Intelligence Network & Secure Platform for Evidence Correlation and Transfer

D8.8 Guide on privacy and ethics-by-design in law enforcement technology

Document Summary Information

Grant Agreement No	833276	**Acronym**	INSPECTr
Full Title	**I**ntelligence **N**etwork & **S**ecure **P**latform for **E**vidence **C**orrelation and **Tr**ansfer		
Start Date	01/09/2019	**Duration**	42 months
Project URL	https://www.inspectr-project.eu		
Deliverable	D8.8 Guide on privacy and ethics-by-design in law enforcement technology		
Work Package	WP 8		
Contractual due date	28/02/2023	**Actual submission date**	28/02/2023
Nature	R	**Dissemination Level**	PU
Lead Beneficiary	TRI		
Responsible Author	Dr. Joshua Hughes		
Contributions from	Dr. David Barnard Wills; David Wright, Ben Howkins, Irma Poder		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 833276.

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	01/12/2022		Initial draft	Joshua Hughes, David Barnard-Wills, David Wright, Ben Howkins, Irma Poder
V0.2	16/01/2023		Section 2, 3, 6	Joshua Hughes
V0.3	17/01/2023		Internal Review	David Barnard-Wills
V0.4	01/02/2023		Peer review	Ray Genoe
V0.5	08/02/2023		Peer review	Melania Tudorica
V0.6	25/02/2023		Peer review	Daragh O Brien
V0.7	27/02/2023		Update following reviews	Joshua Hughes
V0.8	28/02/2023	100	Final edits	Joshua Hughes

Disclaimer

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the INSPECTr consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the INSPECTr Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the INSPECTr Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

Table of Contents

List of Tables

Glossary of terms and abbreviations used

Abbreviation / Term	Description
CCI	UCD Centre for Cybersecurity and Cybercrime Investigation
DMP	Data Management Plan
DPA	Data Protection Authority
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EAB	Ethics Advisory Board
EC	European Commission
EDPS	European Data Protection Supervisor
GDPR	General Data Protection Regulation
LEA	Law Enforcement Authority
LED	Law Enforcement Directive
LIA	Legitimate Interests Assessment
LSG	Law Enforcement Authority Steering Group
POPD	Processing of Personal Data
TRI	Trilateral Research
UCD	University College Dublin
WP	Work Package

1 Introduction

The aim of this deliverable is to introduce the concepts of Ethics-by Design and Privacy-by-Design to the INSPECTr project, to explain how these design approaches are incorporated into the INSPECTr project, and to show some of the design solutions that have been developed using these approaches so far in the project.

1.1 Mapping INSPECTr Outputs

The purpose of this section is to map INSPECTr Grant Agreement commitments, both within the formal deliverable and task description, against the project's respective outputs and work performed.

Table 1: Adherence to INSPECTr GA Deliverable & Tasks Descriptions

INSPECTr GA Component Title	INSPECTr GA Component Outline	Respective Document Chapter(s)	Justification
DELIVERABLE			
D8.8 Guide on privacy and ethics-by-design in law enforcement technology	*Final report on ethical, legal and social issues assessment and guidance targeted to external stakeholders for law enforcement technology. Includes best practices applied in LEA and digital forensics contexts.*	*2, 3, 4, 5*	*Ethical, legal, and societal issues are considered throughout the document. Specific guidance for external stakeholders on law enforcement technologies is provided in Sections 2, 3, and 4. Best practices are provided in Sections 4, and 5.*
TASKS			
T8.3 *Privacy-by-design and Ethics by design for INSPECTr tools and platforms*	*Provide Privacy-by-Design and Ethics-by-Design support to infrastructure and analysis tool development (primarily in WP3 to WP5) as needed, in an ongoing, responsive and agile basis. This is an iterative process following the known best practices, and parallel case studies that deal with potential privacy, ethical and social impacts. Identify potential impacts at the different levels of the design process and mitigate negative impacts. This will include:*	2, 3, 4, 5	Much of this work was provided in D8.7. However, Privacy and Ethics-by-Design support is provided in Section 2. Best practices and case studies are provided in Section 5. Impacts arising from the design process are identified throughout the document, with specific mitigation measures explored in sections 2, 3, 4, and 5. In response to concerns regarding ethical queries on: data usage, Section 4 was developed; exploitation of INSPECTr technologies, Section 5 was developed. In identifying emergent privacy and ethics issues, Section 2 was developed as the need to adapt current methodologies to the LEA context was

	• *Responding to legal or ethical queries as they emerge from the design and development process* • *Monitoring the technology design and development processes to identify any emergent privacy or ethics issues and collaborating to produce design solutions.* • *Acting as an internal stakeholder for privacy and ethics related issues in the project.* • *Identifying relevant resources of designers and developers (e.g. privacy-protecting design patterns).*		identified during the project. As part of acting as an internal stakeholder, TRI discussed issues related to each section of the document, especially data lifecycles and exploitation covered in Section 4 and 5, respectively. In identifying relevant resources, TRI identified the best approaches to privacy and ethics-by-design available prior to the project and worked on enhancing them, as shown in Section 2. Resources to assist partners in an exploitation risk assessment are provided in Section 5.

1.2 Deliverable Overview and Report Structure

Following this introduction, Section 2 explains how Privacy and Ethics-by-Design can be adapted for LEA technologies, with specific reference to work done in the INSPECTr project.

Section 3 explores how the regulation of LEA use of AI systems could learn from LEA use of firearms, and whether this could be an alternative model for AI ethics that is already used within LEA institutions.

Section 4 considers specific issues around the processing of personal data in terms of rules of criminal procedure, and how this can affect data processing and data lifecycles within technologies such as INSPECTr.

Section 5 identifies that exploitation of technologies like INSPECTr can pose significant risks where they are made available to ‘bad actors’, and provides a risk assessment methodology that can be used to determine whether exploitation of such technologies presents risks of misuse and mass surveillance, or to human rights, that are (un)acceptable.

Section 6 concludes the document.

2 Adapting Ethics and Privacy-by-Design for Law Enforcement Technologies

Ethics and Privacy-by-Design have been developed with a primary focus on commercial technologies.[1] So they are not conceptually focussed on the LEA context. In the commercial context, Privacy-by-Design is focussed on protecting the privacy of the end-user from an intrusive technology or excessive data-gathering by the technology provider, while Ethics-by-Design is focussed on minimising harm to the end-user and ensuring they are treated as ethically as possible. Often in the commercial context, the end-user is the procurer of the technology, and so Privacy and Ethics-by-Design become selling points for technologies. For example, Apple presents its products as being at the forefront of privacy protection.[2] However, in the LEA context, the end-user is likely to be an analyst or an investigator but the procurer might be an administrator who is advised by the analysts/investigators but also needs to include other factors into their procurement decisions, such as cost.

When applying Privacy and Ethics-by-Design in the LEA context, the focus shifts from the end-user to data-subjects whose data is captured as part of an investigation or intelligence gathering exercise. This does not mean that the LEA officers as end-users should not be considered. Indeed, where technologies have user accounts and functions for logging uses of tools, there is the potential for end-users to be monitored and surveilled whilst at work. In many situations, this would be unnecessary, intrusive, and illegitimate. However, not all LEA officers abide by the law,[3] so there is an important reason to retain these capabilities for professional standards or misconduct investigations. Further, logging of tool uses can facilitate transparency. This example highlights a major issue in applying Privacy and Ethics-by-Design strategies to LEA technologies: the nature of investigations are, in ethical and privacy terms, exceptional because most societies would not normally accept intrusive examinations of an individual's private life unless it is as part of a criminal investigation where there is reasonable suspicion of them being an offender.[4]

Recognising that the ability of LEAs to intrude on people's private lives is part of a modern democratic society does not mean that we can simply remove all privacy concerns and trust LEA officers; as noted above, many LEA officers have acted in unlawful ways. Rather, we can recognise that LEA investigations should focus on what people have done, not who they are. So, having investigations focused on whether a person of interest has committed a crime, and acknowledging that this will incidentally uncover private information about topics beyond their potential criminality, would likely be acceptable. But, initiating an investigation, or unnecessarily continuing an investigation, to uncover information about a person's private life that is not relevant to a criminal inquiry would not be acceptable. The challenge for the application of Privacy and Ethics-by-Design in the LEA context then, is to understand and allow for an appropriate level of investigation and intrusion into information about a person's private acts and whether they are criminal, whilst reducing the privacy and ethical impact to prevent intrusion or harm into unrelated personal spaces that is unnecessary and therefore

[1] See, for example, the standard published on Privacy by Design in 2023: International Standards Organisation 'ISO 31700-1 Consumer protection — Privacy by design for consumer goods and services — Part 1: High-level requirements', ISO, 2023. Available at: https://www.iso.org/standard/84977.html

[2] See, for example, Apple, 'Apple advances its privacy leadership with iOS 15, iPadOS 15, macOS Monterey, and watchOS 8', 7 June 2021, Apple. Available at: https://www.apple.com/newsroom/2021/06/apple-advances-its-privacy-leadership-with-ios-15-ipados-15-macos-monterey-and-watchos-8/

[3] Puddister Kate and Danielle McNabb, "When the Police Break the Law: The Investigation, Prosecution and Sentencing of Ontario Police Officers", *Canadian Journal of Law and Society / Revue Canadienne Droit et Société,* Vol.35, 2021, p,381.

[4] In legal terms LEA investigations are not exceptional because they are regulated by a legal framework that is intended to apply in everyday situations (which criminal investigations are seen in legal terms), and is not *Lex Specialis*. Contrast this with the application of international humanitarian law as an entire legal regime that only applies in armed conflict.

illegitimate. Indeed, it is important to recognise that being a subject of investigation can, in and of itself, be harmful.[5] Privacy harms can be generated by uncovering secrets, or the subsequent impacts of their revelation.[6] Where LEA officers do not act ethically or use technologies where ethical impacts have not been adequately considered, then individual and societal harms can be raised. For example, biased training data can lead to biased tools and discriminatory policing; this is especially the case where histories of racialised policing are used as a basis for technologies or policies used by police today.

In order to understand how Privacy and Ethics-by-Design could be applied in the LEA context, TRI has been working alongside technology and LEA partners in INSPECTr to discuss design choices and evaluate the implications of them from operational, privacy, ethical, and other perspectives with the intention of finding an appropriate balance of facilitating legitimate investigations whilst creating appropriate limitations. To demonstrate some results from this, each Privacy-by-Design strategy[7] and Ethics-by-Design principle[8] is discussed in-turn below. Tensions that are raised by applying the strategy/principle to the LEA context are explained and examined, with some potential design patterns that could mitigate the different competing perspectives in those tensions are also suggested.

2.1 Adapting Privacy-by-Design

Privacy-by-Design has 8 strategies that are separated into data-orientated and organizational-orientated strategies. The following section explores each strategy in terms of how it can be adapted from a commercial focus to the LEA context. The 8 strategies are:

Data-orientated strategies:

1. Minimise; limit the processing of personal data as much as possible.
2. Separate; separate the processing of personal data as much as possible.
3. Abstract; limit the detail with which personal data is processed as much as possible.
4. Hide; protect personal data or make it unlikable or unobservable. Make sure it does not become public or known.

[5] Feeley, Malcom M., *The process is the punishment: Handling cases in a lower criminal court,* Russell Sage Foundation, New York, 1992; Kolber, Adam J., ”Unintentional punishment”, *Legal Theory*, Vol.18, Issue.1, 2012. pp.1-29; Nathan, Christopher ”Principles of policing and principles of punishment”, *Legal Theory*, Vol.22, Issue 3-4, 2016. pp. 181-204; Hanna, Nathan, “Punitive intent”, *Philosophical Studies*, Vol.179, Issue 2, 2022. pp.655-669.

[6] Keats Citron, Danielle, and Daniel J. Solove “Privacy Harms”, Boston University Law Review, Vol. 102, 2022.

[7] See Hoepman, Jaap-Henk, “Privacy Design Strategies”, in Cuppens-Boulahia, N., Cuppens, F., Jajodia, S., Abou El Kalam, A., Sans, T. (eds) ICT Systems Security and Privacy Protection. SEC 2014. IFIP Advances in Information and Communication Technology, vol 428. Springer, Berlin, Heidelberg, pp.446-459; Danezis, George, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Métayer, Rodica Tirtea, Stefan Schiffner, Privacy and Data Protection by Design, ENISA, December 2014, pp.18-22.

[8] These come from the SHERPA project and EU’s High-Level Expert Group on Artificial Intelligence (which use the same requirements) and a study by The Ethics Centre. Philip Brey, Björn Lundgren, Kevin Macnish, and Mark Ryan, ‘D3.2 Guidelines for the development and use of SIS’, SHERPA project, 2019, p.1 (hereafter: ‘SHERPA Guidelines’). Available at: https://dmu.figshare.com/articles/D3_2_Guidelines_for_the_development_and_the_use_of_SIS/11316833; High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, p.14. Available at: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf; Beard, Matthew, and Simon Longstaff, *Ethical by Design*, The Ethics Centre, Sydney, Australia, 2018, p.59.

Organisational-orientated strategies:

5. Inform; inform data-subjects about the processing of their personal data in a timely and adequate manner.
6. Control; provide data-subjects adequate control over the processing of their personal data.
7. Enforce; commit to processing personal data in a privacy-friendly way, and adequately enforce this.
8. Demonstrate; demonstrate you are processing personal data in a privacy-friendly way.

2.1.1 Data-orientated strategies:

1. *Minimise; limit the processing of personal data as much as possible.*

Minimising the amount of personal data available to an investigation presents a tension between privacy and the need to fully investigate criminal activity; it would not be possible to say *a priori* what data could or could not be examined in an LEA investigation, as LEAs do not know what, or how much, relevant information they will find when they begin an investigation. Rather, applying this privacy design strategy should be thought of in terms of limiting the unnecessary processing of personal data as much as possible. This could involve reviewing all data that comes into an investigation file, and removing or sequestering data that seems unnecessary and only examining those data where the investigation suggests that it would be beneficial. For example, there has been controversy over LEAs subjecting people who make allegations of being sexually assaulted to a 'digital strip search' of their mobile phone in order to examine their relationship with an alleged offender.[9] Minimising the unnecessary use of personal data in such a situation could involve only examining the data if physical evidence is inconclusive, and even then only examining data from applications that are likely to provide relevant information. However, there is also the issue that investigations would not know which applications are relevant until they look. LEA officers in such a situation could be aided by data-driven technologies that highlight specific documents of likely relevance to an investigation; whilst this could involve extensive processing of potential irrelevant personal data, having personal data analysed by an AI tool can be seen as less intrusive than having an LEA officer examine all potentially relevant data manually.

2. *Separate; separate the processing of personal data as much as possible.*

Separating the processing of personal data creates a tension with the realities of policing often being under-funded. Ideally, from the privacy perspective, data processing activities could be separated between many investigators so as to prevent any individual investigator from unnecessarily piecing together a mosaic of disparate data and potentially gaining an intrusive insight into a person of interest (the 'mosaic effect').[10] Investigative teams are often small teams, and many have limited capacity. So separating data processing across a team might not be possible, or might have limited effect if each member of a small team is still processing a lot of data about each person of interest. A practical way of applying this strategy for LEA investigations would be that any data sharing beyond the investigative team is only done for specific tasks, for example having a specialist forensic examiner answer specific questions about particular information, rather than asking them to engage in a wide-ranging examination in the hope they might find something of investigative relevance. Whilst this might still

[9] Denham, Elizabeth, "Mobile phone data extraction by police forces in England and Wales: An update on our findings", ICO, June 2021. Available at: https://ico.org.uk/media/about-the-ico/documents/2620093/ico-investigation-mpe-england-wales-202106.pdf

[10] Gray, David, and Danielle Keats Citron, "A Shattered Looking Glass: The Pitfalls and Potential of the Mosaic Theory of Fourth Amendment Privacy" *North Carolina Journal of Law & Technology*, Vol.14, Issue 2. p.381

allow for an intrusive examination of private information across an investigation, it should limit the extent of intrusion by individual officers. This does raise questions around whether the data should be stored and categorised by LEAs on the basis of investigations or task/roles. For example, if all data from each investigation is stored together, then this would increase the likelihood of privacy intrusion via the mosaic effect. An alternative would be for an administrator to provide data access on the basis of tasks that are allocated to an individual officer, such as reviewing bank statements or emails. This would limit the risk of the mosaic effect being realised. However, it would also limit the likelihood of links being drawn across investigations. As has become clear from speaking to LEA partners in INSPECTr, linking investigations, or sharing information across them, can be highly beneficial for dealing with the complexity of contemporary organised crime. Similarly, as Arendt has highlighted, preventing people from seeing how their work fits into the bigger picture is not useful for encouraging those persons to recognise ethical concerns about their work or the organisation they are a part of.[11]

Other patterns such as enabling/disabling functions, and selective access controls could allow particular tools, or functionalities, to be enabled or disabled for specific officers for specific investigations, whilst access could be controlled to specific data on an as-needed basis. Having flexibility on how data can be accessed seems to be optimal in order to navigate between operational and privacy needs.

3. *Abstract; limit the detail with which personal data is processed as much as possible.*

In general, abstracting details in personal data is highly beneficial from a privacy perspective. Instead of referring to the home address of an individual, referring to their street or town reduces the exposure of their private details. However, details can be very important to an LEA investigation. For example, confirming the exact location of a person at the time an offence was committed could mean they continue as a suspect, or are ruled out and removed from the investigation. Abstracting data of people who are found to be innocent might seem like a workable option. However, investigators might also need to know details about innocent people in order to prove an alibi of a suspect, for example.

Still, it could be possible to use data minimisation techniques where a view across the investigation is not needed. Specialist examinations by forensic experts might not need to include specific details about persons of interest in an investigation to answer the specific questions posed by LEA officers. For example, references to particular locations recorded by wiretaps might be essential if a forensic expert is conducting a geolocation analysis, but might not be necessary to be included if a speech expert is trying to find matches in voice samples.

4. *Hide; protect personal data or make it unlikable or unobservable. Make sure it does not become public or known.*

Hiding personal data can be a privacy-protecting approach in many cases. However, much like with other data-orientated strategies, implementing it completely would impair a legitimate investigation as investigators would be unable to link information about offenders working together. There is a tension between enacting this principle and the need for investigators to legitimately be able to link data. Potentially, having data hidden by default that can easily be revealed where investigators have legitimate reasons to look at the data, could be implemented. However, linking criminals can be important in organised crime investigations. Where links are not already known, then hiding data by default could mean that such links are missed, resulting in cutting off the line of enquiry prematurely and unnecessarily. Rather than enforcing the hiding of data within an investigation by technical default, it might be better to focus on organisational policy so that investigators are trained to

[11] Arendt, Hannah, *Eichmann in Jerusalem: A Report on the Banality of Evil*, Penguin, Oxford, 2006.

recognise where their investigations are intrusive and to appraise whether continuing to explore the data would be appropriate.

The second part of the hide principle relates to making sure data is not made public or known in an irresponsible way. LEAs already have processes and security measures to prevent investigative data from becoming accessible. Afterall, making a suspect aware that they are being investigated would likely lead to the suspect stopping any unlawful behaviour to frustrate the investigation. Also, victims might not wish for information about what has happened to them to become known. However, data breaches from LEAs are rare, suggesting that such organisations often place cyber-security as a high priority and so there should be limited need to change these practices to comply with privacy standards.

2.1.2 Organisational-orientated strategies

5. *Inform; inform data-subjects about the processing of their personal data in a timely and adequate manner.*

Organisational-orientated strategies are good examples of privacy-preserving approaches that need to be changed for the LEA context. For example, as mentioned in relation to the *hide* principle, it would frustrate an ongoing investigation to make suspects aware that LEAs are investigating them. The Law Enforcement Directive (LED) does provide data-subjects with a right to access their personal data processed by LEAs in some circumstances. However, Art.15(1) provides for rights of access to be restricted, specifically (a) provides for this to avoid obstructing investigations. A general alternative could be for LEAs to publicly provide privacy policies that are as detailed as practicable to give people an understanding of how their data might be processed if they ever become a person of interest, bearing in mind the ability for LEAs to restrict this information where it could obstruct investigations or other legitimate LEA activities under Art.13, LED.

As noted earlier, the different approach for Privacy-by-Design in the LEA context means that rather than a commercial entity protecting the privacy of the end-users of their technologies, LEA officers will be protecting the privacy of the people included in their investigations and LEA organisations will be protecting the privacy of their staff. Therefore, we could see the *inform* principle as requiring investigators to be informed about the privacy impact on people whom they are investigating, and themselves, by making them aware of the potential privacy harms associated with the tools they are using. This could come via notices in the technologies and training materials. Warning 'labels' similar to those used on high-calorie foods could be used to indicate highly sensitive data, although the appropriate labelling might depend on labelling at the data input stage, the data structures, or whether the data could be pre-analysed by an algorithm to identify its sensitivity.

It is also important to note that the LEA officers themselves are likely to be data-subjects. For example, AI tools commonly have member accounts and logging functions for accountability purposes so the tools would be processing the personal data of the end-users themselves. In such circumstances, some privacy design patterns could be lifted directly from commercial approaches. For instance, notifying end-users about what data is going to be collected about them, and how it could be used and shared; generally, we would expect that data collected about LEA officers could be shared within, or across, LEAs where there are professional standards or safeguarding concerns about a particular staff member and their activities.

6. ***Control*; provide data-subjects adequate control over the processing of their personal data.**

Allowing people control over whom their personal data is exposed to is a key category of privacy (in comparison to confidentiality, and contextual privacy in practice).[12] Indeed, the EU's General Data Protection Regulation (GDPR) dedicates the entirety of Chapter III to enabling and facilitating the rights of data-subjects, and provides several rights to them. However, as noted above, the processing of persona data in LEA investigations is governed by the LED. Whilst the LED also dedicates its Chapter III to the rights of data-subjects, these rights are not as comprehensive as under the GDPR so as not to frustrate legitimate and lawful investigations unnecessarily. We can expect LEAs to provide for, and facilitate, the rights of data-subject rights as far as possible. However, we cannot expect data-subjects to have the same level of control where they are a person of interest to an ongoing investigation.

Yet, we can consider control from a different angle. In LEA investigations, those protecting the privacy of data-subjects are the senior officers running and managing investigations, and so we can place a responsibility on these persons to ensure that data access is appropriately controlled. This could be implemented in such a way that the relevant data is no more accessible, or subject to analysis, than is necessary for an adequate and effective investigation to take place.

In INSPECTr, the case management system uses a concept of tokens and wallets to control who can access certain datasets. For example, data from a child sexual abuse material (CSAM) investigation would not be accessible to investigators who only have a 'Fraud' token; rather a 'CSAM' token would only be assigned to persons who have received the necessary training for such investigations.

The Cortex capability provides for administrative controls that can restrict certain tools within organisations. For example, a particular team, or specific personnel, could be provided use of certain tools where they have carried out the necessary training. Or, could be restricted from using it where they have not received such training. Further, some LEAs might have policies against using certain tools, or might not be allowed to uses them due to particular national legislation. INSPECTr can facilitate access or restrictions to different tools depending on the needs of each LEA.

Further, a Traffic Light Protocol has been implemented to restrict the sharing and re-sharing of sensitive data. This assigns a rating of red, amber, green, or white (most sensitive to no sensitivity) to a particular dataset. For example, a disk image found in an organised crime investigation might be moderately sensitive (amber) and could be shared across jurisdictions under certain conditions, whereas a politically sensitive investigation might be highly sensitive investigation might class some data as red, preventing it from being shared at all.

Similarly, a maximum permissible actions protocol (MAX PAP) would prevent particular actions taking place automatically for especially sensitive documents. For example, some digital forensics investigators share information about malware they are examining through online services to get a better understanding of the software, where it comes from, and how to respond to it.[13] A high MAX PAP level would restrict any information being shared beyond the investigative team in case criminals deploying the malware monitor such services and would be made aware that their malware has been discovered and examined, for example.

The implementation of these processes allows investigations teams to control who can access specific data/tools and information. Whilst this might often be done from a security perspective (whether in terms of person having security clearance to view certain materials, or an information security perspective), it would not be much additional thought to incorporate the privacy perspective into the thinking of investigation managers for this and to add or augment existing controls.

[12] Gürses, Seda, "Can You Engineer Privacy? The Challenges and Potential Approaches to Applying Privacy Research in Engineering Practice", *Communications of the ACM 57,* No.8, August 2014, pp.20-23; Hopeman, Jaap-Henk, *Privacy is Hard and Seven Other Myths*, MIT Press, Cambridge, Massachusetts, 2021, p.15

[13] For example, services link VirusTotal allow users to upload suspicious files for analysis by tools from different organisations. VirusTotal, "How it works", 2023. Available at: https://support.virustotal.com/hc/en-us/articles/115002126889-How-it-works

7. *Enforce*; commit to processing personal data in a privacy-friendly way, and adequately enforce this.

A strong privacy-culture is imperative for successful LEA investigations: leaks beyond the investigation team risk exposing ongoing investigations to criminals, and wasting the efforts of the investigators. LEAs are also security-conscious organisations that commonly restrict access to certain information for operational reasons. Therefore, LEA officers are already thinking in privacy-respecting ways, but their focus is often on ensuring successful operations rather than the impact of those operations.

Enforcing a privacy-respecting culture in LEAs is not a significant jump from what they are already doing, it is an expansion of their consideration of privacy impacts. Therefore, in terms of enforcing privacy-awareness, LEAs should ensure their data protection offices are well-funded and respected so that not only can they perform their job to adequately protect the personal data of data-subjects, but also can foster privacy-respecting cultures in their organisations that are not just internally-focussed.

In order to adequately foster such a culture, data protection officers (DPOs) could expand their privacy policies and training provision to make investigators aware of the privacy impacts and limits of their work. Further, including investigators in data protection work can also be a useful exercise. Discussing privacy issues relevant to investigations with people who regularly deal with those issues can enable DPOs to understand better what is going on during investigations and also to convey what issues could arise and work out an optimal way of dealing with them.

It could be useful for DPOs to meet investigators and leverage the security-conscious nature of LEAs and begin privacy-related discussions on subjects such as avoiding data breaches and ensuring adequate information security practices are followed. Discussing the implementation of logging activity on LEA tools can also be a useful way to make users of such tools aware of the potential processing of their own personal data in addition to that of others. Further, collaboration on a data protection impact assessment for new AI tools is likely to require expertise from across an LEA organisation and is an opportunity for increasing privacy awareness amongst investigators. At every point available to a DPO, or other privacy professional working with LEAs, it can be useful to raise privacy-related discussions so as to foster a privacy-conscious culture.

8. *Demonstrate*; demonstrate you are processing personal data in a privacy-friendly way.

After a privacy-conscious culture has been built, it can be difficult to prove this externally to the public, especially where trust in the police is at a low ebb,[14] and LEAs themselves are often security-focussed. In order for LEAs to (re-)build trust from the public, from a privacy-perspective, greater transparency should be pursued. This does not mean that risks of security breaches as a result of revealing confidential information need to be taken or considered. Rather, it means that additional steps can be taken by LEAs to make the public more aware of how they investigate, how and when they process personal data, and what the privacy impacts might be. These do not need to be specific, but could be general so that the public has an idea of what investigators might do with their data. After all, the societal-level model of policing by consent requires the public to know and understand what they are consenting to before they can do so.

Additional steps that could be taken might involve LEAs publishing more detailed, but plain-language, privacy policies, or providing 'explainer' documents so that the public can understand LEA activities in more detail. Further, it is important that LEAs take accountability for privacy failures. This could involve publishing reports on privacy practices, or data breaches, by DPOs or external experts (these could be

[14] For example, in the UK, people who think the police do a 'good' or 'excellent' job has fallen to 55% in 2019/20. See Brown, Jennifer, "Policing in the UK", House of Commons Library, 2021, p.4. Available at: https://researchbriefings.files.parliament.uk/documents/CBP-8582/CBP-8582.pdf

redacted where necessary) so that the public can be aware of the procedures of their local LEA, rather than just the policy aims of the organisation.

In order that a privacy-culture is maintained, and is shown to be maintained, internally as well, it is important that DPOs are able to conduct a range of activities. Implementing regular trainings, especially with new investigators is imperative to ensure that they are aware of the implications of their work on privacy. Considering the technological advancements in projects like INSPECTr, these should also cover the use of AI tools. As with demonstrating a good privacy culture externally, plain-language policies for internal use could also be beneficial so as to show all staff the seriousness with which privacy is taken by the organisation.

Overall, it has been shown that Privacy-by-Design strategies can be adapted for the LEA context. This does not mean that privacy is considered in the same way as in the commercial context, or that the same level of privacy can be expected. Rather than data-subjects being able to control and restrict the exposure of their personal data, the application of these strategies to the LEA context is done through carefully thinking through how data can be processed in a way that enables legitimate and lawful investigations to continue whilst avoiding or minimising unnecessary processing. Sometimes this involves questioning whether certain processing is always necessary, or whether it needs to be done in a particular way. This might be disruptive for some investigators, but considering the current cultures of LEAs in terms of privacy and security, it should not be a significant shift in thinking.

2.2 Ethics-by-Design High-level requirements

Privacy-by-Design is a process of taking privacy design strategies and turning them into design patterns by thinking through their practical application. Ethics-by-Design is a younger process (as noted in D8.7), and so the process is primarily related to taking high-level ethics requirements and making them more specific to the LEA context and then developing ethics design patterns where possible. The below list of requirements merges those from SHERPA/the EU's AI High-Level Expert Group and work by Beard and Longstaff.[15] The high-level requirements are:

1. Human agency, liberty, and dignity/non-instrumentalism/self-determination
2. Technical robustness and safety
3. Privacy and data governance
4. Diversity, non-discrimination, and fairness
5. Individual, societal, and environmental wellbeing
6. Accountability/responsibility
7. Transparency
8. Ought before can
9. Net benefit
10. Accessibility
11. Purpose

[15] Philip Brey, Björn Lundgren, Kevin Macnish, and Mark Ryan, 'D3.2 Guidelines for the development and use of SIS', SHERPA project, 2019, p.1 (hereafter: 'SHERPA Guidelines'). Available at: https://dmu.figshare.com/articles/D3_2_Guidelines_for_the_development_and_the_use_of_SIS/11316833; High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, p.14. Available at: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf; Beard, Matthew, and Simon Longstaff, *Ethical by Design*, The Ethics Centre, Sydney, Australia, 2018.

1. *Human agency, liberty, and dignity/non-instrumentalism/self-determination*

Complying with this requirement in the LEA context could primarily be seen in terms of avoiding technical developments that would prevent investigators from exercising their own agency. In commercial technologies, there are concerns around 'dark patterns' that nudge people to select, or avoid changing settings of a technology that would mean the end-user expose more private information than they actually want to do.[16] Some of these 'dark patterns' are so common place as to become the default in some areas. For example, many social media platforms automatically place privacy controls in difficult to find areas of their settings, making it hard to have real control over one's privacy when using these services. In the LEA context, we could imagine developers who contrast a technology without thinking about the nudges that might affect end-users, simply because they are commonly done.

In INSPECTr, the platform has been developed with a keen awareness of the importance of the end-users in decision-making. For example, outputs of several tools provide a confidence level rather than a definitive result so that the end-user is fully aware of what the outputs of each tool mean and can determine their importance for their investigation themselves. A confidence level output of 51% might mean that an INSPECTr end-user recognises the need for further analysis 'by hand' to make a decision, whereas other systems that might present this output as the most likely and supposedly conclusive result would nudge the end-user to accept this output when that would not be responsible or appropriate.

2. *Technical robustness and safety*

Technical robustness and safety is normally considered in terms of the accuracy, precision, and reliability of AI technologies. These are, of course, important. However, in LEA contexts, it is important that the meaning of these terms are considered in the appropriate way. For example, a technology that is robust for commercial use might not meet forensic or investigative standards. For example, there are several facial analytics applications that can be used for 'face swapping' images that are intended to be shared on social media.[17] Anyone who has used these types of apps will be aware that they regularly fail (sometimes in amusing ways). Yet, a facial analytics tools for forensic or LEA purposes that regularly fails would not be appropriate to use and it could not be relied upon at the appropriate standard. Indeed, these technologies are being used in a process that can affect people's liberty, and so it is important that they are developed to an appropriate standard.

Although INSPECTr is a research project and the final platform is not intended to be sold as a product, substantial work has been done to ensure that the tools are usable. Indeed, the tools in INSPECTr have been developed with input from LEAs, especially GN, so that they can be used in the different court systems and instigative procedures used by the different LEA partners.

3. *Privacy and data governance*

Much of the possible discussion on privacy has taken place during the discussion on Privacy-by-Design. However, it is important that data governance is also considered specifically. In cybercrime or digital forensic investigations, sufficient data governance is critical to successful investigations and prosecutions; a poor approach to data management is likely to result in missing data, investigative leads being missed, and openings for defence lawyers to damage a prosecution case.

[16] Forbrukerrådet (Norwegian Consumer Council), Deceived by design, Forbrukerrådet, 2018. Available at: https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf

[17] See, for example, Microsoft Research, "Face Swap", Microsoft Store, 2023. Available at: https://www.microsoft.com/en-us/p/faceswap/9wzdncrdqkn6?activetab=pivot:overviewtab

In INSPECTr, data lifecycles have been considered in detail (see Section 4). Indeed, ensuring that data and tools are accessible to the right individuals has been taken very seriously. The Case Management System has been developed with a token-based approach so that investigation managers can assign tokens to across their investigators, in light of the levels of training and authorisation and the sensitivities of data so that only appropriate people have access to certain capabilities and technologies. There are different ways of approaching this, but it is crucial for the LEA context that data governance is well thought through.

4. *Diversity, non-discrimination, and fairness*

Technologies that have biased processing, or end-users who do not have a diverse understanding of the world can create discriminatory effects in investigations. It is well known that many technologies have demonstrated bias, which have on some occasions led to discriminatory impacts in the criminal justice system.[18] Considering the gendered[19] and racialised histories that many LEA policies and strategies have demonstrated,[20] it is imperative that steps are taken in the LEA context to recognise, and deal with these risks. It should not be acceptable in modern society for similar impacts to be created. Further, a discriminatory investigation is also a wrong investigation: it does not reveal the (whole) truth of the criminal circumstances, but only an incomplete or misdirected perception.

In INSPECTr, partners have engaged in sensitisation workshops to identify and deal with risks of discrimination, especially with respect to gender (see *D8.7: Privacy and Ethics-by-design in the INSPECTr platform, Section 4.2.1.3 Gender and AI*). Work in this direction has continued, with partners taking steps to minimise bias in training data, and consider whether models or tools could be altered or adapted to minimise biased outputs (see *D8.6: Ethical, Legal and Social requirements for the INSPECTr platform and tools – Final Report* for more details on each tool.)

5. *Individual, societal, and environmental wellbeing*

As noted above, the process of investigating an individual can create privacy harms, and the process of arrest and interview can be traumatic. So clearly there are potentially significant impacts for a person's wellbeing that can be caused by investigative decisions taken because of the outputs of investigative tools. Recognising that these are, at the moment, fundamental parts of LEA investigations, it would be difficult for a project like INSPECTr to be able to impact this. However, through making a platform and suite of tools that are as robust as possible, this should reduce the unnecessary harms that could result from the project technologies if they are made available for use.

Where issues with technologies are regularly repeated, this can mean that the same impacts can also be repeated. Where such systemic problems occur, this can create societal issues. A clear example is a tool that demonstrated bias toward or against a particular social group and has discriminatory effects which are repeated in many interactions with members of that group, leading to discrimination on the level of societal interaction between the institution of the police and particular groups.[21] As a result of the steps taken to avoid discrimination, it is expected these sorts of issues should be avoided with INSPECTr.

Contributing to environmental wellbeing is important in AI projects, where tools and platforms can use lots of energy through high volumes of data processing. It is difficult for a research project that is

[18] See, for example, State v. Loomis, 881 N.W.2d 749 (Wisconsin, 2016) (USA).

[19] See, for example, Corsianos, Marilyn, *Policing and Gendered Justice: Examining the Possibilities,* University of Toronto Press, Toronto, 2009.

[20] See, for example, Long, Lisa J., *Perpetual Suspects*, Palgrave, 2018.

[21] See, for example, Lum, Kristian, William Isaac, "To predict and serve?", *Significance*, Vol. 13, Issue 5, October 2016, p. 14.

not explicitly aimed at contributing to environmental wellbeing to do so. Further, there could be competing considerations regarding possible needs to process large volumes of data to conduct an effective investigation. As such, AI research in the LEA context can mostly aim to develop efficient and non-wasteful tools. INSPECTr has been developed with this in mind.

6. Accountability/responsibility

Accountability and responsibility are important for LEA technologies. Due to the nature of LEA investigations infringing on people's privacy, and potentially also people's liberty, it is important that uses of LEA tools can be reviewed and wrongdoers can be held accountable. Logging is a key feature that can enable this as having a clear record of how a tool has been used can enable wrongful uses to be identified and the people behind them held to account. As noted in D8.7, the INSPECTr platform logs uses on a private blockchain. However, as private blockchain can be completely controlled, this means that the blocks are not immutable, and can be subject to a 51% attack where an entity controlling more than 50% of computing power related to a blockchain can effect changes to blocks.[22] So that such activities can be identified and discovered, INSPECTr has also developed the functionality to export hashes of logs to the public Ethereum blockchain.

7. Transparency

Transparency in AI ethics can mean different things in different situations. For example, a computing expert who wants to understand the mathematical interactions between layers of a neural network would need different information to the end-user of a technology who want to know a particular recommendation was made.

In the LEA context, transparency and explainability of technologies are important as the criminal justice system is a high-risk domain where 'black box' systems are undesirable.[23] Focussing on the technology, it is important that they are understandable to end-users so that investigators can interpret results correctly and make the right decisions about how their investigations should proceed. In most INSPECTr tools, many outputs are presented in a prioritised list so that an investigator can clearly see the most likely results, as well as many others. This should enable ends users to see what factors have been identified by the technologies as being relevant and will be able to determine whether this is correct. Investigators will be able to think about whether they wish to incorporate this into their thinking about how to progress the investigation, and explain their reasoning for their decisions.

Thinking toward court trials, it is important that investigators and forensic experts can explain how a technology works for a judge and jury for them to understand what decisions they took and why in case this is a crucial part of a case. Typically, such persons have experience in presenting evidence they have analysed using technologies like those in INSPECTr in court, so this might not be a highly-prioritised area of research. However, as victims continue to have increased focus in the criminal justice system,[24] an expectation might also develop for investigative procedures to be explained to victims and their friends/families. This might require different language and focus from experts; it is

[22] Goyal, Swati, "51% Attack Explained: The Attack on A Blockchain", FX Empire, 2021. Available at: https://www.fxempire.com/education/article/51-attack-explained-the-attack-on-a-blockchain-513887

[23] Selbst, Andrew, Solon Barocas, AI Now 2017 Report, AI Now Institute, 2017. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf

[24] For instance, see the development of Ministry of Justice, Code of Practice for Victims of Crime in England and Wales (Victim's Code), UK Government, 2021. Available at: https://www.gov.uk/government/publications/the-code-of-practice-for-victims-of-crime/code-of-practice-for-victims-of-crime-in-england-and-wales-victims-code

worth considering if technology research can provide a contribution to this in terms of increased explainability of algorithmic processes.

8. Ought before can

The potential for harm to arise from technologies like those in INSPECTr has been mentioned above. However, it is also important to recognise that there is a clear utility for these technologies in criminal investigations. If there were no benefit to be gained from bringing a potentially harmful artefact into the world, then, from the perspective of harm prevention, it should not exist. As such, it is worth determining clear and legitimate purposes for potentially harmful technologies before they are developed.

INSPECTr is designed to assist LEA officers and investigators in their work to combat crime, primarily complex internet-enabled and organised crimes where it can be difficult to link sufficient evidence to offenders. The overall purpose of this activity is to enhance public safety and security, which is something generally desirable by society. This is shown by society allowing LEAs to conduct investigations that infringe on rights to privacy and liberty, amongst other rights, where they are lawfully, ethically, and appropriately conducted. As such, there is a clear and legitimate purpose for using technologies like those in INSPECTr, and the platform overall.

9. Net benefit

Linked with ensuring that there is a beneficial purpose for technologies, it is also important that the benefits of the technology outweigh any negative aspects. We can recognise the potential risks, and the tensions that exist between the benefits and risks. It is important to do this where, for example, LEAs, and potentially the state apparatus they represent, might favour development of a technology, or developing a technology in a particular way to enhance their use-cases but increase the negative aspects for others. Recent proposals to 'break' end-to-end encryption to better facilitate investigations into child sexual abuse material are a good example of this;[25] the presence of 'backdoors' in encrypted services might be useful for LEAs engaged in investigations, but are hugely detrimental to the privacy of ordinary users and also present significant risks of mass surveillance.

INSPECTr, however, does not attempt to develop anything in this area. Rather, questions around the possible tensions presented by INSPECTr relate primarily to privacy. By focussing on developing the platform and tools in ways that are designed for LEAs to use in lawful and legitimate investigations, and focussing exploitation efforts in Europe where there are generally high-standards of human rights compliance, the benefits of the INSPECTr platform should outweigh the potential negativities.

10. Accessibility

Accessibility is important for technology development as it can both avoid people being unnecessarily prevented from using a particular technology, and enable more people to access and use such systems. For example, many investigative roles in contemporary policing can be largely office-based, in comparison to a few decades ago where many investigations required officers to be 'in the field'. As such, as contemporary policing focusses more on technologically enabled investigations and criminality, this should allow a wider range of differently abled people to engage in roles that were previously unavailable to them. Further, technologies like those in INSPECTr could enable such persons to take significant roles in investigations as the importance of the type of investigations where INSPECTr could be used continues to grow. In order to increase the availability of such roles to people

[25] Bertzzi, Luca, "The EU's temptation to break end-to-end encryption", International Association of Privacy Professional, 2022. Available at: https://iapp.org/news/a/the-eus-temptation-to-break-end-to-end-encryption/

who are differently abled in terms of sight or hearing, for example, additional considerations could be to facilitate greater accessibility in this context through alternative interfaces or integration with other accessibility tools.

11. Purpose

From the perspective of cybernetics, the purpose of a system is what it actually does in the real world.[26] We cannot yet know what INSPECTr will actually do in real world use without actually seeing it being used over the long term. Indeed, there are many technologies developed for one purpose but are used for others. An obvious example is CCTV, which was originally developed for scientists to watch rocket test launches at safe distance,[27] and is now used as part of systems of authoritarian oppression.[28]

However, the INSPECTr platform has been developed with the intention to be used by LEAs in their investigations, has been tailored to these LEA needs through meeting with and discussing with stakeholders, and is accompanied by a comprehensive training programme that is intended to show LEAs how the platform should be used in the intended direction. As such, the INSPECTr partners have made significant efforts to ensure that the use of the INSPECTr platform, and so it's purpose from the cybernetics perspective, contribute to lawful and legitimate LEA investigations. The hope is that this will make a significant contribution to increasing public safety, public security, and a generally positive contribution to European criminal justice.

2.3 Affordances

The above discussion notes how technologies can be developed and used in more privacy-respecting and ethically-compliant ways. However, it is difficult to enforce these approaches as investigations can follow unpredictable trajectories and each investigation can vary considerably form others. Therefore, enforcing a set of rules as to how technologies can be used could prevent ethical uses of that technology in a way that was unforeseen by the technology developer at the time of release. An obvious example might be for technologies to only process personal data of persons who are determined to be suspects. However, as many investigative technologies help identifying suspects, applying such a rule would be detrimental to the usability of the technology. Even some controversial technologies, such as facial recognition might have use cases that would be ethically justifiable to most people: identifying a person who has planted explosives in busy public areas, for example.

The point here is not to say that these technologies are 'ethical' in and of themselves. Indeed, the potential for harm discussed above shows that this is not the case. Rather, it is to show that technologies can be used for more or less ethical uses-cases. A challenge for future research is to encourage more ethical uses, and discourage those that are less ethical.

The 'dark patterns' mentioned above are generally used to 'nudge' a person to share more personal data with commercial companies than they would otherwise be comfortable with. This is generally unethical where a person is encouraged to give up more privacy than is needed in order to receive goods and services. However, it could be permissible where there is a net benefit, and the impact is minimal. Affordances are design choices that influence what a user does with a particular artefact.[29] A lumbar-support chair that encourages the user to sit in a position that is beneficial to their posture,

[26] Beer, Stafford, "What is cybernetics?", Kybernetes, Vol. 31, No. 2, 2022, pp. 209-219.

[27] Monatrix, "The History of CCTV", 2021. Available at: https://www.monatrix.com/the-history-of-cctv/

[28] Gershgorn, Dave, "China's 'Sharp Eyes' Program Aims to Surveil 100% of Public Space", OneZero, 2021. Available at: https://onezero.medium.com/chinas-sharp-eyes-program-aims-to-surveil-100-of-public-space-ddc22d63e015

[29] Beard, Matthew, and Simon Longstaff, *Ethical by Design*, The Ethics Centre, Sydney, Australia, 2018, p.47.

even if this is sometimes uncomfortable to begin with, would generally be seen as ethically justifiable. There are 6 types of affordance identified by Davis and Chouinard that could be applied to technologies like INSPECTr:[30]

- **Requests** recommend actions to the user. Presenting analysis results in INSPECTr with a confidence level requests the end-user to think if the provided confidence level is sufficient to make a particular choice.
- **Demands** require particular uses of a technology. INSPECTr requires users to ingest data in a particular way for it to be analysed by the INSPECTr technologies.
- **Encouragement** promotes a particular action over others. A list of prioritised outputs from an INSPECTr tool encourages the end-users to think about the first few results.
- **Discouragement** occurs where there are extra efforts needed to take a particular action. Part of the benefit of the INSPECTr platform is its ability to link evidence across investigations. However, this is not done automatically and requires end-users to go through several steps to create links. To some degree, end-users are discouraged from attempting this where it is not needed.
- **Refusals** prevent some activities from happening. All users of INSPECTr will be prevented from accessing data or technologies that they do not have the correct authorisations for.
- **Allowances** are indifferent to particular actions. INSPECTr can provide analysis of many different datatypes.

Further research could be done on affordances as a way to extend by-Design approaches beyond recommendations about what a technology can do, to how it should do it in the LEA context. This could be particularly fruitful as LEA technologies contain inherent risks of harm. A clear criticism of such work could be that harmful behaviours would still be possible for these technologies. Yet, harmful activities continue to be possible today and application of affordances would present an opportunity to make them less likely, which is ultimately beneficial.

2.4 Conclusion

Overall, the work done in INSPECTr shows that Privacy-by-Design strategies can be adapted to limit unnecessary impacts raised by the privacy-invading nature of LEA investigations, and Ethics-by-Design can be made specific to enhance the ethical development and use of LEA technologies. Both approaches can have a protective effect by preventing and limiting unnecessary harms, but this is more than just technology design and includes thinking about processes and polices that are relevant to the use of the technologies in question. Affordances could be explored to further promote respect for privacy and ethics in the development and use of LEA technologies.

[30] Davis, Jenny L., and James B Chouinard, 'Theorizing Affordances: From Request to Refuse', Bulletin of *Science, Technology & Society*, Vol. 36, Issue 4, 2016, pp.241-248

3 What if we governed police use of data analytics and Artificial Intelligence like firearms?

3.1 Introduction

Law enforcement and police services around the world are increasingly using and seeking to make use of big data technologies, digital forensics and artificial intelligence (AI) tools and technologies. Some of these are relatively experimental, being developed in research projects or through internal trials, whilst others are being deployed in day-to-day policing.

This section will take seriously the potential danger posed by these tools, and consider their use and operation by drawing an analogy from another area where police services make use of tools that are dangerous, but are provided to police under particular circumstances: Firearms. Firearms are a lethal technology by design, yet society has determined that in some circumstances these lethal technologies should be usable by police: in this context, governance frameworks and practices for the control of these dangerous tools. In comparison, we are still at the start of developing such frameworks for the police use of data analytics and artificial intelligence (for example, with the EU publishing its plans for an artificial intelligence regulation[31]), and as such have much to learn from other examples.

By drawing this analogy, we by no means intend to imply or suggest that police use of firearms is perfect, or perfectly managed. We recognise this is a live, sensitive and very political issue. Black Lives Matter is the latest in a series of high-profile campaigns to draw attention to the potential lethality of encounters with police, to police related violence and its racialised dimension. Police killings of innocent civilians in the USA have even been described as a failure of governance, with state and federal governments failing to collect reliable data to investigate the causes of high death rates or develop administrative standards to reduce unnecessary killings.[32] We do not therefore consider that governance of the police use of lethal weapons is a perfect model, to be imported wholesale into AI governance, but rather that experiences here offer us a set of real-world existing models, which can be evaluated for what they can offer. We also acknowledge the role that history, gun ownership cultures, lobbying and marketing activities of firearms manufacturers and experiences of harmful events have played in the development of different governance regimes.[33]

Our hope is that by drawing upon a model of governance that is culturally familiar to police forces, we can show that regulation, monitoring and control of the use of potentially harmful technologies by law enforcement is not an alien concept for police forces, but rather an area in which they already have some experience and have developed practices. Often, the use of innovative, disruptive digital technologies is accompanied by hype and exaggerated claims of the novelty of the tools, suggesting their very nature somehow necessitates a complete break from prior ways of operating.[34] We believe that rather than assuming that all regulation of digital technologies should start *tabula rasa*, with a blank slate, we think there is insight to be gained from mining existing practices of governance. We are not advocating simply replicating firearms governance measures with any mention of a firearm

[31] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act), EU, 2021. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (Hereafter: AI Act proposal)

[32] Franklin E. Zimring, "Police Killings as a Problem of Governance", Annals American Association of Political & Social Science, 687, January 2020, pp.114-123, https://journals.sagepub.com/doi/pdf/10.1177/0002716219888627

[33] Peter Squires, *Gun Culture or Gun Control? Firearms and Violence: Safety and Society*, Routledge, London, 2000. https://doi.org/10.4324/9780203187753

[34] Gemma Milne, *Smoke & Mirrors: How hype obscures the future and how to see past it*, Robinson, London, 2020.

replaced with 'AI', but rather using these existing governance structures to consider potential methods for the governance of law enforcement use of AI and data analytics.

This reflection could serve to prepare the ground for codes of conduct for LEA use of data analytics and artificial intelligence tools, and contribute to the emerging literature on AI governance. At this stage, we justify why we consider such tools potentially dangerous at all, and that the level of danger and risks is sufficiently similar to firearms that the applicable governance structure can be used as a resource for governing AI use.

3.2 How is data analytics and AI dangerous at all

Cathy O'Neil coined the term "Weapons of Math Destruction" to describe the math-powered applications powering the data economy, that however encode human prejudice, misunderstandings, and bias into software systems. The models in these systems are often opaque, meaning that their outputs and verdicts, even when wrong or harmful get placed beyond dispute or appeal.[35] For O'Neil, the harm from such systems comes from their impact on the individuals subject to algorithmic decision making, but also from corrupting impacts upon social and political processes, and the way that they tend to accumulate power to the already powerful, and have greatest impacts upon the already vulnerable.

The use of AI technologies by LEA regularly involves the processing of personal data about suspects, victims, witnesses, convicts, and other person of interest. This raises the possibility of privacy and data protection risks. Keats, Citron and Solove have explored the notion of 'Privacy Harms' in detail. Generally, this is where the release of private information results in some sort of injury to a person.[36] A common aspect of criminal investigations conducted by law enforcement is the uncovering of previously private or hidden information about a crime. Such details can be discovered by police officers during investigations and are often made public during a criminal trial, even if the suspect is found to be innocent. As such, revelation of private information, even if it is legitimate, could cause harm to a person's economic situation, reputation, psychological wellbeing, their relationships, or other aspects of their life that might not be warranted, especially where they are innocent. Yet, this can be an unavoidable part of an LEA investigations. So, we can see that investigations can be fundamentally harmful to privacy.

Further, the UK's Information Commissioner's Office has provided a taxonomy of data protection harms which could occur through the misuse of personal data. These include harms to individuals and society, and include examples relevant to the criminal justice system such as suicide or self-harm, location tracking leading to physical assault, impacts of biased decision-making, unwarranted surveillance, injury to peace of mind, detriment to mental health, chilling effects of victims of/witnesses to crimes, and mistrust in public offices.[37] The potential for serious physical and mental harm to be created through data protection harms is important, as it draws a clear parallel to the level of harm that can results from firearms. AI systems can massively increase the speed and scale of processing personal data, thereby potentially increasing the impact and scope if any of these risks actually manifest.

Proposed legislation recognises the potential for harm arising from policing use of artificial intelligence, particularly in terms of the potential impact upon fundamental rights. The European Union's proposals for the Artificial Intelligence Act categorise AI systems by the potential risk they pose to health and safety or fundamental rights of persons, introducing requirements that apply to

[35] Cathy O'Neil, *Weapons of Math Destruction: How Big Data increases inequality and threatens democracy*.

[36] Keats Citron, Danielle, and Daniel J. Solove "Privacy Harms", Boston University Law Review, Vol. 102, 2022.

[37] Information Comissioner's Office, "Overview of Data Protection Harms and the ICO's Taxonomy', 2022. Available at: https://ico.org.uk/media/about-the-ico/documents/4020144/overview-of-data-protection-harms-and-the-ico-taxonomy-v1-202204.pdf

high-risk AI systems deployed in the EU. Law enforcement use of AI for purposes such as making risk assessments for potential offenders or victims; polygraphs or emotion detection; detection of deep fakes; the evaluation of the reliability of evidence; prediction of crime based on profiling or assessment of personality traits or past criminal behaviour; and searching through large and complex data sets for unknown patterns and hidden relationships are all to be considered high risk. The use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes is perceived by the legislators as *so* risky, that it is placed in the prohibited AI systems category, apart from in a set of qualified uses, where its use is strictly necessary.[38] This covers many of the policing uses of AI and data analytics that we are concerned about in this report and given the nature of these tools, largely frames these impacts in terms of fundamental rights, rather than health and safety harms.

The combination of predictive or recognition technologies with firearms should also be seen as particularly dangerous. In firearms contexts, mistaken identity can be lethal. Knowing that AI systems can sometimes fail so often as to undermine their very purpose,[39] the very realistic prospect of LEAs potentially combining fallible AI systems, end-users taking operational decisions without critically interrogating the results of AI systems, and the potential deployment of armed officers raises significant risks for people targeted by policing operations. This is more worrying considering recent examples of large numbers of innocent people killed as a result of poor decision-making by police (US officers '*wantonly and blindly fir*[ing] *10 rounds*' during a raid on the house of Breanna Taylor being a prime example).[40]

Of course, the use of firearms by police does have utility and has been accepted by society for some time. Indeed, from the perspective of human rights law, states are under a '*positive obligation … to take preventive operational measures to protect an individual whose life is at risk from the criminal acts of another individual*'.[41] This might involve use of firearms by LEAs against a person posing life-threatening danger to protect the lives of others. Yet, human rights law also recognises the potential harm that can arise from such objects and the need to place stringent controls in place:

> '*Given the particularly high level of risk to life involved in any misuse of firearms, … it is essential … to put in place and rigorously apply a system of adequate and effective safeguards designed to counteract and prevent any improper and dangerous use of such weapons*.'[42]

We now explore the governance structures that have been applied to firearms, and what, if anything, can be drawn for them and applied to the use of potentially harmful AI systems by LEAs.

3.3 Firearms governance structures and their relevance for AI and data analytics

The primary aim of governance structures for firearms are typically to reduce deaths and injuries, but also '*the woeful impact that police woundings and killings have on citizen' perception of the fairness and decency of police agencies*'.[43] There is a strong link between governance structures around

[38] AI Act proposal

[39] Inioluwa Deborah Raji, et al., The Fallacy of AI Functionality, *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA.

[40] BBC, "Breonna Taylor: What happened on the night of her death?", 8 October 2020. Available at: https://www.bbc.co.uk/news/world-us-canada-54210448

[41] Case of Osman v The United Kingdom, (App no. 23452/94) 28 October 1998, para.115.

[42] Case of Kotilainen and Others v. Finland, (App no. 62439/12) 17 December 2020, para.88

[43] Patrick V. Murphy, Foreword in James J. Fyte (Ed.) *Readings on Police use of Deadly Force,* Police Federation, 1982.

firearms and the principle of law enforcement by consent rather than by force in the paramilitary sense.

In this section we proceed through several governance mechanisms and assess what lessons they can teach us about the governance of the police use of AI. We discuss each governance mechanism at a high-level in terms of its applicability to AI, and then provide more specific requirements in relation to adapting the United Nations *Basic Principles on the Use of Force and Firearms by Law Enforcement Officials*[44] to AI.

3.3.1 Treaties and case law – legal and administrative frameworks

Human rights treaties provide an overarching approach to regulating firearms in terms of how they are used by state agents, and how firearms are treated in legislation. All human rights treaties provide protection for the Right to Life in one form or another.[45] The European Court of Human Rights has developed a significant body of case law on the right to life, particularly in relation to use of force by state agents. The content of European case law is not the sole topic of this work, and so a detailed examination will be left for later works. However, some broad requirements can be described.

The basic requirement of states in protecting the right to life demands '*an appropriate legal and administrative framework defining the limited circumstances in which law enforcement officials may use force and firearms, in the light of the relevant international standards*'.[46] In this line of thinking, the Court has suggested that states should consider the United Nations *Basic Principles on the Use of Force and Firearms by Law Enforcement Officials* (discussed in more detail below).[47] The Court also requires national legislation governing LEA activities to '*secure a system of adequate and effective safeguards against arbitrariness and abuse of force and even against avoidable accident*'.[48] However, outdated legislation on firearms was deemed unacceptable.[49]

What lessons can this give us for police AI and data analytics?

In terms of AI, we can suggest that an appropriate legal and administrative framework should be developed for dangerous LEA AI systems. The development of the AI Act and national AI policies show that this is already in progress, but there is much more to be done, especially in terms of effective safeguards against arbitrariness, abuses, and accidents. Due to the pace of progress in AI development, it would be worth national legislatures creating regularly convened, or permanent, bodies to monitor development and recommend advances for regulation.

[44] Basic Principles on the Use of Force and Firearms by Law Enforcement Officials (adopted 07 September 1990), Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Havana, Cuba, 27 August to 7 September 1990, UN Doc. A/CONF.144/28/Rev.1, p. 112.

[45] See, for example, Art.6, International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171

[46] Giuliani and Gaggio v. Italy, (App no. 23458/02) 24 March 2011, para.209; Makaratzis v Greece, (App no. 50385/99) 20 December 2004, para.57–59; Bakan v. Turkey, (App no. 50939/99) 12 June 2007, para.49.

[47] Atiman v. Turkey, (App no. 62279/09) September 2014, paras. 23, 30; Błońska v. Poland (dec.), (App no. 26330/ 12, 1 April 2014, para.55; Benzer and Others v. Turkey, (App no. 23502/06) 12 November 2013, para.90; Gorovenky and Bugara v. Ukraine, (app nos 36146/05 and 42418/05) 12 January 2012, para. 22; Makaratzis v Greece, (App no. 50385/99) 20 December 2004, para.30; Hugh Jordan v. the United Kingdom, (App no. 24746/94) 4 May 2001, para.88; McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.139; Huohvanainen v. Finland, (App no. 57389/00) 13 March 2007, para.75.

[48] Giuliani and Gaggio v. Italy, (App no. 23458/02) 24 March 2011, para.209; Makaratzis v Greece, (App no. 50385/99) 20 December 2004, para.59.

[49] Erdoğan and Others v. Turkey, (App no. 19807/92) 25 April 2006, paras. 77–78.

3.3.2 Treaties and case law – operations

Operationally, use of firearms by LEAs requires '*a careful assessment of the situation'* before they are employed.[50] Exhaustive policies on situations where firearms could be used as a last resort, and procedures for moving from warning shots (which should be employed wherever possible)[51] to aiming at persons can also be implemented.[52] Indeed, through several key cases,[53] the European Court has been clear that uses of force must be absolutely necessary.[54]

What lessons can this give us for police AI and data analytics?

First, LEA officers should carefully consider whether there is a need to employ AI systems in LEA operations; what issues should be considered would be a worthy area of future research. It is unlikely that an AI equivalent of 'warning shots' would be practical in the sense of informing a suspect of an ongoing investigation they are unaware of, as this would alert them to the investigation, leading to them to change their behaviour and frustrate the investigation. Rather, they could be implemented in the sense of LEA officers announcing and recording their intentions to use AI systems during an investigation. Whether it would be practical to exhaustively list all situations where AI could be implemented, and how, could be impractical due to the numerous variations in data and investigative circumstances.

Some AI systems are less dangerous than others. A system that identifies the languages spoken in an intercepted phone call is likely to have less potential for harm than a system that attempts to determine a person's emotional state, for example.

Thinking about the absolute necessity requirement for use of force, we can recognise that not all AI systems are as dangerous as firearms. Indeed, in the case of *McCann* regarding the right to life of terrorist suspects killed by UK soldiers, the European court stated:

> *'use of the term "absolutely necessary" indicates a stricter and more compelling test of necessity must be employed from that normally applicable when determining whether State action is "necessary in a democratic society" …In particular, the force used must be strictly proportionate to the achievement of [a legitimate aim].'*[55]

As such, where AI systems are especially dangerous, the necessity of using it must be very strongly justified. Where AI systems are not assessed to be as dangerous as firearms, a lower level of necessity requirement could be applied, along with less stringent needs for justifications. In any case, they should not be used without good reason.

This could be considered in terms of a requirement to only use AI for the benefit of protecting potential victims of crime. Of course, a criticism here is where a victim is deceased, they no longer need protecting. However, we can consider the friends of a direct victim as victims as well. Even if a victim was alone in the world, society in general could be seen as a victim in need of protecting from uncaptured offenders.

[50] Giuliani and Gaggio v. Italy, (App no. 23458/02) 24 March 2011, para.209.

[51] Kallis and Androulla Panayi v. Turkey, (App no. 45388/99) 27 October 2009, para.62.

[52] Bakan v. Turkey, (App no. 50939/99) 12 June 2007, para. 51.

[53] McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995; Makaratzis v Greece, (App no. 50385/99) 20 December 2004; Nachova and others v Bulgaria (Apps no. 43577/98 and 43579/98) 6 July 2005.

[54] McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.148

[55] McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.149

3.3.3 Legislation

Several countries make the rules around police use of lethal weapons explicitly part of national law, thereby subjecting it to specific regulation. The reasons for this can be manyfold, but harm reduction is likely to be a major motivating factor. Due to considerations of length, we have not engaged in a detailed examination of each legal framework of the 27 EU countries. However, it is important to note that whilst there are commonalities between the regulation of police use of firearms across different countries, there are also clear differences based on what is acceptable in different societies.

What lessons can this give us for police AI and data analytics?

The firearms example suggests some benefit in giving the use of controversial and potentially harmful tools a firm legal basis which also incorporates appropriate constraints and limitations. There are already attempts to put police use of AI and data analytics on a legislative footing. The EU has the LED regulating use of personal data by law enforcement, and the proposed AI regulation bringing many advanced policing activities within the scope of high-risk AI. As a directive, the LED is not applied uniformly across the EU. Rather, it is for Member States to incorporate into their national law, whilst adapting it to their particular situation. It would not be unimaginable that different countries would accept different standards for the use of AI by their LEAs. For example, France does not allow processing of personal data relating to some personal characteristics in most situations due to the processing of such data by the Vichy government during collaborations with the Nazi regime as part of the Holocaust.[56] As such, one could expect different types and uses of AI systems to have a different reception in different societies, and so should be regulated differently in order to take account of the views of those societies.

3.3.4 Codes of conduct

In enacting the Basic Principles relevant to LEA policy, the UK College of Policing code of practice on armed police and police use of less lethal weapons[57] sets out responsibilities of the chief officer, basic principles related to the selection, evaluation, approval, authorisation, acquisition and use of firearms specialist munitions and less lethal weapons by police; the manner in which those principles are implemented within the police service; provide a statement on standards of competence, accreditation and operational practice; aims to ensure that the these principles lead to a systematic programme of continuous development of policy, practice and capability; promote compatibility of operating procedures and cooperation between officers across forces; and foster the identification of good practice.

What lessons can this give us for police AI and data analytics?

Development and application of a similar code would be useful for AI, especially to enact the adapted principles into LEA policy. It would seem to be highly beneficial for LEAs to: clarify the responsibilities relating to AI systems of their officers; determine how AI systems should be procured, used and overseen; elaborate what standards should be applied to users of LEA AI systems; continually develop

[56] Bleich, Erik, "Race Policy in France", *Brookings*, 1 May 2001. Available at: https://www.brookings.edu/articles/race-policy-in-france/

[57] College of Policing, "Code of Practice on Armed Policing and Police use of Less Lethal Weapons, presented to Parliament pursuant to Section 39A(5) of the Police Act 1996, as amended by Section 124 of the Anti-social Behaviour, Crime and Policing Act 2014", January 2020. Available at: https://library.college.police.uk/docs/appref/CCS207-CCS0120853800-001-Code-of-Practice-on-Armed-Policing.pdf

AI policies; promote common standards between LEAs using AI systems; identify good practices. All of these activities would seem to complement the more detailed requirements outlined below (see *Section 3.3.13 Basic Principles*) by providing a clear oversight and policy structure for the use of potentially harmful AI systems by LEAs.

3.3.5 Training

Typically, before an officer is allowed to carry a firearm they have to undergo some training; this must include training on assessments about whether use of firearms is an absolute necessity[58] and must have the objective of complying with human rights standards for use of firearms.[59] The training must be specific to the situations that are to be expected for LEA officers; in *McCann*, the Court held that the militaristic training of the soldiers involved indicated that the operation lacked the '*the degree of caution in the use of firearms to be expected from law enforcement personnel in a democratic society*'.[60] Training requirements can have a time or complexity dimension: perhaps officers need to refresh their training on a regular basis to remain certified. Perhaps they need special further training to carry certain categories of weapon. As part of this training and selection process they may undergo screening for appropriate moral or psychological traits. Training should include ethics and alternatives to the use of force.

Training is also important for senior officers so that they are aware of the types of operations and activities armed officers can and cannot perform safely and understand the tactical options available to them.[61]

What lessons can this give us for police AI and data analytics?

Where training on the ethical use of specific law enforcement AI tools is part of research projects, this training could be delivered alongside the practical training on the new tool created by technical partners in those projects (See D6.3 Capacity Building Programme – Final Materials). This could be built upon by LEAs to develop sufficient human rights-based training programmes for the use of AI systems that are specific to the circumstances that their officers find themselves in. As AI becomes a more significant part of LEA operations, general training on data analytics and AI could also be offered as part of basic training for new officers as they are introduced to other technical systems used by LEAs. However, training programmes are often victims of budget cuts, and recipients also forget content as time progresses. It would be important that these training programmes are not only maintained, but are also regularly updated by trainers and revisited for updates by operational officers, especially where new systems are procured.

3.3.6 Deployment

Deployment of new weapons for LEAs can take place in stages. For example, in the United Kingdom, the rollout of X-26 tasers provides an example of this cautious approach to a new tool. Initially trialled in five regional forces, Chief Officers were allowed to make tasers available to trained firearms officers across all forces in England and Wales in 2004. In July 2007, the use of tasers by non-firearms trained

[58] Nachova and others v Bulgaria (Apps no. 43577/98 and 43579/98) 6 July 2005, 97; Sašo Gorgiev v. the former Yugoslav Republic of Macedonia, (App no. 49382/06)),19 July 2012, para.51.

[59] Güleç v. Turkey (App no. 21593/93) 27 July 1998, para.71; Şimşek and Others v. Turkey, (Apps no. 35072/97, 37194/97) 26 October 2005, paras.109, 117

[60] McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.212.

[61] Peter A J Waddington, "Arming an Unarmed Police Policy and Practice in the Metropolitan Police", Police Foundation, London, 1988.

officers was trialled. When this was seen as a success and use of tasers was subsequently extended to officers who were specially trained in their use (although they need not be firearms trained).[62]

What lessons can this give us for police AI and data analytics?

Trials would need to have specific qualities in order to properly inform chief officers of the risks and rewards offered by AI systems for investigations. Before engaging in such trials, there would need to be a clear conception of what counts as 'success' in this context: enhancing compliance with applicable legal and ethical standards and reducing risks of infringements would seem to be a good place to start. These successes would need to be subject to honest evaluations by LEA officers. Further, as AI systems can have a significant impact on communities, it would be worth including different stakeholder groups, especially people from the society that would be policed by the AI systems to engage in the process of honest evaluation.

3.3.7 Question of routine use

Several countries, including the United Kingdom, Ireland, Iceland, Norway, and New Zealand do not routinely arm police officers on duty with firearms.[63] Reasons given for this include the risks of escalatory arms races with criminals, risks to officers, and links to particular policing traditions and cultural values.[64] Essentially these societies have concluded that these tools, despite their availability are not appropriate for routine, day-to-day use. For these countries, these decisions have been quite stable, despite occasional tragic gun violence events. Indeed, many LEA officers themselves in the UK are against the routine arming of police officers: *'The last survey of individual officers by the Police Federation of England and Wales, the representative body of all police officers up to and including the rank of Chief Inspector, showed only 22% of rank and file officers were in favour of routine arming.'*[65]

What lessons can this give us for police AI and data analytics?

This shows us two things. Firstly, that just because a tool exists does not mean that police services need to adopt it, particularly if its use clashes with other priorities and strategic considerations. Second, that if a police service does adopt a risky tool, it does not necessarily need to issue that tool to all officers for use in all circumstances, and rather that the use of such tools can be restricted to specialists and specific contexts of use. In the context of AI, there might be potentially harmful AI systems that could only be accessible to specialists who can fully understand their risk and how to use them appropriately, and that they may be only deployed when their use is particularly necessary and justified.

[62] Turner, Ian David, "Arming the Police in Britain: A Human Rights Analysis", *the Police Journal: Theory, Practice & Principles*, 90(2), 107-127. p

[63] Noack, Rick, "5 countries where most police officers do not carry firearms — and it works well", W*ashington Post*, 8 July 216. Available at: https://www.washingtonpost.com/news/worldviews/wp/2015/02/18/5-countries-where-police-officers-do-not-carry-firearms-and-it-works-well/

[64] Ibid.

[65] Turner, Ian David, "Arming the Police in Britain: A Human Rights Analysis", *the Police Journal: Theory, Practice & Principles*, 90(2), 107-127, citing Johnston, Philip, "Do we really want to arm our police?" *The Telegraph*, 19 September 2012. Available at: www.telegraph.co.uk/news/uknews/law-and-order/9553111/Do-we-really-want-to-arm-our-police.html

3.3.8 Routine suspensions after critical incidents

In some contexts, the act of discharging a firearm will result in the officer being put on a temporary suspension or administrative leave, even when they have every legal power to use that weapon. The idea behind this is that the use is, in some sense, a failure of alternate methods of policing, public safety, planning or de-escalation of a situation. The pause is to allow for that potential failure to be investigated and better understood, but also to assess and respond to post-shooting stress and emotional impact on the involved officer.[66]

What lessons can this give us for police AI and data analytics?

We don’t think that deployment of a data analytics tools will often have an impact upon an officer akin to shooting; an exception might be where data analytics processes might confront an officer with images of abuse. This approach is deployed because of the potential impact from critical incidents, that can have strong emotional impacts. This is potentially one area where the analogy between AI and firearms is weakest. AI and data analytics tools are less likely to be tools for critical incidents are more likely to be part of the background, day-to-day knowledge work of policing. Their technological affordances appear to encourage their use for activities around improving situational awareness, mapping trends, quickly sifting through data or assessing risks. The tempo of use is likely different, as the justification for their use or deployment will rarely be an immediate threat to life.

However, if AI were to ever be deployed by police forces in a manner where it essentially took or supported critical incident decisions then a suspension-and-stock taking process would certainly be appropriate. Examples would be if an automated system were used for synchronising the release of force from human-controlled weapons[67] or more likely an AI identification system identifying a suspect which then leads to a critical incident. In those contexts, it may be appropriate to consider suspension and evaluation of the system in question.

3.3.9 Investigations into mistakes and misuse

Where investigations into the LEA use of firearms take place, whether after a perfectly lawful weapons discharge or otherwise, ideally the findings from these investigations should be fed back into training and used to inform future practices. This would be especially useful where officers are faced with novel situation not normally included in training programmes or where offenders use a new approach to their offending.

A failure to carry out an effective investigation into deaths caused by LEA use of firearms is likely to cause a violation of procedural requirements of substantive obligations of states in relation to the

[66] Charoen, Patrick P., "Officer involved shooting: The emotional impact and the effective coping strategies" (1999). UNLV Theses, Dissertations, Professional Papers, and Capstones. 288. http://dx.doi.org/10.34917/1481139

[67] It is unlikely that a weapon system with autonomy could release force autonomously in compliance with human rights legislation during a law enforcement situation, but Heyns suggests that an AI system could lawfully be used to synchronise the release of force when snipers have a clear shot during a hostage situation. See Christof Heyns, “Human Rights and the use of Autonomous Weapon Systems (AWS) During Domestic Law Enforcement”, *Human Rights Quarterly*, Vol.38, No.2, pp.350-378, 375-376.

right to life of the victim(s).[68] Indeed, the European Court has laid out the following requirements for an effective investigation:

> *'…those responsible for carrying out the investigation must be independent from those implicated in the events; the investigation must be "adequate"; its conclusions must be based on thorough, objective and impartial analysis of all relevant elements; it must be sufficiently accessible to the victim's family and open to public scrutiny; and it must be carried out promptly and with reasonable expedition*.'[69]

Investigations into LEA firearms incidents are often reviewed by a professional standards unit, or a wholly separate organisation to provide independence to the nature of the investigation. For example, in the UK, all police incidents where death or serious injury occur are automatically referred to the Independent Office for Police Conduct.[70] Further, officers involved in firearms incidents are separated after the event so as to maintain the independence of their recollection of the event and to prevent conferring/colluding between officers involved. [71] It is important that these investigations are rigorous to avoid any doubts about their quality or independence.[72]

Where wrongdoing is found in LEA use of firearms, officers can be subjected to disciplinary procedures, or even criminal punishment. Investigations into these matters might take place in parallel to an investigation of events. Further, investigations might discover that failures are at an institutional level, rather than individual.[73]

What lessons can this give us for police AI and data analytics?

The requirement for investigation could be applied to AI systems in terms of regular audits or reviews of uses. It could also form a technical requirement for the systems to be auditable and reviewable by officers from professional standards units, for example. This could pose a challenge for highly-complex AI systems as they would need to be understandable to an ordinary professional standards officer, who might have little expertise with AI. Of course, it could be useful to have such systems understandable by ordinary investigators too. LEAs procuring AI tools should have high requirements around transparency and explainability.

Continual learning for LEA users of AI systems would generally seem to be beneficial. It is unlikely that systems developed and deployed will remain as useful and as applicable many years into the future. Indeed, we expect technologies to be continually updated, and so feedback from end-users could be provided to technology developers for future updates. Updates to trainings should also be made, whether in conjunction with the technology developers (if issues could impact operations in multiple countries, for example), or for the internal training programmes at each LEA.

Where an LEA AI system produces real harm, it would be beneficial for an independent organisation to engage in an investigation of events so as to provide an account of why this happened, and make recommendations so that similar harms do not occur in future. During such investigations, it would

[68] Armani Da Silva v. The United Kingdom, (App no. 5878/08) 30 March 2016, para.229-230; McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.161.

[69] Armani Da Silva v. The United Kingdom, (App no. 5878/08) 30 March 2016, para.240.

[70] IOPC, Assessing referrals, 2023. Available at: https://www.policeconduct.gov.uk/complaints-reviews-and-appeals/statutory-guidance/assessing-referrals

[71] Turner, Ian David, "Arming the Police in Britain: A Human Rights Analysis", *the Police Journal: Theory, Practice & Principles*, 90(2), 107-127. http://clok.uclan.ac.uk/15527/1/Ian_Turner_Armed_Police_Article%20pdf.pdf, p.7

[72] Decker, Rick, "You've been in an officer-involved shooting...", *Lexipol*, 19 May 2022. Available at: https://www.police1.com/patrol-issues/articles/youve-been-in-an-officer-involved-shooting-ZAN5MG0VXak6qog0/

[73] Armani Da Silva v. The United Kingdom, (App no. 5878/08) 30 March 2016, para.284.

also be worth separating the LEA officers involved to prevent discussions that could affect the outcome of the investigation.

As the police are public servants, even the potential perception that such an investigation gives a charitable view of the LEA officers, or an LEA as an institution, can lead to viewing the process, and therefore the LEA, as potentially corrupted. Indeed, the implementation of impartial justice being done, and seen to be done, is an important part of society being able to trust in the police, especially where there could be disciplinary or criminal punishments given to LEA officers who (mis)used an AI system.

3.3.10 Uniformity in administration, testing, and access procedures

In the UK, police forces must be able to show an audit trail for the procurement of any firearms, less lethal weapons, and specialist munitions they purchase. Selection and acquisition of firearms is the responsibility of chief officers, based on operational requirements arising from threat and risk assessment processes.[74] This assessment is supposed to take into account the training implications – for example, will there be sufficient facilities to adequately train officers in the use of the new equipment. Firearms equipped officers are typically not allowed to privately procure or alter their own equipment. For example, under the Caribbean Human Rights and Use of Force Model Policy, law enforcement officers are prohibited from modifying any instrument of force, and where such instruments are assigned to an officer on a longer-term basis, they must be checked annually for modifications and be verified as still fit for purpose.[75]

Whilst this is likely a result of procurement practices, economies of scale and the desire for interoperability, it also works to prevent the use of unsafe or unreliable weapons or weapons that are disproportionately powerful for the envisaged contexts of use. Indeed, poor overall planning of firearms operations has been severely criticised by the European Court,[76] showing that the circumstances surrounding use of force are intrinsically linked and are important to be considered.

When LEA officers acquire an LEA firearm, they are typically required to sign for it to show they have taken responsibility for the firearm whilst it is in their possession. Depending on LEA processes, there can be separate administrative and storage processes for training firearms, operational firearms, ammunition, and other equipment.

These processes prevent unauthorised access to dangerous firearms, by members of the public, criminals, or irresponsible officers and facilitates accounting practices. It may also play a role in preventing unwarranted escalation of a situation – for example, where larger and more dangerous weapons are secured in a locker in a police car this can reduce the likelihood of them being used in an incident.

What lessons can this give us for police AI and data analytics?

These processes show that AI systems used by LEA officers should be owned and controlled by LEA. There are many experts working for law enforcement and forensic agencies who, due to their level of expertise, are able to determine, for example, whether an AI system or tool that they could acquire from an online repository is sufficient and robust enough to be used in their daily work. However, in

[74] College of Policing, "Weapons and equipment", 23 October 2013. Available at: https://www.college.police.uk/app/armed-policing/weapons-and-equipment

[75] Independent Commission of Investigations (Jamaica), "Caribbean Human Rights and Use of Force Model", 2018. Available at: https://www.policinglaw.info/assets/downloads/Caribbean-Human-Rights-and-Use-of-Force-Policy-Final-June-2018.pdf

[76] McCann and Others v. The United Kingdom (App no.18984/91) 27 September 1995, para.201-208.

much the same way that LEA officers used to be able to bring their own gun to work, such practices would likely struggle to generate confidence from the public that the highest standards are being applied. Indeed, even if only a handful of people using AI systems fall below acceptable standards, that is likely to have a detrimental impact on the public perception of LEAs using AI systems.

Yet, it is important to recognise the expertise that many investigators have, and bring to their work. Where investigators are recognised experts (they provide expert witness testimony in court trials, for example), and they see a need to modify an AI system so that it can work better for their purposes, processes could be developed so that modifications can be proposed to an independent reviewer who could authorise testing of a modified version of a technology for future use. Whether such a system is developed, it is important that LEAs consider the surrounding circumstances of the uses of AI, and are not focussed solely on assessments of specific, narrow cases.

Where access to AI systems are provided centrally by LEA administrators providing sufficient oversight, this can enable implementation of accountability processes. Where systems are in place for LEA officers to be held responsible for their access to, and use of, AI systems, this should prevent any risks of 'responsibility gaps' emerging where it is unclear whom should be held responsible for the use of an AI system. Further, where an LEA officer knows they will be held responsible for their uses of a particular system, they are likely to be more risk-averse in situations where there is potential for harms to manifest.

3.3.11 Provision and use of Non-lethal / Less-lethal options and self-defence equipment

Firearms officers are often trained to use less-lethal options as well as potentially lethal firearms. This is so that LEA officers have options that could resolve a situation without having to take a life (although some supposedly 'less-lethal' weapons have been used with fatal effects).[77] The ambition for the UK College of Policing is that these tools might reduce reliance on conventional firearms or ammunition, without compromising the safety of police officers or others who might be affected.[78] Indeed, where there is time to plan operations, it is '*the duty of the police to devise a realistic plan of action which* [makes] *it possible to arrest the suspect without using lethal force.*'[79]

However, the focus on use of lethal force by police has often resulted in poor consideration of less-lethal weapons; they are not as regulated, meaning that the circumstances which they should be used are subject to less rigour and guidance.[80]

What lessons can this give us for police AI and data analytics?

The model this provides is the concept of escalation. It shows that the use of a potentially harmful tool is not simply binary – permitted or not permitted, but is related to the needs of a context, and that a range of tools allows for a range of responses. AI and data analytics tools might have different harm

[77] Dymond-Bass, Abi, and Neil Corney, 'The use of 'less-lethal' weapons in law enforcement' in Stuart Casey-Maslen, (ed) Weapons under International Human Rights Law, Cambridge University Press, Cambridge, 2014, p.33.

[78] College of Policing, "Weapons and equipment", 23 October 2013. Available at: https://www.college.police.uk/app/armed-policing/weapons-and-equipment

[79] Joint dissenting opinions of Judges Karakas, Wojtyczek and Dedov, Armani Da Silva v. The United Kingdom, (App no. 5878/08) 30 March 2016, para.7.

[80] Dymond-Bass, Abi, and Neil Corney, 'The use of 'less-lethal' weapons in law enforcement' in Stuart Casey-Maslen, (ed) Weapons under International Human Rights Law, Cambridge University Press, Cambridge, 2014, pp.37, 40-47

and risk profiles, some might be based upon less representative data than others, or require more invasive data to use, or have more or less impact upon stigmatised populations. Having a range of tools, and most importantly, an understanding of the different risks of those tools, allows officers to select from this set, working from principles of necessity and proportionality, and in response to the context they are working in. Considering that the use of AI systems in investigations is rarely urgent, this should allow LEAs to plan their use of AI systems so that harms can be reduced as much as possible.

3.3.12 Outsourcing

LEAs are almost always public servants. However, some roles are outsourced to companies whose employees are private citizens (whilst there are some situations where a private citizens or organisations might act as *de facto* state agents,[81] such consideration does not add to the discussion here). Where private security companies have provided security services, this has led to questions over the responsibilities of such organisations and their staff; a key example of this in policing is anti-piracy operations.[82]

What lessons can this give us for police AI and data analytics?

Many LEAs employ in-house experts for digital forensics using AI systems, for example.[83] However, there is variety cross different models of providing forensic capabilities in different countries. If access to specialist AI tools or capabilities is required for an investigation, for example, and these are only available through outsourced analysis of a particular specialist, then this could raise concerns about their ability to be subject to the same, or a comparable, responsibility framework as LEA employees. Outsourcing should not be used as a method to avoid scrutiny. Rather, LEAs could require adequate oversight as an integral part of their outsourced contracts.

3.3.13 Basic Principles

In the system of the United Nations, the application of the right to life to policing has influenced the *'Basic Principles on the Use of Force and Firearms by Law Enforcement Officials'*. This document also draws attention to the special weight upon the policing of lawful assemblies, given the impact policing of these situations can have on democratic participation, freedoms of expression and rights to free association. This may likewise indicate that we should give special attention to the police use of AI and data analytics in relation to lawful assembly as a particular source of high risks towards rights and freedoms.

As a soft law instrument, the *Basic Principles* encourage governments and LEAs to follow these principles. Due to the specificity of the *Basic Principles*, they are provided in the left-hand column of the table below and the adapted requirements for LEA use of AI are provided in the right-hand column. An overall comment on adapting these principles is provided afterwards

[81] Arts. 5 and 8, International Law Commission, ' Report of The International Law Commission on the Work of its Fifty-Third session' (2001) UN Doc. A/CN.4/SER.A/2001/Add.1 (Part 2) p.31

[82] Priddy, Alice, 'The use of weapons in counterpiracy operations', in Stuart Casey-Maslen, (ed) Weapons under International Human Rights Law, Cambridge University Press, Cambridge, 2014, p.144.

[83] Science and Technology Select Committee, "Forensic science and the criminal justice system: a blueprint for change", House of Lords, 2019, para.42. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldsctech/333/333.pdf

Basic Principles	*Requirements adapted for LEA use of AI*
General provisions	
1. Governments and law enforcement agencies shall adopt and implement rules and regulations on the use of force and firearms against persons by law enforcement officials. In developing such rules and regulations, Governments and law enforcement agencies shall keep the ethical issues associated with the use of force and firearms constantly under review.	*1. Governments and law enforcement agencies using AI systems should adopt and implement rules to regulate the impact of such systems upon people they are used on, or against. In developing such rules and regulations, Governments and law enforcement agencies shall keep ethical issues associated with AI use under constant review.*
2. Governments and law enforcement agencies should develop a range of means as broad as possible and equip law enforcement officials with various types of weapons and ammunition that would allow for a differentiated use of force and firearms. These should include the development of non-lethal incapacitating weapons for use in appropriate situations, with a view to increasingly restraining the application of means capable of causing death or injury to persons. For the same purpose, it should also be possible for law enforcement officials to be equipped with self-defensive equipment such as shields, helmets, bullet-proof vests and bullet-proof means of transportation, in order to decrease the need to use weapons of any kind.	*2. Governments and law enforcement agencies should develop a range and means as broad as possible and equip law enforcement officials with various AI systems to allow for a differentiated use of AI. These systems should allow for minimal harm to data-subjects. For the same purpose, law enforcement officials should be provided with appropriate means to avoid needing to use AI systems, such as knowledge about and access to other, potentially less harmful, means of investigation.*
3. The development and deployment of non-lethal incapacitating weapons should be carefully evaluated in order to minimize the risk of endangering uninvolved persons, and the use of such weapons should be carefully controlled.	*3. The use of AI systems that have less potential for harm than others should be carefully evaluated to minimise the risks, and the use of such systems should be carefully controlled.*
4. Law enforcement officials, in carrying out their duty, shall, as far as possible, apply non-violent means before resorting to the use of force and firearms. They may use force and firearms only if other means remain ineffective or without any promise of achieving the intended result.	*4. Law enforcement officials, in carrying out their duty, shall, as far as possible, apply the least harmful means available to them before resorting to harmful AI systems. They may use harmful systems only if other means remain ineffective or without any promise of achieving the intended results.*

5. Whenever the lawful use of force and firearms is unavoidable, law enforcement officials shall: **(a) Exercise restraint in such use and act in proportion to the seriousness of the offence and the legitimate objective to be achieved;** **(b) Minimize damage and injury, and respect and preserve human life;** **(c) Ensure that assistance and medical aid are rendered to any injured or affected persons at the earliest possible moment;** **(d) Ensure that relatives or close friends of the injured or affected person are notified at the earliest possible moment.**	*5. Whenever use of harmful AI systems is unavoidable, law enforcement officials shall:* *(a) Exercise restraint and act in proportion to the seriousness of the offence and the legitimate objective to be achieved;* *(b) Minimize damage and injury, and respect and preserve human life;* *(c) Ensure that assistance, support, and medical aid are rendered to any injured or affected persons harmed by uses of AI at the earliest possible moment.* *(d) Ensure that relatives or close friends of the injured or affected person are notified at the earliest moment.*
6. Where injury or death is caused by the use of force and firearms by law enforcement officials, they shall report the incident promptly to their superiors, in accordance with principle 22.	*6. Where injury or death is caused by the use of AI systems by law enforcement officials, they shall report the incident promptly to superiors, in accordance with principle 22.*
7. Governments shall ensure that arbitrary or abusive use of force and firearms by law enforcement officials is punished as a criminal offence under their law.	*7. Governments shall ensure that arbitrary or abusive use of dangerous AI systems by law enforcement officials is punished as a criminal offence under their law*
8. Exceptional circumstances such as internal political instability or any other public emergency may not be invoked to justify any departure from these basic principles.	*8. Exceptional circumstances such as internal political instability or any other public emergency may not be invoked to justify any departure from these basic principles.*
Special provisions	
9. Law enforcement officials shall not use firearms against persons except in self-defence or defence of others against the imminent threat of death or serious injury, to prevent the perpetration of a particularly serious crime involving grave threat to life, to arrest a person presenting such a danger and resisting their authority, or to prevent his or her escape, and only when less extreme means are insufficient to achieve these objectives. In any event, intentional lethal use of firearms may	*9. Law enforcement officials shall not use dangerous AI systems against data-subjects except to protect victims of crime, to prevent the perpetration of a particularly serious crime involving grave threat to life, to enable arrest of a person presenting such a danger and resisting their authority, or to prevent his or her escape, and only when less extreme means are insufficient to achieve these objectives. In any event, intentional harmful use of AI systems may only be made when strictly unavoidable for protective purposes.*

only be made when strictly unavoidable in order to protect life.	
10. In the circumstances provided for under principle 9, law enforcement officials shall identify themselves as such and give a clear warning of their intent to use firearms, with sufficient time for the warning to be observed, unless to do so would unduly place the law enforcement officials at risk or would create a risk of death or serious harm to other persons, or would be clearly inappropriate or pointless in the circumstances of the incident.	*10. In the circumstances provided for under principle 9, law enforcement officials shall identify as such when using AI systems and make clear their intentions and justifications for uses of (harmful) AI systems, with sufficient time for objections to be raised, unless to do so would unduly place law enforcement officials at risk or would create a risk of death or serious harm to other persons, or would be clearly inappropriate or pointless in the circumstances of the incident.*
11. Rules and regulations on the use of firearms by law enforcement officials should include guidelines that: **(a) Specify the circumstances under which law enforcement officials are authorized to carry firearms and prescribe the types of firearms and ammunition permitted;** **(b) Ensure that firearms are used only in appropriate circumstances and in a manner likely to decrease the risk of unnecessary harm;** **(c) Prohibit the use of those firearms and ammunition that cause unwarranted injury or present an unwarranted risk;** **(d) Regulate the control, storage and issuing of firearms, including procedures for ensuring that law enforcement officials are accountable for the firearms and ammunition issued to them;** **(e) Provide for warnings to be given, if appropriate, when firearms are to be discharged;** **(f) Provide for a system of reporting whenever law enforcement officials use firearms in the performance of their duty.**	*11. Rules and regulations on the use of AI systems by law enforcement officials should include guidelines that:* *(a) Specify the circumstances under which law enforcement officials are authorized to use AI systems and prescribe the types of systems permitted;* *(b) Ensure that AI systems are used only in appropriate circumstances and in a manner likely to decrease the risk of unnecessary harm;* *(c) Prohibit the use of those AI systems that cause unwarranted injury or present an unwarranted risk;* *(d) Regulate the administration, data storage and access to AI systems, including procedures for ensuring that Law enforcement officials are accountable for the AI systems used by them;* *(e) Provide for warnings to be given, if appropriate, when AI systems are used;* *(f) Provide for a system of reporting whenever Law enforcement officials use AI systems in the performance of their duty.*

Policing unlawful assemblies	
12. As everyone is allowed to participate in lawful and peaceful assemblies, in accordance with the principles embodied in the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights, Governments and law enforcement agencies and officials shall recognize that force and firearms may be used only in accordance with principles 13 and 14.	*12. As everyone is allowed to participate in lawful and peaceful assemblies, in accordance with the principles embodied in the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights, Governments and law enforcement agencies and officers shall recognize that AI systems may be used only in accordance with principles 13 and 14.*
13. In the dispersal of assemblies that are unlawful but non-violent, law enforcement officials shall avoid the use of force or, where that is not practicable, shall restrict such force to the minimum extent necessary.	*13. In the dispersal of assemblies that are unlawful but non-violent, law enforcement officials shall avoid the use AI systems, where that is not practicable, shall restrict use of AI systems to the minimum extent necessary.*
14. In the dispersal of violent assemblies, law enforcement officials may use firearms only when less dangerous means are not practicable and only to the minimum extent necessary. Law enforcement officials shall not use firearms in such cases, except under the conditions stipulated in principle 9.	*14. In the dispersal of violent assemblies, law enforcement officials may use AI systems only when less dangerous means are not practicable and only to the minimum extent necessary. Law enforcement officials shall not use AI systems in such cases, except under the conditions stipulated in principle 9.*
Policing persons in custody or detention	
15. Law enforcement officials, in their relations with persons in custody or detention, shall not use force, except when strictly necessary for the maintenance of security and order within the institution, or when personal safety is threatened.	*15. Law enforcement officials, in their relations with persons in custody or detention, shall not use AI systems, except when strictly necessary for the maintenance of security and order within the institution, or when personal safety is threatened.*
16. Law enforcement officials, in their relations with persons in custody or detention, shall not use firearms, except in self-defence or in the defence of others against the immediate threat of death or serious injury, or when strictly necessary to prevent the escape of a person in custody or detention presenting the danger referred to in principle 9.	*16. Law enforcement officials, in their relations with persons in custody or detention, shall not use AI systems, except to protect against the immediate threat of death or serious injury, or when strictly necessary to prevent the escape of a person in custody or detention presenting the danger referred to in principle 9.*
17. The preceding principles are without prejudice to the rights, duties and responsibilities of prison officials, as set out in the Standard Minimum Rules for the Treatment of Prisoners, particularly rules 33, 34 and 54.	*17. The preceding principles are without prejudice to the rights, duties and responsibilities of prison officials, as set out in the Standard Minimum Rules for the Treatment of Prisoners, particularly rules 33, 34 and 54.*

Qualifications, training and counselling	
18. Governments and law enforcement agencies shall ensure that all law enforcement officials are selected by proper screening procedures, have appropriate moral, psychological and physical qualities for the effective exercise of their functions and receive continuous and thorough professional training. Their continued fitness to perform these functions should be subject to periodic review.	*18. Governments and law enforcement agencies shall ensure that all law enforcement officials are selected by proper screening procedures, have appropriate moral, psychological and physical qualities for the effective exercise of their functions and receive continuous and thorough professional training. Their continued fitness to perform these functions should be subject to periodic review.*
19. Governments and law enforcement agencies shall ensure that all law enforcement officials are provided with training and are tested in accordance with appropriate proficiency standards in the use of force. Those law enforcement officials who are required to carry firearms should be authorized to do so only upon completion of special training in their use.	*19. Governments and law enforcement agencies shall ensure that all law enforcement officials are provided with training and are tested in accordance with appropriate proficiency standards in the use of force. Those law enforcement officials who are required to use AI systems should be authorized to do so only upon completion of special training in their use.*
20. In the training of law enforcement officials, Governments and law enforcement agencies shall give special attention to issues of police ethics and human rights, especially in the investigative process, to alternatives to the use of force and firearms, including the peaceful settlement of conflicts, the understanding of crowd behaviour, and the methods of persuasion, negotiation and mediation, as well as to technical means, with a view to limiting the use of force and firearms. Law enforcement agencies should review their training programmes and operational procedures in the light of particular incidents.	*20. In the training of law enforcement officials, Governments and law enforcement agencies shall give special attention to issues of police ethics and human rights, especially in the investigative process, to alternatives to the use of AI systems, including the peaceful settlement of conflicts, the understanding of crowd behaviour, and the methods of persuasion, negotiation and mediation, as well as to technical means, with a view to limiting the use of AI systems. Law enforcement agencies should review their training programmes and operational procedures in the light of particular incidents.*
21. Governments and law enforcement agencies shall make stress counselling available to law enforcement officials who are involved in situations where force and firearms are used.	*21. Governments and law enforcement agencies shall make stress counselling available to law enforcement officials who are involved in situations where AI systems are used.*
Reporting and review procedures	
22. Governments and law enforcement agencies shall establish effective reporting and review procedures for all incidents referred to in principles 6 and 11 (f). For incidents	*22. Governments and law enforcement agencies shall establish effective reporting and review procedures for all incidents referred to in principles 6 and 11 (f). For incidents reported*

reported pursuant to these principles, Governments and law enforcement agencies shall ensure that an effective review process is available and that independent administrative or prosecutorial authorities are in a position to exercise jurisdiction in appropriate circumstances. In cases of death and serious injury or other grave consequences, a detailed report shall be sent promptly to the competent authorities responsible for administrative review and judicial control.	*pursuant to these principles, Governments and law enforcement agencies shall ensure that an effective review process is available and that independent administrative or prosecutorial authorities are in a position to exercise jurisdiction in appropriate circumstances. In cases of death and serious injury or other grave consequences, a detailed report shall be sent promptly to the competent authorities responsible for administrative review and judicial control.*
23. Persons affected by the use of force and firearms or their legal representatives shall have access to an independent process, including a judicial process. In the event of the death of such persons, this provision shall apply to their dependants accordingly.	*23. Persons affected by the use of AI systems or their legal representatives shall have access to an independent process, including a judicial process. In the event of the death of such persons, this provision shall apply to their dependants accordingly.*
24. Governments and law enforcement agencies shall ensure that superior officers are held responsible if they know, or should have known, that law enforcement officials under their command are resorting, or have resorted, to the unlawful use of force and firearms, and they did not take all measures in their power to prevent, suppress or report such use.	*24. Governments and law enforcement agencies shall ensure that superior officers are held responsible if they know, or should have known, that law enforcement officials under their command are resorting, or have resorted, to the unlawful use of AI systems, and they did not take all measures in their power to prevent, suppress or report such use.*
25. Governments and law enforcement agencies shall ensure that no criminal or disciplinary sanction is imposed on law enforcement officials who, in compliance with the Code of Conduct for Law Enforcement Officials and these basic principles, refuse to carry out an order to use force and firearms, or who report such use by other officials.	*25. Governments and law enforcement agencies shall ensure that no criminal or disciplinary sanction is imposed on law enforcement officials who, in compliance with the Code of Conduct for Law Enforcement Officials and these basic principles, refuse to carry out an order to use AI systems, or who report such use by other officials.*
26. Obedience to superior orders shall be no defence if law enforcement officials knew that an order to use force and firearms resulting in the death or serious injury of a person was manifestly unlawful and had a reasonable opportunity to refuse to follow it. In any case, responsibility also rests on the superiors who gave the unlawful orders.	*26. Obedience to superior orders shall be no defence if law enforcement officials knew that an order to use AI systems resulting in the death or serious injury of a person was manifestly unlawful and had a reasonable opportunity to refuse to follow it. In any case, responsibility also rests on the superiors who gave the unlawful orders.*

What lessons can this give us for police AI and data analytics?

Overall, the *Basic Principles* do seem to be adaptable to AI systems. Indeed, the identification of potential harms raised by the *Basic Principles* are highly-useful for areas where AI systems could also cause harm. However, as these *Basic Principles* are intended to be applied in the context of tactical situations that are 'life or death', the focus on lethal uses of force and firearms does skew the potential requirements for AI systems. As noted above, whilst harms raised by (mis)use of AI systems can cause serious injuries to physical and mental integrity, and death, not all uses of AI systems are as harmful as firearms. Of course, not all uses of firearms cause fatalities, but the understanding of risks associated with firearms assumes that this is always a possibility, whereas it is unlikely to always be a possibility with every AI system.

3.4 Some potential findings

The concepts of lawfulness, necessity, proportionality and accountability should not be alien to police forces, both in Europe and around the world. Indeed, the models used for applying these concepts in the context of regulating use of firearms are already embedded in LEAs. It is clear that AI systems can cause harms, especially in the LEA context which is, by its very nature, a high-stakes domain.

The way that a police force manages the balance between risk of harm and seizing the utilities of AI systems tells us something about how it works, its organisational priorities and structures, and the political and legal context that surround it. We see that firearms are treated with a special focus as an inherently risky and dangerous artefact with special policies and procedures applied to minimise harm and maximise the safety of LEA officers and the public.

We are early in the development and use of AI and data analytics by police forces, whereas the police use of firearms is much further down the line. Yet, we can see that there is not only a need, but also the possibility, for LEA use of AI systems to be regulated in a much more comprehensive way. We already have several frameworks for applying to AI systems in terms of ethics (indeed, as shown in Section 2, they can be adapted to the LEA context). But, thinking about AI ethics from the perspective of firearms regulation provides an alternative frame for viewing and managing AI ethics that is already based on LEA policies and procedures.[84]

Granting the police the ability to use a tool or a weapon is not the end of the story. Improper and unlawful use of firearms by police officers can be prosecuted. Police officers have stood trial for murder[85] or been the subject of independent investigations into their activities using firearms.[86]

A critical question remains where the controls that are in place are not enforced. This approach is worthless if these controls are solely words on paper in policies and best practices guides, either for firearms currently, or for AI tools in the future. An ideal set of measures would operate at multiple levels and in harmony to increase the safety of this use of risky tools.

Governance must equally not only serve the policing interest. The eventual set of governance mechanisms around potentially harmful tools must also serve the interests of policed communities and be perceived by them as appropriate and sufficient.

[84] Hagendorff, T. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines***30,** 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8

[85] Dodd, V. (2015) 'Azelle Rodney Shooting: Police Marksman Cleared of Murder' The Guardian 3 July <https://www.theguardian.com/uk-news/2015/jul/03/police- marksman-anthony-long-cleared-azelle-rodney-murder>, cited in Turner, Ian David, "Arming the Police in Britain: A Human Rights Analysis", *the Police Journal: Theory, Practice & Pricniples*, 90(2), 107-127.

[86] Independent Police Complaints Commission (2007) Stockwell One: An Investigation Into the Shooting of Jean Charles de Menezes at Stockwell Underground Station on 22nd July 2005, 8 November www.ipcc.gov.uk/stockwell_one.pdf - 2007-11-08,

4 Dealing with Criminal Procedure Rules in LEA Technologies

4.1 Introduction

Legislative power within the criminal justice systems in the EU still lies with the member states and therefore there are many discrepancies between the rules from country to country. However, there are similar baseline requirements that any LEA technology design in Europe will have to meet, to which it would then be possible to add additional specifications arising from national or organisational criminal procedure rules.

As established by the European Charter of Fundamental Rights, every person has a right to a trial within a reasonable time.[87] Technologies such as those developed in the INSPECTr project are capable of facilitating that right by assisting LEAs across Europe that are struggling to cope with reduced funding and lack of human resources. For example, the natural language processing (NLP) tools and relational navigation tools would be capable of recognising patterns more efficiently at the evidence gathering phase, increasing speeds in at least one step in the investigative process. That is of course assuming that meeting all other requirements that these types of pattern recognising technologies must adhere to, such as additional authorisation or lengthy data sharing protocols, do not elongate the process of using the tools beyond the feasibility of their use. As a preventative measure, some LEA technologies can be designed to meet those additional requirements by design.

This Section will first look at the legal requirements that LEA technologies, similar to those developed within the project, and their practitioners are expected to meet as part of the rules of criminal procedure. It will give suggestions on how the tools developed in INSPECTr could meet some of these requirements by design and consider what the criminal procedure that includes rules that could be applied to AI tools might look like in the future. It will also consider the data lifecycle in law enforcement and how criminal procedure rules establish data sharing practices within that cycle.

4.2 Legal requirements

4.2.1 For the technologies

In addition to the AI and NLP tools developed in INSPECTr, LEA technologies can also include much simpler tools such as audio and video recording devices. Since they have been around for longer and have many more applications, these tools have become essential to police work. They have also been established as method for meeting some of the criminal procedure requirements and not only having to meet requirements themselves. For example, when an accused child is questioned, the police must be able to provide an audio-visual recording of the questioning to demonstrate the presence of a lawyer (or not) and whether the child has been deprived of liberty or not.[88] The technology is being used to facilitate procedural requirements.

One of the procedural rights of the defendant in a trial is to call and cross-examine witnesses. Under circumstances where 'vulnerable witnesses' are involved, such as minors, the victims are able to provide their statements without directly participating in the court proceedings. Based on case law simply providing a video statement from a vulnerable witness is infringing on the right to cross-examination and could, therefore, not be sufficient evidence. INSPECTr tools have the capability to support the prosecution's case by providing additional evidence to witness statements to protect their

[87] Balsamo, Antonio, "The Content of Fundamental Rights." In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp.100-172.
[88] Balsamo, Antonio, "The Content of Fundamental Rights." In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp.100-172.

health and mental wellbeing; for example, in cases of child exploitation, AI tools might be able to enhance testimonies by linking information across different datasets without the need for children to experience difficult cross-examinations. These tools would not replace the witness statement's value as evidence but would enhance the quality of evidence by diversifying the accounts of the crime, especially in cases where the decision of the court would often depend on the strength of the witness statements.

Many of the requirements the audio-visual technologies will be expected to meet themselves relate to the source of data, especially where this is from surveillance. While the INSPECTr tools do not do surveillance themselves, they do analyse products of surveillance and should therefore be informed of the criminal procedure rules around interceptions and privacy to prevent dismissal of evidence due to analysing unlawfully gathered surveillance data. One of the fundamental rights in the criminal procedure, as outlined by the European Court of Human Rights (ECtHR), that is most relevant for technologies used for surveillance, is the right to privacy. Art.8(1) of the European Convention on Human Rights (ECHR) states: '*Everyone has the right to respect for his private and family life, his home and his correspondence.*'.

European case law of recent years has had a huge influence on governing communication interceptions or "wiretapping" as it goes beyond following criminal procedure rules and asks questions of wider human rights in a democratic state.[89] Art.8(2) of the ECHR provides conditions for interference with the right:

> *'There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.'*

As the tools used in INSPECTr are designed to be used against organised crime, such as those outlined in the use-cases (terrorism, fraud and child exploitation) it can be argued that the third requirement is met and if the tools are used in the intended way then the first requirement would also be met. The second requirement demands a proportionate response to the legitimate aims being pursued. It must also be based in domestic law. The ECtHR has held that where the surveillance is secret, it remains unchallengeable for the defendant and it is therefore in violation of their right to privacy. This could mean that for example a behavioural pattern recognition tool would have more legitimacy when analysing the product of overt surveillance, as covert surveillance requires increased supervision of the national judicial authority. If secret surveillance is performed in a contracting state it also becomes a jurisdictional issue as the margins for justification of secret surveillance are more restricted.[90]

4.2.1.1 Jurisdiction restrictions

A related barrier also stemming from jurisdictional restrictions is that digital evidence can be tracked back to servers in various different locations across the globe. This can introduce jurisdictional challenges to the investigation as the investigating officer will not have the authority to request the evidence from the server located abroad. Newly proposed EU legislation would allow judicial authorities to directly request access to evidence from service providers where the data is stored on

[89] Balsamo, Antonio, "The Content of Fundamental Rights." In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp.100-172.
[90] Zoltán Varga v. Slovakia, (Apps no. 58361/12, 25592/16, 27176/16) 22 November 2021, para.151.

a server located outside the EU but operates within any EU member state. This would ease the processes as the request would no longer have to go through the judicial authority of the other state.[91]

AI is a field that's evolving at great pace and therefore the legislation and policies around using it are also constantly changing. As highlighted before, criminal procedure rules require evidence to be collected in a lawful manner, meaning that unlawfully collected evidence would be dismissed from court. That leads to practitioners being hesitant when using AI and NLP solutions as what might be lawfully collected one day, could be dismissed from court a few years later when the case finally makes it to trial,[92] so these state-of-the-art tools might be favoured less in jurisdictions or investigation fields where the case takes longer to build (oblivious to the irony that the aim of these tools is to assist and speed up the collection of digital evidence).

4.2.2 For the practitioners

As the nature of the crimes that INSPECTr tools will be looking to help prevent could involve covert operations, it is important to ensure that the investigation techniques remain essentially passive: criminal procedure rules require that police do not incite the crime[93]. By analysing the patterns of behaviour and the patterns in language used, the INSPECTr tools remain passive in their role in the investigation as they will not be influencing these behaviours of the suspect(s) but rather analysing and making speculative predictions based on past behaviours. Nevertheless, if the AI tool analyses the product of covert operations it is affected by the same criminal procedure rules as all subsequent analysis and could still be dismissed from court due to improper methods collecting the original evidence. The European Court requires there be a clear procedure in place for authorising covert measures as well as proper supervision (in most cases judicial).[94]

Similarly to AI applications used in other fields, it is also important in the criminal justice system to establish accountability for the tool: who is responsible for the decisions the AI makes, for the way it makes decisions and for when it fails to make the correct decisions. Doubts still remain whether accountability of the system lies with the judicial authority, the company providing the solution or if it trickles all the way down to the developers who trained the algorithm.[95] Often the AI solution the LEAs use is bought in, meaning the algorithms have been trained by a private company claiming intellectual property rights which may make them hesitant to share the methods for training their algorithms.[96] However, greater transparency behind the algorithm construction would make sharing or establishing accountability with the person or legal body authorising the use of the solution easier to manage and less complex to implement in cases where the system fails. Transparency behind the

[91] Proposal for a Regulation of the European Parliament and of the Council on the European Production and Preservation Orders for electronic evidence in criminal matters, {SWD(2018) 118 final} - {SWD(2018) 119 final}, 17 April 2018; European Commission, "Security Union Facilitating Access To Electronic Evidence", 2018. Available at: https://ec.europa.eu/commission/presscorner/api/files/attachment/855819/Factsheet%20E-evidence.pdf;

[92] Goodison, Sean E., Robert C. Davis, and Brian A. Jackson, "Digital Evidence and the U.S. Criminal Justice System: Identifying Technology and Other Needs to More Effectively Acquire and Utilize Digital Evidence", Santa Monica, California, RAND Corporation, 2015.

[93] Balsamo, Antonio, "The Content of Fundamental Rights." In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp.100-172.

[94] Balsamo, Antonio, "The Content of Fundamental Rights." In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp.100-172.

[95] Carrera, Sergio, Valsamis Mitsilegas, and Marco Stefan, "Criminal Justice, Fundamental Rights and the Rule of Law in the Digital Age: Report of a CEPS and QMUL Task Force", Centre for European Policy Studies, Brussels, 2021.

[96] CEPEJ, European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, Council of Europe, Strasbourg, 2018.

algorithm also makes it easier for the defendant (or their legal team) to determine if the technologies have been used appropriately, and contest the findings where practicable.

4.3 Data lifecycle

The stages of the data lifecycle that raise the most ethical questions around LEA technologies are storage, sharing and destruction. The purpose for collection is usually for an investigation and criminal procedure rules are lenient in the limitations set on the data collected and stored during that period – the main ethical considerations are around what happens after the end of the investigation.

4.3.1 Data storage

In most cases, data is moved to long term storage at the point of the court decision.[97] At this point the distinction is made between irrelevant files and relevant files classed as "intelligence". Files classed as intelligence could be kept on file for 6 years in the UK.[98] Files that are classed as forensically relevant could be on file for 80 years in the UK and 110 years after the offender's birth in the US.[99] Automatic storage of clearly irrelevant data for more than 6 months is not considered justified under the right to privacy.[100]

The discretion of classing material as relevant is available to the investigator, the officer in charge of the investigation or to the disclosure officer. They will judge whether the material could have any bearing on another case under investigation, the person under investigation or the surrounding circumstances of the case, and unless deemed fully irrelevant to all of the above, it will be deemed as incapable of having any impact on a case.[101] Retention of personal data, e.g. images of the offender, indefinitely would harm the offender's right to privacy and LEAs are required to review stored data at regular intervals for compliance.[102]

There may be some discrepancies between policies for storing investigation files and forensic files: in the UK forensic files may be kept for up to 20 years longer than police evidence depending on the seriousness of the crime it was related to.[103] Indeed, EU member states should develop time limits of data storage that are appropriate for their jurisdiction.[104]

[97] National Police Chiefs Council (UK), "National Digital and Physical Evidence Retention Guidance", NPCC Digital and Physical Evidence Group, 2021.

[98] Elkins, Matt, "What information does the police hold on me?", Police Cautions , 2019.Available at: https://policecautions.uk/2019/02/09/what-information-does-the-police-hold-on-me/.

[99] National Police Chiefs Council (UK), "National Digital and Physical Evidence Retention Guidance", NPCC Digital and Physical Evidence Group, 2021; Elkins, Matt, "What information does the police hold on me?", Police Cautions , 2019.Available at: https://policecautions.uk/2019/02/09/what-information-does-the-police-hold-on-me/.

[100] Roman Zakharov v Russia (App no. 47143/03), 4 December 2015, para.255.

[101] Ministry of Justice (UK), "Criminal Procedure and Investigations Act Code of Practice", Ministry of Justice, 2015.

[102] S. and Marper v. the United Kingdom [GC], (Apps no. 3052/04, 30566/04), 4 December 2008.

[103] National Police Chiefs Council (UK), "National Digital and Physical Evidence Retention Guidance", NPCC Digital and Physical Evidence Group, 2021.

[104] Art.5, European Parliament and Council, Directive (EU) 2016/680 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, OJ L 119/89, Vol.59, 4 May 2016 (Law Enforcement Directive, hereafter: LED)

4.3.2 Destruction of data

Data moved to long term storage will likely be reviewed at set intervals depending on the nature of the crime.[105] If found to no longer be relevant the data will be destroyed. All data should be shredded to the point where, for all practical purposes, it cannot be recreated. In the UK the forensic unit is required to maintain records to demonstrate all destruction of case data.[106] This rule may be different in member states where the equivalent of a forensic unit is not a part of the LEA.

From a data protection perspective, there was nothing found on any differentiation made between content data and metadata, and there was nothing specifically mentioned at the EU level which would suggest that a distinction should be made in how these types of data should be handled. The only possible distinction to be made at this general level is that irrelevant data should be removed.[107]

4.3.3 Data sharing

The lack of common criminal procedure rules across the EU leads to two options when sharing evidence or data transnationally: 1) the principle of mutual assistance where the cooperation comes from the governments working together rather than the individual LEAs in each country, and 2) mutual recognition where cooperation is done directly between the judicial authorities.[108] Mutual assistance is too limited as a principle to base evidence sharing on when the one crime is being committed in multiple countries – particularly relevant to the use cases for the tools being developed in this project as terrorism, fraud and child exploitation often have pathways that go beyond borders, and therefore the principle of mutual recognition is favoured in these types of investigations. Having fewer bodies involved in authorising the use and sharing of the evidence made available by AI and NLP tools allows for creating clearer lines of accountability for the systems.

Transmission of data between states must adhere to principles of proportionality, especially in the case of sharing personal information, e.g. DNA profiles.[109] For full availability of the profiles, transmission must be done through a state authorised national contact point. The contact point should not be able to identify the DNA profile directly – only after the investigating bodies are able to match the DNA to an existing profile, may the contact point request the transmission of additional personal information. This procedure will then follow the national legislation of the state to whom the request was made. Many national laws in the EU require the criminal offence to be of a severe enough nature to proceed. This is an example of privacy principles and confidentiality of evidence restricting the amount of data that is able to be shared. As AI tools are not as equally established across jurisdictions, sharing personal data for the purpose of it being analysed by the AI may be resisted as it’s not exactly clear where this measure lies on the scale of proportional methods of investigation. However, no specific rules were found in relation to sharing the results of the AI analysis – this would rather fall under the operating LEA’s individual confidentiality policies.

[105] National Police Chiefs Council (UK), “National Digital and Physical Evidence Retention Guidance”, NPCC Digital and Physical Evidence Group, 2021.

[106] National Police Chiefs Council (UK), “National Digital and Physical Evidence Retention Guidance”, NPCC Digital and Physical Evidence Group, 2021.

[107] National Police Chiefs Council (UK), “National Digital and Physical Evidence Retention Guidance”, NPCC Digital and Physical Evidence Group, 2021.

[108] Daniele, Marcello, and Ersilia Calvanese. 2018. “Evidence Gathering.” In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp., 353-392;

[109] Daniele, Marcello, and Ersilia Calvanese. 2018. “Evidence Gathering.” In Handbook of European Criminal Procedure, by Roberto E. Kostoris (ed), Springer, 2018, pp., 353-392.

4.3.4 Organisational requirements

There is little information currently available on the LEA's organisational level requirements that is universal for handling AI technologies and whether there are restrictions applied to the use stemming from the jurisdiction, crime severity, the department, the investigation field, or other factors. As discussed earlier as part of data sharing principles, that might be due to the confidentiality policies of individual LEAs.

The organisational level requirements for LEA technologies, that are most available, are around the practices used for facial recognition technologies. These tools have been surrounded by a lot of controversy and discussion around privacy violation with some cases being taken to court (for example, use of facial recognition by South Wales Police)[110] and are therefore receiving more press attention. The concerns around it still relate to the right to privacy, as mentioned multiple times before, and whether the use of these tools are in violation of this right. Each LEA looking to use these tools will have to consider a data protection impact assessment that considers the proportionality of the measure alongside the perceived impact with some basis in the national law.[111] The distinctions in national law, and different impacts that could be raised for the different societies being policed could mean that there could be a lot of variation between the assessment made by LEAs across Europe. Indeed, in some EU member states the use of facial recognition technology is more forcefully opposed than in others.[112] Other AI tools that have the potential to infringe on the right to privacy through high-risk processing would also be expected to provide a data protection impact assessment demonstrating that the risks have been considered and appropriate measures are in place to protect people's privacy.

The constant press around LEAs taken to court for the use of facial recognition technologies highlights the lack of trust (perhaps justifiably) around the use of these measures and the ambiguity that still exists around the impact these tools have on the privacy rights of individuals. For better trust, the impact to privacy needs to be clear from the training stage of the algorithm, demonstrating that any societal bias has been considered and accounted for.

4.3.5 Considerations for the future

Currently, the ECtHR recognises that the storage of cells, footprints and DNA samples for an undetermined time in the name of preventing future crimes infringes on a person's right to privacy.[113] At the moment there is no existing case law at the European level that would apply those same principles on patterns in images, language, or behaviours detected by AI. The AI is able to learn from these patterns and apply them to future patterns or prediction models. Could this become a question of infringing on the defendant's right to privacy in the future? One could argue that the patterns of one person committing a stand-alone crime should not influence the evidence collected against another separate perpetrator in such a way as to predict the outcomes of actions based on decisions made by completely different people and the likelihood of those same decisions being made again by a second individual. Whether this constitutes an infringement will, of course, depend on the training models behind the AI and its ability to make more sophisticated predictions than finding the probability of a scenario.

As in many other fields, where AI tools are used to support decision making, the transparency of the system becomes important – it is not necessarily important to understand the reason why or how the system draws its conclusions or finds the patterns, but how the system was trained, what type of

[110] R (Bridges) v. The Chief Constable Of South Wales Police [2020] EWCA CIV 1058
[111] Art.27, LED
[112] Goujard, Clolthilde, "Europe edges closer to a ban on facial recognition." Politico, 20 September 2022. Available at: https://www.politico.eu/article/europe-edges-closer-to-a-ban-on-facial-recognition/
[113] Gaughran v. the United Kingdom, (App no. 45245/15) 13 June 2020

material it was shown and equally importantly what material was it not shown – understanding how the training and testing of the AI influences its pattern making capabilities can help to ensure the rights of the accused are not infringed upon.

4.4 Conclusions

A call for more transparent AI tools is found when analysing a number of these criminal procedure requirements. If it is challenging to understand why an AI tool made a particular recommendation, or how the data leading to the recommendation was treated, then it could be difficult to demonstrate that rules of criminal procedure have been followed. For example, poor transparency could lead to problems ensuring that chain of custody has been maintained. This can also link to ensuring that meaningful human oversight is a feature of the overall operation of AI systems in the LEA context, so as to increase trust and overall transparency in the processing of investigative data.[114] Meaningful, in this case, not necessarily meaning a human who can explain the detail of the mathematical processes within the AI system, but a person who understands how it was trained and can interpret how the decision was made.

Another principle of AI design relevant to a number of these criminal procedure rules is the need for clear accountability. Whether that be for the whole product of its analysis or divided up by the various applications or stages of the solution is not determined by case law as of yet. A commonly understood hierarchy for responsibility is another way for establishing greater trust in the systems, making their use more favourable among both the practitioners weary of the ever-changing rules around their use as well as among the general public.

With respect to data lifecycles, it is clear that there is limited high-level guidance in this area, and that adapting details from other, similar, areas of forensic or law enforcement enquiry is difficult as there are few commonalities between different countries, at least in openly available literature, and most high-level recommendations require individual states to determine 'appropriate' measures without giving further guidance. As such, LEA technologies could be developed with adjustable approaches to data lifecycles that can be specific in more detail by LEAs who intend to use the technologies operationally.

[114] European Commission, "Building Trust in Human-Centric Artificial Intelligence", 2019. Available at: https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/building-trust-human-centric-artificial-intelligence

5 'Know Your Customer' Exploitation Risk Assessment

This section provides an explanation of how a 'Know Your Customer' (KYC) assessment from the domain of financial services has been adapted for technology exploitation. The research work for this approach has taken place across both the ROXANNE and INSPECTr projects, and has been tailored for each project. KYC is often used to evaluate risks of providing financial services where there is a risk that those services might be used for, or to enable, unlawful activities. The conducting of a KYC assessment, arguably, facilitates sharing of responsibly for preventing criminal activities between the organisation providing financial services and the organisations that could abuse those services. With AI systems, a similar approach can be considered as the moral responsibly of an AI manufacturer does not end when they provide tools to end-users. Conceivably, a KYC assessment for AI systems could eventually become a legal duty to help avoid and mitigate risks of the AI system being misused or used for mass surveillance. Below, an approach to doing this assessment is provided.

It is important that we recognise and appreciate the risks of harm that can arise from AI systems, and that the potential harm can change with the context. For example, using an AI system to evaluate personal data and reveal information that the individuals usually keep secret regarding their sexuality, ethnicity, or political opinions, would normally be seen as an intrusion into the privacy of a person that would be generally unacceptable to society. However, during a criminal investigation, uncovering such information might be relevant to a legitimate investigation and so would be acceptable to society where it is lawful, necessary, and proportionate. But, processing such data at a scale including lots of people who are irrelevant to the investigation would also not be acceptable. The developers of such technologies bear some responsibility for the impacts that their tools create – unacceptable actions that are perpetrated using the tools might not be possible without them and so the developers enable such uses. It is difficult to state clear lines in advance around how technologies can and cannot be used, especially where the acceptability of a certain action might be in a grey area. An alternative is to recognise the risks of unacceptable use of tools, and try to avoid or mitigate those risks. Ultimately though, there should not be an entirely open market for powerful AI systems that have the potential to cause harm because those harms could be hugely significant.

It is important to note that those implementing a KYC approach are not trying to prove a case akin to a legal argument about whether an organisation should provide a service. Rather, it is about understanding the risks that could be created by sharing AI technologies where there are threats and vulnerabilities that could result in misuse or mass surveillance. Some risk assessments only assess risk based on wrongful acts that people are performing, or have performed and assume that previous behaviour is a guide to future behaviour.[115] Whilst a backward-looking approach can be informative, and could be instructive in some circumstances, it is insufficient as risks can only manifest in the future and so a forward-looking approach is needed. This includes the reputational risks that could arise for a company that provides technologies that can analyse and output highly sensitive information to customers who use it in an irresponsible way, and risks to people whom such technologies are used on or against.[116] The latter concern is like a data protection impact assessment, where risks to a data-subject are considered by the data controller. In this situation, however, the risks are not just data protection risks but also risks of misuse, mass surveillance, and human rights violations.

[115] See, for example, the Premier League Owners' and Directors' Test. *Premier League, Handbook Season 2021/22*, The Football Association, 2021,pp.141-148. Available at: https://resources.premierleague.com/premierleague/document/2022/04/07/c0d0f725-3fe3-4470-ba6c-ed83c7aa75fe/PL_Handbook_2021_22_DIGITAL_07-04-22.pdf

[116] See, for example, Benmeleh, Yaacov and Eliza Ronalds-Hannon, "NSO Group explores shut down of its Pegasus spyware unit, sale", *Al Jazeera*, 2021. Available at: https://www.aljazeera.com/economy/2021/12/14/nso-group-explores-shut-down-of-its-pegasus-spyware-unit-sale

To explain in more detail, KYC approaches involve potential customers completing questionnaires about their situation, and this is then followed by a risk assessment done by the technology provider to determine whether the risk profile of a particular customer is acceptable. Following this introduction, the document provides an understanding of risk, provides a short threats and vulnerabilities assessment where particular areas of interest are highlighted. These then lead onto specific questions for both the prospective end-user and the technology provider themselves. The responses to the questionnaires can then be analysed with additional background research to complete the risk assessment and come to a conclusion on whether the risk of providing technologies to the prospective end-user is acceptable or not.

The intention is that this risk assessment can be used by partners exploiting technologies from projects like INSPECTr to determine the risk of misuse, mass surveillance, and human rights infringements that would be posed by providing a technology to a particular customer. The work done in developing this risk assessment in the project might not be comprehensive for all types of exploitation and could be adapted by different organisations providing different goods or services for their specific needs; TRI intends to expand this work in future projects to deal with more ethical or human rights risks.

5.1 Risk

Risk can be understood in different ways depending on the domain in question. However, the risks considered in the below assessment are multi-faceted and include the reputation of the technology provider, the likelihood that technologies will be used for purposes beyond those intended by the project partners, and the impact of those uses on individuals. Thus, a general understanding of risk needs to be used. One is provided by the International Organization for Standardization's work on risk management, and defines risk as the effect of uncertainty on meeting objectives due to incomplete knowledge about how a decision will impact on future events or circumstances.[117]

The objective of exploiting project technologies is for them to be used by LEAs during organised crime investigations so that offenders and their wrongdoing can be identified more easily and sooner than with existing technologies, speeding up the LEA response to such criminality. Uncertainties for meeting this objective in the context of this analysis would be recipients of the project technologies who use them for purposes other than organised crime investigations.

It is important to acknowledge that where the technologies developed in INSPECTr are used for purposes beyond legitimate and responsible criminal investigations, they can pose a hazard to individuals and to society. Individual persons can be harmed by LEA technologies where tools are used to expose private information that then creates privacy harms which might be physical, emotional, psychological, reputation, economic, relational, discriminatory, or impact on a person's autonomy.[118] LEA technologies can also have impacts on a wider, societal level where, for example, biases in the technology might result in people from a particular group being treated differently than others. Due to these potential hazards, it is important that we take into account the possible risks of providing the technology to customers or clients. This is an area that is generally absent in the EU's proposed AI regulation which focuses on ensuring technologies meet a conformity assessment and are then used appropriately, but does not acknowledge the part in-between where technologies are provided to end-users. The present risk assessment can fill this gap in a practical way and potentially suggest a possible policy response for the future.

[117] ISO, Guide 73:2009 Risk Management – Vocabulary, 2009. Available at: https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en

[118] Keats Citron, Danielle, and Daniel J. Solove "Privacy Harms", Boston University Law Review, Vol. 102, 2022.

In risk assessments for financial services, which inspired the below assessment process, risks are seen as the culmination of threats, vulnerabilities, and consequences.[119] As such, a risk assessment is a process for identifying, analysing, and evaluating the uncertainties that could arise from different threats, vulnerabilities, and consequences.[120] This can be followed by a risk management plan to direct and control an organisation with regard to risk and risk treatments to modify risk (in order to reduce/minimise them).[121] Below, possible threats, vulnerabilities, and consequences that could impact on the exploitation of project technologies are outlined prior to explaining how these relate to the risk assessment process.

5.1.1 Threats, vulnerabilities, and consequences

A **threat** is something that can cause harm.[122] In this analysis, it is persons or organisation who would use project technologies for misuse or mass surveillance. In this context, they include:

- Nefarious actors (e.g., Groups actively involved in harming people/groups directly or indirectly such as an organisation facilitating a government policy of apartheid, or LEAs over-policing a particular group as a matter of specific policy).
- Irresponsible end-users employing technologies for misuse or mass surveillance (e.g., an individual end-user using technologies for personal purposes such as stalking).
- Responsible end-users who allow access by undesirable end-users (e.g., LEA with technology transfer programme to organisations that operate under different standards for exceptional situations such as the military; university researchers providing technology transfer to LEAs with poor human rights record; end-user providing access to a trusted individual who then allows access to an irresponsible colleague).
- Well-intentioned end-users not thinking about unintended impacts of technology use, or are subject to external pressures (e.g., LEAs using COVID-tracking technologies for criminal investigations; researchers succumbing to political pressure to use/share technologies with a particular organisation; funded researchers with an obligation to exploit their results thinking that they should provide their technologies to anybody who is interested).
- Well-intentioned but naïve technology providers not adequately considering the impact of their actions (e.g., an exploitation partner not conducting proper due diligence before providing access to technologies).
- Hackers to the platform, and LEA systems (e.g., people gaining access to the technologies and platform due to inadequate safeguards and security measures).

The above threats lead to questions being asked of the end-user about who the potential end-users are, what they intend to do with the technologies, the situation in their country, how they control

[119] Financial Action Task Force (FATF), *National Money Laundering and Terrorist Financing Risk Assessment*, Paris, 2013, p.7. Available at: https://www.fatf-gafi.org/media/fatf/content/images/National_ML_TF_Risk_Assessment.pdf

[120] ISO, Guide 73:2009 Risk Management – Vocabulary, 2009. Available at: https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en

[121] ISO, Guide 73:2009 Risk Management – Vocabulary, 2009. Available at: https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en

[122] Financial Action Task Force (FATF), *National Money Laundering and Terrorist Financing Risk Assessment*, Paris, 2013, p.7. Available at: https://www.fatf-gafi.org/media/fatf/content/images/National_ML_TF_Risk_Assessment.pdf

access to their technologies, and their information security protocols. Further, questions can be asked of the technology provider themselves whether they wish to interact with end-users presenting higher levels of threat. The responses to these questions can then be used to determine if the proposed end-users are an appropriate type of organisation for the technology provider to do business with.

A **vulnerability** is something that can be exploited by the threat-actor, or facilitate their activities.[123] In this analysis, it is a weakness in the exploitation process whereby a threat-actor can gain greater access to project technologies than they should have. In this context, they can include exploitable or facilitative vulnerabilities.

Exploitable vulnerabilities:

- Limited ability to restrict exploitation of project technologies (e.g., the Model Grant Agreement requires project results to be made available to public authorities in the EU and this could include organisations with less respect for human rights standards; organisations from third countries might be entitled to participate in EC-funded projects and gain access to project results that might not otherwise be made available to them).
- Encouragement to exploit technologies from the EC without adequate guidance (e.g., pressure to exploit project results to fulfil a Grant Agreement obligation might be prioritised above the appropriateness of a particular exploitation route allowing a threat-actor to gain access to technologies).
- Lack of oversight in the end-user organisation (e.g., an LEA with poor oversight/management structures might enable nefarious individuals to misuse technologies without being discovered).
- Abandonment of formerly high-standards (e.g., an end-user might be subject to changes in oversight and governance that mean lower-standards are implemented providing space and opportunity for misuse and mass surveillance to take place).

Facilitative vulnerabilities:

- Unrestricted open-source access to technologies (e.g., making project technologies available open-source without restriction would allow threat-actors to gain access to technologies and use them for misuse or mass surveillance.)
- Poor consideration of end-user track-record (e.g., little research is done on how potential end-users operate, or it is assumed that all LEAs in Europe abide by high-standards when there are well known instances of some LEAs engaging in harmful practices in violation of human rights legislation).
- Absent or inadequate security processes by the end-user (e.g., nefarious individuals without authorisation could gain access to the technologies where they are not properly restricted).
- Inexperience in marketing technology in the security (or other relevant) domain (e.g., a technology provider might not be aware of the risks posed by providing technology in a new domain, or key issues they need to consider when conducting business in a different area).
- Lack of knowledge of potential negative or unwanted impacts of technologies that are intended for, or are used by, LEAs or other relevant organisations (e.g., a technology provider

[123] Financial Action Task Force (FATF), *National Money Laundering and Terrorist Financing Risk Assessment*, Paris, 2013, p.7. Available at: https://www.fatf-gafi.org/media/fatf/content/images/National_ML_TF_Risk_Assessment.pdf

is not aware of the potential misuses of their technology that could be performed by their customers).

- Lack of knowledge about how end-user operates, or how they use technologies.
- Release-and-forget problem (i.e., no awareness of technology use after exploitation).

The vulnerabilities that can allow risky actors greater access to tools are implicitly included in the below risk assessment. Vulnerabilities regarding internal processes at the potential end-user leads to questions about whether their internal processes provide adequate safeguards. The responses to these questions can be used to determine if the potential end-users can be trusted to treat and use the project technologies with the required level of respect.

With respect to safeguards, some discussion has taken place in the INSPECTr project about whether remote monitoring of technologies should be implemented so that technology providers can determine if their tools are being used in abuses and then 'pull the plug'. The technical possibility of remote monitoring has not been developed in INSPECTr. It is still an open question whether this is something that should be researched by technology developers. Remote monitoring and alternatives, such as removal of licenses, are explored in more detail below.

A **consequence** is the impact or harm resulting from the activities that the risk assessment seeks to avoid.[124] In this situation, consequences cover undesirable uses of the technologies, primarily those that are unlawful, unethical, or socially unacceptable. In this context, they can include both those for the technology provider and for the persons subjected to the technologies:

Consequences for the technology provider:

- Reputational damage (e.g., stakeholders discover that project technologies are used for undesirable purposes causing a negative impact on how they view the technology provider and the project).
- Scrutiny regarding funding (e.g., where technologies are funded by the taxpayer, increased scrutiny about whether public money should be used to create such technologies might take place with the potential for future policy aims to be moved away from developing technologies that have been used in unethical ways.)[125]
- Alienation of staff (e.g., where staff begin to realise the negative impacts of the technologies they work on and force institutional change or leave the organisation taking important institutional and technological knowledge).[126]

Consequences for persons subjected to technologies:

- Being a victim of their data being used for undesirable purposes (e.g., a nefarious end-user uses the technologies to acquire data, or insights, to profit from them).
- Being a victim of illegitimate targeted surveillance (e.g., a nefarious end-user deploys the technologies to acquire information on, and insights about, innocent individuals for purposes such as stalking, to facilitate blackmail, or voyeurism).

[124] Financial Action Task Force (FATF), *National Money Laundering and Terrorist Financing Risk Assessment*, Paris, 2013, p.7. Available at: https://www.fatf-gafi.org/media/fatf/content/images/National_ML_TF_Risk_Assessment.pdf

[125] Note, for example, the controversy around the iBorderCtrl project. See, Breyer, Patrick, "EU-funded technology violates fundamental rights", about:intel, 2021. Available at: https://aboutintel.eu/transparency-lawsuit-iborderctrl/

[126] Griffith, Erin, "Google Won't Renew Controversial Pentagon AI Project", Wired, 2018. Available at: https://www.wired.com/story/google-wont-renew-controversial-pentagon-ai-project/

- Being a victim of mass surveillance (e.g., a technology is used for large-scale indiscriminate monitoring of citizen without specific reason or justification).

The consequences can be understood as things that should be avoided both for the technology provider, but also for the persons subjected to the technologies. These points lead to questions for the technology provider themselves regarding their risk appetite. Often when people think about consequences such as people being subjected to illegitimate surveillance, they do not consider themselves as potential victims. However, it is important that technology providers personalise the risk as something applicable to them, especially when there is a risk of mass surveillance.[127] Thus, questions about the attitude of the technology provider to their own technologies are also included.

5.2 Questionnaires and risk assessment

Having identified potential threats, vulnerabilities, and consequences that could impact on exploitation of project technologies, the analysis of risks is now explained. In this process, questions are asked of potential customers, and of the technology provider themselves in order to understand and analyse the risks of selling or transferring technologies to prospective end-users. So that the link between the responses offered by prospective end-users and the evaluation of this by technology providers in the risk assessment is clear, both are explained together for each group of questions. As the analysis of each potential end-user must be done on a case-by-case basis, it is not possible for all aspects of risks regarding future users of the INSPECTr technologies to be understood at this stage. As such, suggested values for risk severity, likelihood, and the impact of safeguards are not given and the technology provider should judge this intrinsically (i.e., by instinct) during their assessment. It could be possible to evaluate several case studies to generate a more standardised assessment of risk, but that would take up a disproportionate amount of time considering the other work that needs to be completed in the project. Adding more quantitative elements to the analysis could be conducted in a future project.

The question areas below are key indicators for evaluating the risks of misuse, mass surveillance, and human rights abuses. The intention is that the technology provider would complete a risk indicator table for each of the different question areas. The severity, likelihood, and safeguards for each risk are scored out of 10 and are assessed cumulatively. Severity and likelihood are assessed positively (i.e., the more severe and likely risks have a higher score), and safeguards are assessed negatively (i.e., better safeguards have a lower score). This would then provide a score for each risk that can be compiled at the end of the assessment to give an overall risk score.

Where evaluation of risks uncovers a definitive 'no-no', this should be considered a supra-maximal risk. The assessment would end at this point and exploitation would not proceed. For example, if it is discovered that a person requesting technologies, or their country, are subject to sanctions that would make it unlawful to provide the technologies to them, then this would be a supra-maximal risk and exploitation should not take place.

It is important that the questionnaire gives possible end-users the opportunity to explain their activities and organisation so that a fuller view of the potential customer is presented to the technology provider. It also saves potential time and effort for a technology provider who might

[127] Personalising potential privacy harms to fundamentally change the focus of discussion is a useful approach, as demonstrated by Sheila Colclasure who testified to presenting highly-sensitive modelling data of individual company leaders who then agreed that such data was too sensitive to sell. See Transcript of the May 25, 2016 NCVHS Subcommittee on Privacy, Confidentiality & Security Hearing of: "De-Identification and the Health Insurance Portability and Accountability Act (HIPAA)". Available at: https://ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-may-25-2016-ncvhs-subcommittee-on-privacy-confidentiality-security-hearing/

otherwise need to conduct in-depth research on their own into specific organisations prior to providing them with access to technologies. Allowing end-users to author answers does create a risk that they will present their organisation in a more positive light than it deserves, but this process is complemented by additional research by the technology provider to answer questions directed at themselves. It is important that both these elements are present so that the subjective approach taken by end-users can be understood alongside a (hopefully) more objective view of their practices from the outside.

The questionnaire is intended to get end-users to provide an insight into their organisation, their processes, and their activities. This information can then inform a technology provider about whether they are a type of organisation they want to be involved with. Further, it might be possible to corroborate or test some information with results from additional research to determine the accuracy of what the end-user is saying, as dishonesty on behalf of the end-user represents an unethical practice and is a clear issue of concern for exploitation.

The approach of gathering information from end-users implicitly asks the question of '*if you have done nothing wrong, you have nothing to hide.*' This is usually asked in terms of surveillance over citizens, and assumes that privacy and security are somehow tradeable.[128] In the situation of providing AI technologies to potentially risky end-users, it assumes that confidentiality and reputation are somehow tradable. However, both of these are false cost/benefit analyses. The concepts do not need to conflict, but can exist in parallel. The question is also fundamentally flawed as it presents privacy or confidentiality as only being relevant where a person/organisation wishes to conceal wrongful behaviour and offers powerlessness in the face of intrusion as a solution.[129] However, privacy is multi-faceted and people have many legitimate reasons why they might wish to keep something private or confidential.[130] In the case of LEAs, common and legitimate reasons include ensuring the security and confidentiality of their ongoing investigations and covert techniques. It is also important to note that regulatory bodies for LEAs already exist, and technology providers do not need to replace them. But, it is still important for technology providers to recognise that their moral responsibility does not end when they finalise a sale or technology transfer and actively take steps to continue their moral engagement with their products and clientele.

5.3 Questions to potential customers and their assessment by technology providers

In this section, the numbered questions should be answered by potential customers. These questions could be provided as part of an application form when prospective end-users want to demonstrate their intent to purchase a product, or after they have already engaged in a discussion about procuring the technology with the provider. In the discussion for each group of questions, additional questions that technology providers can ask themselves are also included. The purpose of these additional questions is to encourage additional research to complement the responses of the prospective end-users, and to ensure they are correct. Further, the additional questions encourage self-reflection by the technology provider to ensure that they have considered many different issues regarding potential uses of the technology.

[128] Moore, Adam D., *Privacy Rights: Moral and Legal Foundations*, Penn State Press, 2011, p.204

[129] Solove, Daniel J., ""I've Got Nothing to Hide" and Other Misunderstandings of Privacy", San Diego Law Review, Vol.44, 2007, p.745, pp.764, 766

[130] Koops, Bert-Jaap and Maša Galič, "Unity in Privacy Diversity: A Kaleidoscope View on Privacy Definitions", South Carolina Law Review, Vol.73, 2022 (forthcoming)

5.3.1 Organisations and their location

1. *What type of organisation do you represent? LEA, private security, university, research organisation, etc.?*
    a. *Please give a brief summary of your organisation, its history and what it does. Please mention if your organisation is one arm of a larger corporate group, or similar.*

A response explaining the organisation and its history is useful as it can help determine what the organisational culture might be like. Researchers are likely to present low risks as it is most likely they would use the technologies to contribute to scientific knowledge by engaging in comparison testing but could engage in other riskier activities, for example. An LEA would present a higher risk as the technologies are likely to be used on/against people. A private security company would present an even higher risk as the technologies are likely to be used on/against people and they might not have the same level of scrutiny as an LEA who are a public authority. Due to the civil focus of the project, a military or paramilitary organisation is likely to present a supra-maximal risk and technologies should not be provided to such organisations; though there could be conceivable exceptions for UN-mandated peacekeepers, for example.

Using any details provided on company history, which could be augmented by research on the company website or a national register of companies if available and relevant, the technology provider can begin to understand whether the organisation has acted in acceptable ways. The technology provider can ask themselves '*Has this customer recently changed names or only recently been constituted as a company?*' This allows the provider to determine if they really know who they are dealing with.

It is also important to ask '*If the organisation has done unethical things in their past, have they done enough to distance, or redeem, themselves from this?*' For example, many medical institutions explored concepts of eugenics during and prior to the early 20th century but have acknowledged that this was wrong to do,[131] thus demonstrating that they are no longer promoting unethical practices. Further, it is important to understand if the organisation has collaborated with others for unethical purposes, whether for business purposes where decision-makers ignored ethical concerns or they were specifically nefarious. For example, IBM helped produce computing machines used specifically for the Holocaust[132] but have made clear efforts toward fulfilling requirements of corporate social responsibility. Where a clear effort has been made, this can reduce the likelihood of a risk manifesting. However, a lack of evidence of organisational change could increase the likelihood.

Whilst the focus here is on use of technology, how the prospective end-user operates in other ways is also relevant as merely being linked to an unethical organisation can be enough to damage the reputation of other organisations. For example, Amazon has surveilled its own employees[133] and attempted to discredit staff, and fired them, for engaging in legitimate organising for fair conditions.[134] As such, a technology provider should ask themselves '*Is this the type of organisation that we want to be involved with?*' If the answer is no, then the technologies should not be provided to them.

[131] See, for example, UCL "UCL makes formal public apology for its history and legacy of eugenics", UCL, 2021. Available at: https://www.ucl.ac.uk/news/2021/jan/ucl-makes-formal-public-apology-its-history-and-legacy-eugenics

[132] Black, Edwin, *IBM and the Holocaust*, Dialog Press, Washington D.C., 2001.

[133] Gurley, Lauren Kaori, "Secret Amazon Reports Expose the Company's Surveillance of Labor and Environmental Groups" Motherboard, 2020. Available at: https://www.vice.com/en/article/5dp3yn/amazon-leaked-reports-expose-spying-warehouse-workers-labor-union-environmental-groups-social-movements

[134] Blest, Paul, "Leaked Amazon Memo Details Plan to Smear Fired Warehouse Organizer: 'He's Not Smart or Articulate'", Vice, 2022. Available at: https://www.vice.com/en/article/5dm8bx/leaked-amazon-memo-details-plan-to-smear-fired-warehouse-organizer-hes-not-smart-or-articulate

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g. Private security company that is a new arm of an existing company that has engaged in various private security operations for 20 years.*	
Potential risk impact and effects	*e.g., Use of technology to identify known criminals in public space, which could lead to discrimination and over policing of persons from ethnic minorities who are more prevalent in database due to history of institutional racism in the criminal justice system. Parent company was involved in a previous instance of discriminating against persons from ethnic minorities in a similar situation 5 years ago.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., Could lead to preventing persons in the database from moving freely through the monitored space*	9/10
Likelihood of risk	*e.g., Persons in the database are expected to be monitored most days*	9/10
Safeguarding measures	*e.g., 3 years ago a new policy to deal with discrimination issues for the whole company group, but there is limited evidence of its effectiveness yet.*	4/10
Overall risk	*e.g., This is a new arm of a company that has made efforts to deal with previous history of discrimination.*	22/30
Actions/ recommendation	*e.g., Ask for more information on the enforcement of the discrimination policy.*	

2. *What country is your organisation based in? What countries do different arms of your organisation or corporate group operate in? What country will the technology be used in?*
    a. *Are any of the countries subject to any international sanctions?*
    b. *If the technology is subject to export controls, have you made contact with a national contact point?*
3. *What locations in the country does your organisation expect to use the technologies? Are there any particular cities/areas you intend to use the technology at?*

This facilitates the technology provider understanding where their technologies could end up and what standards, and level of respect for them, might be in place. A more detailed evaluation of standards is carried out below, but this question is useful for understanding the organisational attitude toward standards. Asking about different arms of an organisation operating in different countries is important as keeping potentially unethical practices at a distance from the main company can be done on purpose to 'draw the corporate veil' and shield the parent company from moral or legal responsibility.[135]

Generally, a high risk would be allocated to countries:

- With a poor human-rights records, as this demonstrates that the technology could be used in such human rights abuses.
- That are engaged in armed conflict, as armed conflict has different standards that are not appropriate for civilian technologies and the technology could be requisitioned for military activities.
- That are non-democratic, demonstrate authoritarian tendencies, or exercise oppressive practices, as this suggest the technology could be used for oppression, whether that is *ad hoc* or widespread and systematic.
- That have disputed territory, as this could lead to the technology being used to target persons supporting a breakaway territory.
- That are heavily securitised (e.g., Kashmir, Palestine, Western Sahara), as the technology could be used as part of a discriminatory security apparatus.
- That are impacted by domination, as the technology could be used as part of domination by one group over another.
- That have very lax regulation or are highly corrupt, such as tax havens, as there is likely to be limited ability for effective regulation to be enforced.

In writing this question, TRI did consider whether to include an outright question to prospective end-users on whether the countries in question were engaged in conflict, are disputed, otherwise securitised, subjected to forms of oppression, or are otherwise impacted by domination. However, it is possible that this could put the respondent in a difficult situation. For example, if such a question were asked during the 2022 Russian invasion of Ukraine it could be politically difficult for a Russian organisation to acknowledge that their country is engaged in an armed conflict rather than a 'special military operation', as the conflict is referred to by the Russian government, due to political and peer pressure.[136] Further, citizens in an oppressive state are unlikely to be able to acknowledge the existence of oppression without incurring a response from the authorities.[137] Yet, the status of the country can be objectively assessed without the need to place respondents in a difficult situation, and so an assessment of a country's status can be performed by the technology provider as the risks can still be acknowledged and understood without a response from end-users. So, the technology provider should ask themselves if their technologies will be used in the types of countries highlighted above. If the answer is yes to any of these aspects, then this will present a high risk.

[135] See, for example, Mind The Gap, "Case study: Shell denying responsibility for Nigerian oil spills", Centre for Research on Multinational Corporations, 2020. Available at: https://www.mindthegap.ngo/harmful-strategies/avoiding-liability-through-judicial-strategies/shielding-parent-companies-from-liability/case-study-shell-denying-responsibility-for-nigerian-oil-spills/

[136] Al Jazeera, "Do not call Ukraine invasion a 'war', Russia tells media, schools", 2022. Available at: https://www.aljazeera.com/news/2022/3/2/do-not-call-ukraine-invasion-a-war-russia-tells-media-schools

[137] See, for example, Wang, Maya, "China's campaign of intimidation against human-rights lawyers has to be stopped", Human Rights Watch, 2015. Available: https://www.hrw.org/news/2015/07/28/chinas-campaign-intimidation-against-human-rights-lawyers-has-be-stopped

Where a country is engaged in armed conflict and there are active hostilities in the same territory as the expected technology deployment, this would present a supra-maximal risk due to the heightened chance that the technology could be requisitioned for a war effort. Further, if the destination country is subject to high levels of oppression, then this would be a supra-maximal risk as there is a high likelihood that the technology will be used for political purposes.

The sub-question on sanctions is also important, a technology provider might not be aware of sanctions if they were initiated a long time ago, for example. If there are sanctions in place that would impact on the provision of the technologies, then this should be considered a supra-maximal risk where it could involve the technology provider breaking the law. The sub-question on export controls is also important as it is imperative that both parties take the export control regime seriously. An end-user that has a good relationship with their national contact point for export controls would present a lower risk that end-users who are not aware of, or actively avoid, their export control authority.

In order to evaluate the status of a country, an in-depth analysis by experts on the different aspects mentioned here would be ideal. But that is unlikely to be cost-effective. Instead, the technology provider could consult various resources to get an idea of the status of a country. This could include:

- For sanctions: https://www.sanctionsmap.eu/
- For conflict status: https://ucdp.uu.se/?id=1&id=1
- For the status of democracy: https://freedomhouse.org/countries/freedom-world/scores; Internet freedom could also be considered as a major indicator of democratic status,[138] see https://freedomhouse.org/explore-the-map?type=fotn&year=2021
- For country stability: https://fragilestatesindex.org/
- For tax havens: https://www.consilium.europa.eu/en/policies/eu-list-of-non-cooperative-jurisdictions/
- For the perception of corruption: https://www.transparency.org/en/cpi/2020/
- For human rights: country information and Universal Periodic Reports published by the UN human rights bodies (https://www.ohchr.org/en/countries; https://www.ohchr.org/en/hr-bodies/upr/documentation), country or thematic reports by Human Rights Watch (https://www.hrw.org/publications), Amnesty International (https://www.amnesty.org.uk/issues), or other reputable human rights organisations. Some NGOs and civil society organisations focus on specific issues, often because they have a particular agenda and so their recommendations and writings should be assessed critically.

It is important to note that where a country's human rights record is compared to others as part of determining an appropriate risk severity or likelihood, this must be done in with the appropriate context to give an accurate understanding. One could compare the number of cases brought at the European Court of Human Rights against different countries to give a relative understanding of risks surrounding human rights violations, but this is unlikely to be a useful comparison. For example, the European Court has ruled against Turkey almost 10 times more than they have against Slovenia.[139] This does not necessarily mean that Turkey is 10 times worse than Slovenia in human rights terms. Several different circumstances might have an impact: a larger population can have more incidents requiring legal remedy; the legal system of some countries might facilitate, or force, appeals to

[138] Stoycheff, Elizabeth, G. Scott Burgess, and Maria Clara Martucci, "Online censorship and digital surveillance: the relationship between suppression technologies and democratization across countries" Information, Communication & Society, Vol.23, Issue 4, pp.474-490.

[139] European Court of Human Rights, "Violations by article and by State', 2021. Available at: https://www.echr.coe.int/Documents/Stats_violation_1959_2021_ENG.pdf.

Strasbourg more frequently than others; or many cases might relate to particular situations. For example, security situations in Turkey and Russia have contributed to a large number of cases against them, and cases regarding length of judicial proceedings make up almost two thirds of Italy's cases.[140] It is also important to consider that lost cases can lead to legal reform. For example, the old UK bulk surveillance framework was judged not to be in compliance with the rights to a private life or freedom of expression, but this has since been replaced with a new framework.[141] As such, when evaluating a country's human rights record, it can be much more complex than it first appears and so should be evaluated on a case-by-case basis.

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g., The arm of the company using the technologies will operate in Morocco, but the parent company is based in Spain.*	
Potential risk impact and effects	*e.g., Risk that by operating technologies in Morocco, lower ethical and legal standards could be applied.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., Morocco disputes territories (Western Sahara) and has foreign enclaves that create political tensions (Melilla and Cueta, which are part of Spain). However, there have not been any recent armed clashes over any disputed or tension-creating territories. There are no sanctions on Morocco. Morocco ranks as 'Partly free' in terms of democracy. In terms of political stability, Morocco is at a 'warning' state.*	8/10
Likelihood of risk	*e.g., The technology is expected to be used in Marrakesh and Tangier which are away from the disputed territories.*	3/10
Safeguarding measures	*e.g., The legal office of the organisation has a good relationship with their national contact point regarding export controls, but no contact has been made for this issue yet.*	5/10
Overall risk	*e.g., Generally, Morocco has a concerning record on human rights and democracy. However, the disputed territories seem generally peaceful currently. Use of the technologies could be less concerning if this is away from disputed/tension creating territories.*	17/30

[140] European Court of Human Rights, "Violations by article and by State', 2021. Available at: https://www.echr.coe.int/Documents/Stats_violation_1959_2021_ENG.pdf

[141] Case Of Big Brother Watch and Others V. The United Kingdom, Apps Nos. 58170/13, 62322/14 and 24960/15 (ECtHR, 25 May 2021).

Actions/ recommendation	*e.g., Try to acquire guarantees that the technologies will not be used in disputed or tension creating locations.*

5.3.2 Uses of technologies

4. *What do you intend to use the technologies for? How will they benefit your organisation? Could they be useful to your organisation in other ways?*
    a. *Has your organisation used similar technologies previously? What were they used for?*
    b. *What safeguards are there in your organisation to ensure that any re-purposing of technologies is appropriate and ethical?*

This allows the technology provider to understand whether their technologies will be used as they are intended, whether there are alternative uses that they might be unaware of. The risk of using a technology as intended is likely to be lower than if it is unintended. For example, a facial analytics technology that is intended to provide biometric validation for access to secure areas that is re-purposed for recognising people on a watch-list would likely present a greater impact on people and so result in a higher risk. However, it might be possible that these examples could be inverted, and an unintended use presents less impact and lower risk.

When analysing the responses to these questions, technology providers could ask themselves '*Is there any other purpose that this customer could use this technology for other than their stated purpose?*' A similar question is asked directly of the prospective end-user, but there is a chance that end-users could give a narrow or unimaginative response. This could be to purposefully shroud real risks, or through genuine ignorance. Either way, a poor understanding of unintended uses could be considered as raising the severity of risk in supplying the end-user. Linked with this, the technology provider could consider '*Has this customer used technologies in an irresponsible, nefarious, or unexpected way before?*' This is similar to the earlier question on whether the organisation has engaged in unethical practices in the past, but being specific to the types of technologies involved provides additional context and so is worth considering discretely. If the end-users have used similar technologies in unethical ways, and there is little evidence of real and effective safeguards being implemented to prevent the same from happening again, then this would raise the risk.

So that the concerns of the society where the technology will be used are considered, it is important that the technology provider ask themselves '*What will the people who the technology is used on think about it?*' In answering this question, it is insufficient to respond with the common suggestion that the technology will be opposed by criminals who are more likely to have their criminality detected. Societies might oppose the use of a particular technology for a variety of reasons, and so it is important that technology providers take steps to understand whether there is a general opposition to the use of the technology in a particular location or situation. Unless there is a very strong justification provided by the end-users, then a technology provider should not go against the general view of the society in question. To gain these insights, technology providers could seek out population surveys, interviews, or focus group results that deal with the technology in question, or at least similar technologies/use-cases.

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g., Technologies will be used for identifying known criminals, to monitor their activities so that criminality can be prevented and public security can be increased in Marrakesh and Tangier. But, the technology could be re-purposed to monitor other types of people. It would generally be acceptable to monitor criminal activities, but not to monitor and profile people based on personal characteristics.*	
Potential risk impact and effects	*e.g., Profiling of known criminals would impact on their ability to build a new life and redeem themselves to society.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., Due to previous instances of discrimination, the potential for discriminatory profiling to occur raises significant risks that people from protected groups will be negatively impacted.*	8/10
Likelihood of risk	*e.g., There is no specific indication that the technologies will be used for anything other than identifying known criminals, which the organisation has stated. The parent company has an anti-discrimination policy, but it is not clear how this would have corrected previous activities that resulted in discrimination.*	5/10
Safeguarding measures	*e.g., It is not clear how the anti-discrimination policy from the parent company would be operationalised for this use of the technology.*	8/10
Overall risk	*e.g., There is no indication that the technology will be used for purposes other than we have been told. However, other technology companies have been given the same promises that were broken and the impact of the anti-discrimination policy is not clear.*	21/30
Actions/ recommendation	*e.g., Ask for assurances on how the anti-discrimination policy will be implemented in this instance*	

5.3.3 Regulation of technologies

5. *What is the legal framework that will regulate use of the technologies by your organisation?*
6. *Are there any legal judgement applicable to your country, especially human rights cases, relevant to your intended use of the technologies?*
7. *Has your organisation ever been subject to a legal case, or regulatory action, that is relevant to your intended use of the technologies? What are the details of the case?*

These questions allow the technology provider to understand the level of legal regulation applicable to the end-user and whether they have fallen afoul of it previously. Questions on legal cases are important as it demonstrates if, and how, the legal framework is enforced. Specific consideration of remedial actions taken against the end-users themselves are also important as it indicates the level of respect for the law in the institutional culture. The questions here clearly link to the earlier consideration of a country's human rights record and the focus here is on specific legal cases that impact on use of the technology by the end-user. Nevertheless, the assessment of relevant human rights aspects might need to be done across both questions as needed.

It is important to note, however, that the number of legal cases brought in relation to a legal framework, including where judges have ruled against the framework, need to be considered in context. A strong legal framework could have developed through case law and so a large number of cases could be a good thing. Conversely, a legal framework in legislation that has been shown to be in violation of human rights legislation in many cases would not be good.

Generally, a limited legal framework, or limited understanding of it on behalf of the end-users, a large number of legal rulings against the framework intended to regulate use of the technologies, and legal rulings against the end-user organisation regarding their compliance with the legal framework would indicate a higher risk. A well understood legal framework with limited case law against it, or against the end-users, would generally result in a lower risk as long as the legal framework has been in place long enough to be legally challenged. In assessing the legal framework, the technology provider should ask themselves '*Is this legal framework effective in preventing and punishing uses of technologies in misuse, mass surveillance, or other undesirable ways?*'

8. *What is the governance structure of your organisation?*
9. *What is the oversight/compliance structure of your organisation?*
    a. *Do you have an ethics board (or similar) that would oversee the use of the technologies?*
    b. *Is it independent? How is this guaranteed?*
    c. *What are its powers?*
    d. *Is the advice offered by the ethics board (or similar) followed? How are concerns of the ethics board generally taken into account?*

These questions allow the technology provider to understand the amount and quality of oversight in the end-user organisation. Generally, the presence of an ethics board, and following their advice would present a lower risk. As there are companies that engage in ethically compliant practices without an ethics board, it is difficult to say that the absence of an ethics board would lead to a higher risk. Therefore, the technology provider will need to determine if the absence of an ethics board will create a higher risk. For the INSPECTr project, it is unlikely to increase risk as ethics boards are only present at a small number of LEAs.[142] However, the prevalence and the associate risk of not having an LEA ethics board might change in future.

In line with these concerns, a technology provider should ask themselves is '*Are the governance structures in this organisation effective at preventing individuals from engaging in misuse, mass surveillance, or other undesirable activities?*' It is also important to determine whether the concerns and advice of the ethics board are properly considered, or if the existence of the board is an exercise

[142] West Midlands Police and Crime Commissioner, "Ethics Committee", 2022. Available at: https://www.westmidlands-pcc.gov.uk/ethics-committee/

in ethics washing, i.e., paying lip service to ethical standards whilst not actually doing anything to change their actions.[143] The presence of ethics washing implies an aspect of dishonesty as there is clear deceit in pretending to take ethics seriously, so this raises the potential risk and an effective ethics board would lower the risk.

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g., The use of technologies in Marrakesh and Tangier will be subject to the legal framework of Morocco, and the parent company will be subject to the legal framework of Spain. The Moroccan legal framework is not at the European standard. It seems that the company's legal office has a good understanding of both legal frameworks.* *The prospective end-users have a corporate board that makes key decisions, this includes major ethical considerations but there is no ethics board and no specific corporate officer with responsibility for ethics.*	
Potential risk impact and effects	*e.g., The use of technologies in a country with lower human rights standards does pose concerns about outsourcing and off-shoring of legal risk.* *The lack of specific ethical expertise in the company is concerning when there are many ethical issues to consider regarding deployment of the technology in the suggested situation.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., There is a high likelihood that Moroccan law will be complied with, but limited expectation that the higher standards from the Spanish legal framework applicable to the parent company will be applied. There is a low expectation that a strong ethical assessment will be conducted. Therefore, there is a risk that practices that would be unacceptable in Europe could be conducted.*	8/10
Likelihood of risk	*e.g., There is nothing to suggest that unacceptable practices are planned by the company. However, there are limited safeguards to prevent this happening in future, and limited organisational expertise to identify such occurrences.*	5/10

[143] Mintz, Steven, "What is Ethics washing?", Work Place Ethics Advice, 2021. Available at: https://www.workplaceethicsadvice.com/2021/04/what-is-ethics-washing.html#:~:text=Ethics%20Washing%20refers%20to%20the,it%20is%20occurring%20in%20practice.

Safeguarding measures	*e.g., Ethical considerations are made at the corporate board level, but it is not a specialist ethics board and so the effectiveness at ensuring compliance with ethical standards is expected to be limited.*	7/10
Overall risk	*e.g., There are no intentions to do anything that would be unacceptable or unlawful in Europe. However, the company does not appear to have developed safeguards that would prevent acting below European standards.*	20/30
Actions/ recommendation	*e.g., Ask about whether the company considers ensuring that it's corporate arms comply with Spanish law, and whether they intend to create any ethical compliance process.*	

5.3.4 Onward provision of technologies

10. *Does your organisation engage in technology transfer? i.e., if technologies from the project are provided to your organisation, are there processes to share these technologies with another organisation, or another arm of your organisation?*
    a. *Does your technology transfer programme provide technologies to military, paramilitary, or private security organisations?*
    b. *What safeguards does your technology transfer programme have?*
    c. *Does your organisation retain any responsibility for technologies you transfer to other organisations? How does this work?*

These questions allow the technology provider to understand if they need to consider risks that their technology will be used by one end-user or if it could be used by many more whose access to the technology they will not control. As the INSPECTr project is focussed entirely on civil applications, a risk of technologies being transferred to military, paramilitary, or private security organisations could represent a supra-maximal risk. For other technology providers, this could be an acceptable risk, but the expectation that the technologies could likely contribute to a military or security operation with lethal effects needs to be considered.

An end-user might have a series of safeguards to prevent irresponsible technology transfer, in which case this would lower the risk associated with providing the technologies. Further, the retention of ongoing responsibility would present a lower risk if it is accompanied with some level of control over the use of technologies. For example, technology results from INSPECTr might be provided to researchers who then adapt the technologies but add a monitoring or license removal capability as discussed below.

When considering the risk of onward transfer, technology providers should ask themselves '*Do I trust this company not to pass the technology on to another party who might have low standards?*' Where there is limited trust, this should increase the risk and *vice versa*.

Risk indicator table	
Consideration	**Details**

Risk/ failure mode	*e.g., The company does share technologies within the corporate group, but does not transfer technologies to other companies.*	
Potential risk impact and effects	*e.g., Whilst the company does not have corporate arms in states with very low legal/ethical standards, the company is expanding into other countries which could mean that another corporate arm is place in a country with low standards.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., If the company replicates it's Moroccan activities in a country with lower standards then there are risks that breaches of European standards could take place.*	8/10
Likelihood of risk	*e.g., There are no known plans to expand the company to a jurisdiction with very low standards. This seems unlikely as the company is primarily focussed on operations in Southern Europe, with the Moroccan activities seemingly being an exceptional case.*	1/10
Safeguarding measures	*e.g., Corporate expansion needs to be approved by shareholders. Ethical concerns could be explained at a shareholder meeting.*	2/10
Overall risk	*e.g., Expansion to a very low regulation jurisdiction is possible and this could result in the technology being used in such a place. However, there are no concrete plans to do so yet.*	11/30
Actions/ recommendation	*e.g., Request plans for corporate expansion if available, and suggest that an assessment of human rights standards is commissioned by the board when determining possible locations for expansion.*	

5.3.5 Information security and data protection

11. *What information security measures are implemented by your organisation? How will you ensure that nefarious actors do not gain access to the technologies?*
12. *What data protection measures are implemented by your organisation? How will data processed by the technologies be protected?*

These questions allow the technology provider to understand whether their technologies, and any associated data accompanying, or part of,[144] the technology will be kept safe. Fundamentally, there is little point in engaging in a process such as this if a technology provider is willing to give technologies to organisations with poor cyber security and data protection practices; a nefarious organisation whose custom is rejected could simply steal the technologies from an organisation who do receive the

[144] For more on the retention of personal data in machine learning algorithms as 'algorithmic shadows', see Li, Tiffany C., "Algorithmic Destruction", *SME Law Review* (2022) (forthcoming).

technologies. Further, it is important that the information security and data protection measures taken are not just sufficient for the intended uses by the end-users, but that they also can deal with uses that might not initially be intended or expected by the provider. For example, an end-user could take a face verification technology intended to be used for regulating access to secure areas and repurpose it for facial recognition to monitor the activities of certain people. These activities process different data types of different sensitivities and so the safeguards taken would need to be appropriate for other uses of the technologies that are reasonably possible.

The technology provider should ask themselves '*Do I trust the end-users to keep the technology and data safe?*' Self-evidently, strong information security and data protection measures would lower risks of providing the technologies to the end-user in question. Poor safeguards would increase the risk.

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g., The parent company has a good standard of information security and data protection. The same standards will be applied in the Moroccan arm of the organisation. However, additional infrastructure is more exposed, and the quality of staff training could be lower, when it is away from the corporate centre due to less oversight capability.*	
Potential risk impact and effects	*e.g., Where infrastructure is more exposed to cyber-attack in Morocco, and training quality is lower, this could result in lower standards than those applicable to the parent company. Morocco does have a data protection law and a data protection authority.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., A successful cyber-attack, or failure to uphold data protection standards could result in highly sensitive biometric information about known criminals being exposed.*	9/10
Likelihood of risk	*e.g., Assuming that corporate standards from the parent company are adequately applied, there is a low chance of this risk materialising.*	2/10
Safeguarding measures	*e.g., As the parent company has good data protection and information security expertise, it is likely that good advice could be sought in case any issues are raised by the Moroccan arm of the organisation.*	1/10
Overall risk	*e.g., Generally, the risk of a failure of information security or data protection are higher for the Moroccan arm than the Spanish parent company, but support should be available from the parent company in any case.*	12/30

Actions/ recommendation	*e.g., Gain assurances about processes for staff working for the Moroccan arm of the company to get support from the Spanish parent company.*

5.3.6 Misuse and mass surveillance

13. *What steps are taken in the country that the technologies will be used (and the country of the parent organisation, if different) in to prevent mass surveillance? Are they effective?*
14. *What steps are taken in your organisation to avoid engaging in mass surveillance? Are they effective?*
15. *What steps are taken in your organisation to avoid misuse of technologies (e.g., use of technologies for personal or criminal purposes)? You can note internal discipline policies here. Are they effective?*
16. *What steps are taken in your organisation to prevent biased effects of technologies from impacting on different parts of the population? Are they effective?*
17. *What remedial action could someone who is a victim of misuse or mass surveillance, or are otherwise wronged by your use of the technology take against your organisation/country? Are they effective?*

These questions allow the technology provider to understand the nature of safeguards taken by the prospective end-user to prevent misuse, mass surveillance, discrimination, and violations of human rights linked with these activities. The reference to internal discipline gives an opportunity for end-users to demonstrate how they will deal with wrongful use of the technologies. However, the final question here relates to how the end-users might respond to wrongdoing if a victim comes forward. Legal responsibility and ethical accountability are key aspects of responsible use of technology, and so it is important that end-users demonstrate that they have considered how they can provide redress for wrongs that might have been carried out as a company or by individuals within the company.

Where end-users have clear policies or can demonstrate having plans to prevent misuse or mass surveillance that are actionable, then this would lower risks considerably. Where such policies or plans are not completed or actionable then the potential for this to lower risks is minimal. If policies or plans are non-existent, then this should be treated neutrally as the absence of documentation does not mean that people will act unethically. Similarly, the presence of plans and policies does not guarantee ethical behaviour, but does make it more likely and so presents and increased chance of lowering the risk. Of course, if end-users respond to these questions in such a way that indicates they do not acknowledge any risks of misuse or mass surveillance, or are reckless toward then, then this should increase the risk. If the responses indicate a level of contempt for safeguards by suggesting that safeguards are a waste of time, for example, then the technology provider could consider this to be a supra-maximal risk because there will be limited ability for a technology provider to be able to trust that the end-user will act ethically. In line with this thinking, the technology provider can ask themselves '*Are the safeguards, plans and policies to prevent misuse and mass surveillance any good?*'

Risk indicator table	
Consideration	**Details**

Risk/ failure mode	*e.g., Significant human rights abuses have taken place in Morocco. However, the situation appears to have improved slightly in recent years. Still, there is an extensive legal framework that can facilitate surveillance by the state.*[145] *Surveillance tools, including NSO's Pegasus, have been used to invade private spaces and intimidate activists.*[146] *Some violations of the right to privacy have been ignored or rejected by the Moroccan authorities, despite clear evidence.*[147]	
Potential risk impact and effects	*e.g., By providing technologies that can be used for surveillance purposes to Morocco, there is a risk that the technologies could be requisitioned by the state, or the data processed by the technologies could be accessed by state agents. As such, there is a risk that the technologies could contribute to a system of mass surveillance.*	
Consideration	**Details**	**Score**
Severity of risk	*e.g., There is a clear risk that any surveillance technology, or the data from it, could be collected by Moroccan authorities as part of systematic surveillance. This could lead to intimidation, punishment, and arbitrary detention.*	10/10
Likelihood of risk	*e.g., There have been several instances of clear and serious violations of privacy rights, and there is no evidence that the Moroccan state is going to alter it's practices any time soon.*	10/10
Safeguarding measures	*e.g., As a technology provider we could only issue licenses for one year at a time, which could be removed where there is a reasonable suspicion that the technologies, or the data processed by them, is being used for mass surveillance or misuse. However, if the surveillance activities are classed as a state secret, we are unlikely to be informed about such activities. Therefore, the ability to implemented effective safeguards is hampered.*	10/10
Overall risk	*e.g., There is a clear risk of mass surveillance and misuse in Morocco. It is not clear that wrongdoers are effectively*	30/30

[145] Privacy International, "State of Privacy in Morocco", 2019. Available at: https://privacyinternational.org/state-privacy/1007/state-privacy-morocco

[146] Stangler, Cole, and Abdellatif El Hamamouchi, "Morocco's Surveillance Machine" The Intercept, 2021. Available at: https://theintercept.com/2021/10/21/morocco-pegasus-surveillance-journalists/

[147] Privacy International, "State of Privacy in Morocco", 2019. Available at: https://privacyinternational.org/state-privacy/1007/state-privacy-morocco

	punished. There is limited ability to implement effective safeguards.	
Actions/ recommendation	*e.g., Look at other options for safeguards, or else do not continue with sale.*	

5.3.7 External monitoring

Optional question:

18. *Would your organisation be willing to accept remote monitoring of your use of the technologies?*

For this question to be useful the technical capability first needs to exist, and have been chosen for implementation ahead of less invasive measures such as licensing reviews. It is important to note that, despite the below concerns, Art.61 of the EU's proposed AI Regulation includes post-market monitoring. It is not yet clear what this will involve, how this could be done in a proportionate way, and whether this will make it into the final version of the AI Regulation. As such, this question might need to become mandatory for high-risk AI systems if this article is included in the AI regulation, and it might need to be adapted if the proposed article changes before the Regulation is finalised.

Assuming monitoring is implemented, a technology provider would need to be prepared to take on a monitoring role. If the provider is not prepared for this role, then this question should not be asked, as the monitoring capability would be worthless and would not need to be included in the technology itself. If the question is asked, then an end-user's willingness to accept external monitoring would lower their risk.

Even with a technology provider willing to monitor use of the technology, remote monitoring is not an ethical panacea. Monitoring would allow technology providers to protect their reputation and people whom the technologies are used against. However, it also provides a potentially intrusive window into an environment of potentially sensitive data processing. A more easily amenable procedure for monitoring uses of technology by customers might be to repeat the below risk assessment after a particular period or time, or when a specific situation appears, to determine if risks have risen and a license should be removed, thereby preventing the platform from working. This would enable technology providers to protect their reputation and possible victims of abuse without needing a potentially intrusive insight into the use of the technologies.

Here, technology providers can ask themselves '*Is the external monitoring capability sufficient to report all reasonably expected acts of misuse and mass surveillance?*' Where it is not, then this would mean the efficacy of the safeguarding measure is reduced.

Risk indicator table	
Consideration	**Details**
Risk/ failure mode	*e.g., Use of technologies for misuse or mass surveillance*
Potential risk impact and effects	*e.g., Use of the technologies to invade people's privacy*

Consideration	**Details**	**Score**
Severity of risk	*e.g., If the tool is used for misuse or mass surveillance the impact would likely be significant*	9/10
Likelihood of risk	*e.g., There is a record of poor compliance with Morocco's human rights legislation by the state, so there is a reasonable likelihood*	6/10
Safeguarding measures	*e.g., There is remote monitoring of technology access, meaning that any external attempts to acquire data would be notified to us and we could halt the license if we are concerned enough to do so. We could not remove the technology once it is in the possession of the end-users.*	2/10
Overall risk	*e.g., There is a significant risk of misuse and mass surveillance, but this can be monitored and assessed. Evidence of state agents interfering with technologies can be made available to investigators.*	17/30
Actions/ recommendation	*e.g., Ensure responsibilities for monitoring activities are clear before providing technology.*	

5.3.8 Political oversight and outside influence

19. [for LEAs] *What is the structure for political oversight of LEAs in your country?*
20. [for universities/researchers] *Is your organisation subject to any political control?*
    a. *How is the independence of your organisation ensured?*

These questions are differentiated between LEAs and universities/researchers as they need to be aimed differently. Both are intended to determine whether supplying technologies to the prospective end-users could lead to the end-users being 'overruled' by government, meaning that their responses to questions in this risk assessment would be worthless where a government exerts external control. Of course, some level of political direction is to be expected: Ministries of Interior in most democratic countries can still set high-level policy for policing, for example, but the risk here is political influence on investigations and use of the technologies. A politician with the ability to set a policy aim for police to target organised crime groups is not likely to increase risk of misuse and surveillance in and of itself, but the ability to set a policy aim to investigation and discredit opposition politicians for the purposes of damaging their election prospects would be a significant risk.

Further, it is important for technology providers to determine whether supplying the end-users could lead to support for government initiatives that they might not wish to be associated with. For example, a technology provider might be willing to supply an independent organisation in a country with a poor human rights record, but might not be willing to do so where the organisation is effectively a state organ that is complicit in human rights abuses.

Political independence should be expected for end-users, so there are no safeguarding measures here. The level of risk should be increased depending on the level of political control, from a low risk where politicians have no ability to direct police investigations, to a very high risk where they have the ability to directly intervene in ongoing investigations.

21. [for universities and private organisations] *Where does the funding for your research or your organisation come from?*

   a. *Does any of this funding come with 'strings attached', if so, what are they?*

   b. *Does your organisation act as a contractor for a government? Do they carry out functions normally performed by public authorities (e.g., privatised policing)?*

This question allows the technology provider to understand whether there are any contractual or other influences over how the end-user might operate, or use their technology. This is particularly relevant in two ways. First, a university researcher might be given funding on the condition that government agents can access all aspects of research, including technologies or software used in the research. This would present a larger risk as there is greater uncertainty about how the technology will be used, and how.

Second, it is also relevant where the end-user acts on behalf of the government, potentially generating state responsibility. For example, a private security contractor who is provided with public powers to carry out a particular task.[148] This could potentially raise greater risks if a private contractor does not have the same level of safeguards normally associated with performing the public powers in question. However, it could also be possible that by acting as a contractor for the government that the end-user organisation is required to put additional requirements in place to be eligible to bid for government contracts.[149] As safeguards are not included at this point of the assessment, the existence of any requirements at organisational level should be considered in the likelihood of risk manifesting. On this point specifically, is would be appropriate for a technology provider to ask themselves '*Is this organisation working on behalf of a government that has used similar technologies in unethical ways before?*' If the answer is yes, then this would contribute to a high risk.

Overall, for these questions, technology providers should ask themselves '*Do I trust that the organisation we will be providing technologies to will be the organisation deciding on how to use them?*' and, if the answer is yes, '*Have external controls over the organisation been used for unethical purposes in the past?*' The responses to these questions would allow an understanding of who the real decision-makers are for technology use.

Risk indicator table		
Consideration	**Details**	
Risk/ failure mode	*e.g., The funding for the Moroccan activities includes money from the Moroccan state.*	
Potential risk impact and effects	*e.g., There is a risk that the Moroccan state might expect access to the technology in exchange for funding.*	
Consideration	**Details**	**Score**

[148] Art.5 and associated commentary, International Law Commission, 'Report of The International Law Commission on the Work of its Fifty-Third session' (2001) UN Doc. A/CN.4/SER.A/2001/Add.1 (Part 2), p.31

[149] For example, s.6, Human Rights Act 1998 makes it unlawful for any public authority to act in contravention of human rights, and public 'acts' includes procurement and purchasing of services.

Severity of risk	*e.g., State access to the technology would pose serious risks to human rights.*	9/10
Likelihood of risk	*e.g., It is not clear whether state agents have exerted influence to impact on uses of technology previously. In any case, the Spanish parent company has a well enforced policy on rejecting outsider control, including leaving the business activity.*	3/10
Overall risk	*e.g., There is a potential risk of state influence, but no concrete evidence of this. Further, there are policies to deal with attempted outside influence.*	12/20
Actions/ recommendation	*e.g., Ask how policies will be enforced in Moroccan context.*	

5.3.9 Bringing the risk home

This is the final stage of the risk assessment. As noted above, where people are evaluating risks for others they might have a higher risk appetite than if the risks are going to impact on themselves. Therefore, it is appropriate for the technology provider to ask questions relevant to them individually as this helps contextualise the other responses. This stage should also be used as an opportunity to evaluate the risks overall. In terms of the overall risk score, it has not been possible in INSPECTr to explore enough case studies so that response ranges could accurately be given to help contextualise the result. Generally, a good rule of thumb would be to any overall risk scoring above 50% (i.e., a score of 135 if every question is asked) should be subject to serious consideration before technologies are provided to end-users.

First, the technology provider should consider '*How do our staff feel about our technologies being used in the way proposed by the end-user?*' Generally, organisations should not do things that are opposed by their staff. It is not appropriate to ask staff to perform tasks that they are opposed to, especially where they have an opposition for ethical reasons that arise from their conscience. Further, *in extremis*, staff could protest, or resign, about a particular business decision which would be deleterious from a business reputation perspective.

Second, the technology provider should consider whether they would accept being subject to the technology in the proposed way. This should be considered by the highest-level decision-maker that is appropriate. They should ask themselves '*How would you feel if you, or your data, were subjected to analysis by the prospective end-user using your technology?*' If the answer is negative, then the technology should not be provided to the end-user. Where the answer is conditional (i.e., '*I would feel okay about it if X safeguard was in place and it was for Y reason*'), then the technologies should only be supplied where those conditions are fulfilled. It is not ethical to allow others to be subjected to something that we would not accept ourselves, so it is important that we take steps to ensure the technologies are used in an acceptable way.

Risk indicator table		
Consideration	**Details**	**Score**
Personal risk understanding	*e.g., Our staff are opposed to providing the technology to the prospective end-user due to the high risk that our*	10/10

	technology will be used in mass surveillance activities. I would not accept my data being processed by the technology where the is a high-risk of state interference.	
Overall risk	*e.g., There is a high-risk that the technology could be misuse or abused in Morocco. There are safeguards promised by the Spanish parent company but they do not seem to be sufficient considering the risks.*	182/270
Final assessment	*e.g., Do not provide the technology to the prospective end-user.*	

5.4 Conclusion

Having explained the nature of risks associated with providing technologies intended for LEA investigations, and acknowledged the threats, vulnerabilities, and consequences of providing such technologies to undesirable organisations, this document has outlined a risk assessment that technology providers can use to understand the risks associated with providing technologies to a prospective end-user. The example used concludes in the risks being too great, and so the technology would not be provided. However, where risks are otherwise higher than the technology provider would like, it might be possible to manage those risks through various means such as additional safeguards. For example, a technology provider could write certain safeguards, such as review of technology use by an independent ethics board, into their contracts with end-users. How risks can be managed to mitigate or reduce them for different possible end-users across different domains could be considered in future research. Still, the above risk assessment methodology would provide an initial step to reducing risks of misuse, mass surveillance, and human rights violations by prospective end-users if it is implemented by the technology providers from the project.

6 Conclusion

Overall, this document has explored many ethical, legal, and societal issues relevant to the research and development of technologies like INSPECTr, and has considered how these issues can be mitigated or dealt with. These mitigations can be applied during the process of technological design (as in Section 2 and 4 exploring issues in the design process), before technologies are provided to end-users (as in Section 5), or whilst the technologies are in use (as in Section 3).

Building on the work in *D8.7: Privacy and Ethics-by-design in the INSPECTr platform* that explored how relevant design strategies were applied in the INSPECTr technologies, this document demonstrates how the concepts can be adapted for the LEA context. In doing so, it begins to re-develop these approaches for LEA technologies, specifically. More work can be done in future to develop specific LEA-orientated Privacy and Ethics-by-Design patterns tailored to the LEA context, and to move beyond those applicable to commercial technologies.

As demonstrated in Section 3, this process can learn from other areas of regulation that are already embedded in LEA processes. Firearms are the example used here, and there are certainly important results that can influence how LEA uses of AI systems are governed. Other areas where LEA use artefacts in potentially dangerous ways could be explored in future to determine if they have useful strategies that could be applied to AI systems, driving LEA vehicles in pursuit of fleeing offenders could be one avenue.

A key requirement for LEA systems for data analytics is maintaining compliance with the rules of criminal procedure, especially in terms of data lifecycles. Section 4 provides guidance on this, though is hampered by the significant variance provided for states in the high-level requirements arising from human rights law and the Law Enforcement Directive, and what individual states and agencies decide to apply. This work develops a platform from which further research can be done on individual requirements in different countries in future projects.

Of course, spending significant resources on developing approaches to technology development that are privacy-respecting and ethics-compliant is not beneficial if the end-users decide to engage in malicious activities regardless. Section 5 provides a significant step toward being able to identify and avoid the risks that some of these malicious activities could manifest. The risk assessment could be adapted for different use-cases, types of technology provider, or recipient. Indeed, it is not just related to LEA technologies, but could be used for others.

In summary, this document has taken both a broad and narrow view of Privacy and Ethics-by-Design as they can be applied to LEA technologies: issues are approached on the conceptual level and on the level of technological details, and issues beyond the technology design process but within the overall circumstances of providing LEA technologies (such as exploitation) are also considered. This is a useful contribution to adapting and re-considering the processes of Privacy and Ethics-by-Design so that further work can be done to develop tailored approached for LEA technology development.

7 Annex 1: Blank 'Know Your Customer' Exploitation Risk Assessment Tables

Here, we provide a blank risk assessment table that could be used to evaluate exploitation risks.

Organisation

Questions to end-users:

1. What type of organisation do you represent? LEA, private security, university, research organisation, etc.?
    a. Please give a brief summary of your organisation, its history and what it does. Please mention if your organisation is one arm of a larger corporate group, or similar.

Questions to the technology provider:

Has this customer recently changed names or only recently been constituted as a company?

If the organisation has done unethical things in their past, have they done enough to distance, or redeem, themselves from this?

Is this the type of organisation that we want to be involved with?

Consideration	**Details**
Risk/ failure mode	
Potential risk impact and effects	

Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Location

Questions to end-users:

1. What country is your organisation based in? And what countries do different arms of your organisation or corporate group operate in?

a. Are any of the countries subject to any international sanctions?

b. If the technology is subject to export controls, have you made contact with a national contact point?

2. What locations in the country does your organisation expect to use the technologies? Are there any particular cities/areas you intend to use the technology at?

Questions to the technology provider:

Will the technology be used in a country:

- With a poor human-rights records;
- Engaged in armed conflict;
- That is non-democratic;
- That has disputed territory;
- That is heavily securitised;
- That is impacted by domination;
- That has very lax regulation or are highly corrupt?

Consideration	**Details**
Risk/ failure mode	
Potential risk impact and effects	

Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Use of technologies

Questions to end-users:

3. What do you intend to use the technologies for? How will they benefit your organisation? Could they be useful to your organisation in other ways?

a. Has your organisation used similar technologies previously? What were they used for?

   b. What safeguards are there in your organisation to ensure that any re-purposing of technologies is appropriate and ethical?

Questions to the technology provider:

Is there any other purpose that this customer could use this technology for other than the one they claim they will use it for?

Has this customer used technologies in an irresponsible, nefarious, or unintended way before?

What will the people who the technology is used on think about it?

Consideration	**Details**
Risk/ failure mode	
Potential risk impact and effects	

Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Regulation of technologies

Questions to end-users:

4. What is the legal framework that will regulate use of the technologies by your organisation?
5. Are there any legal judgement applicable to your country, especially human rights cases, relevant to your intended use of the technologies?
6. Has your organisation ever been subject to a legal case, or regulatory action, that is relevant to your intended use of the technologies? What are the details of the case?
7. What is the governance structure of your organisation?
8. What is the oversight/compliance structure of your organisation?
   a. Do you have an ethics board (or similar) that would oversee the use of the technologies?
   b. Is it independent? How is this guaranteed?
   c. What are its powers?
   d. Is the advice offered by the ethics board (or similar) followed? How are concerns of the ethics board generally taken into account?

Questions to the technology provider:

Is this legal framework effective in preventing and punishing uses of our technologies in misuse, mass surveillance, or other undesirable ways?

Are the governance structures in this organisation to prevent individuals from engaging in misuse, mass surveillance, or other undesirable activities?

Consideration	**Details**	
Risk/ failure mode		
Potential risk impact and effects		
Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Onward provision of technologies

Questions to end-users:

9. Does your organisation engage in technology transfer? i.e., if technologies from the project are provided to your organisation, are there processes to share these technologies with another organisation, or another arm of your organisation?
    a. Does your technology transfer programme provide technologies to military, paramilitary, or private security organisations?
    b. What safeguards does your technology transfer programme have?
    c. Does your organisation retain any responsibility for technologies you transfer to other organisations? How does this work?

Questions to the technology provider:

Do I trust this company not to pass the technology on to another party who might have low standards?

Consideration	**Details**

Risk/ failure mode		
Potential risk impact and effects		
Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Information security and data protection

Questions to end-users:

10. What information security measures are implemented by your organisation? How will you ensure that nefarious actors do not gain access to the technologies?
11. What data protection measures are implemented by your organisation? How will data processed by the technologies be protected?

Questions to the technology provider:

Do I trust the end-users to keep the technology and data safe?

Consideration	**Details**	
Risk/ failure mode		
Potential risk impact and effects		
Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Misuse and mass surveillance

Questions to end-users:

12. What steps are taken in the country that the technologies will be used (and the country of the parent organisation, if different) in to prevent mass surveillance? Are they effective?
13. What steps are taken in your organisation to avoid engaging in mass surveillance? Are they effective?
14. What steps are taken in your organisation to avoid misuse of technologies (e.g., use of technologies for personal or criminal purposes)? You can note internal discipline policies here. Are they effective?
15. What steps are taken in your organisation to prevent biased effects of technologies from impacting on different parts of the population? Are they effective?
16. What remedial action could someone who is a victim or misuse or mass surveillance, or are otherwise wrong by your use of the technology take against your organisation/country? Are they effective?

Questions to the technology provider:

Are the safeguards, plans and policies to prevent misuse and mass surveillance any good?

Consideration	**Details**
Risk/ failure mode	
Potential risk impact and effects	

Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

[optional] External Monitoring

Questions to end-users:

17. Would your organisation be willing to accept remote monitoring of your use of the technologies?

Questions to the technology provider:

Is the external monitoring capability sufficient to report all reasonably expected acts of misuse and mass surveillance?

Consideration	**Details**	
Risk/ failure mode		
Potential risk impact and effects		
Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Safeguarding measures		x/10
Overall risk		x/30
Actions/ recommendation		

Political oversight and outside influence

Questions to end-users:

18. [for LEAs] What is the structure for political oversight of LEAs in your country?
19. [for universities/researchers] Is your organisation subject to any political control?
    a. How is the independence of your organisation ensured?
20. [for universities and private organisations] Where does the funding for your organisation come from?
    a. Does any of this funding come with 'strings attached', if so, what are they?
    b. Does your organisation act as a contractor for a government? Do they carry out functions normally performed by public authorities (e.g., privatised policing)?

Questions to the technology provider:

Is this organisation working on behalf of a government that has used similar technologies in unethical ways before?

Do I trust that the organisation we will be providing technologies to will be the organisation deciding on how to use them? Have external controls over the organisation been used for unethical purposes in the past?

Consideration	**Details**
Risk/ failure mode	

Potential risk impact and effects		
Consideration	**Details**	**Score**
Severity of risk		x/10
Likelihood of risk		x/10
Overall risk		x/20
Actions/ recommendation		

Bringing the risk home		
How do our staff feel about our technologies being used in the way proposed by the end-user? How would you feel if you, or your data, were subjected to analysis by the prospective end-user using your technology?		
Consideration	**Details**	**Score**
Personal risk understanding		x/10
Overall risk		x/270
Final assessment		