Centre for Emerging Technology and Security

RESEARCH REPORT

Privacy Intrusion and National Security in the Age of AI

Assessing proportionality of automated analytics

Ardi Janjeva, Muffy Calder and Marion Oswald

May 2023

About CETaS

The Centre for Emerging Technology and Security (CETaS) is a research centre based at The Alan Turing Institute, the UK's national institute for data science and artificial intelligence. The Centre's mission is to inform UK security policy through evidence-based, interdisciplinary research on emerging technology issues. Connect with CETaS at cetas.turing.ac.uk.

This research was supported by The Alan Turing Institute's Defence and Security Programme. All views expressed in this report are those of the authors and do not necessarily represent the views of The Alan Turing Institute or any other organisation.

Acknowledgements

The authors are very grateful to Jonathan Hall KC, Daryl Burns, Rob West, James S and the anonymous reviewers for their valuable feedback on an earlier version of this report.

Executive Summary

This CETaS Research Report explores the complex issue of privacy intrusion arising from the use of automated analytics, with specific focus on artificial intelligence (AI). The research focuses on UK national security and law enforcement agencies with access to legal powers that incur some degree of intrusion into individuals' private lives. As automated methods are increasingly deployed to process the data collected through the use of such powers, there is a need to understand the additional privacy considerations that could arise as a result of this automated processing.

The report's ultimate objective is to develop a structured analytical framework for assessing proportionality of privacy intrusion arising from the use of automated analytics. The framework considers the whole lifecycle of automated analytics, including data collection, training, testing processes, and use. It aims to introduce a common language and taxonomy that will assist stakeholders in identifying, comparing, and assessing the potential impact of relevant privacy considerations in a structured and evidence-based way. This framework is not intended to replace any existing authorisation or compliance processes, but rather to provide an additional layer of rigour and assurance to supplement and futureproof existing processes.

The research is informed by semi-structured interviews and focus groups with stakeholders across the UK government, national security and law enforcement, and legal experts outside government, as well as an understanding of the literature on proportionality in English law and critiques of the application of the proportionality test. Particular attention is paid to the distinctive aspects of automated processes and artificial intelligence, and the requirements for making a structured analytical framework useful in practice.

The structured framework presented in Section 3 consists of questions that probe six factors that are relevant when developing or assessing proportionality arguments related to the use of automated analytics:

- Datasets
- Results
- Human inspection
- Tool design
- Data management
- Timelines and resources

As discussed in the report, different subsets of questions are applicable depending on the process stage, scenario, and aspects of the proportionality test being considered. Practitioners are encouraged to adopt the framework as a guide to support with the application of the legal proportionality test, in cases where automated techniques are being deployed on previously collected data.

Additional findings and recommendations are as follows:

1) There is a need to better understand, map and monitor the cumulative intrusion risk of multiple, connected, automated systems feeding into each other over an extended period.

2) There should be specific consideration of the potential of developing automated systems which may incur a degree of privacy intrusion in the short run, but reduce privacy intrusion in the long run. In this instance, a way of measuring the potential *future reduction* in privacy intrusion would need to be developed.

3) The ubiquity of big data analytics and automated systems in modern society means that shifting public expectations of privacy need to be understood in a more rigorous and representative manner, to promote transparency and public trust.

The proportionate and effective use of automated analytics is central to the ability of the national security and law enforcement community to operate in a rapidly changing technological environment. This is recognised by recently published and ongoing reviews of investigatory powers in the UK. The framework and supporting recommendations proposed in this report will contribute to enabling organisations deploying automated analytics to assess proportionality of privacy intrusion in a more systematic and empirical way, while ensuring appropriate steps are taken to minimise privacy intrusion throughout the analytic lifecycle.

Introduction

The context

National security and law enforcement agencies are tasked with protecting the country's people and institutions from the most severe forms of harm. Meeting this obligation entails a degree of monitoring and surveillance, which in turn incurs a degree of intrusion into some people's private lives. In a liberal, democratic society, there is a responsibility on the state to reduce this intrusion to the minimum level required to keep people safe.

The statutory functions of the UK intelligence agencies are set out in the Security Service Act 1989 and Intelligence Services Act 1994, which restrict the agencies' power to obtain and disclose information to that which is necessary for their functions.[1] Although directed and intrusive surveillance and the use of covert human intelligence sources continue to be governed by the Regulation of Investigatory Powers Act 2000 (RIPA), the framework governing most digital investigatory powers is now laid out in the Investigatory Powers Act 2016 (IPA), under the supervision of the Investigatory Powers Commissioner's Office (IPCO).[2] The Act specifies that the techniques used by national security and law enforcement agencies for surveillance purposes must be necessary in a democratic society and proportionate in the interests of national security, economic wellbeing or the prevention and detection of serious crime.

Each instance of intrusion is therefore regulated by these two criteria of *necessity* and *proportionality*. However, the question that national security and law enforcement agencies increasingly face is the extent to which the context of automated analytics – often augmented by artificial intelligence (AI) – changes the nature of the necessity and proportionality assessment and the way in which these criteria are interpreted.

The principle of proportionality guides not only the security community but most individuals during daily life. Adjusting our behaviour in accordance with the risk or reward we perceive in a given situation is an instinct intrinsically connected to what we judge to be fair or just. However, the stakes in the national security context are different. Matters of life and death

[1] Alexander Babuta, Marion Oswald and Ardi Janjeva, "Artificial Intelligence and UK National Security," *RUSI Occasional Paper* (April 2020): 20.

[2] Further information: https://www.legislation.gov.uk/ukpga/2016/25/contents/enacted.

are constantly in play, and the capacity to intrude on individuals' lives is often far greater. High stakes call for correspondingly high standards of clarity and accountability.

Previous research has emphasised the additional privacy and human rights considerations that need to be assessed when AI is used in the national security context.[3] These come from two main sources: the requirement to collect training data to develop an effective AI tool, and the inherent uncertainty of probabilistic AI outputs. In a world where emerging technologies are making the work of national security and law enforcement agencies simultaneously easier and harder, meeting the highest standards of rigorous assessment is an increasingly complex task, yet also essential if the security community is to maintain its license to operate in the eyes of the public.

This report addresses the challenge of assessing proportionality of privacy intrusion of automated analytics, with specific focus on AI. It aims to complement ongoing reviews of investigative powers and governance of AI more broadly, including Lord Anderson's independent review of the IPA[4] (still ongoing at the time this report was written), the UK Government's consultation on its recent AI regulation proposals,[5] and the House of Commons Science and Technology Committee's inquiry into the governance of AI.[6] These reviews represent an important milestone to re-assess the operational, legal and ethical considerations raised by increased use of automated analytics for intelligence work. This report's contribution is to further develop best practice for assessing the different dimensions of proportionality in this changing operational environment.

Purpose and objective

This report considers the matter of privacy intrusion during the lifecycle of automated analytics tools. This lifecycle is illustrated in Figure 1 below.

The lifecycle diagram is intended to depict the different stages involved in automated analytics – to better understand where privacy considerations could arise. As outlined below, these stages include input data being *collected* and *prepared,* followed by

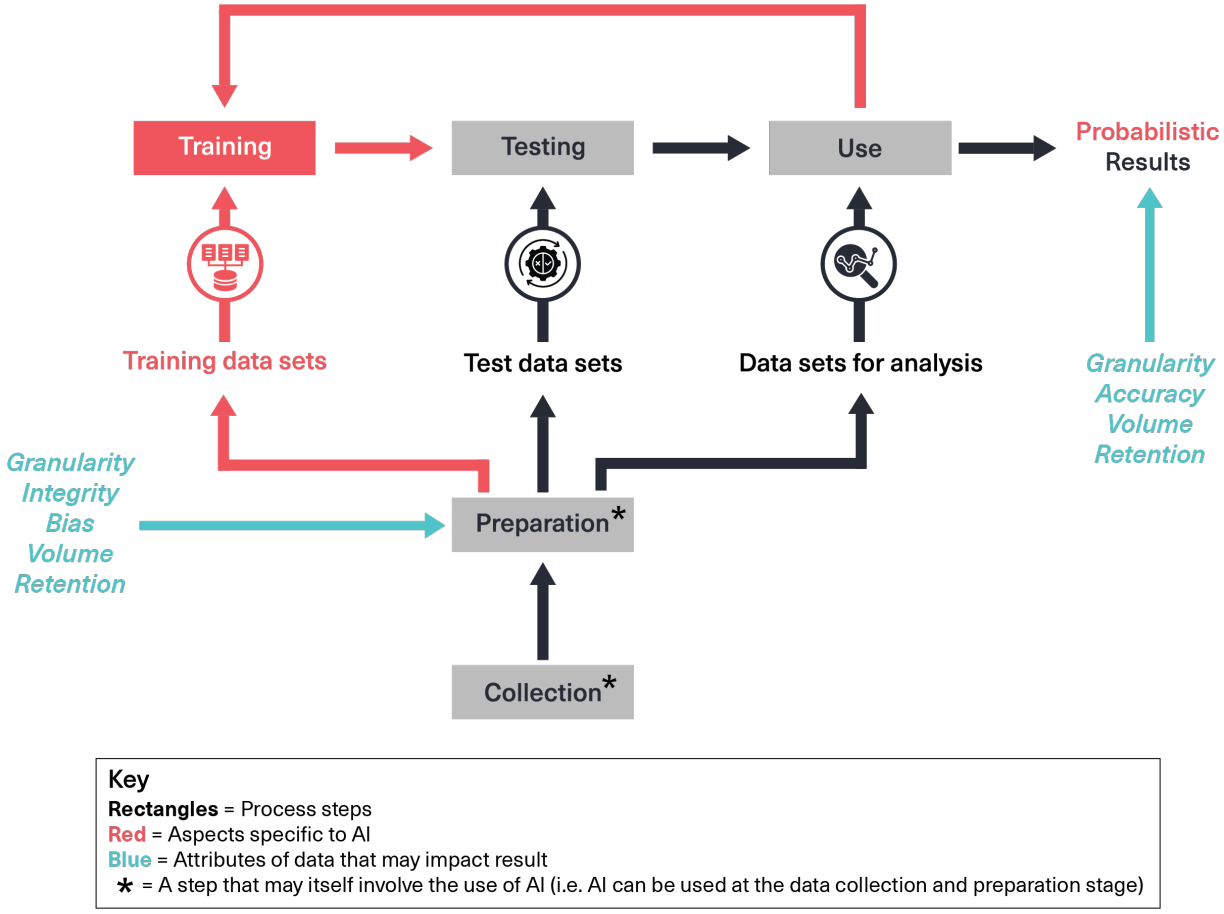
[3] Alexander Babuta et al., "Artificial Intelligence and UK National Security," *RUSI Occasional Paper* (April 2020).
[4] Further information: https://www.gov.uk/government/news/lord-anderson-appointed-to-review-the-investigatory-powers-act.
[5] Further information: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.
[6] Further information: https://committees.parliament.uk/committee/135/science-and-technology-committee/news/173701/mps-to-examine-regulating-ai-in-new-inquiry/.

subsequent *training*, *testing* and ultimately *operational deployment* of the analytic – which produces certain *results* or *outputs.*

There are three caveats to note when viewing the diagram:

- **Training** and **testing** steps are separated due to the possibility of employing pre-trained, third-party models. Non-machine learning automated systems will not typically involve a training stage.
- **Collection** and **preparation** may include querying third-party data at rest and/or also involve AI.
- Data may be **analysed** for the purpose of identifying other relevant data sources, which are then targeted for **subsequent collection**. This feedback loop is not indicated in Figure 1.

Figure 1: Automated analytics lifecycle. Note that a wide range of input data may be involved at the 'collection' stage, and there may be significant variation in the level of sensitivity of this data.



This report's main objectives are twofold: i) to develop a structured analytical framework that allows practitioners to systematically assess the relevant proportionality factors at each

stage of the lifecycle of an automated tool, and ii) to introduce a common language and taxonomy for discussing proportionality, both within the national security and law enforcement community, and between this community and the public.

The framework presented in Section 3 is intended to *supplement* and *futureproof* existing approaches to proportionality assessment, providing an additional layer of assurance that relevant proportionality factors are considered at each stage of the analytic lifecycle. It is not intended to replace existing authorisation processes within the national security and law enforcement community. The framework is a tool for enabling richer evaluation of whether there are less intrusive measures available to achieve the same intelligence aim, and whether there is a fair balance being struck between individual rights and the interests of the community at large.

The intended user community for this framework is illustrated in Figure 2 below.

Figure 2: Intended user community.

These stakeholders will be found across a wide range of public authorities with access to IPA powers. Proportionality considerations are also central to the activities of defence (including the Ministry of Defence), law enforcement agencies (including police forces, the National Crime Agency, and Border Force), HM Revenue and Customs, and a range of Wider Public Authorities and Local Authorities. The framework proposed in this report has been developed with the diversity of this community in mind. Nonetheless, much of the context and literature is understood through the lens of the national security and law enforcement communities, and it is these stakeholders to which the findings and recommendations predominantly refer.

Additionally, it is important to note that not all these authorities have equal access to the full range of warrants and authorisations, and departments may vary widely both in their technological maturity and their data access requirements.

While IPCO is responsible for approving targeted interception (Part 2 IPA), targeted equipment interference (Part 5 IPA), bulk warrants (Part 6 IPA), and bulk personal dataset warrants (Part 7 IPA), The Office for Communications Data Authorisations (OCDA) is responsible for approving communications data warrants (Part 3 IPA). Similar proportionality considerations apply for these as to other forms of intrusive powers. This report therefore refers to 'oversight body/bodies' to collectively encompass both IPCO and/or OCDA.

Definitions and scope

The following definitions of data science terminology are referred to throughout the report.

- **Automated analytics**: The use of algorithmic methods to analyse data and generate insights. This includes both fully automated systems, and systems that are not fully automated but entail some degree of human inspection. Not to be conflated with automated decision-making as defined in Article 22 of the Data Protection Act.
- **Artificial intelligence (AI)**: Use of machine learning (ML) and statistical models to enable computer systems to learn and improve through experience, thereby finding patterns, deriving insights, or making predictions. Excludes non-machine learning examples of AI.
- **Accuracy**: Degree of correctness of the data or prediction.
- **Data granularity**: Level of detail in the dataset.
- **Data integrity**: Overall completeness, consistency, and accuracy of the dataset.

The scope of the report is limited to the following:

1) Automated analytics (as defined above). Nonetheless, elements of the findings and the contents of the proportionality framework may also be useful for structured assessments of non-automated processes.

2) Proportionality of intrusion of *all forms of automated analytics* applied to *collected data*, whether AI-enabled or otherwise. As such, the research is assessing the proportionality of *analysis* processes, rather than *collection* (while noting that analysis may inform further collection).

3) *Privacy intrusion* as defined in Article 8 of the European Convention on Human Rights (ECHR), as opposed to other forms of intrusion or human rights concerns. The report's focus is proportionality in this context, rather than in relation to data

protection, although many similar concerns arise in respect of data protection principles.

Methodology

The findings presented in this report are based on a literature review and legal analysis focused on proportionality and human rights, and semi-structured interviews and focus groups conducted between September 2022 and January 2023 with 14 participants across government agencies and law enforcement, and non-government lawyers with relevant expertise. The findings also draw directly on the authors' recent work engaging with relevant agencies and oversight bodies to advise on the 'factors of intrusion' relating to the use of automated analytics, for instance through the IPCO Technology Advisory Panel (TAP) and its recent research into privacy intrusion metrics.[7]

A semi-structured interview format allowed a broadly consistent line of questioning across interviews, with certain areas expanded upon or omitted depending on the specific expertise and experience of individual participants. As well as standard interview questions, participants were invited to consider three scenarios developed by the research team which placed the participant in the position of a practitioner making necessity and proportionality judgements in a range of fictitious contexts.

Any references in this report to comments made by participants reflect their own opinion and should not be interpreted to represent the official position of any government department, agency or other organisation.

This study has some limitations. Firstly, it is partly based on a relatively small interview sample, and therefore has limited external validity. The specificity and context of the research questions under consideration – motivated by the desire to conclude the research with a usable practitioner framework – meant that current or prior experience in relevant practitioner roles was desirable (although not necessary) for a formal interview. This narrowed the pool of potential participants and meant that representatives from academia and those without practitioner experience comprised a smaller proportion of the interview sample. Follow-on research could capture these views more comprehensively.

[7] IPCO Technology Advisory Panel, "Metrics of Privacy Conference" (November 2018). Further information: https://www.ipco.org.uk/publications/technology-advisory-panel/page2/; IPCO Technology Advisory Panel, "Privacy Intrusion Metrics: concepts and considerations" (October 2021). Further information: https://www.ipco.org.uk/publications/technology-advisory-panel.

The remainder of this report is structured as follows. Section 1 provides an overview of the academic literature on proportionality in English law and considers some of the critiques of the application of the proportionality test, including through the lens of a particular case study. Section 2 analyses the proportionality considerations arising from automated analytics and is broken down into five sub-sections: assessing the nature of privacy intrusion; the implications of AI; the relevance of assurance, handling, and retention policies; the prospect of an ongoing approach to assessing cumulative intrusion; and the merits or otherwise of practitioner-focused versus metrics-based approaches to proportionality. Section 3 proposes a new structured analytical framework for assessing 'factors of proportionality' in the operational context. Section 4 concludes with a set of key findings and recommendations for stakeholders across relevant public sector organisations.

Section 1. Proportionality in English Law

Assessing the relevant academic literature and case law is central to understanding external perspectives of the assessment of proportionality, and the various human rights implications arising from the use surveillance techniques.

In the words of one interviewee, 'the law is an art not a science. If it were, lawyers would be redundant. But because it is an art, people will disagree.'[8] Many cases considering the proportionality of privacy intrusion involve interpretation, which can have profound implications for operational approaches.

As Murray points out, although national security and law enforcement agencies must apply key legal concepts in their operational and policy activities, our understanding of the law is based largely on cases and determinations of regulators after the event.[9] Meanwhile, in their paper on European Union counterterrorism law-making, De Londras and Tregidga draw attention to the interplay between proportionality as a legal concept and as an instrument operating in a dynamic policymaking environment.[10]

However, these scholars further highlight a risk that proportionality impact assessments conducted before the event 'might be shaped in a way to reinforce, rather than challenge, the starting assumptions that underpinned the initial policy formation.'[11] To the extent that this risk is borne out, it further highlights the importance of this report's proposed framework as a tool to challenge starting assumptions throughout the deployment of an automated tool.

While this section is focused on the interpretation of the proportionality test by the courts and related academic commentary, the aim of this report is to inform the processes of making proportionality decisions and the oversight of those decisions, specifically in relation to automated technology deployed for the purposes of national security.

[8] Interview with non-government lawyer, 14 November 2022.
[9] Daragh Murray, "Using Human Rights Law to Inform States' Decisions to Deploy AI," *AJIL Unbound* 114 (2020): 158-162.
[10] Fiona de Londras and Jasmin Tregidga, "Rights, proportionality, and process in EU counterterrorism lawmaking," *International Journal of Constitutional Law* 19, no. 2 (2021): 665–693.
[11] Ibid.

1.1 The legal proportionality test and critiques of its application

The human rights-based analysis in this report is conducted through the lens of Article 8 of the ECHR, the right to respect for private and family life, home and correspondence, and its implementation in English law.

The European Court of Human Rights (ECtHR) Guide usefully sets out how Article 8 can be invoked and how an alleged infringement will be assessed.[12] An applicant (the victim) must show that their complaint falls within the scope of Article 8. Then, the Court would examine whether there has been an interference with that right or whether the State's obligations to protect the right have been engaged. Article 8 is a qualified right, meaning the state may interfere with the enjoyment of this protected right in the interests of national security, public safety, or the economic wellbeing 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.

Limitations on Article 8 rights are permissible if they are 'in accordance with the law' or 'prescribed by law' and are 'necessary in a democratic society' for the protection of one of the above objectives.[13] In the assessment of the test of 'necessity in a democratic society', the Court often needs to balance the applicant's interests protected by Article 8 against a third party's interests protected by other provisions of the Convention and its Protocols.[14] The Court will further consider whether the intrusive measures were *proportionate* to the legitimate aims pursued.[15]

This section focuses upon the structured 'full proportionality analysis' approach to cases involving fundamental rights, formally established through the case of *Bank Mellat v HM Treasury*.[16] This test is used to review and assess the State's interference with ECHR rights

[12] European Court of Human Rights, "Guide on Article 8 of the European Convention on Human Rights," updated on 31 August 2022. Further information: https://www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf.

[13] In this report, we do not address the 'prescribed by law' and 'legitimate aim' elements of the test, instead focusing on 'necessary in a democratic society'. The ECtHR has clarified that "necessary" in this context does not equate to expressions such as "useful", "reasonable", or "desirable" but implies the existence of a "pressing social need" for the interference in question.

[14] European Court of Human Rights, "Guide on Article 8 of the European Convention on Human Rights," updated on 31 August 2022.

[15] *Z v. Finland*, 25 February 1997, Reports of Judgments and Decisions 1997-I; European Court of Human Rights, "Guide on Article 8 of the European Convention on Human Rights," updated on 31 August 2022.

[16] *Bank Mellat v HM Treasury (No 2)* [2013] UKSC 39.

and is generally regarded as a more 'intensive' review than the traditional *Wednesbury* judicial review test in English law.[17]

The 'necessity and proportionality' test in cases where the Human Rights Act 1998 is pleaded is laid out in four parts:

> (a) is the objective sufficiently important to justify limiting a fundamental right (a pressing social need)?
> (b) are the measures which have been designed to meet it rationally connected to it?
> (c) are the means used to impair the right or freedom no more than necessary to accomplish the objective? (Is a less intrusive measure available?)
> (d) does the measure strike a fair balance between the rights of the individual and the interests of the community?[18]

The proportionality test is described as structured because of the need for judges to work through the above four stages in their reasoning. Although there will still be disagreement on where the balance lies,[19] this structured approach ensures clear and defensible lines of reasoning and easier identification of points of disagreement, increasing the transparency of the overall decision-making process.[20]

Despite this, proportionality reasoning has been criticised by some as 'a highly discretionary exercise of judicial policymaking',[21] which is potentially susceptible to legally irrelevant preferences. For example, one study found that legal experts' prior policy preferences appeared to affect their propensity to judge whether an action was proportionate.[22]

It has been argued that rights protection can be increased by applying the proportionality analysis in terms of the necessity test (requiring that the least restrictive/intrusive means be chosen), rather than applying the 'strict' proportionality test (comparing potential positive effects with potential negative effects and weighing the benefit to the public against the harm to the individual). This is because the comparative element of the necessity test

---

[17] A.C.L. Davies and J.R. Williams, "Proportionality in English Law." In *The Judge and the Proportionate Use of Discretion: a Comparative Study*, eds. Sofia Ranchordás and Boudewijn de Waard (Routledge, 2016).
[18] *Bank Mellat v HM Treasury (No 2)* [2013] UKSC 39.
[19] See the disagreement between Lord Reed and Lord Sumption in Bank Mellat on the 3rd and 4th stages: Lord Reed at para 125/6 of the judgment Bank Mellat v Her Majesty's Treasury (No. 2) [2013] UKSC 39 (19 June 2013) (bailii.org).
[20] A.C.L. Davies and J.R. Williams, "Proportionality in English Law." In *The Judge and the Proportionate Use of Discretion: a Comparative Study*, eds. Sofia Ranchordás and Boudewijn de Waard (Routledge, 2016).
[21] Julian Rivers, "The presumption of proportionality," *Modern Law Review* 77, no. 3 (2014): 409-433.
[22] Ranaan Sulitzeanu-Kenan, Mordechai Kremnitzer and Sharon Alon, "Facts, Preferences, and Doctrine: An Empirical Analysis of Proportionality Judgment," *Law and Society Review* 50, no. 2 (2016): 348-382.

creates an enhanced awareness of potential alternatives and thereby reduces the likelihood of unjustified rights limitations.

Our interview data has shown that sometimes the term 'necessity' is used in relation to the question of 'whether there is a pressing social need' (the first part of the *Bank Mellat* test); and sometimes 'necessity' is a shortcut for the question of 'whether there is a less intrusive means available' (the third part of the test). Ensuring that both aspects are understood and addressed – both the *objective* and the *least intrusive means* – will be important in any application of the proportionality test.

Understanding the potential impact of practical applications of automated analytics is essential in weighing up utility and harm, and therefore to assessing proportionality.[23] But making this assessment is difficult in practice. First, it has been argued that proportionality is not an objective concept, as it is not possible to make a like-for-like comparison between individual rights and public policy goals.[24] Second, it may be impossible to assign precise values to rights and interests. However, the proportionality test does not depend on exact mathematical formulations, and values and weights can still be assigned without being precise or identical. This discussion is revisited in Section 3.1.

Interviewees also commented upon the challenges of interpreting and applying the concept of proportionality. Generally, proportionality was understood to mean that the intrusiveness of the proposed activity is balanced against the interests of others who might be affected, and against the operational advantage that the investigating party expects to gain.[25] If the level of intrusion or interference into the right to privacy *is* genuinely necessary in the circumstances, and *is* capable of objective justification, then the scales will tip in favour of the proposed activity being judged proportionate.[26] In practice however, complexities may emerge around the notion of 'objective justification'. Different activities may engage individuals' rights in different ways, making it difficult to account in advance for all possible contextual factors that may be relevant in each individual case.[27] In turn, this emphasises the acute challenge of assessing minimal levels of intrusion in some instances.

In practice, there are numerous overlapping, context-specific factors that need to be considered if assessing proportionality in relation to the four-stage test. Use of automated

[23] Daragh Murray, "Using Human Rights Law to Inform States' Decisions to Deploy AI," *AJIL Unbound* 114 (2020): 158-162.
[24] A.C.L. Davies and J.R. Williams, "Proportionality in English Law." In *The Judge and the Proportionate Use of Discretion: a Comparative Study*, eds. Sofia Ranchordás and Boudewijn de Waard (Routledge, 2016).
[25] Interview with government lawyer, 25 October 2022.
[26] Interview with government lawyer, 10 November 2022.
[27] Ibid.

analytics introduces additional considerations in this regard, and the purpose of our proposed framework is to provide a structured approach for assessing the most relevant factors which may arise in the national security context.

1.2 The impact of intrusion on the individual and on others

This section concludes by using an example of case law to examine the 'scaling' effects of intrusion, and the extent to which quantitative assessments are relevant to a legal human rights analysis.

Case Study: Metrics in the case of 'Bridges v The Chief Constable of South Wales Police [2020] EWCA Civ 1058'

This case concerned a complaint by Mr Bridges about the use of live facial recognition (LFR) technology by South Wales Police, and an appeal from the Divisional Court decision to the Court of Appeal. The Court of Appeal was concerned with the fourth stage of the proportionality test in this case (the 'fair balance' question).

The issue of the number of people impacted

Mr Bridges submitted that the Divisional Court had been wrong not to consider the impact of the LFR on other members of the public, when considering the 'cost' side of the balance. However, the court pointed out that the complaint made by Mr Bridges was in respect of the impact on him alone. The court agreed with the police that:

> '...the impact on each of the other members of the public who were in an analogous situation to this Appellant on the two occasions with which we are concerned for present purposes (in December 2017 and March 2018) was as negligible as the impact on the Appellant's Article 8 rights. An impact that has very little weight cannot become weightier simply because other people were also affected. It is not a question of simple multiplication. The balancing exercise which the principle of proportionality requires is not a mathematical one; it is an exercise which calls for judgement.' (Paragraph 143)

Machine intrusion and population level concerns

The Court in *Bridges* took the view that processing an unmatched face is a passive function of the LFR system, meaning an individual's privacy is only materially intruded upon if they are flagged by the system as a potential match and subsequently reviewed by a human officer.

However, it has been argued that this fails to account for the scaling effect of privacy intrusion – the fact that automated systems such as LFR essentially place large groups of the population under surveillance, rather than just individuals within a population. Individual-level privacy intrusion may appear negligible, but cumulatively may scale to a disproportionate level.[28] This remains a matter of open debate between academics, lawyers, and privacy campaigners.

These concerns reflect the individualistic and victim-focused approach to human rights claims (the question is whether an *individual's* rights have been infringed – regardless of the impact on wider groups of individuals). This contrasts with the singling out of bulk personal datasets (BPDs) in the IPA as deserving of particular consideration, due to the privacy intrusion of such datasets. By including large numbers of individuals unlikely to be of intelligence interest, BPDs inevitably entail a large degree of *collateral intrusion*.[29] On the other hand, the intrusion caused by the retention and examination of BPDs is often more static and predictable than for other IPA data types, where warrants authorise the continuous acquisition of data.[30]

Crucially, the amount of collateral intrusion that is deemed proportionate in any given scenario will be directly correlated to the risk of *not* identifying the right 'needle in the haystack', and the potential adverse outcomes that could result. For a high-priority threat to life investigation, the acceptable level of collateral intrusion will inevitably increase (particularly if there is no less intrusive means to acquire the information needed within the required timeframe).

Assessing collateral intrusion must involve consideration of the wider implications of using or acquiring a method or dataset, even if these reflect potential future complaints or concerns. Those judging the proportionality of an automated method need to have sufficient breadth of perspective, training, and awareness to anticipate the contextual circumstances

[28] Bernard Keenan, "Automatic Facial Recognition and the Intensification of Police Surveillance," *Modern Law Review* 84, no. 4 (2021): 886–897.
[29] IPA (s199(1)(b)).
[30] Home Office, *Report on the Operation of the Investigatory Powers Act 2016* (Home Office, 2023), 15.

and subsequent compound action that might give rise to adverse impacts, a complaint or public concern. The proportionality of the *method* cannot be assessed in isolation without considering the subsequent action (including by other organisations) that may result. We explore this further in the next section.

Section 2. Proportionality Considerations arising from Automated Analytics

Building on insights elicited from the primary research, this section analyses proportionality considerations arising from the use of automated analytics. This will lay the groundwork for the subsequent section, which proposes a new analytical framework for assessing 'factors of proportionality' relevant to automated analytics.

2.1 The nature of privacy intrusion

The *nature* of the privacy intrusion caused by an automated analytic technique was a central talking point throughout the research. Developing a holistic understanding of this requires possessing a contextual appreciation of four main areas: expectations of privacy; the distinctions between targeted and collateral intrusion; distinctions between human and machine intrusion; and the reconciling of competing human rights in the national security context. These aspects are each considered in turn.

2.1.1 Expectations of privacy

Interviewees suggested that the public's expectations of privacy are dynamic and responsive to a wide range of contextual factors. Most individuals will be aware of the trade-off they make when walking through a busy high street in the UK – a limited intrusion into their privacy courtesy of CCTV systems, in return for deterring those individuals who would seek to put their safety at risk. The expectations of privacy in this basic scenario – and the violation that an individual may feel – differ compared to a hypothetical scenario where an individual goes to a place of worship (where they may not expect to be surveilled) and find out that CCTV is being deployed.[31]

The nature of the data being collected is a central part of this conundrum. In the words of one interviewee, 'not all data is equal. Some is inherently deeply intrusive, and some far less so.'[32] Moreover, the IPA itself does not expressly apply the same safeguards to all material acquired under all Parts of the Act – 'the differing levels of intrusion associated with separate powers, and the way in which they are used in an operational context, explains

[31] Interview with law enforcement lawyer, 21 October 2022.
[32] Interview with government lawyer, 10 November 2022.

many differences in the level of applicable safeguards.'[33] One interviewee made a comparison between the electoral register and a dataset containing private medical records, emphasising the level of personal information contained in the latter would entail a much higher degree of privacy intrusion. Using the medical dataset for automated processing is therefore likely to be more intrusive from the outset, regardless of the type of analysis deployed.[34]

However, other examples will be more contested. For instance, communications data (sometimes called 'metadata') has sometimes been presented as a privacy-enhancing approach when compared to accessing the *content* of private records or conversations. Yet others argue that the extrapolations that can be made from the metadata portion of communications data (in combination with other techniques) can end up being just as intrusive as reviewing the content of communications.[35]

In the Home Office's recent IPA review, it was pointed out that communications data applications are likely to continue rising due to the continued importance of communications data in all types of criminal investigations. This stems from an improved understanding of the benefits communications data can bring to investigations, and a trend of digital investigations supplanting conventional surveillance as a more efficient investigative tool. Furthermore, emerging technologies such as vehicle telematics and the growth in Internet of Things (IoT) devices mean that communications data is available from an increasingly diverse range of sources.[36] Indeed, successful prosecutions have already incorporated smart watch, smart speaker, vehicle, and video doorbell data.[37]

As algorithms and advanced analytics become increasingly embedded in daily life, the argument for normalising their use in security and policing (with the appropriate safeguards in place) becomes stronger.[38] Some interviewees argued this may gradually shift expectations of privacy within a population, although this will likely vary depending on what a capability is used for, and how informed the public feels about those use cases. Public expectations of privacy should therefore be monitored on an ongoing basis to inform proportionality assessments made in relation to privacy intrusion.

[33] Home Office, *Report on the Operation of the Investigatory Powers Act 2016* (Home Office, 2023), 10.
[34] Interview with government lawyer, 10 November 2022.
[35] Ibid.
[36] Home Office, *Report on the Operation of the Investigatory Powers Act 2016* (Home Office, 2023), 10.
[37] Ibid, p.21.
[38] Interview with law enforcement lawyer, 21 October 2022; Interview with law enforcement lawyer, 10 November 2022.

2.1.2 Targeted intrusion and collateral intrusion

The distinction between targeted privacy intrusion upon individuals of intelligence interest, and collateral intrusion upon people not of intelligence interest, is important to judgements about proportionality of automated analytic techniques.

Some would argue that targeted intrusion demands greater scrutiny than collateral intrusion, which often involves *acquisition* but very minimal or no *processing* or *analysis* once data is deemed not to be of intelligence benefit. On this basis, the justification for subjecting one specific individual to the more intrusive applications of technology should be more strenuous compared to bulk data which may not be reviewed in the same detail (if at all).[39] This argument would be reinforced by existing handling arrangements[40] ensuring that data that stops being operationally relevant is discarded at the earliest opportunity (additionally, the discarding of unwanted or low-grade material is closely linked to system volume and cost considerations, as well as proportionality).[41]

In time-pressured scenarios, setting data retention thresholds becomes more challenging. New information about a terror group planning an imminent attack, for example, may justify the use of a technique with a relatively large number of potential errors (and therefore higher potential for collateral intrusion), that would not be justified in the context of a long-term investigation whose outcome was less time critical.[42] Moreover, in a fast-moving, high-threat investigation it may be deemed proportionate to deploy a range of techniques with varying degrees of intrusion, further emphasising the complexity in these scenarios.

Previous CETaS research has studied in detail the issue of 'discovery failure risk' and intelligence analysts' thresholds for false positives and negatives:

> 'The research found that false negatives are generally considered to be the costliest type of error in an intelligence context. Across all decision-making settings (whether ML-assisted or otherwise) analysts' risk appetite for any false negatives is very low. Furthermore, interviews revealed that the risk tolerance for false positives is likely to increase in high-stress, time-constrained or high-stakes decision-making contexts. This is because the consequences of not taking action in an urgent or high stakes

[39] Interview with government lawyer, 10 November 2022.

[40] The importance of handling arrangements for review, retention and deletion was emphasised in the decision of the Investigatory Powers Tribunal in "(1) LIBERTY, (2) PRIVACY INTERNATIONAL- and - (1) SECURITY SERVICE, (2) SECRETARY OF STATE FOR THE HOME DEPARTMENT, IPT/20/01/CH" (30 January 2023).

[41] Interview with government lawyer, 25 November 2022.

[42] Interview with non-government lawyer, 14 November 2022.

> situation could be significantly worse than the consequences of incorrectly taking action on the basis of a false positive.'[43]

Nonetheless, it is important to remember that false positives could have greater implications for the individual whose privacy rights are engaged, and too many false positives could divert investigative resources away from genuine culprits.

Ultimately, the maximum acceptable degree of targeted or collateral intrusion cannot be quantified solely on the volume of data acquired or the sensitivity of that data; it must be assessed in the context of the wider investigative aims, the urgency of decision-making, and the consequences of missing potentially pertinent material or investigative opportunities. This emphasises the need for maintaining a high degree of professional judgement throughout the proportionality assessment process, rather than relying solely on a quantitative or mathematical approach.

2.1.3 Human vs machine intrusion

The relative intrusion of human and machine-based approaches to surveillance is a contested topic. As noted by Babuta, Oswald and Janjeva (2020):

> 'The use of AI arguably has the potential to reduce intrusion, both in terms of minimising the volume of personal data that needs to be reviewed by a human operator, and by resulting in more precise and efficient targeting, thus minimising the risk of collateral intrusion. However, it has also been argued that the degree of intrusion is equivalent regardless of whether data is processed by an algorithm or a human operator. According to this view, the source of intrusion lies in the collection, storage and processing of data. The methods by which this is achieved – whether automated or manual – are immaterial.'[44]

Over the course of this project, some interviewees were more steadfast in their view that automated processes can reduce intrusiveness, when compared with manual human review of data. One commented, 'if selection is carried out by machine, although that does invade privacy, it does so without (in the initial stages) being apparent to human consciousness (...) and for me that is important.'[45]

---

[43] Anna Knack, Richard Carter and Alexander Babuta, "Human-Machine Teaming in Intelligence Analysis," *CETaS Research Report* (December 2022): 17.
[44] Alexander Babuta et al., "AI and National Security: Policy Considerations," *RUSI Occasional Paper* (April 2020): 24.
[45] Interview with government lawyer, 25 November 2022.

Other responses, however, demonstrated less certainty, referencing the importance of algorithms being able to perform functions that a human could not: 'if you wouldn't let a human do it, you shouldn't let a computer do it. However, if you've passed that initial test, is it more proportionate or less intrusive for a computer to do it? Perhaps.'[46] Another interviewee pointed to the *Bridges* case which shed light on the matter in a legal context: '*Bridges* is a useful starting point in terms of automatic deletion – they concluded that the intrusion of being processed by the camera is 'negligible', but they conclude that it is there.'[47]

Although the *Bridges* case gives useful guidance, and the interview data presented here raises important considerations, the question of machine intrusion relative to human intrusion remains a matter of open debate. This should continue to be kept under review as the performance and operational use of automated analytics develops.

2.1.4 Reconciling competing rights

Article 8 ECHR is qualified by positive obligations under Articles 2 and 3 of the ECHR, which a state may infringe if it fails to 'take measures within the scope of their powers which, judged reasonably, might have been expected to avoid' a real and immediate risk to an individual or society.[48] Beyond the legal obligations imposed by Articles 2 and 3, there is also strong societal pressure on governments and law enforcement agencies to avoid letting any potential security threats 'slip through the net'. One law enforcement interviewee commented, 'we need to work with government to understand risk appetites and what people are worried about. Is it the intrusive nature of the work or a child dying because we didn't exploit the technology in the way we could have done?'[49]

The national security and law enforcement community face a daily challenge to make sure that both obligations are being upheld to the maximum possible level, summarised in the comments of one interviewee: 'you have a positive right under Article 2, and you have collateral subjects under Article 8. You cannot just add up all the collateral and assess that it outweighs the right to life – it's about judgement in context. This is something we come up against repeatedly.'[50]

[46] Interview with law enforcement lawyer, 21 October 2022.
[47] Interview with law enforcement lawyer, 21 October 2022.
[48] Emma Lazarovna Tagayeva and Others v Russia, Application Nos. 26562/07, 49380/08, 21294/11, 37096/11, 14755/08, 49339/08, 51313/08, (13 April 2017): para. 482.
[49] Interview with law enforcement representative, 3 November 2022.
[50] Interview with law enforcement lawyer, 21 October 2022.

The process of reconciling competing rights is one that evolves over time and responds to public sentiment. The challenge is perhaps most pronounced in policing – particularly units such as Counter Terrorism Policing (CTP), which not only possess intrusive collection powers but also officers with coercive powers and firearms. Their 'opportunity to intrude' and directly affect an individual's life is therefore heightened; a misdirected action can have substantial and irremediable consequences.[51] In the law enforcement context, these consequences potentially extend beyond Article 8. For example, Article 11 (the freedom of assembly and association) may be engaged by police activity that restricts the movements and actions of individuals or groups. Despite this report's focus on Article 8 rights, the authors note that proportionality calculations must often weigh up positive obligations against the potential impact of subsequent activity on other human rights and freedoms, beyond simply the risk of privacy intrusion.

This also brings to light a potential issue of diminishing marginal returns: after a certain point surveillance can continue to be increased but for ever-decreasing reductions in harms, until a point is reached where the reduction in harm does not justify the level of intrusion occurring. The difficulty here lies in identifying the specific points where these thresholds are crossed, particularly as such calculations are deeply bound up in societal expectations of privacy and understanding of the threat landscape (or lack thereof).

2.2 Implications of AI

The previous section considered the nature of privacy intrusion of automated analytic techniques in general, but it is important to note the additional implications introduced by AI in this context. While the core decision-making process and proportionality test remain the same regardless of whether AI is used, the 'way that the AI algorithm causes intrusion can vary significantly', meaning the depth of questioning involved is different.[52]

The figure below illustrates the four main additional proportionality considerations when using AI-enabled analytics.

[51] Interview with law enforcement representative, 3 November 2022.
[52] Interview with law enforcement lawyer, 21 October 2022.

Figure 3: Additional proportionality considerations when using AI-enabled analytics.

Sources: Inscrutability: Christoph Molnar, Giuseppe Casalicchio and Bernd Bischl, "Interpretable Machine Learning – A brief history, state-of-the-art and challenges", Communications in Computer and Information Science, (2020): pp. 417-431; Finale Doshi-Velez and Been. Training data: Interview with government lawyer, 25 October 2022.

Elaborating further on point 4), one interviewee raised the question of whether the use of AI could result in an increase in *authorised collection of data*, and whether this would be significant in its own right.[53] There is also a further question about the *extent of processing required* by the AI system in order for it to 'learn' to a suitable level.[54]

On the other hand, some interviewees held that if the purpose of data processing is to enhance or enrich a knowledge base, this is arguably less intrusive than the tool being used to conclude something about a particular individual in a live investigation.[55] Data collected

[53] Interview with non-government lawyer, 10 November 2022.
[54] Interview with non-government lawyer, 14 November 2022.
[55] Interview with government lawyer, 10 November 2022; Interview with law enforcement lawyer, 21 October 2022.

via warrants for research or capability development rather than strictly operational purposes may therefore carry a lower risk of intrusion.[56]

Furthermore, there may be circumstances where an organisation wishes to develop an automated system that could reduce *future* intrusion, despite such a system incurring some degree of *present* intrusion as it entails the collection of training data. An example is the use of AI in a collection process, resulting in the collection of less, but more relevant, data. As such, the proportionality of capability development activities must also be considered in the context of potential reductions in future intrusion that the planned capability could provide. While this report does not seek to make a definitive statement on the proportionality of intrusion of internal capability development activity, its relevance is highlighted here and in the analytical framework proposed in Section 3.

By way of comparison, publicly available datasets are seen as a crucial resource for training AI models in the commercial context. To train their large language model GPT-3, OpenAI scraped 45 terabytes of text data from the open web, while Meta's LLaMa is trained on 1.4 trillion tokens (collections of pieces of words).[57] Relying on the open web as a training source, however, comes with certain risks regarding data provenance, and potential biases within datasets that have not been verified line by line or curated with a particular purpose in mind. It should be noted for example that the Italian Data Protection Authority has recently imposed a limitation on the processing of personal data in ChatGPT. Of particular concern was the absence of an appropriate legal basis for the collection and processing of personal data by OpenAI for the purposes of training the model, and a breach of the 'accuracy' principle as the output provided by ChatGPT does not always amount to accurate personal data.[58] Although this is a different context and jurisdiction, it will be important for the implications of this decision and others to be kept under review.

2.3 Assurance, handling and retention

Checks and balances for data handling were perceived by several research participants as integral to weighing up the intrusion of automated analytics.

Taking appropriate steps to ensure that the procurement, configuration, implementation, monitoring, and potential decommissioning of automated analytic tools are all conducted in

[56] Interview with law enforcement lawyer, 21 October 2022.
[57] OpenAI, "Language models are few-shot learners," *Advances in neural information processing systems* 33 (2020): 1877-1901; "Introducing LLaMa: A foundational, 65-billion-paramater large language model," Research Blog, Meta AI, 24 February 2023, https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.
[58] Further information: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832.

an auditable way is an essential part of the investigative process. The end-to-end nature of these criteria is important to highlight, as they provide a birds-eye view of potential errors or when an activity that is initially proportionate gradually stops being so. One interviewee emphasised, 'we have stopped projects because we no longer thought they were going to achieve what we wanted them to achieve. That is not just about doing proportionality at the beginning but continuing to do it.'[59]

One non-government interviewee outlined three internal risk assessments – i.e. not required by the IPA – that a national security practitioner might carry out to impress upon the regulator the role of their assurance, handling and retention mechanisms: assessing the risk of an adverse outcome; the risk of litigation and challenge; and the risk of undermining the institution if publicity is thrown on an adverse outcome.[60] In the eyes of practitioners, the Authorities granted by oversight bodies then carry significant weight in creating a safe space and affording greater protection from challenge in potential grey areas of proportionality.[61]

Policies pertaining specifically to the retention and deletion of data received particular attention over the course of the research. Some interviewees placed emphasis on handling arrangements, although these arrangements may have historically been designed with human processes in mind rather than automated analytical processes. Nonetheless, these arrangements aim to contain the intrusion involved in the initial acquisition as soon as feasibly possible by identifying and destroying material irrelevant to a line of investigation. Emphasis was placed on the volume of raw data that is discarded without being analysed, and the speed with which the judgment is made that a piece of data is no longer useful. Although it should be noted that it is usually difficult to know how much intelligence value a given dataset will provide when first acquired, meaning this process could also involve the discarding of potentially valuable data.

The strength of the retention policy should be the counterweight that reassures oversight bodies that action will be taken swiftly when intelligence value (or lack thereof) is later ascertained.[62] Practitioners generally felt this played out well in practice: 'I've found our overseers quite sympathetic when we want to do something new, but you need appropriate protections, and you cannot come across as gung-ho. If you show you have engaged colleagues all around the organisation and maybe externally as well, then you should be

[59] Interview with law enforcement lawyer, 21 October 2022.
[60] Interview with non-government lawyer, 27 October 2022.
[61] Ibid.
[62] Interview with government lawyer, 10 November 2022; Interview with government lawyer, 25 November 2022.

afforded a degree of latitude.'[63] With this being said, it is important to recognise that datasets used to train ML models may need to be retained longer than other types of operational data. This is also reflected in the framework presented in Section 3.

Another factor that might affect data retention is the feasibility of querying data at source rather than needing to make a copy of that data. Privacy-enhancing technologies[64] (PETs) offer one way of achieving this: previous CETaS research has outlined numerous potential use cases for PETs in the national security context.[65] However, some scepticism remains concerning the practicalities of deploying PETs due to current technical resource requirements. A Royal Society report published in January 2023 stressed that 'the PETs value proposition remains abstract and the business case for adopting PETs is unclear for potential users.'[66] Given the questions surrounding technological readiness and computational requirements, it is unclear whether PETs will shift the dial on retention concerns in the short term, although further research persists with the aim of reducing those barriers.

2.4 An ongoing approach to proportionality assessment

There are differences in the way that sections of the national security and law enforcement community conduct proportionality assessments. This includes the stage at which the proportionality assessment is conducted and the frequency with which it is repeated. For some agencies (particularly those that deal with bulk data), the proportionality assessment process may be ongoing and iterative, perhaps because new techniques are applied to bulk data previously collected under IPA authorisation. For example, the IPA includes safeguards relating to the retention and disclosure of material acquired through bulk interception, including ensuring that 'the selection of any intercepted content or secondary data for examination is necessary and proportionate *in all the circumstances*' (emphasis our own).[67] The requirement to ensure necessity and proportionality 'in all the circumstances' implies that this assessment must be made on an ongoing basis, whenever previously intercepted data is selected for examination (whether by human review or automated techniques).

[63] Interview with government lawyer, 10 November 2022.
[64] PETs are a range of technologies designed to address and mitigate privacy risks through encryption, data minimisation, anonymisation, and pseudonymisation. Further information: https://cetas.turing.ac.uk/publications/privacy-and-intelligence.
[65] George Balston et al., "Privacy and intelligence: implications of emerging privacy enhancing technologies for UK surveillance policy," *CETaS Research Reports* (July 2022).
[66] The Royal Society, *From privacy to partnership: The role of privacy enhancing technologies in data governance and collaborative analysis* (The Royal Society, 2023), 6.
[67] IPA Part 6, s 152(1)(b).

Irrespective of differences between organisations and contexts, there is a broader opportunity to embrace a longer-term outlook on proportionality assessment. Currently, proportionality frameworks are focused on case-by-case assessments. But in an environment where automated analytics become more ubiquitous, there should be a rigorous attempt to assess the *cumulative effect* of multiple automated systems feeding into each other over time. These cumulative effects include the possibility of wider societal impacts which, when assessing a single technique (and its associated data) in isolation, are difficult to comprehend, but when assessing the compound effect of multiple automated techniques may become easier to identify and prepare for.

At a basic level, this could involve establishing a function that quantifies how many automated systems are in commission in an organisation compared to previous years, establishes the amount and rate of growth of data processing, and maps how these systems feed into each other both within and across organisations. This may provide an additional layer of assurance beyond *activity-level* proportionality assessments.

2.5 Practitioner judgement vs a metrics-based approach

This research evaluated the potential merits and pitfalls of relying on practitioner judgement when assessing proportionality of privacy intrusion, and the feasibility of a metrics-based approach given the various context-specific considerations involved in these judgements.

Some are critical of the 'metrification of intensities of interference' – describing a situation of '30% infringement' and then comparing it to a situation of '31% infringement' is deemed impossible because of the mathematical difficulty of determining what counts as a cost or benefit, and the endless considerations which can only be settled through value judgements.[68] In the words of Anne Peters, 'the multifactorial and interconnected social problems and the underlying conflicts between various groups or individuals cannot be "resolved" by relying on checklists and ordinal numbers. In an open society, they can only be addressed through constant deliberation and debate in which reasons are formulated, discussed, and challenged.'[69]

This report's research surfaced a range of perspectives regarding the feasibility and desirability of a metrics-based approach to proportionality assessment. Generally, interviewees from a legal background tended to raise the most concerns and caveats (although this was not a unanimous stance). This reflects academic concern that metrics

[68] Ibid.
[69] Ibid.

may tell us little about an individual case and could be unrepresentative of real-world cases where AI is used.

Apprehension was expressed that a mathematical approach might provide 'false confidence', especially in the 'middle ground' where consideration of detail is required,[70] or where intrusion may be 'fleeting or momentary' or difficult to predict.[71] A 'formulaic approach' was rejected by one interviewee, in favour of 'a straightforward, old-fashioned assessment as to whether or not evidence supporting the line of inquiry is sufficiently present.'[72] While this may be overly simplistic, the significance of context ('the circumstances of the case') in the assessment was emphasised by many interviewees. In the words of one:

> 'Attaching percentages or equations can make things easier to articulate. And it might help with articulating the issue to the Judicial Commissioners for example. But given that every scenario is different, it depends on the facts and circumstances. Something that is justifiable in one scenario is unjustifiable in another. So it depends whether that metrics-based approach would be able to take into account those nuances.'[73]

This concern was also reiterated in the context of legally privileged information such as confidential journalistic information. The obligation to consider a higher level of protection for such material is applied through the overarching provisions of Section 2 of the IPA and the existing Codes of Practice.[74] One interviewee summarised the interpretive nature of this process when saying, 'there is instruction in legislation that provides guidance. But things may crop up that don't neatly fit into those categories – and there one must exercise professional judgement.'[75]

One interviewee suggested that it could be possible to assign a numerical value to a level of intrusiveness per technique, but that this would have to be adapted for 'cocktails of techniques that do more than the sum of their parts than if deployed individually.'[76] This reiterates the additional cumulative intrusion risk discussed in Section 2.4.

[70] Interview with law enforcement representative, 3 November 2022.
[71] Interview with law enforcement lawyer, 21 October 2022.
[72] Interview with government lawyer, 25 November 2022.
[73] Interview with non-government lawyer, 14 November 2022.
[74] Home Office, *Report on the Operation of the Investigatory Powers Act 2016* (Home Office, 2023), 11.
[75] Interview with government lawyer, 10 November 2022.
[76] Interview with non-government lawyer, 10 November 2022.

Generally, interviewees were cautious about 'one-size-fits-all' models: 'a framework needs to be able to expand and contract as necessary to respond to what the scenario is.'[77] One interviewee highlighted an experimental approach to assessing new data-driven technologies and techniques in their context based on four main criteria: strategic purpose; lawfulness; technical validation; and ethical considerations. This was said to use a sliding-scale, risk-based approach to determine the level of internal governance or escalation required for a particular project.[78]

In summary, the research has found that both the meaning and desirability of a 'mathematical', 'numeric', or 'formulaic' approach to proportionality assessment are subject to different views. The most useful approach is one that can allow those making proportionality assessments to do so in a more consistent, objective, and systematic manner, without unduly restricting their ability to exercise discretion and professional judgement. The following section attempts to formalise such an approach.

[77] Interview with law enforcement lawyer, 21 October 2022.

[78] Interview with law enforcement representative, 3 November 2022.

Section 3. A Structured Framework for Assessing Proportionality of Privacy Intrusion

This section builds on the previous analysis by presenting a structured analytical framework of the factors most important to the assessment of proportionality of automated analytics. The questions are designed to inform judgements about whether a less intrusive method is available, and the balance being struck between individual rights and national security interests. They define a taxonomy and common vocabulary that should enable more clarity, transparency, and consistency in and between proportionality arguments. Readers should note that the factors are also informed by prior research conducted by IPCO TAP, which has published reports on 'Privacy Intrusion Metrics' and concepts for assessing proportionality of privacy intrusion in data collection and analytics.[79]

The factors and associated questions are intended as guidelines (or prompts) when developing and assessing proportionality arguments in the context of the automated analytics lifecycle, reflecting aspects of 'good data science and AI practice'. Different subsets of questions will apply depending on the scenario and on the aspect(s) of the proportionality test being considered. The framework recognises that privacy intrusion can occur at any stage, perhaps due to the nature or volume of the data sets, or inconsistencies between attributes of data sets. For example, if the training data is more accurate than the data for analysis, or contains different biases, this could lead to inaccurate results, which in turn could result in unnecessary intrusion.

The factors are organised into six broad categories: data inputs; results; degree of human inspection; analytic techniques; data management; and timelines and resources required. While these last two categories may be viewed as second order with respect to intrusion, they are important considerations from a practical perspective and can ensure better comparison of methods, repeatability, and audit – a less intrusive method is still impractical if it cannot compute the results within the required timescales.

There are two caveats to bear in mind when navigating the questions. Firstly, they are designed to be guidelines only, for identifying areas of concern in respect of a proposed technique or scenario, or when comparing automated analysis against an alternative technique. They do not represent an exhaustive list, bearing in mind the context-specific nature of the assessment. Secondly, they are not intended to provide a determinative

[79] IPCO Technology Advisory Panel, "Metrics of Privacy Conference" (November 2018); IPCO Technology Advisory Panel, "Privacy Intrusion Metrics: concepts and considerations" (October 2021).

'proportionate/not proportionate' evaluation, but instead to help direct a deployment strategy with a greater degree of rigour, with higher-risk proposals triggering a greater degree of internal scrutiny. We note the recent warning given by the Investigatory Powers Tribunal (in the context of data safeguarding requirements) that categorising *data management failures* amounting to unlawfulness as '*high legal risk'* was not appropriate:

> 'Statements in the form of risk factors could not be relied upon as excusing any actual compliance breaches.'[80]

It would therefore be vital that any internal assessment using the factors below is combined with processes to identify and prevent error, misuse or at worst non-compliance, particularly bearing in mind the increased data handling complexity that is likely to come with the development of AI and the use of its outputs.

Two areas that are not covered in detail here are expectations of privacy, and distinctions between metadata and content (which are increasingly difficult to determine with the proliferation of new communications infrastructure and data streams). Factors that were either infrequently or not mentioned at all by interviewees include: scaling effects of intrusion (see Section 1.2); whether a tool is prototype or final product; computational resources required; and details of data volume.

[80] "(1) LIBERTY, (2) PRIVACY INTERNATIONAL- and - (1) SECURITY SERVICE, (2) SECRETARY OF STATE FOR THE HOME DEPARTMENT, IPT/20/01/CH" (30 January 2023): para 106.

Factors when considering intrusiveness of automated analytic methods

1. Datasets

1.1 Datasets for analysis

- What are the datasets?
- What are the sensitivities and are there any data types of concern e.g. legally privileged?
- What are the granularities and is it possible to infer detailed information regarding specific individuals?
- Are there any issues in terms of collateral intrusion?
- Are there concerns related to integrity?
- Could the volume be reduced?

1.2 Datasets for training (questions only applicable for AI methods)

- What are the datasets and do the volumes meet requirements?
- What are the sensitivities and are there any data types of concern e.g. legally privileged?
- What are the granularities and is it possible to infer detailed information regarding specific individuals?
- Are there any issues in terms of collateral intrusion?
- Is the data consistent with the data for analysis in its granularity, integrity, and biases and if not, what are mitigations?
- What is the quality of data labelling (if applicable)?
- Can the purpose be achieved using anonymised or synthetic data (as opposed to real citizen data)?
- If real citizen data is used, could the volume be reduced?

1.3 Datasets for testing

- What are the datasets and do the volumes meet requirements (i.e., yield statistically meaningful results)?
- What are the sensitivities?
- Are there any issues in terms of collateral intrusion?
- Is the data consistent with the data for analysis in its granularity, integrity, and biases and if not, what are mitigations?

- Is the data consistent with the training data in its granularity, integrity, and biases and if not, what are mitigations?
- Can the purpose be achieved using anonymised or synthetic data, as opposed to real citizen data?
- If real citizen data is used, could the volume be reduced?

2. Results

- Are results produced regularly or upon request?
- Is this a step change in the scale of results that can be generated?
- Does the analysis have the potential to result in further data being collected?
- What is the granularity, accuracy of prediction, and explainability of results?
- Will they support a specific/discrete investigation or a strategic/general purpose solution?
- Will they alone be used to inform subsequent decision making? If not, what other factors will be taken into account?
- Who/which systems require or will have access to results, or reports based on the results? Is there automatic chaining of analytics?
- How does the intrusion scale from individuals to different populations? For example, is it constant, additive, or multiplicative?
- How could the intrusion affect communities with protected or sensitive characteristics?
- Could any inaccuracies, as a consequence of bias, uncertainties, or mismatch between training, testing, and analysis data, lead to adverse outcomes?
- What are the error reporting processes in the event of adverse outcomes or misdirected actions deriving from these results and subsequent decisions?

3. Human inspection

- When is human inspection of intermediate results expected to occur and at which points? For example, does it happen after one or several automated filtering steps?
- To what extent are such intermediate results understandable by a human and potentially actionable?
- Does any automated filtering inform another decision or automated system (before human inspection)?
- Are levels of uncertainty in the algorithms known and if so, what is their impact on the volume and sensitivity of data requiring human inspection?

- Is the expected human inspection feasible? For example, is there an upper bound on the amount of data that can be reviewed and number of people required?

4. Tool design

- What biases and constraints exist within the algorithm design (excluding training data), including assumptions about data for analysis and uncertainties of prediction?
- What is the amount of data required to train and test model(s), is it minimal?
- What is the expected lifetime before retraining is required and how is performance monitored?
- How much control does the user have over thresholds within the algorithm(s) (if applicable)?
- Is this automating a new capability or an existing process?
- Is this an established or a prototype tool and is it novel in this domain, business as usual in this domain, or business as usual in another domain?

5. Data management

- Who/which systems have access to the datasets and are you assured there is suitable access control?
- What are the retention and deletion policies for all the datasets and is the training data extended retention policy consistent with the retraining lifecycle and any requirement to revert to a previous model?
- Are you assured by whoever is retaining the data that it is protected from loss and corruption?

6. Timelines and resources

- What is the urgency, gravity, and extent of potential harm?
- What are the timescales for each of the steps and is there any flexibility?
- Are there adequate computational resources (processing power, storage) and training data suitably labelled (if required), to meet those timescales?

3.1 Visualising relationships between factors

Some factors listed in the framework are naturally quantitative in nature, such as timescale, uncertainty, accuracy, and volume of data. For other factors, a user may be able to judge relative quantities. For example, points in the process where there is *more* (or *less*) collateral intrusion, or where one dataset is *more* (or *less*) sensitive than another. When this is

possible, it may be useful to visualise the relationships between factors, or more specifically, to consider the *impact* of one factor on another.

Four examples of graph-based visualisations help to illustrate. In each, two factors are plotted against each other resulting in a *curve* that can take a wide variety of shapes, including a single straight line (a linear relationship), multiple straight lines with inflection points, or an exponential curve.

These graph-based visual aids do not signify a 'mathematical' approach to assessing the factors, nor do they address *how* factors should be judged and weighed against each other in different contexts. Rather, they should serve as additional prompts in discussions of proportionality. Some key questions to ask when viewing the graphs include:

- To what extent does the actual shape of the curve matter and why?
- If it does, what level of detail is important? For example, does it matter if the shape is linear or exponential, or simply that it is increasing or decreasing?
- What would need to happen for the curve to have a different shape, and/or what changes to the factor(s) would result in a different shape?

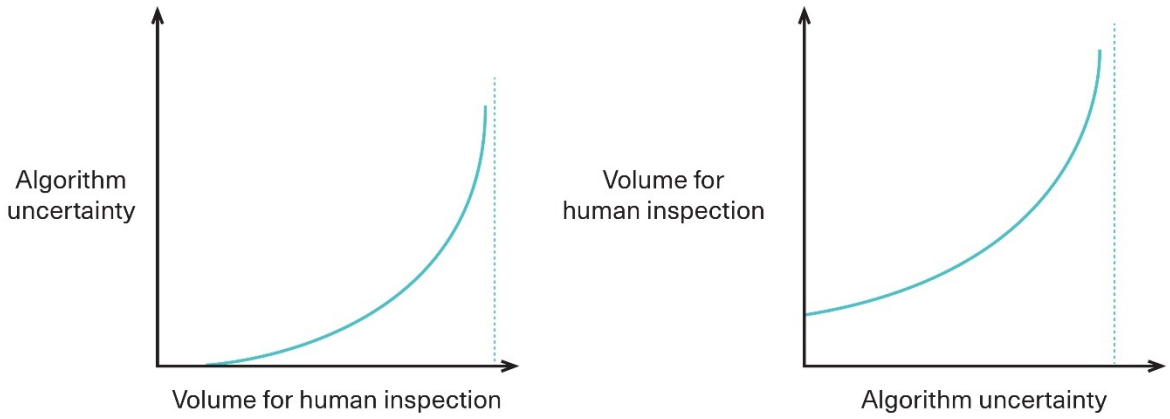
Finally, inflection points and asymptotes[81] are likely to be of particular interest as they may indicate key thresholds that could be dialled up or down, depending on the scenario.

Example 1: Impact of algorithm uncertainty on the volume of data requiring human inspection

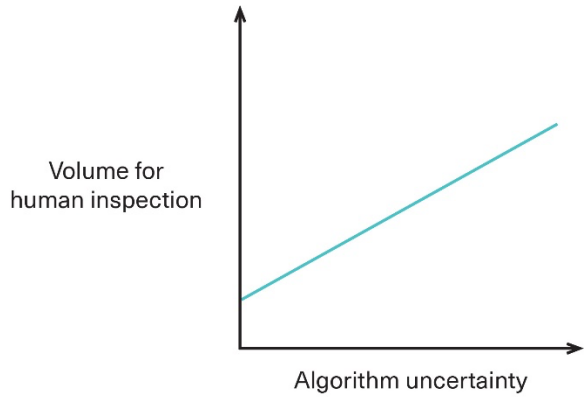
Usually, *increased* algorithm uncertainty would lead to an *increased* volume of results requiring human inspection. Two example curves are depicted below; both curves are exponential, but the axes are transposed. On the left, uncertainty grows much faster than volume. The curve approaches an asymptote (a value that is never reached, as indicated by the dashed vertical line), which is the upper bound on the total volume of data that needs to be reviewed. In general, this scenario is unlikely, especially the steeper parts of the curve, which mean that when uncertainty increases, hardly any more results require to be inspected. On the right, we have the converse: volume grows faster than uncertainty and the asymptote represents the maximum uncertainty that can be considered. This is a more likely scenario. Note both graphs indicate the results *always* require some degree of human

[81] An asymptote is a value that is never reached, which is approached by the curve but never touched by the curve.

inspection, even when there is no algorithm uncertainty. This is indicated by the curves starting *past* the origin (the intersection of the horizontal and vertical axes).



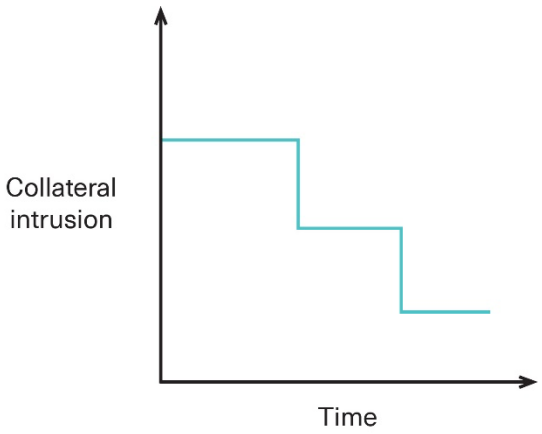
Another example, below, is a linear relationship. Depending on the steepness of the curve, this may be the most desirable scenario, especially when there may be high values of algorithm uncertainty.



Example 2: Impact of iterative training sets on the type of intrusion

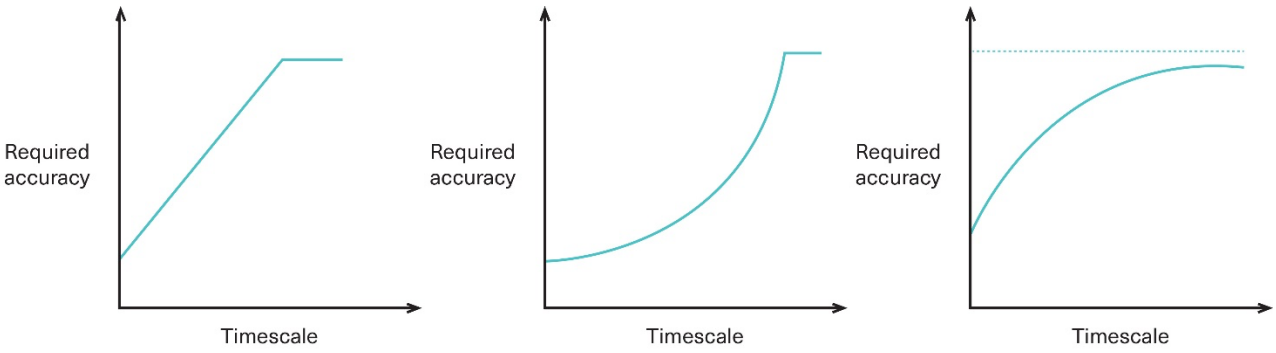
Iterative AI training sets would be expected to enable *better* filtering of data: after the introduction of each new training set (iteration), collateral intrusion of the analytic process would be *reduced*. This is depicted by the stepwise curve below, each step representing the discrete training steps (it may be useful to consider when and how often the steps occur). This type of filtering can occur both during the overall analytic process and within the collection process. An example of the latter is the use of facial recognition to exclude images from being stored during collection.

It is possible that a new training set introduces new collateral intrusion – that scenario is not depicted by this curve.



Example 3: Impact of timescales on accuracy requirements

We would expect the accuracy requirements of the technique applied to data (for analysis or training) and/or results to *increase* as the timescale for returning the results (the time available for the analysis) *lengthens*. As the need for results become less urgent (perhaps because the level of potential harm decreases) the expectation of accuracy would increase. Three curves are shown below. In the leftmost and middle case, at some discrete point there is a maximum level of required accuracy/maximum possible accuracy. In the rightmost case, there is no discrete point indicating the maximum has been reached, but rather an asymptote. In each case, there is a vital assumption that there are computational resources to perform the analytics within the timescale.

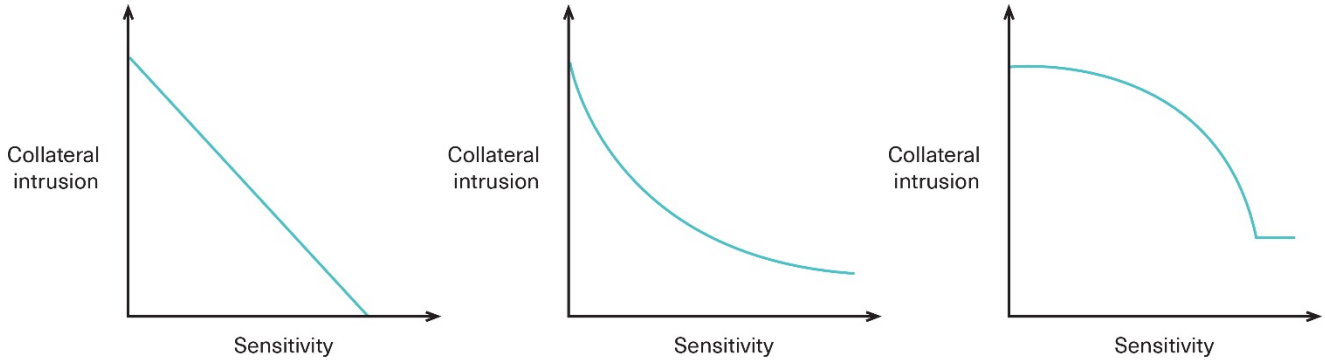


Example 4: Impact of sensitivity on collateral intrusion

In some scenarios collateral intrusion *decreases* as the sensitivity of data employed (for analysis or training) *increases* (though it is worth noting that this does not preclude a high level of *targeted intrusion* because of the nature of the datasets employed). For example, as

one considers increasingly sensitive characteristics, individuals who are of no interest are less likely to be identified or present in the data sets.

The leftmost curve below depicts the scenario where the impact is linear and eventually there comes a point at which there is no collateral intrusion. The middle and rightmost curves depict scenarios in which there is always some level of collateral intrusion. This may represent the intrusion resulting from the training set(s), for instance. In the middle scenario the level of intrusion reduces rapidly when there is still low sensitivity. Whereas in the rightmost scenario the level of intrusion reduces only gradually, but a significant leap in sensitivity starts to reduce intrusion very quickly. For these reasons it may be useful to consider what represents 'low' and 'high' sensitivity in the given scenario.



Section 4. Conclusions and Recommendations

This report has presented a new structured analytical framework for assessing proportionality of privacy intrusion arising from the use of automated analytics. This framework is intended to complement existing authorisation and compliance processes with an additional layer of rigour, to provide assurances that all relevant considerations have been accounted for at every stage in the automated analytics lifecycle.

The report's main recommendation is as follows, and is relevant to all organisations with access to IPA powers:

1) **Adopt the proposed framework** in practice across varying contexts within national security and law enforcement, and **assess the extent to which the framework assists with the structured legal proportionality test**, in particular the question of whether a less intrusive means is available.

The research also prompted the following additional recommendations based on the opportunities and challenges highlighted throughout the report:

2) **Processes are required to understand and monitor the risk of cumulative intrusion caused by multiple, connected, automated systems** feeding into each other over an extended period. Stakeholders across the national security and law enforcement community should **consider the methods at their disposal to develop a more robust approach to monitoring this risk**, consulting with internal legal advisors and oversight bodies.

3) There may be circumstances where an organisation wishes to develop an **automated system that potentially reduces *future* intrusion, but doing so will incur some degree of *present* intrusion** as it entails the collection and extended retention of training data. **Proportionality assessment should provide specific consideration for this scenario, particularly in respect of accuracy and utility of results**. In these circumstances, it will be important to measure the extent to which these future reductions in privacy intrusion are delivered in practice.

4) **Public expectations of privacy** are dynamic and responsive to a range of contextual factors. Achieving a more rigorous understanding of this will be beneficial for both assessing intrusion and **promoting transparency and public trust**. One way of doing this could be to commission regular **studies surveying public perceptions of intrusion from automated analytics in different national security scenarios**.

About the Authors

Ardi Janjeva is a Research Associate at CETaS. His research interests include technology-enabled threats in the 21st century; the future of intelligence innovation; technology-based geostrategic alliances and competition; and the relationship between technology and economic resilience.

Professor Dame Muffy Calder is a Senior Research Consultant at CETaS. She has been Vice-Principal and Head of the College of Science and Engineering at the University of Glasgow since 2015 and was previously the Chief Scientific Adviser for Scotland. She is Chair of the Investigatory Powers Commissioner's Office Technology Advisory Panel and a member of the Prime Minister's Council for Science and Technology. Dame Muffy is a computer scientist with research interests in modelling and automated reasoning for complex, interactive, and sensor-driven systems.

Dr Marion Oswald MBE is a Senior Research Associate at the Alan Turing Institute and Associate Professor in Law at Northumbria University. Marion is a lawyer with over 30 years' experience spanning several contexts: law firms, international technology businesses, central government including national security, academia, and oversight functions. She sits on the Board of the Centre for Data Ethics and Innovation and chairs the West Midlands Police data ethics committee. She has a particular research interest in the human rights, ethics and use of data analytics within policing and intelligence agencies.

CETaS

Centre for Emerging Technology and Security

RESEARCH REPORT