Helping organisations to explain decisions made with AI

Organisations are increasingly using AI to make or assist decisions that directly affect people, from diagnosing disease and approving bank loans to assessing job applications and recommending products. From a legal and ethical perspective, it is important that those affected by these decisions understand how and why the decisions are made. Moreover, when organisations are transparent about how they use AI, it helps to build trust within the workplace and the wider public, especially where the decisions raise possibilities of discrimination against protected characteristics such as age, disability or race. Striving to make AI systems explainable can also help to flag up potential biases within the systems.

Since 2018, the Turing's public policy programme and the government's Information Commissioner's Office (ICO)

have been working to produce **a co-badged guidance document** for organisations, providing advice on how to clearly explain AI decisions to those affected by them. Published in May 2020, it is the most comprehensive practical guidance on AI explanation produced anywhere to date. It gives four key principles for organisations to follow when explaining AI: be transparent, be accountable, consider the context you are operating in, and reflect on your AI system's impacts on the individual and wider society.

David Leslie, Ethics Theme Lead at the Turing and co-author of the guidance, has given several lectures and workshops about the work, including a presentation to the US National Institute of Standards and Technology (NIST). He has recently begun research to gauge how organisations are using the guidance to improve their practices.

"As well as providing key expertise, the Turing helped us to ensure that we consulted with a wide range of voices within the AI community, so that the final guidance was as accessible and useful as possible."

Abigail Hackston
Co-author of the guidance
and Senior Policy Officer
ICO