2020 - Information Commissioner consultation on explaining AI

About Jisc

Jisc is the UK's digital lifelong learning and research agency. We operate critical shared digital infrastructure and services for all publicly funded universities, FE colleges and research facilities. In addition, we offer trusted advice and guidance on using technology to improve all aspects of the education and research ecosystem: from teaching and learning to student experience, research excellence and institutional efficiency.

Education 4.0

It is widely agreed that digital technologies such as artificial intelligence (AI), the Internet of Things and machine learning are changing industry and the workplace. However, we believe the potential benefits of these technologies are yet to be fully realised across tertiary education. Education needs to take advantage of technology to provide a capable workforce and a flexible, lifelong learning experience that benefits future decades of learners.

Guided by our vision of 'Education 4.0', we're working with our members to help them embrace digital transformation. Of relevance to this consultation, we have worked with around 100 institutions to develop the world's first national learning analytics service that uses AI to:

- Improve student retention
- Enhance the student experience
- Make the organisation itself more productive

The service uses the data from institutional IT systems to give learning providers indicators of student engagement over time and correlation with learning outcomes. We are currently exploring, within the ICO's Regulatory Sandbox, how similar approaches might safely be applied to supporting student wellbeing.

Substantive response

We welcome the ICO/Turing Institute's draft guidance on Explaining AI Decisions [1], and believe that it could be useful well beyond the narrow question of when and how decisions need to be explained. However, as a regulatory tool we suggest that it needs a clearer, and objective, definition of which systems are, and are not, covered by the term "AI". We also have some suggestions to improve the usability of the guidance.

We consider the most significant contribution of the document to be its identification and analysis of six different types of explanation: rationale, responsibility, data, fairness, safety and performance, and impact. The guidance also provides helpful clarification that some of these occur before processing takes place and apply to the whole system, while others occur after processing and apply to individual decisions. We believe that this analysis could usefully be applied to most systems involving complex flows or large amounts of data, whether or not they involve "Artificial Intelligence". Considering rationale, responsibility, data, fairness, safety and performance, and impact throughout the design, development, implementation and operation of large-scale data processing systems should be good practice to improve their safety for operators and individuals alike.

To deliver its full benefit, we therefore consider that the guidance should both make this broader scope explicit and, within it, provide a clear, objective, definition of "AI". The draft contains only a statement (on Part 1, page 4) that "AI is an umbrella term for a range of technologies and approaches that often attempt to mimic human thought to solve complex tasks. Things that humans have traditionally done by thinking and reasoning are increasingly being done by, or with the help of, AI". This appears to accept that "AI" is often used purely as a marketing term, and to leave it to individual marketing departments to decide whether their product or service falls within the guidance. Those that do not wish to follow the guidance may simply reduce the prominence of the term "AI" in their marketing materials, or replace it by some other term. Conversely, unrealistic expectations may be raised among data subjects that anything labelled "AI" will provide the explanations described in the guidance, even when these may not be relevant or necessary.

Our principal recommendation is therefore that the ICO/Turing Institute adopt an objective definition that can be applied consistently and objectively to determine which systems are, and are not, "Artificial Intelligence". We have found the definitions used by DSTL [2] helpful: in particular that AI consists of "Theories and techniques developed to allow computer systems to perform tasks normally requiring human or biological intelligence."

Alongside that objective definition, we consider that the broad applicability of the guidance would be made clearer by changing the order of the Legal Framework section in Part 1. At present it begins with the narrow set of legally-significant decisions covered by Article 22. The much wider group of data controllers who may be required to provide explanations under Article 5 Fairness might easily conclude that the guidance does not apply to them. To avoid this, we would suggest presenting the Article 5 requirement, then Article 22, then the situations where explanations are good practice.

Part 2 is currently a very long block of text and, as a result, hard to navigate. Some sort of graphical representation would help readers find the sections most relevant to their application. It would also be helpful to provide a graphical indication when discussing types of algorithm that are inherently non-explainable.

Part 2 also contains an important point, which we consider should also be raised in the introductory or management Parts, on the need to train the humans who will be working with AI as an assistant. This involves striking a tricky balance between unquestioning over-reliance on the AI's recommendations and encouraging humans to substitute their own biased judgments. Both training on when to over-ride the algorithm, and support systems to ensure this facility is not (consciously or unconsciously) misused are likely to be needed.

Part 3, page 17 mentions the Article 21/Article 17 right to object, but without explaining its scope and nature. In our experience the application of this right to AI has been widely misunderstood by both data subjects and data controllers. In the most extreme form of this misunderstanding we have heard model builders assert that it requires them to keep all the personal data used to build a model, in case one person exercises their "right to object" and they are required to rebuild the whole model from scratch omitting that individual's data. This guidance would be a good place to counter such high-risk practices by providing an authoritative statement of what the right does, and does not, require.

Finally, we welcome the recognition that for some applications of AI, "gaming the algorithm" is a positively desirable feature. Jisc has done considerable work on applying analytics to various fields in education and research. By examining data generated by teaching and research processes we hope not merely to predict likely outcomes, but to identify changes that can result in actual outcomes being better than those predicted. This presents new challenges throughout the lifecycle of such systems: algorithm developers must not just explain why a certain prediction was made, but also what needs to be done to improve it; users and regulators must understand that predictions that turn out to be inaccurate may actually be a sign that that the system is achieving its objective.

Links
[1] https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultation-on-explaining-ai-decisions-guidance/
[2] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/850129/The_Dstl_