Reflections on AI for Humanity: Introduction

Bertrand Braunschweig, Malik Ghallab

Reflections on AI for Humanity: Introduction

Bertrand Braunschweig[1] and Malik Ghallab[2]

[1] Formerly Inria, Paris. bertrand.braunschweig@bilab.fr
[2] CNRS, LAAS, Toulouse. malik.ghallab@laas.fr

Abstract. This chapter briefly surveys the current situation of AI with respect to its human and social effects, and to its risks and challenges. It presents a few global initiatives regarding ethical, social and legal aspects of AI. It introduces the remaining chapters of the book and briefly discusses a global cooperation framework on AI and its governance.

1 Context of the book

Over the last two decades, Artificial Intelligence has moved from a technical area of interest to a focused community of specialists, to a widely popular issue, making the media headlines and bringing daily to the limelights new computational functions and applications. The effectiveness and potential of AI techniques became highly visible, attracting vast private investments and national R&D plans.

The social interest in AI is naturally amplified since its techniques are the mediating means between users and the digital world, which plays a predominant role in personal, social, and economic relations. Comparisons to and competitions with human in games and several tasks, sometimes transposed and exaggerated uncritically, have boosted the general attention. This interests is matched with a growing concern over several risks and infringements related to, for example, security, confidentiality, exploitation of personal data or opinion manipulation.

The concerns about AI have been expressed in numerous forums and programs seeking to steer the technical developments toward social good, to mitigate the risks and investigate ethical issues. This is illustrated through the initiatives taken by international organizations, such as the United Nations and its specialized agencies [24,39], the European Union [18,42], or the Organisation for Economic Cooperation and Development [30]. Many other initiatives have been taken by technical societies [17], NGOs, foundations, corporations, and academic organizations [22,25,16,21,15,36,14,20].

At the political level, statements from several leaders have placed AI as a geopolitical issue, a matter of power competition in international relations. Calls for cooperation have been delivered. Recent G7 summits promoted the idea of setting up a permanent *Global Partnership on AI* (GPAI), relying on international working groups and annual plenary meetings. In that perspective, the *Global Forum on AI for Humanity*, held in Paris in October 2019, gathered a large interdisciplinary audience over five workshops and eight technical sessions.

Its purpose was to provide an initial input to the GPAI working groups. This book results from the contributions and discussions help at this Global Forum. It is written by the organizers and moderators of the Forum debates.

2 What is AI today

Academic controversies about a proper definition of AI, as a science or as a technology, about its weak versus various versions of strength, or its symbolic old fashioned flavor versus its deep numeric one, may have their interest but are not very relevant to our purpose here. It is sufficient to say that AI techniques have demonstrated convincing results and a significant potential in the mechanization of cognitive functions, for *perceiving*, *reasoning*, *learning*, *acting* and *interacting*.

These techniques prosper on and enrich a large interdisciplinary background, mainly from computer science, mathematics, cognitive and neurosciences. They rely in particular on *(i)* data-based approaches, from probability, statistics, and numerical optimization, *(ii)* model-based approaches, from logic, ontologies, knowledge representations and structures, *(iii)* heuristic search and constraint propagation methods, and *(iv)* the fruitful synergies of their algorithmic integrations. They benefit from the tremendous growth of electronics and communication systems.

AI achievements already cover a broad set of capabilities such as image, speech and scene recognition, natural language processing and interaction, semantic information handling and search, automated planning, scheduling, and diagnosis, or computer aided design and decision making. Significant progress has been witnessed in almost all academic competitions and challenges which allow to compare approaches to these capabilities and structure developments.[1]

Successful applications of AI techniques can be found in almost every area of industry and services. Medicine and health have attracted significant developments. The very recent COVID-19 pandemic has already seen numerous proposals, for example in diagnosis and prognosis from medical imaging, protein structure planning for drug discovery, virus nucleic acid testing, epidemiology modeling and forecasting, and in text mining and analysis of the scientific literature.[2] Transportation is another area of significant AI developments and investments, e.g., in autonomous vehicles. Manufacturing and logistics implement AI over a broad spectrum of deployments, from the design and planning stages to the production stage with millions of robots in operation integrating more and more AI techniques. Similarly for mining, e.g., to support deep drills exploration or automated open-pit mining. Space applications are among the early success stories of AI, e.g., [5]. Defense and military applications are a matter of huge investments, as well as concerns. Precision and green agriculture relies on a range of sensing, monitoring and planning techniques as well as on versatile

[1] These are, for example, the challenges in image recognition [23], in question answering [35] and other natural language processing tasks [29], in automated planning [26], in theorem proving [34], and in logistics and other robotics competitions [33].

[2] See [2], an early survey on April 2020 of 140 references.

robots for weeding and crop management tasks. AI has been adopted very early in e-commerce for automated pricing, user profiling and (socially dubious) optimizations. Similarly in finance, e.g., in high frequency trading. Learning and decision making techniques are extensively used in banking, insurance, and consulting companies. Education institutions are routinely using advanced data and text management tools (e.g., timetabling, plagiarism detection). Personal tutoring techniques start being deployed.[3] Automated translation software and vocal assistants with speech recognition and synthesis are commonly marketed. This is also the case for very strong board, card and video games. Motion planning and automated character animation are successfully used by the film industry. Several natural language and document processing functions are employed by the media, law firms and many other businesses. Even graphical and musical artists experiment with AI synthesis tools for their work.

Key indicators for AI show a tremendous growth over the last two decades in research, industry and deployments across many countries. For example, the overall number of peer-reviewed publications has tripled over this period. Funding has increased at an average annual growth rate of 48%, reaching over $70B world wide. Out of a recent survey of 2360 large companies, 58% reported adopting AI in at least one function or business unit [28]. The AI labor demand vastly exceeds trained applicants and leads to a growing enrollment in AI education, as well as to incentives for quickly augmenting the AI schooling capacities.[4]

3 AI risks and challenges

AI techniques have clearly demonstrated their great beneficial potential for humanity. Numerous scientific and technical bottlenecks remain to be overcome, but progress is accelerating and the current state of the art is already providing approaches to many social challenges. This is illustrated in particular through several projects addressing with AI techniques the United Nations Sustainable Development Goals (SDGs) [38]. AI use cases have been identified for about half of the 169 SDG targets by a UN initiative on big data and artificial intelligence for development, humanitarian action, and peace [37].

However, as for any other technology, the development of AI entails risks. These risk are commensurate with AI impact and potential. Moreover, rapid technology developments do not leave enough time to social evaluation and adequate regulation. In addition, there are not enough incentives for risk assessment, in research as well as in industrial development; hence there are many more studies of new techniques than studies of their entailed risks.[5]

[3] e.g., [27,31], the two winner systems of the Global Learning XPrize competition in May 2019.

[4] These and other indicators are detailed in the recent AI Index Report [8].

[5] E.g., according to the survey [28] 13% companies adopting AI are taking actions for mitigating risks.

The main issues of AI are how to assess and mitigate the human, social and environment risks of its ubiquitous deployments in devices and applications, and how to drive its developments toward social good.

AI is deployed in safety critical applications, such as health, transportation, network and infrastructure management, surveillance and defense. The corresponding risks in human lives as well as in social and environmental costs are not sufficiently assessed. They give rise to significant challenges for the verification and validation of AI methods.

The individual uses of AI tools entail risks for the security of digital interaction, the privacy preserving and confidentiality of personal information. The insufficient transparency and intelligibility of current techniques imply other risks for uncritical and inadequate uses.

The social acceptability of a technology is much more demanding than the market acceptance. Among other things, social acceptability needs to take into account the long term, including possible impacts on future generations. It has to worry about social cohesion, employment, resource sharing, inclusion and social recognition. It needs to integrate the imperatives of human rights, historical, social, cultural and ethical values of a community. It should consider global constraints affecting the environment or international relations.

The social risks of AI with respect to these requirements are significant. They cover a broad spectrum, from biases in decision support systems (e.g., [7,10]), to fake news, behavior manipulation and debate steering [13]. They include political risks that can be a threat to democracy [6] and human rights [9], as well as risks to economy (implicit price cartels [4], instability of high frequency trading [11]) and to employment [1]. AI in enhanced or even autonomous lethal weapons and military systems threatens peace, it raises strong ethical concerns, e.g., as expressed in a call to a ban on autonomous weapons [19].

4 Worldwide initiatives on the societal impact of AI

Many initiatives, studies and working groups have been launched in order to assess the impacts of AI applications. There are also a few meta-studies that analyze and compare these initiatives. In this section, we briefly look at four transnational initiatives backed by major organisations that may have a significant impact on the development and use of AI, and we discuss two relevant meta-studies.

The Partnership on AI. This partnership was created by six companies, Apple, Amazon, Google, Facebook, IBM, and Microsoft, and announced during the Future of Artificial Intelligence conference in 2016. It was subsequently extended into a multi-stakeholder organization which now gathers 100 partners from 13 countries [32]. Its objectives are "to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society". Since its inception, the Partnership on AI published a few

reports, the last one being a position paper on the undesirable use of a specific criminal risk assessment tool in the COVID-19 crisis.

UNESCO Initiatives. In 2017, the World Commission on the Ethics of Scientific Knowledge and Technology of UNESCO mandated a working group to develop a study on the ethics of AI. This led to the publishing in 2019 of a Preliminary Study on the Ethics of AI [41]. This study has a broader scope than other similar document as it addresses UNESCO priority issues such as education, science, culture, peace and the development of AI in less-favored countries. It concludes with a list of eleven principles to be included in the requirements for AI applications, such as, human rights, inclusiveness, democracy, sustainability, quality of life in addition to the usual demandes on transparency, explainability, and accountability. Following this report, UNESCO created an *ad hoc* expert group of 24 specialists from 24 different countries and backgrounds to develop recommendations on the ethics of AI; the outcome of its work is still pending.

The European Commission's HLEG. The High Level Expert Group on AI of the European Commission is among the noticeable international efforts on the societal impact of AI. Initially composed of 52 multi-disciplinary experts, it started its work in 2018 and published its first report in December of the same year [18]. The report highlights three characteristics that should be met during the lifecycle of an AI system in order to be trustworthy:"it should be lawful, complying with all applicable laws and regulations; it should be ethical, ensuring adherence to ethical principles and values; and it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm". Four ethical principles are stressed: human autonomy; prevention of harm; fairness; explainability. The report makes recommendations for technical and non-technical methods to achieve seven requirements (human agency and oversight; technical robustness; etc.).

A period of pilot implementations of the guidelines followed this report, its results have not yet been published. Meanwhile, the European Commission released a White Paper on AI [42], which refers to the ethics recommendations of the HLEG.

The OECD's Expert Group and Observatory. OECD created an AI Group of Expert (AIGO) in September 2018, within its Committee on Digital Economy Policy, composed of approximately 50 delegates from OECD countries, with invited experts and other contributors in subgroups. The AIGO published a report [40], which makes recommendations on national policies and sets a few "principles for responsible stewardship of trustworthy AI", similar to those of other organisations, such as

- Inclusive and sustainable growth and well-being,
- Human-centered values and fairness,
- Transparency and explainability,
- Robustness and safety,

- Accountability.

The OECD's initiatives are pursued within a Network of Experts in AI, established in February 2020, as well as an Observatory on AI [30].

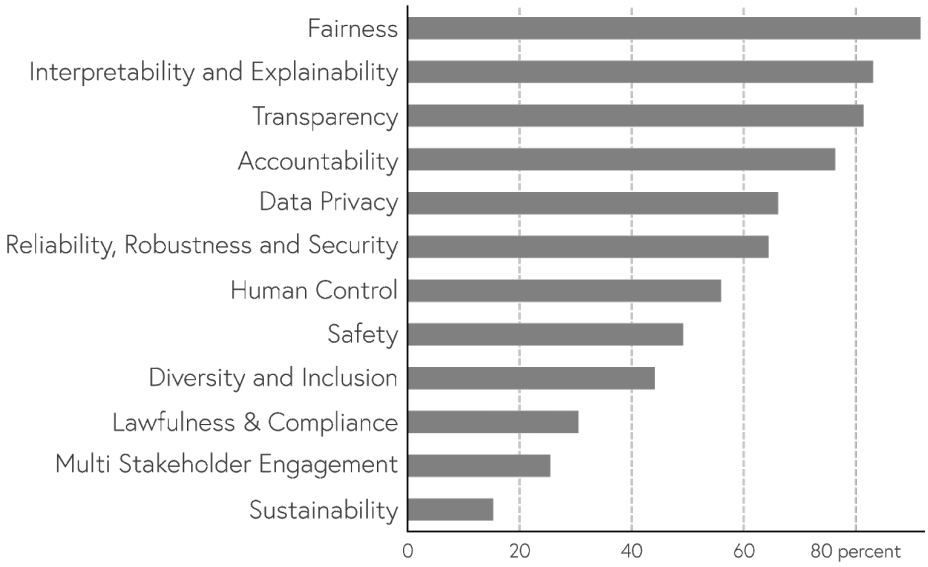

Fig. 1. Ethical AI Challenges identified across 59 documents (from [8], p.149).

Meta-studies: research devoted to analyzing and comparing diverse initiatives. The general AI principles discussed in 74 document are analyzed in [12]. The principles are grouped into ten categories (e.g., fairness, transparency, privacy, collaboration, etc.); the analyzed documents were published between 2017 and 2019 by various organisations. The corresponding website gives access to a 2D-table with links to referred documents for each analyzed category, for example:

- for the category "Fairness", the *Beijing AI Principles* contains the following: "making the system as fair as possible, reducing possible discrimination and biases, improving its transparency, explainability and predictability, and making the system more traceable, auditable and accountable".
- for the category "Privacy", the *Montreal AI Declaration* states that "Every person must be able to exercise extensive control over their personal data, especially when it comes to its collection, use, and dissemination."

Another meta-study analyzes and maps 36 documents from government, companies, and others groups related to AI ethics [3]. The map is designed along eight dimensions: safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility,

promotion of human values, international human rights. It allows for convenient comparisons over these dimensions between the documents. The final version of this analysis shows that most AI ethics documents address all eight key themes, showing a global convergence on the issues currently of concern to society.

Finally, let us go back to the AI Index [8] has been monitoring the advancement of AI over several aspects: how science and technology are progressing; how companies are investing; what is the employment situation in AI; how different countries are placed in the global competition, etc. In its 2019 report, the Index also covers 59 documents from associations, government, companies, and think tanks about ethical AI principles. It summarizes the main topics addressed; the most popular being fairness, interpretability and explainability, transparency, accountability, and data privacy (see Figure 1).

5 Outline of the book

This book develops the issues discussed at the Global Forum on AI for Humanity. Each chapter synthesizes and puts into perspective the talks and debates presented either at a plenary session (for chapters 2 to 10, and 15) or a workshop (for chapters 11 to 14) of the Forum.

In chapter 2, Raja Chatila and colleagues discuss the motivations for trustworthy AI. Human interactions with devices and systems, and social interactions are increasingly mediated through AI. This entails strong requirements to ensure trust in critical AI applications, e.g., in health or transportation systems. Techniques and regulations for the explainability, certification and auditing of AI tools need to be developed. The final part of the chapter examines conditions and methods for the production of provably beneficial AI systems.

In chapter 3, Sylvie Delacroix and colleagues look at ethical, political and legal issues with Data governance. The loop from data to information, to knowledge, action and more data collection has been further automated and improved, leading to stronger impacts, already effective or potential. It is of critical importance to clarify the mutual dependence of bottom-up empowerment structures and top-down rules for the social governance of personal data, metadata, and aggregated data. The chapter ends by exploring the role of data trusts for such purposes.

Yuko Harayama, Michela Milano and colleagues examine in chapter 4 the impact of AI on the future of work. The effectiveness of AI in the mechanization of complex physical and cognitive task has strong economic impacts, as well as social disruptive capabilities, given in particular its rapid progress. Proactive measures may be needed. This requires a good understanding of the likely effects of AI on the main economic channels and the transformation of work. The chapter presents complementary views on economy, job quality and policies as discussed at the Global Forum.

Rebecca Finlay and Hideaki Takeda report in chapter 5 about the delegation of decisions to machines. Delegating simple daily life or complex professional decisions to a computerized personal assistant, to a digital twin, can amplify

our capabilities or be a source of alienation. The requirements to circumvent the latter include in particular intelligible procedures, articulate and explicit explanations, permanent alignment of the machine's assessment functions with our criteria, as well as anticipation of and provision for an effective transfer of control back to human, when desirable.

In chapter 6 Françoise Fogelman-Soulié, Laurence Devillers and Ricardo Baeza-Yates address the subject of AI & Human values such as equity, protection against biases and fairness, with a specific focus on nudging and feedback loop effects. Automated or computer aided decisions can be unfair, because of possibly unintended biases in algorithms or in training data. What technical and operational measures can be needed to ensure that AI systems comply with essential human values, that their use is socially acceptable, and possibly desirable for strengthening social bounds.

Chapter 7, coordinated by Paolo Traverso addresses important core AI scientific and technological challenges: understanding the inner mechanisms of deep neural networks; optimising the neural networks architectures; moving to explainable and auditable AI in order to augment trust in these systems; and attempting to solve the talent bottleneck in modern artificial intelligence by using automated machine learning. The field of AI is rich of technical and scientific challenges, as can be seen from the examples given in this chapter.

In chapter 8, Jocelyn Maclure and Stuart Russell consider some of the major challenges for developing inclusive and equitable education, improving healthcare, advancing scientific knowledge and preserving the planet. They examine how properly designed AI systems can help address some of the United Nations SDGs. They discuss the conditions required to bring into play AI for these challenges. They underline in particular that neither neither pure knowledge-based approaches nor pure machine learning can solve the global challenges outlined in the chapter; hybrid approaches are needed.

In chapter 9, Carlo Casonato reflects on legal and constitutional issues raised by AI. Taking many examples from real-world usage of AI, mainly in justice, health and medicine, Casonato puts the different viewpoints expressed in the previous chapters into a new perspective, regarding regulations, democracy, anthropology and human rights. The chapter ends with a proposal for a set of new (or renewed) human rights, in order to achieve a balanced and constitutionally oriented framework for specific rights for a human-centered deployment of AI systems.

The question of ethical charters for AI is discussed in chapter 10 by Lyse Langlois and Catherine Régis. Looking at the current ethical charters landscape which has flourished extensively in the last years, the chapter examines the fundamentals of ethics and discusses their relations with law and regulations. It concludes with remarks on the appropriateness of GPAI, UN and UNESCO to take the lead in international regulatory efforts towards globally accepted ethics charters for AI.

Continuing on ethical issues related to AI, Vanessa Nurock and colleagues propose in chapter 11 an in-depth analysis of the notion of "ethics by design", as

compared to other framing such as, for example, privacy by design, or responsible innovation. The chapter examines current approaches for applying ethics to AI and concludes with guidelines for an ethics by design demanding to answer four questions on "care".

AI with respect to humanities and social sciences is discussed by Alexandre Gefen in chapter 12 from two perspectives: as an important topic of investigation and as a new mean for research. The questions about AI and its human and social consequences are invading the public sphere through the multiple issues of acceptability, privacy protection or economic impact, requiring the expertise and strong involvement of every area of Humanities and Social Sciences. AI offers also new tools to social sciences, humanities and arts, including massive data extraction, processing, machine learning and wide network analysis.

In chapter 13 Andreas Dengel and Laurence Devillers report on the state of the art of Human-Machine Co-Creation, Co-Learning and Co- Adaptation, and discuss how to anticipate corresponding ethical risks. Human ambiguous relationships with symbiotic or autonomous machines raise numerous ethical problems. Augmented intelligence and superintelligent AI are main topics for the future of human society. The robotic simulation has the virtue of questioning the nature of our own intelligence. Capturing, transmitting and mimicking our feelings will open up new applications in health, education, transport and entertainment.

Chapter 14, by Nicolas Miailhe and colleagues, is devoted to "AI Commons", a global non-profit initiative which aims to democratize responsible adoption and deployment of AI solutions for social good applications addressing the seventeen UN SDGs. This project brings together a wide range of stakeholders around innovative and holistic problem "identification-to-solution" frameworks and protocols. Its ultimate objectives are to pool critical AI capabilities (data, algorithms, domain specific knowledge, talent, tools and models, computing power and storage) into an open and collaborative platform that can be used to scale up the use of *AI for Everyone*.

Finally, Pekka Ala-Pietilä and Nathalie Smuha conclude the book with a framework for global cooperation on AI and its governance. This is certainly an essential issue in a critical period for AI. The chapter clarifies why such a governance is needed jointly with international cooperation. It lists the main areas for which international cooperation should be prioritized, with respect the socio-technical environment of AI in a transversal manner, as well as with respect to the socio-technical environments of data and digital infrastructure, these two dimensions being are tightly coupled. It concludes assessing how global cooperation should be organized, stressing the need to balance speed, holism and contextualism, and providing a number of guiding principles that can inform the process of global cooperation initiatives on AI and its governance.

This book collects views from leading experts on AI and its human, ethical, social, and legal implications. Each chapter is self-contained and addresses a specific set of issues, with links to other chapters. To further guide the reader

about the organization of the covered topics, a possible clustering (with overlaps) of these "Reflections on Artificial Intelligence for Humanity" is the following:

- chapters 7, 13 and 14 are mainly devoted to technological and scientific challenges with AI and at some developments designed to address them;
- chapters 5, 6, 10, and 11 focus on different ethical issues associated with AI;
- chapters 2, 3, 4, 5, and 6 cover the social impacts of AI at the workplace and in personal applications;
- chapters 7, 8, 12 and 13 discuss the possible benefits and risks of AI in several area such as health, justice, justice, education, humanities and social sciences;
- chapters 3, 9, 14, and 15 addresses legal and organizational issues raised by AI.

6 What's next: an opening for GPAI

The GFAIH forum was a step in the preparation of GPAI, the Global Partnership on Artificial Intelligence. Launched by France and Canada on the sidelines of the Canadian presidency of the G7, this initiative aims to organize an independent global expertise on the ethical regulation of AI.

Following the Franco-Canadian Declaration on AI of June 7, 2018, and the production of a mandate for an international group of experts in artificial intelligence (G2IA), France and Canada jointly decided to include the GPAI on the agenda of the French presidency of the G7, in order to place this initiative in a multilateral framework. The G7 digital ministerial meeting on May 2019 helped secure the support of Germany, Italy, Japan, the United Kingdom, New Zealand, India and the European Union for the launch of the GPAI. The G7 summit in Biarritz on 24-26 August 2019 made it possible to obtain the support of the G7 States for this initiative, renamed the Global Partnership on AI (GPIA) and of the four invited countries (India, Chile, South Africa and Australia) and New Zealand, giving a strong political mandate to the initiative thanks to the Biarritz Strategy for an open, free and secure digital transformation. Canada and France also agreed on a tripartite structure for the PMIA, consisting of two centres of expertise in Paris and Montreal and a secretariat hosted at the OECD in Paris to avoid work duplication and maximize synergies, while maintaining a strict independence of the experts' work. A major step was taken on June 15th, 2020, when fifteen countries - among which all G7 - members simultaneously announced the launch of the Partnership and their commitment to make it a success.

This initiative will permit an upstream dialogue between the best scientists and experts and public decision-makers, which is a key condition for designing effective responses and recommendations necessary to cope with current and future challenges faced by our societies. The GPAI will produce, on a comprehensive, objective, open and transparent basis, analyses of scientific, technical and socio-economic information relevant to understanding the impacts of AI, encouraging its responsible development, and mitigating its risks. This work will

follow a project-based approach, with a strong technical dimension. Complementary to other approaches such as the four initiatives mentioned above, the work of GPAI will be mostly driven by science and will include representative experimentation to support its recommendations.

Four working groups have been initially identified in GPAI on, respectively, the issues of responsible AI, data governance, future of work, innovation and commercialization. A fifth working group on the response to the current pandemic situation and to other possible pandemics has been created as a subgroup of "Responsible AI". There is a clear link between the topics of the Global forum, the chapters of this book and the four main working groups of GPAI: the "data governance" and "future of work" themes are direct matches, whereas several chapters contribute to "Responsible AI" (chapters 2, 5, 6, 7, 11 in particular) and to "Innovation and commercialization" (chapters 2, 7, 8, 15 in particular). The first plenary meeting of GPAI experts took place online in early December 2020,[6] the second will take place in Paris in 2021.

It has become crucial to consolidate democracies at a time when technological competition is intensifying, while the risks of Internet fragmentation and AI social impacts are deepening. GPAI aspires to bring together like-minded countries, sharing the same democratic values in order to promote a socially responsible, ethical vision of AI.

References

1. Arntz, M., Gregory, T., Zierahn, U.: The Risk of Automation for Jobs in OECD Countries. OECD Social, Employment and Migration Working Papers (189) (2016). https://doi.org/https://doi.org/https://doi.org/10.1787/5jlz9h56dvq7-en, https://www.oecd-ilibrary.org/content/paper/5jlz9h56dvq7-en
2. Bullock, J., Luccioni, A., Pham, K.H., Lam, C.S.N., Luengo-Oroz, M.: "mapping the landscape of artificial intelligence applications against covid-19". arXiv (2020), https://arxiv.org/abs/2003.11336
3. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M.: "principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for a"i. Tech. Rep. 2020, Berkman Klein Center Research Publication (2020), http://dx.doi.org/10.2139/ssrn.3518482
4. Gal, M.S.: Illegal Pricing Algorithms. Communications of the ACM **62**(1) (January 2019)
5. Muscettola, N., Nayak, P.P., Pell, B., Williams, B.C.: Remote Agent: to boldly go where no AI system has gone before. Artificial Intelligence **103**, 5–47 (1998)
6. Nemitz, P.: Constitutional democracy and technology in the age of artificial intelligence. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**(2133) (Oct 2018)
7. O'Neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Random House (2016)
8. Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., Niebles, J.C.: "the ai index 2019 annual repor"t. Tech. rep., Stanford University (2019), http://aiindex.org

[6] See http://gpai.ai

9. Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., Kim, L.Y.: Artificial Intelligence & Human Rights: Opportunities & Risks. SSRN (September 2018)
10. Skeem, J.L., Lowenkamp, C.: Risk, Race, & Recidivism: Predictive Bias and Disparate Impact. SSRN (2016)
11. Sornette, D., von der Becke, S.: Crashes and High Frequency Trading. SSRN (August 2011)
12. Zeng, Y., Lu, E., Huangfu, C.: "linking artificial intelligence principles". arXiv (2018), https://arxiv.org/abs/1812.04814v1
13. Zuboff, S.: The Age of Surveillance Capitalism. PublicAffairs (2019)
14. Ai for good foundation. https://ai4good.org/about/
15. Ai now institute. https://ainowinstitute.org/
16. Ai4people. http://www.eismd.eu/ai4people/
17. Ethically aligned design. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf
18. Eu high level expert group on ai. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
19. The future of life insitute. https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1
20. The global challenges foundation. https://globalchallenges.org/about/the-global-challenges-foundation/
21. Human-centered ai. http://hai.stanford.edu/
22. Humane ai. http://www.humane-ai.eu/
23. Image net. http://image-net.org/
24. Internationa telecommunication union. https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTS.V1I1-2017-1-PDF-E.pdf
25. International observatory on the societal impacts of ai. https://observatoire-ia.ulaval.ca/
26. International planning competition. http://icaps-conference.org/index.php/Main/Competitions
27. Kitkit school. http://kitkitschool.com/
28. Mckinsey global institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact
29. Nlp competitions. https://codalab-worksheets.readthedocs.io/en/latest/Competitions/#list-of-competitions
30. Oecd ai policy observatory. http://www.oecd.org/going-digital/ai/oecd-initiatives-on-ai.htm
31. Onetab. https://onebillion.org/
32. Partnership on ai. https://www.partnershiponai.org/research-lander/
33. Robocup. https://www.robocup.org/
34. Sat competitions. http://satcompetition.org/
35. Squad explorer. https://rajpurkar.github.io/SQuAD-explorer/
36. Uk center for the governance of ai. https://www.fhi.ox.ac.uk/governance-ai-program/
37. Un global pulse. https://www.unglobalpulse.org/
38. Un sustainable development goals. https://sustainabledevelopment.un.org/?menu=1300
39. Unesco. https://en.unesco.org/artificial-intelligence
40. Deliberations of the expert group on artificial intelligence at the oecd. https://www.oecd-ilibrary.org/ (2019)

41. Preliminary study on the ethics of artificial intelligence. https://unesdoc.unesco.org/ (2019)
42. Eu white paper on ai. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (2020)