview KOBAN Framework

Set 1: Driving Values
Public Safety & Security
Effectiveness & Efficiency
Set 2: Underpinning Values
Data Bias & Fairness
Privacy & Mass Surveillance
Cultural & Gender Sensitivity
Digital & Environmental Inclusion
Set 3: Procedural Values
Accountability
Transparency & Explainability
Human Oversight

3.1 SET 1: DRIVING VALUES

Public Security & Safety

'Making neighbourhoods safe and secure' is one of the goals when it comes to community policing. Creating a safe society refers to the central role of the police, where surveillance by police officers in neighbourhoods and creating partnerships with community members, for example, should contribute to local safety and security. In community policing, AI can be used to improve the prevention and detection of crime and disorder, but it can also be a tool for helping communities to communicate more effectively with local police forces.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

When using AI to improve public safety, it should always be considered whether the AI tool 'works' (*what works*) and whether the way it works (*how it works*) is commensurate with the results. The simple fact that something is technologically possible does not make it immediately desirable. Herein, anticipation is necessary: early identification of the impact of an AI tool and the potential ethical, legal and social harms it may lead in practice (Forsberg, 2014; Steen, 2022). It should be considered what the accuracy of the tool is, whether this is sufficient and justified, and what the acceptable risk is in regards to using the tool in combination with the expected benefits (Sorell, 2024).

In each of the six pilots in the EU-countries, questions that must be asked are:

- What is the added value of a specific AI tool for community policing?
- Is it benefit/added value of using an AI tool greater than using other types of applications – or no application at all?
- What are the potential negative impacts of the use of the AI tool?
- Are the potential negative impacts larger when using AI than when not using AI?
- If the benefit is larger when using AI, but there are additional risks, can these be appropriately addressed by mitigating measures and are these measures feasible?
- Do the potential benefits of using a specific AI tool outweigh the potential negative impact that it could create, taking into account all the safeguards and mitigating measures that will be implemented?

EFFECTIVENESS & EFFICIENCY

AI tools and technology can be developed to improve the effectiveness and efficiency of public administration and public services (art. 59 AI Act). 'Effectiveness' is about the ability to meet pre-determined targets. 'Efficiency' concerns the attainment of a specified target using the fewest possible resources (WRR, 2011: 68). Effectiveness and efficiency have always been important in policing, but they have come to play an increasingly important role in processes of digitalization of police-work. As such, effectiveness and efficiency are also increasingly being advocated and deployed in connection with community policing and finding new ways to improve support and legitimacy for police-work, e.g., by automating and streamlining the documentation and reporting process of police officers in order to reduce their workload by saving them time and effort. Here, an important aspect is to be financially cost-efficient with public funds. The costs incurred for procurement/development, deployment and management must be transparent and properly justified by police organisations (UK Government, 2020).

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

3.2 SET 2: UNDERPINNING VALUES

Data Bias & Fairness

Data is not without its own bias, a concern which relates to the importance of ensuring high-quality data and access to high-quality data for community policing and criminal prosecution purposes in particular. This is the issue of 'dirty data'. In the broadest sense, the term 'dirty data' refers to 'missing data, wrong data, and non-standard representations of the same data' (Richardson, Schultz & Crawford, 2019). If 'dirty data' enters the used AI tools and technologies for community policing, the results will also be contaminated.

According to the AI Act, 'data sets for training, validation and testing, including the labels, should be relevant, sufficiently representative, and to the best extent possible free of errors and complete in view of the intended purpose of the system' (recital 67 AI Act). Since data is the foundation of any AI system (advice or decision) that is used in the context of community policing, data that in itself is 'incorrect' or 'derived from or influenced by corrupt, biased, and unlawful practices' (Richardson, Schultz & Crawford, 2019) can lead to an output in the form of decisions or recommendations of actions that are also problematic. This can have a severe negative impact on people's lives, such as discrimination against minorities, stigmatisation and 'overpolicing' as a result of (hidden) bias and feedback loops (e.g., Harris & Burke, 2021; Ferguson, 2017; Benjamin, 2019; Schuilenburg, 2024).

In addition to data used to train, validate and test AI applications, sometimes LEAs may have control over the datasets they use as input for their AI tools. A good example is an AI tool that recommends routes for patrols, where the tool provider will require police-forces using the tool to populate the tool with relevant data sources, e.g., historical data, demographic data and other relevant factors. The AI Act requires LEAs to make sure that also this input-data is relevant and sufficiently representative to avoid biased outcomes.

Biases can exist in many shapes and forms. Bias can be inherent in datasets, for example when 'historical data' is being used. Bias can also occur from the way a population is defined and sampled to create a dataset ('representation bias'). Biases can also occur as hidden proxies, even when other measures are taken to negate the use of specific indicators (Donatz-Fest, 2024). It should not be forgotten that even if an AI application uses 'clean data', professionals, too, have unconscious prejudices that can affect their actions and decisions in concrete situations. For example, when police officers decide whether to accept or disregard a model's prediction, they can seek and select information that supports existing beliefs, while paying less attention to information that contradicts it ('confirmation bias'). This also can lead to erroneous decisions and run into issues with fairness (see also: *human oversight*).

In theory, there are – at least – two filters for keeping 'dirty data' out of data-driven analyses in community policing: at the front-end and at the back-end of the process. At the front-end, in the design of the AI tool by identifying 'dirty data' and removing it from the datasets to be used, including transparently document which datasets are used by the police organisation – and why. It also implies that the data-ontologies ('vocabularies') used for algorithmic analyses are kept up to date (Ghioni, Taddeo, Floridi, 2024). At the back-end of the process, it might be

down to the judge to decide that the investigative authorities, for example, have obtained the data illegitimately or have used dirty data.

The question remains whether both filters – front-end and back-end – are actually sufficient to keep bias out of the LEA's datasets and offer citizens effective protection against the risk of dirty data will make its way into analyses of police organisations.

PRIVACY & MASS SURVEILLANCE

The use of technology, and in particular data-driven AI tools by police may at times have an impact on the privacy of citizens. A balance must always be found between security and law enforcement goals, which necessitate that privacy is encroached upon to a certain extent, and the reasonable privacy expectations of citizens, not only in their own home, but also in public spaces (including: online), as per the consistent case law of the Court of Human Rights and the Court of Justice of the European Union.

Certain technologies are in particular capable of being abused to encroach upon privacy more than is legally permitted, or more than is societally desirable. Such tools, in its most extreme form even amounting to mass surveillance, can hence be in breach of fundamental rights of local citizens, including in particular the right to privacy. Examples of technologies that present such risk for mass surveillance if not designed and used appropriately are diverse, ranging from evident examples like CCTV monitoring and live facial recognition, and email interceptions, but also covering many other use cases, such as tools for legal hacking, forensic data acquisition and analysis (e.g. of a smartphone that contains an individual's whole life), the use of extensive DNA banks and Passenger Name Recognition systems, to AI tools used for crowd control, and other forms of data-intensive tools.

A specific category of data-driven – and often AI-driven – tools that are available to LEAs are open-source intelligence applications (OSINT), namely tools that use freely accessible sources of data (typically from the internet) to help LEAs with intelligence and investigations. Some authors have indicated that, if used incorrectly, certain OSINT tools can present a threat of social control through state surveillance (Ghioni, Taddeo & Floridi, 2024). Generally, when using AI tools that may limit an individual's privacy, the applicable legal framework must be kept in mind throughout the application's life cycle. This includes the national legal framework relating to policing powers, data protection and criminal procedure, as well as other relevant laws, such as the AI Act and the EU Charter of Fundamental Rights.

According to the AI Act, the right to privacy and the protection of personal data must be guaranteed throughout the entire lifecycle of an AI system (recital 69 AI Act). Reiterating the principles also found in the Law Enforcement Directive, this means that, when dealing with personal data, the principles of 'data minimisation' and 'data protection' by design and by default must be applied when personal data are processed. So, while AI needs data to function, it cannot be a justification to gather and process data without limit. LEAs must here, as in other activities, determine to what extent data can be gathered without being excessive. After all, the Law Enforcement Directive does allow LEAs to go slightly beyond what the GDPR requires (where data must be 'necessary,' rather than 'not excessive' in relation to the goal

pursued). Depending on the type of data, additional rules may apply, which is specifically the case for special categories of data, e.g., biometric data (Ng et al., 2023).

Data protection requires the integration of technical and organisational measures into the processing activities and daily practices, from the design stage right through the lifecycle of an AI tool, in such a way that safeguards privacy and data protection principles right from the start ('data protection by design'). 'By default' means that police organisations should ensure that personal data is processed with the highest privacy protection, so that by default personal data is not made accessible to an indefinite number of persons (art. 20 LED). There should be found a right fit between the AI tool and data-sets used, in respect to the required data reduction techniques (e.g., Khoei & Singh, 2024).

Violating the right to privacy can have chilling effects, meaning that that citizens do not feel completely free and behave differently when they believe they are being watched, regardless of whether this is actually the case (Murray et al., 2023). As such, the chilling potential of facial recognition vis-à-vis protest rights is acknowledged by the European Court of Human Rights (Glukhin v. Russia, 2023). Recent empirical research of protest policing shows that chilling can also lead to feelings of hyper-transparency and hyper-alertness by observed and observers (Storbeck et al., 2024). This can lead, among others, to less trust in authorities by citizens, including the police, which can lead to resistance of citizen to community policing (and thereby reducing its effectiveness).

CULTURAL & GENDER SENSITIVITY

Community policing requires that policing practices fit the local customs, culture and norms and values. Due to geographic and socio-economic differences between and in the pilots in the European countries, this also requires local approaches towards developing and using AI tools for community policing. Throughout the lifecycle of an AI tool, there should be special attention towards potential cultural and gender biases that might occur. Here, there are many different challenges and risks in respect to diversity and inclusion, ranging from a lack of minority groups being involved in the lifecycle of the tool to, using stereotypical or inaccurate and incomplete gender concepts (e.g., sex instead of gender), or homogenous development teams (Fosch-Villaronga & Poulsen, 2022; HLEG, 2019; Shams, Zowghi & Bano, 2023) (see also: *digital & environmental inclusion*).

There should be explicit attention for identifying these potential risks within the specific context of the community policing applications in which they are being used. Especially due to local differences, generic applications are not necessary immediately suitable for the socio-economic and cultural context of the target communities in the EU-countries. For instance, the use of a robot dog by the New York Police Department led to high concerns within the community in regard to the militarisation of the Law Enforcement Agency (Harris & Burke, 2021). Both technical, procedural and social measures need to be considered to mitigate risks in developing inclusive solutions (Shams, Zowghi, Bano, 2024). Technical measures and methods can be directed to mitigate the risks of biases, whilst other measures are directed towards creating development and evaluation settings in which diversity and inclusion gain

explicit attention. For instance, by making police officers aware of their cognitive biases (see also: *data bias & fairness*), but also establish a development process in which relevant local communities and organisations are involved and represented.

DIGITAL & ENVIRONMENTAL INCLUSION

Digital inclusion is an important public value when designing AI tools and technology for community policing. Digital inclusion means involving different forms of knowledge ('epistemic inclusion') and layers of the population ('social inclusion') in the design and implementation of new tools and technology (see also: *cultural & gender sensitivity*). This will partly depend on the type of AI application and the purpose for which it is used. But from the perspective of transparency, for example, it is conceivable to include as diverse and inclusive a team as possible in terms of gender, age and background when designing new AI applications (Schuilenburg, 2024).

These forms of inclusion are also beneficial for drafting AI regulations (Cath, 2018). In this way, a voice can be given to all those who do not yet have or do not have enough in this area, for example minorities, elderly, or people with vulnerabilities. Efforts to include stakeholders can also shed light on perspectives and values that otherwise remain unexplored by conceptual, empirical and technical investigations (Friedman et al., 2013; Aizenberg & van den Hoven, 2020). Moreover, it also forces designers to better empathise with how these groups think about deploying smart sensor applications in neighbourhoods for increased safety, for example (TNO, 2021).

Related to the issue of environmental well-being, the *Ethics Guidelines for Trustworthy AI* speaks of the principle of prevention of harm: 'AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings.' (HLEG, 2019) Here, it is possible to go one step further and include non-human entities such as animals, plants and forests as stakeholders in the design of new technologies. The energy consumption of AI is growing rapidly and has a major impact on our climate. About six per cent of global CO2 emissions come from data centres, using smart devices and training self-learning algorithms. Social scientists (e.g., green criminologists) can play an important role here, in the sense that they act as representatives of the interests of these 'speechless' actors and speak more or less on their behalf when designing responsible AI tools and technology for community policing (Latour, 2017; Schuilenburg & Peeters, 2024).

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

3.3 SET 3: PROCESS-BASED VALUES

Accountability

Fundamental to the legitimacy of LEAs is their ability to account for their practices to relevant stakeholders. Accountability is a relational concept in which an actor needs to explain and justify their conduct to a forum (Bovens, 2007). This requires organisations to establish adequate policies and transparent processes to cater to the questions and needs from these different stakeholders. AI tools used by police organizations are posing risks and challenges to their accountability and transparency of policing practices and decisions. Accountability for community policing practices requires traceability of the decisions and actions performed, in which relevant stakeholders are able to ask for explanation and are able to redress decisions made.

Two perspectives on accountability are required to ensure the use of AI tools and technology for community policing: i) accountability on the *choice* for policing applications, and ii) accountability for the complete *lifecycle* of the AI tools. First of all, it requires a clear (traceable and explicit) formulation of arguments why to opt for a specific AI tool and what the *high-level design* of the application is. Herein it should be made clear what the security problem is that is being addressed and how the AI-led solution can potentially contribute to solving that issue, and why the AI solution is the appropriate way to address said issue (see also: *public safety & security*).

This requires to determine what is expected from the high-level design of the AI tool, and what the foreseen risks are. Herein, it should be made clear what the actual driving values (e.g., public safety) are of the AI tool, and weigh these against the underpinning values (e.g., right to privacy). It should be made explicit, within the context of other measures, why the AI tool is indeed *net positive* in terms of potential benefits and risks. This initial step enables to make the added value of an AI tool explicit for different societal stakeholders, and to be able to reflect on these public values throughout the development and use of the AI tool.

Second, ensuring the accountability of LEAs requires full attention to the whole sociotechnical lifecycle of AI applications. It requires a clear insight in the manner how the tool is being designed, what data is being used and how it is embedded in the policing practices (Goldberg 2022; Wessels, 2024; Wieringa, 2020). Herein, it should be clear what decisions are made in the development of the AI tool, and how potential risks are addressed. Effective procedural fundamental rights for citizens – such as fair trial and presumption of innocence – need to be maintained by LEAs whilst using data-driven applications (recital 59 AI Act). This is especially prevalent when people are under investigation, in which their abilities to defend themselves should be protected by providing sufficient and meaningful information on the algorithms used (recital 59 AI Act). This also includes the human factor within the policing process: it should be clear how unintended or undesirable interaction effects are mediated within the governance of the processes of the LEAs (see also: *human oversight*).

Accountability requires continuous oversight. Within police organisations, adequate AI governance should be in place that helps guiding the procurement, development & verification,

and the monitoring and evaluation throughout its use. Especially when the AI tool has a continuous feedback learning loop, this is of extra importance in which the driving values (as intended throughout development) are actually achieved. There should be clear policies that also enable to phase out specific tooling (see also: *human oversight*). When there is an indication that the system presents a risk, procedures need to be in place for notification to relevant authorities and investigate causes (art. 20 AI Act). This means, among others, that deployers are required to keep sufficient documentation and logs throughout its use (art. 26 AI Act). Sufficient oversight is both needed through intra-organisational checks- and balances, whilst also sufficient external auditing capabilities (art 20 AI Act Art; art. 26 AI Act; Landers & Behrend, 2023).

Transparency & Explainability

Transparency and explainability of community policing tooling is a prerequisite for accountability of LEAs. This means that police organisations should be able to provide insight in the functioning of the used AI tools and make them traceable. It requires explainability on the technical functioning of the system, and the complete policing process. It needs to be elucidated how the AI tool operates and what datasets are being used (Sorell, 2024) (see also: *data bias & fairness*). The AI Act describes how providers of high-risk AI systems are required to keep and provide sufficient technical documentation and notify relevant bodies when changes are made to their systems (art. 13 AI Act; art. 18 AI Act). Depending on the case, LEAs can be both the provider and deployer when developing the tool in-house.

'Sufficient' transparency is context-dependent. Depending on the technology, its purpose, and the different stakeholders involved, there may be different requirements for the degree and types of transparency needed *and* possible. Hayes, Poel & Steen (2020) argue how opacity of algorithms can emerge because of i) intentional purposes (e.g., protection of intellectual property or security reasons), ii) illiteracy of stakeholders regarding the algorithms' technical components, and iii) intrinsic opacity due to the complexity of the algorithmic tools (e.g., machine learning based tooling). In the context of community policing, all three sources of opacity should be sufficiently addressed, and the trade-offs must be made explicit.

Transparency is not only needed about the algorithms themselves. Transparency requires also to communicate about the manner how community policing is designed and executed in practice, referred to as 'business model transparency' (HLEG, 2019). Special attention should be for the communication towards different stakeholders, in which information is tailored to the public. This implies that the information must fit the different target groups to be reached and how and by whom it is being explained, and in which contexts it should be explained (Balasubramaniam, 2023; Köhl et al., 2019). For community policing, this implies that communicating and explaining about the AI tools fits with the local context. Herein, it can be important to have a clear communication strategy regarding the different target groups. There should be distilled in what phase of the lifecycle different audiences are informed, but also what is being explained and by whom and through what communication channels and means are being used.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

Human Oversight

Human oversight deals with what professionals are being asked to do and who is doing it (capabilities of the professional), both at the street level (police officers) and police management. The practice of performing human oversight is categorized as either human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC). This entails a 'human-centric approach' which requires implementing AI tools safely and reliably to benefit humanity, with the aim of protecting human rights and dignity by keeping a 'human-in-the-loop'.

The AI Act requires that AI-designers allow human control or interference with an AI-system to achieve effective human oversight. Article 14 of the AI Act states, for example, that high-risk AI systems must be designed and developed in such a way that they can be 'effectively overseen by natural persons during the period in which they are in use'. This obligation implies that AI-designers integrate in high-risk AI systems a human control function in different phases of the lifecycle as part of a safeguard against malfunctions of AI. This could involve fixed design decisions manifested through specific instructions to automatically interrupt the process and redirect the case or issue to human oversight – at certain instances or as a response to certain impulses (Enqvist, 2023).

In addition, LEAs using AI ('deployers' under the AI Act) must also implement human oversight measures of their own. This includes at least having appropriately trained personnel with at the right level of authority able to understand the instructions for use, to put in place all necessary technical and organizational measures required by the provider, and to monitor the functioning of the AI tool. Some level of human oversight must be done both in relation to each use of a system, as well as for the system as a whole. For example, for an AI system recommending patrol routes, human oversight per use could consistent of providing pertinent information to the patrol officers or patrol commander in charge indicating on the basis of which factors the recommendation is made, and to allow them to accept or reject the recommended route.

In addition to this level, governance processes within the LEA should verify at regular intervals that the system as a whole does not produce adverse outcomes, e.g., always the same routes in disadvantaged neighbourhoods, leading to over-policing, always suggesting routes that are routinely rejected, overreliance on the system by patrol officers (e.g., indicated by having no or very few rejections), etc. This can be done by, among others, analysing logs, surveying users on discrepancies between AI recommendations and real-life daily experiences, statistical analysis, comparisons with results before the tool being used.

The AI Act only provides this obligation for high-risk systems. Examples of high-risk systems being used in the field of public safety are systems for profiling people or assessing their risk of committing a crime (unless prohibited under art. 5 AI Act). High-risk could also be AI systems operating robots or drones in local neighbourhoods. Despite this, human oversight is an essential element of good policing, accountability and good governance, also where the AI

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

system is not high-risk under the AI Act, and therefore not regulated (with the exception of systems directly interacting with humans).

Attention must be paid to the development of AI-skills of police officers and other LEAs-personnel, through all layers of the organization. Regarding the operationalization of human oversight, and depending on the designed AI tools for community policing, this should entail a cooperation between tool providers and the LEA users, leading to a good common understanding. It is important in this context to acknowledge the knowledge gap between the tool provider, who designed the tool and understands it very well, and the LEA deployer, who is often reliant on the information provided and must get sufficient information to allow effective human oversight on their end (see also: *transparency & explainability*). This highlights the importance of proper due diligence in the selection of tools to be taken into use and the extent to which a provider may or may not facilitate the AI governance and human oversight measures on the LEA's end.

According to the AI Act, and this can be considered good practice also where the Act does not apply, LEAs must ensure they assign human oversight to 'natural persons who have the necessary competence, training and authority, as well as the necessary support' (art. 26 AI Act). Moreover, the provider must indicate in the instructions for use which measures are necessary, but it remains up to the LEA to ensure that these instructions are translated into appropriate technical and organisational measures and that this governance is realized in practice in terms of personnel and resources. This includes the possibility to opt-out of the local police officers by not following the recommended advice or decision. Should police officers opt to not accept, for example, further insight is necessary about the design of the AI tool (e.g., to identify shortcomings) and the use of it in the daily work of police officers (e.g., to detect discrepancies between the function of the AI tool and their daily work).

As such, it is important to verify to what extent providers of tools are able and willing to facilitate this, and how this relates to the LEAs own AI maturity in terms of personnel and training, and available resources.

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

4. CONCLUDING REMARKS & FUTURE REVISION

The proposed KOBAN Framework provides an oversight of the ethical, societal, cultural and gender aspects relating to the use of AI tools and technology for community policing in the six EU-countries. The KOBAN framework was created on the basis of an extensive literature review. First, we looked at existing frameworks in the field of AI and the public values distinguished in them. Next, the most important public values were elaborated in relation to community policing and the use of AI.

A number of things stand out here. In the analysed AI frameworks, many different definitions are used to describe public values. Unambiguous definitions are lacking, which means that depending on the AI tool and the local context of the pilots in the EU-countries, it will have to be seen how a specific public value is applied through extensive deliberation. It is also striking that the existing AI frameworks pay little or no attention to public values of safety and efficiency (*set 1*). However, these are also values that should be taken into account when deciding whether or not to make certain AI tools. With regard to the public value of public safety, the question of proportionality and subsidiarity should always be raised. Is a technological tool the best way to achieve something? Or are there other, less intrusive ways that achieve the same goal? These are problems to be addressed further on, and as required by the six pilots in the EU-countries. This means that the pilots should find ways to address/research the outcomes of such questions in a qualitative and/or quantitative manner.

This is the first version of the KOBAN Framework. Based on the findings of the pilots in the EU-countries (WP4) and implementing it in all other work packages, the framework will be updated to ensure a careful and responsible way of designing and dealing with AI-led solutions in community policing. In doing so, the following actions will be taken throughout the KOBAN project:

- A seminar to LEA's and other interested partners within KOBAN will be provided regarding the KOBAN framework to familiarize them with its contents;
- After every KOBAN pilot, there will be reflected in a session on the KOBAN framework regarding its comprehensiveness in respect to the pilot (as part of WP4);
- Within KOBAN, an advisory board will be established with high-level experts who will be actively advocating the KOBAN outcomes and ambitions. The KOBAN Framework will be an instrument that can be used for this board, which can lead to additions and revisions to the framework needed.

The experiences with the KOBAN Framework gained throughout the actions listed above will be compiled and lead to a new revision of the Framework (D1.2b), expected halfway through the KOBAN project (expected after completion of the third pilot).

KOBAN is funded by the EU Horizon Europe research and Innovation programme under GA 101168312

5. REFERENCES

Aizenberg, E., & van den Hoven, J. (2020), Designing for human rights in AI, *Big Data and Society, 7*(2), 1-14.

Benjamin, R. (2019), *Race After Technology: Abolitionist Tools for the New Jim Code*, Medford, MA: Polity Press.

Bovens, M. (2007), Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447-468.

Brey, P., Jansen, P., Maas, J., Lundgren, B., & A. Resseguier (2020), *An Ethical framework for the development and use of AI and robotics technologies*, https://www.sienna-project.eu/w/si/robotics/ethical-framework/

Busuioc, M. (2021), Accountable artificial intelligence: Holding algorithms to account, *Public Administration Review, 81*(5), 825-836.

Casanovas, P., Teodoro, E., Guillén, A., Hashmi, M. & N. Morris (2021), *Legal and Ethical Framework and Risk Analysis*, https://webs.uab.cat/idt/wp-content/uploads/sites/35/2023/04/D9.6-SPIRIT.pdf.

Cath, C. (2018), Governing artificial intelligence: ethical, legal and technical opportunities and challenges, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 1-8.

Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & N. de Oliveira (2023), Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance, *Patterns*, *4*(10), 1-15.

Directive (EU) 2016/680 of the European Parliament and of the Council.

Dechesne, F., Dignum, V., Zardiashvili, L., & J. Bieger (2019), *AI & ethics at the police: towards responsible use of artificial intelligence in the Dutch police,* https://scholarlypublications.universiteitleiden.nl/handle/1887/85954

Donatz-Fest, I. (2024), The 'doings' behind data: An ethnography of police data construction, *Big Data and Society*, 11(3), 1-12.

DRIVER+ (2019), *A guide on assessing unintended societal impacts of different CM functions - Version 2*, https://www.driver-project.eu/wp-content/uploads/2019/11/DRIVER_D913.41_A-guide-on-assessing-unintended-societal-impacts-of-different-CM-functions-Version-2-Final.pdf.

Enqvist, L. (2023), 'Human oversight' in the EU artificial intelligence act: what, when and by whom? *Law, Innovation and Technology*, *15*(2), 508–535.

Europol (2024), *AI and policing: The benefits and challenges of artificial intelligence for law enforcement*, https://www.europol.europa.eu/publication-events/main-reports/ai-and-policing.

Ferguson, A.G. (2017), *The Rise of Big Data Policing. Surveillance, Race, and the Future of Law Enforcement*, New York: NUY-Press.

Forsberg, E.-M. (2014), Institutionalising ELSA in the Moment of Breakdown?, *Life Sciences, Society and Policy*, 10(1), 1–16.

Fosch-Villaronga, E. & A. Poulsen (2022), Diversity and inclusion in artificial intelligence, in Custers, B. & E. Fosch-Villaronga (eds.), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, pp. 109-134. The Hague: T.M.C. Asser Press.

Francis, B., Brey, P., Porcari, A. & Schepers, T. (2023), *Ethical Sensitivity Tools: Social ReadinessTool and Ethical Role-playing Too*l, https://www.techethos.eu/societal-readiness-tool.

Friedman B, Kahn PH, Borning A (2013), Value Sensitive Design and information systems, In: Doorn N, Schuurbiers D, Van de Poel I, et al. (eds), *Early Engagement and New Technologies: Opening up the Laboratory,* pp. 55-95, *Berlin: Springer.*

Ghioni, R., Taddeo, M., & L. Floridi (2024), Open-source intelligence and AI: a systematic review of the GELSI literature, *AI & Society, 39*(4), 1827-1842.

Goldberg, Z.J. (2022), How to Conduct an Ethics Assessment of AI in Policing, *14th ACM Web Science Conference 2022, Barcelona, Spain*. ACM: New York.

Government of the Netherlands (2024), *The government-wide vision on Generative Al of the Netherlands*, Government-wide vision on generative AI of the Netherlands | Parliamentary document | Government.nl.

Harris, H. & A. Burke, (2021), Artificial Intelligence, Policing and Ethics–a best practice model for AI enabled policing in Australia, *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW),* pp. 53-58), IEEE.

Hagendorff, T, (2020), The ethics of AI ethics: An evaluation of guidelines, *Minds and machines*, *30*(1), 99-120.

Hayes, P., I. van de Poel & M. Steen (2020), Algorithms and values in justice and security, *AI & Society,* 35, 533–555.

High-Level Expert Group on Artificial Intelligence (2019), *Ethics guidelines for trustworthy AI*, Brussels: European Commission.

Khoei, T.T. & A. Singh (2024), Data reduction in big data: A survey of methods, challenges and future directions, *International Journal of Data Science and Analytics,* 1-40.

Köhl, M. A., Baum, K., Langer, M., Oster, D., Speith, T., & D. Bohlender (2019), Explainability as a non-functional requirement, *2019 IEEE 27th International Requirements Engineering Conference (RE),* 363-368.

Landers, R. & T. Behrend (2023), Auditing the AI Auditors: A Framework for Evaluating Fairness and Bias in High Stakes AI Predictive Models, *American Psychologist, 78*(1), 36-49.

Latour, B. (2017), *Où atterrir? Comment s'orienter en politique*, Paris: La Découverte.

Murray D. et al. (2023), The Chilling Effects of Surveillance and Human Rights: Insights from Qualitative Research in Uganda and Zimbabwe, *Journal of Human Rights Practice*, https://doi.org/10.1093/jhuman/huad020.

Ng, L. H., Lim, A. C., Lim, A. X., & A. Taeihagh (2023), Digital Ethics for Biometric Applications in a Smart City, *Digital Government: Research and Practice, 4*(4), 1-6.

Richardson, R., J. Schultz & K. Crawford (2019), Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, *New York University Law Review Online*, 94(192), 192-233.

Ryan, M. & A.F. Inguanzo (2021), Big Data Ethics, *Encyclopedia of Business and Professional Ethics* (pp. 1-5). Springer.

Salgado, N., Meza, J., Vaca-Cardenas, M., & L. Vaca-Cardenas (2024), Current and emerging trends in the use of AI for community surveillance, *Journal of Infrastructure, Policy and Development*, 8(8), 6135.

Schuilenburg, M. (2024), *Making Surveillance Public: Why You Should Be More Woke About AI and Algorithms*, The Hague: Eleven International Publishing.

Schuilenburg, M. & R. Peeters (2024), Voorbij de system-level bureaucratie: over datastromen, algoritmes en Inclusieve AI, *Beleid en Maatschappij*, (51)3, 278-293.

Shams, R. A., Zowghi, D., & M. Bano (2023), AI and the quest for diversity and inclusion: A systematic literature review, *AI and Ethics*, 1-28.

Sorell, T. (2024), AI-related data ethics oversight in UK policing, *Policing: A Journal of Policy and Practice*, *18*, 1-9.

Steen, M. (2022), *Ethics for people who work in tech.* Boca Raton, FL: Routledge / CRC Press.

Storbeck, M., G. Jacobs, M. Schuilenburg & R. van den Akker (2024), Surveillance Experiences of Extinction Rebellion Activists and Police: Unpacking the Technologization of Dutch Protest Policing, *Big Data & Society*, *forthcoming*.

TNO (2021), Op zoek naar de mens in AI: Betrek de burger en experimenteer op verantwoorde wijze, Den Haag.

Trevisan, F. & C. Zednik (2023), *Ethical frameworks for the use of LEAs*, https://www.pop-ai.eu/wp-content/uploads/2024/01/popAI-D2.4-Ethical-frameworks-for-the-use-of-LEAs.pdf.

UK Government (2020), *A guide to using artificial intelligence in the public sector*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/964787/A_guide_to_using_AI_in_the_public_sector__Mobile_version_.pdf.

UK Government, (2021), *Ethics, transparency and accountability framework for automated decision-making*, https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making.

Unicri & Interpol (2020), *Towards responsible AI innovation*, UNICRI: United Nations Interregional Crime and Justice Research Institute.

United States White House (2023), *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,* https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Victoria Police, (2024), *AI Ethics Framework: Enabling principles*, https://www.police.vic.gov.au/victoria-police-artificial-intelligence-ethics-framework/enabling-principles

Wessels, M. (2024), Algorithmic policing accountability: Eight sociotechnical challenges*, Policing and Society, 34*(3), 124-138.

Wieringa, M. (2020), What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability, *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 1-18.

WRR (Scientific Council for Government Policy) (2011), *IGovernment*. Amsterdam, Amsterdam University Press.