

Breakout Session Summary: Trustworthy and Explainable AI for Clinical Decision Support

Moderator: Dr Alexander Hill (LIV.INNO Data Science Fellow)

Participants:

- Dr Finn Allison (Virtual Engineering Centre, AI Architect)
- Mr Sean Austin (CHIL, Market Research and Insights Analyst)
- Dr Deepti Bhargava (Walton Centre, Consultant Neurosurgeon and Innovator)
- Ms Nicki Allen (Haptiv8, CEO, Clinical Entrepreneur, Physiotherapist)
- Dr Andrew Davies (Founder, Healthcare Innovation Consultancy)
- Dr Matt Bright (PDRA, Distributed Algorithms CDT)

Overview

The integration of Artificial Intelligence (AI) into healthcare is accelerating, with clinical decision support systems (CDSS) emerging as a critical area of application. However, challenges remain around trust, transparency, explainability, and ethical deployment. This breakout session convened experts from clinical, technical, and commercial domains to explore opportunities and risks associated with explainable AI (XAI) in supporting clinical decisions, with the overarching aim of ensuring better, safer, and more efficient patient care. The conversation aimed to bridge real-world clinical needs, data science capabilities, and market insights, with a specific focus on identifying actionable and trustworthy applications of XAI.

Summary of Discussion

Purpose and Value of AI in Clinical Decision Support

There was broad agreement that AI systems in healthcare must ultimately improve patient outcomes - either by enabling better decisions or by delivering current standards of care more efficiently. Explainability and transparency were recognised not as ends in themselves but as critical enablers of trust, safety, and integration.

Ethical Boundaries and Use Cases

Ethical concerns, such as potential for profiling, were raised. Participants discussed a distinction between clinician-facing and patient-facing AI tools, noting that the ethical bar may be higher for direct-to-patient applications.

One example of a low-risk, high-impact application was ambient transcription: real-time recording and summarisation of clinical consultations. This could free up clinician attention for more human interaction during appointments, and enable longitudinal analysis of unstructured data in Electronic Health Records (EHRs). Clinicians would remain in control by proofreading and approving transcripts.

AI in At-Home Care and Physiotherapy

Applications in physiotherapy and home-based care were discussed, especially around monitoring treatment compliance, flagging issues like pain increase, and suggesting minor plan adjustments. This could improve the effectiveness of care, increase the specificity of reviews, and overall enable more responsive, personalised treatment.

In the private sector, use cases were discussed where AI personas assist patients using medications (e.g. for weight loss), offering real-time feedback, reducing the need for large human teams, and potentially encouraging more open symptom disclosure.

Data Integration and Wearables

Participants highlighted the potential of integrating data from wearables, mobile phones, and sensors to support clinicians in offering more personalised care. This was recognised as an area where AI can synthesise multi-modal inputs into meaningful and actionable summaries.

Emerging Collaborative Ideas

The group explored opportunities for interdisciplinary collaboration, particularly around real-time AI-guided interventions for movement disorders and neurorehabilitation:

- Dr Deepti Bhargava described a neurorestoration device that detects and prevents falls via real-time muscle stimulation.
- Ms Nicki Allen discussed a device used by non-verbal autistic individuals to correct gait using positive feedback mechanisms.

There was excitement about the potential for sensor data fusion, AI-based adaptive control, and integration with population-based data sets, though a precise project was not finalised. Further discussion was proposed to define a collaborative pilot or research proposal, perhaps targeting the Global Challenges Research Fund or HDR UK Driver Projects.

Explainability as a Tool for Human Learning

The session concluded with a reflection that XAI doesn't only support regulatory compliance or clinician trust, but can also enhance human understanding of patients, care patterns, and systemic issues. In some cases, this may be more valuable than short-term gains from high-performing but opaque models. Issues of data ownership and patient consent were also flagged as unresolved areas needing clarity and co-design with stakeholders.