EUROPOL

INNOVATION LAB

ASSESSING TECHNOLOGIES IN LAW ENFORCEMENT

A method for Ethical decision-making

EUROPEAN CLEARING BOARD

Strategic Group on Technology and Ethics

ASSESSING TECHNOLOGIES IN LAW ENFORCEMENT
A method for ethical decision-making

PDF Web | ISBN 978-92-95236-96-7 | DOI: 10.2813/1291864 | QL-01-25-002-EN-N

Luxembourg: Publications Office of the European Union, 2025

Cite this publication: Europol (2025), Assessing Technologies in Law enforcement. A method for ethical decision making, Publications Office of the European Union, Luxembourg.

This publication and more information on Europol are available on the Internet.

www.europol.europa.eu

Assessing technology in law enforcement: a method for ethical decision-making

Presented to the EuCB on 22 March 2024 by its Strategic Group on Technology and Ethics. Revised in January 2025.

Contents

Introduction

Europol Innovation Lab, EuCB & Strategic Group on Technology and Ethics

Europol was mandated in 2019 by the EU Justice and Home Affairs ministers to create an Innovation Lab to support the law enforcement community in the area of innovation. The Lab aims to identify, promote and develop concrete innovative solutions in support of the EU Member States' operational work. These will help investigators and analysts to make the most of the opportunities offered by new technology to avoid duplication of work, create synergies and pool resources. The activities of the Lab are directly linked to the strategic priorities as laid out in Europol Strategy 2020+, which states that Europol shall be at the forefront of law enforcement innovation and research.

The European Clearing Board for 'Tools, Methods and Innovations in the field of technical support of operations and investigations' (EuCB) was launched by the Heads of Europol National Units (HENUs) in their meeting of 5 November 2020. It is composed of Single Points of Contact (SPoCs) from the Europol Innovation Lab, all EU Member States and the four Schengen-associated countries. SPoCs meet regularly in plenary meetings, during which they update each other on innovative projects and tools and decide on new joint collaboration activities.

The Strategic Group on Technology and Ethics was founded in 2021 under the umbrella of the EuCB. Currently, the group is composed of representatives from Australia, Netherlands, Norway, Slovenia, Spain, Sweden and the UK. One of the objectives of the group has been to create these guidelines 'Assessing technology in law enforcement: A method for ethical decision-making' for the benefit of all EuCB members.

A method for applying ethics in practical decision-making

Digital transformation and technology are vitally important in enhancing order and security but may also pose a threat to fundamental rights and freedoms. This document presents a method for assessing novel technology from a perspective of widely accepted values and principles. The guidelines contain a description of the central values and ethical principles, and give examples in the form of use cases, illustrating how they may be applied in structured decision-making and evaluation, in situations involving new technology in law enforcement. The use cases show how the method can be helpful in forming transparent and understandable arguments for trustworthy decisions about the adoption and use of various types of technology in law enforcement.

The values and principles discussed here are also valuable when cooperating European law enforcement authorities are in search of common moral ground in their respective practices.

The work of the Strategic Group on Technology and Ethics is based on methods used in clinical ethics committees and insights from value-based practices in policing – in addition to studies of ethical guidelines for technology and a survey of values central to European law enforcement authorities.

Part I of the guidelines explains the seven steps of the method, while Part II sets out use cases to illustrate the application of the method in practice.

Precautions regarding the use cases

It should be noted that the use cases in Part II have not been legally vetted. They serve to illustrate the present method, and although the conclusion of a use case may be that it is ethical to use the technology (under certain conditions), this should NOT be understood as a conclusion concerning its legality. Legal regulation of law enforcement's use of technology exists on many levels, both nationally and internationally, and it is beyond the mandate and resources of the Strategic Group on Technology and Ethics to perform a legal assessment of each use case scenario.

It is a virtue of the present guidelines that they provide for transparency concerning the principles and values taken into consideration by the decision-maker, and ensure that the making of a decision is specific with respect to local, political, social, cultural and economic contexts, and the technology in question. This also implies that the use cases will never merely be copied to a domestic setting. While they may provide guidance and inspiration, the decision-maker is always responsible for producing an original assessment that takes account of the concrete circumstances in the actual situation.

Ethics and the law

To the members of Europol, it is fundamental that any development and deployment of technology in law enforcement must be lawful. For the purpose of these guidelines, it is thus assumed that, in a concrete case, issues of legality have already been duly addressed according to appropriate procedures.

In the field of new technologies in law enforcement, however, the law may sometimes lag behind, leaving grey

areas that are open for interpretation. A structured ethical approach, as presented in these guidelines, may shed light on the values and principles involved, and suggest which interpretation is ethically defendable.

In the same vein, the present value-based reflective method may play a constructive role in the legislative process, by making visible – and more understandable – the ethical concerns that legislators should take into account when striking the balance between freedom and security in the field of law enforcement. Law enforcement's development and/or use of new technology also regularly raises issues related to fundamental rights, where law and ethics are closely intertwined and the lines between the two may become blurred. This, too, leaves space for the present method to contribute with new perspectives and enrich our understanding of the issues at stake.

A 'living' document

The intention is to make this a 'living' document, that is, a document that captures new technology as well as novel applications of technology already in use. This is provided for by the expansion over time of the collection of use cases, which may integrate further developments in this area on European and national levels. This should also be reflected in Initiatives for training in the use of the method. By its dynamic character the document aims to be a durable resource for law enforcement authorities and policy makers.

Part I: Method and central values

In an ideal world, technology should not only promote instrumental, technical values, but also 'substantive social, moral, and political values to which societies and their peoples subscribe' (Flanagan et al., 2008). Granted, we do not live in an ideal world. In practice, we need to align, to the best of our abilities, our use of technology with our central values[1]. In this part of the report, a method is presented in which values play a central role in the systematic assessment of technology and its application.

The aim of the method is to overcome the alignment problem that may exist between facts, values, rightness and goodness (consequences). It is applicable to high-level decision-making regarding the introduction of emerging technology in LEAs, as well as tactical decision-making in concrete cases.

This value-based method[2] will only succeed if the central values and the facts of the case are scrutinised properly.

The following seven steps are intended to provide support for law enforcement in making ethically robust decisions about using innovative technology. The first three steps provide a description of the case that in part helps decide which normative values matter to the case (step 4), before options and possible best practices/solutions are assessed in steps 5-7. In the following sections, each step is described in more detail.

1. A description of the moral problem

The purpose of Step 1 is to capture the initial moral framing of the case. From the outset, it is important to be aware that the use of new technology can entail ethical problems. To capture this problem, one can examine different moral reactions to the intended use of the technology. Spelling out conflicting concerns ('this is clearly a case of care vs justice'), or briefly describing someone's feeling of discomfort, uncertainty or moral (affective) reaction ('this is disgusting', 'this is unjust')[3], is enough at this stage. Likewise, listing points raised in the public debate ('facial recognition is violating basic human rights!') may also be a way to frame the problem. The basic idea is to record the various initial viewpoints'.

2. The relevant facts

The second step consists of making a note of the relevant facts about the case, such as facts about the technology and how it is (or might be[4]) applied in the present case – in addition to information about the context and relevant legislation. One should also identify types of information that could be useful to further clarify the situation, increase awareness of uncertainty, and help point out blind spots in people's perception of the situation.

3. The parties affected by the technology

The purpose of step 3 is to map different perspectives on the case at hand. This is done by making a list of affected persons (or groups) and their viewpoints, for instance in terms of their immediate interests/ideas, concerns and expectations (ICE)[5]. In a law enforcement context, the parties will typically include offenders, victims, witnesses, next of kin, law enforcement officers and the public.

The viewpoints of the parties are based on communication, observation or, if necessary, educated guesses. Making explicit reasonable assumptions may improve the situational awareness, reveal misconceptions and prove valuable for enhancing transparency.

4. Normative values that matter to the case

The value landscape represents a normative, long-term perspective – as opposed to the descriptive, agent-centred interests of Step 3. The purpose of Step 4 is to explicate a set of values that a possible intervention ought to express. Although an objective view from nowhere is unattainable in practice, the task is to establish a general 'moral point of view' to the best of one's abilities. One technique consists of asking which values an 'impartial spectator' would be likely to emphasise in the case at hand[6].

Values often listed in connection with emerging technology are transparency, fairness, privacy and accountability. Others include honesty, autonomy, beneficence, non-maleficence, social justice, etc. One may also include relevant professional values (i.e. values relevant to law enforcement). Having to deal with a large number of values may complicate the reasoning, so it is necessary to single out the most important ones. Typically, three to five central values are sufficient to describe the value landscape.

A short summary of four central values and principles relating to the use of technology is presented below. They are chosen because of their prominence in meta-studies of ethical codes and guidelines (Laas et al., 2022), as well as in a survey of European police organisations conducted by the Strategic Group on Technology and Ethics in 2021.

a) Transparency

Transparency is important to most public services. Without transparency, actions and practices cannot be discussed, and the absence of knowledge makes abuse of power more likely. The Nolan principles for good governance state that '[h]olders of public office should act and take decisions in an open and transparent manner. Information should not be withheld from the public unless there are clear and lawful reasons for so doing.'[7] Transparency is a precondition for the oversight and accountability of law enforcement authorities. Transparency fosters legitimacy and is a precondition for avoiding polarisation between law enforcement and citizens, and for creating and sustaining a trusting relationship.

On the other hand, transparency may render lawful, yet invasive, surveillance methods ineffective, as it may invite undesirable countermeasures. Law enforcement is also obliged to protect private information. Legal obligations to keep information confidential (e.g. business secrets) may hinder demands for transparency. In addition, being transparent on suboptimal parts of the service may not always generate public trust. Law enforcement cannot always be transparent to all parties[8].

There are, in other words, both good reasons for not informing about some technological capabilities, as well as a duty for doing so. Still, the transparency norm indicates that LEAs ought to be as open as possible about actions, practices and technology. That is, reasons must be given for not being transparent, not vice versa. It may be possible to inform the public of when and where some technology is used, without providing all the technical details.[9] At other times, the transparency norm may be observed through oversight committees representing the public.

Transparency towards the public is particularly challenging when law enforcement agencies themselves are unfamiliar with the inner workings of the technology in question (e.g. third party applications for predictive policing).

b) Fairness

Fairness – as a form of justice – implies that the needs, rights and interests of others matter in decision-making. This may concern both outcomes (in terms of equality and impartiality) and procedural justice. Fairness implies the weighing of different (moral) concerns, a process which requires application of moral rules or principles, or experience and natural decision making (Kahneman & Klein, 2009). In both cases, cognitive and systemic biases may influence decision-making and cause disagreement.

Difference of opinion may concern discretion in the way the situation is understood, what to do, and the manner in which an action should be performed (Kleinig, 1996, p. 82 f).

In well-defined situations, clear and transparent rules may be sufficient (e.g. traffic speed controls, which to some extent simple algorithms may handle adequately). For more complex tasks, AI systems may provide guidance, finding patterns that are hard to detect for humans and at times mitigating human biases. However, AI systems may also yield biases through modelling and/or training databases, as well as incorrect use of the systems.

In order to fulfil the demands of procedural justice, human competence is required to explain the output of AI in meaningful ways. This issue is also emphasised in data protection laws, which require data controllers to provide certain information to individuals whose data they hold and use. A privacy notice, also known as a 'fair processing notice', is one way of providing this information[10].

c) Privacy

Privacy is valued as it provides a sphere within which a person can be free from interference by others. Though considered essential, privacy is not an absolute value. Infringements of privacy pose a threat to the autonomy and social integrity of an individual. The right to privacy encompasses a person's behaviour and actions, communication, data and image, thoughts and feelings, location and space, and association.[11] The right to privacy is challenged online, as personal information (preferences, location, behaviour, etc.) has become the central commodity under 'surveillance capitalism'[12]. Today, being able to control one's personal information is considered an essential dimension of privacy[13]. The sharing of personal information produces vast amounts of information useful for LEAs. Controlling how information is shared is more or less impossible given the structure of the internet. Therefore, data protection is central to most contemporary discussions of privacy.

On the other hand, calls for privacy may also be made in order to hide criminal activities. The right to privacy may be interfered with in criminal investigations, and more generally in intelligence work. Within the applicable legal frameworks, LEAs often interfere with privacy, for instance by using DNA matching, face recognition, GPS tracking (phone, car), grand scale internet data mining, re-identification of anonymised data, or thermal imaging. Still, to protect the privacy of others, law enforcement should only collect personal data that is strictly necessary for the purpose, referred to as the data minimisation principle.

d) Accountability[14]

LEAs must be accountable for their use of technology. In the present context, this means that they are morally responsible for promoting and balancing the central values (transparency, fairness, privacy) when using the technology in question. Technological tools can account for ('log') their output, and some tools can assess the integrity of data sets. However, accountability in the technological sense lacks the moral quality of human responsibility and integrity. In a law enforcement context, responsibility indicates a willingness to protect the citizens' human worth and dignity, privacy and inalienable rights, regardless of the citizens' role. Moral integrity means adhering to central values (being 'principled'), and balancing these values in an acceptable manner[15]. Thus, accountability serves as a protection against the temptation of an 'anything goes' approach when technological opportunities become available. Accountability also requires insight into the limitations of one's competence and available resources.

5. Formulating value-based solutions

In step 5 the aim is, based on the circumstances described in steps 1-3, to imagine options that express the set of values suggested in Step 4. The reflections are typically forward-looking: can we justify the use of the technology in question – in general or in certain contexts?

If a suggested intervention expresses the set of values in an acceptable manner, the suggestion qualifies as 'value-based'. Typically, several options qualify, even though not all the suggested options may fit every selected value equally well. If the set of values is impossible to include, the option must be rejected. Alternatively, the option might be modified, for instance by expanding the human control/contribution.

6. Assessment of the solutions and justification of a choice

In Step 6, the value-based options identified in the previous step are further scrutinised by considering their rightness and consequences. First, rightness is assured by asking the following four questions[16]:

- **CONSISTENCY:** is the suggested line of action (here, the use of technology) always appropriate under similar conditions?
- **DIGNITY:** does the use of the technology imply that LEAs use their professional authority in the best interest of the persons, and not only as a means to fulfil other goals?

- **PUBLIC ACCEPTANCE:** is the use of the technology acceptable to the public if it becomes generally known?
- **ACCOUNTABILITY:** do the necessary competencies and resources exist to take responsibility for the use of the technology?

If the answer is 'no' to any of these questions, the suggested use of the technology should be rejected or modified. If all four questions can be answered positively, the option is permissible.

Secondly, if the suggested option is considered permissible, its consequences are estimated[17]. Addressing the consequences involves estimating how the various proposed interventions affect the parties involved in both the short and the long term, keeping in mind possible side effects, and considering the probabilities of the outcomes. The optimal choice is the one that shows the most positive aggregated outcome.

$Option_{1,2..n}$	$Involved\ party_1$	$Involved\ party_2$	$Involved\ party_n$
Short-term consequences			
Long-term consequences			

It is difficult accurately to quantify the outcome of consequentialist assessments. Nevertheless, what is important is that the LEA considers the different ways in which the interventions will affect the parties and, if there are several permissible interventions, compares these in terms of consequences. Sometimes, consequentialist assessments may also help reveal biases (for instance, that one of the parties reaps all the benefits).

7. Short summary

Finally, the process should be summarised to ensure the coherence/consistency of the reasoning and choice. The following format is one way of summarising the reasoning: in the case where <based on Step 2>, the moral concern was initially considered to be <Step 1>. The most important involved parties were <Step 3>, and the most central values to the case were <Step 4>. Based upon these values, several actions <Step 5> were proposed (A1…An). Ax was deemed permissible <Step 6> according to the rightness test and was deemed to produce the best overall consequences.

This value-based method does not produce definitive answers, but generates transparently construed and justified, criticisable solutions. Therefore, a value-based method is also suitable as a tool for discerning and explaining differences of professional opinion, as it helps explain the reasons for disagreement, referring to (i) framing, (ii) relevant facts, (iii) involved parties, (iv) values or (v) interventions (use of technology in the present context). In other words, disagreements are discussed in terms of differing premises and inferences, not as differences in personal preferences or character. The aim is thus discursive rather than decisive, and following the steps outlined above should increase the insight and understanding of the participating law enforcement officers.

Part II: Use cases (Model analysis)

This part contains use cases to illustrate the method. Apart from providing examples which can be transferred to similar uses of technology elsewhere, the aim is educational, as they demonstrate the application of the method as such.

The use cases are not connected to concrete initiatives in any Member State. They have been brought up by the members of the strategic group and discussed during our meetings. Within the group, there is a wide range of expertise and specialism from members across Europe and beyond. Some members have brought use cases that represent current, real life ethical debate in their home country, to seek wider views and discuss the application of the method. Others have sought to consider 'the art of the possible' and use more hypothetical use cases to test the methodology. Readers should therefore be aware that some of the technology or analytical processing may not exist or be available for use at this time. The use case will, nevertheless, provide a valuable example of how the methodology can be applied.

It is important to note that the conclusions presented for each use case are the conclusions reached by the Strategic Group on Technology and Ethics based on the information considered at the time of the exercise. For each use case, the time of assessment is indicated. In addition, as mentioned in the introduction, the cases have not been legally vetted.

New case descriptions, or full or partial reflections, are welcome in light of the intention to make this a living document (see introduction). In particular, the hope is that the method proves helpful in the context of discussions of specific technology in the core groups of the EuCB.

The limited number of use cases attached to this first version of the guidelines hardly covers the full spectrum of relevant technology. It is too early to come up with overarching categories correctly, as we do not know which types of cases will come up in the future. Preliminarily, we have grouped the cases into three categories: 1) LEA efficacy measures, 2) Particular types of crime, and 3) Counterterrorism.

The overview sets out the problem, the values and the conclusion. The full reflection is found after the overview.

LEA effectivity measures			
Case	The moral problem	Central values	Conclusion
1: Video analytics technology (VAT) VAT is considered to complement existing CCTV capability, for searching for particular objects, persons by description, vehicles etc. more effectively than by current manual analysis. VAT does not involve facial recognition/ matching.	Balancing individuals' right to privacy against concerns about wide-scale monitoring of citizens' behaviour in public places.	Public safety, privacy, and transparency	To introduce VAT software after a period of public engagement, only for defined use cases and with approval of a senior officer (assuming the public are largely supportive of the technology).
3: Model analysis of open source data scraping Stolen power tools are often resold on online marketplaces. Should the LEA scrape such sites for data?	Is automated research of online marketplaces using 'scraping' tools acceptable?	Public safety, fairness, privacy and transparency.	As automated investigations violate the terms of service, open source data scraping is unacceptable.
5: Automated analysis of large and complex datasets Examines a range of problems connected to analysis of large and complex datasets	Analysis of large and complex datasets is vulnerable to function creep, sensitive combination of sources, and evolution of data.	Privacy, fairness, transparency	Acceptable as a middle ground between 'anything goes' and 'forbidden'. Different measures are required depending on the specific technology.

Particular types of crime			
Case	The moral problem	Central values	Conclusion
2: Measuring the risk of reoffending in cases of gender-based violence AI is used to measure the risk, but a human review by trained personnel is always carried out.	Transparency issues caused by concerns about reliability, explainability and fairness.	Public safety, privacy, fairness, transparency	Proper development with extensive tests and simultaneous use with the current system, transparently engaging with the public, is considered the best solution.
4: Using a chatbot to prevent child sexual abuse The chatbot detects sexualised speech, indicates age and gender, performs sentiment analysis and detects linguistic fingerprints, allowing a human operator to intervene.	Risk of excessive surveillance as all chat data in the forum is processed. Real-life testing is problematic. There is also the black box problem of deep learning.	Privacy, children's safety online, fairness, transparency and accountability.	A limited version of the chatbot with a large age threshold (age difference between the interlocutors) is acceptable.

CASE 1: VIDEO ANALYTICS TECHNOLOGY

(March 2024)

Law enforcement are considering an enhancement to the capability of their existing CCTV cameras to introduce video analytics technology (VAT). This technology will enable law enforcement to research CCTV footage that has already been captured far more quickly than by using current manual processes, as it will allow them to search for particular objects, persons by description, vehicles, etc. There is no live facial recognition or facial matching capability in this particular use case.

The technology will not be introduced to support a particular investigation but rather in order to be available if required.

Examples of its use could be a search for a missing person by a description of their clothing or research to identify how many persons entered a particular address in a drugs or people trafficking enquiry. The technology is widely used in the private sector (for example, in retail) to identify patterns of behaviour consistent with shoplifting. However, the supplier of the technology has not provided any information on its accuracy, for commercial reasons.

Some senior officers view the use of the technology as a mere extension of existing manual CCTV provision, while others have concerns about the use of AI to monitor behaviour, regardless of whether individuals can be identified or not.

1. THE MORAL PROBLEM

Law enforcement need to find ways of using technology to enhance their ability to conduct enquiries efficiently, especially where there is a time-critical element – for example in a search for a missing child. However, it must do so in a way that balances individuals' right to privacy and recognises concerns about wide-scale monitoring of citizen's behaviour in public places. The public are largely accepting and supportive of law enforcement's use of CCTV but less is known about their attitudes to the use of analytics to automate research of imagery.

2. THE FACTS

The facts relevant to the situation are set out below.

- CCTV imagery has been captured lawfully by law enforcement across the world and this is widely accepted by the public;
- current processes require officers and analysts to manually review hours of CCTV footage when dealing with certain enquiries;
- the use of video analytics technology could speed up the process of research and provide valuable information sooner to support investigations;
- all matches and objects of interest suggested by the video analytics software will be manually reviewed by an officer;
- a data protection impact assessment has been completed and approved, meaning that the introduction of the technology would be lawful;
- the supplier of the analytics software has not provided any information about its accuracy and performance;
- some sections of society have expressed concern about the use of AI to analyse CCTV imagery.

3. PARTIES INVOLVED

By defining an initial understanding of the problem, and stating the facts of the case and the immediate interests of the different parties involved, a broad understanding of the case and how using the technology may affect the case is established.

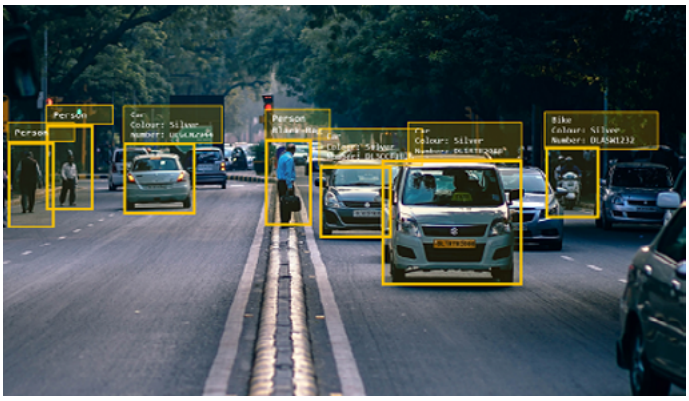
LAW-ABIDING CITIZENS who wish to go about their business in areas captured by CCTV, although it is likely that there will be a wide variety of opinions on the use of video analytics software. Some may see this as an unacceptable heightened intrusion on their privacy (even when compared to existing CCTV capture), while others may expect law enforcement to take any reasonable steps

to process CCTV in a more efficient manner and may expect that this capability already exists.

LAW ENFORCEMENT are interested in processing CCTV imagery as quickly and efficiently as possible to reduce the time taken to carry out this activity and remove the risks of human error.

CIVIL LIBERTIES GROUPS may be interested in the use of video analytics by law enforcement and may challenge its use.

VICTIMS OF CRIME OR RELATIVES OF MISSING PERSONS will have an expectation that police will be doing all in their power and using all available technology to apprehend suspects or find their loved ones before they come to harm.



4. VALUES

There are many technical issues to be considered in this case, but in terms of purely moral considerations, based on the description above (1-3), the normative considerations ought to be motivated by the following three values:

- **PUBLIC SAFETY:** in this context, enhancing public safety means processing CCTV imagery as quickly and efficiently as possible.
- **PRIVACY:** public safety benefits must be balanced with the right of law-abiding citizens to privacy and to not have their behaviour and movements processed using AI. On the other hand, the technology could be seen as enhancing privacy as research of CCTV imagery will be more efficient.
- **TRANSPARENCY:** this refers to law enforcement striking the balance between being transparent about the use of video analytics and providing information about its use that could be used to reduce its effectiveness by those intending to commit crime. Transparency also refers to the requirement for officers and CCTV operators to understand how the technology is working and what its limitations are.

5. OPTIONS

The following four options are available:

1. Do nothing and continue to manually review CCTV.
2. Introduce the video analytics software with no public consultation or engagement and use this to support any enquiry that has captured CCTV.
3. Introduce the video analytics software after a period of public engagement and use this to support any enquiry that has captured CCTV (assuming the public are largely supportive of the technology).
4. Introduce the video analytics software after a period of public engagement but only for defined use cases and with approval of a senior officer (assuming the public are largely supportive of the technology).

OPTION 1 does not align with the value landscape documented in Section 4, as the tool could enhance public safety.

OPTION 2 could be the most effective in terms of supporting law enforcement enquiries but not informing citizens in advance of its introduction could undermine confidence in law enforcement and is contrary to the values of privacy and transparency.

OPTION 3 assumes that, with proper engagement and messaging about the technology, it would be acceptable to the public. Using it in any case that has captured CCTV could still be viewed as contrary to the values of privacy and transparency.

OPTION 4 also assumes that, with proper engagement and messaging about the technology, it would be acceptable to the public. Using it in only serious cases is more in line with the value of privacy than option 3 but does not go as far in terms of public safety, as opportunities to add value to less serious cases could be missed.

OPTIONS 3 and **4** appear to be the most acceptable, with option 4 arguably the more acceptable (i.e. best in line with the value landscape) of the two.

6. JUSTIFICATION

Rightness

There are four questions that can assess the rightness of the suggested options:

1. Is the suggested line of action (use of technology) always appropriate under similar conditions?

This use case does not concern a particular investigation that would benefit from the use of technology but instead looks at introducing technology to be used when required. It would be appropriate to define a set of criteria about when the technology should be deployed.

2. Does the intervention imply that LEOs use their professional authority in the best interests of the clients and not just as a means to fulfil other goals? Yes, the use of video analytics is justified and in the best interests of citizens.

3. Is the action/practice acceptable to the public? This is difficult to assess as some members of the public may feel very strongly about the privacy implications of any sort of AI processing of images of them going about their lawful business. Others would expect police to use any technology available to enhance public safety and prevent crime. Some will view video analytics as an extension of existing CCTV capability; others will see it as a significant change to this. The key to public acceptance of the technology is transparency and education.

4. Do the necessary competencies and resources exist to take responsibility for the intervention? This will depend on the accuracy of the tool and assumes it does not have a high rate of false positive reports. Law enforcement should ensure adequate resources to quickly assess matches and carry out interventions where appropriate. The supplier should be pushed to provide information about the accuracy of the tool. If this is not possible, it should be tested by law enforcement during a trial period.

Consequences		
OPTION 3	The police	The public
Short-term	Enhanced ability to use CCTV images in support of serious investigations.	May notice impact of more efficient CCTV research in crime rates or if a victim of crime.
Long-term	May provide a pathway to more intrusive technology such as live facial recognition. Public confidence could be impacted.	Erosion of privacy rights and confidence in police if use of AI technology is expanded without proper engagement.

OPTION 4 could be acceptable based on the values and permissibility considerations summarised above. The law enforcement agency must ensure it has robust mechanisms in place to monitor the effectiveness of the technology and must be confident in the accuracy of the tool before it is deployed.

The law enforcement agency should develop guidelines as to the types of circumstances where the use of video analytics is justified.

7. SHORT SUMMARY

In the case where video analytics technology is considered for searching for objects, the moral problem is about balancing individuals' right to privacy against concerns about wide scale monitoring of citizens' behaviour in public places. In addition to law-abiding citizens and law enforcement, the case involves civil liberties groups and victims of crime/missing persons and their next of kin. A solution to this problem must promote public safety, privacy and transparency. After considering five possible options, introducing video analytics software after a period of public engagement, but only for defined use cases and with approval of a senior officer (assuming the public are largely supportive of the technology), was found to align to the set of values, pass the rightness test and lead to the best consequences.

CASE 2: MEASURING THE RISK OF REOFFENDING IN CASES OF GENDER-BASED VIOLENCE

(March 2024)

An EU country has a gender-based-violence IT case management system (hereinafter, 'system') in which data concerning all gender-based aggressions (hereinafter, 'aggressions') are stored, along with offender and victim data. Each aggression entry contains personal data about the offender and the victim (as well as other persons related to the victim who were attacked or threatened at the same time). The purpose of the system is to enable the police to apply adequate protective measures according to the risk of recidivism set by the system.

When the victim reports the aggression to the police, she is interviewed by specially trained personnel that also record data relating to previous aggressions and/or threats, use of drugs and/or weapons, whether the offender has access / a weapons licence, etc. Police officers using this system have been properly trained and undergo mandatory annual training. The risk of reoffending is always evaluated by an officer and this risk level can be upgraded by the officer, but never reduced. In other words, the system is designed to 'overprotect', rather than 'underprotect' the victim. The introduction of AI analyses makes the system generate more precise output. Based on the output, a court will decide which protective measures ought to be taken.

On the basis of the recorded data and previous entries, the system generates a risk assessment (risk of the offender reoffending against his former victim(s)) with five possible scores: undetected risk, low risk, medium risk, high risk, extreme risk.

1. THE MORAL PROBLEM

The protection of victims of gender-based violence is a major concern and has previously been the subject of national campaigns to raise awareness and various countermeasures. Using AI technology to aid law enforcement in preventing aggression seems uncontroversial. If measures are taken without proper public consent and no explanation, further aggression may ensue. AI systems do not perform magic. Which parameters are entered into the database? Are there enough entries to train the system properly? If not, the risk assessment may challenge the integrity of individuals, such as former offenders. Moreover, questions about how potential reoffenders and their victims are treated/protected may surface. Is AI the best answer or are causal indicators possible to find? Are the social mechanisms hard to discern?

2. THE FACTS

The facts relevant to the situation are set out below.

- The scope of the problem is relevant (to estimate aggression over a period of time).
- The records describing offenders, victims and the aggression are legitimately stored in a secure database.
- The software is an in-house development.
- This system is under continuous development and its accuracy has increased.
- The use of AI may provide more accuracy in the risk assessment. The data protection impact assessment has been completed and approved, so the introduction of the technology would be lawful.

Unknowns

- The size of the training database
- Is the AI system based on algorithms or on deep learning machine learning?
- Are the characteristics of the victim included in the assessment (or relational factors, or only offender data)?

3. PARTIES INVOLVED

OFFENDERS whose data is recorded in the system and may be subject to restrictive measures issued by a court.

LAW ENFORCEMENT are interested in having a better tool to assess the risk of the victims under their responsibility and, with that assessment, focus and prioritise the

preventive or protection services, depending on the score.

VICTIMS OF GENDER-BASED VIOLENCE expect that protective measures have been put in place and the risk assessment has been done with the best possible tools to guarantee their safety against future threats.

By defining an initial understanding of the problem, stating the facts of the case and the immediate interests of the different parties involved, we established a broad understanding of the case and how using this technology may affect the case.

4. VALUES

There are many technical issues to be considered in this case, but in terms of purely moral considerations, based on the description above (1-3), the normative considerations ought to be motivated by the following four values:

- **PUBLIC SAFETY:** in this context, enhancing public safety means having a better risk assessment tool to ensure the security of the victims of aggressions (in addition, the EU Member State has a legal framework with specific regulations to fight aggressions and ensure the protection of victims).
- **PRIVACY:** sensitive information about victims and offenders is stored in secure local databases and all processing is carried out by resident systems developed in-house.
- **FAIRNESS:** any system or human reasoning may overestimate or underestimate risks regarding individual offenders. AI may mitigate (or amplify) human biases but may also produce its own biases. If the combination of factors likely to express high risk makes sense and is presentable, it may produce false or hallucinatory outcomes (red hair + birthplace + tattoo on right arm), which is a real possibility if the number of offenders in the database is low. Biases are mitigated by properly trained officers-in-the-loop, and centrally reviewed cases.
- **TRANSPARENCY:** describes the level of public insight into the LEA's activities and measures – for instance, whether the public is informed about the existence of a powerful AI risk assessment tool. Transparency may also refer to whether the system's model or algorithm should be transparent to the operators of the system (to which degree this is possible depends on the type of modelling).

5. OPTIONS

The following three options are available:

1. Do nothing and continue with the old algorithm.
2. Introduce the AI upgraded system after extensive lab tests and live trials in a police unit. No public consultation or engagement. Before replacing the old system, it will be used in parallel with the old one to compare both risk assessments (the non-AI and the AI).
3. Introduce the AI upgraded system after extensive lab tests and live trials in a police unit, in addition to an extended period of public engagement.

All three options express the value landscape to some degree. Option 1 works, but is probably inferior to option 2 and 3, which enhances public safety to a larger extent. Options 2 and 3 seem equal in most respects, but option 3 clearly scores higher in terms of transparency. One may argue that option 2 provides an earlier system launch and may potentially preclude some aggression taking place.

Options 2 and 3 both appear to be acceptable, though option 3 is arguably the best in line with the value landscape of the two. As there is some ambivalence between the two options, the justification for both has therefore been studied.

6. JUSTIFICATION

Rightness

There are four questions that can assess the rightness of the suggested options:

1. Is the suggested way of introducing technology always appropriate under similar conditions? If the existing system addresses the problem, there is urgency but no pending catastrophe that precludes public engagement around the AI evolution of the system. This question clearly favours option 3.
2. Does the option imply that LEOs use their professional authority in the best interests of the clients and not just as a means to fulfil other goals? There is no information that suggests that any of the options are motivated by concerns other than the best interests of victims.
3. Is the introduction of the system acceptable to the public? A robust answer to this question is only secured by option 3, although hypothetically one

may assume that option 2 is partly acceptable, as its aim is to help victims – but the public may also ask why secrecy is chosen regarding this system.

4. Do the necessary competencies and resources exist to take responsibility for the system? The answer is yes to both option 2 and 3, as the competence and resources for system validation, operation, and revision exist. There is a training programme for operators (law enforcement officers) and those responsible for the revision after each case. As we can see, option 3 passes the rightness test and is clearly permissible, whereas option 2 is uncertain from a standpoint of universalisability (i) and publicity (iii).

Consequences

Option 3 is acceptable based on the values and rightness considerations summarised above. The argument for choosing option 3 is already strong, given that the LEA has robust mechanisms in place to monitor the effectiveness of the technology and is confident in the accuracy of the tool before it is deployed. However, an assessment of the consequences is still useful as a point of departure in a public discussion of the introduction.

OPTION 3	The police	The public	Victims	Offenders
Short-term	More precise risk assessment for victims.	Grounds for optimism regarding a safer environment.	Grounds for optimism regarding leading a safer life.	Attention around the new system may in some cases deter aggression.
Long-term	Hope for better crime prevention in a challenging area.	May suggest a more favourable view of law enforcement in its role of ensuring protection.	Increased security to some extent.	The system may aid LE in seeking out candidates for early intervention.

7. SHORT SUMMARY

The use of AI implies the enhanced effectiveness of an existing LE system, which would have an impact on the lives of victims and offenders. With the AI upgrade, the system is expected to provide more accurate risk assessments for all aggressions, and a human review by trained personnel is always carried out.

The central moral values in this case are safety, fairness and transparency, in addition to the more technical ones (reliability/explainability). After considering the options, proper development using extensive testing, simultaneous use with the old system, and transparently engaging with the public are the methods considered to be the best solution.

CASE 3: MODEL ANALYSIS OF OPEN-SOURCE DATA SCRAPING

(March 2024)

Law enforcement are dealing with an increasing number of reports relating to the theft of power tools from sheds, garages, business premises and vehicles. It is known that many of the stolen items will be listed by thieves on online marketplaces such as eBay or Facebook Marketplace for quick sale.

Law enforcement has limited resources to investigate power tool theft. However, there are individual instances of officers being able to identify stolen goods through manual research on online marketplaces.

It is believed that organised gangs with links to serious and organised crime are stealing the tools.

The LEA is exploring whether it would be possible to use data science techniques to 'scrape' adverts for second hand tools on popular online marketplaces to compare these with a database of stolen power tools and thus identify online accounts suspected of selling stolen goods.

Initial testing of the approach has shown that it can add value to an investigation into power tool theft. However, there are concerns about the 'scraping' of data from online marketplaces being against the host's terms of service and also about the impact of having their data processed by law enforcement in this way on innocent users of these sites.

The LEA is also considering whether it is possible to use machine learning techniques to identify patterns of behaviour on suspicious online marketplace accounts in order to proactively identify those who may be involved in selling stolen goods.

Once such accounts have been identified, the law enforcement agency can commence enquiries with a view to ultimately prosecuting offenders and reuniting victims with their stolen goods.

1. THE MORAL PROBLEM

Law enforcement need to find ways to improve the detection rate for this type of crime. Not only is it a key priority to the public, and therefore crucial to maintaining public confidence, but it is also known to have links to serious and organised crime.

Automating research of online marketplaces using 'scraping' tools is undoubtedly more efficient than manual research by a law enforcement officer. However, this is often against the terms of service of the online marketplace and the protections it has put in place for its customers. Further, if the LEA opts for developing automated tools, to what extent should they be open about it?

2. THE FACTS

The facts relevant to the situation are set out below.

- Theft of power tools is increasing, and low detection rates are having an impact on public confidence.
- A significant portion of these thefts is carried out by persons with links to serious and organised crime.
- It is common for stolen power tools to be resold on online marketplaces.
- There have been occasions when victims have identified their stolen items for sale via online marketplaces and then put themselves in danger when trying to recover the goods.
- Law enforcement agencies do have some success when searching online marketplaces manually to identify specific stolen items. This type of research is lawful under data protection legislation and is permitted by the law enforcement agency's own policies on internet research.

3. PARTIES INVOLVED

LEGITIMATE USERS OF ONLINE MARKETPLACES may have the expectation that they are able to go about their lawful business selling second-hand goods without their data being processed on a large scale by law enforcement. They may have an expectation that law enforcement could research the marketplace they use on a case-by-case basis, but not that the terms of service of the site are breached by law enforcement agencies.

LAW ENFORCEMENT wish to address the low levels of detection for this type of crime to enhance public confidence, and want to use the available technology to

assist with this resource-intensive activity. They may also wish to take advantage of the ability to access intelligence about those engaged in criminal activity through, for example, contact telephone numbers provided on online adverts.

VICTIMS OF CRIME will have an expectation that law enforcement will take all available measures to investigate reported thefts and reunite them with their property.

CRIMINALS USING ONLINE MARKETPLACES TO SELL STOLEN GOODS may expect that selling goods in the 'virtual' world may be less risky than in the 'physical' world and will be aware that law enforcement have not traditionally policed online marketplaces on a large-scale basis.

ONLINE MARKETPLACE HOSTS may be concerned about the reputational impact of their sites being utilised to sell stolen goods, but they may believe that most activity is legitimate. They may have an expectation that their terms of service will be honoured by law enforcement and may believe it to be unethical were that not to be the case.

LAW-ABIDING CITIZENS may not have any strong views on the matter: some will expect law enforcement to use all available tools to investigate crime; others will expect that they are afforded a degree of privacy online and will only have their data processed where there is a specific reason for law enforcement to do so. If law enforcement are going to use advanced data science techniques to improve investigative outcomes, law-abiding citizens will have an expectation that these are legal and effective.

4. VALUES

There are many technical issues to be considered in this case in terms of developing effective data science models. However, in terms of purely moral considerations, based on the description above (1-3), the normative considerations ought to be motivated by the following three values:

- **PUBLIC SAFETY:** in this context, enhancing public safety means identifying and prosecuting those involved in the sale of stolen goods and reducing the harm caused by serious and organised crime. It can also refer to the public safety benefits of having effective measures in place that will reduce the likelihood that victims of crime will do their own research and place themselves in danger trying to recover stolen items.
- **PRIVACY:** public safety benefits must be balanced with the right of law-abiding citizens to privacy to conduct their business online, especially when they believe that the terms of service of an on-line marketplace afford them protection from large-scale data scraping.
- **FAIRNESS:** in this context, fairness refers to the LEA using proportional measures to identify stolen items and identify sellers. Proportionality may refer both to the intrusiveness of the measure and the amount of resources allocated. Fairness may also refer to bias in terms of which types of goods are targeted, i.e. which groups are attracted to this type of goods. This aspect is relevant both to automated and manual searches.
- **TRANSPARENCY:** this refers to law enforcement striking the balance between being transparent about the use of data scraping techniques to identify those involved in selling stolen goods on the 'clear' web and not alerting criminals to this intelligence source and pushing them towards the 'dark' web or other methods of disposing of stolen goods.

5. OPTIONS

The following options are available[18].

1. Do nothing and continue to manually research online marketplaces on an ad hoc basis with limited resources when investigating specific crimes.
2. Dedicate more resources to this type of investigation to allow for more frequent and in-depth manual research on online marketplaces.
3. Develop an AI model that can be used to 'scrape' online marketplaces to search for specific stolen goods when investigating specific incidents.
4. Develop an AI model that can be used to 'scrape' online marketplaces to search for specific stolen goods when investigating specific incidents AND that can learn to recognise patterns of behaviour consistent with suspicious accounts allowing for proactive investigation of potential sale of stolen goods.

OPTION 1 does not fit perfectly with the value landscape documented in Section 4, as public safety could be enhanced by using web scraping techniques if these led to a reduction in this type of crime and prevented members of the public from putting themselves in danger. Manual research of online marketplaces is lawful and assumed to be acceptable to the public with no negative impacts on privacy and transparency. This option also seems reasonable in terms of proportionality, as manual research is relatively unintrusive.

OPTION 2 could be the most effective in terms of supporting law enforcement enquiries, and therefore enhancing public safety, but the LEA may have other strategic and operational priorities that mean there are no additional resources available. This option could have an impact on public safety and fairness if resources are diverted from elsewhere to investigate these specific types of crime.

OPTION 3 could be seen to enhance public safety but the large-scale processing of online data, even when it is publicly available, could have a negative impact on privacy and potentially also fairness. The law enforcement agency may choose to limit the information they share with the public about this new and innovative capability to avoid alerting criminals that they are more likely to be identified, which could be viewed as contrary to the value of transparency.

OPTION 4 goes further than option 3 in terms of the potential public safety benefits as it could provide proactive opportunities for law enforcement to identify those involved in the sale of stolen goods, even when the individual thefts themselves have not been reported to police. However, this comes with the risk of further intrusion on privacy and a higher impact on the value of transparency as it is assessed that law enforcement will be even less likely to share the existence of the capability with the public. In line with the value of transparency, it will be important that end users of the tool understand why a particular account has been identified as suspicious and worthy of further enquiry.

There are no ideal options in this scenario as there is a requirement to trade off public safety benefits against privacy and transparency considerations on a sliding scale, i.e. the greater the public safety benefit, the higher the impact on privacy, transparency and, depending on the system, fairness as well.

Despite the impact on public confidence and the links to serious and organised crime, it is assessed overall that option 1) is the only acceptable option at this time as the risks to the values of privacy and transparency introducing data scraping tools to assist these investigations are too great.

6. JUSTIFICATION

Rightness

There are four questions that can assess the rightness of the suggested options:

1. Is the suggested line of action (use of technology) always appropriate under similar conditions? There may be times when the use of the technology would be appropriate, for example in the investigation of more serious crimes.
2. Does the intervention imply that the LEOs use their professional authority in the best interest of the clients and not just as a means to fulfil other goals? Yes, it is assessed that the LEOs who used the tool would believe they were acting in the best interests of citizens even if they were not considering privacy and transparency concerns.
3. Is the action/practice acceptable to the public? This is difficult to assess as some members of the public may feel very strongly about the privacy implications of data scraping by law enforcement on websites where the public go about their lawful business. Others would expect police to use any technology available to enhance public safety and investigate crime, and others still would consider developing expensive systems for targeting small-time criminals to be unjust.
4. Do the necessary competencies and resources exist to take responsibility for the intervention? Yes, in terms of competence, as law enforcement interventions will not change significantly. Although the tool will save resources in terms of research of online marketplaces in individual cases, there may also be a requirement to devote extra resources if the tool is used to identify suspicious activity proactively.

Consequences		
OPTION 1	The police	The public
Short-term	No change to crime detection rates or current outcomes of theft enquiries.	Will continue to see low rates of detection and may put themselves at risk to reclaim their stolen property.
Long-term	Lost opportunities to disrupt those involved in serious and organised crime.	May become frustrated at lack of police action for this type of enquiry and perceive an unwillingness to embrace new technology.

OPTION 4	The police	The public
Short-term	Reduction in time spent manually conducting online research. More detections of theft.	May notice improvement in detection rates and be more inclined to report thefts.
Long-term	Enhanced intelligence opportunities and proactive investigation of potential sale of stolen property. Change in crime trends.	Erosion of privacy rights and confidence in police if widespread data scraping becomes known and causes controversy.

From the information available at this time, based on the values and permissibility considerations summarised above, the development of a data scraping tool to support investigations into stolen property has not yet been justified.

The law enforcement agency should take further steps to explore the legality of large-scale data scraping and consider how it could be more transparent about the use of this type of data science technique in general terms without compromising operational security.

Further exploratory work should also be done to consider the resource implications of the tool and whether it will deliver a benefit that justifies the ethical risk.

Law enforcement should also explore other, less intrusive, ways to utilise technology to enhance this type of investigation.

If it can be confirmed that the use of the tool is lawful, and the resources exist to properly utilise the output to enhance public safety, then this assessment could be refreshed and may have a different outcome.

7. SHORT SUMMARY

In this case, the problem is whether automated research of online marketplaces using 'scraping' tools is morally acceptable since it is against the terms of service of the online marketplace and customer protections. The marketplace users, victims, criminals, the LEA and the hosts are affected by this decision, which should be based on public safety, fairness, privacy and transparency. As it is, this automated method of investigation is unlawful and therefore unacceptable.

CASE 4: USING A CHATBOT TO PREVENT CHILD SEXUAL ABUSE ONLINE

(March 2024)

The LEA is considering using 'PrevBot' – a machine learning tool applied to natural language processing – to prevent child sexual abuse (CSA) online. The idea is to detect grooming occurring in chat channels and transmit a warning to the adult in the chat, hoping this will make the person desist from further attempts of grooming. PrevBot alerts the human operator about indications of grooming, who then decides about intervention.

PrevBot is trained to identify conversations that are sexually charged and predict participants' age and gender. It may also perform sentiment analysis and author identification by computing 'linguistic fingerprints' that can be matched against linguistic fingerprints of previous CSA convicts stored in a reference database kept by the LEA.

The reason why the LEA wants to deploy PrevBot is that CSA is an increasing problem, and that criminal investigation and prosecution has proved inadequate to handle online CSA. New technology such as PrevBot is needed to maintain an effective police presence in online forums. PrevBot adds capacity as it can monitor a great number of conversations and yield information that is otherwise not accessible due to the anonymity of online users.

1. THE MORAL PROBLEM

There are several moral problems in this case. The present assessment focuses on two in particular.

Intrusive measures

PrevBot must process the data of all conversations in the chat room to identify those that present a risk of CSA. Consequently, processing chat data of harmless users will take place. Using PrevBot means secret surveillance, which, from a control perspective, is normally regarded as a threat to privacy and freedom of speech.

Uncertainties in the form of transparency and effectiveness

- A (preventative) warning does not remove the threat as the perpetrator may easily switch to a different forum and resume his grooming efforts.
- Given the sensitive nature of the problem, PrevBot cannot be tested in a real-world environment. The accuracy of the tool cannot be established in advance. False positives may affect innocent (young) persons, and too many false negatives means that it is not effective.
- The opaque quality of deep learning involves a black box problem.

2. THE FACTS

The facts of the situation are set out below.

- Children online are easily accessible to CSA perpetrators, and their parents are usually unaware of the abuse, as children seldom report the abuse to anyone.
- CSA perpetrators often have many victims, sometimes in the hundreds.
- PrevBot is deployed by a human operator who decides in which domain/forum/chatroom it will be present. When PrevBot enters the domain, its user profile is shown. If contacted, PrevBot will provide brief replies, but will not initiate contact with anyone.
- If PrevBot is triggered, the preventative warning also contains a link for complaining to the police, if the target feels that the warning is misplaced.
- PrevBot has been tested in an environment including both adults and teenagers. The overall accuracy was deemed satisfactory.
- The LEA has participated in a sandbox process with the National Data Protection Authority to

evaluate all privacy issues related to the tool.

- PrevBot's Tsetlin-based models for identifying sexualised speech and performing sentiment analysis are considered transparent, i.e. the input/output correlation is explainable.

Missing information:

- The accuracy of the tool when deployed online. Its features may also vary with respect to accuracy, as research shows high accuracy for content detection and age/gender categorisation, while the results for sentiment analysis and linguistic fingerprints are more uncertain.
- The stability of linguistic fingerprints over time is not known.
- Whether the tool will be fast enough to enable the human operator to intervene while the suspicious chat goes on.

3. PARTIES INVOLVED

CSA PERPETRATORS. Their interest is to remain anonymous and keep on abusing children.

PERSONS WHO WRONGFULLY RECEIVE A PREVENTATIVE WARNING. They will want to be presumed innocent and to be able to chat freely with other persons in the chatroom. If wrongfully accused, they will probably want to contest the claim and receive an explanation and even an apology.

CHILDREN ONLINE. It is in children's interests to be safe when they explore the internet, even when they are exploring their sexuality online. PrevBOT may prove helpful in that respect

PARENTS. In relation to their children, parents have a strong interest in a safe internet. They expect the police to make use of available measures to protect children from abuse. They also expect the police to be technologically competent and able to exploit new technology in a responsible manner.

CITIZENS. Citizens have an interest in being able to express themselves freely online and are concerned about the potential for surveillance.

LEA. Law enforcement officers are concerned about their inability to counter CSA online effectively. However, they are also concerned about the negative impact on citizens' trust if PrevBOT is seen as 'overpolicing' by processing everyone's personal data on different forums (and not just perpetrators' data).

4. VALUES

Step 1 identified two general moral problems, each of them challenging several moral values. These values are the point of departure for discussing a relevant set of values, which is identified as the following:

PRIVACY. Data protection is a dimension of privacy that is affected by PrevBOT's processing of the chat data of all participants in the forum. However, data not triggering any alert is deleted immediately, making the processing resemble a fleeting observation in a publicly accessible space. Data that triggers an alert must be reviewed by the human operator and remain stored until processed. In the case of a preventative warning, data must be stored for a fixed period due to the need for documentation.

SAFETY. Children's safety online is an important value. This value is challenged if CSA perpetrators move to different platforms after they receive a warning or remain undetected by PrevBOT because of lack of accuracy. To reduce the perpetrator's opportunities, it is important that PrevBOT is scalable and can monitor many platforms at the same time. PrevBOT's categorisation of user's age and gender breaks through online anonymity and is thus vital to the effectivity of the tool.

FAIRNESS. Fairness is threatened due to the uncertainty about PrevBOT's level of accuracy. Praising the LEA for being proactive would not be fair if the children were still no safer due to a high error rate. The problem of targeting a wrong person may be remediated by a complaint service and an apology. Feedback could also provide useful data to recalibrate the tool. It is important that PrevBOT only targets adults. While sexually aggressive youngsters should receive a reaction, the phenomenon is different from that of adults grooming children, and might thus have to be countered with other measures than use of PrevBOT.

TRANSPARENCY. The capabilities of categorising a participant's age and gender and computing linguistic fingerprints hinge on deep learning processes. The way in which the output is generated in a concrete case cannot be explained, as it is not known how the neural network combines and weighs its vectors. Detection of sexualised speech and sentiment analysis is identified by the machine learning algorithm the Tsetlin Machine. The Tsetlin Machine is transparent in the sense that, in a concrete case, the LEA may explain how the output was generated.

ACCOUNTABILITY. Given that PrevBot is a tool for secret surveillance of chat rooms, accountability becomes paramount. Use of the tool must be logged at all times, and the identity of the human operator must be known. It is also important that the personnel tasked with using PrevBOT have been given adequate training.

5. OPTIONS

In this scenario the following options are considered:

1. Not using PrevBot as its accuracy is not known.
2. Start using PrevBot and prepare for recalibration as soon as possible.
3. Start using PrevBot in a limited version, only applying content detection and categorisation of age and gender. In addition, it should be calibrated for targeting persons above the age of 30.

The following options are assessed to determine whether they promote the values of Step 4.

OPTION 1: This alternative promotes the privacy of citizens online, at the cost of doing nothing to promote the safety of children online. It thus represents an LEA that only protects a class of citizens which is already more resourceful than children. This seems unfair and fails to promote transparency and accountability. This option should therefore not be considered further.

OPTION 2: This option involves using PrevBot to the maximum of its capabilities and recalibrating the tool once sufficient experience and data are gained. Although the option cannot be said to promote citizen's privacy, the interference can be minimised by immediate deletion of data unrelated to CSA. Moreover, a safer internet benefits the dimension of children's privacy relating to the development of one's personality by entering into and exploring new social relationships. By protecting this vulnerable group, fairness is also promoted. LEA can be transparent about using the system, about how output from the Tsetlin Machine is generated and about the assessments of the human operator. Accountability is promoted in the sense that the system is operated by trained personnel and robust procedures for logging etc., are implemented. However, there is a deficit in deploying the tool while knowing that its accuracy is in question.

OPTION 3: This option involves use of a limited version of PrevBot. By setting the age threshold to 30, the risk of targeting youngsters is minimal. The limited tool will miss adult perpetrators below the age of 30 yet promote children's safety with respect to perpetrators that are older. The efficiency is generally lower compared to the full version of PrevBot, but the vital function of age and gender categorisation is used. Accountability comes out better than option 2) because of the reduced risk of targeting youngsters. Apart from this, the considerations are the same as for option 2.

Options 2 and 3 are both in alignment with the set of values stated in Step 4 and will be scrutinised further in Step 6 in terms of rightness and consequences.

6. JUSTIFICATION

Rightness

There are four questions that can assess the rightness of the suggested options.

1. Is the suggested line of action (use of technology) always appropriate under similar conditions? It is problematic that PrevBot has not been tested in a real-world environment. Testing in the form of an experiment does not guarantee the same output as in real life, for instance due to exaggerations and lack of knowledge about online jargon. To target persons using technology that is not adequately tested is obviously not a line that can be held over time. The question is whether there are extraordinary circumstances that could nevertheless justify the use of PrevBot. It seems relevant that the CSA problem calls for new innovative approaches from LEAs. It could be better first to deploy a limited version of PrevBot to gain experience. This implies that option 2) is acceptable and option 3) the stronger alternative.
2. Does the intervention imply that the law enforcement officers use their professional authority in the best interest of the clients and not just as a means to fulfil other goals? LEAs feel impotent in the face of online CSA and are seeking new measures to deal with it. Of course, successful deployment of PrevBot would gain much praise, but there is no reason to assume that the LEA does not have the best intentions in this case.
3. Is use of the technology acceptable to the police if it becomes generally known? The LEA will inform the public before deployment of PrevBot and may be open about its features. Much information is already public, resulting from the sandbox process with the Data Protection Authority. Some would probably react as a matter of principle, and worry about a slippery-slope effect.
4. Do the necessary competencies and resources exist to take responsibility for the intervention? Operators of the system must be trained for the task. They must be aware of the uncertainties of the system and be able to independently assess the output in context, and also overlook alerts if they are not convinced of a risk of grooming in the concrete case.

Both options 2 and 3 are thus justifiable, provided the training required in iv) is performed.

Consequences		
OPTION	Option 2	Option 3
Short-term	The full-scale system might detect many perpetrators but may also make many errors. It puts a heavy burden on the human operator. If the operator becomes sceptical about the output, PrevBot may not be used according to its potential. The short-term gain is uncertain.	The limited version of the system may not produce as many alerts as the full-scale version. It will still be more effective against CSA than the alternative of manual patrolling. There is a short-term gain.
Long-term	Mistrust of the police may be caused if the LEA often makes mistakes when targeting persons. It may also create a sense of unwanted surveillance that could have a cool-down effect on speech. These negative effects are serious even if the system is capable of targeting many CSA perpetrators.	The limited version carries less risk of errors, causing less concern about the negative effects associated with the full version. However, using the system is hard to defend if it is not effective.

We are lacking some important information about PrevBot's effectivity online. However, the risk of serious long-term effects should weigh in heavily, as the loss of trust in the police may be hard to remediate. At present, option 3 thus seems to be the better choice.

7. SHORT SUMMARY

In the case where the LEA wanted to use PrevBot to prevent CSA online, the moral concerns were initially that the chat data of all persons in the forum would be processed, that a mere preventative intervention would not incapacitate a CSA perpetrator, the uncertainty about accuracy due to lack of testing, the black box problem of deep learning and the risk of unwanted surveillance. The most important parties were the CSA perpetrators, persons who are wrongfully targeted by the LEA, children online, parents, citizens in general and the LEA. The most important values were privacy, children's safety online, fairness, transparency and accountability. Based upon these values, three options for further action were considered, of which option 1 was rejected. Option 3 was considered to be in line with the suggested values, permissible by the rightness test, and having better overall consequences than option 2.

CASE 5: AUTOMATED ANALYSIS OF LARGE AND COMPLEX DATASETS

(March 2024)

Today, the potential benefit of data analysis for intelligence is widely accepted by governments and national security services, as well as businesses. The extent of data analysis has grown in recent years, as more data is available and tools for conducting analysis have become more powerful. Automated tools have created enhanced opportunities for crime detection and investigation, but have also led to apprehension within communities, governments and even among law enforcement authorities. Have the potentials of surveillance become excessively intrusive? Beyond the traditional issues of privacy and consent associated with the collection and processing of large and complex datasets, what exactly is automation, and how is it used?

The present case study does not refer to some specific technology but in general discusses automation, and large and complex datasets.

1. THE MORAL PROBLEM

Collection and aggregation of multiple data sources is potentially problematic for numerous reasons, including evolution of purpose or 'function creep', meaning where data in third party sources is collected for one purpose but is used for another. For instance, CCTV recordings made at a retail chain's outlet, ostensibly for security purposes, are repurposed for customer analysis. In addition, datasets are rarely static but evolve in line with changing business requirements or relevant legislation. This can subtly change the data coming through, undermining the assumptions and controls put in place in the first place. We might therefore also talk about evolution of nature regarding large and complex datasets.

The intelligence value of combined datasets can also outweigh the sum of individual sources, making such aggregations of high interest not only to law enforcement but also to crime groups and hostile state actors. Combined datasets can therefore be considered to be high-value 'honeypots'. The users handling the aggregated dataset should be able and willing to provide explanations to the public regarding the usage of the data, for instance, by providing explanations regarding the purpose of data usage and details on how the sources are obtained and managed. However, this is not always the case.

The evolution of data sets and lack of transparency create moral problems, particularly when these phenomena are found in a context of law enforcement using automated tools. For one, when things go wrong, who is accountable for the automated tool – the user, the host, the seller/supplier? In the case of open-source software, is there a supplier? Further, law enforcement practitioners may be called on publicly to explain their actions and methods. What if an underlying tool is so complex that it verges on being unexplainable, or if the author of the tool is unwilling to provide an explanation? Are the operators sufficiently trained to understand the processes and tools being used for analysis, including known weaknesses and the potential for new ones? Are they able to sufficiently implement suitable safeguards for the mission at hand?

2. THE FACTS OF THE CASE

Rather than mere logical size such as number of bytes or rows/records, we interpret large and complex datasets to refer to the number and nature of sources combined to form a dataset, with the potential aggravating factor of rapid change – more stuff arriving and things evolving at a speed faster than humans can keep up with.

'Automated' analysis is the entrusting of machines to undertake tasks such as extracting context and meaning from the data on our behalf.

3. PARTIES INVOLVED

The nebulous and dynamic nature of large and complex datasets makes it almost impossible to list all the parties involved. But broadly, the following are directly relevant to the problems mentioned above in a law enforcement context.

PERSONS APPEARING IN POLICE INDICES: individuals and organisations recorded within existing police information systems, be they suspects, offenders, witnesses or even law enforcement personnel themselves. Whereas such datasets' use and confidentiality tend to be strictly regulated, persons recorded in them can be reasonably expected to hold a degree of interest and concern regarding how their information is being used beyond the initial purpose – for example, a witness whose records are considered in subsequent searches related to a separate event/offence.

PERSONS IN SECONDARY SOURCES: individuals and organisations appearing in other datasets not owned/operated by law enforcement, but whose records may be ingested under the authority of data sharing agreements, licence or warrant/court order. This can therefore include concepts such as open-source datasets published under 'Creative Commons' type licences. Crucially, given the third party and opaque nature of some datasets, it is entirely feasible that persons are recorded without their knowledge or consent. This is a known issue within academia, with cases of intimate (and potentially illegal) imagery identified in 'trusted' sources such as Imagenet[19]. It is reasonable to assume that persons whose information is included within such sources would have an expectation of the appropriate use of their data, regardless of whether they actually know it is happening.

DATASET/SYSTEM OWNERS: owners/operators of datasets have an interest in ensuring law enforcement's access and use of their data is both lawful and within the expectations of the persons whose data is recorded. Whereas legislation such as the GDPR sets strict legal requirements, commercial considerations also play a part, particularly if such law enforcement access leads to a loss of customers/contributors.

VICTIMS OF CRIME: victims of crime have a right to expect law enforcement to undertake all reasonable efforts to investigate and prosecute offenders. If technology exists that enables law enforcement to automate analysis at a scale otherwise unfeasible, then arguably they may expect it will be undertaken.

GENERAL PUBLIC: the public has a right to expect police to enforce laws equally, utilising all lawful and reasonable means to do so. The reasonableness of police sourcing additional data for investigative use will be open to debate, with the anticipation that specific uses of such data will heavily influence this measure.

Law enforcement: police have a duty to make all reasonable efforts to enforce laws, and the electronic nature of modern data capture makes the collection of potential intelligence and evidence far more technically feasible. It is inherent upon police to adopt and utilise technology capable of aiding in their duties and mission.

4. VALUES

Regardless of which technology is used in automation, the moral values of concern largely mirror those in guidelines for developing AI (see for instance ANZPAA Artificial Intelligence Principles)[20]. The following moral norms in particular need thorough consideration in automation, given the characteristics of large and complex datasets:

- **PRIVACY AND SECURITY:** is the data under analysis adequately secured from internal and external threats, accurate, and accessible only to authorised users for approved purposes?
- **TRANSPARENCY (IN TERMS OF ACCOUNTABILITY):** beyond those for policing operators, what measures have been taken to ensure that the data sources being used are ethical, accurate and lawful? What right do persons recorded in these sources have to know that their data has been stored and is now being used by the police?
- **FAIRNESS (IN TERMS OF PROPORTIONALITY):** is the inclusion and use of specific data sources reasonable in the circumstances? The value of different applications varies, even when the technical implementation is identical. Compare, for example, the reasonableness of utilising facial recognition to identify a jaywalker to, say, a suspected war criminal.

5. OPTIONS

Since the emphasis here is not on any specific technology, we will simply start by exploring the most extreme options and a middle-ground approach:

1. Do nothing. Bar the use of external data sources and/or automated analysis.
2. Free for all. Allow the unrestricted use of automated tools across lawfully accessible datasets.
3. Middle ground. Identify an acceptable balance of self-regulation and external oversight, ensuring usable efficacy whilst also maintaining restraint.

Option 1 is obviously unacceptable, as it denies law enforcement a series of potentially valuable capabilities that could detect and investigate criminality. Likewise, option 2 is unacceptable, due primarily to the undermining of public trust in law enforcement. In addition, investigations may run into severe problems if challenged in court if option 2 is followed.

A middle-ground approach that addresses concerns of privacy, fairness and transparency thus seems to be the only morally permissible option. There is no 'one size fits all' approach to what forms regulation and oversight should take, nor to their underlying implementation. The 'middle ground' has a range of several options, as long as they are based on the set of values. Further, the specifics of acceptable middle-ground solutions will vary between jurisdictions, but also over time – datasets and capabilities evolve, as do the community's expectations around their use. Nevertheless, the set of values must be promoted as far as possible.

The full reflection model requires justification of a choice in terms of rightness (deontology) and goodness (consequences). This is impossible to accomplish without having some particular technology and a context at hand. However, the middle-ground approach may as such be endorsed as 'right' as it seems like a responsible line to choose under normal circumstances, at least if it is for the benefit of the citizens, and because it is probably publicly acceptable – at least if competent personnel are overseeing the data and the automation processes. Consequences may also be considered similarly from a perspective of rule utilitarianism, where the middle ground seems like the more sustainable option.

References

Aristotle. (1985). Nicomachean Ethics (T. Irwin, Trans.). Hackett publishing company.

Beauchamp, T. L., & Childress, J. F. (2009). Principles of Biomedical Ethics (6 ed.). Oxford University Press.

Christian, B. (2020). The Alignment Problem: Machine Learning and Human Values. W.W. Norton & Company.

Flanagan, M., Howe, D. C. H., & Nissenbaum, H. (2008). Embodying Values in Technology. In J. v. d. Hoven & J. Weckert (Eds.), Information Technology and Moral Philosophy (pp. 322-353). Cambridge University Press. https://nissenbaum.tech.cornell.edu/papers/embodying_values.pdf

Haidt, J. (2012). The righteous mind : why good people are divided by politics and religion (1st ed.). Pantheon Books.

Kahneman, D., & Klein, G. (2009). Conditions for Intuitive Expertise. A Failure to Disagree. American Psychologist, 64(6), 515-526. https://doi.org/DOI: 10.1037/a0016755

Kleinig, J. (1996). The ethics of policing. Cambridge University Press.

Laas, K., Davis, M., & Hildt, E. (Eds.). (2022). Codes of Ethics and Ethical Guidelines. Springer.

Matthys, J., Elwyn, G., Van Nuland, M., Van Maele, G., De Sutter, A., De Meyere, M., & Deveugele, M. (2009). Patients' ideas, concerns, and expectations (ICE) in general practice: impact on prescribing. Br J Gen Pract, 59(558), 26-36. https://doi.org/10.3399/bjgp09X394833

Mazerolle, L. G., Sargeant, E., Cherney, A., Bennett, S., Murphy, K., Antrobus, E., & Martin, P. (2014). Procedural justice and legitimacy in policing. Springer.

Nagel, T. (1986). The view from nowhere. Oxford University Press.

Scheffler, S. (2010). Equality and tradition : questions of value in moral and political theory. Oxford University Press.

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). Ethics of Artificial Intelligence. Case Studies and Options for Addressing Ethical Challenges. Springer. https://doi.org/10.1007/978-3-031-17040-9

Troyer, J. (Ed.). (2003). The Classical Utilitarians: Bentham and Mill. Hackett Publishing Company.

Williams, B. (1993). Ethics and the limits of philosophy. Fontana Press.

Endnotes

1 More on the alignment problem in Christian, B. (2020). The Alignment Problem: Machine Learning and Human Values. W.W. Norton & Company.

2 The method is based on insights from John Dewey (1910, p. 72), Ralph B. Potter's so-called 'Potter Box' (Potter 1965), the more recent 'National Decision Model' of the British Police Code of Ethics (2014), and the values-based practice strategy of Fulford (2008). Above all, it is inspired by the method of Clinical Ethics Committees in Norway (Ruyter, Førde, & Solbakk 2014), and practical experience with adjusting and applying a similar manner of moral reasoning to the LE context (Paulsen 2019, 2020, 2021). For the purpose of the present ethical guidelines the method is modified to a technology assessment tool.

3 See Haidt, J. (2012). The righteous mind: why good people are divided by politics and religion (1st ed.). Pantheon Books.

4 In what if-scenarios, e.g. the hypothetical application of technology, step 2 is typically prior to step 1.

5 See Matthys, J., Elwyn, G., Van Nuland, M., Van Maele, G., De Sutter, A., De Meyere, M., & Deveugele, M. (2009). Patients' ideas, concerns, and expectations (ICE) in general practice: impact on prescribing. Br J Gen Pract, 59(558), 26-36. https://doi.org/10.3399/bjgp09X394833.

6 Nagel, T. (1986). The view from nowhere. Oxford University Press. , Williams, B. (1993). Ethics and the limits of philosophy. Fontana Press.

7 Good Governance Institute. (2020, 1 June). The Nolan principles. Good-governance.org.uk. https://www.good-governance.org.uk/publications/insights/the-nolan-principles. The Nolan principles apply to public services generally, however, transparency as a principle is also included in European ethical codes of police and prosecutors adopted by the Council of Europe, see Recommendation Rec(2001)10 The European Code of Police Ethics, 19 September 2001 (ECPE), point 19 ' Police organisations shall be ready to give objective information on their activities to the public ...', explained by Commentary: 'The police should be as transparent as possible towards the public. A readiness by the police to disclose information on its activities is crucial for securing public confidence'; and Commentary to point 59 about accountability and control of the police: 'Generally, openness and transparency of the police are (...) basic requirements for accountability/control to be effective.' Regarding prosecutors, transparency is a check against abuse of the independence and autonomy ensured to them in performing their duty, see Recommendation Rec(2000)19 The Role of Public Prosecutors in the Criminal Justice System, 6 October 2000, Commentary to point 11: '[a]ll public prosecutors - because they act on behalf of society - must give account of their work at local or regional level, or indeed national level if the service is highly centralised. These regular accounts must be made to the general public ...'. This is followed up in the Rome Charter, Opinion No. 9 (2014) of the CCPE on European norms and principles concerning prosecutors, notably point VII 'Transparency in the work of prosecutors is essential in a modern democracy."

8 Reflected in ECPE point 19: 'Police organisations shall be ready to give objective information on their activities to the public, without disclosing confidential information' (italics added) and Commentary : ' the police must respect confidentiality for a number of reasons; integrity of persons, crime investigation reasons, the principle of the presumption of innocence, security reasons etc. Obviously, even if situations like those described are well regulated in most states, there will always be a margin of appreciation left to the police in striking the balance between the two...'; The Rome Charter point IX mentions the 'confidentiality of investigations' alongside 'the principle of transparency'.

9 LEAs must think carefully about the information given to the public in order to minimize the risks of being misunderstood, or that the information is misused.

10 The General Data Protection Regulation (2016/679) (GDPR) affords the data subject a right to information from the data controller and a right to access information held by the data controller, cf. GDPR Chapter 3. Corresponding rights are included in the Law Enforcement Directive (2016/680) (LED) although in a circumscribed fashion to protect the purposes of LE action, public and national security and the rights and freedoms of others, cf. LED Chapter III. The term 'fair processing notice' is used e.g. in the UK National Health Service (https://www.nelft.nhs.uk/fair-processing-notice/).

11 Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). Ethics of Artificial Intelligence. Case Studies and Options for Addressing Ethical Challenges. Springer. https://doi.org/10.1007/978-3-031-17040-9.

12 The term 'surveillance capitalism' was coined by Zuboff, S. (2019) The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power. Profile Books.

13 The right to privacy is stated in the Charter of Fundamental Rights of the European Union (2012/C 326/02) Article 7, and in the European Human Rights Convention Article 8. Data protection is recognised as a dimension of the right to privacy, yet also as a distinct right, as expressed in the Charter Article 8.

14 Reference is made to the Accountability Principles for Artificial Intelligence, (AP4AI, ap4ai.eu) project. This project develops solutions to assess, review and safeguard the accountability of AI usage by internal security practitioners in line with EU values and fundamental rights. AP4AI offers a robust and application-focused Framework that integrates security, legal, ethical as well as citizens' positions on AI to the internal security community, in particular through its spin-off, CC4AI (Compliance Checker for Artificial Intelligence, cc4ai.eu). CC4AI is a web-based tool to support internal security practitioners assess compliance of their AI systems with the requirements of the EU AI Act. This will allow users to evaluate whether, existing or future applications, meet the criteria set by the new regulatory framework. AP4AI and the Guidelines at hand provided by the Strategic Group on Technology and Ethics, complement each other as AP4AI is designed specifically for the development or procurement of AI based tools whereas the Guidelines could be applied in a broader set of circumstances (not only AI but also other technology); and AP4AI focuses on documenting the findings in a structured self-assessment, whereas the SG ethics focuses on the (interactive) steps to be followed.

15 E.g. the principles-based framework for accountability set forth in the project Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain. Akhgar, B. et al. (2022) AP4AI Framework Blueprint, Version 22 February 2022, CENTRIC.

16 The four questions are inspired by Immanuel Kant's categorical imperative.

17 Inspired by a Benthamite act-utilitarian approach Troyer, J. (Ed.). (2003). The Classical Utilitarians: Bentham and Mill. Hackett Publishing Company.

18 There are other options available in the wider context of reducing rates of acquisitive crime however these options are focused specifically on the identification of stolen items on online marketplaces.

19 https://www.theregister.com/2019/10/23/ai_dataset_imagenet_consent/ (Visited 26 Feb 2024).

20 See https://www.anzpaa.org.au/homepage-announcements/australia-new-zealand-police-artificial-intelligence-principles (Visited 26 Feb 2024).