The ethical implications and challenges of using AI in public services

AI is already being used across parts of the public sector to help to allocate resources, keep records, analyse patterns, and predict needs and risks. So, when algorithms are being used to inform or influence decisions about real lives and situations, it is critical that practitioners are aware of how AI makes conclusions and what the ethical implications of this are.

One of the main issues to be aware of is system bias. AI doesn't create bias - it learns it from the data that teaches it. And this bias may be entirely unintended or even invisible in small data sets but could become apparent in big data. For example, in children and families' services, there may be differences in how families living in rural areas and families living in towns and cities have received support in the past that reflect their reliance on infrastructure such as transport, access to education and their social mobility. If, historically, certain communities have received less support due to barriers or inequalities, AI systems might misinterpret this as 'learning' that some families need fewer services.

This could be a result of an absence of data as well: either leading to AI not understanding that data is missing or working to compensate for this,

filling in the gap by ‘hallucinating’ a false
but seemingly coherent narrative.

In this example then, these factors could mean some families are incorrectly assessed as needing less support or low risk, with resources then unfairly being directed away from the very communities that actually need more support.

Poor quality data leads to poor quality decisions. And this is where the human input has a clear responsibility: If notes and records are inconsistent, incomplete, or biased, AI systems may amplify these biases, inconsistencies or draw ill-informed conclusions...

Human behaviour and professional practice is key here. For example, a practitioner who knows that in the service they work in, AI will scrutinise any records kept, might then alter how they write their notes. This could positively or negatively lead to conscious or unconscious manipulation of the data that AI uses to examine the notes and suggest actions. Understanding this matters.

All these circumstances could create unfairness in the system being used. Whilst AI has been given the data to use or has generated new data based on existing information, it doesn’t have the experience, judgement, or indeed human empathy, to be able to make qualified decisions about people needing support, for example. That is why a practitioner needs to always be in the process to use their judgement to assess what actions might deliver the most appropriate outcomes.

Practitioners should also be alert to issues around privacy and transparency. For example, in social work services, individuals often share sensitive information with social workers and, of course, there are already strong safeguards to protect people’s privacy and data. The idea that the use of AI to analyse data in new ways that reveal patterns that might help to shape services and outcomes to better support people is a very attractive one. But this introduces a number of considerations: should people be asked to consent to their information being used in these ways? Do people have agency in these contexts and circumstances to do so?

Transparency is a broader consideration here too: practitioners need to understand the AI tool or system well enough to see how it has produced a recommendation or suggested course of action.

Practitioners and organisations will want to ensure that they are making the determination of suitable next steps or signing off decisions that are in line with best practice and the ethics and values the service intends to work from.

People using AI in public services then have a number of areas to consider. For health and social care practitioners and decision-makers concerned with responding to the needs of people, there are some central critical questions to keep in mind:

- What data trained this system?
- How do we know the data isn't not biased?
- Where is any new data being generated or fed into the system being kept?
- If we're using a tool to recommend actions, can it explain how it has reached these recommendations?
- Who is accountable if our AI tool gets it wrong?

This all means that AI has the potential to benefit social work, but it should be used to enhance professional judgment, rather than replace it.