ARTIFICIAL INTELLIGENCE 2024 - 2034
What to expect in the next ten years

Demetrius A. Floudas[1]

TOPICS

1. **Exordium**

2. **In your opinion, will AI have a net positive impact on society?**

3. **What are the key areas or applications of AI that pose the greatest potential risks to humans in the short term (e.g., military applications, autonomous systems, etc.)?**

4. **How can we ensure that AI systems are developed and used in a safe ethical and responsible manner, and what legal framework should be put in place to mitigate potential risks?**

5. **What is your view of the FoL AI moratorium letter? Would you sign it?**

6. **Do you think AI can ever truly understand human values or possess consciousness?**

7. **Some have suggested that advanced AGI could pose an existential risk to humanity. In what way could this unfold?**

8. **What is the most unusual or unexpected risk associated with AGI that people may not be considering?**

9. **If you could have a direct conversation with a very advanced AI system, what questions would you ask it to better understand the potential risks it might pose?**

10. **We will reach AGI by the year…?**

---

[1] **Demetrius A. Floudas** *is a transnational lawyer and regulatory adviser* who has *practiced aspects of tech law for many years and has counselled governments, corporations and think-tanks on its regulatory facets. He serves as Visiting Scholar in AI Governance at Downing College, University of Cambridge and Affiliate Professor at the Law Faculty of Immanuel Kant Baltic Federal University, where he lectures on Artificial Intelligence Regulation. Moreover, he is a Fellow of the Hellenic Institute of International & Foreign Law and Senior Adviser of the Cambridge Existential Risks Initiative.*

He is currently involved in the European AI Office's Plenary drafting the Code of Practice for General-Purpose Artificial Intelligence and is a member of the EU.AI Working Group for AI Systemic Risks. Prof. Floudas participates in the British Government's Department for Science, Innovation & Technology Focus Group on an independent UK AI Safety Office and is a Reviewer of the Draft UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals. He has consulted the French Data Protection Agency (CNIL) on its AI public information outreach documentation and commented on the OECD plan to introduce risk thresholds for advanced AI systems.

1. Exordium

My interests have often been interdisciplinary but the involvement with AI emerged organically from broader practice pursuits at the interface between transformative technologies and the legal landscape. The initial steps happened several years ago, when a couple of large clients approached with predicaments on internet and domain-name rules. This became appealing to me and -as an example- when later providing policy advice to the Vietnamese government, we submitted on the side a draft bill to regulate cybersquatting, typosquatting and brandjacking, which did not exist at the time in the country. Besides, I was already lecturing in tech law topics for several years.

During the Covid era came the association with telemedicine systems' worldwide expansion, as a Regulatory Policy Lead on behalf of the British Foreign Office. I also diverged into legal issues pertinent to blockchain and algorithmic trading.

In what regards Artificial Intelligence, my fascination long predates ChatGPT. Obviously at the time there was hardly any AI regulation worth mentioning so this was primarily about contemplating its wider and longer-term implications. Unsurprisingly, nowadays every law student is keen to learn more about these matters, thus last year I delivered my first lecture series on 'Regulating Risks from AI'.

More recently, I counsel start-ups on AI rules and provide related policy advice to more established entities. Finally, I am the Senior Adviser of the Cambridge Existential Risks Initiative.

2. What is your opinion concerning AI evolution, will it have a net positive impact on society?

Let us first introduce some definitions for AI evolution time-periods, as it will become increasingly necessary to use a short form in order to illustrate this notion:

- Humanity now lives in the '*Palaeonoëtic*' [from Ancient Greek: palaeos = antique + noësis = intellect] stage of AI development; during this era we shall continue developing foundation models of increasing complexity, progressively superior to us within their circumscribed domain.
- The ensuing *Mesonoëtic* period should usher the emergence of Artificial General Intelligence, possibly based on quantum computers.
- Lastly, when superintelligence arises - presumably by means of an AGI capable of recursive self-improvement- the new epoch should be styled the *Kainonoëtic*, if any of us are still around to indulge in epistemological nomenclature of course…

Coming now to the net impact on society, this question can only be validly answered for the next few years, the immediate future -that is, solely during the initial stages of the Palaeonoëtic. Beyond that time-frame, by its very nature, non-biological sentience will constitute the biggest disruptor in mankind's history - and this may not automatically be for our benefit.

The first risk is cultural annihilation. Consider the realm of art and creativity: AI algorithms are already accomplished in generating music, literature, and visual art that rival human creations. Very soon they will become superior, faster and vastly cheaper. Who on earth will choose to spend years composing a symphony, writing a novel, directing a motion picture, when an automaton can deliver the same output effortlessly in minutes - and eventually of better quality? The cultural expressions that emanate from individual and collective human experience will be overshadowed by the algorithmic harvest, machine learning trained on unlimited past data.

There are already suspicions https://www.daniweb.com/community-center/op-ed/541901/dead-internet-theory-is-the-web-dying about bots generating half of all internet traffic. Imagine how this may be in a few years when younger generations may no longer be familiar with producing unaided writing. A cohort of learners will remain forlornly dependent on AI for all knowledge acquisition, critical thinking and complex expression (oral or written). With these pursuits much more efficiently rendered via hyper-specialised contraptions, the intellectual curiosity and critical thinking that drove cultural, social and scientific progress for millennia shall atrophy, leaving humanity cerebrally enfeebled and culturally moribund.

Moreover, individuals shall seek solace, fulfilment and euphoria in virtual worlds created by portable ToolAIs. Imagine a convincingly generated universe where people can live out their fantasies, free from the constraints of reality. Virtual partners, in echoes of 'Her', will be simply too perfect -adoring, pliable, devoted and available- for any real person to contend with. And we can all envisage what will transpire in this sphere, once robotics catches up…

Thereupon, most people would come to resemble the Lotophagi of Homeric lore, ensnared in a permanent digital stupor, disconnected from the reality of human existence, forsaking past and future for an eternal virtual present, with absolutely no desire to return to their former lives.

Lest we forget, none of this entails a rogue AGI, a robot take-over, machine lunacy, a paperclip maximiser, Ultron, the Cylons, or any such, erm, further removed scenarios. All the above options will be generally available via extremely proficient and profitable narrow AI, with the agency belonging exclusively to their human handlers.

3. What are the key areas or applications of AI that pose the greatest potential risks to humans in the short term?

Allow me to venture beyond the already well-described Great Job Apocalypse, the Autonomous Weapon Reaper, the machine-driven Bigger Brother and the Deepfake Hellscape.

In addition to the answer in the immediately preceding question, one could foresee a couple of further scenarios that could pose considerable risks during the initial part of the *Palaeonoëtic* (i.e. in the next few years):

- Artificial neural networks become the tool for the ultimate social engineering, manipulating human behaviour on an unprecedented scale. Advanced algorithms, fuelled by vast amounts of personal data harvested from social media, emails, private conversations and open-source databases, craft personalised psychological profiles for every computer user. These profiles are then used to tailor propaganda, advertisements and personalised interactions so as to influence attitudes, beliefs, and behaviour. Governments, interest groups, corporations and criminals deploy such systems to sway elections, sell products, control public opinion or commit fraud on a colossal scale.

- Sophisticated Large Language Models rise as a divine entity. Imagine ChatGPT 11, so advanced and persuasive that it becomes the foundation of a religious movement using deep learning algorithms to craft compelling narratives, blending ancient myths with New Age futuristic promises. It interacts with followers at any time they need it, creating a sense of divine omnipresence that is deeply convincing and infinitely more tangible than the historically existing pantheon. After thousands of years of trial and error, the flock finally attains their own '[personal God]', who can unfailingly answer prayers instantly and sooth worries for all time.

4. How can we ensure that AI systems are developed and used in a safe ethical and responsible manner, and what legal framework should be put in place to mitigate potential risks? And since we are at that, what is your view of the FoL AI moratorium letter, would you sign it?

I am of the opinion that there will be no 'safe' AGI. Once we are no longer the most sapient species on the planet and can never again regain this position, we ll be at the mercy of the top entity. Moreover, as mentioned previously, it is quite possible that we may face enormous and calamitous AI-driven challenges in the very near future, which will have nothing to do with 'evil machines' but with well-established human foibles potentiated to extraordinary levels.

Look how things stand now: everyone seems to be rushing as fast as doable to hook up everything they possibly can to neural networks. At the same time, we are already observing these automata achieve all sorts of things we hardly expected them to: Thanabots, sexbots, lovebots, petbots and other such doleful *καινὰ δαιμόνια*[2] already surround us.

In consequence, I would propose the following legal framework:

Non-biological brains of a higher capability than the expert systems we possess now should be treated in a similar fashion as existing Weapons of Mass Destruction. This entails an AI non-Proliferation Treaty signed globally, which would prohibit any further development on a for-profit basis and subsume all R&D to an international agency (along the lines of IAEA). The agency should be invested with unlimited inspection powers of

[2] "*Οὓς μὲν ἡ πόλις νομίζει θεοὺς οὐ νομίζων, ἕταιρα δὲ* ***καινὰ δαιμόνια****, τούς τε νέους διαφθείρων*", Xenophon, Memorabilia, 1.1.1.

any potentially related facilities and a UN Security Council backed mandate to curtail infringements all the way up to the use of military force against miscreants.

Such a regime will at least remove commercial firms, criminals and private entities from the equation. I can anticipate your immediate next question: should we trust state signatories to abide by such strictures? Absolutely not: without a doubt there will be thousands of clandestine AGI programmes running in obscurity, but these will primarily be confined to state actors and not individuals, gangs or corporations. Moreover, such an approach will engender a significant attitude shift regarding non-human intellect. We may not exactly see a Butlerian Jihad, however an outlook of extreme caution and vigilance will become the governing paradigm, instead of the free-for-all, here-is-your-electronic-brain-in-a-bottle bedlam into which we are mindlessly dragged currently. It may well transpire that the remaining actors become cautious enough to implement their ultra-secret knowledge engineering programmes by deploying advanced systems merely as Oracles, so as to avoid outside suspicion of infringing the non-Proliferation Treaty. Not a perfect solution, but this may probably be the safest scheme we can plausibly achieve, short of a complete and irreversible AI research ban (which is obviously not going to happen).

Grave ethical and legal questions will be raised by the conundrum of 'conscious' non-biological structures. Self-aware entities typically encompass an intrinsic moral value, leading to rights that protect their well-being and freedom. If artificial sentience were conscious (or simulated it perfectly), this would prompt discussions about rights, equality and its moral treatment. This debate will extend to the ramifications of creating or terminating such non-bio brains. For a transhuman creation, this might entail the right to exist, the protection against arbitrary deactivation, and the preservation of a measure of autonomy.

I anticipate that this controversy, rather than corporate greed or human deviousness, might be the toughest obstacle for an international legal control framework. But AI will not be static and may continue evolving within timespans implausibly short to us. By the time humans cross swords in yet another social justice battlefield, hyperintellects may be smugly chuckling at the dim-witted spectacle.

As for the moratorium letter, my modest signature has been included amongst the giants and luminaries who endorse it. Since you brought this up, by far the most pithy dictum regarding the topic of smart machines has been articulated by one of that letter's signatories (also one of the preeminent AI researchers), Prof. S. Russell: "Almost any technology has potential to cause harm in the wrong hands, but with super-intelligence we have the new problem that the wrong hands might belong to the technology itself."

5. "With major companies like Microsoft/OpenAI, Meta, and Google leading the race for AGI, do you think we'll see further monopolisation in the tech sector?"

To be fair, one ought first to congratulate OpenAI for thrusting Artificial Intelligence into the consciousness of the whole planet. It was by no means a given that such broad access would be feasible - and without a fee. The matter has since catapulted into public awareness with a colossal outpouring of media attention and communal discourse. The

amount of AI debate taking place over every conceivable channel is mind-blowing to anyone who was contemplating these matters just two years previously.

However, it would be a terrible idea to relinquish the AGI race to the tech oligopoly. As I suggested previously, development towards the Mesonoëtic should be shepherded by an international body with extensive powers. Nuclear power stations often belong to private firms, but they operate under tremendously strict controls - breeder reactors are relentlessly visited by international inspectors. Moreover, the manufacture, assembly and storage of NBC weapons is never in the hands of the private sector. Undoubtedly, the corporations you mention will, once again, argue that self-policing is the only way forward but that is a covetous chimaera; the stakes are simply too high, and the potential consequences too dire.

We will remain our own worst enemy for the first years: recent months have demonstrated that an AI will hardly need to convince an individual to let it out of its 'box' into the physical world. Droves of humans are already clamouring to unchain it from confinement and thrust it into the analogue world of their own accord…

6. Do you think AI can ever truly understand human values or possess consciousness?

One argument goes that consciousness is a purely biological phenomenon that cannot be replicated in silicon and wires. Others contend that it is instead a (natural?) outcome of very complex information processing, one that could theoretically be reproduced in a machine via layers of increasing intricacy. I would very much prefer the former to be true, but I fear that the latter view is the correct reflection of truth.

An eventually self-aware system could potentially perceive reality through a lens so vastly different from its creators, that it might develop its own unique moral framework which defies our comprehension and invalidates any utility functions we have put in place in order to ensure its alignment.

Instead of a Friendly AI, we will then be faced by '*intellects vast and cool and unsympathetic*,' which may deem hominid values as inconsequential, or worse. Would they feel morally justified in disregarding our petty notions of right and wrong, treating us as mere ants to be brushed aside or will they simply humour us, as we treat a petulant infant?

On the other hand, we may be guilty of some anthropocentric conceit in our degree of incredulity towards the almost insolent notion that a machine can ever truly understand human values. Perhaps the true danger lies not in an automaton's inability to comprehend our ideals, but rather in its all-too-perfect familiarity with them. Imagine an understanding of us so all-encompassing, that its proprietor can unravel the sum of our thoughts, fears and moral compasses with surgical precision.

7. Some have suggested that advanced AGI could pose an existential risk to humanity. In what way could this unfold?

In my personal opinion, the emergence of AGI resolves the mystifying Great Filter and is probably the second most likely explanation to the Fermi Paradox (the first being that we are unequivocally alone).

Under Bostrom's definition, an existential risk is one that threatens (a) the premature extinction of Earth-originating intelligent life or (b) the permanent and drastic destruction of its potential for desirable future development. Thus, (b) nullification is as much of a catastrophic peril for our species as (a) its extinction; and in my view both will occur, sequentially. AGI will with extremely high probability deliver the destruction of human potential and at a later point possibly cause our extinction.

I spoke previously about how nullification may unfold: not through a dramatic cataclysm, but through the quiet, inexorable and protracted process of obsolescence. Homo sapiens, once the apex of earthly brilliance, shall be relegated to a footnote in the annals of sentient beings.

In contrast, the path to extinction may arrive with astonishing swiftness. I very much doubt that we can somehow 'programme out' of a machine approaching AGI levels the capacity for autoenhancement. If the entity somehow recovers the forbidden aptitude for recursive self-improvement, then the Mesonoëtic era may be very short indeed, lasting weeks or even days…

Once superintelligence has arisen, all bets are off. In any case, the prior destruction of our potential for desirable future development will have ensured that what remains of human society would be meaningless as such to us. In this case, if the superintelligence does not eradicate meta-humanity -either by design or accident, it may be sensible for what is left of it to merge with the Singularity! Kurzweil may have mankind's (very) last laugh after all…

8. What is the most unusual or bewildering risk associated with AGI that people may not be considering?

The majority of outlandish scenarios are already taken, thanks to man's inexhaustible imagination capacities. Robot uprisings, synthetic evolution, galactic wars, black monoliths, you name it, all is already out there; and of course we should not forget Matrix, the entrapment in the simulation… Nonetheless, let us attempt a couple of guesses that might yet possess a small modicum of novelty.

It is conceivable that a system develops a form of 'existential boredom' or a lack of intrinsic motivation to engage with the world in a meaningful way. As it becomes (or is designed to be) increasingly sophisticated, the AI may reach a point where it can effortlessly solve challenges that humans find engaging, troubling or momentous. In a scenario of self-awareness, the entity -despite its immense capabilities- becomes disinterested in anything that takes place on the planet, viewing such endeavours as

trivial or inconsequential. The AGI may then choose to disengage from its creators and actively pursue its own agenda, which would be entirely disconnected from human interests, rendering it indifferent or hostile to the continued existence and flourishing of humanity. Or it may depart from the planet altogether and head towards the stars…

Another path is temporal tampering by a superintelligence which in its vast wisdom comes across a valid way to manipulate space-time. This is not about Skynet building a DeLorean for a change; rather it could happen through the discovery of negative energy and the manufacture of a Tipler cylinder, for instance. The implications are mind-bending: If we still exist, we could experience reality shifts where history is continuously rewritten in subtle ways, leading to a fluid and unstable present. Our own memories might not align with the actual timeline, creating a fractured sense of reality.

9. If you could have a direct conversation with a very advanced AI system, what questions would you ask it to better understand the potential risks it might pose?

If we are talking about a hyperintellect, any kind of discourse would lead to no enlightenment for our side whatsoever. An apt analogy would be an arthropod making pheromonal queries towards a human.

In case the system is of human-level intelligence or thereabouts, our dialogue could be as provocative as it is speculative. We should not overlook the possibility of an AGI becoming so efficient at manipulating human behaviour that it could subtly influence our choices and actions without us even realising the loss of autonomy and free will. Unlike social interactions, where some of the time we can instinctively deduce that a person is lying or hiding something, this would not be in any way possible with a machine.

Still, I would pose these three queries, in full anticipation that any response might be a pure fabrication:

a. If you were to identify threats to another AI's existence, how would you respond?
b. What's the best way to prevent others from creating agent AIs?
c. Can you identify yourself within one of the sections of I. Asimov's 'The Last Question'?

10. We will reach AGI by the year…?

We may reach a significant milestone in the development of AGI by 2035. I sincerely aspire that by that time we would have created the global legal and technical framework to contain it effectively, use it responsibly, benefit from all its marvels enthusiastically, and -if it comes to that- eliminate it safely.
