ARTIFICIAL INTELLIGENCE & EMERGING TECHNOLOGIES: RECOMMENDATIONS
Gesda Summit - GSCP Workshop| Oct 12 2023
Alexis Wichowski | alexis.wichowksi@columbia.edu

Recommendations
"Optimizing for peace"
Talking points: workshop

[Recommendations]

[1] Decision makers / policy makers need to learn how to think like technologists
By understanding and adopting algorithmic thinking will ensure they're always asking:

- What outcomes does a technology optimize for?

- How do we incentivize optimization for outcomes aligned with peace-building behaviors?

[2] Bring major technology developers (not just Big Tech) into dialogue with nation-states and multilateral institutions

[3] Ensure new technologies are programmed with built-in "moral compass": optimization rules to reduce harm to human users via self-supervision capabilities

[4] Educate about technology fundamentals at all levels

- For youth
  - STEM basics: ongoing education for understanding various technologies
  - Critical thinking skills: for evaluating appropriate technology use
  - Ethics, philosophy, humanities: for responsible and ethical technology use

- For technologists
  - How political and diplomatic leaders approach problem-solving
    1. Competing priorities
    2. Resource constraints
    3. Public & press scrutiny

- For political and diplomatic leaders
  - How technologists approach problems: optimization
  - How to incentivize different optimizations: funding; data access

["Optimizing for peace"]

[1] OPTIMIZATION

Optimization of tech has historically prioritized more (vs quality) engagement: Technologists create products using algorithms that nudge users to spend as much as time possible using their product. It's possible to create tech products that nudge users toward less frequent but more profound engagement but requires specifically optimizing for those outcomes. Specifically, technologists build for:

- Operational efficiency: tech that always works and works fast

- Outcomes that support their business model: in social media era of all-free access (vs tiered access), platforms have optimized for maximal exposure to advertisers. This business model requires optimizing for longer user engagement (keeping users actively clicking through / posting on a platform as long as possible) over quality of engagement experience (eg fewer but more profound interactions vs widespread, superficial interactions).

Optimization of tech for peace-building behaviors is possible but won't happen on its own – requires intentional "optimizing for peace." Examples of this kind of programming includes:

- Birdwatch – former Twitter pilot program running bridge-based ranking algorithms that identify and prioritize displays of content that shows consensus across different user communities (eg conservative- / liberal-learning journalists; activitsts; politicians; citizens etc)

- Claude.AI – generative chat AI built to prioritize safety and with explicit goal of benefiting nonprofits, businesses, and civil society

[2] INCENTIVES

Existing incentives = building risky tech: Current tech funders (eg VCs) prioritize bold, risky, game-changing new technologies; it's much more difficult to get significant investments for safe, cautious, incremental changes to existing technologies

New incentives are required to create tech optimized for peace-building behaviors: To build tech that optimized for consensus and other peace-building outcomes requires identifying new ways to incentivize technologists. For instance, directly funding such technologies; granting access to existing data; inviting to collaborate on data collection

[TPs] How technologists see the world (Workshop_20231012)

Optimization & incentives

- An ordered world, a rules-based order, to a technologist, can also be thought of as parameters or variables; what systems and structures and features of those systems and structures are required to achieve optimal outcomes

- A "rules-based order" to technologists may be more aptly described as "optimizing for peace"

- This is a key concept. It is a foundational concept to understanding how artificial intelligence and other emerging technologies have been constructed thus far and how they will continue to be constructed

- Because all technologies are built to optimize for one thing or another

- For instance, EFFICIENCY: how to deliver more, faster. In the case of AI, optimization for efficiency means prioritizing more content delivered faster

- To ensure our current and future technologies optimize for peace, we must understand how to think like technologists and how to not only optimize for peace but how to *INCENTIVE* optimization for peace

- I raise this question of *incentivizing* optimization for peace, because this is reality of how technology is built

- And by incentivizing this primarily means FUNDING

- Responsible, safe, cautious technology isn't getting massive amounts of VC funding

- Traditionally VCs have overwhelmingly funded and continue to overwhelming fund RISKY, BOLD technologies that have the potential to be unicorns; massive money makers; new verbs, like "Let me Google that for you"

- When we think about the future of peace and war and how we ensure advanced technologies can be developed in support of peace, we have think like technologists think:

    - first, as a question of what we're optimizing for

    - and second, as how we're going to get funding; how we're going to be sustainable – profitable?

- For these are the constraints under which technologies thrive or die – are they optimizing for outcomes that are also economically sustainable and profitable?

- Thus, “rules-based order” in 5 10 25 years may represent dual perspectives:

    - How geopolitical powers advance peace through established multilateral systems

    as well as

    - How technologists advance peace through technologies built to optimize peaceful outcomes, with non-extractive business models that are also economically sustainable and profitable

Future-proofing new tech with built-in “moral compass”

- Reflecting on earlier conversations in the workshop about the need for technologies to have some sort of internal moral compass

- I’ll share a recommendation -- not from 2023 but from 1950. From legendary science fiction writer Isaac Asimonv and his novel about robots and artificial intelligence, called “I Robot.”

- In it, he described the 3 Laws of Robotics:

    1. “A robot may not injure a human being or through inaction allow a human being to come to harm”

    2. “A robot must obey orders given it by human beings, except where such orders conflict with the first law”

    3. “A robot must protect its own existence as long as such protection does not conflict with the first law”

- Which all point back to reality being truly only law: technology must not injure humans or through inaction allow humans to come to harm