ATINER's Conference Paper Proceedings Series
SOS2023-0313
Athens, 4 October 2023

The Emergence and Spread of the AI for Good Movement in Organisations

Hélène Jeannin

Athens Institute for Education and Research
9 Chalkokondili Street, 10677 Athens, Greece

Hélène Jeannin, Sociologist Researcher, Orange – Humanities and Social Sciences Research Department, France

The Emergence and Spread of the AI for Good Movement in Organisations

ABSTRACT

Artificial Intelligence (AI) algorithms are taking an ever growing place in our lives. Their uses provoke fears, questions and high expectations due to the scale of the promises made. In 2018, Unesco warned the international community against the risks of AI. "It is our responsibility to lead a universal and enlightened debate – not a technical debate, but an ethical one – in order to enter this new era with our eyes wide open, without sacrificing our values, and to make it possible [...] to establish a common global foundation of ethical principles," said its Director-General, Audrey Azoulay . The years 2015-2016, in particular, marked a real turning point in realising the importance of the ethical challenges of artificial intelligence. They demonstrate a transition that is both technological and societal. It is about defining values and requirements to be respected to ensure that artificial intelligence serves the interests of human beings. The initiatives in favour of AI Ethics observed from 2015 onwards (Friedenberg, 2015) have been so numerous and vibrant that they have given rise to a craze, which some have dubbed an "ethics rush" (Georges, 2019), or a "tidal wave" (Wright & Schultz, 2018). What do these initiatives involve? What form do they take? Why and how are they being put in place? In this article, we will retrace the major steps of the path that led to the emergence and then the spread of the "responsible AI" movement, particularly in three spheres: science, media and politics, and businesses and organisations.

Keywords: *artificial intelligence – ethics – algorithms - risks*

Triggers from the Scientific Sphere

Isolated but Influential Voices

In 2014, a decisive essay by Nick Bostrom, "*Superintelligence",* which analysed scenarios for the extinction of the human species[1], revived the controversy surrounding the concept of the *singularity,* whereby the human mind will be overtaken by artificial intelligence. A polymath with expertise in physics, computational neuroscience, mathematical logic and philosophy, Nick Bostrom was at the time a Professor of Applied Ethics at the University of Oxford and the director of several institutes studying the technologies of the future. He has just been included in Prospect magazine's list of world thinkers, where he was the youngest person in the top 15. He was the top-ranking analytic philosopher across all fields (according to the website The Conversation)[2]. With some 200 publications and more than 500 Interviews for TV, film, radio and the written press to his name, he has extensive and varied national and international media experience[3], making him an extremely influential scientist.

His essay became an international bestseller. Translated into around twenty languages, it contributed to the development of the theories of the technological singularity and took the debate on AI Ethics global. It was endorsed by Elon Musk and astrophysicist Stephen Hawking[4]. The former called AI "the biggest existential threat"[5], while the latter declared that AI could put an end to humanity[6]. Bill Gates joined them, warning of the threats posed by advances in AI[7].

These voices were rare, but their influence considerable. They were to find major resonance.

The Technical and Scientific Communities

At the same time, following the *NIPS* conference in December 2014[8], scientists met to discuss the consequences of these negative statements on their research (Fondation Telecom, 2016). In their eyes, researchers, aware of the transformative effect of their work, have a fundamental role to play. The question of the influence and impact of their work and practices must remain a central concern for them, because it is an ethical dimension (Paoletti, 2014). As

---

[1] https://www.nickbostrom.com/existential/risks.html
[2] https://theconversation.com/profiles/nick-bostrom-139186
[3] https://nickbostrom.com/media.pdf
[4] died 14 March 2018
[5] https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat ; https://www.vox.com/future-perfect/2018/11/2/18053418/elon-musk-artificial-intelligence-google-deepmind-openai
[6] https://www.bbc.com/news/technology-30290540
[7] https://www.cnet.com/science/bill-gates-is-worried-about-artificial-intelligence-too/
[8] NIPS: Neural Information Processing Systems Conference, 8-13 December 2014, Montreal Convention Center, Canada, http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=21362

Jacques Testard[9] shows, contemporary science is an activity that brings multiple interests (scientific, economic, industrial, military) into play, interests that are sometimes contradictory due to their close interrelations, competition between teams, and directions of travel decided on by structured, hierarchical institutions. There are, however, some researchers who take responsibility for what will become of their knowledge. Xavier Guchet puts it this way: "A *responsible* science is a science capable of anticipating the impacts of its technical applications on nature and on society" (Guchet, 2016).

The technical and academic communities are therefore showing growing concern about the possible effects of misuse of AI. Since 2010, AI techniques have vastly improved, with the introduction of a new category of algorithms, increased computing power, and the availability of ever more and better-quality data (Big Data). The combination of these elements, together with progress in AI's constituent disciplines, explains how ever more sophisticated tasks will be able to be performed, and at ever faster speeds, using machine learning techniques that are capable of improving autonomously. While capacities in this field and others are moving beyond the laboratory research stage to become economically viable technologies, a virtuous cycle is taking root, in which even small improvements in performance lead to major financial gains, a state of affairs which encourages greater investment in research. It is now widely accepted that AI research is progressing steadily and that its impact on society is likely to grow.

Whistleblowers and Open Letters

In January 2015, at the suggestion of Stuart Russel, an AI specialist and professor at the University of California at Berkeley, a dozen eminent researchers signed an open letter to the *Future of Life Institute* website, calling on their peers to look beyond the historic objective of technical performance for AI. The letter stipulates that it is not enough for AI systems to become more and more robust; they must also maximise their societal benefits.

In two months, the signatories assembled more than 300 research groups comprising computer scientists, physicists and philosophers from the worlds of academia and industry. In June 2016, more than 8,600 of them, enthusiastic and fearful in equal measure, called for work to be done on an ethical direction for AI work[10] and new objectives. This research was necessarily interdisciplinary, as it concerned both society and AI.

---

[9] http://jacques.testart.free.fr/pdf/texte794.pdf
[10] https://futureoflife.org/ai-open-letter

Concerns Reach the Public and Political Spheres

Media Coverage of Spectacular Events

In 2016, several major events drew public attention to AI:

o A game of Go between the program AlphaGo and the South Korean world champion Lee Sedol, organised by Google DeepMind and live-streamed on YouTube, showed the algorithm's supremacy over humans.
o Tay, a chatbot, came out with inappropriate racist and sexist tweets after a learning phase, which caused Microsoft to pull out the service completely.
o The first mortal accident involving a self-driving Tesla rekindled the controversy about self-driving cars and issues of liability.

The above-mentioned examples all sent signals that concerned public opinion and prompted controversies. This has been exacerbated by the fact that some Hollywood films[11] and works of fiction model a certain way of thinking about artificial intelligence. In these works, AI is often presented as a source of anxiety and disappointment[12].

It is becoming vital to establish a framework for AI developments on the international stage and to assert some principles to control its varied implications in everyday life.

International Organisations

In June 2017, the ITU (International Telecommunications Union[13]) and the XPrize Foundation[14] co-organised a summit in Geneva devoted to "Artificial Intelligence for Good", inviting 500 representatives of governments, industry, academic and research institutes, the United Nations and civil society to "explore the potential of AI to accelerate progress in fields such as health, education, energy, and the protection of our environment[15]". Convinced that AI will help resolve one of the most pressing challenges of our planet and its populations, they affirmed their faith in its promises for a better life for all, provided that there is cooperation to develop the positive aspects of the technology and control the risks it entails. With the *Association for Computing*

---

[11]i.e. the industrial, commercial US cinema which is responsible for most big-budget productions distributed worldwide.

[12]*A.I. Artificial Intelligence*, Steven Spielberg, 2001; *Her,* Spike Jonze, 2014; *Ex Machina*, Alex Garland, 2015

[13]United Nations agency specialising in Information and Communication Technologies, comprising 193 member countries and nearly 800 academic and private-sector organisations; https://www.itu.int/en/about/Pages/default.aspx

[14]https://www.xprize.org/

[15]https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf

Machinery (ACM) and several United Nations agencies[16], the digital platform "AI for Good" was launched[17]. It aims to collectively implement a programme of concrete actions to meet the United Nations' 17 Sustainable Development Goals for 2030.

The phrase *AI for Good* struck a chord and was picked up by many other organisations, helping to spread the *AI for Good* movement, and making AI "a force for Good" (Floridi & Taddeo, 2018). France, too, picked up on the concept, dubbing it "Responsible AI", with responsibility defined as the obligation of a person to answer for their actions by virtue of the role and duties they take on, and then bear all the consequences.

The Difficulties of Practical Implementation

Contemporary debates often centre around ethical or moral issues, also known as "conflicts of values". These are situations in which values or principles enter into conflict and make decisions difficult, because the individual – or the group – is torn between several principles that they find important. For a problem to be moral or ethical, it must bring into play ideals that give meaning to our lives or rules of behaviour that we feel obliged to follow. These principles and values are liable to change. This is why ethics is an eminently sociocultural domain.

Establishing ethical practices is not self-evident, as the notion of ethics poses a certain number of limits: philosophical (the very raison d'être of ethics is to pose dilemmas for which there is, by definition, no good outcome), scientific and technical. Implementing the chosen ethical theories can run into operational difficulties, creating a risk of a possible chasm between theoreticians and practitioners, if the two do not engage in dialogue. In business, "ethics always faces a conflict of values. It is therefore necessary to classify and prioritise principles or values based on individual or collective interests, in order to take the decision that seems the most appropriate. This requires complex reasoning based on contextual analysis, the concept of a sense of values and duties, and a capacity to anticipate" (CIGREF, 2016, p. 49).

All these elements lead to a need to equip ourselves with technical tools and governance structures to deal with the ethical issues raised by AI.

In addition to AI Ethics, we then have Data Ethics, which is becoming a new branch in its own right, studying and assessing more particularly the issues linked to data (generation, recording, processing, dissemination, sharing and use), algorithms and related practices (including responsible innovation, programming, and professional codes). The quality of datasets or biases – which could reinforce discrimination and prejudices while outwardly appearing neutral and objective – are fundamental in AI developments. A group of experts led by Luciano Floridi and Mariarosaria Taddeo has identified three areas to consider in this field: data ethics (the issues raised by the collection,

---

[16]https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx
[17]https://aiforgood.itu.int/about-ai-for-good/

analysis and use of wide datasets); algorithm ethics (the issues raised by the complexity and autonomy of algorithms), and finally, the ethics of practices (the responsibilities of individuals and organisations). These three aspects must be dealt with jointly, as they are interdependent (Floridi & Taddeo, 2016).

Applied Ethics in Organisations

The Rise of Ethics

Applied ethics deals with practical situations that raise ethical issues. The focus is often placed on supporting decision making to meet specific challenges, both in terms of the form and the decision-making process, and with regard to the values and principles at stake. Often, the applied ethics approach involves documenting and detailing the problematic situation using factual information: Who is likely to be affected by this situation? What are the applicable laws? What are the options available? What are the risks and possible consequences of each option for the stakeholders involved? The aim is to lay the groundwork for assessing the different options with regard to values and principles.

It is to answer these complex questions that studies and initiatives concerning the ethical and social stakes of the impact of artificial intelligence are emerging and multiplying in various spheres (Georges, 2019): grey literature, publications[18], scientific conferences, the creation of research chairs, observatories and specialist research centres addressing legal, technical and social issues, such as *The Ethics and Governance of Artificial Intelligence Initiative* by the MIT Media Lab and the Harvard Berkman-Klein Center for Internet and Society[19] (2017), the University of Guelph (Canada) and its AI Ethics centre[20] (2018), and 4 interdisciplinary institutes with the 3iA[21] label in France. They show the urgent need to establish an ethical framework around AI.

In AI, public-sector investments are not sufficient to reflect the reality of the situation. In the United States, most come from the private sector, with companies such as Google, Facebook, Twitter, Amazon, Netflix, Microsoft, IBM, Uber, Salesforce, Spotify and Apple – who also have the data and the skills – and an entire ecosystem of start-ups supported by substantial venture capital funding (Mercer & Macaulay, 2019). In China, too, the state has built close ties with private-sector players of the same kind (Nocetti, 2019, p. 17).

---

[18]A number of them can be found listed on the site: http://www.impact-ai.fr/publications/
[19]https://cyber.harvard.edu/events/2019-05-14/everyday-chaos
[20]https://news.uoguelph.ca/2018/12/u-of-g-launches-ethical-artificial-intelligence-centre/
[21]https://www.inria.fr/fr/instituts-interdisciplinaires-dintelligence-artificielle-3ia-les-resultats-definitifs

Ethics in Business

This is why companies are keen to establish their "AI for Good" credentials[22], publishing guidelines, appointing ethics committees, and rolling out training. Ethics guidelines, which generally fall within the scope of CSR (Corporate Social Responsibility) have often been criticised as whitewashing, useful for assuaging the conscience, saving face, getting off the hook, reassuring the consumer, or even for a communications or public relations operation in the event of a crisis. It seems they replace the legislator, saving them the trouble of doing their job. In the case of technologies, things are not always so simple, as we could also argue that the legislator, who is much less *au fait* than companies with the advances in R&D and experimentation happening in laboratories, is not as informed or competent as an expert in the field. This could explain why alarm bells are being sounded by companies and why, in the absence of regulation, these companies are anticipating by attempting to introduce safeguards in the form of ethics guidelines.

In some cases, companies are teaming up with universities, or with one another. In September 2016, Facebook, DeepMind and Google, Amazon, IBM, Apple, and Microsoft formed the *Partnership on AI*. Their stated aim was to recommended best practices and to conduct and publish research in various fields, including ethics[23].

Its European counterpart, *Impact AI*[24], was launched in France in 2018. This non-profit association is made up of 30 founding members – including Orange – with a shared vision and ambition: to be the leading authority on ethical artificial intelligence in France and to have a positive impact. The collective promotes the sharing of tools and the raising of public awareness, and supports AI development projects. To this end, working groups (Observatory, Responsible AI, AI for Good, Education) have been set up to open up Impact AI to ecosystems and potential partners in France and beyond. Today, its members include major companies, DSPs (Digital Service Providers), strategic consulting firms, AI ecosystem players, startups and schools.

Impact AI: Concrete Commitments

Above and beyond its desire to reflect on AI Ethics, Impact AI has set itself the mission of acting pragmatically. To do this, the collective selects socially innovative organisations and works with them to develop artificial intelligence solutions designed to increase their social impact. But it sometimes has a more political role. In 2019, it was approached by the AI HLEG (High Level Expert Group), a group of around fifty independent experts[25] appointed by the European Commission in 2018 to advise it on its AI strategy. In April

---

[22]https://www.prnewswire.com/news-releases/organizations-are-gearing-up-for-more-ethical-and-responsible-use-of-artificial-intelligence-finds-study-300714476.html
[23]https://partnershiponai.org/
[24]http://www.impact-ai.fr/
[25]https://digital-strategy.ec.europa.eu/fr/policies/expert-group-ai

2019, this group produced the report "Ethics Guidelines for Trustworthy AI[26]", which constituted a pioneering effort to establish rules for the application of AI for different uses, based on the Charter of Fundamental Rights of the European Union. In particular, it invited various stakeholders to beta-test an assessment list that was intended to guide the different roles (political, managerial, operational) involved in the assessment of a future AI-based product or service. This was where Impact AI came in.

I led workshops within my company (Orange) with ten or so people with complementary profiles internally. Starting from a use case (Alloscope, a tool to help those in vulnerable situations[27]), we developed a series of recommendations designed to improve that initial assessment list. Then, after comparing our points of view with those of two additional teams from other companies, we produced a summary report detailing the shared conclusions. To lend it greater weight, it was sent to the HLEG in the name of Impact AI. In July 2020, the HLEG published the final Assessment List for Trustworthy AI (ALTAI) in pdf format. It also made an operational tool available on the web free of charge, in the form of a dynamic self-assessment checklist for professionals in the field. These two deliverables[28] take our recommendations into account.

This commitment and participation show the importance of an approach based on sharing and dialogue as well as on the combination and cross-fertilising of ideas and influences.

Conclusion

The years 2015 and 2016 were pivotal in terms of ethics and AI, marking the starting point for a series of concrete initiatives from multiple players.

On the one hand, many reports and white papers on AI for Good or AI and ethics reflected the concerns of scientists and entrepreneurs and manifested a desire to avoid any excesses or wrong turns. On the other, ethics are a source of distrust. They are accused of being a whitewash, promising a better world, but actually disguising economic interests and a resistance to regulation. Under this reading, ethics would serve as an excuse for any legislation that would get in the way of business.

Explicitly setting out the issue of sincerity (feigned or otherwise) and/or opportunism in this way is ultimately only secondary, as it merely shifts the focus of the debate.

The question is therefore not so much to expose any insincerity as to consider the reach and the medium- and long-term consequences of these actions. Ethics is also a factor in competition between businesses, for both the customer and employees, as the rebellion at Google on the Maven military

---

[26]https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[27]https://hellofuture.orange.com/en/alloscope-a-tool-to-help-those-in-vulnerable-situations/
[28]https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

program showed (Jeannin, 2019). The consistency of the company's actions with its declarations is fundamental, as trust is a factor that is essential but easily affected by circumstances and events, requiring vigilance at all times. Reflections about ethics must also be considered in the context of a general climate characterised by a battle for political and geostrategic influence: commercial tensions and conflicts between the US and China are growing, as they struggle for economic hegemony through innovation.

The debate must take place throughout all of society… in every household. It is now important for every player on the ground (companies, public authorities, civil society) to set out its priority principles to be followed, in order to make ethics a part of decision-making mechanisms, algorithms and practices (Barocas, 2017).

Bibliography

Barocas S., Boyd D. (2017). Computing Ethics. Engaging the Ethics of Data Science in Practice: Seeking more common ground between data scientists and their critics. *Communication of the ACM,* 60(11), pp. 23-25

.Bostrom N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford : OUP Oxford.

CIGREF (2016). Gouvernance de l'intelligence artificielle dans les grandes entreprises. Enjeux managériaux, juridiques, éthiques. Paris: Cigref.

Floridi L., Taddeo M. (2016). What is data ethics? *Phil. Trans. R. Soc.* A 374: 20160360, http://dx.doi.org/10.1098/rsta.2016.036

Floridi L., Taddeo M. (2018). How AI can be a force for good. *Science,* 361(6404), pp. 751-752.

Fondation Telecom (2016). Intelligences Artificielles : Quelles promesses ? Quels défis, *Cahier de veille n°8*, Paris, Fondation Télécom.

Friedenberg M. (2015). A.I. Ethics Emerge. *CIO,* 28(11), p. 6

Georges B. (17 January 2019). *La ruée vers l'éthique de l'intelligence artificielle.* [Online] Available at: https://www.lesechos.fr/tech-medias/intelligence-artificiel_le/la-ruee-vers-lethique-de-lintelligence-artificielle-441344

Guchet X. (2016), L'éthique des techniques, entre réflexivité et instrumentalisation*, Revue française d'éthique appliquée*, n°2, pp. 8-10.

Jeannin H. (2019), Le projet Maven, entre pressions politiques et postures éthiques, *A Contrario,* n° 29, pp. 141-155.

Jeannin H. (2020), L'émergence du mouvement IA responsable dans les organisations: structuration et enjeux , *Communication & management*, vol. 17, pp. 105-120.

Mercer, C., Macaulay T. (2019). *How tech giants are investing in artificial intelligence.* [Online] Available at: https://www.techworld.com/picture-gallery/da_ta/tech-giants-investing-in-artificial-intelligence-3629737/

Nocetti J. (2019). *Intelligence artificielle et politique internationale. Les impacts d'une rupture technologique,* Paris, IFRI - Institut Français des Relations Internationales.

Paoletti I. (2014), Ethics and the Social Dimension of Research Activities, *Human Studies,* 37(2), pp. 257-277.

Wright S., Schultz A. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. Business Horizons, vol. 61, n° 6, pp. 823-832.