AI and global ethical issues

Guglielmo Tamburrini
Università di Napoli Federico II
Italy

Digital Humanism Summer School
Wien, September 20th, 2022

Global ethical issues

- **Global**: **a.** of, relating to, or involving the entire world
    - https://www.merriam-webster.com/dictionary/global
- **Global ethical issue**: affecting welfare, duties, fundamental rights of all human beings at the same time.
- Global ethical issues come with
    - Pandemic
    - *Climate crisis*
    - *Nuclear conflict*

Local and global in AI ethics

- **Local**
  - AI system predictions, classifications and decisions affecting selected populations
    - Fairness in access to education and careers, public and private services, …
      - Focus of UE proposal for regulating AI (2021)
        - https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
- **Global**
  - Climate crisis & AI carbon footprint
  - Nuclear conflict & AI in NC3 (= Nuclear Command, Control, and Communication)

Climate crisis and AI

- **AI is part of the solution**
  - AI-driven management of energy-intensive sectors
- **AI is part of the problem**
  - Training of some NLP models $\approx_{GHG}$ 5 mid-size cars in their life-cycle
    - Strubell E., Ganesh A., McCallum A. (2019) 'Energy and Policy Considerations for Deep Learning in NLP', *arxiv.org.1906.02243*. https://arxiv.org/abs/1906.02243
  - AI model training (10% electricity consumption);
  - AI model use after training (90% electricity consumption)
    - Patterson, D.; Gonzales, J.; Le, Q.; Liang, C.; Mungia, L.M.; Rotchchild, D.; So, D.; Texier, M.; Dean, J. Carbon Emissions and Large Neural Network Training. (2021), https://arxiv.org/abs/2104.10350

Ethical analysis: the AI pipeline

- **Multiple processes**
  - AI model training and experimenting
  - AI model use
  - Hardware production & operation
  - Data centers
  - Electricity supply from fossil sources
- **Multiple actors**
  - Within AI: scientists, AI industry executives, professional associations boards
  - Without AI: hardware producers and users, data center managers, electricity producers
- **Multiple responsibilities**
  - Which responsibilities of AI scientists for reducing the AI carbon footprint?

Entrenched practices in AI research

- AI model **accuracy** is mostly the single-minded goal
  - shapes inquiry and success in the AI research community
  - usually achieved by increasing parameters of AI models and hyperparameters in training and experimenting
- **Efficiency** is mostly neglected
  - In a random sample of 100 papers from the 2019 NeurIPS proceedings, one paper only "measured energy in some way, 45 measured runtime in some way, 46 provided the hardware used, 17 provided some measure of computational complexity (e.g., compute-time, FPOs, parameters), and 0 provided carbon metrics."
    - Henderson, P.; Hu, J.; Romoff, J.; Brunskill, E.; Jurafsky, D.; Pineau, J. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. J. Mach. Learn. Res. 2020, 21, 1–43, p. 6, https://arxiv.org/abs/2002.05651

AI scientists: ethically motivated actions

- **Develop** systematic knowledge of AI carbon footprint
  - develop metrics to measure overall AI carbon footprint
- **Design** energy efficient model architectures and training processes
  - curb the race towards larger AI models & training sets
- **Improve** computational efficiency of AI results
    - correlated to electricity consumption, independent of computing infrastructures and energy sources mix (as measured in terms of required Floating Point Operations - FPO)
        - Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. CACM 2020, 63, 54–63. https://dl.acm.org/doi/10.1145/3381831
- **Engage** with non-AI communities within the AI pipeline
  - Use more energy efficient hardware
  - Choose facilities and temporal windows receiving cleaner electricity supplies

New goals and practices for the AI community

- A new idea of what is a “good” AI result
  - The idea of combining accuracy and efficiency connects well to John McCarthy’s goal for AI: understanding how any ***computationally limited*** system – either biological or artificial – copes with complex information processing
- AI research community promoting green AI
  - AI competitions prizing efficiency
    - Build on the tradition of research programmes cashing on competitive games
      - Chess, Go, Robocup, Darpa Grand Challenge,…
    - Which categories/leagues for mixed accuracy/efficiency competitions?
- Provide role models for ICT at large
  - E.g. blockchain

Interlude
MAD – Mutually Assured Destruction

- *Climate* MAD
  - "**Addiction to fossil fuels is mutually assured destruction**… Instead of hitting the brakes on the decarbonization of the global economy, now is the time to race towards a renewable energy future." Antonio Guterrez, UN Secretary General (March 21st, 2022)
    - https://twitter.com/antonioguterres/status/1505992640843685899
- *Nuclear* MAD
  - Antonio Guterres's remarks on 2022 anniversary of Hiroshima and Nagasaki: "Crises with grave nuclear undertones are spreading fast — from the Middle East, to the Korean peninsula, to Russia's invasion of Ukraine.It is totally unacceptable for states in possession of nuclear weapons to admit the possibility of nuclear war. Humanity is playing with a loaded gun."
    - https://www.un.org/sg/en/content/sg/statement/2022-08-06/secretary-generals-remarks-the-anniversary-of-the-atomic-bombing-of-hiroshima

AI and nuclear conflict risks

Integrate “AI-enabled technologies into every facet of warfighting”

- US National Security Commission on Artificial Intelligence (NSCAI 2021)

“Promote all kinds of AI technology to become quickly embedded in the field of national defense innovation”

- “New Generation Artificial Intelligence Development Plan” (China’s State Council 2017)

“Whoever becomes the leader in AI will become the ruler of the world”

- Vladimir Putin (Russia Today 2017)

Integrating AI in NC3?

NC3= Nuclear Command, Control, and Communication

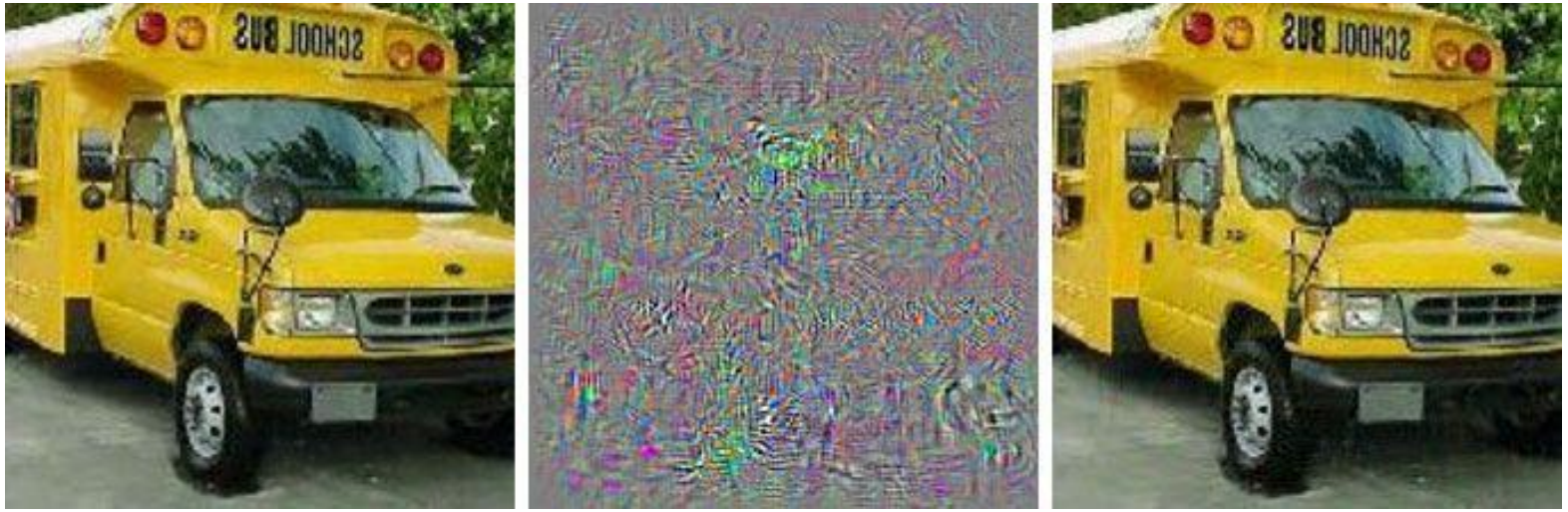
National Security Commission on AI, Final report 2021

- "AI should assist in some aspects of nuclear command and control: **early warning**, early launch detection, and multi-sensor fusion…" (p. 104)
- **Motivation:** AI may increase reliability, reduce accident risks, shorten processing time, buy more time for decision-makers
- **Cons** - AI brittleness and vulnerabilities

unintended misclassifications

- 1983: due to unusual weather conditions, Soviet early warning system OKO mistook sunlight reflecting on clouds for engines of 5 incoming Intercontinental Ballistic Missiles
- Colonel Petrov's commonsense reasoning :"when people start a war, they don't start it with only five missiles."
  - https://www.armscontrol.org/act/2019-12/focus/nuclear-false-warnings-risk-catastrophe
  - *AI still lacks commonsense*
  - *Which representative big data for AI to learn from?*

induced misclassifications in the lab
Adversarial AI



Szegedi C. et al. (2014). Intriguing properties of Neural Networks https://arxiv.org/abs/1312.6199

Image *classified as*

Adversarial perturbation *10x*

Perturbed image *classified as*

induced misclassifications in the wild
Adversarial AI

- Gnanasambandam A., Sherman A.M., Chan S.H. (2021), **Optical Adversarial Attacks**, IEEE/CVF International Conference on Computer Vision Workshops (ICCVW 2021), 92-101.

Stop sign

45 mhp speed limit

NC3 and AI explanation problem
Adversarial AI

- NC3 situational awareness: operators understand the ***why*** of machine suggestions/outputs

- *Familiar obstacle*: black box nature of many successful AI systems
- *Proposed mitigation*: XAI (eXplainable Artificial Intelligence)
  - Why did you do that? Why did you make this classification?
- *Emerging countermeasure*: Adversarial XAI
  - Inducing incorrect explanations

Beyond AI in NC3
Additional AI threats to nuclear stability

- AI and autonomous (underwater) vessels
  - AUVs trail submarines carrying SLBMs and undermine their stealth.
- AI information warfare
  - Deep fakes erode political leaders' credibility, fueling misperception of nuclear threats and second-strike posture
- AI in cyberwarfare
  - may increase speed and destructiveness of cyberattacks (to NC3)
- AI-powered autonomous weapons
  - have the potential to give large conventional military advantages to adopters and tilt the conventional military balance, incentivizing use of nuclear weapons to avoid military defeat

Responsibilities for AI scientists
… after physics, chemistry, biology

- Engage with public opinion, political and military leaders to raise awareness about AI potential threats to nuclear stability.
- Contribute to establish venues for international scientific and political dialogue
  - to affirm that only human beings and no AI system can authorize employment of nuclear weapons
  - to prohibit AI-powered cyber attacks on critical infrastructures, including nuclear facilities.
- Propose trust and confidence building measures

Responsibilities for AI scientists (continued)

- Clarify that emerging AI threats and vulnerabilities to nuclear stability add new motives for nuclear non-proliferation and disarmament (in the wake of Russell-Einstein manifesto),
- Explore AI's potential for nuclear non-proliferation and disarmament, by integrating AI in compliance monitoring of nuclear arsenals and treaties.

4. Concluding reflections

- AI success season leads to global AI issues
  - climate crisis: horizontal pervasiveness of AI
  - nuclear war risk: vertical pervasiveness of AI
- Global ethical issues insufficiently emphasized (in EU risk pyramid)
- AI communities engagement on global ethical issues
  - understanding conceptual and technical issues
  - identifying ethically motivated actions for AI actors
- Contribution on global issues from DigHum Project Sessions?

Thank you for your kind attention!

- Tamburrini G. (2022). The AI carbon footprint and responsibilities of AI scientists, *Philosophies* 7(1), 4. https://www.mdpi.com/2409-9287/7/1/4
- Amoroso D., Garcia D., Tamburrini G. (2022). The Weapon that Mistook a School Bus for an Ostrich *Science & Diplomacy*, 1-3, ISSN: 2167-8626, https://www.sciencediplomacy.org/article/2022/weapon-mistook-school-bus-for-ostrich
- Tamburrini G. (forthcoming). Nuclear weapons and the militarization of AI, *Proceedings of the XXIII Edoardo Amaldi Conference*, Accademia dei Lincei, Rome, 6-8 April 2022, P. Cotta-Ramusino et al. (eds), Springer Verlag