Fellows' AI Interest Group

Shaping Our AI Agenda
21st July 2023, 12:00 – 14:00, Online

In attendance

Fifteen Fellows of the Royal Society for the Encouragement of the Arts, Manufactures and Commerce (RSA), including David Atkinson, Tania Duarte, Michael Allan Galvez, Marc Goblot, Ismael Kherroubi Garcia, Aranghan Lingham, Michaela Moher, Silviu Pirvu, Julie Samuels and Ian Stobie.

Acknowledgements

Author of this report: Ismael Kherroubi Garcia

Reviewers: Aranghan Lingham and Silviu Pirvu

Brief Overview

On 21st July 2023, the Fellow-led Artificial Intelligence (AI) Interest Group conducted their second workshop, building on the insights gained on the Steps at RSA House on 7th July 2023. On this occasion, the workshop was delivered online, allowing for Fellows from across the UK and the US to join. Across both workshops, 21 Fellows have been involved in total.

The session lasted two hours and had two distinct parts, with one break in the middle. Acronyms used in the summary are "AI" (short for "artificial intelligence") and "FRSAs" (short for "Fellows of the RSA"), and "the Group" is used as shorthand for "the Fellow-led AI Interest Group." The present summary has three parts:

1. Introducing workshop methodology and tools; namely, the World Café Method for structuring the discussion, Zoom for meeting online, and Miro for note-taking;
2. Reproducing attendees' responses to the question "What three big areas should we work on?" followed by a brief thematic analysis; and
3. Presenting the discussions that unfolded around the three pillars of the Design for Life mission: *Building Capabilities*, *Growing Hubs* and *Developing Infrastructure*; respectively.

Part One: Methodology

The workshop was structured following "The World Café Method" by arranging breakout groups that would discuss different topics through three rounds; and, by the end of which, all participants would have discussed all topics.

As we found on 7th July, the Design for Life mission's three pillars enormously helped Fellows focus their discussions. Thus, we created three breakout rooms – each targeting one of the three pillars – and participants were to move from one to another (freely) between rounds. In turn, each room was facilitated by one person throughout to ensure that each round built on the last.

The three pillars of the Design for Life mission are *Building Capabilities*, *Growing Hubs* and *Developing Infrastructure*; and they were facilitated by Ian Stobie, Marc Goblot and Sivliu Pirvu, respectively.

The tools employed for the workshop were Zoom for meeting online, and Miro for taking notes and making the event more interactive. The Zoom account was facilitated by the RSA, and the Miro board was made available by Marc, designed by Ismael, and introduced to attendees by Michaela.

Part Two: Areas of Interest

Everybody at the event was asked to share their name and response to the question: **What three big areas should we work on?**

Anonymised responses are offered in the appendix. Some answers were given by people before the event but then couldn't attend, and some didn't provide answers (hence some inconsistency with the number of attendees at the top of this document). Responses in italics have been written according to Ismael's understanding of what was said during the event, rather than attendees themselves.

The ideas shared during part one can be roughly categorised into six themes of questions. These include five themes already explored on 7th July, with one addition: "A" for "activism." Furthermore, whilst the "M" for "misinformation" was not explicitly brought up in the answers on the 21st, it has been maintained:

- **Regulation:** What do AI tools mean for intellectual property and copyright? How can we bridge the gap between fast-paced technological advancements and much slower

regulatory developments? What role does self-regulation play, and how do we standardise risk taxonomies?

- **Activism:** How do we ensure the inclusion of diverse voices in the design, development and deployment of AI systems? How are commonly marginalised voices given a platform to shape advancements in AI?
- **Creativity:** What creative methods are there to identify novel applications of AI tools? How can creativity – as a human capacity – help guide future AI research and development to meet real social needs? Conversely, what threats and opportunities do AI advancements pose for the arts?
- **Education:** How can we best inform the general public about both the potential and limitations of AI technologies? What does tech education look like for children, workers of all ages, non-workers, and retirees? Conversely, how can AI tools support education?
- **Misinformation:** What are the societal impacts of convincing texts, images and sounds resulting from generative AI tools? How can we combat the spread of misinformation and the crumbling of our socioepistemic institutions?
- **Economy:** How do we enable a more equitable distribution of the economic growth enabled by technological advancements? How can this counterbalance the current widespread violation of intellectual property rights, and ultimately inform better practices? What impact do AI tools have on diverse workforces?

Part Three: Design for Life

As mentioned the event followed the World Café Method. Three rooms were facilitated during three rounds. The rooms were focused on one of the Design for Life pillars: *Building Capabilities*, *Growing Hubs*, or *Developing Infrastructure*. Below are summaries of the discussions held throughout the three rooms.

Building Capabilities

Description: Change starts with supporting people to learn and lead. At the heart of Design for Life is the commitment to build the skills, connections and confidence people need to fulfil their potential.

Facilitator: Ian Stobie

The group focused on the potential for developing educational material on AI, as this was found to be of interest during the workshop of the 7th. The two main topics discussed were on

content and audience. "Meta" questions about how the efforts are framed for FRSAs were also brought up.

In terms of content, suggestions included:

- "Core competencies for AI" that can then be adapted to different contexts and industries
- Foundational technical knowledge
- Real uses and limitations
- Responsible use of AI systems
- "Human-centric educational materials"

Different audiences were mentioned as possible consumers of this educational material:

- "The general public," as was the case with the "Living with AI" course led by We and AI
- Primary schools
- 18-to-25 year-olds

"Meta" questions were proposed, including:

- **Can we work on a framework for developing AI education resources?** For example, should we seek to answer questions such as: what are the key indicators, competencies, and mental models needed to be covered in any course about AI? What needs to be explained about how AI systems work? Why is this knowledge necessary? What work has already been done on educating about AI? What sources are credible to draw on for such material?
- **How do we ensure the accessibility of AI courses?** Should they be free? How do we make sure nobody gets left behind?
- **Who needs to be involved in the development of such materials?** Two parties were mentioned: experts and those who are usually digitally excluded.

Growing Hubs

Description: People effect change by coming together. Design for Life grows the connections that organisations, communities, industries and regions need in order to flourish.

Facilitator: Marc Goblot

Building on findings from the workshop of the 7th, this group focused on mapping the networks that the workshop participants are part of. This was a pilot of the exercise to hopefully be carried out with the entire AI Interest Group's membership.

The idea is a visual representation of the Interest Group's membership and the networks they are part of. Participants were invited to share organisations they are affiliated to in some way, and then asked to reflect on what information should be captured in the network map. It was suggested that relevant information about the organisations includes:

- Size
- Purpose and
- Contact details or individuals known to Fellows.

The map is intended to help uncover potential collaborators, and identify clusters of organisations with shared goals.

For the RSA, the map can capture individual Fellows' skills, as well as the areas that individual Fellows are keen and able to pursue.

Underpinning the map, a "unique selling point" should emerge. During the event, the question of what the RSA brings to the "AI table" that is distinct from other groups was raised. The multidisciplinary nature of the AI Interest Group's members was reiterated on several occasions.

Developing Infrastructure

Description: We are aiming for systemic, long-term change. Design for Life develops and influences the conditions needed to enable people, communities and organisations to regenerate our world.

Facilitator: Silviu Pirvu

On 7th July, two definitions of *infrastructure* were discussed: technological and regulatory. This debate ensued during the workshop on the 21st, with added emphasis on the democratic process. On the one hand, we discussed widening access to technological infrastructure to decentralise AI for the public good. On the other hand, AI systems were suggested as infrastructure for democracy.

The Group discussed the government's role in making hardware and software available to the broader public. This could help decentralise the sort of compute power needed for AI developments, which is currently accessible mostly by big tech, or through certain academic affiliations. Key to this idea is framing AI infrastructure as a public good.

Regulatory infrastructure was seen as the potential for FRSAs to inform policy. One policy brought up was on *disclosing* the usage of AI systems – both in the public and private sectors. This discussion revolved around particular applications of AI tools, and considered their potential implications for democratic processes.

An example use case of AI as democratic infrastructure is to support the collection and analysis of large numbers of constituents. Narrative opinions from huge numbers of constituents on policy could be rapidly gathered and summarised with the right tools. This could improve diverse inclusive representation of view points to politicians.

Reflections and Next Steps

This section compiles reflections from feedback on the event's online delivery and tools, as well as action points for the future.

The event was the AI Interest Group's second Fellow-led event and was delivered online to be accessible to Fellows who can't easily reach London. Making the event accessible to different geographies was key to tap into the Group's international network. However, running the event online required greater resourcing than the in-person event on the 7th did. Some particularities of the online event include needing:

- **More facilitators**, although this is partly due to going with the World Café Method, which required one facilitator per room;
- **Online tools** – such as Zoom and Miro – which would ideally be sourced by the RSA; and
- **Note-takers** to support facilitators and participants in capturing discussion points (in this case, on the Miro board).

These considerations can inform future Group events.

The attendees all shared their views passionately, and the diversity of experiences and disciplines represented led to deeply insightful discussions. It was acknowledged, once again, that the Group cannot become merely a place for exhilarating discussion, but a space that fosters new ideas and effective interventions. Below are identified action points, attributed to the group they came from. Fellows are welcome to draft proposals for the areas they are most interested in.

- **Building Capabilities:** Develop a glossary to ensure a shared understanding of sociotechnically charged words and phrases, such as "defining AI."
- **Building Capabilities:** Develop an AI-focused brief for Student Design Awards.
- **Growing Hubs:** Create a "network map," visualising the different organisations and networks its members are part of.
- **Growing Hubs:** Articulate what makes us distinct from the many other communities in this space.

- **Developing Infrastructure:** Clarify what AI infrastructure means as a first step to focus efforts or even define different work streams or applications (e.g.: AI infrastructure for climate, for healthcare, for governance, and so on).

Appendix

Three big areas we should work on
Future of Work. Impact on employment (ie. WGA/SAG-AFTRA strike)
Regulation. Ensuring AI as a tool isn't abused by any sector or bad actor. (Creation of an RSA Guide to Responsible AI Use)
Education. AI literacy of capabilities, benefits, and issues
Regulating who does what with code and QA certification for their output
Education of users in their niches - students - care workers- voters- how to spot fake info
Ring fenced knowledge bases which can still use LLM techniques to provide value with out danger of poor QA and downright dishonesty
Identification: What constitutes an AI Application, of what type and/or application and how is it implemented/contextualised for use (private or public); and how is it to be identified by others? Think food allergens and their hidden use in food products. They must be known about so people can make informed choices.
Education: At all levels of education, the types, uses and limitations of identified AI applications must be (critically) understood by a greater proportion of society, again so informed choices can be made.
Justice: A framework of legislation governing identification and use must be implemented, with the means to govern a just use of the technology in the interests of society and the individual. Again we can think of food allergens and the requirements around food safety. Allergens are not prohibited, but in food safety we must properly identify their presence and (critical) potential.
AI education in schools - integrated teaching of AI as a sociotechnical system across the curriculum rather than just computer science

Three big areas we should work on
National AI literacy - a commitment to programmes fostering and measuring key principles about AI decision making, development and impact - accessible and tailored to the whole UK public. An essential extension to digital literacy (intersecting also with Data and Media literacy). To enable safeguarding, critical thinking, responsible use, active participation and inclusive participatory innovation for public good, democratic discourse
AI disclosure - labelling - a charter, guidelines or industry standards to communicate where, how and when AI is used
Ways for disabled and other people outside the mean to realise the potential of AI and equitably redistribute benefits - especially to enhance alternative models of learning, communication and expression.
Personal profiles that machine learn from you. Data sovereignty and reclaiming self determination.
Open Source and transparent approaches to ensure access for all and not only determined by Big Tech and how they want you to use it. How we make public and government literate to balance the power relations.
Future of work: Impact on employment and wealth concentration (e.g.: à la Luddite)
Lifelong learning: Employers' role in developing staff's AI knowledge
Education: Enriching the tech syllabus with critical thinking
Developing Infrastructure - Natural language processing with Large language models will revolutionise how we co-create a valid understanding of the world around us. We have, over the past few hundred years, depended on quantitative metrics due to scalability, ease of access and understanding to guide resource allocation (cost, dividends, etc). This has resulted in driving the market economy out of proportion to other areas we as humans value but find more challenging to measure (core economy based around family - spending time with children, public purpose economy - services, climate, etc
Other qualitative measures are expensive, prone to bias, and are not scalable (interviews, coding and thematic analysis, surveys)
Imagine a world where an MP could ask 1000 constituents for their free text opinion on a topic before a vote in 2 days time, where an LLM could immediately generate key themes for those 1000 responses instantly to guide the MP's viewpoint and discussions in parliament. Or where healthcare commissioners and councils could have free text based opinions of what is really important to their population to guide how we design services (ultimate form of co-production that is scalable). Do we put more money into new hospitals or improve child poverty?
Cultural capital - can AI provide opportunities for individuals to gain cultural capital thus enabling social mobility?

Three big areas we should work on
Lifelong learning - how can we develop a 'relationship' with AI so that it supports our lifelong learning journey (which includes our formal education, career/s, interests and 'retirement')?
Creativity - what is the potential for AI to help us access and nurture our creativity and develop our creative thinking e.g. finding solutions to combat climate change?
Education: Understanding AI - foster a better understanding of what AI is, its capabilities and its limitations, and prospects for the future. Feed in from "Opportunities and Threats" to provide education on where to use, where not to use and risk assessment.
Identify Opportunities and Threats - develop a taxonomy of opportunities and threats to humanity from the deployment of AI across a number of segments such as the arts and creative services, business, health, finance, education, transport, politics, defence etc. Create a spectrum of risk (similar to German Data Ethics Commission). Balance economic benefits and societal risk. [*given RSA's purpose one focus area could be "the Arts"*]
Safeguarding - identify what needs to be done to protect society both in terms of self regulation through education and legislation.
Licensing and regulatory mechanisms for data used for AI training - Where the regulations cannot do anything about it?
Progressing the Art & Science of AI that solves real world issues. Refocusing the public conversations on the opportunities to create value and save lives, not only recurring concerns and perspectives of ethics (especially ones that are independent of the tech, and mostly dependent on human uses).
Data standards for purposeful uses. - Which sectors are the highest priority to adopt data standards. Are there data standards in place (e.g. RICS for surveying) or new ones are needed?
AI in Health + AI for Health
AI in Education + AI for Education
Data Privacy and Regulation
Disability, race and gender: these are characteristics that AI tools can discriminate against
How will we challenge and tackle discrimination?
How are AI systems being developed, and who will they represent?

Three big areas we should work on
Create RSA **Foundational AI knowledge learning module** - MVP for Fellows Networks use - feedback - then formalise as RSA tool for wider use. Possible commercial product to RSA corporate commercial & NFP orgs usage.
AI Usage Framework Assessment Tool - MVP Fellows/RSA use in DfL initiatives. (will include best practice and compliance / regulation). Possible commercial product. May need Fellow partnering services to deliver follow-up
Economy & Business Network Tech4Good - Competition - What good use of AI looks like' - DfL initiatives with use of AI - challenge to Fellows & all Networks for project initiatives/ideas. Who tech assesses & votes on these? Sources of funding and partner tech co support & Fellows time / expertise.