Delivering a Good AI Society
Governance of AI and AI for Governance

Nour El Kadri

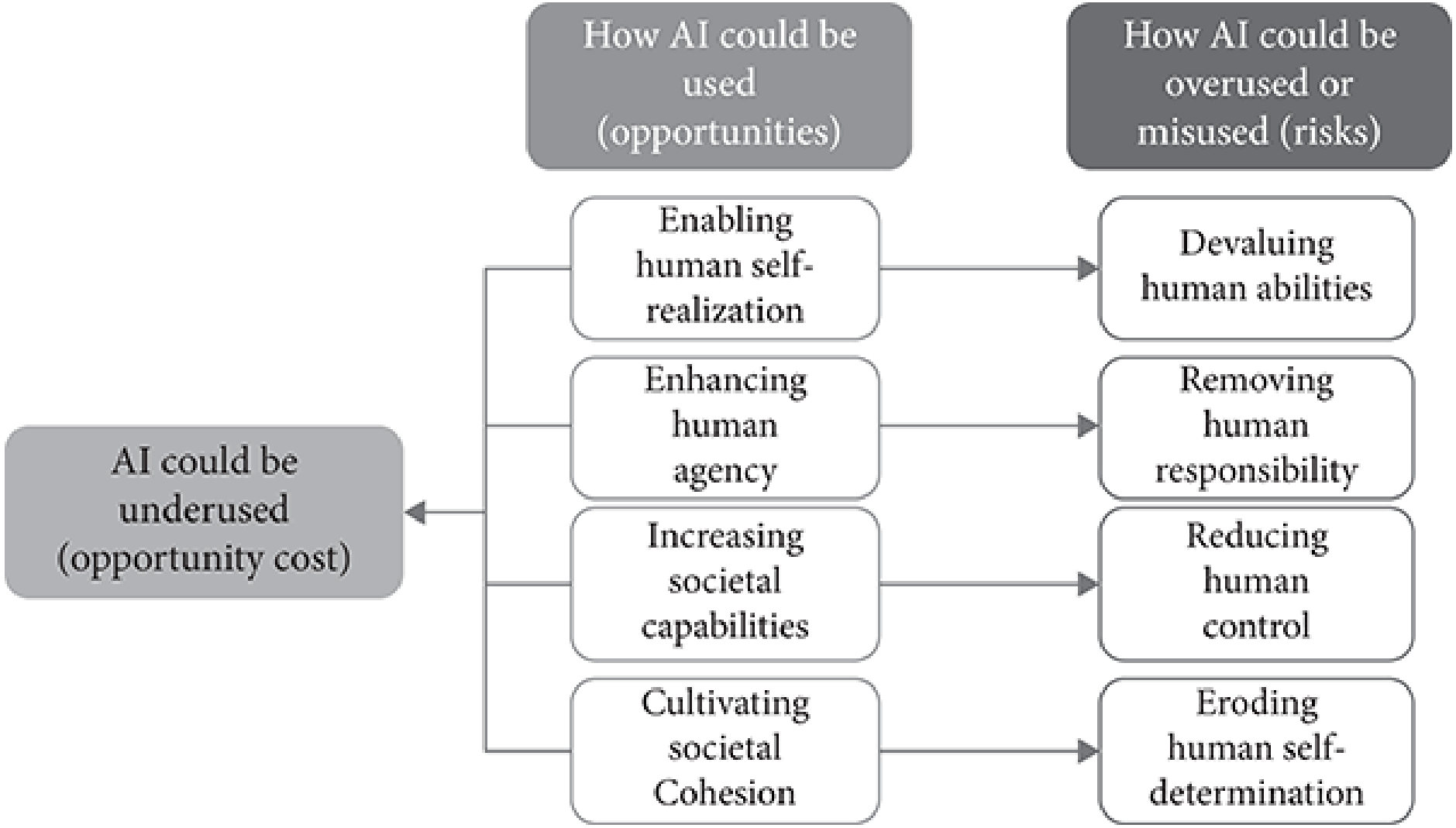
ISACA AGM Talk

Philosophical Anthropology

Our understanding of human dignity and flourishing:

1. autonomous self-realization, or who we can become;
2. human agency, or what we can do;
3. individual and societal capabilities, or what we can achieve; and
4. societal cohesion, or how we can interact with each other and the world.

Opportunity Costs and Risks



Who We Can Become: Enabling Human Self-Realization without Devaluing Human Abilities

- AI may enable self-realization, that is, the ability for people to flourish in terms of their own characteristics, interests, potential abilities or skills, aspirations, and life projects.
- Much like inventions such as the washing machine, which liberated people (particularly women) from the drudgery of domestic work, the ‘smart’ automation of other mundane aspects of life may free up yet more time for cultural, intellectual, and social pursuits, and more interesting and rewarding work.
- More AI could easily mean more human life spent more intelligently.
- The risk in this case is not the obsolescence of some old skills and the emergence of new ones per se, but the pace at which this is happening and the unequal distributions of the costs and benefits that result.

What We Can Do: Enhancing Human Agency without Removing Human Responsibility

- AI is providing a growing reservoir of ‘smart agency’. Put at the service of human intelligence, such a resource can hugely enhance human agency.
- We can do more, better, and faster, thanks to the support provided by AI. In this sense of ‘[human] augmented intelligence’, AI could be compared to the impact that engines have had on our lives.
- The larger the number of people who will enjoy the opportunities and benefits of such a reservoir of smart agency ‘on tap’, the better our societies will be.
- Responsibility is therefore essential, in view of what sort of AI we develop, how we use it, and whether we share with everyone its advantages and benefits.
- AI offers the opportunity to improve and multiply the possibilities for human agency.

What We Can Achieve: Increasing Societal Capabilities without Reducing Human Control

- AI offers many opportunities for improving and augmenting the capabilities of individuals and society at large.
- Whether by preventing and curing diseases or optimizing transportation and logistics, the use of AI technologies presents countless possibilities for reinventing society by radically enhancing what humans are collectively capable of.
- More AI may support better collaboration, and hence more ambitious goals.
- Human intelligence augmented by AI could find new solutions to old and new problems ranging from a fairer or more efficient distribution of resources to a more sustainable approach to consumption.
- Precisely because such technologies have the potential to be so powerful and disruptive, they also introduce proportionate risks.

How We Can Interact: Cultivating Societal Cohesion without Eroding Human Self-Determination

- From climate change and antimicrobial resistance to nuclear proliferation, wars, and fundamentalism, global problems increasingly involve high degrees of coordination complexity.
- This means they can be tackled successfully only if all stakeholders co-design and co-own the solutions and cooperate to bring them about.
- AI can hugely help to deal with such coordination complexity with its data-intensive, algorithmic-driven solutions, supporting more societal cohesion and collaboration.

Twenty Recommendations for a Good AI Society

Taken together along with their corresponding challenges, the four opportunities outlined above paint a mixed picture about the impact of AI on society and the people in it, and the overall environments they share.

Accepting the presence of trade-offs and seizing the opportunities while working to anticipate, avoid, or minimize the risks head-on will improve the prospect for AI technologies to promote human dignity and flourishing.

Ensuring that the outcomes of AI are socially preferable (equitable) depends on resolving the tension between incorporating the benefits and mitigating the potential harms of AI—in short, simultaneously avoiding the misuse and underuse of these technologies.

Good AI - Principles - Practices

- The assumption is that, to create a Good AI Society, the ethical principles should be embedded in the default practices of AI.
- It is especially important that AI be explicable as explicability is a critical tool for building public trust in, and understanding of, the technology.
- Creating a Good AI Society requires a multi-stakeholder approach. This is the most effective way to ensure that AI will serve the needs of society by enabling developers, users, and bto all be on board, collaborating from the outset.
- Inevitably, different cultural frameworks inform attitudes to new technology.

No matter where we live in the world, we should all be committed to the development of AI technologies in a way that secures people's trust, serves the public interest, strengthens shared social responsibility, and supports the environment.

Recommendations: 1 & 2

1. Assess the capacity of existing institutions, such as national civil courts, to redress the mistakes made or harms inflicted by AI systems. This assessment should evaluate the presence of sustainable, majority-agreed foundations for liability from the design stage onwards to reduce negligence and conflicts (see also Recommendation 5).
2. Assess which tasks and decision-making functionalities should not be delegated to AI systems using participatory mechanisms to ensure alignment with societal values and understanding of public opinion. This assessment should consider existing legislation and be supported by ongoing dialogue between all stakeholders (including government, industry, and civil society), to debate how AI will impact society (in concert with Recommendation 17).

Recommendation: 3

3. Assess whether current regulations are sufficiently grounded in ethics to provide a legislative framework that can keep pace with technological developments. This may include a framework of key principles that would be applicable to urgent and/or unanticipated problems.

Recommendation: 4

4. Develop a framework to enhance the explicability of AI systems that make socially significant decisions. Central to this framework is the ability for individuals to obtain a factual, direct, and clear explanation of the decision-making process, especially in the event of unwanted consequences. This is likely to require the development of frameworks specific to different industries; professional associations should be involved in this process alongside experts in science, business, law, and ethics.

Recommendation: 5

5. Develop appropriate legal procedures and improve the digital infrastructure of the justice system to permit the scrutiny of algorithmic decisions in court. This is likely to include the creation of a framework for AI explainability (as indicated in Recommendation 4) specific to the legal system.

Recommendation: 6

6. Develop auditing mechanisms for AI systems to identify unwanted consequences, such as unfair bias. Auditing should also (perhaps in cooperation with the insurance sector) include a solidarity mechanism to deal with severe risks in AI-intensive sectors. Those risks could be mitigated by multi-stakeholder mechanisms upstream.

Recommendation: 7

7. Develop a redress process or mechanism to remedy or compensate for a wrong or grievance caused by AI. To foster public trust in AI, society needs a widely accessible and reliable mechanism of redress for harms inflicted, costs incurred, or other grievances caused by the technology.

Such a mechanism will necessarily involve a clear and comprehensive allocation of accountability to humans and/or organizations. The development of this process must follow from the assessment of existing capacity outlined in Recommendation 1. If a lack of capacity is identified, additional institutional solutions should be developed at national levels to enable people to seek redress. Such solutions could include:

- an 'AI ombudsperson' to ensure the auditing of allegedly unfair or inequitable uses of AI;
- a guided process for registering a complaint akin to making a Freedom of Information request; and
- the development of liability insurance mechanisms that would be required as an obligatory accompaniment of specific classes of AI offerings in every jurisdiction and other markets. This would ensure that the relative reliability of AI-powered artefacts, especially in robotics, is mirrored in insurance pricing and therefore in the market prices of competing products.

Recommendation: 8

8. Develop agreed-upon metrics for the trustworthiness of AI products and services. These metrics could be the responsibility of either a new organization or a suitable existing one. They would serve as the basis for a system that enables the user-driven benchmarking of all marketed AI offerings.

In this way, an index for trustworthy AI can be developed and signaled in addition to a product's price. This 'trust comparison index' for AI would improve public understanding and engender competitivenesss around the development of safer, more socially beneficial AI (e.g. 'IwantgreatAI.org').

In the longer term, such a system could form the basis for a broader system of certification for deserving products and services—one that is administered by the organization noted here, and/or by the oversight agency proposed in Recommendation 9. The organization could also support the development of codes of conduct (see Recommendation 18). Furthermore, those who own or operate inputs to AI systems and profit from it could be tasked with funding and/or helping to develop AI literacy programs for consumers, in their own best interest.

Recommendations

9. Develop a new oversight agency responsible for the protection of public welfare through the scientific evaluation and supervision of AI products, software, systems, or services.
10. Develop a country-wide observatory for AI. The mission of the observatory would be to watch developments, provide a forum to nurture debate and consensus, provide a repository for AI literature and software (including concepts and links to available literature), and issue step-by-step recommendations and guidelines for action.

Recommendations: 11 & 12

11. Develop legal instruments and contractual templates to lay the foundation for a smooth and rewarding human–machine collaboration in the work environment. Shaping the narrative on the ‘Future of Work’ is instrumental to winning ‘hearts and minds’. Championing ‘inclusive innovation’, and efforts to smooth the transition to new kinds of jobs an AI Adjustment Fund could be set up to help flatten the curve.
12. Incentivize financially, the development and use of AI technologies within the country that are socially preferable (not merely acceptable) and environmentally friendly (not merely sustainable, but actually favourable to the environment). This will include the elaboration of methodologies that can help assess whether AI projects are socially preferable and environmentally friendly. In this vein, adopting a ‘challenge approach’ (see the Defense Advanced Research Projects Agency, DARPA, challenges) may encourage creativity and promote competition in the development of specific AI solutions that are ethically sound and in the interest of the common good.

Recommendations: 13 & 14

13. Incentivize financially a sustained, increased, and coherent country-wide research effort tailored to the specific features of AI as a scientific field of investigation. This should involve a clear mission to advance AI4SG to counterbalance AI trends with less focus on social opportunities.
14. Incentivize financially cross-disciplinary and cross-sectoral cooperation and debate concerning the intersections between technology, social issues, legal studies, and ethics. Debates about technological challenges may lag behind the actual technical progress but if they are strategically informed by a diverse multi-stakeholder group, they may steer and support technological innovation in the right direction. Ethics should help seize opportunities and cope with challenges, not simply describe them. It is thus essential that diversity infuses the design and development of AI, in terms of gender, class, ethnicity, discipline, and other pertinent dimensions, to increase inclusivity, toleration, and the richness of ideas and perspectives.

Recommendations: 15 & 16

15. Incentivize financially the inclusion of ethical, legal, and social considerations in AI research projects. In parallel, create incentives for regular reviews of legislation to test the extent to which it fosters socially positive innovation. Taken together, these two measures will help ensure that AI technology has ethics at its heart and that policy is oriented towards innovation.
16. Incentivize financially the development and use of lawfully deregulated special zones within the country. These zones should be used for the empirical testing and development of AI systems. They may take the form of a 'living lab' (or Tokku), building on the experience of existing 'test highways' (or Teststrecken). In addition to aligning innovation more closely with society's preferred level of risk, sandbox experiments such as these contribute to hands-on education and the promotion of accountability and acceptability at an early stage. 'Protection by design' is intrinsic to this kind of framework.

Recommendation: 17

17. Incentivize financially research about public perception and understanding of AI and its applications. Research should also focus on the implementation of structured mechanisms for public consultation to design policies and rules related to AI.

This could include the direct elicitation of public opinion via traditional research methods (such as opinion polls and focus groups), along with more experimental approaches (such as providing simulated examples of the ethical dilemmas introduced by AI systems, or experiments in social science labs). This research agenda should not serve merely to measure public opinion. It should also lead to the co-creation of policies, standards, best practices, and rules as a result.

Recommendation: 18

18. Support the development of self-regulatory codes of conduct, for both data and AI-related professions, with specific ethical duties. This would be along the lines of other socially sensitive professions, such as medical doctors or lawyers.

In other words, it would involve the attendant certification of ‘ethical AI’ through trust labels to make sure that people understand the merits of ethical AI and will therefore demand it from providers. Current attention manipulation techniques may be constrained through these self-regulating instruments.

Recommendation: 19

19. Support the capacity of corporate boards of directors to take responsibility for the ethical implications of companies' AI technologies.

This could include improved training for existing boards, for example, or the potential development of an ethics committee with internal auditing powers. It could be developed within the existing structure of both one-tier and two-tier board systems, and/or in conjunction with the development of a mandatory form of 'corporate ethical review board'. The ethical review board would be adopted by organizations developing or using AI systems. It would then evaluate initial projects and their deployment with respect to fundamental principles.

Recommendation: 20

20. Support the creation of educational curricula and public awareness activities around the societal, legal, and ethical impact of AI. This may include:
    - school curricula to support the inclusion of computer science among the other basic disciplines that are taught;
    - initiatives and qualification programmes in businesses dealing with AI technology to educate employees on the societal, legal, and ethical impact of working alongside AI;
    - a country-level recommendation to include ethics and human rights within the university degrees for data and AI scientists, as well as within other scientific and engineering curricula dealing with computational and AI systems;
    - the development of similar programmes for the public at large. These should have a special focus on those involved at each stage of management for the technology, including civil servants, politicians, and journalists;
    - engagement with wider initiatives, such as the AI for Good events hosted by the International Telecommunication Union (ITU) and NGOs working on the UN SDGs.

Conclusion: The Need for Concrete and Constructive Policies

- Humanity faces the emergence of a technology that holds much exciting promise for many aspects of human life. At the same time, it seems to pose major threats as well.
- These recommendations seek to nudge the tiller in the direction of ethically, socially, and environmentally preferable outcomes from the development, design, and deployment of AI technologies.
- The recommendations build on the set of five ethical principles for AI and on the identification of both the risks and the core opportunities of AI for society.
- They are formulated in the spirit of collaboration and in the interest of creating concrete and constructive responses to the most pressing social challenges posed by AI.

Conclusion: The Need for Concrete and Constructive Policies

- With the rapid pace of technological change, it is tempting to view the political process of contemporary liberal democracies as old-fashioned, out of step, and no longer up to the task of preserving the values and promoting the interests of society and everyone in it.

I disagree.

- The recommendations offered here, which include the creation of centres, agencies, curricula, and other infrastructure, support the case for an ambitious, inclusive, equitable, and sustainable programme of policymaking and technological innovation. This will contribute to securing the benefits and mitigating the risks of AI for all people, as well as for the world that we share.

Thank You!

Q &A

nelkadri@uottawa.ca Tel: 613 240-3181

Nour Kadri on LinkedIn and other social media.