ico.
Information Commissioner's Office

Upholding information rights

Wycliffe House, Water Lane, Wilmslow, Cheshire, SK9 5AF
T. 0303 123 1113 F. 01625 524510
www.ico.org.uk

The Information Commissioner's response to the Draft AI Ethics Guidelines of the High-Level Expert Group on Artificial Intelligence

About the ICO

The Information Commissioner has responsibility in the UK for promoting and enforcing the EU General Data Protection Regulation (GDPR), the UK Data Protection Act 2018 (DPA) and the Privacy and Electronic Communications Regulations 2003, as well as the Freedom of Information Act 2000 and the Environmental Information Regulations 2004.

The Commissioner is independent of government and upholds information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The Commissioner does this by providing guidance to individuals and organisations, solving problems where she can, and taking appropriate action where the law is broken.

The Commissioner's response to each chapter of the draft AI Ethics Guidelines, produced by the European Commission's High-Level Expert Group on Artificial Intelligence, is as follows:

Introduction: Rationale and Foresight of the Guidelines

As the regulator for information rights and data protection in the UK, and current Chair of the International Conference of Data Protection and Privacy Commissioners (ICDPPC), the ICO welcomes the work of the High Level Expert Group and the opportunity to respond to the working document on draft ethics guidelines for Trustworthy AI. In doing so, the Information Commissioner recognises the importance of a shared ethical framework underpinning the international landscape of AI governance, building on the Declaration on Ethics and Data Protection in AI agreed at last year's ICDPPC conference.

We support the identification of Trustworthy AI as the 'north star' of the High-Level Expert Group, and particularly the requirement that AI be 'demonstrably worthy of trust.' The ICO has found evidence of low levels of trust by the public in how organisations use personal data and this represents a potential barrier to the development of AI. It is only when data controllers and processors are in a position to demonstrate that they are worthy of the trust that may be placed in them that the benefits of AI can be fully and ethically realised.

This chapter also references a future mechanism to enable stakeholders to sign up to the guidelines, and the ICO would be interested to learn more about the role such a mechanism is expected to play. We welcome a greater degree of co-ordination and co-operation in this space, recognising the connection between digital ethics, governance and regulation in the realisation of trustworthy AI.

Chapter I: Respecting Fundamental Rights, Principles and Values - Ethical Purpose

The ICO welcomes the rights-based approach to AI Ethics in the draft guidelines (which complements the protection of human rights and freedoms operationalised in the GDPR) and the derivation and development of ethical principles from these rights, These help to reinforce key data protection principles, as recommended in the ICO's 2017 paper *Big Data, Artificial Intelligence, Machine Learning and Data Protection*.

Section 3.4 on equality talks of "inclusion of minorities, traditionally excluded, especially workers and consumers" (p7). It seems a little unusual to categorise workers and consumers in this way. We would recognise that it is possible to have an imbalance of power between workers and consumers on the one hand and data controllers on the other, but it does not seem correct to refer to these groups as minorities.

The introductory paragraphs to section 4 (p8) advise "the presence of an internal and external (ethical) expert". Some organisations deploying AI will have limited resources and the guidelines should present an approach which is scalable to their needs. This statement could therefore perhaps be qualified with a phrase such as 'wherever practicable'

The ethical principles articulated in section 4 reflect those used in the field of bioethics. We make no comment on how successfully they have been implemented in that field, but we do note the addition of a fifth principle: explicability. It could be argued that explicability would not be universally recognised as a normative principle, but nevertheless we think it is appropriate to add it, provided it is interpreted broadly to include concepts such as explainability, intelligibility, transparency and accountability. We see it as a principle which can enable the application of the other principles and provide an assurance that they are being followed. There are also important linkages with the GDPR principles of transparency and accountability, and the GDPR requirements for meaningful explanation of automated decision-making. In this context, the ICO is currently working with the UK's Alan Turing Institute on producing guidance to assist organisations in explaining decisions made by AI systems.

Chapter II: Realising Trustworthy AI

The explanation of accountability in the list of ten requirements for realising trustworthy AI seems to focus on mechanisms for compensating for error or wrong-doing, rather than pro-actively ensuring compliance. This doesn't cohere with the meaning of accountability in the GDPR, where it is understood in a wider sense as being responsible for, and able to demonstrate compliance with, the data protection principles. It may not be helpful to use the same term in a more limited way in the ethical guidelines, and we would prefer it to be used in a wider sense here.

Regarding non-discrimination (section 5), it may be worth distinguishing between two kinds of unintentional bias in data. The first is the bias that arises when the data is not drawn from a statistically representative sample of the population of interest (e.g. containing proportionately fewer women), which results in less accurate models. The second kind of bias concerns data which accurately represents the population (e.g. containing a proportionate number of each gender), but where this in turn reflects the results of direct or structural discrimination (e.g. workplace assessments which reflect the gender biases of managers or unfair maternity arrangements).

Reference to the GDPR in the list of requirements for trustworthy AI is limited to the requirement for *Respect for Privacy*, but the scope of the GDPR extends to a number of the other categories, and it may be helpful to acknowledge this. In addition to accountability, there is a clear requirement for transparency (1st data protection principle), non-discrimination (e.g. recital 75), robustness (accuracy, 4th data protection principle; controller obligations, e.g. Articles 25, 35). These are legal requirements for data processing under the GDPR as well as ethical requirements. It would be helpful if reference to compliance with the GDPR were not limited only to the requirement for *Respect for Privacy*.

Following on from this, some of the technical and non-technical methods for achieving trustworthy AI align with the legal requirements under the GDPR for data controllers and processors to ensure that data protection principles and data subject rights are complied with by design and by default. In particular, a Data Protection Impact Assessment (DPIA) is likely to be a requirement for AI projects processing personal data. The ICO would expect UK data controllers to demonstrate their compliance as part of a DPIA using at least some of these methods, as part of an ongoing process and in line with their legal obligations.

Section 2.1 *Testing & Validating* argues that 'bounty hunting' may be considered, whenever feasible, as a technical method of achieving Trustworthy AI. In the context of the GDPR, this may not be advisable. Depending on how the 'bug bounty' program is organised, vulnerabilities exposed by even 'white hat' hackers may still constitute data breaches that need to be reported to the respective data protection authority (DPA).

Chapter III: Assessing Trustworthy AI

The draft Assessment List for trustworthy AI, although not proposed as mandatory for AI developers and companies, would sit alongside the likely legal requirement for a DPIA under the GDPR. We would expect a number of the questions on the list to be addressed by data controllers as part of a DPIA for an AI project, so there is a broader question here about how these assessments sit alongside one another – or might be brought under a single form of assessment – where the appetite among some organisations to carry out two discrete assessments may be low. Outside the GDPR jurisdiction, the

Ethical Data Impact Assessment model developed for the Hong Kong Privacy Commissioner is an interesting example of bringing together ethical and data protection assessments.

Under the assessment questions for *Respect for (& Enhancement of) Human Autonomy* (p26), there is no reference to consideration of a right to opt out or withdraw from AI systems and decision making, although these rights are mentioned under the principle of autonomy in Chapter I. Perhaps this ought to be part of the Assessment List when considering autonomy, and this may lead to a consideration of how such rights can be actualised, given how pervasive AI systems and decision-making may become.

The ICO is interested to understand the process by which such assessments would be carried out, with the inclusion of “specific metrics” (p.24), and looks forward to learning more in the next iteration of this document.

General Comments

The ICO welcomes the HLEG’s framework for trustworthy AI, as comprising ethical purpose and the need to be technically robust. Europe is already a global leader in the regulation of information rights, reinforced by the introduction of the GDPR, and additional compliance with agreed ethical guidelines will help further position Europe and the UK to reap the benefits of AI.

We are keen to see the realisation of these benefits, both for individuals and for society as a whole, encouraging innovation that is compliant with data protection law and with people’s rights and freedoms. Having said that, the statement in the Executive Summary, that “on the whole, the benefits of AI outweigh its risks” seems rather too generalised to be meaningful. It may be better to make a statement to the effect that AI can bring enormous benefits to society and to individuals, but its development requires a respect for fundamental rights.

We believe that a drive towards the ethical development of AI will serve to support the work of DPAs in protecting personal data rights. Assessing fairness in the context of AI increasingly raises issues to do with the societal impacts of the processing which are difficult to resolve within the scope of data protection legislation, and the development of ethical standards can help there. Furthermore, the adoption by organisations of ethical approaches to data use will serve to assist their compliance with legal requirements.

The ICO is willing, within the limits of its remit, to contribute to the development of this framework, working together with partners in data ethics to “develop a unique brand of AI” (p ii). In the UK the creation of the national Centre for Data Ethics and Innovation is an important step in the same direction and we are planning to work closely with them.

We recognise that these draft guidelines aim to foster reflection and are a starting point for discussion on ‘trustworthy AI made in Europe.’ While our remit is to be the regulator for the data protection laws that protect UK citizens, the development of a wider international consensus on ‘the common good’ in relation to AI technologies would be a welcome result of such discussions.