

Phase 2 – Codebook

Theme 1 Patient Safety and Ethical Accountability

Captures how participants conceptualise patient safety as a primary motivation for engaging with XAI, and how ethical accountability is negotiated across socio-technical systems.

Code	Description	References	Data Example
Patient Safety	Concerns about preventing harm to patients, including misdiagnosis, inappropriate treatment, and system failure in real-world clinical use	10	S5: “Doctors have this statement, “do no harm”, basically everything they do should first put patient's safety at the top priority”
Accountability and Liability	Discussions of responsibility distribution when AI systems cause errors, including developers, clinicians, institutions, and regulators	3	S6: “Another one is liability when something goes wrong. You need to find who is liable for the damage occurring.”
Responsibility	Reflections on professional and moral responsibility when deploying or relying on AI systems in healthcare	2	S1: “I would tend to think that the responsibility is maybe shared by many different actors. For the medical practitioner part, I think they are definitely still important agents because they are the ones who make the ultimate decision.”
Autonomy	Considerations of clinician autonomy in decision-making when AI systems are introduced	1	S4: “These systems have far-reaching consequences for people’s autonomy, for me this is clearly something that UX people can clearly contribute to, making sure people still feel in control of their lives while interacting with these AI systems.”

Theme 2 Bias, Fairness, and Power

Captures ethical risks related to bias, discrimination, and power asymmetries embedded in AI systems and their deployment contexts.

Code	Description	References	Data Example
Bias and Fairness	Concerns about biased datasets, algorithmic bias, and unequal performance across populations	11	S6: “Then you would have everything that is related to patient rights and fundamental rights—I mean in the public health approach—because there are these risks of discrimination.”
Representative Data	Emphasis on the need for training data that reflects the target clinical population	5	S6: “So, it means that you can only access and process data or collect data only where it is necessary, but no more. And now, you know, that's this big data issue with AI wanting to access and process more and more data—the most possible data.”
Power Dynamics	Reflections on asymmetries between stakeholders, including patients, clinicians, developers, and institutions	5	S1: “I think what’s most dangerous sometimes is when people see how powerful AI is and see the positive sides that AI can be done, and think that once we adopt it, we bring the benefit to everyone, but not realizing there can still be a gap, and not realizing that the gap can be either replicated or exacerbated by the existing disparity.”
Values	Explicit references to ethical, professional, or societal values shaping AI design and use	2	S1: “I’m thinking of this kind of thing—how it’s designed as a default. Whether that already encodes some values and whether those values should be made more salient and presented to the user when they begin to engage with the AI.”

Theme 3 Data Governance and Consent

Captures legal and institutional constraints shaping healthcare AI, including data protection, consent, device regulation, and cross-institutional sharing.

Code	Description	References	Data Example
Data Protection and Scarcity	Accessing sensitive health data under data protection and minimisation principles	6	S6: “So, it means that you can only access and process data or collect data only where it is necessary, but no more. And now, you know, that's this big data issue with AI wanting to access and process more and more data—the most possible data.”

Code	Description	References	Data Example
Information and Consent	Informing patients, consent withdrawal, secondary use, and reuse constraints	2	S6: “here’s two—there’s information and consent. It’s one thing to be informed and say, okay, usually when you go to the doctor, you will not consent. You will give general consent to be treated and have the surgery and all, but you will not give your consent for each tool being used.”
Accountability and Liability	Attribution of responsibility when harm occurs across the AI development–deployment chain	3	S6: “Another one is liability when something goes wrong. You need to find who is liable for the damage occurring.”

Theme 4 Trust, Reliability, and System Behaviour

Captures how trust in AI systems is constructed, challenged, and calibrated in clinical contexts.

Code	Description	References	Data Example
Trust in AI	Factors influencing clinicians’ trust in AI systems, including reliability and prior experience	14	S1: “But I can also see why a total lack of trust—meaning that people completely reject a tool that might have some value to them or don’t engage with it at all—that they could have some miss-out. But, yeah, I think trust is tricky.”
Risks of Over-Reliance	Concerns about automation bias, over-trusting AI outputs, and reduced critical checking	5	S5: “It could be that in the future people have been working with this AI system in the hospital for years, they don’t check anymore... doctors don’t even think about that result and they just blindly trust the system. It’s like death by GPS.”
Reliability	Expectations that AI systems behave consistently and predictably in practice	3	S3: “When you get it already, then you have to make the decision: “Do I leave it?” The machine is seeing something there, seeing hyperintensities that I have problems identifying. So then you have to decide: “Do I trust this that I’m not seeing properly?” But apparently, they are there. Or do I remove it?” This is a decision sometimes where you have to, as you said, trust the machine.”

Code	Description	References	Data Example
Performance and Accuracy	References to diagnostic performance, metrics, and comparative accuracy	3	S5: “And then it's also from the side of the user to demystify any wrong perception about the system, how the system will figure out, like wrong annotations like the human needs.”
Robustness	Discussions of system stability under varying conditions such as noise, motion, or domain shift	5	S5: “We add noise to the images, we simulate patient motion... we down sample the image like crazy to see if the system still works. You stress the system.”

Theme 5 Explainability, Interpretability, and Transparency

Captures how how explanation-related concepts are understood, valued, and operationalised by different stakeholders.

Code	Description	References	Data Example
Explainability	Expectations that AI systems can provide understandable reasons for their outputs	12	S4: “For me, explainability is a necessary prerequisite for being able to contest the system.”
Interpretability	Emphasis on making system behaviour interpretable rather than fully explainable	5	S6: “Legally, I would rather use interpretability as something which we should try to achieve.”
Transparency	General calls for openness about AI use and system behaviour	1	S5: “The best level of trust comes with transparency.”
Transparency about Limitations	Explicit communication of system limits, failure modes, and uncertainty	2	S5: “Showing the limits of the system... showing pitfalls of the system... That brings transparency.”
Process Visibility	Desire for visibility into how AI outputs are generated within workflows	1	S2: “Transparency... making the system transparent so we can dissect it and see all the components.”
Simplification vs. Relevance	Tensions between simplifying explanations and preserving clinical relevance	1	S5: “Many times we calculate uncertainty values, but we don’t connect that to the clinical relevance.”

Theme 6 Clinical Implementation, Workflow Fit, and Human Collaboration

Captures adoption conditions in practice, including workflow integration, UI/UX requirements, clinician engagement, and training.

Code	Description	References	Data Example
Clinical Relevance	Alignment between AI outputs and clinically meaningful decision points	6	S5: “Many times we calculate uncertainty values, but we don't connect that to the clinical relevance. I think that any information that is provided to an MD should be connected to the impact of that decision.”
UX Design	User interface and interaction design considerations for clinicians	1	S2: “you cannot oversimplify, I agree. Oversimplifying is not necessarily using a minimalistic background—you can even oversimplify with a lot of elements. You need to verify which elements are actually needed for that specific user. This requires iterative testing and evaluating what is effective and what isn't.”
Friction	Practical barriers such as workflow disruption, time cost, or cognitive load	3	S1: “I just think it needs to be more careful about how the explanations are designed. Also, having some friction or alerts to the practitioners or the users of the system to understand how to take the explanations that the AI system presents.”
Training and Education	Training clinicians to understand AI behaviour, limitations, and appropriate use	1	S5: “Showing the limits of the system to a user during a training session for example, showing pitfalls of the system (like let's say) this system for MS (Multiple Sclerosis) is very susceptible to patient motion.”
Engagement	Active involvement of clinicians and other stakeholders in system evaluation or use	5	S2: “This participatory process helps in a way because it uses a common ground, a shared space, where people can interact and give their own interpretations of it.”