ENFIELD

Human-Centric AI - Debate

9 Sept 2025, 09:30 to 17:00 Hours

National University of Science and Technology POLITEHNICA Bucharest

Results of Questionnaires

Interpretation

- Respondents stressed that aligning AI with human values cannot be reduced to technical fixes.
- Human-Centric AI must account for cultural diversity, evolving societal values, and contextual definitions of fairness.
- Ethical alignment requires **participatory approaches** involving diverse stakeholders, beyond the domain of AI specialists. Importantly, sustainability was identified as an ethical dimension, linking human-centricity directly with Green AI.

Strategies

- Enhancing interpretability and explainability.
- Reducing bias and ensuring fairness.
- Designing accountability frameworks and compliance mechanisms.
- Adopting participatory co-design, value-sensitive design, and human-in-the-loop oversight.

Results of Questionnaires

Research Gaps

- Developing contextualized fairness metrics across cultures remains unresolved.
- Quantifying abstract values such as dignity and autonomy in computational systems is a frontier challenge.
- Participatory evaluation methodologies that extend beyond technical validation are limited

Benchmarks Needed

- Explainability, fairness, and ethical performance must be measured dynamically and across contexts, incorporating usability and societal impact in addition to accuracy.

Vision and Roadmap

Near term

- Consolidate and validate novel XAI methods in real-world pilots across healthcare, energy, manufacturing, and space
- Apply multimodal explainable methods to longitudinal biomedical data
- Refine evaluation frameworks for human–AI collaboration, focusing on user reliance, cognitive load, and decision-making quality
- Advance interpretable decision-support tools and neuro-symbolic explainability models that integrate contextual explanations
- Create reproducible algorithms and open benchmarks for explainability in selected domains.

Vision and Roadmap

Long term

- Position Human-Centric AI as the foundation for responsible AI ecosystems where interpretability, inclusiveness, and societal alignment are built in by design.
- Achieve multimodal explanation systems that adapt seamlessly across domains while resisting catastrophic forgetting.
- Establishing frameworks for human–AI collaboration, covering transparency, usability, and accountability.
- Embed bias detection and mitigation frameworks into healthcare, energy, and industrial datasets, ensuring fairness across applications.
- Establish dynamic benchmarks and certification schemes that measure ethical and societal impact alongside technical performance.

Discussions / Debate

ENFIELD

THANK YOU