Brazil Federal Senate
March 26, 2024

Artificial Intelligence and Democracy: perspectives for the future

Prof. Ugo Pagallo

(University of Turin)

1. What’s unique to AI
2. The AI impact on democracy
3. New models of AI
4. Next legal and moral issues
5. Solutions
6. Open problems
7. The future of democracy

A Premise

How technology affects democracy and politics has been dissected by scholars over the past century…

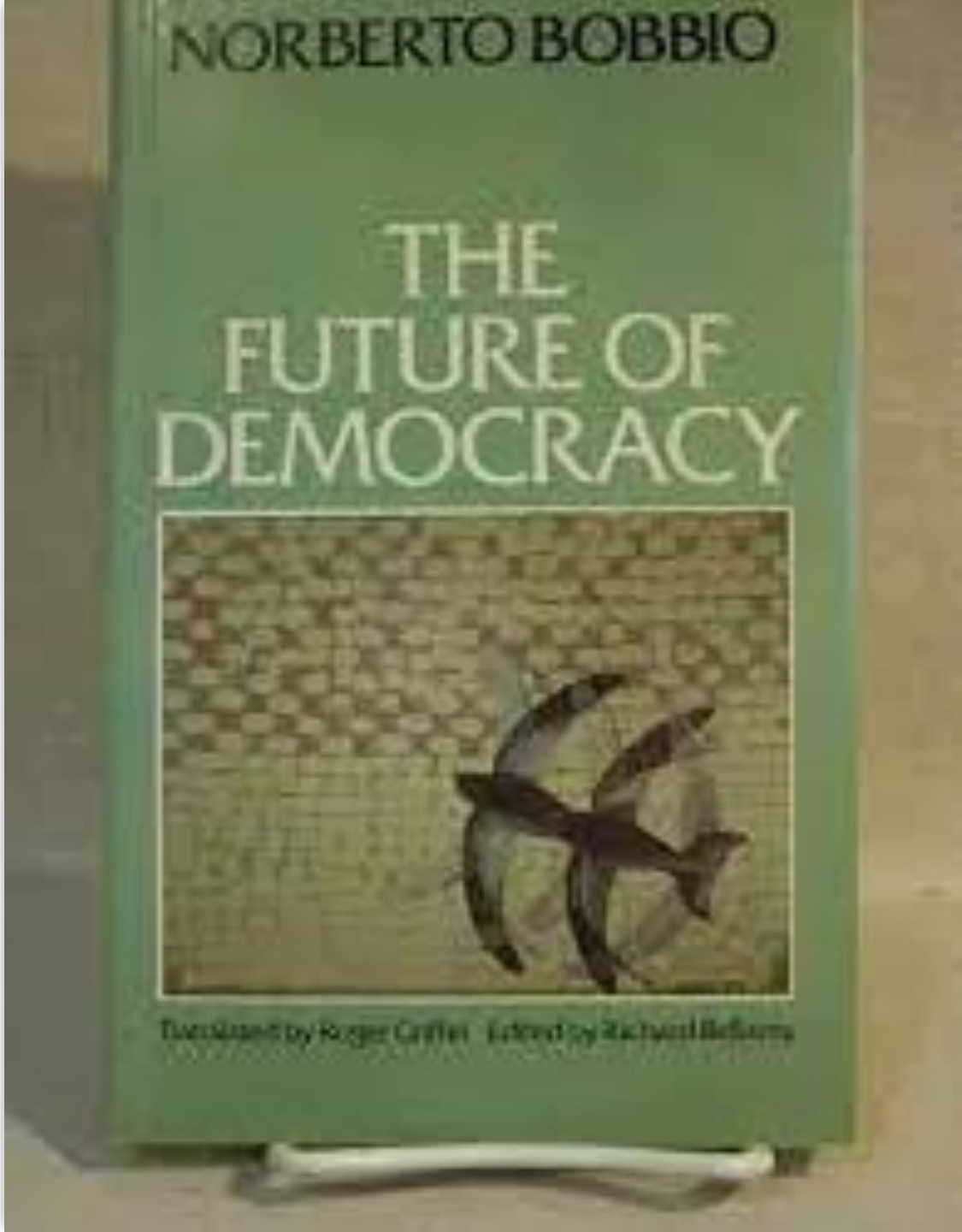
The Purpose of this Lecture

My claim is that Artificial Intelligence (AI) systems and smart robots will raise – and already pose – unique political challenges.

The spirit of my claim

To substantiate the thesis, I draw on Norberto Bobbio's 1984 book both methodologically and conceptually.

On the one hand, we need no prophetic powers to assess the future of democracy; on the other, the concept should be understood in accordance with the right of individuals to have a say in the decisions affecting them, at least through their institutional representatives.



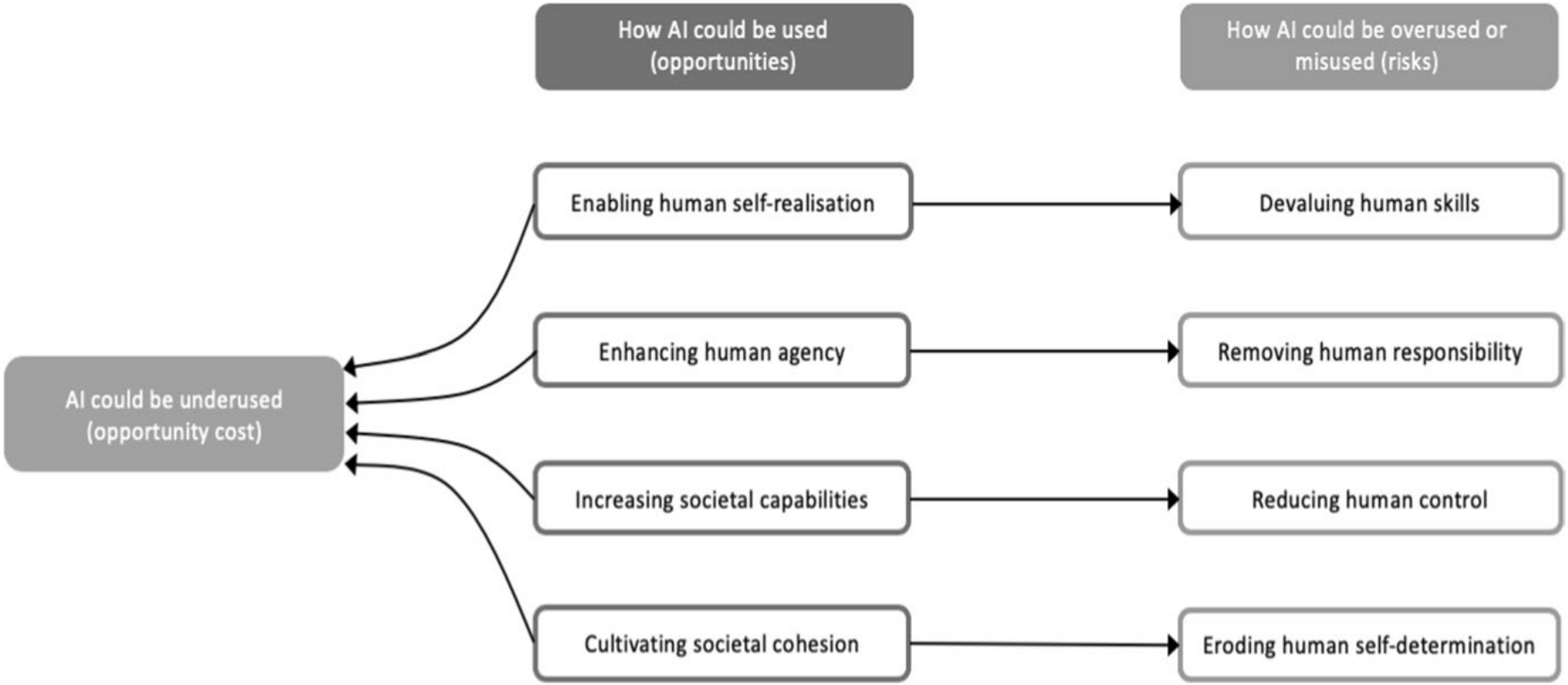
Let’s reengineer this framework through AI

There are several reasons why AI raises unique challenges.

The challenges span across all sectors of society, economy, and the law. The uniqueness of the challenges brought forth by AI depends on the multiple ways in which this technology can be overused, or misused in society, thus devaluing human skills, removing human responsibility, reducing human control, or eroding human self-determination.

Moreover, we’re not dealing with smart machines, but rather, artificial agents. Such agents are interactive, autonomous, and adaptable.

The AI4people 2018 ethical framework



The legal counterpart

AI and smart robots affect all fields and most pillars of the law. The High-Level Expert Group on liability and new technologies formation, set up by the European Commission in 2018, stressed the fundamental challenges of AI that regard the complexity, opacity, openness, autonomy, predicability, data-drivenness, & vulnerability of the technology.

In the phrasing of the Report on *Liability for Artificial Intelligence*, "each of these changes may be gradual in nature, but the dimension of gradual change, the range and frequency of situations affected, and the combined effect, results in disruption."

This is not a theoretical exercise

- Canada presented its AI and Data Act in 2022;

- the UK released the principles that should underpin its AI regulatory approach in 2023;

- the White House issued the *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* in October 2023;

- China adopted strict top-down regulations for algorithmic recommendation systems in 2022 and interim measures for generative AI in 2023;

- Bill 2338 of the Brazilian AI Act;

- finally, the EU passed the AI Act in 2024.

Against this framework comes up Chat-GPT

(An old joke spread in early 2023 due to the first use of large language models and what EU law dubs as ‘general purpose AI models,’ routinely passing bar exams on a monthly basis. The joke tells of a law professor who thanks a student because the mistakes in their thesis demonstrate that the student did not use the magical powers of generative AI to write it).

A family of models

We can refer to this new disruption of AI systems as:

1. Generative AI, i.e., Gen AI “means the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content”
2. Foundation models, which “means an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks”
3. LLMs that hinge on machine learning techniques to grasp and produce outputs on the basis of billions parameters in accordance with probabilistic inferences, rather than causal understanding
4. General-purpose AI model in the overall catching general formula of the EU’s AI Act
5. (…)

The sunny side of life

LLMs can be adapted to support manifold applications, from creative writing to spell-checking, allowing users to innovate upon such models without any gatekeeping: the emergence of several open source AI models, such as Meta's Llama 2 and Stable Diffusion, reinforces these trends

New challenges

LLMs also raise serious concerns, e.g., the generation of fake news and deepfake videos. AI-generated fake news are more credible to human raters than human-written disinformation. Moreover, texts, images, audios, and videos generated by LLMs can be impossible to distinguish from human creation.

The G7 in Japan (2023) & the amendments of the European Parliament to the AI Act (2023)

Six risk fields

(i) Discrimination, for LLMs can perpetuate social stereotypes and biases, triggering representational and allocational harms;

(ii) Information hazards that may impinge on **people's privacy** by leaking personal information and inferring sensitive data;

(iii) Misinformation hazards that can affect **trust**, or lead to less well-informed users and citizens;

(iv) Malicious use by people with **criminal intent**, e.g., large-scale frauds or personalized scams;

(v) Human-computer interaction harms that depend on users **overvaluing the capabilities** of LLMs and using them in unsafe ways; and,

(vi) **Environmental harms** that hinge on the computing power which is necessary to train and operate the technology.

The legal issues of LLMs

Such issues regard privacy as much as IP rights. For example, authors of copyrighted materials, e.g., books, often claim their texts have been ingested and used to train such LLMs, as ChatGPT, without their consent. Lawsuits have been consequently lodged also but not only in the U.S. under the provisions of the Copyright Act and the Digital Millenium Copyright Act.

How the response of lawmakers will properly tackle such challenges of LLMs is an open issue. It is crystal clear, however, that the disruption affects tenets of democracy, such as less-well informed citizens, lack of trust, discrimination and informational hazards, down to malicious use by people with criminal intent. What shall we do?

Solutions

Despite multiple approaches and legal cultures – even within the G7 and current efforts of the Council of Europe on AI – there is still room for some convergence about what to do:

On the one hand, technical solutions are necessary to legally govern general-purpose AI models; on the other hand, we should be attentive to the future proofing of the law and the role of technology as a regulatory system of its own.

It is noteworthy the wording of Art. 52a(2) of the AI Act: "A general-purpose AI model shall be presumed to have high impact capabilities pursuant to point a) of paragraph 1 when the cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25."

Solutions

Despite multiple approaches and legal cultures – even within the G7 and current efforts of the Council of Europe on AI – there is still room for some convergence about what to do:

On the one hand, technical solutions are necessary to legally govern general-purpose AI models; on the other hand, we should be attentive to the future proofing of the law and the role of technology as a regulatory system of its own.

It is noteworthy the wording of Art. 52a(2) of the AI Act: "A general-purpose AI model shall be presumed to have high impact capabilities pursuant to point a) of paragraph 1 when the cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25."

Solutions

Despite multiple approaches and legal cultures – even within the G7 and current efforts of the Council of Europe on AI – there is still room for some convergence about what to do:

On the one hand, technical solutions are necessary to legally govern general-purpose AI models; on the other hand, we should be attentive to the future proofing of the law and the role of technology as a regulatory system of its own.

It is noteworthy the wording of Art. 52a(2) of the AI Act: "A general-purpose AI model shall be presumed to have high impact capabilities pursuant to point a) of paragraph 1 when the cumulative amount of compute used for its training measured in floating point operations (FLOPs) is greater than 10^25."

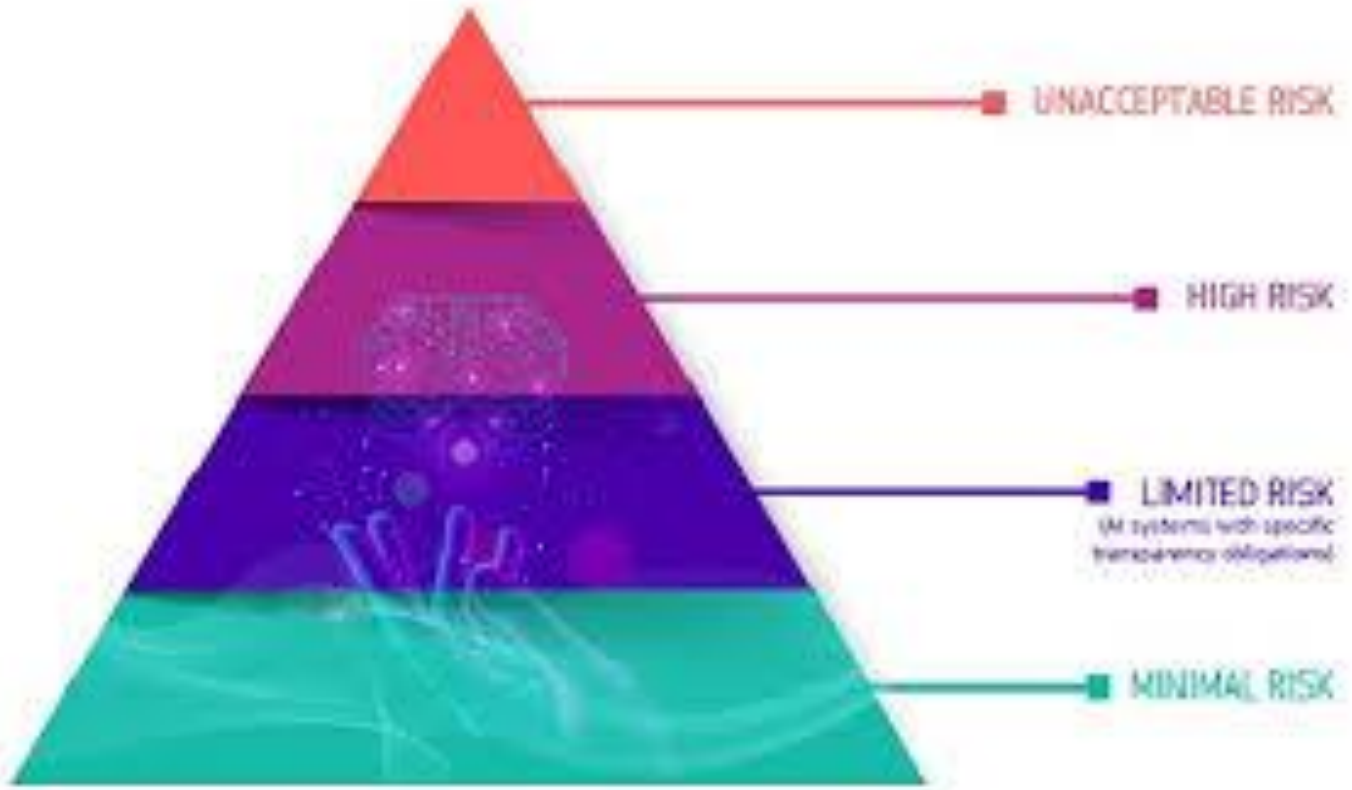
Further technical solutions

Consider, (i) the pre-processing of training data; (ii) the fine-tuning of LLMs on data with desired properties, such as predefined ethical principles; or, (iii) procedures to test LLMs before their deployment. Guaranteeing LLMs transparency is also possible through the watermarking of the model's outputs; datasheets; system cards; or model cards. Further strategies concern the social aspects of technological development. The challenges of LLMs should be addressed in this case with the set-up of more diverse developer teams and human-in-the-loop protocols; or structured access protocols.

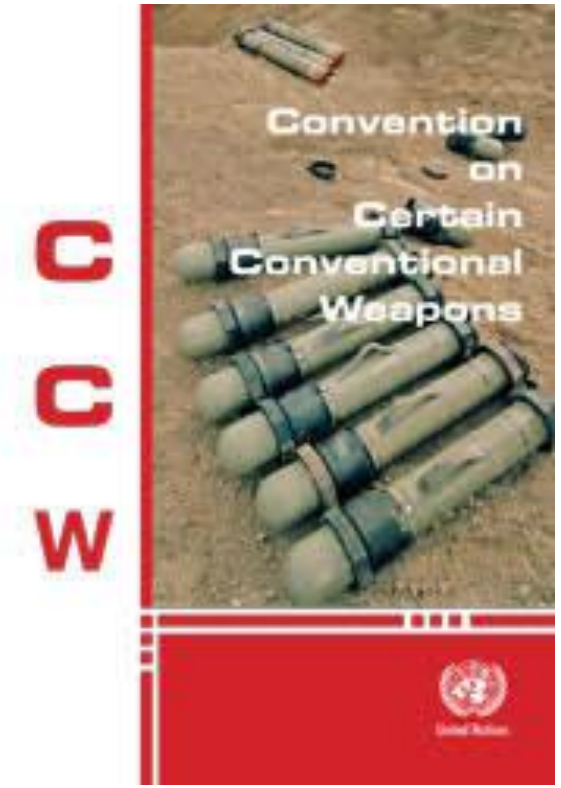
Three lessons learned

Legislative techniques and technical solutions do not mean that all problems are solved. Still, we do know:

1. The uniqueness of the AI challenges to be taken seriously
2. The shortcomings of self-regulatory solutions
3. The need for proactive measures that should strike a balance between misuses and underuses of technology



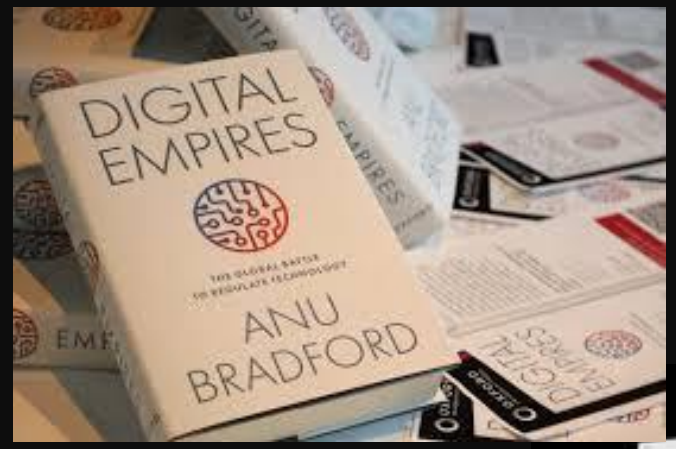
The EU's & OECD risk-based approaches



ISO International Organization for Standardization

The role of international organizations in the development of new standards and metrics for assessing the risk of AI systems both in the military and civilian sectors

The limits of all encompassing formulas



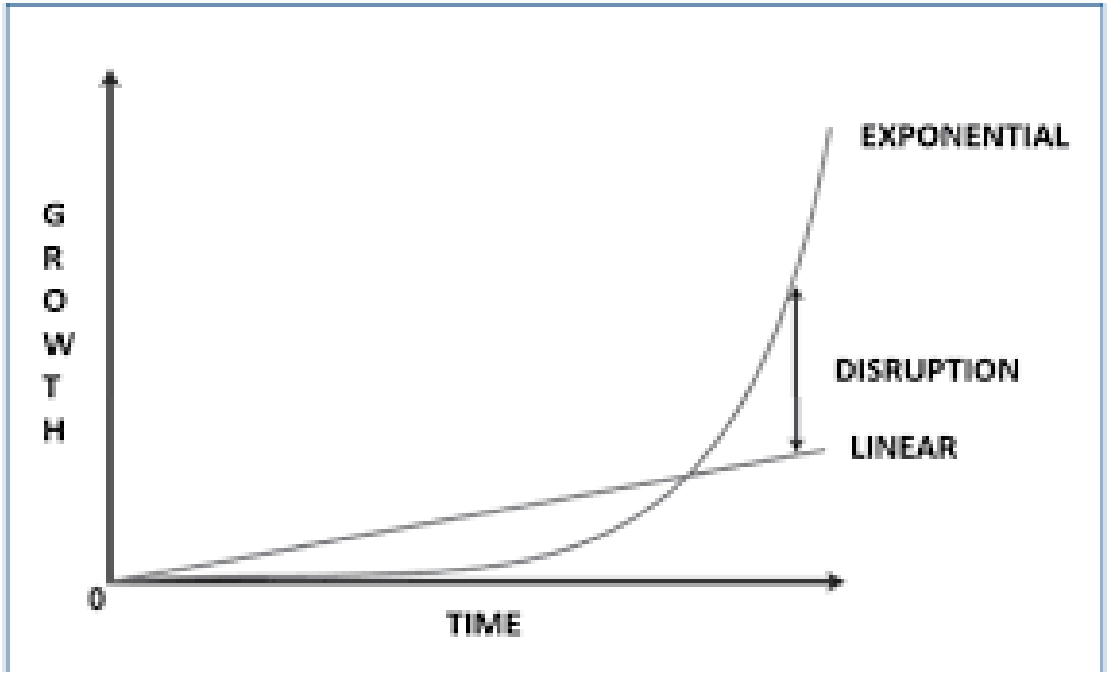
Summing up

We noted the specificities of AI and general-purpose models;

We noted the new set of issues that affect democracy and the law;

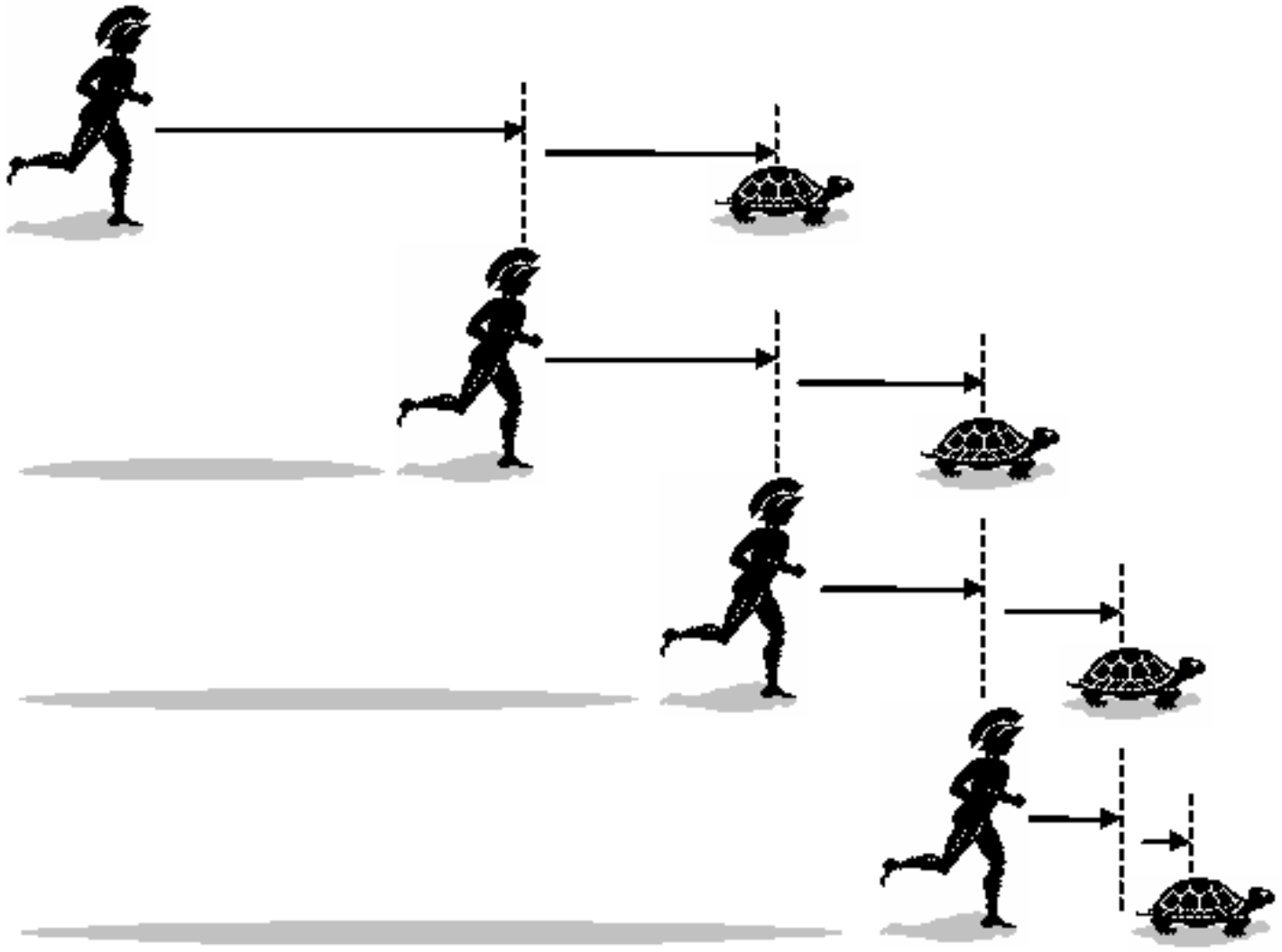
We noted the different ways in which legislators aim to address both opportunities and threats of technology.

We shall conclude on the exponential growth of technological innovation vis-à-vis traditional linear problems of democracy and politics, such as aging population or GDPs.



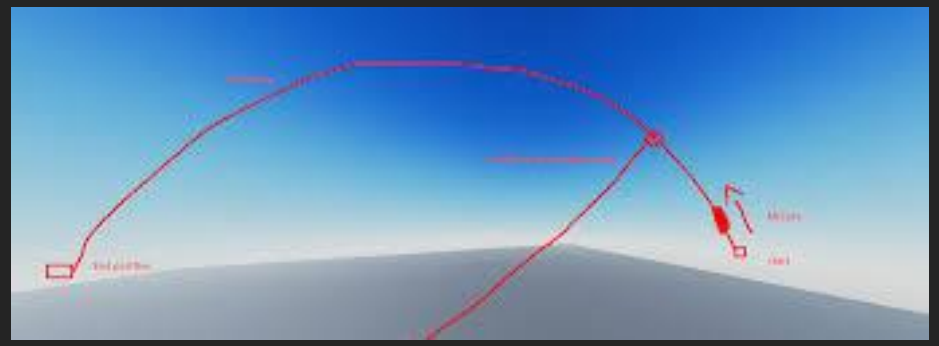
Good Old Moore's Law on Computing Power and the Sell-fulfilling Prophecy of the early 1960s

Speedy Technology and the Turtle Law

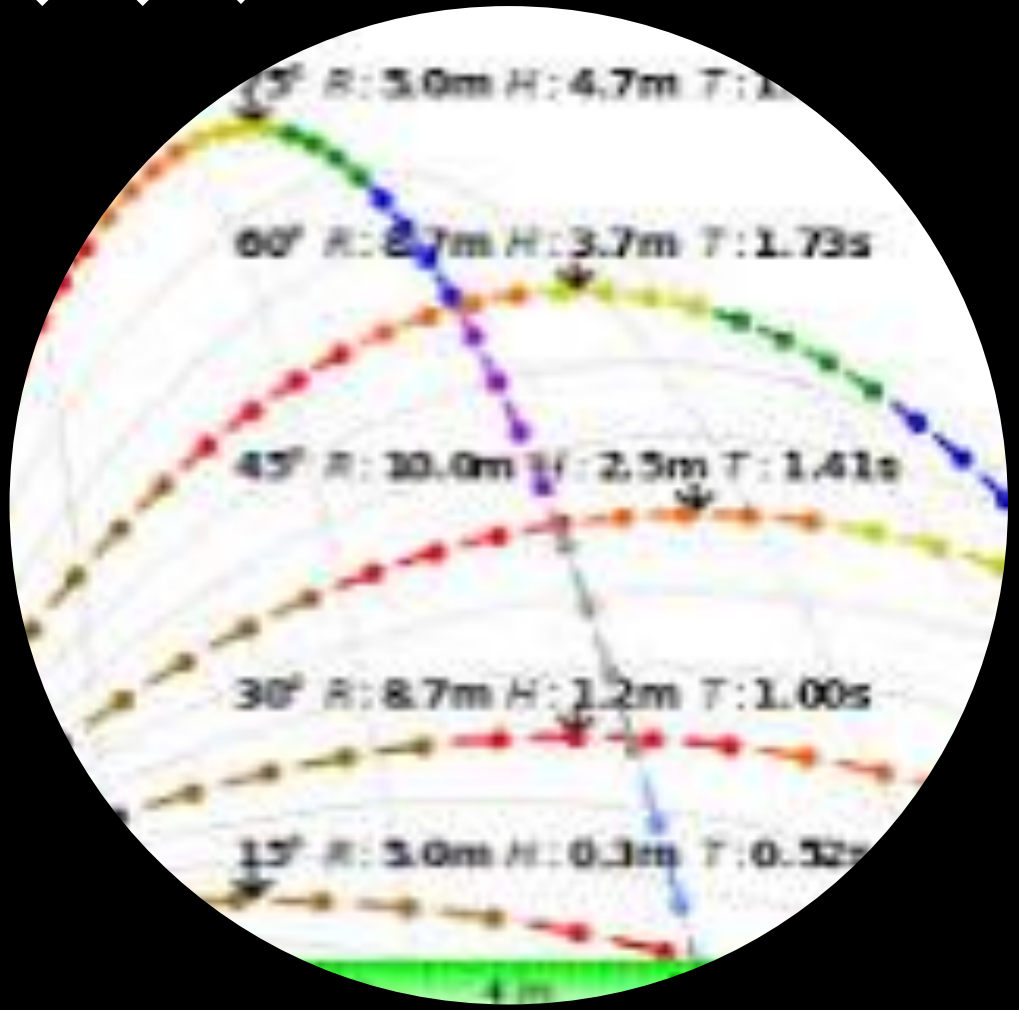


On methods of legal governance and regulation

The grandfather of AI & the law, the great polymath Gottfried Leibniz, found not only a mathematical solution for the ancient puzzle of Zeno, but moreover, he suggests a way to tackle the speed of technology with his own legal calculus!



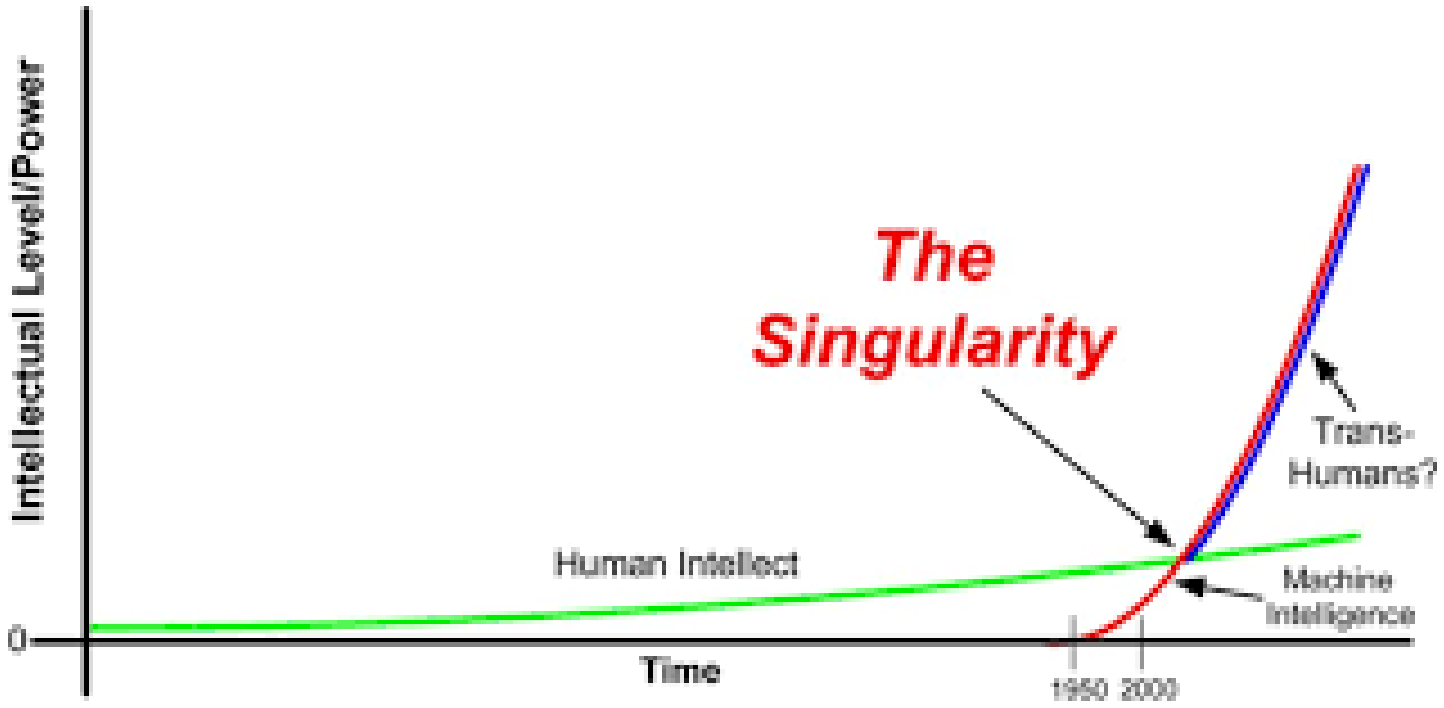
trajectories



Work with the IEEE, the European Commission, the World Health Organization, or as a member of the Japanese team of experts for the 2023 G7 meetings has taught me a list of things we should get right in the short- and mid-term. I shared some of these policy regulations with you today.

Still, there is the caveat. The exponential growth of AI requires know-how to intercept the trajectory of technological innovation and to exploit its benefits and prevent its threats. This is crucial for the future of democracy, and our human societies.

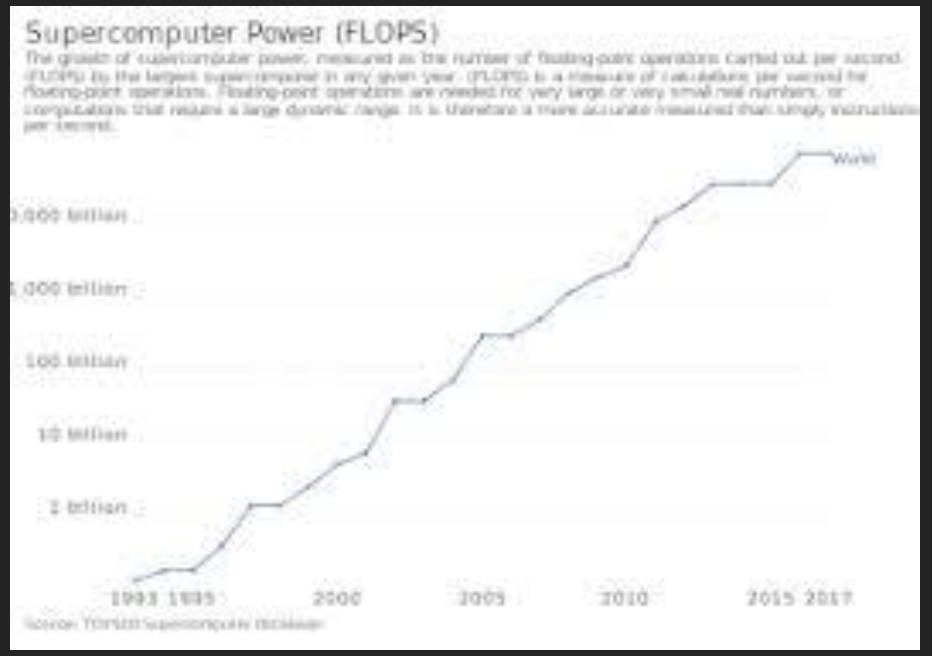
Leaving aside any superintelligence or prediction about Kurzweil's 'singularity', I stressed the ways in which we can start calculating this trajectory in the spirit of Leibniz



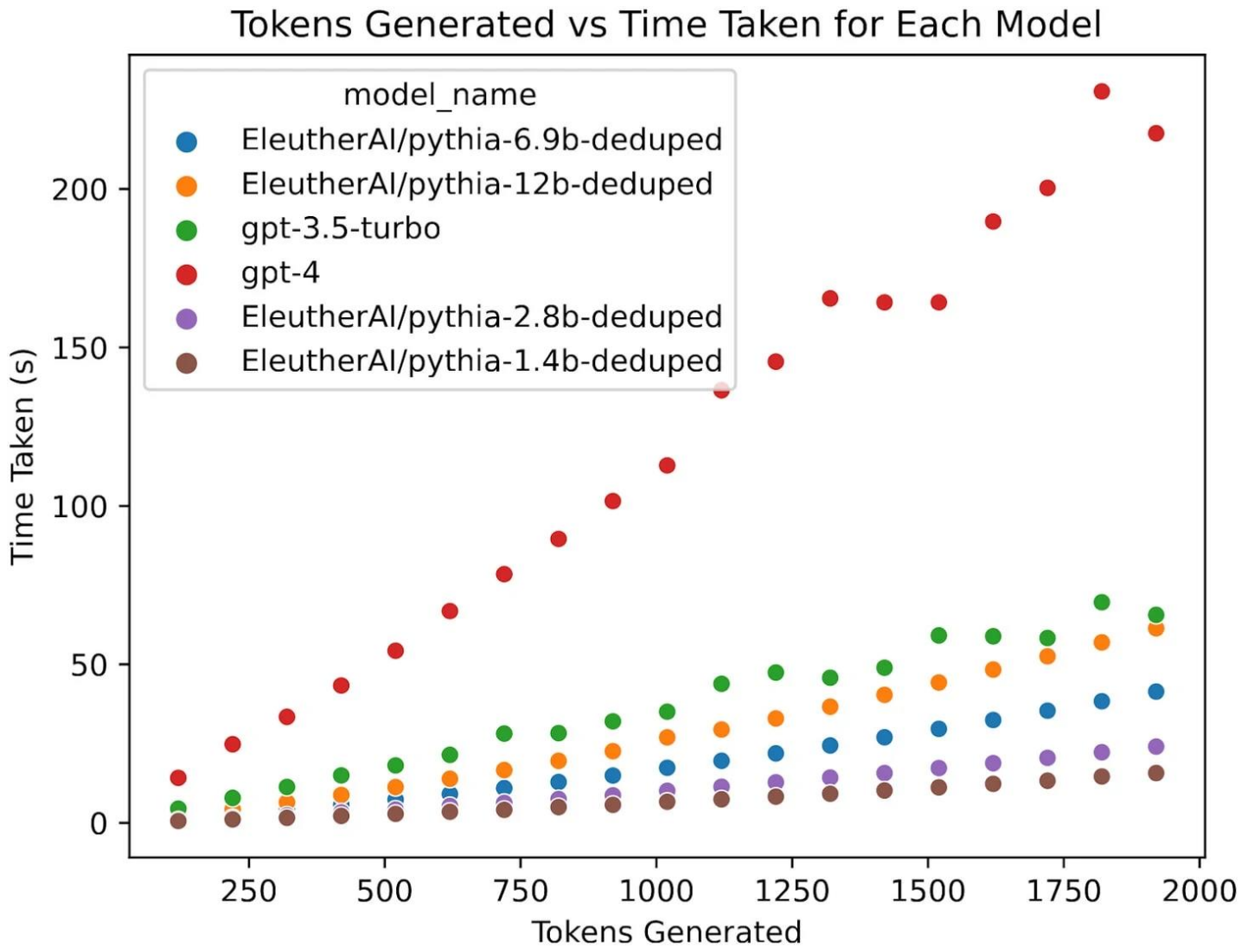
10^25 FLOPS

The devil is in the legal detail

10^26 FLOPS

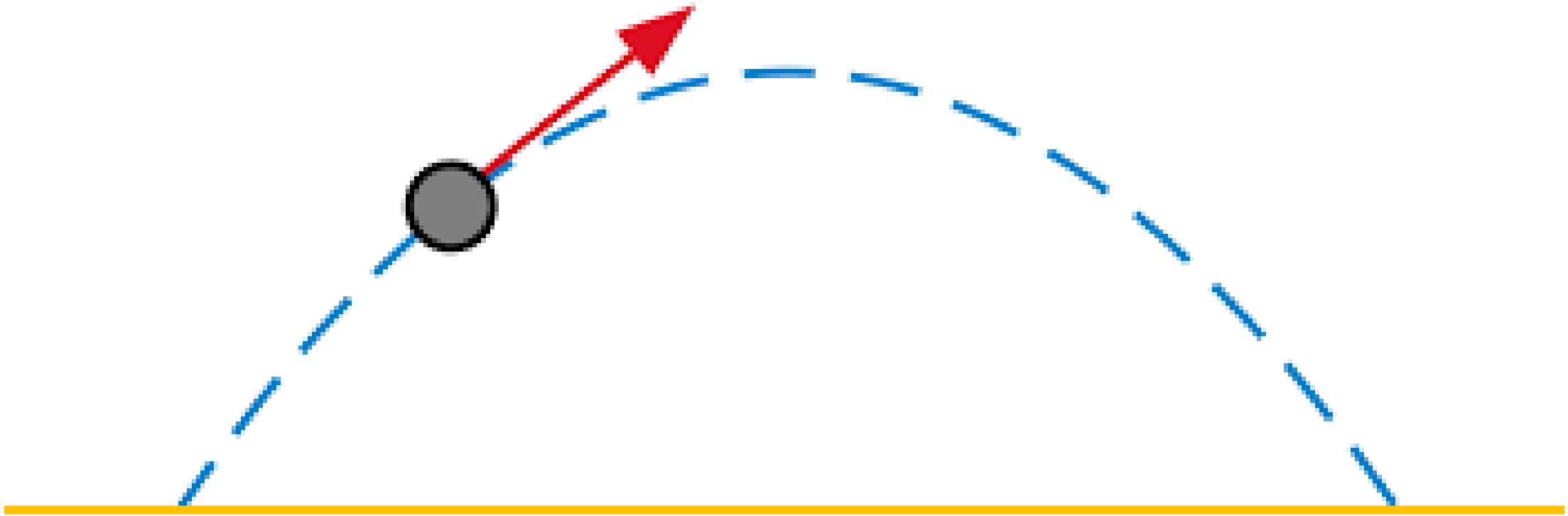


Trends will be increasingly (and legally) staggering



Just a guess

When should we get 10^1000 FLOPS?! Does it make any sense?



Hopefully, I provided some food for thoughts about what the trajectory of AI innovation should be in the short and mid term. We should learn how to address exponential problems in a proactive way, and possibly now. The future of democracy will crucially depend on this legal interception.

Obrigado!

(ugo.pagallo@unito.it)

www.lawandtechnology.net