Requirements for Human-Centered AI

Luís Macedo

University of Coimbra

September 23, 2024

Introduction

- Human-Centered AI (HCAI) focuses on AI systems that prioritize human needs, values, and collaboration.
- AI should amplify human capabilities and align with global ethical standards.
- Key frameworks include the UN SDGs, EU AI Act, and UNESCO AI Ethics guidelines.

Requirement 1: Human-AI Cooperation and Collaboration

- AI should enhance human intelligence, not replace it.
- Emphasis on **Hybrid Intelligence** where AI assists humans in complex decision-making.
- **Collective Intelligence**: AI as part of multi-agent systems with humans.

Requirement 2: Adequate Levels of Automation

- **Adaptive Autonomy**: AI systems should adjust to user needs and task complexity.
- **Human-in-the-Loop (HITL)**: Ensures human control in critical decisions.
- Dynamic adjustment of automation levels based on task demands.
- AI must maintain human oversight to ensure **safety and ethical responsibility**.
- **Human-Centered Control** allows users to override AI when necessary.
- Decision-making processes should be explainable and transparent.

Requirement 3: Ethical Alignment, Accountability, Trustworthiness, and Global Standards

- AI must align with **ethical frameworks** like the United Nations Sustainable Development Goals (SDGs), EU AI Act, and UNESCO recommendations.
- Ensure fairness, transparency, and accountability.
- Promote human rights, privacy, and diversity in AI applications.
- Clear **responsibility** for AI decisions and actions, ensuring developers and users are accountable.
- Ongoing ethical reviews and audits of AI systems.
- Foster trust by ensuring **transparency** in AI operations and decision-making.

Requirement 4: User-Centered Design and Accessibility

- AI systems should be **inclusive and accessible** to all users, regardless of abilities or backgrounds.
- Interfaces must be **intuitive** and adaptable to varying expertise levels.
- Empower users to control and customize AI behavior.

Requirement 5: Adaptability and Continuous Improvement

- AI systems must be **adaptive** based on user feedback.
- Evolve over time to meet changing needs and improve performance.
- Ensure scalability and applicability across different domains.

Conclusion

- Human-Centered AI must prioritize **human well-being**, **ethical alignment**, and **user control**.
- These requirements ensure AI systems are **safe, fair, transparent,** and promote human flourishing.
- Continuous adaptation and alignment with ethical standards are essential for the responsible development of AI.

References

- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D. (2016). *Concrete problems in AI safety*. arXiv preprint arXiv:1606.06565.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv preprint arXiv:1802.07228.
- Floridi, L. (2016). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.
- Friedman, B., Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., Crawford, K. (2018). *Datasheets for datasets*. arXiv preprint arXiv:1803.09010.
- Mitchell, M. (2021). *Artificial Intelligence: A Guide for Thinking*