Modeling and Guiding the Creation of Ethical Human-AI Teams

Christopher Flathmann, Beau G. Schelble, Rui Zhang, & Nathan J. McNeese
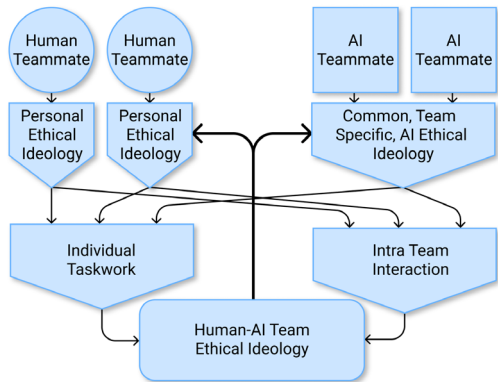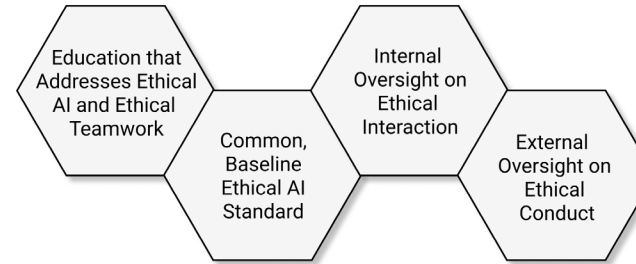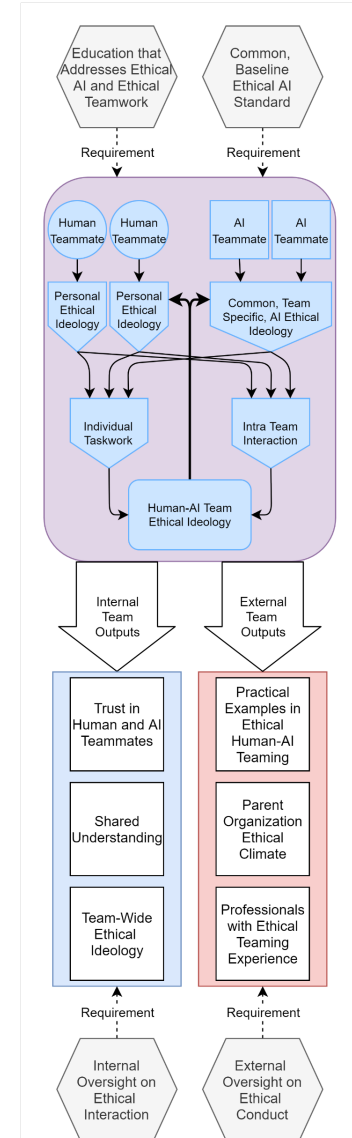Clemson University, United States

TRACE RESEARCH GROUP

Requirements for Ethical Human-AI Teamwork

- Requirements external to a team must be established to ensure a team can both establish an ethical foundational and maintain it through ethical interactions.
- To ensure human-AI teams can be ethical, humans should partake in coursework regarding the ethical use of AI and a common baseline for AI ethics should be established.
- The interactions and outputs of a human-AI team should be evaluated to ensure they comply with internal ethical standards and external ethical regulations[1].
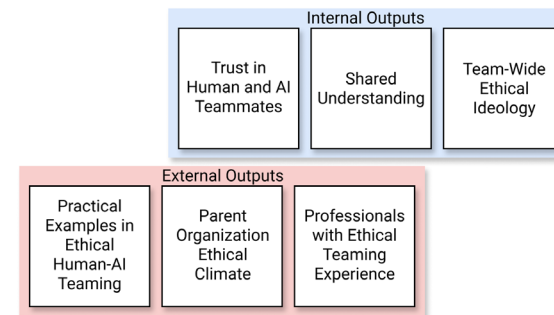

Building an Ethical Human-AI Team through Teamwork Interactions

- Both human and AI teammates are going to possess ethical ideology that influence their work; however[2], AI systems may share a common ideology for standardization.
- Teammates' personal ideologies should interact with each other and be displayed through both principle during normal intra-teaming interactions and through practice during taskwork[3].
- Ultimately, these interactions will form a common ethical ideology that is representative of a team's ethical ideology[4].
- Through continued participation in a team, the team-wide ethical ideology will influence the individual ethical ideologies of the teammates, which means AI should be able to adapt to the ideology of the team they have been placed on.

The Outputs of Ethical Human-AI Teams

- Internal teaming outputs, including mutual trust, shared understanding, and the shared ethical ideology, will help maintain team ethicality over time.
- External teaming outputs, including practical examples of human-AI teamwork and contributions to overall ethical climate, will help improve ethical human-AI teaming in society.
- These outputs will be produced regardless of the ethicality of a human-AI team; however, their quality will be reliant on the ethical quality of the team that produces them.

References
[1]Torresen, J. (2018). A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI*, *4*, 75.
[2]Lau, W. W. F., & Yuen, A. H. K. (2014). Internet ethics of adolescents: Understanding demographic differences. *Computers & Education*, *72*, 378–385.
[3]Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507.
[4]VanMeter, R. A., Grisaffe, D. B., Chonko, L. B., & Roberts, J. A. (2013). Generation Y's Ethical Ideology and Its Potential Workplace Implications. *Journal of Business Ethics*, *117*(1), 93–109.