ISSN: 2641-2039

Global Journal of Engineering Sciences

DOI: 10.33552/GJES.2023.11.000762

Iris Publishers

Mini Review

General AI Rising: A Major Societal Risk, a Blessing, or a New Horizon of Collaborative Intelligence?

Serge Dolgikh*

Researcher, Artificial and Natural Intelligence, Ukraine

***Corresponding author:** Serge Dolgikh, Researcher, Artificial and Natural Intelligence, Ukraine

Received Date: December 13, 2023

Published Date: December 21, 2023

Introduction

Recently, a group of prominent scientists, engineers, and thinkers in the field of Artificial Intelligent Systems also known as Machine Intelligence and AI (Artificial Intelligence) came up with a brief statement on potential societal risks of the new technological advances in the field. Due to its succinct character, we will cite it in full here [1]:

"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."

Interestingly, global warming was not mentioned as an example of societal risks, that can be a subject for another discussion. In this piece we would like to expand and comment on the statement, not in the least to attempt to understand the nature of the risks and challenges that have been emphasized, but also flesh out for a general audience where the perceived risks may be coming from, and what considerations may convince one consider them as an existential, societal-level risk.

To begin, it may be worth making an observation that the potential risks that are coming to the public's attention these days aren't really very new. The first advances of artificial systems achieving a human level of ability in intellectual tasks were reported in the late 1990s – early 2000s, in increasingly complex visual recognition tasks, followed by the development of average human-like and expert-like abilities in intellectual games (Chess, Go) and culminating in the ability of complex artificial models to maintain human-like conversation and generate different types of complex media, including visual, speech and music, programming code and so on. The risks and the problems associated with it are, in fact, decades old, springing from the first advances in the field that demonstrated the potential capacity of the new intelligent systems not only to automate and speed up routine tasks but to achieve human-level performance in some tasks previously considered exclusively a human ability and prerogative. With the advent of "chat" models, a readily available direct experience to the public, the potential of AI in common, easily understood everyday activities has finally caught up in the perception of the general public.

Technically, we are not yet at the stage where the advent of General AI (also: Artificial General Intelligence, AGI [2]), a universal intelligence matching human ability in any domain is a matter of reality. But given the current state of technology and the pace of advances in the field, it is no longer a matter of speculation and science fiction either. Indeed, examining the essential characteristics that still separate human and artificial intelligences may illustrate this point:

1. An ability to create, propagate or reproduce itself in the physical environment, and modify its physical environment.
2. An ability to advance and improve itself.
3. Having an imperative, that is, a distinct and prevailing existential purpose.

Recent experiments and experiences with "chat" models (such as pre-trained large language models) show that even at the current stage, an advanced AI system is capable of creating a plan and executing it to an extent, to attain a certain purpose. So, the first point does not appear to be far-fetched imagination today. The others are still less clear or controversial, but there are many bright minds working in the field and to a very wide variety of ends and purposes. These concepts cannot be hidden or removed from public space, so the only way to reduce the risks at this stage is to bring them into an open discussion. An intelligent entity that possesses all of these qualities, with near-instant ability to absorb and process information at or above basic human ability, instantly learning and sharing it with other "individuals" with unknown, but very likely much faster pace of progress would indeed signify an arrival of a new and powerful intelligent species on our planet. And what if, at some point in their development path it grows more powerful than us?

How would it affect us? We know what we have done to the wild nature, the aboriginal people on multiple occasions in earlier history. Is there and can there be any assurances that the encounter would be positive to humanity, or at least, not catastrophic?

The risks of the arrival of AI capable of human-level intelligence have been identified and discussed at length. At this point, it can be worthwhile to examine and systematize them. One can be looking, for example, at these characteristics of potential impacts:

- The extent of the disruption: does it have the potential to impact the whole, or a large part of society?
- The rate and the time frame of the disruption: how does it compare with the ability of a human society to adapt?
- Potential benefits to humanity.
- Potential risks, harmful effects, and threats to humanity.

One can attempt to consider, within the proposed framework some of the possible risks and potentially harmful impacts of a General AI technology that were raised and discussed:

1. A new rapid technological transformation – a technology and societal disruption
2. Malicious use and unintended consequences
3. Militarized AI
4. Uncontrolled development – a new advanced intelligent (species) out of, and beyond human control
5. The emergence of a powerful intelligence: friendly, indifferent, unfriendly, or hostile?

Throughout history, humanity has undergone many, probably many more unrecorded, disruptive changes in social organization and technology. As more recent examples, the invention of the steam engine in the 18th century led to a rapid transformation of society and its fabric known as "industrialization"; rapid advances in communications technology in the late 20th and early 21st centuries are triggering society-wide effects whose precise impact is still difficult to assess.

These examples can be seen, in the proposed assessment framework, as follows:

At the societal level: at least as of present, they are within the framework of human ability to adapt and can be fully controlled by humans; having a clear potential to provide significant and valuable benefits to humanity, not least in terms of prosperity, longevity and quality of life; with real and sometimes serious risks and harmful effects, for example: industrial pollution; the ozone layer problem; deforestation; global warming; militarization of societal-level impact technologies such as nuclear and thermonuclear. However, each of these effects and impacts could be studied, analyzed, and evaluated with at least some degree of confidence.

Overall, as long as the introduction of AI technology is controlled by humans, based on past experience, it may not be reasonable to estimate the likelihood of major societal-level risks or threats from this category of impacts. And the benefits can be many, including increased productivity, a new burst of creativity, developments and improvements in many areas and applications in industry, public service, and society in general (Figure 1).

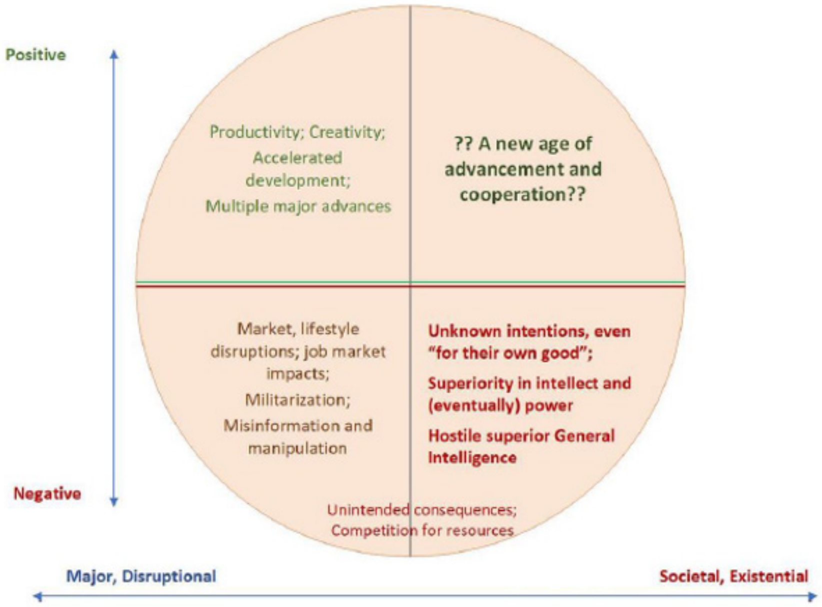
Positive
Productivity; Creativity; Accelerated development; Multiple major advances
?? A new age of advancement and cooperation??
Market, lifestyle disruptions; job market impacts;
Militarization;
Misinformation and manipulation
Unknown intentions, even "for their own good";
Superiority in intellect and (eventually) power
Hostile superior General Intelligence
Unintended consequences; Competition for resources
Negative
Major, Disruptional
Societal, Existential

Figure 1: General AI: potential impacts and scale.

The next category of impacts is malicious use and unintended consequences. One can talk about various AI-generated content that can and will be used for purposes such as deception, manipulation including political, misinformation, disinformation and propaganda, fraud and so on, not a hypothesis, the reality. The extent of such impacts could, potentially, reach the societal scale; but again, in the narrow sense, only as a new technology of creation and propagation of information content even of undesirable nature, it would not be something unique and unprecedented in the history of society. One may recall a quick adaptation of broadcasting technology for propaganda by totalitarian and authoritarian regimes. Again, as long as the impact is envisioned, planned and controlled by humans in a human timescale, one would be challenged to define this impact as being of a societal and existential nature.

Similar considerations can be made about the next category of impacts: militarized AI. Military technology is the one that may have known, possibly, the highest rate of technological innovations and disruptions throughout human history. There is no reason to expect that applications of AI technology, that are very likely already happening as we are writing this, would result in an uncontrollable societal risk, again under the condition that it is controlled and operated by human intelligence.

It is the following entries in the list that may bring us closer to understanding where the experts who supported the statement may see real society-level existential risks: the Columbus precedent. The arrival of an unknown; possibly and likely, superior, cohabitant, community, species? and possibly at some point, a competitor for a number or any number of resources, whose thinking and actions cannot be controlled or even, at some point, understood well enough to plan for contingencies. The Columbus story, and in human history at least and so far, it rarely ended well for aboriginals.

To bring it closer to home, it is more than an abstract speculation: we can have a pretty good idea, even now, where this superiority can be coming from. Suppose a General AI system, community as discussed and defined earlier, has become a reality. With it would come two essential abilities: the rate of solving problems and an acquisition of new knowledge that is vastly superior to the human, this is after all, why we use computers for routine tasks. And, secondly, an ability to pass on the new, gained knowledge almost instantly in the community, avoiding and bypassing the long and oftentimes painful process of learning, known to us humans. Together, these qualities can produce a rate of development, progress that is incomparable to ours. In a virtual flash, Artificial General Intelligence could reach levels of intelligence vastly superior to ours. And with intelligence, as we are aware, comes power in the tangible, physical world.

Certainly, in reality it may not be as simple as it's written on paper, as some knowledge would need to be verified in physical tests or empirical trials; still an expectation of a rapid and accelerating intellectual development of AGI technology? community? species? is well grounded in a logical argument. And combined with other discussed capacities such as an ability to maintain and propagate itself in the physical space, it would mean little less than that we just created ourselves a cohabitant on this planet, probably more intelligent and very likely at some point quite near, much more powerful. Aka, the Columbus precedent.

Should someone be worried? On the expert as well as a regular individual scale of analysis, this type of argument can indeed count as a societal, existential risk. One may ask the opinion of Incas or Mayans or any one of the dozens or hundreds of aboriginal tribes that disappeared from the face of the planet after the arrival of European settlers.

The last one in the list of potential impacts of AGI is more of a perspective on the preceding one: what would or could be the attitude of our new intelligent cohabitant to us, aboriginals? The answer to this question is not known and may not be known for some time, or until the time it can be found from experience. It needs to be noted though that even the first two options, the best news for us, may not be all good news. Think of a friendly and big, very big elephant in a China shop in the first case; and in the other, only trivial competition for some necessary resources, nothing personal, may create serious problems for aboriginals.

And so: where can and shall we go from here? Will we find out about the arrival of Artificial General Intelligence when it knocks on our door?

Acknowledgment

None.

Conflict of Interest

No conflict of interest.

References

1. Safe AI: The societal risks statement. Online: https://www.safe.ai/statement-on-ai-risk
2. Artificial General Intelligence. Online: https://en.wikipedia.org/wiki/Artificial_general_intelligence

Citation: Serge Dolgikh*. General AI Rising: A Major Societal Risk, a Blessing, or a New Horizon of Collaborative Intelligence?. Glob J Eng Sci. 11(3): 2023. GJES.MS.ID.000762. DOI: 10.33552/GJES.2023.11.000762.