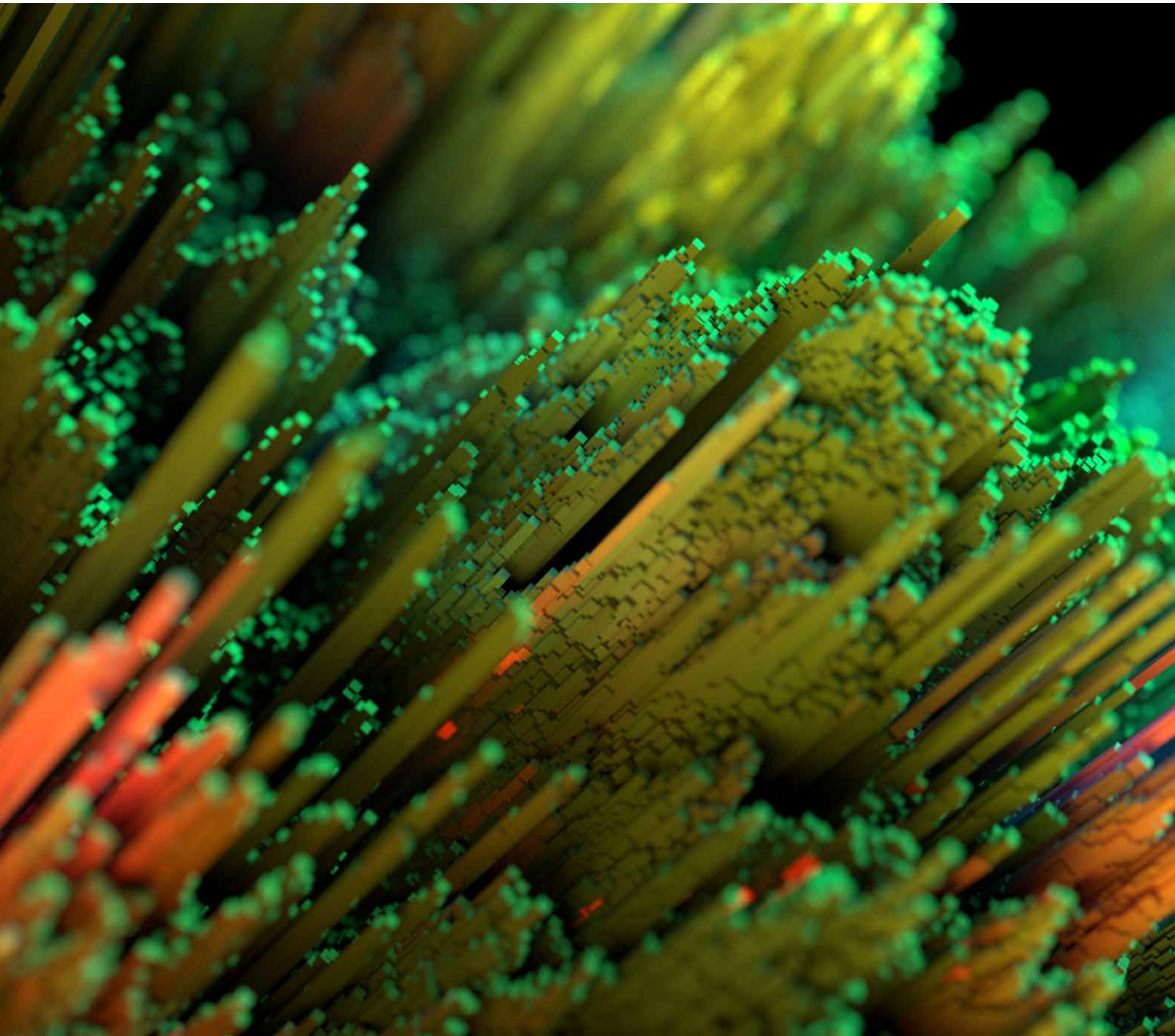


UCL Centre for
Digital Innovation.



UCL

What is Centering the Public Interest in AI?



Abstract

With the rapid development and adoption of algorithmic systems, there is an increasing awareness and need to ensure that such systems are governed appropriately. Indeed, precipitated by high-profile cases of harm, a social concern over the future of AI has arisen—a concern that has led to proposals for policy, standards, and regulatory interventions. Indeed, there are regulatory activities and signalling from jurisdictions across the globe, relating to both the broad applications of artificial intelligence in terms of, for example, risk and privacy, and sector-specific applications, concerning the applications of AI to sectors such as education, finance, and health. In this article, we examine the existing regulation, comparing the more horizontal (broad) approach of the European Union with the vertical (sector specific) approach from the United Kingdom. Through this examination, we determine that the lack of an overarching framework that creates a common understanding and interpretation of what represents the public interest in the context of AI is likely to lead to many different interpretations. We recommend that in order to address the public interest and create a robust regulatory framework, we need to be clear about from which perspective we are considering risk and where the motivations for AI development lie and, in doing so, we examine the approach of relevant stakeholders including small and medium sized enterprises and big tech. We conclude that there is a need for a coherent framework to guide how the competing interests that comprise the 'public interest' ought to be weighted.

Keywords: Regulation, Compliance, Standards, Legislation, Artificial Intelligence, Accountability, Governance, Fairness, Transparency

Contents

Abstract	2
1. Executive Summary	4
2. Legal Frameworks	7
3. EU AI Act - Horizontal par excellence?	9
4. The UK Approach	11
4.1 Sector Mapping	11
4.1.1 Sector Survey	11
4.2 Teasing out a Horizontal Approach?	15
4.2.1 UK National AI Strategy	15
4.2.2 Assurance Roadmap	15
4.2.3 Public Sector AI Transparency Standard	16
4.2.4 AI Impact Assessments - Draft Legislation	16
5. Critique of Vertical vs Horizontal	16
6. What is Driving Public Interest?	17
7. Conclusion	20
8. References	21
Authors	25



TOGETHER
WE
GROW

1. Executive Summary

With the rapid development and adoption of algorithmic systems, there is an increasing awareness and need to ensure that such systems are governed appropriately. Indeed, precipitated by high-profile cases of harm (such as the Facebook-Cambridge Analytica controversy and the GCSE/A-Level Scandal), a social awareness has arisen - an awareness that has led to policy, standards and regulatory proposals.

The legislative debate is perhaps the strongest indicator of the concretisation of algorithmic system regulation. Indeed, there are regulatory activities and signalling from jurisdictions across the globe:

- *Existing Regulation.*

Despite the fact that there are no current laws that specifically regulate the use of artificial intelligence (AI), there exist laws in Data protection regulations such as the General Data Protection Regulation (GDPR) that are relevant to AI. Indeed Data Protection Impact Assessments are heavily referenced in the AI impact assessment literature [1]. Additional laws are applicable to the use of AI, such as equality and anti-discriminatory laws (AI systems must not discriminate), rights to recourse (citizens have a legal right to an explanation and hence, where appropriate, systems must be explainable), and industry-specific rules (financial services, insurance, recruiting). The website of the Organisation for Economic Co-operation and Development (OECD) [2] keeps a database of national strategies and policy initiatives around the world. In Canada, the government developed an Algorithmic Impact Assessment [3] that policy makers and other officials should use to assess and mitigate the risks associated with deploying an automated decision system. In the UK, the Information Commissioner's Office (ICO) published guidance on the AI auditing framework [4], explainability [5] and AI and data protection [6]. Australia published an AI Ethics Framework with a set of voluntary AI Ethics Principles to follow [7]. Most recently, senators in the US have proposed a US Algorithm Accountability Act [8], and the Japanese government [9] has published its current thinking on AI governance.

- *Sector-specific regulations.*

There is some sector-specific guidance existing on the use of AI. For instance, the Department of Defence in the US laid out guidance for use of AI in defence [10], the UK's Financial Conduct Authority is active in the standards debate for AI systems in financial services [11], and the UK's Care Quality Commission for medical diagnostic services [12]. Other application specific standards are being developed such as in Recruitment [13], where algorithms are increasingly being used to score measures [14], and in Facial Recognition [15], a technology that has previously attracted attention due to gender and racial disparities in accuracy [16].

As is clear from a cursory reading of these regulatory proposals, the primary concern is with mitigation and reduction of harm (e.g., EU Risk Tier approach). However, invariably there is a countercurrent in the respective regulatory proposals regarding the need to ensure that the opportunity that AI represents is suitably harnessed. Here we can refer to the regulatory proposals within the context of national data and AI strategies, which embed AI governance (where we read AI governance as the umbrella term encompassing the regulation, policy and standards debate) within broader economic and strategic interest (e.g., UK, EU). As such, we can think of the AI governance agenda as the mediation of harm and opportunity (c.f. EU's original white paper named the advancement towards an 'Ecosystem of Excellence' and 'Ecosystem of Trust' as its two foundational pillars). In abrasto this is not a particularly insightful observation; indeed, all legislative and governance agendas are a mediation of this tension. However, upon further inspection, when the notions of harm and opportunity are unpacked, and situated within a fleshed out notion of opportunity, it becomes clear that when considering the subtleties of method (how to govern, who should govern) and desired outcomes (motivated and realised by competing interests), the mediation is an expression of a value-system that serves stakeholders (citizens, legislations, industry - small and large scale -, research) to differing extents.

In this paper, we will survey regulatory approaches and explore the question of What is Centering the Public Interest in AI? - by probing what we are referring to as the centre of gravity that forms the nucleation point of mediating harm and opportunity. In order to explore this question, we begin with a high-level excursion into the current proposed legal frameworks and proposals. We do so both generically (via the notions of horizontal, vertical and hybrid) and through exposition of case-studies (via exploring the EU as horizontal, UK as Vertical and USA as Hybrid). Following this we focus on the UK's sectoral approach and then offer a brief exposition of the proposed EU AI Act.

Finally, we will tease out some non-sectoral regulator's approach from the UK and use this to tackle the question of public interest. By surveying the competing interests at play in AI regulation, we confirm the 'wickedness' of discerning a coherent view of the public interest in the absence of any guidance. To this end, we suggest that there is a need for formal guidelines for the development of AI regulation.

Before moving forward we make two notes on how we are using the term 'AI' and the relation between AI and Data Protection.

- **AI:** that when we speak of AI we do so in broad terms that appeals to the lay / non-technical vernacular, rather than appealing strictly to the engineering literature [17]. We do so because we take the regulatory discussion to fall within this usage of the term. More specifically we read this as a reference to the use of algorithms in the automation of traditionally human decision making contexts.
- **AI and Data Protection:** The relationship between existing data protection laws and governance procedures is important to understanding the path dependency of proposed AI legislation and governance - this is due to the intricate relationship between data and AI [1]. It is also important to understand that the body of existing data protection legislation can, where appropriate, fall within the remit of 'AI' broadly understood. This is particularly the case when 'data processing' concerns are considered; in practice, this is likely to vary considerably on a case by case basis although, for the purposes

of illustration, we have generalised our approach in order to provide an overview and highlight the potential complexity with the existing legal framework. Building on the discussions of existing regulation above, determining whether AI datasets are personal data is one example. Where datasets constitute personal data as defined by data protection laws, they will be governed accordingly, although this may not always be a straightforward determination to make; information that has been pseudonymised that could be attributed to a person with the use of additional information should also be considered as information about an identifiable natural person [18] (Recital 26, GDPR). The ability to ascertain whether individuals can be identified from a combination of datasets is likely to be challenging and should be one of the drivers for good governance around scoping and using datasets. Nevertheless, there will be projects that are unlikely to fall within data protection laws and the UK Government's Guide to Using Artificial Intelligence in the Public Sector [19] provides a number of examples of AI projects that are very unlikely to require the processing of personal data.

Key Takeaways

- The UK has intimated that it will adopt a sectoral approach to AI regulation
- The absence of a cross-sector framework presents critical interpretive problems
- There is a need for a cross-sector framework to guide AI regulation
- A cross-sector framework must balance the competing interests that together compose the 'public interest'

Table 1. An overview of key regulators of AI in different sectors

Sector	Relevant Regulators	What are they Regulating?	Themes
Education	The Office of Qualifications and Examinations Regulation (Ofqual); Office for Standards in Education, Children's Services and Skills (Ofsted)	The use of AI in exam marking; The potential use of AI in school inspections	Automation; prediction
Business/Finance	Bank of England (BoE); Financial Conduct Authority (FCA)	The FCA-BoE AI Public-Private Forum created to facilitate the transfer of knowledge between the public and private sectors and AI	Automation; prediction
Commerce	Competition and Markets Authority (CMA)	The use of AI in reducing competition and harm to consumers	Minimisation of harm; personalisation; competition
Health	Care Quality Commission (CQC); Medicines and Healthcare products Regulatory Agency (MHRA)	The use of machine learning in diagnostic services	Automation; minimisation of harm; training of users
Utilities	Office of the Gas and Electricity Markets (Ofgem)	The use of machine learning to provide insight into energy demand and plan investments	Prediction; insights; funding
User-to-user platforms	Office of Communications (Ofcom)	The use of AI to minimise harmful content available online	Automation; minimisation of harm
Tax	HM Revenue and Customs (HMRC)	Potential use of AI to ensure compliance and assist with complex tax returns, along with a virtual assistant	Automation; assistance; support
Data	Information Commissioner's Office (ICO)	Technical and governance aspects to risk assessment; advice on explaining algorithmic decisions	Risk assessment; explanation
Rights	Equality and Human Rights Commission (EHRC)	Guidance on the protection of sensitive attributes	Minimisation of harm
Policing / Security	Police Commission; EHRC	Predicting the occurrence of certain crimes and behaviours and facial recognition technology	Prediction; minimisation of harm
Gambling	Gambling Commission	The use of nudging in the context of gambling	Minimisation of harm
Audit	National Audit Office (NAO)	Regulation of finances to enable adequate funding for the use of AI in the running of public bodies	Funding

2. Legal Frameworks

In this section we will sketch three (broad) approaches to regulating AI - namely Horizontal, Vertical and Hybrid. For each we explain how we define it, provide an example, and note advantages and disadvantages. This section thereby sets the conceptual landscape from which we offer our commentary and excursions in subsequent sections.

- **Horizontal:** legislature can enact wide, sector agnostic regulation that encompasses all uses of AI. Such legislation is top-down insofar as it starts with as wide a remit as possible and specificity is brought about through application in differing use-cases, industries, and sectors.
 - Example: the proposed EU AI Act [20] is a piece of legislation that is currently under consideration by the European Commission. It represents the most mature and significant development in the AI legislative agenda globally. Widely recognised as dominating the debate, it is appropriate to refer to the proposal as the central reference and comparative point. We explicate the proposal in later sections, however, for the purposes of this section it suffices to say that with respect to the executive, it is a horizontal form of regulation because it applies across the union and without prejudice to industry or sector.
 - Advantages: the proposal is clear, internally coherent and forms a solid foundation for further debate and iteration by the very fact that the proposal is centralised.
 - Disadvantages: may fail to address particularities that stem from particular sectors (for example, already there has been significant revision via pressures from the national security communities of respective countries and research institutions); There is a real danger that the legislation fails to recognise differences in realms that are too far apart and distinct to be dealt with together (think of AI in security and AI in advertising); another disadvantage is that getting it wrong at the horizontal level is highly costly, both in terms of failing to meet its aims and taking remedial actions through amendment; finally, the legislation is in danger of being diluted by too many (irreconcilable) differences and interests - thereby leading to no-one wins scenarios.

- **Vertical:** legislature can refrain from sweeping regulation and instead mandate (or exercise existing mandates) by sector regulators. Such an approach is a bottom-up approach and starts with the application/ industry to build up an agenda intimately related to the specificities of a particular sector.
 - Example: since 2018 there has been a strong current within the UK policy and regulatory ecosystem that the UK can take an approach that reflects a more pragmatic and innovation friendly approach to AI standards and regulation. In a kind of 'Brand Britain' / 'Made in GB' [21] approach, the UK is to leverage itself as a credible regulatory ecosystem whilst being concerned with respect to a high-standard. With the move away from the EU, the UK has indeed signalled, in the digital realm, that there is an appetite to propose an alternative approach to (digital/technology) regulation. This has manifested itself through the activity of individual regulators publishing and signalling in the space of AI. Although there are signals of an alternative UK approach (one that feels closer to a horizontal approach) [22] we read the current stand as imbuing a sectorial stance.
 - Advantages: a sectoral approach is agile and respects the nuances and complexities of AI in various uses and industries - this also coheres with the fact that these systems are dynamic and rapidly evolving (both in technological terms and in terms of deployment). It is also advantageous to have dedicated resources addressing various sectors and idiosyncrasies; such sector regulators already have sectoral sentience, experience and relationships that they can use to effectively regulate and enforce any proposed regulation. Sectoral regulation might already exist that can be adapted and expanded to AI use cases.
 - a sectoral approach may encourage 'shopping' by industry players who are seeing the least active and/or 'lightest' form of regulation - this is a disadvantage as it encourages bad-faith actors; there may be a significant amount of overlap between regulations resulting in a multiplicity of responsible agents, overlap and undue burden upon relevant parties; a single party may be required to report to multiple regulators regarding the same thing; the regulators may protect their autonomy and remit, thereby restarting knowledge transfer and enforcement across sectors.

- *Hybrid*: legislature can decide upon what horizontal and vertical regulation is appropriate and thus legislate at the appropriate level. Here the dimensions are two fold with respect to the nature of the legislative intervention. For example, hybrid may involve horizontal and vertical legislation, but it also involves legislation by differing legislative bodies i.e., federal vs state legislation.

- *Example*: in the United States of America (USA) [23–25] there are both federal as well as state level legislative proposals that relate directly and indirectly to AI. At the federal level there are initiatives regarding AI and national security (military use, intellectual property theft, procurement and selling) - indeed, recently Democratic legislators have proposed an Algorithm Accountability Act [8]. At a state level there are piecemeal interventions, with Illinois introducing the Artificial Intelligence Video Interview Act [26], which requires that employers analysing video interviews inform candidates of the characteristics that will be used to inform employment decisions and get consent in writing before commencing the interview. Further, perhaps the most high-profile example of an intervention has been the mandatory bias audits of AI in talent management in New York City [27], with frameworks being developed to guide audits of algorithmic recruitment systems [28].

- *Advantages*: respect for the sovereignty of state legislatures as point of principle in liberal and democratic federalism has a value in itself - one expression of that is that the related regulation will respect the popular will at more 'local' level and the federal governments' (assuming a positive value is ascribed to this i.e. respect for subset communities within a pluralistic state); this approach is likely to be more agile than exclusively horizontal and vertical approaches; this approach is also likely to be faster, where speed may be valued given the rapid pace of change and impact of said technologies and the need to react in a timely manner.

- *Disadvantages*: the hybrid approach is ripe for contradiction and confusion, where there is likely to be clarity on both the remit and the demarcation of what can and cannot be legislated for at a federal and state level; given the nature of such technologies (in the fact that they seldom respect geographic boundaries) there is little sense in federal/state distinctions unless clauses regarding procurement origin, use geography, dual-use, sell-on, etc. are introduced - all causing further

complications; this is likely to lead to further burden and confusion regarding applicability of laws in context of use.

The above serves to canvas the conceptual landscape and surface our principal case-studies, namely the EU AI act and the UK's AI governance ecosystem.

In the next two sections we will expand, in greater detail, on the case-studies of the EU and UK. Our primary case-study is the UK because we believe that it presents the legislative ecosystem that has the most mature developments in the vertical (sectoral) approach and it is the jurisdiction that the authors primarily operate within. As such, we will begin with an overview of the EU's approach and use this to situate the UK's difference.

3. EU AI Act - Horizontal par excellence?

Notwithstanding the current lack of explicit 'AI regulation', there is an active debate with proposed legislation. The most substantive of these is the European Commission's recently published draft of the first ever legal framework focused solely on AI [20]. Here we draw on our work in [29, 30]. This draft legislation proposes a risk based approach to AI governance. There are four tiers::

- *Limited risk*: concerns systems that do not pose a threat to the safety and livelihood of persons. Action in this context pertains to 'transparency obligations', i.e., users should be aware that they are interacting with a machine in order to make an informed decision (article 52). Here self-regulation and mechanisms of adhering to codes of practice are appropriate.
- *Minimal risk*: concerns systems that do not pose a threat to the safety and livelihood of persons. Here the right to opt out of the use of such technologies and transparency provisions (e.g., ensuring users are aware they are interacting with a machine), suffice. No action is necessary in this context and it is envisioned that the vast majority of systems will fall into this category. We have grouped these together and will not treat them further as they are of least concern to our interests in this white paper.
- *High-risk*: Here a general criterion is not offered, instead examples of sectors and applications are given (expanded upon below and corresponding to Title III, Chapter 2, articles 9-15). In Annex III to the draft legislation a list of high-risk systems, along with explanandum, is given. Here we list them without expansion:

Biometric identification and categorisation of natural persons; Management and operation of critical infrastructure; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum and border control management; administration of justice and democratic processes; insurance [31].

We infer from the case studies that, similar to unacceptable risk, such systems pose a threat to the safety and livelihood of persons, however, in these cases there are benefits that can be derived and used to justify

deployment through good governance/risk management. In such high-risk cases a number of (legal) requirements are stipulated in terms of justifying the use of these high-risk systems. Indeed, article 9 asserts the need to establish a 'risk management system' that must be acted upon and maintained, including adequate documentation. It is suggested that this is a 'continuous iterative process run throughout the entire [high-risk system's] lifecycle'. Following this, articles 10-15 denote, in more detail, the conditions that have to be met for a system to be justified for use.

- Data and data governance (article 10): Training, validation and testing data sets to ensure that they are of high quality data.
- Documentation (article 11, 12): Provide detailed documentation for third party assessment, including technical documentation and record-keeping i.e., period logging of standards specifications being met.
- Transparency for users (article 13): Comprehensible information regarding contact details of provider, purpose, accuracy, security, data used, human-oversight measures and expected life-cycle of a system should be reported.
- Human-oversight (article 14): Must ensure high-level of human oversight in development and deployment, through appropriate interfaces. The overseers must be able to understand the capacities and limitations of a system, avoid automatically accepting recommendations of a system, and be able to intervene effectively. Decisions should be taken after at least two people have overseen the system.
- Accuracy, robustness and cybersecurity (article 15): such relevant metrics must be declared, including failsafe mechanisms, mitigation strategies against vulnerabilities and for cybersecurity attacks.

- *Unacceptable risk*: Here concern is with systems that pose a direct and clear threat to the safety, livelihoods and rights of people. The action for such systems is an outright ban. Three use cases are named, these are:
 - Social Scoring systems: in opposition to systems that have been used in China, inferring character judgements from social behaviour is banned. Cases where a person incurs traffic incidents or engages in other kinds of antisocial behaviour should have no bearing on other (public) services/benefits they may receive.
 - Manipulation: Social Scoring systems: in opposition to systems that have been used in China, inferring character judgements from social behaviour is banned. Cases where a person incurs traffic incidents or engages

in other kinds of antisocial behaviour should have no bearing on other (public) services/benefits they may receive.

- Remote biometrics: the use of indiscriminate scanning and use of identifiable characteristics (e.g., facial recognition, audio scanning, sentiment analysis in the public sphere, etc.) are banned. The qualifier 'remote' is used to indicate that individual and consensual use of such systems is fine i.e., logging in via fingerprint, face, voice, etc.

In the context of our discussions of horizontal and vertical approaches, there is a key thread worth teasing out. This is, namely, that the approach leans heavily into a case study - that no necessary and sufficient conditions were articulated for high-risk systems suggests a significant lacuna in (at least theoretically) conceptualising 'high-risk'. Ironically, the draft can be read as a horizontal approach that is embedded within attributes of a vertical approach. Where the case studies (sectors and use cases) represent the kinds of things one can imagine falling under the remit of a sector regulator. Indeed, the approach begs questions regarding reporting, enforcement and knowledge sharing i.e., the disadvantages of the sectoral approach.

It is necessary to pause on this point and speculate as to why this is the case. One reason could be because the approach reflects the tensions of various inputs and thus, with such a composition, the presence of a multiplicity

of sectors and use-cases is inevitable. Another, perhaps more compelling reason, is because *regulating AI is a particularly hard problem*. Some reasons for this are to do with the complexity of the subject matter (it is not trivial to track and digest developments in AI) and the fact that it is an ever evolving field.

This speculative point also presents a counterpoint to claims that the horizontal approach is clearly a better approach i.e. that there are strong arguments in favour of a horizontal approach. As we have seen with the EU's proposed AI Act, this is far easier said than done and manifests in a proposal that has had to accommodate what a sector purportedly covers.

Rather than positioning the UK's approach, which we present as a case-study of the vertical legislative intervention, as simply 'anti-European', we take the sectoral approach on its own merits.



COLLABORATE

4. The UK Approach

This section is comprised of two subsections, namely:

Sector Mapping and Sectoral regulators: this non-exhaustive mapping of sectors and sectoral regulators in the UK context serves to highlight the scope and expanse of the problem of regulating AI.

Horizontal UK Approach: in this section we cursorily treat what we read as examples from the UK AI governance ecosystem as potentially representing possible horizontal approaches.

4.1 Sector Mapping

As a point of departure, we must note that it is not straightforward to accurately map the UK regulatory landscape (in terms of the regulatory remit, scope and enforcement jurisdiction). Some (most that is relevant in this area) regulation is reserved to the UK Parliament but some is devolved and therefore there may be differences in Northern Ireland, Scotland and Wales. In the penultimate section we turn to the public interest and this is likely to be particularly relevant to discussions around public interest where there will be devolved legislation as well as UK legislation that needs to take the public interest into consideration. Two examples are the Freedom of Information (Scotland) Act 2002 and the Environmental Information (Scotland) Regulations 2004, with some exemptions and exceptions being subject to the public interest. These legislation are both regulated by the Scottish Information Commissioner (not to be confused with the Scottish Office of the ICO).

(2015) (RPSI), which are on occasion used to obtain datasets, remain a reserved matter and are regulated by the ICO. Decisions under these regulations are therefore potentially complex as RPSI does not apply to information that would be exempt under FOI or EIR legislation (across the UK), and consequently both the SIC and the ICO could be involved in a decision on a particular case. RPSI is intended to encourage the re-use of public sector information [32] although it is unclear to what extent they are regularly used as the source of data for AI projects. Continuing with the public sector as an example, another example would be the National Audit Office and Audit Scotland. They perform similar roles in different jurisdictions and could potentially oversee very similar initiatives, although it is difficult to identify a framework

(legal or otherwise) that would ensure that there was any consistent oversight with respect to approaches to AI governance.

A final example that is relevant in this context is the Equalities and Human Rights Commission (EHRC), who has a responsibility to "...encourage equality and diversity, eliminate unlawful discrimination, and protect and promote the human rights of everyone in Britain" [33]. They demonstrate further complexity with regards to a legal framework, having a jurisdiction to enforce equalities legislation across the whole of the UK under Equalities Act (2010) but also being required to monitor compliance with subordinate legislation passed by the devolved parliaments.

The potentially adverse impact of AI on the protection of fundamental rights has been recognised by the EU's proposal for legislation referenced above and is discussed further below in the section on public interest, although our research suggests that we have yet to see any substantive guidance from the EHRC about how equalities legislation interfaces or interacts with the issues highlighted by AI. This is discussed further below.

4.1.1 Sector Survey

There are 90 sector regulators in the UK, with varying degrees of scope, reach and remit. In this section we map what we consider to be the crucial regulators with respect to AI regulation. This is a non-exhaustive list and serves to both highlight what a sectoral approach to AI regulation would look like (including relevant areas of concern) and also to indicate activity that may have already taken place by the said regulator. Table 1 has an overview of some key regulators in different sectors, and what they are regulating in terms of the applications of AI, which we expand upon below.

Table 1. An overview of key regulators of AI in different sectors

Sector	Relevant Regulators	What are they Regulating?	Themes
Education	The Office of Qualifications and Examinations Regulation (Ofqual); Office for Standards in Education, Children's Services and Skills (Ofsted)	The use of AI in exam marking; The potential use of AI in school inspections	Automation; prediction
Business/Finance	Bank of England (BoE); Financial Conduct Authority (FCA)	The FCA-BoE AI Public-Private Forum created to facilitate the transfer of knowledge between the public and private sectors and AI	Automation; prediction
Commerce	Competition and Markets Authority (CMA)	The use of AI in reducing competition and harm to consumers	Minimisation of harm; personalisation; competition
Health	Care Quality Commission (CQC); Medicines and Healthcare products Regulatory Agency (MHRA)	The use of machine learning in diagnostic services	Automation; minimisation of harm; training of users
Utilities	Office of the Gas and Electricity Markets (Ofgem)	The use of machine learning to provide insight into energy demand and plan investments	Prediction; insights; funding
User-to-user platforms	Office of Communications (Ofcom)	The use of AI to minimise harmful content available online	Automation; minimisation of harm
Tax	HM Revenue and Customs (HMRC)	Potential use of AI to ensure compliance and assist with complex tax returns, along with a virtual assistant	Automation; assistance; support
Data	Information Commissioner's Office (ICO)	Technical and governance aspects to risk assessment; advice on explaining algorithmic decisions	Risk assessment; explanation
Rights	Equality and Human Rights Commission (EHRC)	Guidance on the protection of sensitive attributes	Minimisation of harm
Policing / Security	Police Commission; EHRC	Predicting the occurrence of certain crimes and behaviours and facial recognition technology	Prediction; minimisation of harm
Gambling	Gambling Commission	The use of nudging in the context of gambling	Minimisation of harm
Audit	National Audit Office (NAO)	Regulation of finances to enable adequate funding for the use of AI in the running of public bodies	Funding

- Education:

- *Relevant Regulators:* Ofqual; Ofsted.
- new digital technologies, in particular AI, are rapidly being researched and deployed in education, representing acute ethical risks in particular when the target group are children [34, 35]. For example, it has been proposed that Ofsted inspectors could be replaced by algorithms [36]. A more infamous example is the use of algorithms to score GCSE and A-levels for students in the Covid-19 pandemic [37], which was extremely controversial, particularly since some students received lower grades than was expected. To the best of our knowledge, sector regulators have not publicly signalled any position or activity, however there is an active debate about what appropriate forms of intervention should be targeted [38].

- Business/Finance/Commerce

- *Relevant Regulators:* Bank of England (BoE); Financial Conduct Authority (FCA); Competition and Markets Authority (CMA).
- Note: the sector of business, finance and commerce is a particularly active space regarding regulation and standards. Indeed, irrespective of AI, these are areas that are heavily regulated. Examples of activity include the FCA-BoE AI forum [39], which seeks, among other things, to 'Gather views on potential areas where principles, guidance or good practice examples could be useful in supporting safe adoption of these technologies'. More explicitly the CMA have published [40] on the role that algorithms can play in reducing competition and harming consumers. In this piece, there is an explicit discussion of the role of regulators and a call for further empirical research [41]. See also [42, 43].

- Health

- *Relevant Regulators:* Care Quality Commission (CQC); Medicines and Healthcare products Regulatory Agency (MHRA)
- healthcare is an acute realm of concern given the nature of the potential harm. This aligns with the significant levels of activity with respect to AI development and use in the healthcare sector, including both delivery of care and management of systems. Within this context the CQC and the MHRA conducted a trial regulatory sandbox and published their findings [12], which included recommendations for strong governance and effective regulation of suppliers.

- Utilities

- *Relevant Regulators:* Office of the Gas and Electricity Markets (Ofgem)
- Note: energy demand has historically been predictable due to its stability but recent shifts towards demand for low carbon electricity and the instability of renewable energy availability has led to trials of machine learning software to provide data-driven insights into demand. Through the deployment of these programs, it is hoped that investment in infrastructure can be targeted to facilitate flexibility in distribution of low carbon energy, towards the government's targets of being net zero by 2050 [44]. Although, to the best of our knowledge, how this application of AI will be regulated has not been publicly announced, Ofgem are planning on taking steps towards greater transparency and require all the regulated network companies to publish their digitalisation strategies [45].

- Social media

- *Relevant Regulators:* Office of Communications (Ofcom)
- Note: it has been proposed that AI can improve the moderation of online content at multiple stages of the moderation pipeline to judge whether content can be considered harmful, including before the content is posted for users to see, and when content is flagged as harmful by users. There are a number of proposed benefits to this approach to regulating content, including increased accuracy in identifying harmful content, increasing human productivity by removing this burden from them, and the protection of moderators from exposure to harmful content [46].

- Tax

- *Relevant Regulators:* HMRC - HM Revenue and Customs
- Note: HMRC was reportedly exploring the potential for AI to be used to ensure compliance and assist with complex tax returns [47], although a progress update has not been given since the announcement in 2018. However, they have introduced a virtual assistant Rita to assist customers with inquiries [48]. While, to the best of our knowledge, HMRC have not provided much detail how they plan to regulate this, we speculate that there are some internal steps being taken towards this.

- Data

- *Relevant Regulators:* ICO

- Note: according to our reading, the ICO is the leading regulator with respect to their output and activity concerning appropriate AI governance. A natural reason for this is the necessary and intimate relationship between data and AI, where one can think of data as enabling AI. We may highlight two principal texts, ■ 'Guidance on the AI auditing framework: Draft guidance for consultation' here the aim was to produce guidance that encompasses both technical (e.g., system impact assessments) and governance (e.g., Reporting, assessing, human oversight) dimensions to risk assessing AI [49].

- Co-badged by the Information Commissioner's Office and The Alan Turing Institute, 'Explaining decisions made with AI' [5] is an expansive excursion into the practical translation of calls for accountability and transparency in organisations that use AI systems. The report aims to give organisations practical advice to help explain the processes, services and decisions delivered or assisted by AI, to the individuals affected by them. The guidance is divided into three sections, namely: Part 1 The basics of explaining AI; Part 2 Explaining AI in practice; and, Part 3 Explaining AI means for an organisation [50].

- Given the significance of the ICO in this space, foreshadowing the following section somewhat, should the UK decide to take a horizontal approach the ICO is best to assume the role (disregarding the possibility of a new AI regulator being established [51].

- There are other regulators within the UK enforcing rights in relation to data, information and records. They have not been discussed here as their current significance to the AI agenda appears to be minimal, although this could change and evolve with time as AI use becomes more widespread, particularly by public bodies.

- Rights

- *Relevant Regulators:* EHRC - Equality and Human Rights Commission

- Note: As discussed above, the EHRC enforces equalities legislation across the UK and works to protect and promote human rights. Equalities legislation relates to nine protected characteristics defined by equalities legislation and while some of these protected characteristics are special categories of personal data under data protection laws, it should be noted that some are not. The Government Response to the Committee

on Standards in Public Life's 2020 Report on AI and Public Standards [52] made a recommendation that the EHRC public guidance on "AI and Equalities Act" and the response to the recommendation states that the EHRC will be developing guidance for public bodies in conjunction with others in the next financial year. Such a document will be helpful in shaping the direction of travel although, as it will be for public bodies only, leaves a significant gap for organisations other than public bodies including those looking to exploit AI for commercial gain.

- Policing / Security

- *Relevant Regulators:* Police Commission; EHRC

- Although controversial due to claims of racism and disparities in the accuracy of technology for darker skinned individuals [16], AI is being used in policing to predict the occurrence of certain crimes, behaviours and characteristics, as well as in facial recognition. The application of these technologies is scrutinised by the EHRC, who have raised concerns about a lack of transparency and testing for bias, as well as over-reliance on automation [53].

- Gambling

- *Relevant Regulators:* Gambling Commission

- Note: in broad terms much of the ethical debate and concern has centred on algorithmic nudging [54], i.e., manipulation of persons in order to realise a particular end in their action. The most common example of this is the use of AI in advertising. A particularly ethically acute example of algorithmic nudging is the use of AI in the gambling industry. Where gambling is considered an addition, this falls into wider public health concerns. Here the concerns centre on design (luring people, including the vulnerable onto platforms) and maximizing time spent/money spent on such platforms [55].

- Audit

- *Relevant Regulators:* National Audit Office (NAO)

- Note: The NAO is the UK's independent public spending watchdog and "...helps Parliament hold the government to account for the way it spends public money" [56]. While it is unlikely to have a significant role in regulating AI, the role it has will remain vital in ensuring value for money with public sector projects, which is likely to require a focus in the AI context on whether projects are achieving the required outcomes and, for most, supporting the statutory a public body's statutory purpose. Indeed, AI is likely to increasingly be used to

help improve the running of public bodies in areas such as the detection and prevention of fraud, where the sharing of information for this purpose already has some enabling legislation under Part 5 of the Digital Economy Act 2017.

4.2 Teasing out a Horizontal Approach?

In the introduction we have read and interpreted the UK's position as leaning towards a sectorial/vertical approach to AI regulation. In the previous subsection we have cursorily explored what this may mean to particular sectors and their associated regulators. However, according to our reading there is also a narrative that can be forwarded that has the UK as positioning itself for horizontal regulation. To be clear, the language of 'the UK positioning itself' is not a claim that there is a coherent and singular intentional force driving the agenda. Indeed, naturally, there are competing interests and forces that motivate their respective agendas, seeking to realise aims that conform to what they would envision as appropriate and adequate AI governance (whether that be in the form of self-regulation, representation, regulation (horizontal and vertical), etc.). Below we will tease out a narrative that can be thought of as implicitly (if not explicitly) leaving open the possibility of a horizontal form of UK regulation. The documents we will look at are:

- UK National AI Strategy
- AI Assurance Roadmap
- Public Sector AI Transparency Standard

4.2.1 UK National AI Strategy

The UK's National AI Strategy [57] can be seen in terms of a (potential) horizontal approach to regulation. We read this into the text because of the sections within the strategy that relate to AI governance. In fact one of the three pillars of the text is 'Governing AI effectively', which has two themes woven into it, namely stimulating innovation and enterprise, and a standards and regulatory regime that reflects this innovation agenda [58]. Within this framework, drawing on the work of [58] there are a number of themes namely that:

- removing some existing regulatory burdens where there is evidence they are creating unnecessary barriers to innovation
- retaining the existing sector-based approach, ensuring that individual regulators are empowered to work flexibly within their own remits to ensure AI delivers the right outcomes.

- introducing additional cross-sector principles or rules, specific to AI, to supplement the role of individual regulators to enable more consistency across existing regimes.

Note that these themes read as an endorsement of the sectoral approach i.e., contrary to the argument we are making in this section. In fact, as a corollary to the final point, as we have noted above, we have read the ecosystem as signalling that the UK will take a sector-specific approach, as such the inclusion of the cross-sector regulatory policy is to be read as an important shift. Further, the content and realisation of the themes are all congruent with a horizontal approach. Regarding removing existing regulatory burdens, the reading can quite readily be one of moving towards a more principled, national level, regarding regulation. Regarding retaining the sector-based approach, a reading can be made that appeals to the imperative that the regulators should be flexible i.e., the sector regulator burdens may be in setting standards and monitoring rather than in codification of a specific regulatory requirement - something that would readily align with a horizontal approach. As such, an interesting argument can be given that the sectorial/vertical approach can serve a horizontal agenda insofar as the associated regulators of the sectorial/vertical serve as effective policy, standards and monitoring institutions.

4.2.2 Assurance Roadmap

Mentioned in the UK's National AI Strategy text (discussed above), the Centre for Data Ethics and Innovation (CDEI) were mandated to publish an AI Assurance Roadmap. The roadmap is a detailed excursion of both the state of play regarding AI assurance and an intervention with one intent being to demarcate a robust UK "AI assurance ecosystem". We read this text within the discussion of potentially supporting a horizontal approach to regulation; we do primarily because the text is clear insofar as the aim is to realise mechanisms of ensuring trustworthy AI to establish trust and set the conditions for AI innovation, industry and adoption (very much in the AI as an opportunity genre). In the section that directly addresses 'The Role of Regulators', instruction is given regarding the need for the regulators to provide guidance and expertise, reflecting the nuances of the respective sectors. In closing the section the text reads 'as the use of AI systems becomes more widespread across sectors, clear regulatory scope will need to be established between regulators with similar

and overlapping mandates. On top of this, regulators will need to decide in what contexts assurance be directly sought by regulators (regulatory inspection) and when assurance should be delegated to assurance providers, or where assumed conformity is appropriate (emphasis ours)' - here the reading is one of indeterminacy. Although there is clearly an implicit view that the landscape will evolve in a sectorial one (insofar as the sectoral regulators and assurance bodies will be tasked in their respective domains), it is still an open question how this will translate to the specificity of regulatory regimes with respect to specific sectors (assurance and compliance are not one and the same thing).

4.2.3 Public Sector AI Transparency Standard

Published by the Central Digital and Data Office (CDDO) - part of the UK Cabinet Office - the Algorithmic Transparency Standard seeks to forward the AI standards agenda via the power of the public sector as exemplar. It is a tiered approach that encourages transparency as targeted to distinct audiences, with a view to facilitating trust. For our purposes, most relevant is that there is an implicit use of government procurement in driving standards through industry (through purchasing power); the core aim is public transparency, where the two-tier approach can be thought of as a reasonable approach for catering to distinct audiences and addressing the problem of communicating to different stakeholders. Most critically, there are calls for pilots and testing, all of which is sector and use-case agnostic [59]. In other words there is nothing particular that lends this to a vertical approach - through codification this can readily lend itself to a horizontal approach.

In concluding this section we note that there are plenty of policy and regulatory proposals that are ostensibly rooted in a particular sector or with particular concern (both in terms of specific regulators and in terms of special interest, such as future of work). However, as we have suggested, despite being framed in terms of particular sectors and sets of interests, these proposals can readily be appropriated for use in a horizontal approach. Indeed, the proposed regulatory solutions, such as the provisions for innovation enabling governance (UK AI Strategy), generating trust (AI Assurance Roadmap) and transparency and accountability requirements (Public Sector Standard), can readily be adapted for the purposes of horizontal governance. The same is true of proposals we have not considered here: for instance, the

Institute for the Future of Work's proposed framework for algorithmic impact assessments can also be applied through horizontal regulation [22], as has been proposed in the US Algorithmic Accountability Act (Wyden et al., 2022).

5. Critique of Vertical vs Horizontal

The question of public interest is also present in legislators' choice of legal mechanism for regulating AI. Saliently, legislators must decide whether to use horizontal regulation that emanates from the central government and apply to applications of AI across all sectors, or vertical regime in which industry regulators regulate AI according to the specific circumstances of their sector. We contend that this question has significant implications for public interest [Trengove & Kazim, 2022].

In its proposed AI Act, the EU has adopted a form of horizontal regulation, as part of its rights-oriented approach to AI. The benefit of this type of regime, from a public interest perspective, is that it guarantees a minimum set of safeguards against rights-violations by AI that protect all members of a state. By contrast, the UK has suggested that it favours a vertical approach, which will likely leave industry actors with greater freedom since they are unencumbered by central regulation. Such a vertical regulatory regime offers no universal safeguard for rights, but it does provide the flexibility for industry regulators to adapt regulations to meet the specific needs of their industry, encouraging greater innovation and enterprise.

It is the imperative of a government to balance the cultivation of innovation and economic growth as well as the protection of human rights. As we argue in the following section, these conflicting goals can all fall within the remit of the 'public interest', since the latter value is composed of the competing interests of different actors and groups within a state.

6. What is Driving Public Interest?

The concept of determining public interest is core to many of the legal frameworks in the UK and further afield however it is often referenced in debates about AI without any further information to help us to understand whose interest is being represented and what it may look like.

The Attorney General Jeremy Wright QC MP [60] spoke on his role in relation to the public interest saying “It is essentially a way in which the Attorney ... upholds a well-functioning and fair justice system”. In the same speech, the Attorney General quoted an earlier Attorney General, Sir Elwyn Jones, who said “The Attorney is the protector ... of the public interest generally”. This view arguably differs slightly from the direction given by the judge in a case that was centred on public interest, the Ponting case in 1985, which was that “the public interest is what the government of the day says it is” [61]. This case concerned a whistleblower, Ponting, who was charged under the Official Secrets Act of 1911. Despite the jury being given the direction on public interest, Ponting was acquitted and it remains difficult to be clear about the tangible factors of what drives public interest when looking at the different applications in practice.

So what are the implications when considering where the public interest lies for AI? And are there any indicators to help us establish what the public interest means in practice? Ultimately, as suggested by the Attorney General, it is going to be for the courts to decide although it presents many challenges for progressing an AI agenda: how can each of the different regulators that have a role to play consistently determine and uphold the public interest? Without an overarching framework that creates a common understanding and interpretation of what represents the public interest in the context of AI, it is likely to be subject to many different interpretations. Considering the current landscape in the UK, the National Data Strategy [62] refers to public interest only when determining purposes, with the National AI Strategy [57] not referencing the public interest at all.

In contrast with the UK's approach, the EU's proposals reference public interest when discussing topics of “health and safety, consumer protection and the protection of other fundamental rights” which they term responsible innovation (p11, European Commission, 2021) [20]. The proposal recognises AI generating risks to the public interest as well as the need to balance public interest (and substantial public interest) in some circumstances and also begins to explore the relationship between risk

and factors that determine risk such as special categories of personal data, as discussed above.

Turning back to the UK, the way in which we address public interest needs to be multifaceted if it is going to protect and safeguard the various stakeholder groups that are identified below. Any framework to assist with determining public interest needs to be cognisant of the dynamism of AI; so many factors relating to the use of AI are unlikely to be static, from the dataset itself to the outcomes it is capable of producing. It also needs to ensure the desire to drive innovation and commercial development has appropriate checks and balances when considering the fundamental rights of individuals. One model may be to incorporate an approach similar to that in the EU's proposal of basing safeguards for different activities on risk, regardless of whether the legal framework is horizontal, vertical or hybrid. This will introduce its own challenges although such a framework is likely to assist with the consistent application of standards for safeguards providing a set of common risk indicators could be agreed. The profile of risk was increased significantly in the text of the GDPR in comparison to previous data protection legislation in the UK, with the GDPR providing some tangible identifiers for higher risk processing, activities, or datasets. As an example, Recital 38 [63] talks about the need for specific protection for children “as they may be less aware of the risks, consequences and safeguards” in relation to the processing of their personal data. There are other parts of the legislation that specifically refers to the need to be cognisant of risk, such as Article 32 (Security of Processing) [64], as well as recognition of the need for certain restrictive conditions to be in place to process special categories of personal data (Article 9). However, as discussed above, data protection is unlikely to provide a holistic approach: should, for example, all protected characteristics as defined by equalities legislation be considered higher risk data rather than using special categories? Additionally, should the conversation focus solely on special categories of data as opposed to the potential harms arising from the misuse of ‘standard’ types of personal data? The discussion is likely to become even more complex when thinking about activities: can all potential applications and uses for AI be fully understood and defined in a framework, and then consistently interpreted for practical application when it is evolving so quickly? Any framework would also need to be able to accurately reflect risks inherent in potential relationships between datasets that could be established through AI.

In order to address the public interest and create a robust regulatory framework, we need to be clear about from which perspective we are considering risk and where the motivations for AI development lie. Where the motivations are related to those organisations that are funding initiatives, they are likely, at least in part, to be commercially incentivised which is likely to require a different regulatory framework to other types of motivation. This brings us full circle to the existing regulatory framework and whether any of the existing regulators have the motivation or appetite to identify a more holistic solution and, in some cases, to even enter the AI debate. These questions sit at the centre of the regulatory debate, irrespective of the horizontal and/or sectorial approach.

Stakeholders

- *Public writ large*: here we refer to the general public and citizens, where, at a minimum the standards of human rights and civil liberties should be respected. In this domain there is an increasing concern amongst citizen groups regarding manipulation of the democratic process, surveillance, general use in security services and policing, consent, fairness and discrimination, and addiction. This can be cashed out both in terms of the 'public' and the 'consumer', where the impact is felt at a societal level and by non-technical, non-business (B2B) end-users.
- *NGO*: here we refer to organisations, citizen groups, unions, activists, journalists and any others speaking on behalf of, or purportedly representing, citizens i.e., without the individual citizen speaking on their own behalf. This category would have to be further unpacked and investigated in terms of what is motivating the non-governmental organisations (NGOs) themselves (who are their constituents), where funding is coming from, and to what extent can they really be speaking on behalf of, or purportedly representing, citizens.
- *Research*: research can be conducted independently, within the context of universities, think-tanks, citizen groups, NGOs and industry. Each institution (in cases where it is appropriate to speak of institutions) will have its own processes, procedures and ethical standpoint (implicit and explicit) and hence it is important to note that research takes place within a matrix of interests (and itself is an expression of value propositions - why was a particular problem selected? How was the question formulated? What evidence or supporting material was sought and how was this selected and justified?

Why and how was the information used to address the research woven together and contributed into a coherence? Who was the intended audience, why were they chosen (excluding analysis of whether this was implicit or explicit)? What are the intended consequences and/or possible unintended consequences of the research? The relevant here with respect to AI regulation concerns research as it pertains to shape and influence the horizontal/vertical forms of regulation, as well as the regulation of AI research. With respect to the latter this is likely to be a contested space where the intuition will be to protect research from undue burden, however, there is likely to be a difference of opinion when that research is being done in a university, or a start-up, or big-tech. For example, sufficiently large enterprises may be able to 'off-shore' (borderline unethical/contentious) research (the regulatory shopping concern flagged earlier) and then 'import' the compliant end-product.

- *SME*: small and medium-sized enterprises are critical for innovation and enterprise, often championed by governments and ecosystems as forms of dynamism, employment and growth in an economy. These stakeholders will both need guidance and understanding (that may be expensive and burdensome), as well as an ecosystem where they can experiment and 'disrupt'. Indeed, we can envision that a strong voice in this space may be to 'leave us alone: we are too small', which may be encouraging to big tech because they are the only ones who can afford it. As such the interests of SME are difficult to ascertain.
- *Big Tech*: the popular depiction of big tech is that they are interested in lowering any regulatory burden. Drawing on the influence and jurisdictional shopping of big tech with respect to tax efficiency, an analogy may be drawn with the interest of Big Tech lowering standards and regulatory burdens. However, given the investments and depth of engagements with research and ecosystems of talent in countries/locations that are likely to exist within robust regulatory regimes, this straightforward relationship between Big Tech interest and 'light' regulation is tenuous. A critical interest is likely to be the need for clear and universal standards - for example, it is likely that Big Tech will adopt the highest standard and then universalise it. A good example of this is the way European GDPR became the de facto global standard for data protection.

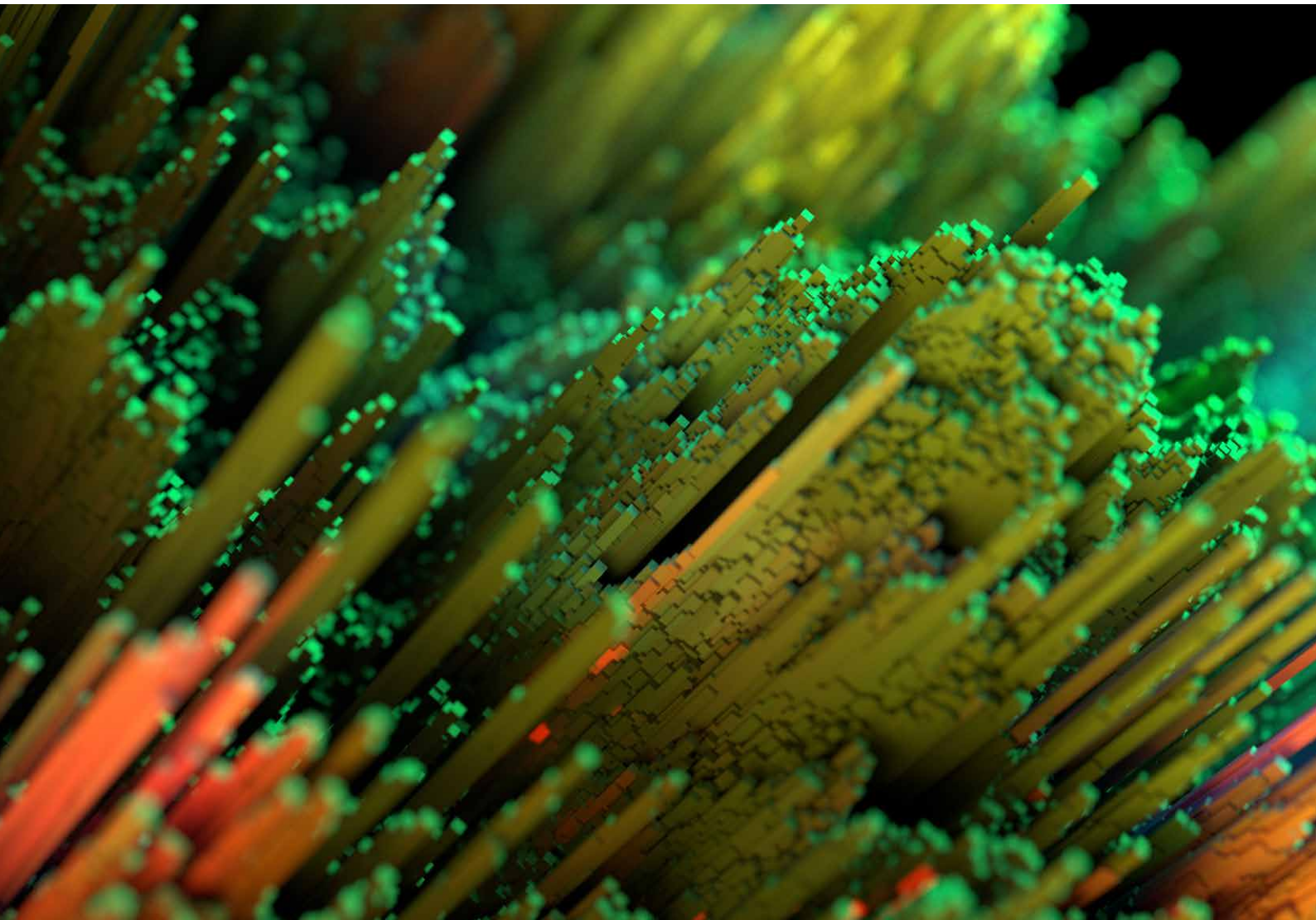
Serving the public interest is by nature a 'wicked' problem: it is a solution problem that, by its nature, inevitably requires trading off the interests of competing parties, with no right or wrong answers but only degrees of better and worse, and in which the problem and its causal implications cannot be mapped comprehensively or distinctly (Rittel and Webber, 1973). In this section, we have mapped some of the competing interests at play in the AI ecosystem. There are attempts to square this circle in AI legislation: the EU, for instance, suggests that its horizontal establishment of basic rights is a means of ensuring a 'human-centric' approach without overburdening the development of AI. Although this approach seeks to balance various competing interests, we cannot properly appraise whether it meets the public interest standard in the absence of a coherent framework for how these interests and preferences should be traded off.



COMMUNITY

Conclusion

In this article, we have examined the existing regulation, comparing the more horizontal (broad) approach of the European Union with the vertical (sector specific) approach from the United Kingdom. This examination confirms the 'wickedness' of the public interest: we demonstrated that the lack of an overarching framework that creates a common understanding and interpretation of what represents the public interest in the context of AI is likely to lead to many competing and contradictory interpretations. We recommended that in order to address the public interest and create a robust regulatory framework, we need to be clear about from which perspective we consider risk and where the motivations for AI development lie. In this analysis, we examined the positionality of relevant stakeholders including small and medium-sized enterprises and big tech. We concluded that there is a need for a coherent framework to guide how the competing interests that comprise the 'public interest' ought to be weighted.



Declaration of interest

The authors have no competing interests to declare

8. References

1. Kazim, E., & Koshiyama, A. (2021). The interrelation between data and AI ethics in the context of impact assessments. *AI and Ethics*, 1(3), 219–225. <https://doi.org/10.1007/s43681-020-00029-w>
2. OECD AI's live repository of over 260 AI strategies & policies. (2022). Retrieved April 5, 2022, from <https://oecd.ai/en/dashboards>
3. Government of Canada. (2021, March 22). Algorithmic Impact Assessment Tool. guidance. Retrieved January 24, 2022, from <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
4. Information Commissioner's Office. (2020, July 20). AI Auditing Framework. ICO. Retrieved April 21, 2022, from <https://ico.org.uk/about-the-ico/news-and-events/ai-auditing-framework/>
5. Information Commissioner's Office. (2021, December 1). Explaining decisions made with AI. ICO. Retrieved April 21, 2022, from <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/>
6. Information Commissioner's Office. (2022, February 28). Guidance on AI and data protection. ICO. Retrieved April 21, 2022, from <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/>
7. Department of Industry, Science, Energy and Resources. (2021, June 17). Australia's AI Ethics Principles. Department of Industry, Science, Energy and Resources. Text, Department of Industry, Science, Energy and Resources. Retrieved January 24, 2022, from <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>
8. Clarke, Y. D. Algorithmic Accountability Act of 2022 (2022). Retrieved from <https://www.congress.gov/bill/117th-congress/house-bill/6580/text>
9. Ministry of Economy, Trade and Industry. (2021). Governance Guidelines for Implementation of AI Principles. Retrieved from https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_9.pdf
10. Lopez, C. T. (2020). DOD Adopts 5 Principles of Artificial Intelligence Ethics. U.S. Department of Defense. Retrieved April 21, 2022, from <https://www.defense.gov/News/News-Stories/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/>
11. Mueller, H., & Ostmann, F. (2020, February 18). AI transparency in financial services – why, what, who and when? FCA Insight. Retrieved April 21, 2022, from <https://www.fca.org.uk/insight/ai-transparency-financial-services-why-what-who-and-when>
12. Care Quality Commission. (2020). Using machine learning in diagnostic services: A report with recommendations from CQC's regulatory sandbox. Retrieved from https://www.cqc.org.uk/sites/default/files/20200324%20CQC%20sandbox%20report_machine%20learning%20in%20diagnostic%20services.pdf
13. Centre for Data Ethics and Innovation. (2020). Review into bias in algorithmic decision-making. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf
14. Hilliard, A., Kazim, E., Bitsakis, T., & Leutner, F. (2022). Measuring Personality through Images: Validating a Forced-Choice Image-Based Assessment of the Big Five Personality Traits. *Journal of Intelligence*, 10(1), 12. <https://doi.org/10.3390/jintelligence10010012>
15. Centre for Data Ethics and Innovation. (2020). Snapshot Paper - Facial Recognition Technology. GOV.UK. Retrieved April 21, 2022, from <https://www.gov.uk/government/publications/cdei-publishes-briefing-paper-on-facial-recognition-technology/snapshot-paper-facial-recognition-technology>

16. Buolamwini, J. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification *. In Proceedings of Machine Learning Research (Vol. 81, pp. 1–15).
17. Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9), 100314. <https://doi.org/10.1016/j.patter.2021.100314>
18. European Union. (2016). Recital 26 - Not Applicable to Anonymous Data. General Data Protection Regulation (GDPR). Retrieved April 21, 2022, from <https://gdpr-info.eu/recitals/no-26/>
19. Central Digital and Data Office & Office for Artificial Intelligence. (2019). A guide to using artificial intelligence in the public sector. GOV.UK. Retrieved April 21, 2022, from <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>
20. Council of the European Union. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Presidency compromise text (2021). Retrieved from <https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf>
21. Kazim, E., & Koshiyama, A. (2019). Data Ethics Principles: A Comment on the House of Lords Report 'Regulating in a Digital World.' *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3581754>
22. Institute for the Future of Work. (2020). Executive briefing on an Accountability for Algorithms Act. Retrieved April 21, 2022, from <https://www.ifow.org/publications/executive-briefing-accountability-for-algorithms-act>
23. Hendrix, J. (2021, May 27). Senator Markey and Representative Matsui propose Algorithmic Justice and Online Transparency Act. Tech Policy Press. Retrieved April 21, 2022, from <https://techpolicy.press/senator-markey-and-representative-matsui-propose-algorithmic-justice-and-online-transparency-act/>
24. Steven Zeitchik. (2022). The battle to prevent another Jan. 6 features a new weapon: The algorithm. *Washington Post*. Retrieved from <https://www.washingtonpost.com/technology/2022/01/06/jan6-algorithms-prediction-violence/>
25. techregister. (2021, December 9). Oriole Park to test artificial intelligence security product. *Techregister*. Retrieved from <https://www.theintelligencer.com/sports/article/Oriole-Park-to-test-artificial-intelligence-16688241.php>
26. Illinois General Assembly. Artificial Intelligence Video Interview Act (2020). Retrieved from <https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68>
27. The New York City Council. Int 1894-2020. , Int 1894-2020 (2021). Retrieved from <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search>
28. Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., ... Lomas, E. (2021). Towards algorithm auditing: A survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3778998>
29. Kazim, E., Kerrigan, C., & Koshiyama, A. (2021). EU proposed AI legal framework. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3846898>
30. Polle, R., Kazim, E., Carvalho, G., Koshiyama, A., Inness, C., Knight, A., ... Goebel, R. (2021). Towards AI standards: Thought-leadership in AI legal, ethical and safety specifications through experimentation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3935987>
31. European Parliament. (2021). Procedure File: 2021/0106(COD). Retrieved April 21, 2022, from [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en)
32. Information Commissioner's Office. (2020, July 20). What is re-use of public sector information? ICO. Retrieved April 21, 2022, from <https://ico.org.uk/for-organisations/guide-to-rpsi/what-is-rpsi/>
33. Equality and Human Rights Commission. (2021). What we do. Retrieved April 21, 2022, from <https://www.equalityhumanrights.com/en/about-us/what-we-do>

34. Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIED): a high-level academic and industry note 2021. *AI and Ethics*, 2(1), 157–165. <https://doi.org/10.1007/s43681-021-00074-z>
35. The Institute for Ethical AI in Education. (2019). Towards a Shared Vision of Ethical AI in Education. Retrieved from <https://www.buckingham.ac.uk/wp-content/uploads/2020/02/Summary-The-Institute-for-Ethical-AI-in-Educations-Interim-Report-Towards-a-Shared-Vision-of-Ethical-AI-in-Education.pdf>
36. Dickens, J. (2020, February 27). Seldon: AI can replace Ofsted inspectors within a decade. Retrieved April 21, 2022, from <https://schoolsweek.co.uk/seldon-ai-can-replace-ofsted-inspectors-within-a-decade/>
37. Ofqual. (2020). Exploring the potential use of AI in marking. Retrieved April 21, 2022, from <https://ofqual.blog.gov.uk/2020/01/09/exploring-the-potential-use-of-ai-in-marking/>
38. Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education*, 19(1). <https://doi.org/10.14324/LRE.19.1.01>
39. Financial Conduct Authority. (2020, January 22). Artificial Intelligence Public-Private Forum. FCA. Retrieved April 21, 2022, from <https://www.fca.org.uk/news/news-stories/financial-services-ai-public-private-forum>
40. Competition and Markets Authority. (2021). Algorithms: How they can reduce competition and harm consumers. GOV.UK. Retrieved April 21, 2022, from <https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers>
41. Kazim, E., Barnett, J., & Koshiyama, A. (2021). Comments on CMA Report on How Algorithms May Reduce Competition and Harm Customers. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3785647>
42. Financial Conduct Authority. (2019, July 15). The future of regulation: AI for consumer good. FCA. Retrieved April 21, 2022, from <https://www.fca.org.uk/news/speeches/future-regulation-ai-consumer-good>
43. Financial Conduct Authority. (2020, February 18). AI transparency in financial services – why, what, who and when? FCA Insight. Retrieved April 21, 2022, from <https://www.fca.org.uk/insight/ai-transparency-financial-services-why-what-who-and-when>
44. Ofgem. (2021). Case study (UK): Digitalising energy systems for net zero. Ofgem. Retrieved April 21, 2022, from <https://www.ofgem.gov.uk/publications/case-study-uk-digitalising-energy-systems-net-zero>
45. Ofgem. (2021). Ofgem innovation vision 2021 – 2025. Ofgem. Retrieved April 21, 2022, from <https://www.ofgem.gov.uk/publications/ofgem-innovation-vision-2021-2025>
46. Cambridge Consultants. (2019). Use of AI in online content moderation. Retrieved from https://www.ofcom.org.uk/_data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf
47. Trendall, S. (2019, October 1). AI Week: Automation makes work less taxing for HMRC. PublicTechnology.net. Retrieved April 21, 2022, from <https://www.publictechnology.net/articles/features/ai-week-automation-makes-work-less-taxing-hmrc>
48. Digital Marketplace. (2020). Artificial Intelligence Tax and Revenue Chat Bot – Virtual Assistant - Digital Marketplace. Retrieved April 21, 2022, from <https://www.digitalmarketplace.service.gov.uk/g-cloud/services/162218546684837>
49. Kazim, E., Denny, D. M. T., & Koshiyama, A. (2021). AI auditing and impact assessment: according to the UK information commissioner's office. *AI and Ethics*, 1(3), 301–310. <https://doi.org/10.1007/s43681-021-00039-2>
50. Kazim, E., & Koshiyama, A. (2020). Explaining Decisions Made With AI: A Review of the Co-Badged Guidance by the ICO and the Turing Institute. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3656269>
51. Kazim, E., & Soares Koshiyama, A. (2020). No AI Regulator: An Analysis of Artificial Intelligence and Public Standards Report (UK Government). SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3544871>

52. Department for Digital, Culture, Media & Sport. (2021). Government response to the consultation on the National Data Strategy. Retrieved from <https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/outcome/government-response-to-the-consultation-on-the-national-data-strategy>
53. Christie, L. (2021). AI in policing and security. Retrieved from <https://post.parliament.uk/ai-in-policing-and-security/>
54. Clean Up Gambling. (2022). Investigation reveals scale of behavioural surveillance by online gambling firms. Retrieved April 21, 2022, from <https://cleanupgambling.com/news/cracked-labs>
55. Department for Digital, Culture, Media and Sport. (2020). Review of the Gambling Act 2005 Terms of Reference and Call for Evidence. GOV.UK. Retrieved April 21, 2022, from <https://www.gov.uk/government/publications/review-of-the-gambling-act-2005-terms-of-reference-and-call-for-evidence/review-of-the-gambling-act-2005-terms-of-reference-and-call-for-evidence>
56. National Audit Office. (2019). An introduction to the National Audit Office. Retrieved from <https://www.nao.org.uk/wp-content/uploads/2019/06/An-Introduction-to-the-National-Audit-Office.pdf>
57. Office for Artificial Intelligence. (2021). National AI strategy. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf
58. Kazim, E., Almeida, D., Kingsman, N., Kerrigan, C., Koshiyama, A., Lomas, E., & Hilliard, A. (2021). Innovation and opportunity: review of the UK's national AI strategy. *Discover Artificial Intelligence*, 1(1), 14. <https://doi.org/10.1007/s44163-021-00014-0>
59. Kingsman, N., Kazim, E., Chaudhry, A., Hilliard, A., Koshiyama, A., Polle, R., ... Mohammed, U. (2022). Public sector AI transparency standard: UK Government seeks to lead by example. *Discover Artificial Intelligence*, 2(1), 2. <https://doi.org/10.1007/s44163-022-00018-4>
60. Attorney General's Office & The Rt Hon Jeremy Wright QC. (2016). The Attorney General on who should decide what the public interest is. GOV.UK. Retrieved April 21, 2022, from <https://www.gov.uk/government/speeches/the-attorney-general-on-who-should-decide-what-the-public-interest-is>
61. Ian Burrell. (2012). Ian Burrell: Intrusion is unpleasant – but sometimes it's justified | The Independent | The Independent. Retrieved April 21, 2022, from <https://www.independent.co.uk/voices/commentators/ian-burrell-intrusion-is-unpleasant-but-sometimes-it-justified-7622360.html>
62. Department for Digital, Culture, Media & Sport. (2021). UK National Data Strategy. Retrieved December 7, 2021, from <https://www.gov.uk/government/publications/uk-national-data-strategy>
63. European Union. (2016). Recital 38 - Special Protection of Children's Personal Data. General Data Protection Regulation (GDPR). Retrieved from <https://gdpr-info.eu/recitals/no-38/>
64. European Union. (2016). Art. 32 GDPR – Security of processing. General Data Protection Regulation (GDPR). Retrieved from <https://gdpr-info.eu/art-32-gdpr/>
65. Carvalho, G., & Kazim, E. (2022). Themes in data strategy: thematic analysis of 'A European Strategy for Data' (EC). *AI and Ethics*, 2(1), 53–63. <https://doi.org/10.1007/s43681-021-00102-y>

Authors

Emre Kazim

Research Fellow in the Department of Computer Science, University College London

Elizabeth Taylor

PhD Candidate in the Department of Information Studies, University College London

Adriano Koshiyama

Research Fellow in the Department of Computer Science, University College London

Markus Trengove

Senior Research Fellow, Holistic AI

Airlie Hillard

PhD Candidate at the Institute of Management Studies, Goldsmiths University

Denise Almeida

PhD Candidate in the Department of Information Studies, University College London

Elizabeth Lomas

Associate Professor in the Department of Information Studies, University College London

Sara Zannone

Head of Research, Holistic AI

Ethan Barrance

Research Intern, Holistic AI

Nigel Kingsman

AI Audit and Assurance Officer, Holistic AI

Corresponding author: e.kazim@ucl.ac.uk