**Report on the nTAIL/TAS workshop - “Making AI Understandable”**

**Introduction**

Much of the work in the Explainable AI (XAI) research field is aimed at practitioners and policy makers, little at a more general audience. Similarly, very few AI-enabled products explain how and why they use this technology. From our research we’ve seen users suggest that they would like to know more about AI in a system, but when shown interfaces to explain recommendations they were nonplussed. We also see millions of users frequently using AI-powered apps without a second thought to understanding how they work. As the recent AI Council Roadmap says...

*“It is about knowing enough to be a conscious and confident user of AI-related products; to know what questions to ask, what risks to look out for, what ethical and societal implications might arise, and what kinds of opportunities AI might provide. Without a basic literacy in AI specifically, the UK will miss out on opportunities created by AI applications, and will be vulnerable to poor consumer and public decision-making, and the dangers of over-persuasive hype or misplaced fear."*

What could we make that demystifies AI for the public, across digital products, education, media, language, imagery and culture? How could we help make better mental models of how AI works? Could we make people care about understanding AI? How can we show where it touches peoples’ lives and affects individuals? How can we better show the benefits and limitations of AI systems? How can we ensure explaining AI benefits the individual and society, and enables trust in AI systems where appropriate?

**Scene setting**

We started off by presenting some of our existing work on understanding AI, including an overview of explainable AI, a demonstration of an ML-powered bird identifier that explains itself, and our thinking on the different approaches you can take to explaining AI. We discussed whether we should focus on how the technology works, or the interactions with the technology, or the impact of it. We also heard that asking people what they fear about AI can prove a good starting point.



**Images of AI**

We then ran an interactive session around the images that are used to represent AI in society and culture. News stories or press releases about AI are often illustrated with stock photos of shiny gendered robots, glowing blue brains or the Terminator. We think this is potentially harmful and ultimately affects public understanding and useful critical discourse around this increasingly influential technology. But we also think there could be better, less clichéd, more accurate and more representative images and media to represent this technology.

Participants were asked to bring along some images that represented AI, and we had curated a set of "helpful" and "unhelpful" ones. These formed the basis of our discussion. The group’s thoughts about "good" or "helpful" images included pictures that shows AI around us, invisible and pervasive, illustrations showing creative uses of AI and the collaborations and relationships between the technology and people. The discussion around unhelpful images focused on images implying general intelligence where it doesn't exist and robots replacing jobs. We thought images should avoid anthropomorphising the technology. We had some disagreement about whether algorithms can create order from nothing. We also discussed the style of image. Diagrams can be seen as boring, and illustrations can take more time to process than a photograph-style image.



Following this we asked participants to try to sketch, make or describe a better representation of AI. We ended up with illustrations including:

* The process of taking real-world faces into abstract facial features into AI-generated photos of faces that don't exist.
* Sieves, filters and funnels of data
* Showing AI making mistakes and the selection of the good stuff in AI-assisted creativity.
* Alternative representations of brains, ghosts and souls





**The future of AI**

In our work we've been using "futuring" tools to extrapolate possible signals of change into possible futures. We used an impact wheel tool (or Paul the Octopus as we prefer to call it) to take a trend and extrapolate out some of its implications; and then think about the implications of those implications. We chose 3 trends related to understanding AI and for the final part of the workshop, three groups took on one question each using this process.



Prompt 1 was *"Someone develops an augmented reality interface to YouTube allowing you to manipulate the algorithm and see the results".*

What if this excludes those without the (expensive) technology and those who don't know how to use this feature. Whose algorithm is manipulated? Influencers start influencing algorithms. People can manipulate other peoples' algortihms to your own benefit. It creates more awareness of the algorithm, which is influencing our lives. From the AR angle we may end up with custom visual filters that "clean up" what you see or that splinter what people see into multiple different realities.

Prompt 2 was *"It becomes a legal requirement to respond to FOI requests for the reasons why an algorithm made a decision".* This creates potential for algorithmic "offshoring", so you're not in the jurisdiction. Some companies will become more transparent, new organisations and start-ups develop to support and enable this. It creates tensions between individual and groups rights and public understanding will need to catch up with the regulation

Prompt 3 was *"Statistical ML methods becomes a compulsory GCSE module".* As more people know about ML there could be a wider range of uses in different disciplines and potentially more innovation and creativity as younger people create more things with ML. Maybe people get less excited about AI and there's a risk that it puts people off studying AI and ML. But there is less hype and money is better targeted in the field.

**Conclusions**

We hope people enjoyed the workshop, we are grateful to everyone who attended and kept up their energy levels throughout a long Friday afternoon. We are continuing to work on developing better images for AI and the means of publishing them, and we are actively seeking partners who would like to work on this. We are also looking for collaborations that can help make creative and engaging ways of explaining AI to everyone.

*Tristan Ferne, BBC Research & Development*

tristan.ferne@bbc.co.uk