AI Explainability in Clinical Context

Daniel Gruen, Shruthi Chari, Morgan A. Foreman, Oshani Seneviratne, Rachel L. Richesson, Amar K. Das, Deborah L. McGuiness

Explanations are a core component of collaboration between users and AI technology. Explanations must be designed in context, to address questions and information needs that will arise in actual use.

Our approach for anticipating XAI needs includes:

1. Expert panel to identify usage situations
2. Creation of a rudimentary design mockup
3. Walkthrough with users in specific scenarios including edge cases
4. Design iteration to support explanation needs that emerge

Potential Explanation Types

Case based	Provides information on specific prior cases with features similar to the current one, potentially involving analogical reasoning.
Contextual	Refers to information about items other than the explicit inputs and output, such as information about the user, situation, and broader environment that affected the computation.
Contrastive	Answers the question "Why this output instead of a different output?"
Counterfactual	Indicates what solutions would have been obtained with different inputs.
Everyday	Uses accounts that appeal to users and their general commonsense knowledge
Scientific	References the results of rigorous scientific methods, observations, and measurements.
Simulation	Uses an imitation of a system or process and the results that emerge from similar inputs.
Statistical	Relates to the likelihood of the outcome based on data about the occurrence of events under specified) conditions.
Trace based	Provides information on the underlying sequence of steps used by the system to arrive at a specific result.

Value of Explanations Beyond Establishing Trust

Explanations contribute to the richness of the interaction between the various actors in the overall cognitive system.

- Reveal assumptions, rationales and intermediate results
- Contribute to user's mental model of patient, condition, and system

Illustrative Example: *Diabetes Treatment with New Medication Classes*

① Expert Panel to Identify Use Case and Opportunity for AI Assistance

② Initial Mockup Created

Basic Patient Info (Simulated patient is a somewhat atypical edge case)

Disease and Treatment History

Factors considered

Medication class recommendations

Guideline path

Medication details

Human in the Loop AI
Human users must be able to both request and provide information and explanations to the AI so they can work together to create an effective distributive-cognition system.

③ Explanation needs identified as mockup was "used" for a specific patient

Walkthrough Results
- AI desired to identify situations where guidelines might not apply, with rationales.
- Clinicians invoked rationales or explanations 43 times during walkthrough.
  - Contrastive, contextual, counterfactual used most.
  - Statistical and case based desired for edge cases and non-standard situations.
- "Clinical Pearls" involving rules of thumb or heuristics taught explicitly and shared among practitioners, repeatedly invoked.
- Multiple instances where clinicians wanted to request information or provide knowledge system lacked.

④ Design Iteration to Support Identified Explanation Needs

Exploring Counterfactuals

Learning from User Input

Requesting Contrastive Explanations

Proactively Providing Explanations (Trace, Scientific, Statistical, Clinical Pearls)