

Artificial Intelligence must Responsibly Innovate to Prevent Real-world Problems

Katie Greisiger Frost

College of Global Futures, Arizona State University

PIT 501: Principles of Public Interest Technology

Dr. Bénédicte Callan

February 15, 2021

Artificial Intelligence must Responsibly Innovate to Prevent Real-world Problems

In the mid-20th century, the definition of Artificial Intelligence (AI) would have been drenched in science fiction fanfare, with the simple notion that AI was merely machines *acting* as human intelligence (Helm et al., 2020). Fast forward to modern times and decades of tech innovation with computing, AI has proven itself to be far more complex and versatile than originally imagined. According to Buchanan (2005), AI is “about understanding the nature of intelligent thought and action using computers as experimental devices.” On a simultaneous, upward trajectory with the advancements in computational languages, computer hardware and memory enhancements, the innovations of AI and machine learning technologies have increased exponentially: better and faster machinery in essence *should* equate to solving complex problems in faster and more efficient ways (Helm et al., 2020).

The speed at which these innovations are reaching the public sphere are also moving quickly, so swiftly in fact that many of us are not even aware of AI technologies working in the background of many of our electronic devices, nor are we aware of how the technologies have influenced many facets of society (Helm et al., 2020). Additionally, the devices and applications that we access could be tracking and analyzing *loads* of our data – whether metadata, imagery, online posts, etc. – to the effect that researchers in many fields rely on the access to our data for scientific study, product research, marketing, or even behavioral interventions. As the reliance on big data – *typically a data set of one terabyte or larger* – increases and spreads its reach into social science realms, researchers and engineers handling such data are facing new challenges “outside of their comfort zones” (Zook et al., 2017).

AI isn’t the only technological advancement increasing our fears of potential ethical, societal, and/or legal implications (ELSI). For many researchers in the early 21st century, inside and outside of academia, it was clear that technology at-large was going to have a huge impact on the way people lived. With its arms reaching many spans of industries and day-to-day human life, interweaving ethics into tech

innovation was an opportunity that early engineers of responsible development could not ignore. Thus, international efforts, such as through the Dutch Research Council, have catapulted the idea of Responsible Innovation through collaborative efforts to define, and conceptualize best practices (Grunwald, 2014). Developers and researchers are now being held to new standards, that of which demands “explicit ethical reflection” and democratic processes (Grunwald, 2014). Responsible development:

“can be characterized as the balancing of efforts to maximize the technology’s positive contributions and minimize its negative consequences. Thus, responsible development involves an examination both of applications and of potential implications. It implies a commitment to develop and use technology to help meet the most pressing human and societal needs, while making every reasonable effort to anticipate and mitigate adverse implications or unintended consequences” (Grunwald, 2014).

Grunwald further states that Responsible Innovation “as an active process makes clear that assignments and attributions of responsibility take place in concrete social and political spaces involving and affecting concrete actors in concrete constellations,” and that “Responsible Innovation unavoidably requires a more intense inter- and trans-disciplinary cooperation between engineering, social sciences, and applied ethics” (2014). Therefore, Responsible Innovation defines the: *someone* – the person or group that is held responsible; *something* – the “results or of actions or decisions” formulated by the someone; *body of rules* – the “normative framework” that the someone agrees to; and the *quality of available knowledge* – agreed upon knowledge regarding consequences and results of actions or in-actions (Grunwald, 2014).

It’s no longer enough to rely on familiar tools of Technical Assessment (TA) *if* the movement beyond experts and into public realms does not occur early in the innovation process (van Lente, et al.,

2017). Assessing a technology upstream – meaning engaging with various publics early in the innovation process – for its potential negative effects is a crucial component to ensuring Responsible Innovation (Grunwald, 2014). This type of reflective process also requires a collaboration of stakeholders from different fields. As ethical ramifications are not inherently built into the Scientific Method and processes, experts from fields entrenched with moral and societal compasses are needed to intervene in the innovation process (Callan, February 10, 2021) – to step in with reflective questions, reasoning, and knowledge of what public engagement could be needed in an effort to prevent harm to public systems. Stilgoe, et al., briefly defines this asset of Responsible Innovation as, “...taking care of the future through collective stewardship of science and innovation in the present” (Callan, January 14, 2021).

Efforts from the research community aren’t the only ones providing models or standards for responsible and ethical innovation. Many technical industries rely on membership, training, and guidance from professional associations. The AI field leans heavily on its decades-old organization, the Association for the Advancement of Artificial Intelligence (AAAI), to define a scope of conduct which to adhere by and to shed light on leaders that are paving the way for responsible AI technology development:

“The AAAI Code of Ethics and Professional Conduct is designed to inspire and guide the ethical conduct of all AI professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses AI technology in an impactful way, and is particularly intended to act as a standard of ethical and professional conduct for all AAAI members. Additionally, the Code serves as a basis for remediation when violations occur” (n.d.).

The AAAI makes a clear statement of what is expected of its industry’s members and that there are ramifications if one does not adhere to their ethical standards. It’s not always a given that a professional organization will interweave ethics and responsibility into its expectations, and without the notion of responsibility to interrupt traditional TA, often ELSIs are left untethered. Additionally, it’s not surprising

that new technologies “often fall outside established regulatory structures...” and that “Government may be slow to apply existing regulations or may need to develop new legal frameworks” (Callan, January 14, 2021).

When further illustrating how AI technology innovates with the use of big data and *its* potential implications on society, one must also review its reliance on algorithms, and computing. The medical AI company ArtificialIntelligence.health gives a great explanation of how AI technology works:

“Algorithms are generally understood to produce very specific results based upon input. Typical computer programs do not demonstrate the deductive reasoning, pattern recognition, anticipation, and decision-making that are hallmarks of human intelligence in action. Artificial Intelligence programming allows for this higher level information processing through various systems, allowing intellectual work that typically would require full human control to be automated by computers. Not only does this save time and labor, but it also offers the opportunity to discover patterns in expansive data sets that might otherwise go unnoticed by the human eye”
(n.d.).

The innovation of AI itself is a feedback loop – the results of an algorithm’s computing relay information back to modify its initial algorithm coding and variables to produce finer and more accurate results over time. Like statistical analysis, there is a positive relationship between the amount of data utilized in the innovation of the AI, and the possibility for greater predictability and ingenuity of an expected outcome. However, as Zook et al. illustrates, “Data are people” (2017), and as such, AI technologies that utilize big data should have greater incentive to act responsibly for the protection and wellbeing of the people that formulate said data sets in the first place.

In essence, AI technologists are trying to avoid harm caused by such human-centered technologies as Clearview AI’s facial recognition software. Clearview alleges that the novel technology is

“used by law enforcement agencies to identify perpetrators and victims of crimes” (n.d.). More than 600 law-enforcement agencies use this product in the United States and Canada, and yet it was proven to enforce racial biases towards the improper accusation of crime and/or arrest of African American males, and was sued by the American Civil Liberties Union of Illinois in 2020 for breaches of privacy when the company was found to have scraped social media images and meta-data without proper consent (ACLU v. Clearview AI, 2020). This technology, initially funded by the private sector as a start-up, is in the market stage of development and is having to respond to allegations of harm and value missteps in order to regain integrity, trust of use, and prevent further harm – one could argue having to work backwards to introduce responsible innovation in its processes.

Another ideal example of data equating to the human form is the use of AI to collect data from public video capture to identify, prevent, and/or intervene with potentially dangerous societal events, such as a school shooting. The Center for Homeland Defense and Security (CHDS) defines a school shooting as:

“an incident in which there is an active shooter on school property, using the FBI’s definition of active Shooter,” which is an “individual (or individuals) actively engaged in killing or attempting to kill people in a populated area, no matter the number of victims” (Melgar, 2020).

Although school shootings are still rare compared to daily gun violence experienced in the United States, the average days between school shootings has decreased since the infamous Columbine shooting in 1999, and its presence in the media has continued to increase (Melgar, 2020).

“The idea is that algorithms might be able to better analyze data related to school shootings, and perhaps even identify patterns in student language or behavior that could foreshadow school violence” (Rieland, 2018). One such start up that is focusing on preventing the violence incurred through school shootings, is Sitch AI out of California. According to its website, Sitch AI (n.d.) states that its two main

goals are to “find current changes in human behavior,” and to “predict future behavior with AI and advanced forecasting techniques” (n.d.). The changes in behavior are monitored by data-points that the company has defined as a reason for concern – it gathers information about a student’s behavior, such as falling grades, absences, and disciplinary action, and provides real-time alerts of a student’s “down trend” (Sitch school predictive solutions, n.d.). Sitch (n.d.) details that the predictive analytics use a technology known as Long-Short Term Memory (LSTM), an algorithm technique that allows computing to identify trends in behaviors to predict *future* behaviors.

Sitch AI further explains that “LSTM’s are used in advanced cases such as voice recognition, language processing, financial and economical analysis” (n.d.). Sitch’s highlighted application, the *Mobile Control Panel*, is an interface that allows users to track visual captures of the real world to identify potential harm. The panel is an encapsulation of visual readings from computer vision technologies that read camera imagery as data in an effort to track potential threats and/or weaponry on a person or on a campus, and monitors the assets that are capturing such data to make asset management more centralized and efficient (Sitch school predictive solutions, n.d.).

Sitch hopes that their technologies will better equip schools and law enforcement to be alerted if a student needs help, and to intervene before it’s too late. CEO Matty Squarzoni has said “we are not looking at being able to say, ‘This person’s going to be a shooter.’ We want to be able to say, ‘This person needs help’” (Rieland, 2018). Indeed, scientists have found that AI can be used to successfully identify the risk of a child acting aggressively or violently. According to a study at Cincinnati Children’s Hospital Medical Center, the AI/ML technology used to assess the risk of violent behavior was “just as successful as a team of child and adolescent psychiatrists” (Rieland, 2018).

AI researchers looking to utilize children’s data, even with the best of intentions to prevent violence towards them, raises many concerns. School shooting predictive analytics trudge up such ELSIs

as incorrect predictive policing and human biases, invasion of privacy, and/or furthering disparities that exist in school disciplinary action (Rieland, 2018). As a private entity not held to as high of standards as professional mental health practitioners, for example, how would such an AI company producing applications reliant on children's data incentivize itself to act ethically and morally without the use of Responsible Innovation?

Compounding this argument is the fact that AI technology is increasingly a part of our day-to-day lives and requires little incentive to innovate responsibly. The technology research giant, Gartner, stated that by 2020, "AI technologies will be virtually pervasive in almost every new software product and service" (Moore, 2017). Leading up to 2020, "84% of C-suite executives believe[d] they must leverage artificial intelligence (AI) to achieve their growth objectives, yet 76% report[ed] they struggle with how to scale" (Reilly, 2019), and the Trump administration signed an executive order creating the American AI Initiative – further encouraging a priority to invest in AI technologies in the United States (Vincent, 2019). With a recent increase in federal funding, federal monies' recipients will have requirements to act in responsible and ethical ways, fortunately as a stand in until lawmakers catch up to the expanse.

For the private sector, however, a commitment to innovating responsibly must be established to prevent unnecessary harms on public systems and individuals. AI technology can only go so far as to assess and prevent risk, as the team who built it is aware and reflective regarding said risks. Per another Gartner study, the current demand for AI talent comes from non-IT professionals, further illustrating the need for inter- and trans-disciplinary cooperation that Responsible Innovation would provide, it states:

"that although the IT department's need for AI talent has tripled between 2015 and 2019, the number of AI jobs posted by IT is still less than half of that stemming from other business units...Departments recruiting AI talent in high volumes include marketing, sales, customer service, finance, and research and development. These business units are using AI talent for

customer churn modeling, customer profitability analysis, customer segmentation, cross-sell and upsell recommendations, demand planning, and risk management” (Rimol Gartner, M., & Goasduff Gartner, L., 2020).

It’s evident that even though AI is a technology comprised of machinery and data, the need for morality and ethics of its human builders and their partnerships with public actors, stakeholders, and policy-makers, is increasingly important in order to innovate responsibly. To consider the public interest, actors collaborating to build AI technologies must utilize Responsible Innovation approaches. We can utilize the Sitch AI tool to prevent school shootings as an example of how Responsible Innovation may be utilized. Sitch could partner with an organization like the AAAI to engage with the different forms of the public as defined by Braun & Schultz (2010), using different engagement models theorized by Macnaghten & Chilvers (2014). Suggested engagement, activities, and potential outcomes are outlined in figure 1, and could help definition the necessary components that Grunwald established for Responsible Innovation: the someone, something, body of rules, and quality of available knowledge (2014).

Type of Public	Model of Engagement	Activities	Potential Outcomes
The Affected Public	Issue Advocate Model	Virtual conferences/focus groups – engage with a population of public that has been affected by the AI technology (both positively and negatively)	Receive feedback for how the technology may cause harm, ideas for how to improve said technology, and data points for reviewing what is working well – <i>information collected that is contradictory could be further reviewed by the partisan public</i>
The Partisan Public	The Honest Broker Model (professionals) & Issue Advocate Model (panel)	Virtual conferences & long-term membership – engage professionals in the related AI field Panel – engage other like-minded stakeholders through a panel/working group that is committed to being held accountable	Engagement with professionals and organizations innovating AI can give opportunities for teaching and interventions – i.e., learning events, conferences, and professional society membership; engagement through a panel of experts can allow a centralized management system that can collect, review, and publish information gathered through engagement activities; the panel could also act as policy maker to influence law-maker buy-in and action The panel may establish boundaries for repercussions when those held accountable do not meet expectations
The Pure Public	The Upstream Model	Exploratory/focus groups	Feedback from un-affected populations could give an extra data-point for anticipatory consideration
The General Public	The Upstream Model	Poll	Another anticipatory data point for consideration could be learning what the at-large/unaffected public understands about the way the algorithm works and what its purpose is

Figure 1: Activities in the public domain

Incentives for Sitch and AAAI to create an accountable body that will be responsible to insemminate the above actions into their innovation model are bountiful, such as but not limited to: recognition as a leader in responsible and ethical innovation in AI, potential to receive and increase federal funding, support the definition and creation of new laws to protect humans that are part of large data sets used in AI technology, and to increase the research and training models utilized for responsible innovation.

Even though these steps may be established, a lot of industries still “learn by doing” (Callan, January 23, 2021), thus leaving a link in the chain of iterative processes. Sitch, and related technologies, must remain adaptable and flexible when responding to incidents that they could not predict. As such, the innovation process can sometimes be more fluid-like and radical than incremental (Callan, January 23, 2021), leaving it difficult to provide models for Responsible Innovation in practice. If the team held accountable for this innovation remains intact and listens to customers and stakeholders’ experiences, however, quick responses to harmful actions could in theory be set into motion to prevent further harm to the public’s trust and/or individuals. Further study on the best practices for *how* responsible, and therefore accountable, parties respond to harmful implications is needed. Additionally, policy makers must create laws that incentivize the private sector to act responsibly, and to be held accountable for their follow through.

In addition to not being able to predict all potential risks, and due to the very nature that it innovates, some growth to AI technology may need to be endured in real-time. As discussed earlier, without incentives from policy makers nor private funders for ethical development, AI actors must invoke their own incentives for keeping the public interest in mind. This presents another potential, grave challenge – most incentives for new private business are centered around economical gains (Callan, 2021, February 10). Responsible Innovation in its purest sense offers a model for which morally minded

developers can garner – they must act in collaboration with different facets of the public, commit to interdisciplinary cooperation and shared responsibility, be open to shifting priorities as they receive and respond to feedback, and create professional working groups that will keep the stakeholders engaged and held accountable. Without Responsible Innovation, AI technologies will be left free to cause harm, in nearly any aspect of our daily lives.

References

- ACLU v. Clearview AI. (n.d.). Retrieved February 15, 2021, from <https://www.aclu.org/cases/aclu-v-clearview-ai>
- Braun, K., & Schultz, S. ... a certain amount of engineering involved: Constructing the public in participatory governance arrangements. *Public Understanding of Science* 2010 19: 403.
- Buchanan, B. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine*. 26. 53-60.
- Callan, B. (2021, January 14). *Technology Choices & Values* [Lecture recording]. Canvas.
<https://asu.instructure.com/courses/76364/pages/week-1-learning-material>
- Callan, B. (2021, January 23). *National Systems of Innovation* [Lecture recording]. Canvas.
<https://asu.instructure.com/courses/76364/pages/week-2-learning-materials>
- Callan, B. (2021, February 10). *Principle of Reflexivity* [Lecture recording]. Canvas.
<https://asu.instructure.com/courses/76364/pages/week-5-learning-materials>
- Clearview. (n.d.). Retrieved February 15, 2021, from <https://clearview.ai/>
- Grunwald, A. (2014). Chapter 1: Responsible Innovation: A new Look at Technology and Ethics. In Jeroen van den Hoven, Neelke Doorn, Tsjalling Swierstra, Bert-Jaap Koops, Henny Romijn Editors, *Responsible Innovation: Innovative Solutions for Global Issues*, Springer, pp. 3-13.
- Grunwald, A. (2014). Chapter 2: Technology Assessment for Responsible Innovation. In Jeroen van den Hoven, Neelke Doorn, Tsjalling Swierstra, Bert-Jaap Koops, Henny Romijn Editors, *Responsible Innovation: Innovative Solutions for Global Issues*, Springer, pp. 15-32.

Helm, J., Swiergosz, A., Haeberle, H., Karnuta, J., Schaffer, J., Krebs, V., Spitzer, A., & Ramkumar, P. (2020).

Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions.

Current Reviews in Musculoskeletal Medicine, 13(1), 69–76.

How does artificial intelligence work? (n.d.). Retrieved February 12, 2021, from

<https://artificialintelligence.health/describe-artificial-intelligence.html>

Macnaghten, P. & Chilvers, J. (2014). The future of science governance: publics, policies, practices.

Environment and Planning C: 32: 530–548

Melgar, L. (2020, September 17). Are school shootings becoming more Frequent? We ran the numbers.

Retrieved February 12, 2021, from <https://www.chds.us/ssdb/are-school-shootings-becoming-more-frequent-we-ran-the-numbers/>

Moore, S. (2017, July 18). Gartner says AI technologies will be in almost every new software product by

2020. Retrieved February 13, 2021, from <https://www.gartner.com/en/newsroom/press-releases/2017-07-18-gartner-says-ai-technologies-will-be-in-almost-every-new-software-product-by-2020>

Reilly, A. (2019, November 14). Scaling ai: From experimental to exponential. Retrieved February 15,

2021, from <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments>

Rieland, R. (2018, June 22). Can artificial intelligence help stop school shootings? Retrieved February 12,

2021, from <https://www.smithsonianmag.com/innovation/can-artificial-intelligence-help-stop-school-shootings-180969288/>

Rimol Gartner, M., & Goasduff Gartner, L. (2020, March 18). Gartner says strongest demand for Ai talent

comes from Non-IT Departments. Retrieved February 15, 2021, from

<https://www.gartner.com/en/newsroom/press-releases/03-18-2020-gartner-says-strongest-demand-for-ai-talent-comes-from-non-it-departments>

Sitch school predictive solutions. (n.d.). Retrieved February 13, 2021, from <https://www.sitchai.com/sitch-school-predictive-solutions/>

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.

van Lente, H., Swierstra, T. & Joly, P-B. (2017). Responsible innovation as a critique of technology assessment. *Journal of Responsible Innovation*, 4:2, 254-261.

Vincent, J. (2019, February 11). Trump signs executive order to spur US investment in artificial intelligence. Retrieved February 15, 2021, from <https://www.theverge.com/2019/2/11/18219981/american-ai-initiative-trump-administration-funding-research-data>

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S., Goodman, A., Hollander, R., Koenig, B., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology*, 13(3), e1005399–.