Statement on AI futures
by
Francesca Rossi, IBM
NAIAC meeting, August 3rd, 2023

I would like to thank the NAIAC AI Futures Working Group for this opportunity to address the whole NAIAC, the US Senate and White House delegations, and all those who are joining today from remote.

Let me first briefly introduce myself. My background is in Computer Science and I have contributed to advance AI research for about 35 years, 25 of which in academia in Europe and then at IBM in the US. At IBM, I am also the AI ethics global leader, which means that I help shape all the company does around AI ethics, from high-level principles to their operationalization within the company's business units, including education, playbooks, risk assessment frameworks, and governance, and within the various multi-stakeholder partnerships we founded and joined. I am a board member of the Partnership on AI and a Steering Committee member of the Global Partnership on AI, where I am one of the 3 experts appointed by the US government. I am also co-chairing the OECD Expert group on AI Future, and I have been a member of the European Commission High Level Expert Group on AI, that was used as the basis for the EU AI Act. I am also the current president of AAAI, the international association of AI researchers.

As we all know, AI is a powerful technology with amazing capabilities and potential but also significant usage risks.
AI is not new and we have been using it for many years in our life. We started with narrow, explicit, so-called symbolic approaches, which worked well in many scenarios but showed brittleness in others, and we moved on to add data-driven flexible techniques, based on machine learning. This gave AI the ability to interpret the environment through data and to learn how to behave in it. But it also introduced issues related to data privacy and governance, bias, lack of explainability, and robustness.
The recent advances in machine learning provided the additional ability to generate content and to master human language. Generative AI supports the creation of so-called foundation models, that is, very general AI models that can be used as a basis to build quickly other downstream AI models designed to solve specific problems. All this greatly expands the range of useful applications of AI, with potential to accelerate scientific discoveries and economic growth, increase societal well-being, and solve crucial global problems, such as in health and climate. It also allows machines to interact with us in a much more natural way. But it also expands the already mentioned issues and introduces additional risks related to harmful content generation and the spread of misinformation. It also may have an increased impact on jobs, on education, on creative activities, and on democracy. Such a powerful technology can also be misused by bad actors, and even non-educated good actors may use it in inappropriate ways generating undesired and harmful behavior.

How to take the best of AI and mitigate the risks?

To capture AI's benefits, we need to create an ecosystem of trust in the technology and its uses. This should be done by designing and adopting clear risk-based policies and regulations that associate risk to AI uses rather than to the AI technology, and impose guardrails and obligations to the various actors and providers tailored to the role and capabilities they play in the complex AI lifecycle. As an example, the risk of bias (and therefore of an unfair treatment of some groups compared to others) cannot be adequately identified and tested in upstream AI models, but only on downstream AI systems for which we know the purpose and the deployment scenario. High-risk AI applications should be subject to higher scrutiny and guardrails for providers, deployers, and users.
Risky scenarios existed even before AI was used, and there were regulations to address their risks. Rather than defining new general and broad AI regulations, we should check what needs to be added of modified in the various sector-based regulations because of the adoption of AI solutions. This would also provide a more fine grained analysis of the risks and their mitigation policies.
This approach is what at IBM we call precision regulation.

How to identify the risks?
This can be done only by using a multi-stakeholder approach, where all the societal actors are included: AI builders, AI providers, AI users, AI researchers, civil society organizations, policy makers, social scientists. AI experts alone cannot correctly identify the concerns, the impact, and the implications of a very pervasive use of AI in our society. It may seem slower, but it is rather faster, if the goal is to use technology to accelerate human progress. At IBM we joined several multi-stakeholder AI initiatives. For example, in 2016 we founded the Partnership on AI, with the mission to define best practices to make Ai beneficial for people and society. PAI now has 100 partners, of which only 20 are companies, and all others are academic institutions and civil society organizations. Over the years, it has delivered impactful guidelines, recommendations, and best practices around all major AI ethics issues, the most recent being for the use of synthetic media and of large language models.

Can regulation solve all the problems?
Regulation is needed, but we need contributions from all other actors in the AI ecosystem, also because technology evolves much more rapidly than the legal system. AI companies should define internal AI ethics frameworks to operationalize high-level principles into concrete actions in all business units, supported by powerful governance bodies, education activities, risk assessment processes, software tools, and AI development guidelines. Transparency should be central in communicating the real capabilities, limits, and appropriate uses of an AI system.
At IBM, we have a company-wide AI ethics framework that covers all these dimensions and engages all business units in making sure we build, use, and deploy AI responsibly. This includes a significant investment in AI governance to embed ethics-by-design in the AI lifecycle. The IBM watsonx platform, recently released, includes a governance component to help clients build the appropriate tests and checks within all steps of developing an AI model and solution.

What are now called “voluntary commitments”, IBM has been making them for a long time, and also concretely implementing them, both internally and with our partners.
Companies and governments are not the only actors in the AI ecosystem with an important role to play. Users should be educated about capabilities and risks and should use AI in the most appropriate and intended ways. Standards and certification bodies also have an important role to play, in harmonizing and providing certainty over terms, methods, and processes.

Can AI research help?
AI research, informed by multi-disciplinary and multi-stakeholder consultation, is an essential component to address and mitigate AI issues. Some of these issues are related to current AI limitations, and AI research can overcome some of these limitations. Moreover, AI research can also identify new lines of work and techniques to equip AI models with the ability to reason more similarly to humans and to behave in a way that is aligned to human values. Value alignment is a crucial open problem in current AI research, where the combination of data-driven machine learning and explicit knowledge and logic-based approaches is being increasingly investigated.
The IBM AI research division continuously delivers innovation in AI capabilities, AI risk assessment and mitigation toolkits, and principled aligned AI, with the goal to support the delivery of trustworthy AI to enterprise applications.
It is important to increase support in AI research. The NAIRR is a crucial initiative that should fund and facilitate multi-disciplinary AI research efforts devoted to advance AI capabilities via a variety of AI techniques, also inspired by cognitive theories of human reasoning, while addressing the AI value alignment problem.
AI research needs resources, data, computing power, and transparent models. This can be best supported by open-source innovation that democratizes AI and makes it more accessible, and also allows the many talents in academia to play a role in advancing it. IBM has been long supporting open-source technology with the acquisition of Red Hat and the recent partnership with Hugging Face on large language models.

What is the role of AI in our society?
AI can help us achieve our vision of the future, where human values are protected and supported. The UN Sustainable Development Goals define such a vision and give us a definition of where we want to go with the help of powerful technologies like AI. Risks to impact negatively on human values should be identified, measured, and mitigated, but without technology we cannot achieve these goals. As technology evolves, we also need to evolve our methods, guidelines, and guardrails to mitigate the additional risks, but we also need to use the technology to generate trajectories that lead us closer to the ideal vision of the future, not farther away. This requires a combination of reactive measures and proactive strategies, to build responsible and trustworthy AI and to use it for good purposes.