ukgovcamp

Session: 2
Room: Banqueting Suit 1

Session title : **AI in public services - how do we stay in control**

Session leader : Shaid

Volunteer to continue conversation after :

Notes taken by : Anju Dhir

Notes

Banqueting Suite, Session 2. Shaid.
Session Overview: AI in public services, how do we stay in control in AI. Move from pilots to frontline services (homelessness, care), and what happens when we go live, stay in control, harms earlier, not governed well, need to tackle challenge together. 3 exam questions- using 1, 2, 4 All technique.

Kat: on loan from BCC and now MHLCG.
Raam: AI Local Govt directorate, and A.AI good set of tools and how to take in local govt setting.
Softwire- control sites, self learning, robots, science fiction, Claude Code, what evaluations using at technical level, reduce model drift over time, test and use new models and right set of architecture, and tools for AI services. Need proper insights and maintenance of this and ensure is it still working right. Don't want legacy of wrong tools remaining.

Shaid- improve outcomes. 1. Once AI is live, what should we be watching. 1 = 1 think for 1 min, 2 = conversation for 2 mins, and get into 4 to discuss.

1 min think: hallucinations, credibility, AI and humans make mistakes but also get stuff right, so work together and reduce risk and issues, and see benefits. Best of both.

Discuss in pairs for 2 mins.

Discussion in 4’s for 4 mins. Dip sampling and audit ability, as to are things benign right. Technically adding in AI in a system, is this done right, and is it over performing and underperforming. E.g. library. Of data and archivists and ai follows rules. Top 3 things to play back.

- Best of both as ai and humans are always right, human process and where best I sAIU served.
- Dip sampling and audit
- Technical evaluation and guard rails in the system.

Discussions and feedback
Attendee: what causes harm with any service deploying, engineering and tracking, staying in control, integrated AI, and shifting responsibility

Attendee: how you measure impact and what type of AI using- gen, chatbots. Drift and ask golden questions, to throw in and quality assure, so AI not drifting.

Attendee; micro harms and everyday and ensure how we’re measuring from this use, ensure outcome tracker don’t look sight of micro groups and not changing how groups are affected, e.g. digitally excluded. Don’t fall in trap of quality risks. Trust and societal change and measure proxy’s for trust and maintain this.

Attendee: best both and humans and AI make mistakes and only good on what we build it, where best are we placed. Model drift and audit trails, and dip sampling for explainability, and not use if black box. Technical evaluation and potential self adaptive self systems.

Attendees: domain dependence, measure impact. Measure quantitatively, how quickly does it take to go to the next step, having sessions with users and if working for you and making better. How can share results with social and geographical groups, e.g. what works in London may not work in Wales. Share projects to increase adoption.

Question 2: harms and risks. What are harms and risks that once show when deploy AI in real world. People put trust in AI on mental health concerns, cost lives. How to mitigate this.

Fringe cases as can’t assess for anything. Lots of variables across.

Scaremongering and AI take jobs, and not know how to do a process as take short cuts.

Group 6 discussion
AI is 100% correct, use ChatGPT as a therapist. Real world believe AI as gospel.
Statistical vastness of real world- test in smaller group, and launch in 70m people and won’t see these in test.
Scaremongering of AI taking jobs, and skills being taken away, assessing public and 100% dependant on process.

Catch 22, implement in gov but don't see effect in public, e.g. like social media. How do we identify risks early and clearly label them, mitigate and bring trust. AI- credit score, visa approval, and people will hate it. Test AI in 22 council;s, and what about unintended harms. Keep knowledge and skills in place.
AI making decisions. Humans make mistakes, AI make mistakes. Doing review work and different to doing real work, so train and check like a line manager. Model may drift and give people first class upgrades, and harm of AI compared to people.

Discussion as collective
Attendees: most models are trained outside of UK government institution, ways model constructed tells about bias, training of them, and limited control of them, so dependant on the groups trained model, e.g. open AI and deep seek. Getting addicted to AI, successful implementation, and who is brave enough to stop project, and automation very well become very dependant, and if switch off, then are employees there to provide support.

Attendees: 1. Using ChatGPT as a therapist and suicides. We are tech savvy but not everyone has the same understanding, and trust AI, having disclaimers is vital. 2. Can't test for every possible eventuality and release to a population, and no way test for all fringe cases. 3. Scaremongering, but lose fundamental skills to AI and trust system as can't check as AI, automation of PA's. 4. Catch 22 progressively promoting until too late, and social media is not as great as it was before. 5. AI decision making, humans make mistakes, treat AI as humans, and resp like managers is res.

Attendee: is risk and context happening team needs to consider. Rush to deploy, failures and monitoring, known and unknown risks. Ability to adapt and fall back to previous process.

Attendee: misinformation is constantly changing, incorrect information, AI is more malleable, constant change of AI model. Lack of confidence, chatbot wrong direction and lack of confidence pull back. Local context and not developed in UK, e.g. Bham, regions etc. lacking local context.

Third question- cross sector supplier, how do we deploy cross sector AI- assurance, how we scale.

Shaid to follow up:
Lorraine Ldaley@hotmail.co.uk
Gavin Maguire gavin.maguire@necsws.com