ICRES 2018: International Conference on Robot
Ethics and Standards, Troy, NY, 20-21 August 2018.
https://doi.org/10.13180/icres.2018.20-21.08.008

SIMILARITIES IN RECENT WORKS ON SAFE AND SECURE BIOLOGY AND AI RESEARCH

Pedro Henrique Oliveira dos Santos

Instituto de Informática, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500
Porto Alegre, Rio Grande do Sul 91501-970, Brazil
E-mail: pedrosans@gmail.com
http://www.inf.ufrgs.br

Dante Augusto Couto Barone

Instituto de Informática, Universidade Federal do Rio Grande do Sul
E-mail: barone@inf.ufrgs.br
http://www.inf.ufrgs.br

Given the growth rate of artificial intelligence capabilities, it's natural to pay attention to the risk involved in such unprecedented technology growth, how it affects society today and potential security threats. Such concern is not new and has been the object of scrutiny over the years due to the same threats posed by advances in biology. This article resumes recent and relevant works on both areas and lists their similarities.

Keywords: Biorisk; Artificial Intelligence; Ethic.

1. Recent Works on AI

Technologies made possible by advances in AI can greatly benefit society like by aiding disease diagnosing while giving neuroscientists a better understanding of the brain.[1] As disease diagnosing is thought as a specialized and intelligent task, the artificial mean to perform it, aided by the use of technology, is referred to as AI.[2] Because new technologies enabled by AI research is also creating new risks, like a fail in the self-driving ending in a fatality, a number of institutions are raising concern. To put in perspective, an accident involving a vehicle with autopilot hardware in March 2018[3] is being investigated for failures in the car software. To address new risks made possible by the developments in AI, a number of institutions are coming together to raise questions and create guidelines for AI systems. In February 2017 the Future of Humanity Institute reported on the current AI capabilities, how it can be used with malicious intent, and how this risk can be addressed; it produced the document The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.[2] On April 2017 UK's House of Lords Select Committee on Artificial Intelligence, aiming to lead the international community, proposed five principles for upcoming ethical AI frameworks.[4] On December 2017 IEEE published the second version of the document Ethically Aligned Design which is part of its global initiative on ethics of autonomous and intelligent systems to identify and find consensus on issues of transparency, accountability, and algorithmic bias in implementations of AI systems.[5] In this scenario, US government officials and researchers came together to better understand the possibilities of mutual support that will need to be put in place in the case of a cyber attack and reported their work in the Cyber Mutual Assistance Workshop Report.[6] As a result of the mentioned works some threats were identified, including:

- Repurpose of existing AI systems by terrorists like drones or autonomous vehicles programed deliver explosives and cause crashes

- Automating influence campaigns leveraged by IA analysis of social networks
- Denial-of-information attack leveraged by AI enabled bots capable of publishing false or distracting information next to real information
- Fake news reports with realistic fabricated video and audio
- Many jobs can disapear
- Unclear liability when AI systems malfunction or cause harm to users

Given the new risks and capabilities being made possible by AI, ethical concerns were raised in the mentioned works, including:

- Ethical design should be an integral part of the curriculum of AI system developers and users
- Well being should be prioritized in AI systems
- If an accident involving an autonomous car occur, the AS will need to be transparent to an accident investigator
- Transparency is important to understand what AI systems are doing and why
- Record-keeping of intended use and system parameters to enable investigators to find out the legally responsible for a particular AI system
- Tailored defenses for an attack will need financial backing

2. Recent Works on Biorisk

In September 2017 the Future of Life Institute published three paper to assess global catastrophic and existential biosecurity risk. In the papers, it's notable the concern with financial aspects, the benefits of threat-mitigation efforts and questions on ethics including:

- Should a dual-use research be funded?
- What should be the price for expected risks in a dual-use research?
- Should an estimated price be included as a cost in research grant proposal?

A central point in the ethical discussion on biology is the gain-of-function experiments since a subset of such experiments can be used by malicious actors.[7] To ethically balance the need for better public health by better understanding viruses, and to protect the same public from the risks associated with this research proved a point of controversy. Selgelid[8] developed/proposed in 2016 a framework for gain-of-function research decision making supported by a set of principles - including manageability of risks, justice, engagement - designed to indicate ethically acceptable or ethically problematic ou unacceptable researches. A similar initiative came from the Obama government that tasked, in 2014, the US National Science Advisory Board for Biosecurity (NSABB) to recommend on the deliberative process regarding risks associated with gain-of-function researches. In 2016 NSABB published its recommendations for evaluating gain-of-function researches supporting its values[9] with the Belmont Report,[10] the literature on public health ethics[11],[12] and the ethics analysis by Dr. Michael Selgelid.[8] The recommended ethical values include:

- Non-maleficence: research should consider and apply approaches preventing harm and mitigating potential risks.
- Beneficence: the research should have a beneficial outcome for public health
- Social justice: benefits and risks should be fairly distributed, even on a global scale if it's the case
- Accountability: actions should have a responsible actor and a justification
- Transparency: Uncertainties, controversies, and limitations should be made public and updated as the research develops

3. Finance

The Department of Homeland Security (DHS) lists financial service[13] as one of the 16 critical infrastructure sectors for USA. While the work to seek safe forms of utilizing new technologies does have a cost, in an effort exercise the security of financial systems, the Norwich University worked on a $9.9 million contract, in 2013, to develop the Distributed Environment for Decision-Making Exercises – Financial Sector (DECIDE-FS) tool, awarded by Cyber Security Division of the DHS Security Science and Technology Directorate (CMAWR).[6]

Even though big investments in safe usage of software can be found, The Malicious Use of Artificial Intelligence report does question if existing funding strategies, like to put a bounty on a vulnerability, should be extended to AI technologies, and if such bounties could be offered by third parties like government or philanthropic sources.[2]

The same financial concern can be seen in the papers on biorisk published by the Future of Life Institute. The risk of human extinction is presumably low and reducing this risks chance does have a cost. Historically, costs in public health can be high, like $13 billion on health security-related programs in 2017 projected by US federal government.[14] In the work of Millett,[15] the cost of existential risk prevention is not cost-effective when compared to basic healthcare investments, but it does show up as cost-effective when compared to the benefit of the investment. An initial estimate shows that it takes around 10 cents of a dollar to save 1 life-year. Farquhar[7] goes into details on how to price and charge, were by identifying liability in case of catastrophe and by assessing the risk of a research, research institutions can pay upfront for the risks and be incentivized to minimize its chances of happening. The fair price can be pursued by borrowing methods used by insurance companies that are already pricing risks such as cyberattack.[16]

4. Similarities

Given the work on safety and security in both biology and AI, the following similarities can be listed:

(i) Dual-use research: the concern shown by gain-of-function research funders, that the research can be used with malicious intent, can be seen in recent works of major AI researchers when questioning the risk of existing AI systems being repurposed by terrorist
(ii) Ethical framework: the need for a methodology and a set of parameters and values to reconcile ethical concerns in AI systems, being worked by the IEEE 7000 project,[17] where the goal of the NSABB report published in 2016 on gain-of-function research
(iii) Financial backing: the decision to fund a research in gain-of-function opened the same questions that can be made for dual-use research in AI
(iv) Public interest: the motivation to fund a research, even after its risk assessment, can be seen in both areas. Gain-of-function researches pursued better understanding the virus H1N1 and aided the management of the 2009 pandemic,[7] where the project 7010 by IEEE[18] elaborate well-being indicators also aiming at the public interest.

Given the similarities, it's natural to see the common response where researchers and their institutions came together to ensure the safe development of their areas. The work on safe and secure research on biology had an earlier start and is in a more mature state, already substantiating research fundings cuttings. As outlined in this article, there is a good number of recent initiatives on safe and secure AI research which points to a future that can be beneficially permeated by the technologies we are researching today. To mitigate the risks involved it's important to develop and mature the work on creating guidelines and rules for ethically aligned AI systems and the work on biology can be a reference and inspiration.

Bibliografia

1. B. Gonzales, Riascos, *How Artificial Intelligence is Supporting Neuroscience Research: A Discussion About Foundations, Methods and Applications*, tech. rep. (2017), https://link.springer.com/chapter/10.1007/978-3-319-71011-2_6.
2. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, tech. rep. (2018), https://arxiv.org/pdf/1802.07228.pdf.
3. J. Stewart, Tesla's autopilot was involved in another deadly car crash (2018), https://www.wired.com/story/tesla-autopilot-self-driving-crash-california/.
4. U. Parliament, Ai in the uk: ready, willing and able? (2017), https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm.
5. *Ethically Aligned Design A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, tech. rep., IEEE (2017).
6. *Cyber Mutual Assistance Workshop Report*, tech. rep., Carnegie Mellon University (2018).
7. O. C.-B. Sebastian F and A. Snyder-Beattie, Pricing externalities to balance public risks and benefits of research (2017), https://www.liebertpub.com/doi/pdfplus/10.1089/hs.2016.0118.
8. M. J. Selgelid, Sebastian farquhar, owen cotton-barratt, and andrew snyder-beattie (2016), https://link.springer.com/content/pdf/10.1007%2Fs11948-016-9810-1.pdf.
9. *Recommendations for the Evaluation and Oversight of Proposed Gain-offunction Research*, tech. rep., National Science Advisory Board for Biosecurity (2016), http://www.iucn-whsg.org/sites/default/files/People,%20Pathogens%20and%20Our%20Planet.pdf.
10. *The Belmont Report*, tech. rep., Department of Health, Education, and Welfare. (1979), https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.
11. N. Kass, An ethics framework for public health. (2001), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446875/.
12. New directions. the ethics of synthetic biology and emerging technologies. (2010), http://bioethics.gov/sites/default/files/PCSBI-Synthetic-Biology-Report-12.16.10_0.pdf.
13. Critical infrastructure sectors (2017), https://www.dhs.gov/critical-infrastructure-sectors.
14. People, pathogens and our planet, volume 2: The economics of one health. (2012).
15. P. Millett and A. Snyder-Beattie, Existential risk and cost-effective biosecurity (2017), https://www.liebertpub.com/doi/pdfplus/10.1089/hs.2017.0028.
16. M. M. Daniel Garrie, Cyber-security insurance: Navigating the landscape of a growing field (2014), https://repository.jmls.edu/cgi/viewcontent.cgi?article=1766&context=jitpl.
17. 7000 - model process for addressing ethical concerns during system design (2016), https://standards.ieee.org/develop/project/7000.html.
18. 7010 - wellbeing metrics standard for ethical artificial intelligence and autonomous systems (2017), https://standards.ieee.org/develop/project/7010.html.