Toward a Computational Sustainability for AI/ML to Foster Responsibility

Eva Thelisson
University of Fribourg
eva.thelisson@unifr.ch

Abstract

This paper proposes to develop a new field of research designated as computational sustainability. It takes into account legal and ethical considerations of Artificial Intelligence (AI) and Machine Learning (ML) Technologies. As AI and ML will deeply impact the society within the next decade, this paper raises the awareness that technology is not value neutral and that technologists shall take responsibility for the ethical and social impact of their work. In particular, this paper aims at considering the last AI and ML developments and its convergence with associated technologies like Nanotechnology, Biotechnology, Information Technology, Cognitive Science (NBIC). The challenge is to reflect on the finalities of AI / ML Technologies, while referring to the Philosophy, Ethical Theory, Ethical Principles and Soft Law Mechanisms. Those Mechanisms refer to rules that are not strictly binding in nature (like guidelines or codes of conduct which set standards of conduct). National competetent authorities may encourage their development, rewarding their implementation or making them enforceable. AI Codes of Conducts and Quality Labels may play a key role in developing computational sustainability for AI / ML Technologies, in parallel to the development of Hard Law Mechanisms based for example on an International Convention on Civil Liability for Algorithmic Damages or a Digital Geneva Convention.

Introduction

Artificial Intelligence can be defined as "a tool whose purpose is to understand the mind from a new perspective". (Burton, 2017) It specificity is its "capacity of learning and applying intelligence to a wide variety of tasks: some as robots able to take action in our physical and social world, and some as software agents that make decisions in fractions of a second, controlling huge swaths of the economy and our daily lives" (Burton, 2017). In November 2017, Stephen Hawking raised his concerns about AI/ML Technologies at the opening ceremony of the Web Summit, that was held in Portugal. He qualified the future work in this area as "crucial to the future of our civilization and of our species". "The rise of powerful AI will be either the best, or the worst thing, ever to happen to humanity. We do not yet know which."

Trust is required in the ethical implications of the actions and decisions of AI systems, which shall act in a morally acceptable manner. Therefore, "we need to integrate moral, societal and legal values with technological developments in Artificial Intelligence, both within the design process as well as part of the deliberation algorithms employed by these systems" (Dignum, 2017). This includes social norms and professional codes, that must be embedded into these systems. This is a key element, as the technological developments aims at creating a human-machine symbiotic system with the capability to make optimal decisions. Designing ethical preferences and ethical reasoning frameworks may help defining priorities over actions.

We need to take the time to think carefully about AI finalities, the level of risks at each development of new product or service, without being blinded by ideologies (transhumanism, or security-based ideology) or by the outlook for profitability. Taking into account a risk-based approach: "AI categories shall be defined and only provably beneficial systems of the highest categories may get a certification by a safety authority and be put on the market" (Stuart Russel, 2017). A European Safety Authority dedicated to AI may be created in the EU to implement this approach.

The 47th World Economic Forum (WEF) Annual Meeting held on 17-20 January 2017 in Davos-Klosters, Switzerland was dedicated to the theme Responsive and Responsible Leadership. It put forward the urgent need of considering inclusive principles in designing altruistic and human-centered artificial intelligence systems. How to implement ethical principles to human-designed artifacts? This is of key importance as those artifacts are (or will soon be) capable of making their own decisions based on their perceptions (Burton, 2017). To identify which principles have to be taken into consideration for AI Technologies, we have to refer to the Philosophy.

Ethical Theory

Aristotle developed the first Theory, the Virtue Ethics or Teleological Ethics, in particular in Nichomachean Ethics. (Annas, 2006). This Theory emphasizes the virtues, or moral character. A central component of this theory is "phronesis" ("moral prudence" or "practical wisdom"). The ACM Communication made by Toby Walsh in 2015, namely "The Turing Red Flag" pursues along the same theory. The second theory of Ethics is the **deontological Ethics** that was developed in particular by Immanuel Kant in the 18th century. What makes a choice right is its conformity with a moral norm. This Theory emphasizes moral duties or rules. If an act is not in accord with the moral norm or with the "Right", it may not be undertaken, no matter the Good that it might produce (Stanford Encyclopedia of Philosophy). Jeremy Bentham and John Stuart Mill developed the third Theory, the theory of Utilitarianism, in the late 18th century. The basic question of utilitarianism is "what is the greatest possible Good for the greatest number?" Utilitarianism is the foundation for the game-theoretic notion of rationality as selecting actions that maximize expected utility, where utility is a representation of the individual agent's preference. Games Theory is often used in AI to understand how individuals or groups of agents will interact. (Burton, 2017). An utilitarian will point out the fact that the consequences of helping someone in need will maximize well-being, a deontologist to the fact

that, in doing so the agent will be acting in accordance with a moral rule such as "Do unto others as you would be done by" and a virtue ethicist to the fact that helping the person would be charitable or benevolent. Later the Asimov three laws for Robots as well as Nick Bostrom researches on Ethics raised the level of awareness on Ethics. Which principles can we deduce from those Theories for AI/ML principles?

Ethical principles

We can deduct from those theories, some ethical principles that AI Community shall implement. First, AI Technologies should be designed in a way to provide safeguards, increase social benefit, enhancing fairness, freedom, fraternity and equality among individuals in society. A fairly distribution of AI/ML Technologies across society and across developing and developed countries should be encouraged. AI Technologies "should improve people's lives, placing their rights and well-being at its very heart" (Google Deep Mind, 2017). An embargo on the use of genetic data by insurance companies should be encouraged. Second, all AI applications should remain under meaningful human control, and be used for socially beneficial purposes. Should predictive policing based on algorithms and facial recognittion software be legal? Third, AI research must be evidence-based and explore the opportunities and challenges posed by these technologies. Fourth, AI research must occur in a transparent and open manner. Any funding must be disclosed. AI projects must be interdisciplinary and involve a diverse set of people with various backgrounds to be able to identify a large panel of risks and viewpoints. This is of key importance to build a safe, secure, efficient and reliable AI Technologies, beneficial for Humankind.

Computational sustainability

Computational sustainability is a new field of research. It encompasses computational methods of a sustainable environment, economy and society. This notion was inspired by the emergence of sustainable development, i.e. the development that meets the need of the present generation without compromising the ability of future generations to meet their needs (Bruntland Commission, United Nations, 1983). Computational sustainability analyzes how computational techniques, including AI, can be used to improve planetary sustainability in the ecological, economic and social realms. AI researchers and development engineers potentially have part of the skills required to address aspects of concerns of global warming, poverty, food production, arms control, health, education, the aging population, and demographic issues. If the primary goal of computational sustainability was to tackle the environmental and sustainable challenges facing the planet, we propose to expand the Computational Sustainability to risk management, legal and ethical aspects of AI / ML technologies. The convergence of NBIC raises in particular new challenges for the next generations. Targeted genome editing via CRISPR-CAS 9 Technology), or seamless Robotization of Human and the Humanization of Robots (Abi Ghanem, 2017) are some of the concrete challenges raised by those technologies.

Towards a Responsible AI Label? In order to build trust in AI/ML Technologies, and to foster a responsible AI, the WEF-called for inclusion of values and responsible requirements embedded in the design. Standards for AI / ML design shall be developed. The International Standard Organization, the Royal Society or a new conference similar to the Asilomar Conference of 1975 may play an important role in setting those standards. AI industry shall also adopt AI / ML Soft Law Mechanisms, approved by Competent Authorities

Conclusion

A novel area of research concerned with transparency and ethical accountability is emerging across the science and technology. It advocates a different kind of relationship between innovations, stakeholders and researchers/innovators based on accountability, trust and transparency, as part of a computational sustainability policy. The effectiveness of this approach will depend on its sectorial approach, on the existence of dissuasive sanctions as well as competent authorities' engagement and approvals of Soft Law mechanisms. This co-regulation could foster innovation while promoting a responsible AI at international level, in parallel of a binding legislative framework.

References

ANDERSON, M. and ANDERSON, S. L. (2011). Machine ethics. Cambridge University Press.

ANNAS, J. (2006). Virtue ethics. In Copp, D., editor, The Oxford Handbook of Ethical Theory. Oxford University Press.

ARISTOTLE (1999). Nichomachean Ethics. Hackett. trans. Terence Irwin.

ASIMOV, I. (1950). I, Robot. HenryBennet.

AACH, John, LUNSHOF, Jeantine, IYER, Eswar, et al.Addressing the ethical issues raised by synthetic human entities with embryo-like features. Elife, 2017, vol. 6, p. e20674.

ABI GHANEM,Antoine, The Humanization of robots and the Robotization of the Human Person, Ethical reflections on lethal autonmous weapons systems and augmented soldiers, The Caritas in Veritate Foundations Working Papers, November 2017.

BURTON, Emanuelle, GOLDSMITH, Judy, KOENIG, Sven, et al.Ethical considerations in artificial intelligence courses. arXiv preprint arXiv:1701.07769, 2017

HE, Cheng, YAO, Yun-jin, et YE, Xue-song. An Emotion Recognition System Based on Physiological Signals Obtained by Wearable Sensors. In : Wearable Sensors and Robots. Springer Singapore, 2017. p. 15-25.

FERRY, Luc, la révolution transhumaniste, Paris, 2016.

FRIEDMAN, B. and Yoo, D.(2017). Pause: A multi-lifespan design mechanism. In *Proceedings of CHI 2017*, 460-464. New York, NY: ACM Press. [pdf]

FUKUYAMA, Francis, La fin de l'homme, 2002.

MATHUROS, F. (2016, 09 14). World Economic Forum's 47th Annual Meeting Calls for Responsive Leadership. Retrieved 05 10, 2017, from World Economic Forum: https://www.weforum.org/press/2016/09/world-economic-forum-47thannual-meeting-calls-for-responsive-leadership/

POOLE, D and MACKWORTH, A, Artificial Intelligence, Foundations of computational agents, second edition, Cambridge University, 2017.

WEISER, Mark. Some computer science issues in ubiquitous computing. Communications of the ACM, 1993, vol. 36, no 7, p. 75-84.