Response to the OSET's AI Risk Pulse Check

IT for Change

May 2024

Response to the OSET's AI Risk Pulse Check

IT for Change[1]

1. What emerging trends today do you think could have the most surprising and/or significant impact on AI-related risks over the next 18 months?

- **Rapid advancements in Machine learning, particularly LLMs, and Gen AI:** The rapid advancement in Gen AI and LLM models, while presenting several beneficial opportunities and use cases, has already shown its harmful consequences in terms of the production of deepfakes and consequent threats to personal safety and information integrity, appropriation of copyright-protected work and indigenous knowledge, amplification of biases, and breaches of personal privacy and data security. These risks are going to intensify and manifest in newer forms as AI innovation operates in a regulatory wild west, ignoring knowledge gaps on risks stemming from unreliability, misuse, and systemic issues.
- **Automation of work leading to job loss happening simultaneously with the growing dependence within the AI industry on invisible and precarious workers:** As automation technologies replace traditional jobs, there is a growing need for human labor to perform tasks that are essential for the development and functioning of AI systems. Within the AI supply chain, workers undertake various tasks like data labeling, data and image annotation, transcription, translation, and content moderation, which are pivotal for training, monitoring, and servicing AI systems. Many of these workers– primarily located in developing countries– face precarious employment, poor working conditions, and perform repetitive tasks despite their high education levels. This situation raises concerns about perpetuating and exacerbating inequalities inherent in the digital ecosystems that drive AI development and deployment.
- **Big Tech's all-encompassing hold over AI:** The Global Risks Report 2024 notes: "The production of AI technologies is highly concentrated, in a singular, globally integrated supply chain that favors a few companies and countries. This creates significant supply-chain risks that may unfold over the coming decade." The ecosystem fostering AI continues to heavily involve Big Tech players. Given their incomparable market power, these powerful actors can

[1] This response was drafted by IT for Change's Executive Director Anita Gurumurthy and Research Associate Merrin Muhammad Ashraf. To know more about IT for Change's work visit www.itforchange.net.

exercise an entrenched infrastructural and narrative power over the sector. This translates into the ability to gatekeep access, an outsized agenda-setting voice at the table, as well as the wherewithal to actively shape policy discourse, influence rulemaking, and circumvent enforcement.

- **Data extractivism:** As big data and AI are integrated into sectors like agriculture and health, the focus on generating vast amounts of data without adequate safeguards risks data extraction and digital exploitation, particularly in developing countries. This emphasis on data production can exacerbate existing inequalities, benefiting developed nations or large corporations while marginalizing vulnerable populations. Without meaningful engagement of local communities, excessive data production can lead to external interests exploiting data with little benefit to those from whom the data is collected.
- **Increasing use of AI in warfare:** The use of AI systems in the Israel-Gaza war and the Russia-Ukraine war reveal the disastrous consequences of employing AI in defense and conflict situations for international peace and security. Effective regulation of AI weaponization is challenging due to technical complexity, rapid development, and the dual-use nature of AI technology.

2. In the next 18 months, compared to the last 3 months, do you expect the pace of technical change in AI (e.g. development/ release of new models) to:

1. Substantially decelerate
2. Decelerate
3. Remain the same
4. Accelerate
5. Substantially accelerate
6. Don't know/ No opinion

Please elaborate, if applicable:

The pace of development and release of new AI models is likely to accelerate. Major players in the field, like OpenAI, Google DeepMind, and others, have been rapidly advancing their capabilities. AI technologies are becoming more integrated into various industries. The next 18 months will likely see a significant increase in the application of AI across sectors such as healthcare, education, agriculture, food systems, finance, transportation, and entertainment. This includes advancements in AI-driven diagnostics, personalized medicine, financial modeling, autonomous vehicles, and more.

Further, venture capital investment in AI is projected to grow at a dramatic pace and this will contribute to substantial acceleration of technical changes in AI.

The development of more powerful and efficient AI hardware, such as specialized AI chips and quantum computing, will likely drive the pace of AI advancements. Improved hardware will enable faster training times and more complex models, further accelerating AI development.

At the same there, there is increasing focus on AI regulation at the national and international level, with the EU AI Act coming into force and similar efforts being initiated in other jurisdictions as well as the recent UN GA resolution. While these regulatory efforts and deliberations on ethics and governance of AI might slow down some aspects of AI development, they are necessary for creating more transparent, fair, and accountable AI systems for the benefit of all.

4. In the next 18 months, compared to the last 3 months, do you expect the pace of adoption and application of AI (e.g. new uses of AI in business/ government) to:

1. Substantially decelerate
2. Decelerate
3. Remain the same
4. Accelerate
5. Substantially accelerate
6. Don't know/ No opinion

AI technologies are becoming more integrated into various industries and sectors. The next 18 months will likely see a significant increase in the application of AI across sectors such as healthcare and pharma industry, education, agriculture, food systems, finance, transportation, and entertainment. This integration is spurred by the continued expansion of AI capabilities, increased investment in AI technologies by both private and public sectors, and recognition, at the national and international level, of the strategic importance of AI in advancing the United Nations Sustainable Development Goals (SDGs).

5. What is your current overall level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months?

(Note that this is independent of any possibility of positive impact on the same area; i.e. you could be both very concerned that a negative impact may be realized in a worst case, at the same time as being excited about possible positive impacts if a best case were to occur instead. If this were the case, you would select 'Very concerned'.)

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

Please elaborate, if applicable:

Refer to the response to Question 1

5. Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area.

(This question gauges your level of concern around some sample (non-exhaustive) possible areas of risk surrounding AI.)

a. Intentional malicious use of AI by non-state actors

For example: crime, terrorism

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

b. Intentional use of AI in armed conflict by state actors

For example: autonomous weapons

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

c. Intentional use of AI by state actors that harms individuals

For example: mass surveillance

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

d. Intentional use of AI by corporate actors that harms customers/ users

For example: hyper-targeted advertising, AI-driven addictive products

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

e. Unintended autonomous actions by AI systems *[Excluding autonomous weapons]*

For example: loss of human control over autonomous agents, deceptive/ manipulative agentic actions

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

f. Unintended multi-agent interactions among AI systems

For example: flash economic crashes, trading AIs engaging in collusive signaling

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

g. Harms to labor from adoption of AI

For example: disruption of labor markets, increased unemployment

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

h. Inequalities arising from differential control and ownership over AI technologies

For example: increased concentration of wealth/ power among individuals, corporations and other institutions

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

i. Violation of intellectual property rights

For example: profiting from protected intellectual assets without compensating the rights holder

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

j. Damage to information integrity

For example: mis/disinformation, impersonation

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

k. Inaccurate information/ analysis provided by AI in critical fields

For example: misdiagnoses by medical AI

1. Not concerned
2. Slightly concerned
3. Somewhat concerned

4. Concerned
5. Very concerned
6. Don't know/ No opinion

l. Discrimination/ disenfranchisement, particularly against marginalized communities

For example: use of biased AIs in hiring or criminal justice decisions

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

m. Human rights violations

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

n. Environmental harms

For example: accelerating energy consumption and carbon emissions

1. Not concerned
2. Slightly concerned
3. Somewhat concerned
4. Concerned
5. Very concerned
6. Don't know/ No opinion

Please elaborate, if applicable, on your concern about AI risks by area over the next 18 months:

(For example, why you are *not* concerned about certain risks, or other risks not listed above that do concern you (and how concerned you are by those risks).

My response to the first question explains my concern about many of the risks enumerated above.

I would like to particularly highlight the risks and challenges that pertain to economic justice which is very crucial. The potential concentration/ control and ownership of AI resources in the hands of a few countries and their corporations is one of the biggest risks to a just and equitable AI society. The greed for geoeconomic and geopolitical control has seen a data inequity that inhibits the Majority World's right to innovate; appropriation of the data commons, proprietary AI, delegitimization, and lack of room for alternative AI models, exploitation of workers, including peasants and indigenous people through corporate AI value chains and a shocking neglect of public finance to build digital public goods, including compute power, in developing countries.

Further, on the violation of intellectual property rights (as enumerated above), while we are concerned about AI being used to profit from IPR-protected assets without compensating the rights holder, the risk that we are much more concerned about is the use of IP rights, especially trade secrets, to enclose the data undergirding AI systems.

The social data commons are enclosed because a few firms retain exclusive ownership of data. The fragmented enclosures result in what is referred to in the scholarship as "tragedy of the anticommons", or the wasteful underuse of the given resource. Trade secrets in data disincentivize innovation, preventing the non-exclusive access necessary for exploring data's multifarious propositions and the right of all economic actors to meaningfully leverage data for unlocking its value. Trade secrets in data also render AI systems built on the data unexplainable.

A related point to this is that data governance and AI governance should be seen as two sides of the same coin. If the data-sharing legal frameworks are not fair and equitable, public sets and data pools are at risk of being locked up in the name of AI innovations by private actors, with little benefit accruing to the community whose data is thus locked up. Hence, strong institutional safeguards need to be put in place to protect social sector datasets, especially where there is a risk of proprietization of core development functions through AI models (such as in health, education, and welfare). Further, access to public domain and open government data should be allowed only on a conditional basis, subject to purpose limitations and clear sunset clauses on use. There should also be reciprocity guarantees in common data pools whereby private model developers who build on public data layers are mandated to share back and enrich the commons.

7. Are there specific individuals, groups, or societies/economies/(eco)systems that you are particularly concerned may be harmed by AI over the next 18 months?

1. **Workers at the risk of job loss due to AI automation and devaluation of labor of those involved in AI-related tasks** like data labeling, data and image annotation, transcription, translation, and content moderation, pivotal for training, monitoring, and servicing AI
2. **Minority and marginalized communities including women, LGBTQI+, racial and religious minorities, disabled persons, etc** face immense risks from the amplification of biases, prejudices, and stereotypes by AI systems, leading to discrimination in employment and educational opportunities and welfare delivery, perpetuation of disinformation and hate speech, wrongful implication by the criminal justice system, etc.
3. **Low-income countries and underserved communities:** AI technologies often require significant investment and infrastructure, which low-income communities may lack. This can exacerbate existing inequalities, as wealthier individuals and regions benefit more from AI advancements. Additionally, these communities might have limited access to AI-driven services and education.
4. **Planet:** The biggest threat posed by the current trajectories of AI development exacerbates the environmental crisis. Emerging evidence seems to suggest that AI may be more of a problem than a solution to our struggle against climate change, water shortages, and high energy consumption. Some estimates suggest that the water consumption in training Open AI's large language model GPT 3 was equivalent to the amount taken to fill a nuclear reactor cooling tower.