ACAI 2021
The advanced course on AI on Human Centered AI

Tutorial Outline

Children and the Planet - The Ethics and Metrics of "Successful" AI
John Havens, IEEE; Gabrielle Aruta, Filo Sofi Arts

Introduction

The vision and research agenda for ACAI is, focused "around ethics values and trust (Responsible AI)… intimately interwoven with the impact of AI on society" yet what often lacks clarity in the technological design of AI are the metrics of success for its design, usage and legislation. Where metrics of success for Artificial Intelligence depend on the economic underpinnings of how specific systems are built, financial sales or growth of a company building a system are the stated or implied benchmarks of said success.

Where organizations intend their systems to assist "the common good" or honor principles like beneficence, or are framed as "AI for Good," while potentially helpful as directional frameworks to establish a baseline for Responsible AI, a crucial aspect related to socio-technical design is to ask questions such as "What is the definition of 'good' being used in the design of a system?" and "Whose 'good' are we referring to when these systems are designed but also deployed? Are all stakeholders created equal?"

It is this last question we will explore in, "Children and the Planet – The Ethics and Metrics of "Successful" AI and work to illuminate the following insights:

- Where existing societal metrics of success *(such as GDP or indicators prioritizing exponential growth)* do not include any measure of children *(human wellbeing)* or the planet *(environmental sustainability)*, then the future of humanity is at great risk.

- Where AI is designed using only *(or primarily)* metrics such as GDP, then systems will not inherently measure or value children or the planet, putting the future of humanity at great risk.

- Where "risk" is defined largely or solely as physical risk for current AI design and regulation, or where "risk" is not defined to explicitly include metrics prioritizing children and the planet, AI systems will continue to be designed in ways that may be harming human wellbeing and environmental sustainability, putting the future of humanity at great risk.

- **Unless we define "Responsible AI" (and "Responsible Innovation") to include adequately measured and prioritized metrics of success including children and the planet in all stages of AI system design, the future of humanity, right now today, at great risk.**

*Tutorial Outline:

As this tutorial is designed to be a proactive workshop, "exercises" below note a portion of the program where attendees will be asked to participate on an activity provided by Gabrielle and John.

- **Introduction:** Understanding Individual, Societal and Environmental Metrics of Success for AI

    o (John) <u>A Focus on Flourishing</u> – Why Prioritizing ethically aligned design and "Beyond GDP" Metrics must form the basis for Responsible AI

    o (Gabrielle) <u>The Action of Aesthetics</u> – How applied ethics must form the basis for education and Responsible AI

- ***Exercise One: Evaluating Values***

    o <u>*Goal*</u>*: Help attendees understand the logic of the statement, "How will machines / AI know what we value if we don't know ourselves?" to provide a way of understanding Human-AI interaction (Pillar Three[1])*

    o <u>*Process*</u>*:*
        - *Attendees find a partner*
        - *Partner A asks Partner B, "What are one of your key life values and how do you know?"*
        - *Partner B responds.*
        - *Process repeats with attendees both responding.*

- *(John)* **Ethically Aligned Design**: An Introduction to IEEE's work prioritizing human in the loop, values-driven design at the outset of AI System creation *(Pillar Five).*
    o Ethically Aligned Design: It's history, formation, content and principles
    o IEEE 7000-2020: The standard connecting values-based engineering to traditional systems design
    o IEEE 7010-2020: The standard providing a basis for the identification and application of Beyond GDP / Wellbeing metrics regarding AI Systems
    o IEEE 2089

- *(Gabrielle)* Gabrielle's Section featuring John Dewey's work with a focus on:
    o How applied ethics / aesthetics are active *(Pillar Two).*
    o Art as Experience as inspiration
    o AI analogy – logic is that algorithms and other AI systems affect us viscerally (they access our data, motivate or manipulate our actions) so we should honor our human ability to also actively identify and live to our values.
    o One goal of this section is to segue to John's next section on non-Western / Ubuntu ethics as a way to discuss global wellbeing metrics and the environment.

- ***Exercise Two: Art as Experience (Slow Looking)***
    o *Goal: Model behavior to evolve from subjective opinions to objective analysis while identifying broad human values in the work of art. This helps people arrive t a consensus driven, pragrmatic truth. This models relational thinking and values.*

---

[1]*Pillars refer to Research Roadmap for European Human-Centered AI*

- o *Process: Gabrielle will facilitate a community of inquiry with the audience using works of art as stimuli. We'll split the audience into two groups for this process.*
  - o *Goal is for attendees to realize the active, visceral sense of relationality between a work of art and themselves as well as others.*

- (John and Gabrielle) **KindNext – Relationality and Responsible AI**
  - o *An introduction to non-Western Ethical Traditions*
  - o An introduction to Ubuntu Ethics
  - o An introduction to Virginia Dignum's "Social AI"
  - o An exploration of Social AI and Wellbeing Metrics
  - o Why "Relationality" should be a part of Responsible AI

- (John and Gabrielle) Conclusion: **KindNext – The Ethics and Metrics of Responsible AI (RAI)**
  - o (Gabrielle) Teaching values is a key metric for caring for children / RAI
  - o (John) Recognizing the planet as a set of living systems is a key metric for RAI
  - o (John) Rationality and Relationality together represent humanity in full
  - o (Gabrielle) (Point about action and aesthetics)
  - o (John) KindNext – These things must be prioritized before AR/VR/XR ("next")

- *Exercise Three: A Case for Kindness*

  - o *(Goal): End the workshop with a proactive kindness exercise*

  - o *(Process):*
    - *Attendees are asked to pair up with a partner*
    - *Partner A asks Partner B, "What's one thing that stood out to you about this workshop?"*
    - *Partner B responds.*
    - *Partner A then reflects on Partner B's answer before providing a kind response in reaction to Partner A, eg, "What an insightful point – I hadn't thought about that."*
    - *Process repeats with both partners.*
    - *Before workshop begins, Gabrielle and John encourage all attendees to continue to give kindnesses throughout the day using the phrase, "Can I give you kindness?"*

Articles and Resources to be utilized in the Tutorial:

- How Dewey's View on Aesthetics is Relevant to Philosophical Counseling: Gabrielle Aruta and Morten Fastvold; Journal of the APPA, Volume 7, Number 2, July, 2012.

- **From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance**. Sabelo Mhlambi, Carr Center for Human Rights Policy, Harvard Kennedy School, Spring, 2020.

- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE, 2019. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/ autonomous-systems.html.

- **Responsible Artificial Intelligence: recommendations and lessons learned**. Virginia Dignum, Department of Computing Science; Umeå University, July 30, 2021