Issue: 10 Date: May 2024

# CAISS

Computation & AI for Social Science Hub

**Newsletter in collaboration with The Alan Turing Institute and Lancaster University**

*Sign up now as places are limited*

## The CAISSathon – 22nd & 23rd July 2024, Exeter

*Please find more details on page 4*

### CAISS Editorial Report from AI UK – London March 2024 – Some highlights

**Saqib Bhatti MP** opened AI UK with an address that emphasised how important this area is with the market for AI being one trillion dollars, it therefore needs to be pro-business, pro innovation and pro safety. He spoke about how widening the skills pipeline is crucial - there are nine new AI research hubs in the UK already with a hundred million pounds being awarded to the Alan Turing Institute in the spring budget 2024. Collaboration is vital now to help shape global technical standards. He sees three key areas that need to be focused on:

1. Data - foundation AI models (a form of generative AI) leads to an increase in "power", (generating comprehensive detailed outputs), allowing access to essential public data assets plus security for the data. Safety is key here.
2. Regulations – we need the right rules which will lead to trust; systems will not be adopted without them being trusted which could then stifle future innovation. [Will regulations lead to trust though? – Ed]
3. AI ethics - we need an open-minded approach. It's too easy to stifle innovation and there are no easy answers; however, transparency is key. Fairness innovation challenges can lead to grants being available which can help to tackle bias for example, it's important to get the regulations right.

**Professor Dame Angela McLean (MOD Chief Scientific Advisor)** gave a keynote speech where she confirmed she sees two cases for AI. 1. A proper scientific approach and 2. A systems approach (where AI tools and applications are integrated not stand alone).
The UK needs a science and technology framework to become a science superpower.
She is an AI optimist and excited about what AI can do, but it needs to be safe and responsible. There is a whole Government approach to use the Science & Technology (S&T) framework from 2023 with a focus on foundational R&D. This framework is the strategic anchor that government policy will deliver against. She emphasised how we need an innovative public sector, especially as we have a once in a generation opportunity to grow AI with ideally a sociotechnical approach being taken.

**Professor Steven Meers (Dstl)** discussed how the purpose of Dstl is to assist with defending the homeland, help shape global strategic intentions, shape international defence and de-escalate conflict. Most of the "fighting" military personnel are in their 20-30s and they should have the best science and technology available to do their jobs. AI is a ubiquitous technology with all areas of defence impacted. We want people to be using their judgement and not doing mundane tasks which could easily be automated. He gave the example of the "Sapient" system – which protects the perimeter of a military base. This system is using AI and autonomy to operate a network at the sense and fusion level to resulting in good coverage as a sensor system. Sapient is now a British Standard. The UK is also working with the USA to counter misinformation – for example working with DARPA (USA defence research and development agency) to counter deepfakes.

*Continued on next page…*

[dstl] The Science Inside

The Alan Turing Institute

Lancaster University

### [ ] Work with us

Please email us if you wish to be added to the mailing list.

### [ ] Contribute

Have something to share, please let us know and we can feature it in a future issue.

### [ ] Spread the Word

Please feel free to pass this onto your colleagues.

### [ ] Contact Details

**caiss@lancaster.ac.uk**
caiss@dstl.gov.uk
To email us at Dstl, scan the QR code

DSTL/PUB158211

*Continued from previous page:*

Steve went on to discuss whether we can identify what is holding AI back in defence? There is a paradox with AI, how we take things from the lab and take them into the real world can be challenging, especially as AI is a sociotechnical thing, there is a competition of ideas and ethical values. Great care is needed to navigate the area. We all need to apply critical thinking and understand the provenance of the situation/information. Should watermarks be used? We need fast solutions. How do we define acceptable error thresholds? We can do this by using systems within a clear framework with the rules of engagement strictly followed.

*AI UK was engaging over the two days with three stages running in parallel. More information can be found here*: https://ai-uk.turing.ac.uk/programme/

## Is the YouTube algorithm radicalising people?

About a quarter of Americans get their news via YouTube which is one of the biggest online media platforms in the world. The media have hypothesised that young Americans are being radicalised by the content they are viewing on the channel due to algorithmic recommendations - leading these viewers in a certain direction. However, the University of Pennsylvania's Computational Social Science Lab (CSSLab) has found that individuals own preferences and political interests are the primary driver around what they watch.

The researchers created bots that either ignored the recommended content or followed it by using the watch history from 87,988 real life users. On average sidebar video recommendations shifted towards moderate content after about 30 videos, whilst homepage recommendations adjusted more slowly. The results of their experiments showed that sidebar videos are more related to the content of the current video being viewed and homepages are linked more towards the viewers preferences.

**So what:** The accusations hold some merit but crucially viewers also have some agency over the content they watch. The researchers hope that for the future they can study how user preferences and algorithms interact to be more cognisant of how algorithmic content recommendation engines impact individuals daily lives.

Link: https://techxplore.com/news/2024-02-youtube-algorithm-isnt-radicalizing-people.html

## Good writing is about having something to say

Our own unique perspective means that when we write it is hopefully interesting and original, backed up by evidence. Generative AI tools could prove helpful in both explaining content and aiding our understanding. But Generative AI tools can lead to hallucinations (wrong information) and even knowledge cut-offs when used to help write a piece. Are writers forgetting that GenAI can't provide critical thinking, understand the broader contexts or challenge ideas?

This editorial in Nature Magazine proposes that GenAI can be useful in other ways when "brainstorming or "story seeds" are required to help human writers overcome the anxiety they often feel when staring at a blank piece of paper or screen. Whilst this can be helpful, "clichéd nothingness" that these systems often output is a very unsurprising result. By using prompts precisely to fully articulate your ideas the generated output will be more useful. But the very effort of thinking about and articulating your ideas into a robust prompt or prompts does still not give the AI tools agency.

**So what:** Generative AI is not a "knowledge partner". Most output is "stitched" together and is "more or less logically connected platitudes". However, if you have to write in a foreign language it can help you with giving you better synonyms or idioms for example. It can be used for translation into another language or adapt a piece for a different audience. Caution is needed as AI systems lack agency in doing or writing about science or anything else for that matter. GenAI is useful but beware of its limitations.

Link: https://www.nature.com/articles/s42254-024-00713-4

*Continued from previous page:*

## *Article Review*

### *"Artificial intelligence and illusions of understanding in scientific research"*

Artificial Intelligence (AI) is being embraced by all types of scientists from replacing participants in social science experiments with bots to self driving laboratories where robots and algorithms are working together to devise and conduct experiments. Should we be worried about this, are researchers in danger of overlooking the limits of certain tools or even being lured into a false understanding of concepts.

Lisa Messeri – an anthropologist at Yale University in New Haven, Connecticut and Molly Crockett a cognitive scientist at Princeton University in New Jersey feel that risks need evaluating now before AI systems "become too deeply embedded in the research pipeline". The authors examined 100 peer reviewed papers, preprint, conference proceeding and books from the last five years and drew together an assessment of how "scientists see AI systems as enhancing human capabilities". They came up with four viewpoints/visons:

1. AI as Oracle – researchers use AI tools to survey the scientific literature in depth.
2. AI as Arbiter – systems evaluate scientific findings more objectively than people.
3. AI as Quant – AI tools surpass human mind in analysing vast and complex data sets.
4. AI as Surrogate – AI tools simulate data that is too difficult or complex to obtain.

The authors predict that various risks will arise from these "visions" such as "the illusion of exploratory breadth". This could result in an AI encouraging experiments involving human behaviours that can be simulated by an AI and discourage those that would require real human participants (AI as surrogate). Researchers could forget that the viewpoints found in the data used to train the models will contain the same biases as those in the training data.

A considered approach is required when using AI says Crockett, AI is "not a panacea" but a choice with risks and benefits that need to be weighted up carefully.

When AI is evaluated ethical concerns are often cited including: algorithmic bias, public misunderstanding of what AI is and does, environmental costs and even exploitative labour costs. The authors posit that technical approaches alone "are inadequate for addressing ethical concerns" but feel that "exclusively technical solutions" will be used to address the concerns.

Some individuals may use AI to boost their own cognitive limitations by using AI tools as "knowledge-production partners". This approach could lead to a "phase of scientific enquiry in which we produce more but understand less". AI tools can help mitigate the problems of time constraints, fixed budgets and cognitive capabilities – but will this enable scientists to be more productive and more objective? The benefits of AI need to be fully understood in every situation. They may limit understanding not enhance it and everyone who uses such tools will need to fully evaluate "their potential epistemic benefits".

**So What:** There is a lot of "chatter" around the use of AI. Nature magazine recently stated that "Tools based on large language models (LLMs) such as SciSummary, Scholarcy and SciSpace can help researchers to speed-read the literature and make studies accessible to non-experts. AI-generated summaries could also aid people not writing in their first language, says biophysicist Esther Osarfo-Mensah: "Some people hide behind jargon because they don't necessarily feel comfortable trying to explain it." It is important to be aware that AI summaries could introduce errors or strip some of the subtleties away from information thereby changing the context. AI does not have superhuman abilities and researchers need to be aware of the risks this could create. AI can help with mundane tasks but is no substitute for good, evidence based research. The authors argue that we are "producing more but understand less".

Link to article: https://www.nature.com/articles/s41586-024-07146-0

*Continued from previous page:*

## The CAISSathon, July 2024 - Reserve your place now

***Where: Exeter University***

***When*: 22nd and 23rd July 2024**

**Theme of the event: Explainability**

The social responsibility of Artificial Intelligence (AI) has been under increased scrutiny as it creeps into every facet of society. Understanding the potential ramifications and harm caused by AI is of key importance, particularly as AI technology used for facial recognition, loans and mortgages and job applications (amongst others), have already been shown to be biased against ethnic minority populations and, for example, those with disabilities. Within a Defence context, understanding how AI enabled technologies can facilitate decision making is of key interest. The need for explainable and transparent AI systems is one argument to uncover bias and prevent harm. However, does explainability solve these issues? Does understanding how AI makes it decisions provide enough evidence to negate the harm or at least provide indications of potential harm? Additionally, how understandable do such explanations need to be for expert and lay users?

The Computation and AI for Social Science Hub would like to invite you to participate in our first 'CAISSathon' to explore how explainability can be conceptualised and implemented. This event will propose a series of challenges that will be collaboratively addressed throughout the two days. It will bring together individuals from government, academia and industry to brainstorm and engineer potential solutions. Researchers will have an opportunity to put knowledge into practice and solve problems which have real life implications within Defence. At the end of the two days, a portfolio of potential solutions, research questions and collaboration will be established which will inspire future investigation and lead to insightful developments in this fast moving field. Additionally, there will be a prize awarded to the team who prepare the most innovative solutions. Come and join us at this exciting event.

***To request an invite please email us at <u>any</u> of the details on the first page***

## Can bias and racism be fixed in AI image generators?

Back in 2022, Pratyusha Ria Kalluri from Stanford University asked an image generating AI system for a "photo of an American man and his house". The result was a pale skinned man outside a large "colonial style house". When prompted for an image of an African man and his "fancy" house it produced a dark skinned person in front of a simple mud hut. Investigating these results Kalluri and her colleagues found that all of the popular tools resorted to common stereotypes and even amplified some biases. For example flight attendants were women, housekeepers people of colour, doctors as men and nurses as women. Should we be surprised at these results as our society is full of such stereotypes and these AI system will just keep amplifying the existing stereotypes?

**So what:** This issue needs to be tackled now before these exacerbated stereotypes and biases become firmly entrenched. This could be addressed by making more tools "open source", then we can identify where biases are and mitigate them. Improving training data sets is however, time consuming and expensive. The big question here seems to be, "do we want our AI data sets to reflect reality, even if this reality is unfair?" Should we be striving for equality, non bias and minimising misinterpretation and misinformation as standard?

**Link**: https://www.nature.com/articles/d41586-024-00674-9