ukgovcamp

Session: 4
Room: HMS Daring Room

Session title : Agentic AI & Building for AI Agents (services that humans won't use)

Session leader : Paul & Alex

Volunteer to continue conversation after :

Notes taken by : Darren

Notes

- Lots of hype around AI since the launch of ChatGPT but there's a bit of "magic beans" around AI in general. Agentic services are different to generative (ChatGPT, etc.) and we are now starting to see agentic caseworking systems emerge.
- Session leaders see agentic AI as the real changemaker as they could have a massive impact on the administration of the state, doing all the back-office work needed for decisions to be made.
- Moral dilemma – public services should feel human, e.g. social workers should be doing social work, not paperwork.
- How do we do this well? – technology, ethics, security, job losses, etc.
- The use of AI means we no longer control user interaction – e.g. Search referrals are shifting from Google to ChatGPT, but they are often buried deep in the journey and not how we'd expect the user to find that page or info.
- Agents can also reduce barriers, especially for people traditionally digitally-excluded, because it can do a lot of the grunt work for the user.
- It's important to recognise that we need to maintain good, managed data and accept that Gemini and the like are going to scrape it and regurgitate it. Can we encourage agents to behave (and scrape) in a certain way?

- Analytics: We can only see what the agent did, not the steps it took to get there, so how can product people take sensible decisions with only part of the information?
- We also need to make sure we understand the boundaries – when should a human be doing a task or interaction, and when is it OK for an agent to do it?
- OpenAI now has apps (launched in the UK around a month ago, in the USA earlier)
- What does human oversight of agentic AI look like? We are guiding a series of LLMs through tightly-defined steps and gates to achieve an outcome; lots of checks, balances, and guardrails.
- A new discipline of service design is needed for working out these interactions and workflows, especially when humans and AI agents are in the same loop
- There is a risk that money saved through the use of agents will just be that – savings, not reinvested
- Where does the buck stop for the quality and security of an agent? Who ultimately carries the can if something goes wrong? What monitoring is there, and what level of responsibility does a human in that loop hold?
- Commercially-licensed agents also need to be held accountable and rules set by the gov dept
- Concern over the accuracy of outcomes as well – example given about rewriting a piece of text via three different AI apps and one of them got it completely wrong.
- If something doesn't work properly for everyone, who is going to be brave enough to take a call on whether to launch it, taking into account the pros and cons of something that isn't necessarily right for a subgroup?
- The better an agent gets, the more likely it is that people will start trusting the outputs and outcomes and will reduce or stop the human sense-check at the end. How easy would it be to correct the output?
- Will this always be around, though? The tech is incredibly expensive to build and run and it's easier to quantify the cost of a human doing a task than the volumetric cost of the software and hardware.
- Can we see a future where there are off-the-shelf agents available on government procurement frameworks? A lot of these models are open-source and easily reproducible.