



ukgovcamp

Session: 5

Room: M69

Session title : Mitigating social effects of AI

Session leader : Callum

Volunteer to continue conversation after :

Notes taken by : Ian Ames

Notes

Opener from Callum:

Get ideas on managing the risks inherent in AI adoption in the UK. We have a moral obligation to adopt AI safely. The quality of life and quality of outcome is outsized compared to other technologies at the moment. Life will change so how do we do that ethically?

To inevitable, the diffusion theory of technology, if you have a sufficiently useful technology, it will suffuse. There's no such thing as unbridled and only good technology. When barbed wire was invented, the thought was on containing cattle, tin cans, were for food preservation, both were instrumental to world war 1 "industrialising" war.

It's a down sight more obvious than canning that AI has potential for mis-use. There is room for the establishment of norms and behaviour *and of laws*.

Archivists have values and a code, not falsifying records is their core. What's the values and codes for practitioners?

Is AI being driven by technologists, or by unethical people? These systems are very expensive, what pressures does this introduce?

More visibility of how AI models work would help. It's a black box system. How do we know we're being "evil" if we can't see how it works? Will open data and open source move more into this space when it's so proprietary and expensive?

What's a specific? Job loss, how do we moderate those effects, how do we slow down? I don't think we want to slow down, are we doing all we can for people, no? Can we do better with this tech, yes?

We don't really know what the impacts will be, we don't really know what we're offloading to AI? We can't see or understand the connectiveness/ narrowness of offloading. It has the potential to shift things so much, do we really have the concepts to respond? How do we re-establish concepts that address these things.

For example: Humans think language is descriptive of reality, but it is also performative, as we use it, we are changed, as we change the world changes. LLM's do performative, how does that lens help us manage risk.

Shannon Vallor: The AI mirror, and the technology and the virtues are worth reading.

If all you think your role in life is to be productive, that's not a broad enough question of what it means to be human.

AI being 'confidently wrong' is AI already unethical?

It reflects the way we talk, it shortcuts the way we communicate.

Our services are being pummelled for gov content, with no regard of how they are maintained or funded. Should tax payers fund Open AI's use of services like national archives with no costs. There's a whole side of economics that are currently irresponsible and highly extractive, it's incredibly exploitative right now. This needs the protocol, the money, the standards to solve this, and it's damming on the AI industry that they have not led or taken any responsibility. They've displayed the worst behaviour you see in a gold rush.

With capitalism people are empowered to vote with their wallet, but this isn't in play with the way that AI companies are currently operating. What happens when the money tap ends?

No one is taking accountability on holding organisations to account. How do we address this. Accountability is starting to happen (police guy that lost his job recently, is being held to account for his use of AI). These ways and boundaries for when it's fine and not fine to rely on AI outputs will develop and change. Our norms are always to check and verify in archives, that may vary and develop given the range of how adoption. Some people aren't using it at all, others are heavy users.

How do we build, deploy and scale to help frontline services and we're grappling with this issue? What practical support do we provide and monitor when they're in the real world? Not just of the risks, but of the benefits.

- We've evolved through telling stories, we need to build the culture and feedback, people won't learn from abstract documents and policy. For example when mRNA vaccines came out, everyone knew what they were because they were part of the story of COVID.
- Finding people to take accountability can be difficult. Currently you'll be crucified by judicial review. You can hold a human being to account for a decision. It's not important that they make a right or wrong decision, it's that they are accountable.

- Non-deterministic models result in inconsistent responses, how do we account for that in accountability?

Fragmentation, shared stories are the way to learn as a society, but LLM's take fragments of that narrative, and hand parts of it back, without clarity on how its been fragmented.

How do we manage it?

Humans in the loop - humans are susceptible to trust or laziness.

We're already guinea-pigging this on a generation that's been through covid, homeworking is removing the opportunity to learn from peers, now we're saying entry-level jobs may be replaced.

What is "cool down the pub" in 10 years time? (I'm happy with my involvement in how this has developed)

There's a role for legislation. Do we take the american model of high level of regulatory environment versus low level regulation. Data protection and date privacy legislation is showing us already how companies approach responsibility.

How do we evaluate in the public sector, there's slowly a muscle being built, [the testing and evaluation document](#) published last year is the first we've seen of that.