Peoples AI Stewardship Summit

Ormeau Baths, Belfast, March 5th, 2024

"There's huge potential for benefits and a huge need for balance."

Scribe Capture Document

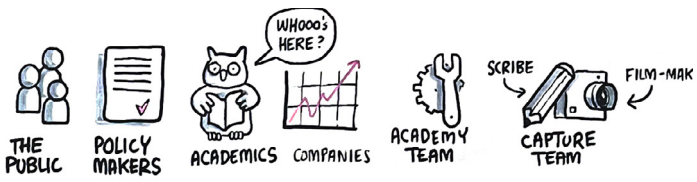
Captured by
We are Cognitive
wearecognitive.com

Royal Academy of Engineering

Royal Academy of Engineering

Peoples AI Stewardship Summit

Ormeau Baths, Belfast, March 5th, 2024



In very recent history, we have seen an explosion in Generative AI technologies. But how does the public feel about these changes? What are their hopes and fears for these new tools? And how do they perceive responsibility, safety, and trustworthiness?

The People's AI Stewardship Summit brought together a mix of diverse voices: the public, industry, policymakers, and academics. The aim was to listen to what the people of Belfast want from AI, ensuring their preferences are heard as we shape the future of these technologies.

The Royal Academy of Engineering (or just 'The Academy') hosted the Summit. Our strategy is all about using engineering to build a sustainable society and inclusive economy.

The Ormeau Baths—a beautiful red-brick building that has brought people together since 1888—served as a fitting space for the conversation. The venue is home to one of a number of Enterprise Hubs set up by The Academy, who provide the support, funding, and network for startups to flourish.

Pre-event Zoom Call: What Is the Summit All About?

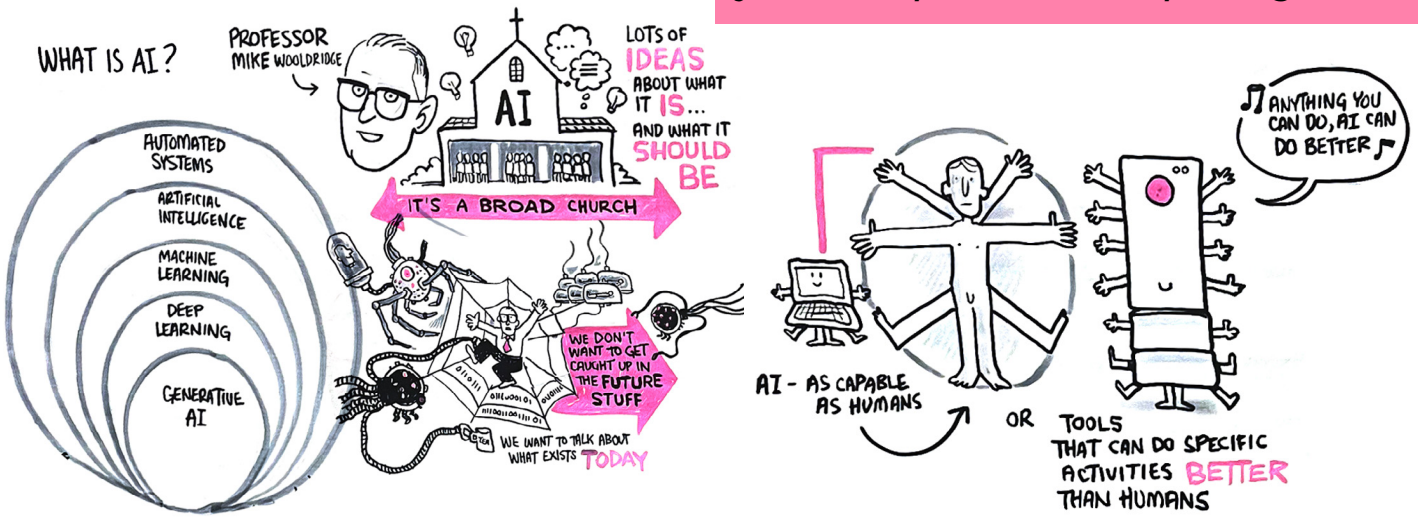
In the lead-up to our in-person gathering, we saw each other's pixellated faces on Zoom, and Facilitator Anna Beckett gave context on the Summit's purpose.

Participants came with varying degrees of familiarity and interest in AI—some had played with generative tools, while others had not.

A Royal Institution video from Professor Mike Wooldridge explained what AI is in the first place. To some, it means the Hollywood dream; AI means machines as capable or even better than humans. To others, himself included, it's about tools that can do specific tasks better than humans.

Eliot Gillings from The Academy then gave a live demonstration featuring the generative AI tools ChatGPT and Stable Diffusion, showcasing their strengths and weaknesses. Although AI tools seem to create something out of thin air, they're instead creating something out of many, many something elses. They work on probability, guessing what the data they're trained on means, without ever understanding it.

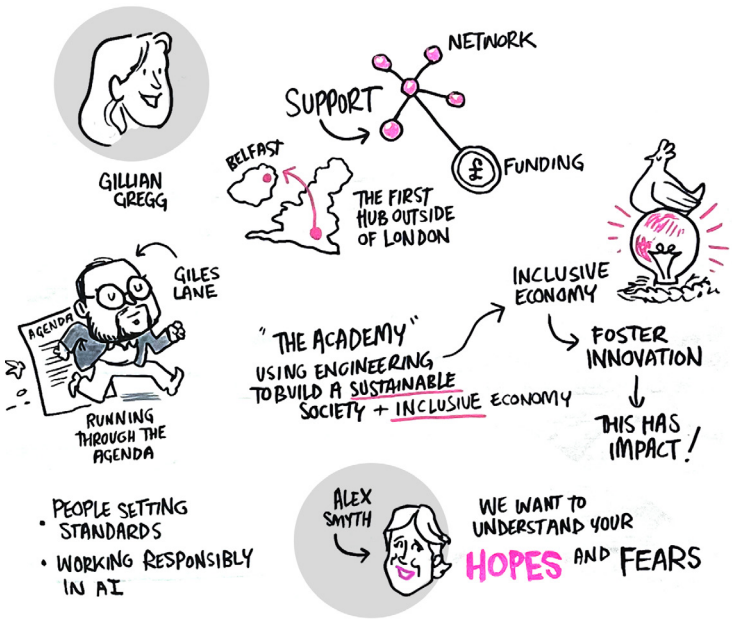
"It's analogous to asking a friend to translate something into a language you don't speak and then passing it on



Launching The Summit: Welcome and Opening Remarks

Gillian Gregg, Alex Smyth, and Giles Lane

Participants streamed into the room, collecting their nametags and exchanging introductions over a spread of sandwiches and spring rolls. Once they had found their seats, they were encouraged to mull over AI's role in their lives.



The Academy's team kicked off proceedings with introductions. The Royal Academy of Engineering is UK-wide, but we decided to visit Belfast because it is home to the Northern Ireland Enterprise Hub. The Enterprise Hub connects local innovators, researchers, and Fellows to the wealth of entrepreneurship, funding, training, and support The Academy offers.

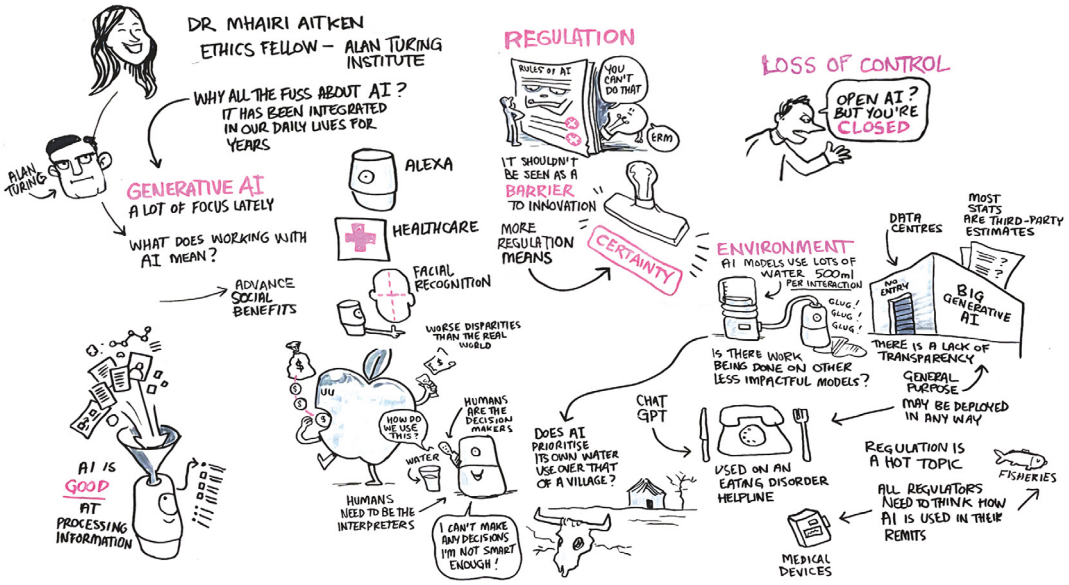
> ***"We're supporting people who are developing AI in Belfast."***

We want The Academy to foster innovation in Belfast, but we want this innovation to be responsible and in alignment with what the public wants from new technologies. Each participant brought their own perspectives, hopes, and fears; this event was about unearthing those.

Setting the Scene: What AI Safety Might Mean

Dr Mhairi Aitken, Ethics Fellow, Alan Turing Institute

Dr Mhairi Aitken researches AI Ethics at the Alan Turing Institute, the national institute for data science and artificial intelligence. She was at the Summit to be an expert voice, framing discussions on AI safety and how AI is already having an impact today.



> ***"When we think about AI safety, we're thinking about individuals, communities, and wider society."***

To Dr Aitken, AI safety means understanding the risks but also maximising the value of technology and innovation.

AI is by no means new. It has already been integrated into systems affecting our lives in some familiar ways, like Alexa and Siri, but also ways we might not be so aware of, like screening job applications, diagnosing diseases, and allocating services. The past 19 months have seen an explosion in interest in AI—generative AI, to be specific.

Dr Aitken highlighted some of the main risks, beginning with bias and representation. Humans, as we all know, are biased. But AI can be worse. The results from a study of Stable Diffusion's AI image generators made this clear. The image sets it generated amplified offensive stereotypes, such as the idea that white male CEOs run the world, women are rarely well-paid, or that men with dark

skin are more likely to commit crimes.

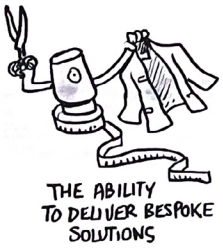
Knowing where information comes from is vital for a healthy democracy. Misinformation and disinformation present another risk. Dr Aitken showed a picture of Donald Trump being arrested. A closer look shows that Mr. Trump's arm is much too short, and his hand resembles a claw. This 'photo' was fabricated.

It surprised some participants to hear that AI has an environmental impact, too. Computer scientists estimate that training OpenAI's giant GPT-3 text-generating model is akin to driving a car to the Moon and back. Moreover, AI chatbots guzzle vast amounts of water, with the average conversation being equivalent to spilling a 500ml bottle of water.

You often hear whispers of AI carrying an existential risk and that one day it will make its own decisions and take over. According to Dr Aitken, such stories are sensationalised and belong in the realm of science fiction.

As the discussion opened to the floor, the issue of transparency arose. Most major tech companies aren't transparent; they don't let us in on their stats, leaving us to rely on third-party estimates.

"Developing paths forward requires public dialogue."



the misuse of generative AI models was another discussion point. Dr Aitken shared a story of an AI eating disorder helpline giving damaging advice, and she suggested that, perhaps, we should be thinking about smaller, more bespoke models, which would have smaller impacts, too.

Lastly, the discussion moved to regulation, its merits, and struggles. Dr Aitken explained how regulation could curb profit-driven behaviour and ensure responsible innovation.

"Regulation shouldn't be seen as a barrier to innovation but as a way to ensure that we innovate responsibly."

Poster Presentations: Visions of a Post-AI Society

The next portion of the day was post-its and flipcharts galore as participants shared their personal hopes and fears for the future, first on an assigned topic, and then on a topic of their choice. They turned their thoughts to government, industry, and academia.

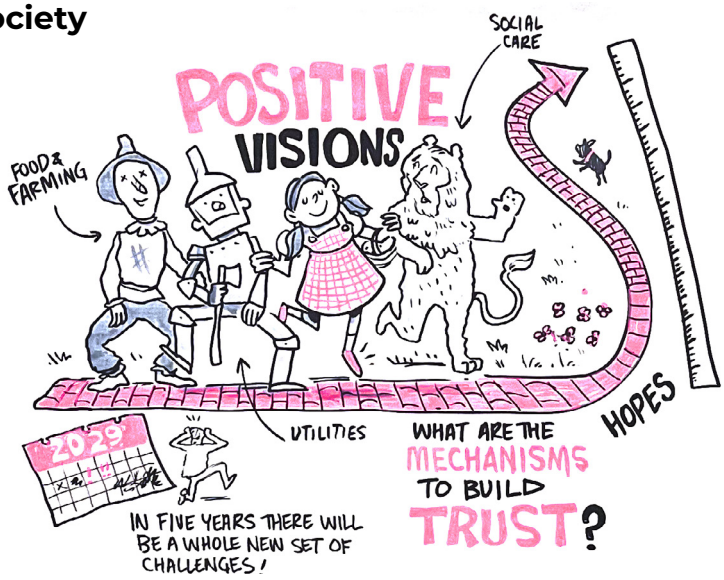
What could they do to make those hopes more likely and those fears less likely?

Groups made an artistic representation of the positive AI-related future they'd like to see. Amidst the sound of ripping paper, laughter, and lively discourse, they created eight collage posters. Presenting those posters to the room, it was clear that several common themes had emerged:

AI Should Bring People Joy

An image of a bot helping an elderly lady to pick something up represented the point that AI should make life easier.

Participants hope that by supporting menial daily tasks, AI could grant us more time to do things we enjoy, like gardening, and will help us to create "home-centred" communities.



"We're hoping for a 3-day working week."

AI Could Make Us Feel Safer

Participants hoped for assurance about how AI is used, with multiple groups advocating for a solid watchdog. With robust oversight, they envisaged that AI could prevent conflict and make society safer.

AI Could Deliver Health Benefits

Groups envisioned that massive medical and well-being advances are on the horizon thanks to AI, with innovations ranging from a wearable "doctor in the pocket" to an AI agony aunt to proactive, preventative data collection supporting public health.

"We want to be able to see a GP right now."

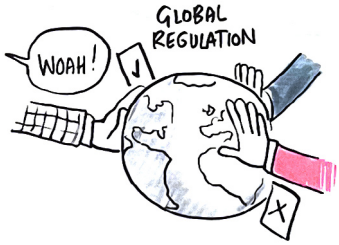
AI Must Be Accurate and Reliable

When fish ingest other fish, they build up mercury. Similarly, as AI models recycle content, more training and data can have unintended consequences. Ensuring the integrity of AI models emerged as a priority, with participants expressing the desire for high-quality, trustworthy AI systems.

"How will AI lead to a 'happy fish' rather than a 'sad fish'?"

AI Should Be Regulated Globally

In their vision boards, participants expressed the desire to vote on regulations related to AI. In their vision boards, the participants expressed the desire to vote on regulations related to AI. And for regulation to be global, a point emphasised by the AI stakeholders in the room.



"There's a need to balance a smart use of money with the risks of privatisation and profitmaking."

AI Must Be Accessible to Everyone

Participants hope that AI can be delivered to all and that everyone will feel equipped to avail of it. In their imagined futures, the wealth AI generates will be delivered across the community, saving and making money for the many, not the few.

"We don't want a niche tool for the already advantaged."

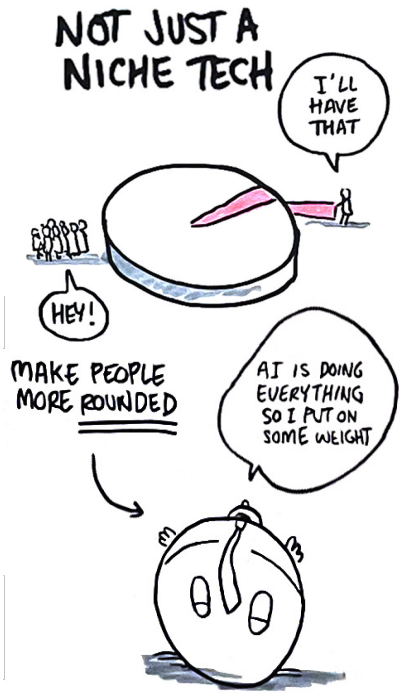
AI Could Make Humans Better

Finally, contributors hope that AI will change us humans and make us better people—people who are more creative, less biased, and more well-rounded, with a better understanding of diverse cultures.

Throughout these discussions, the overarching theme of balancing innovation with responsibility resonated strongly:

"There's huge potential for benefits and a huge need for balance."

SAVE & MAKE MONEY



Panel Session: Expert Reflections

Next, it was time to hear from those using and shaping these technologies. On our expert panel for the afternoon, we had:

Co-founder of Enzai Technologies Ltd and member of the Regional Talent Engine Enterprise Hub. Ryan is a lawyer advising legal firms worldwide about their technological responsibilities. His work is about regulating AI while still allowing the technology to flourish.

Senior AI Engineer at Kainos, a digital transformation consultancy. She is working to understand emerging technologies. (An example of when you may have encountered what Kainos do is renewing your passport online.)

A Lecturer in Diagnostic Radiography who researches uses of AI systems in medical decision-making (in radiography, for example).

Having witnessed the poster presentations, the panel reflected that they were excited to hear such optimistic energy in the room and felt pleased that discussions hadn't gone down a dystopian "terminator path." As people directly working with AI, they share that optimism.

Ryan noted a "fascinating" disconnect between the consensus in the room that AI should be regulated and the government's stance of no regulation. And with apologies for bursting participants' bubble, he noted that we aren't likely to see a 3-day working week any time soon; instead, we'll be expected to deliver more and faster

What Do You Hope to See in Belfast?

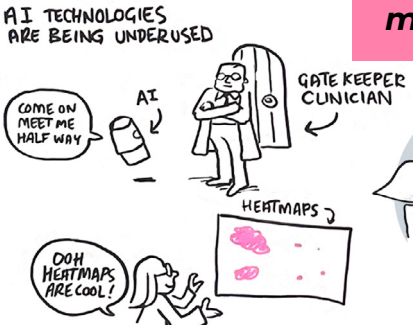
The panel reflected on what makes Belfast unique.

"We have extraordinary tech talent for our small size."

Additionally, there's the unique privilege of access to both UK and European markets. To capitalise on Belfast's advantageous position, the panel hopes to see more cross-domain collaboration, more leveraging of the existing infrastructure, and more AI-driven reskilling within the community.

Clare highlighted that many AI technologies are being underused. In a clinical setting, a tech-savvy clinician is more likely to go out of their way to seek new tools. For those less savvy or less inclined to trust modern technologies, she'd like to see technologies meeting them halfway, giving them the information they need in an interrogatable format.

Holly suggested that a 'scores on the doors' style initiative, like that used in food safety, could instil confidence in AI without requiring deep technical knowledge.



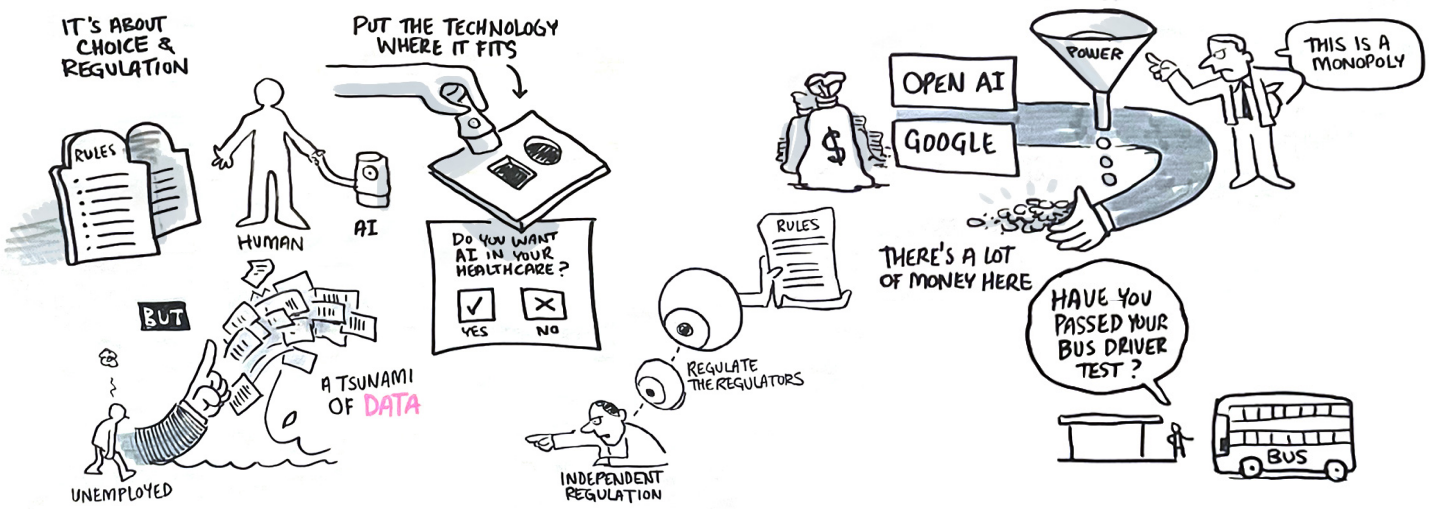
The wide-ranging discussion also touched on bias—and how it isn't always bad— and the possibility of learning from the aviation industry's exemplary decision hygiene.

Participants posed questions ranging from regulation and taxation to moral accountability, with the panel underscoring the need for relentless safety and robustness processes in AI systems.

Participants discussed the existence of regulations that protect against a monopoly and unfair usage. Thankfully, as it stands, there is enormous competition between the big AI technology providers, and that is beneficial, particularly in sectors like healthcare.

"Laws codify our collective morality."

Open Discussion and Reflections



Participants chose from three discussion prompts for our final facilitated discussion.

What tangible 'markers of trust' are needed for you to feel confident in AI and how it is being used?

In answering this question, it became apparent that we tend to trust some organisations more than others. The NHS, for example, was high up on participants' trust rankings. But they questioned if we truly know how they use AI, how accurately they use it, and what measures they take to deal with bias.

The group naturally assumes AI products have been verified and tested, though they'd like to see this overseen by an independent regulator. That said, how can we trust those regulators? One suggestion that arose was a second regulator to oversee the first.

How can local businesses be supported to use AI effectively and responsibly?

This group found it easy to think of how small and medium businesses could use AI, but it was trickier to imagine AI used in microbusinesses like a local butcher.

They reflected on trust. We unthinkingly put our faith in people every day, like bus drivers and manufacturers, assuming they have done their due diligence. Similarly, AI has many unknown safeguards, and we expect it to meet similar standards.

What is the role of the public in setting standards for AI? Where and how should the public be involved in the AI safety assurance process?

The consensus was that it's up to everyone in the sector to foster transparency, providing the information for people to make informed choices.

AI will be globally influential, so they want to see people included from different countries and of all ages. What people wish to see may vary from country to country.

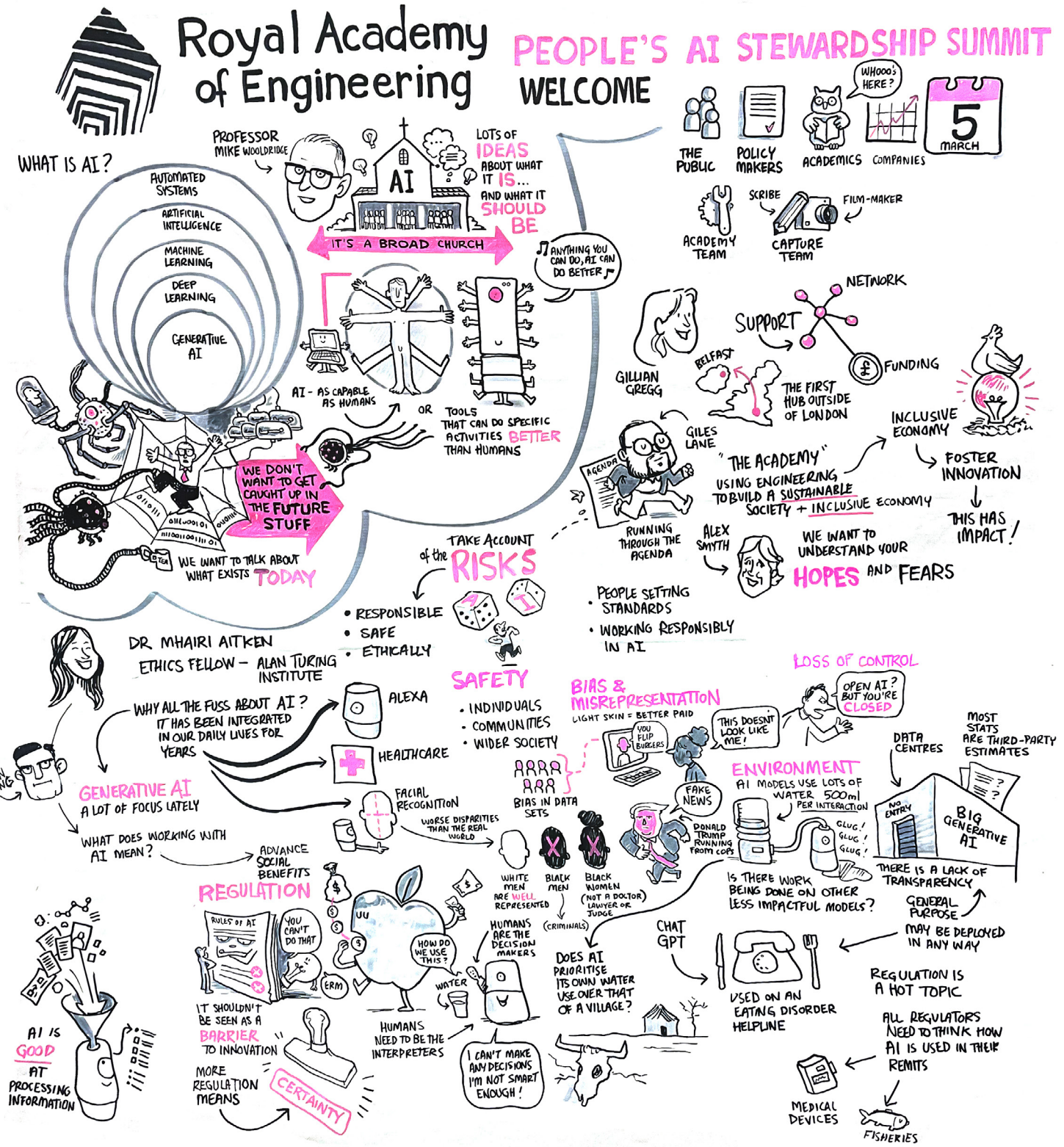
Participants want to know, right from the initial stages, that AI is being built with the correct principles in mind by a person. AI should serve as a tool rather than a substitute for moral judgment.

"The moral responsibility for making decisions should always lie with the human and never with the machine."

That wraps up the formal part of the Summit's dialogue, which was continued informally over refreshments.

We'd like to extend a heartfelt thank you to each participant for enriching the dialogue.

Insights gleaned from the discussion will inform future Enterprise Hub activities and contribute to shaping policy developments, fostering an AI future that prioritises inclusivity and societal well-being.

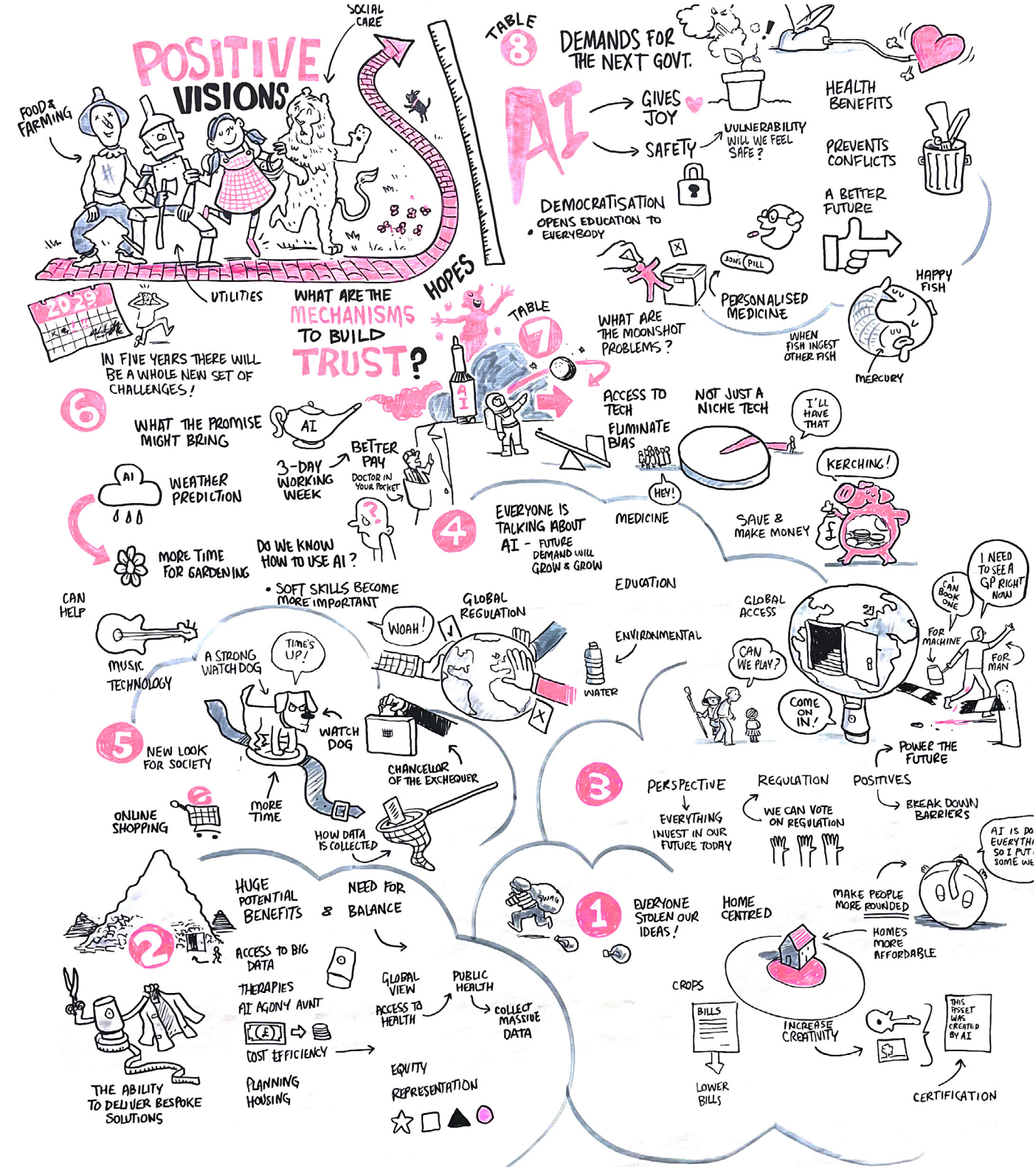


Captured by:

We are Cognitive
wearecognitive.com

The Peoples AI Stewardship Summit
March 5th 2024

The Big Picture

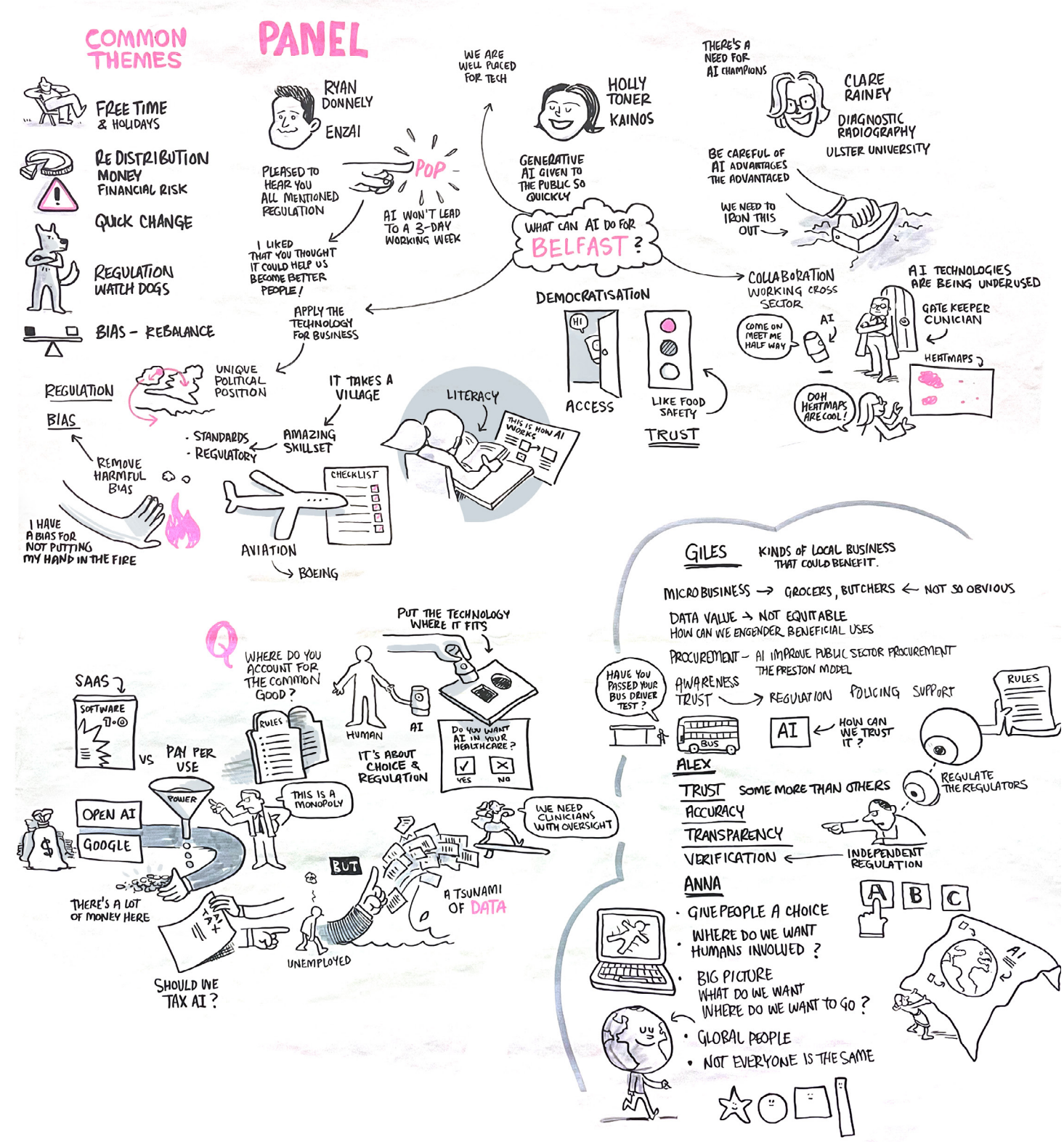


Captured by:

We are Cognitive
wearecognitive.com

The Peoples AI Stewardship Summit
March 5th 2024

The Big Picture



Captured by:

We are Cognitive
wearecognitive.com

The Peoples AI Stewardship Summit
March 5th 2024