Ramblings about future regulation of AI

By Bjørn Remseth
Electronic Frontier Norway (efn.no)
Vice president
rmz@efn.no

Disclaimer

While I am presenting this in my role as vice president of Electronic Frontier Norway (efn.no), what I'm presenting is not *the policy of EFN. It is part of an internal discussion we have, that may or may not end up in elements of it becoming EFN policy, but at this time this is not EFN policy.*

These are my *ramblings.*

What I will talk about today

What to regulate (what is “AI”)

- What we call AI today.
- What we can reasonably expect to emerge in the relatively near future.
- What we can less reasonably, but still plausibly imagine will happen in the near to mid-distant future.
- For the far future: All bets are off.

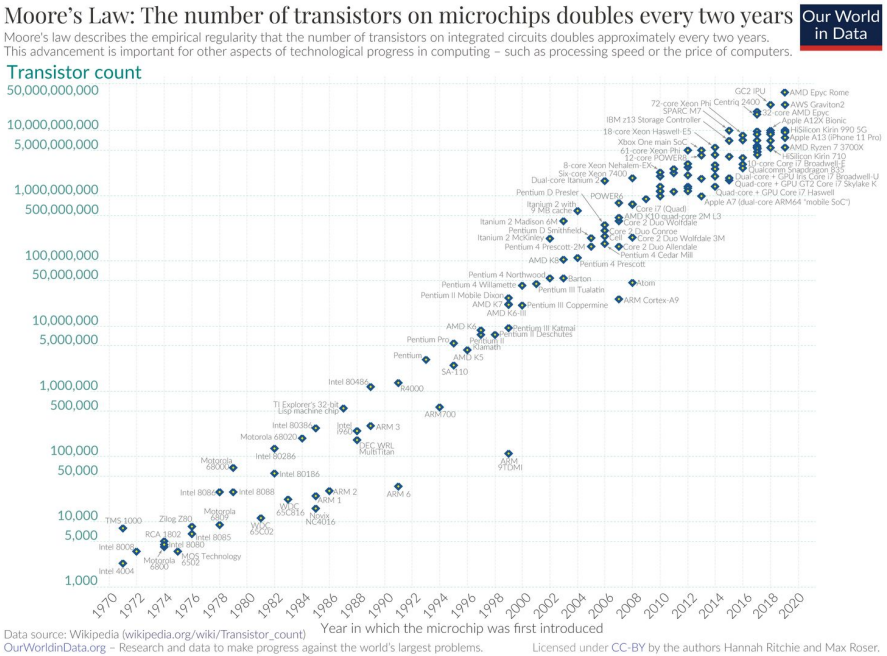
How to regulate

- *Today*: EU AI Act, GDPR, European Convention of Human Rights, Intellectual property law, market access legislation. Applied to *known* phenomena.
- *Reasonable future*: AI Act/ECHR, applied to emerging issues.
- *Somewhat plausible future*: *tbd*
- *Sci.fi. future*: Needs to be regulated in collaboration with AIs.

Evolution of AI

… but first a look the drivers

- **Hardware**
- Algorithms
- Economics/finance
- Philosophy

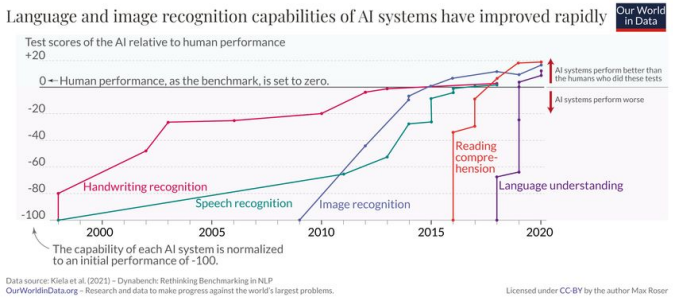


1. Can’t continue forever
2. Won’t stop this year (3nm)
3. Intel indicate 1nm in 2027

… but first a look the drivers

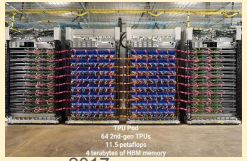
- Hardware
- **Algorithms**
- Economics/finance
- Philosophy

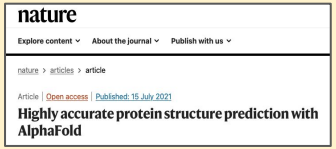
Classification/ Regression



Reinforcement learning


2017 AlphaGo

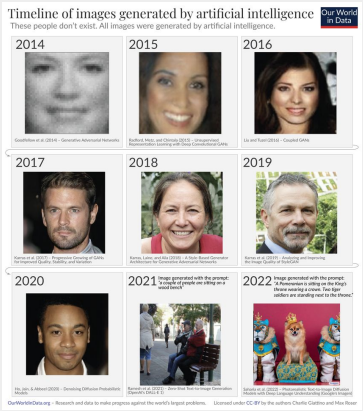

2017 AlphaZero



1997

https://community.cadence.com/cadence_blogs_8/b/breakfast-bytes/posts/alpha-go-and-alpha-go-zero

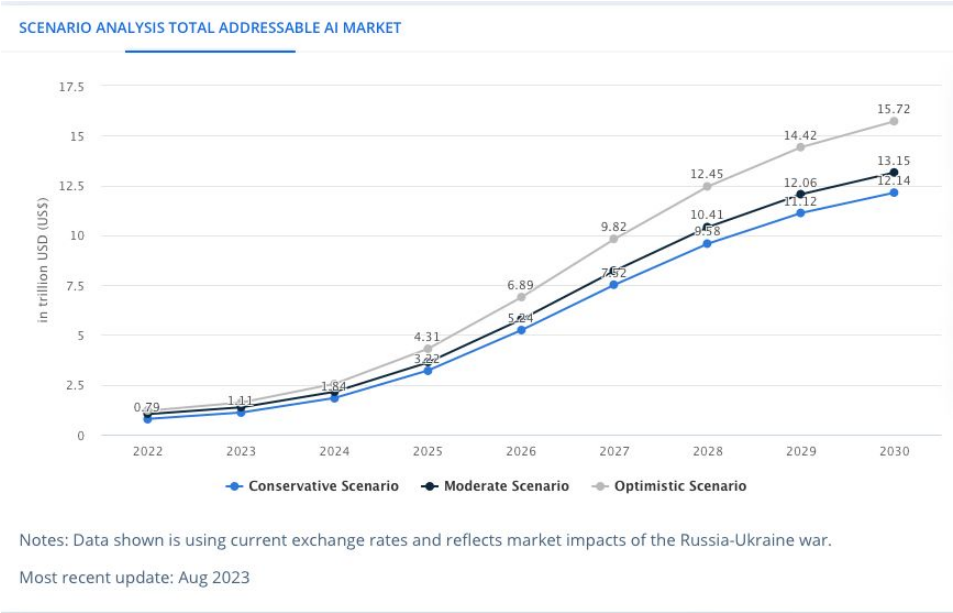
Generative



… but first a look the drivers

- Hardware
- Algorithms
- **Economics/finance**
- Philosophy

Probably not completely true, but the numbers are big and will be bigger



https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

… but first a look the drivers

- Hardware
- Algorithms
- Economics/finance
- **Philosophy**

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia | Search | Create account | Log in

Photograph your local culture, help Wikipedia and win!

Consciousness

100 languages

Contents hide

(Top)
Etymology
The problem of definition
Philosophy of mind
Scientific study
Medical aspects
Outside human adults
Stream of consciousness
Spiritual approaches
Notes
See also
References
Further reading
External links

Article | Talk | Read | Edit | View history | Tools

From Wikipedia, the free encyclopedia

For other uses, see Consciousness (disambiguation) and Conscious (disambiguation).
Not to be confused with Conscientiousness or Conscience.

Consciousness, at its simplest, is awareness of internal and external existence.[1] However, its nature has led to millennia of analyses, explanations and debate by philosophers, theologians, and all of science. Opinions differ about what exactly needs to be studied or even considered consciousness. In some explanations, it is synonymous with the mind, and at other times, an aspect of mind. In the past, it was one's "inner life", the world of introspection, of private thought, imagination and volition.[2] Today, it often includes any kind of cognition, experience, feeling or perception. It may be awareness, awareness of awareness, or self-awareness either continuously changing or not.[3][4] The disparate range of research, notions and speculations raises a curiosity about whether the right questions are being asked.[5]

Examples of the range of descriptions, definitions or explanations are: ordered distinction between self and environment, simple wakefulness, one's sense of selfhood or soul explored by "looking within"; being a metaphorical "stream" of contents, or being a mental state, mental event or mental process of the brain.

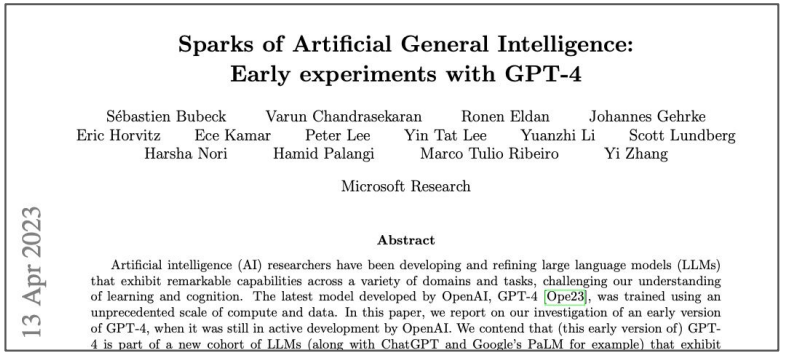
Etymology [edit]

The words "conscious" and "consciousness" in English date to the 1600s and the first

Representation of consciousness from the seventeenth century by Robert Fludd, an English Paracelsian physician

As far as I can tell:

- No sharp definition of neither intelligence nor consciousness.
- Hence it it is hard to say how far we are from it.



Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

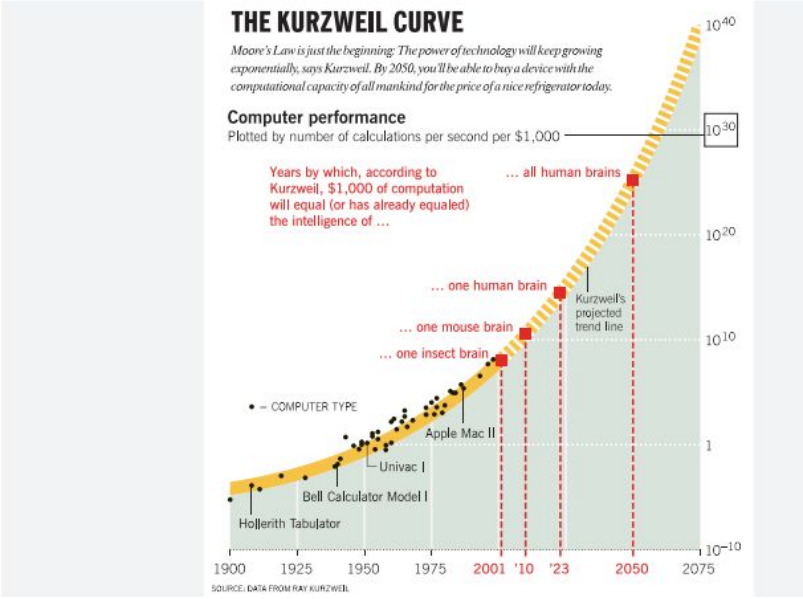
13 Apr 2023

Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a variety of domains and tasks, challenging our understanding of learning and cognition. The latest model developed by OpenAI, GPT-4 [Ope23], was trained using an unprecedented scale of compute and data. In this paper, we report on our investigation of an early version of GPT-4, when it was still in active development by OpenAI. We contend that (this early version of) GPT-4 is part of a new cohort of LLMs (along with ChatGPT and Google's PaLM for example) that exhibit

Regulation

Maybe not completely wrong?



Ray Kurzweil—The Smartest (or the Nuttiest) Futurist on Earth | Fortune

Regulating today's AI

What to regulate (what is "AI")

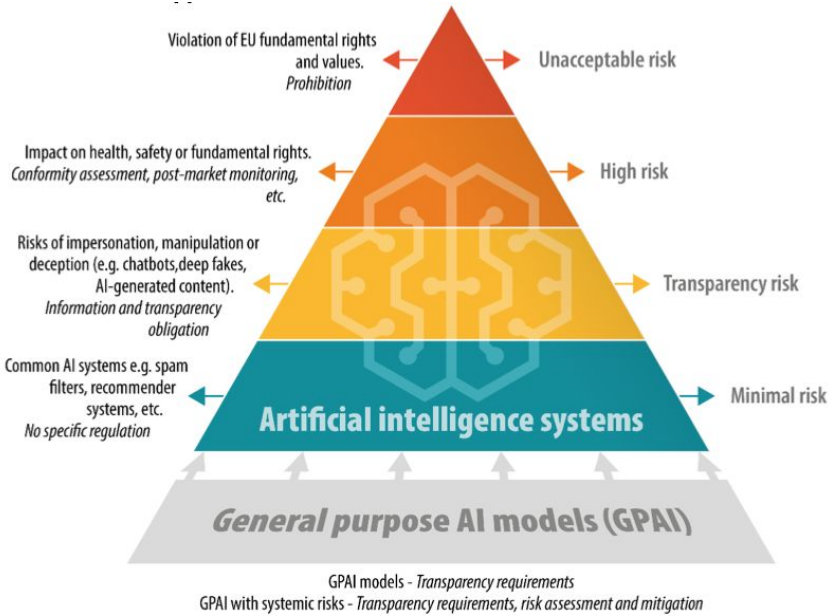
- Classification/regres…
  - Facial reco…
  - Medica…
  - …
- Ge…
  - …arize documents
  - …anslate and adapt media content
  - …

Too much to fit on a slide already

How to regulate

- *Today*: EU AI Act, GDPR, European Convention of Human Rights, Intellectual property law, market access legislation. Applied to *known* phenomena.

Regulating today's AI

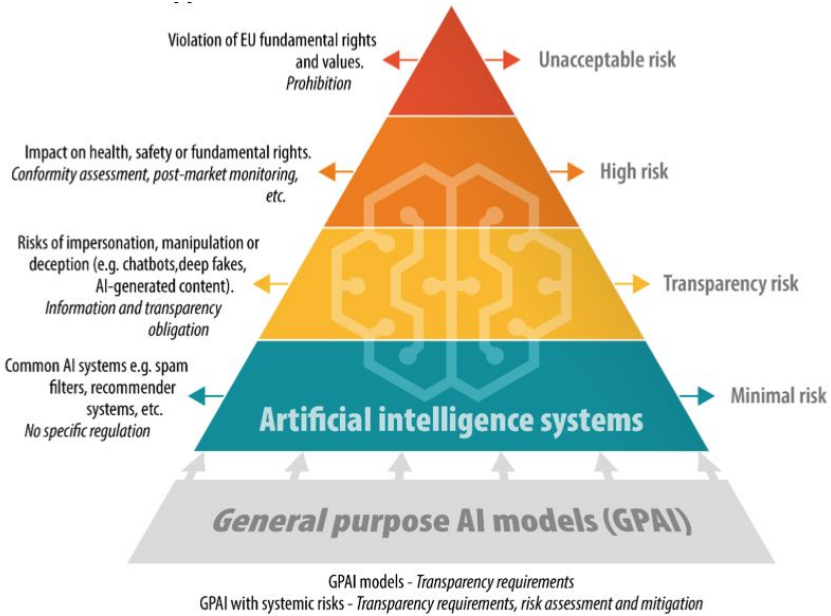


How to regulate

- *Today*: **EU AI Act**, The European Digital Markets Act (DMA), GDPR, European Convention of Human Rights, Intellectual property law, market access legislation. Applied to *known* phenomena.

<- Briefing doc for EU Parliment

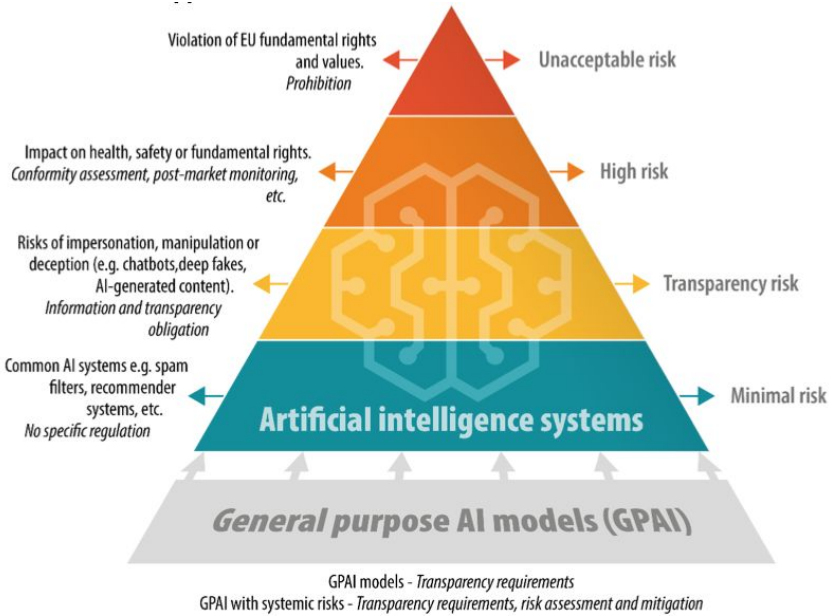
Regulating today's AI



GENERAL PURPOSE AI MODELS (GPAI)

- *Must draw up & present up to date tech doc & make available to downstream providers of AI systems.*
- *Must have policy that respect EU law, including copyright and GDPR*
- *Systemic risk: High impact capabilities must -> must warn EU (10^25 FLOP).*
- *Constant cybersecurity mitigation.*

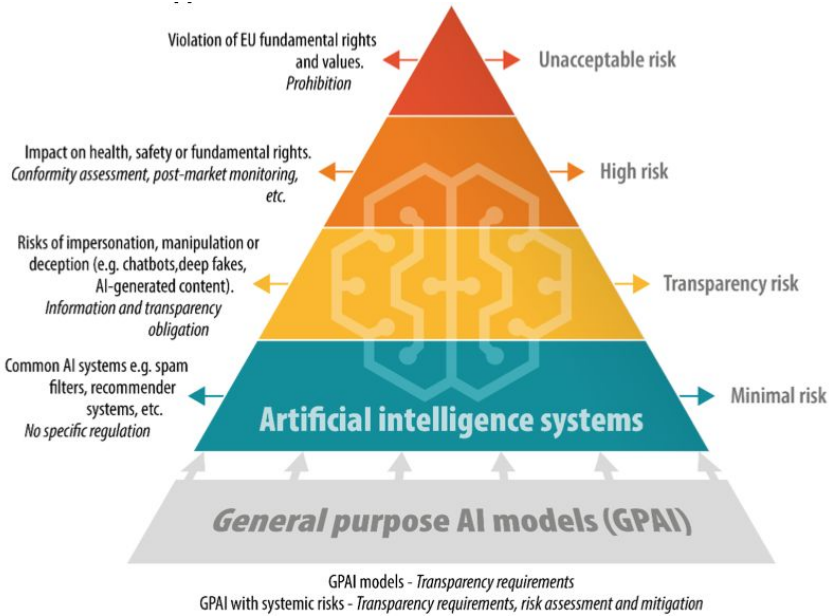
Regulating today's AI



MINIMAL RISK

- *Spam filters etc.*
- *Subject to GDPR, DMA etc., but no further obligation.*

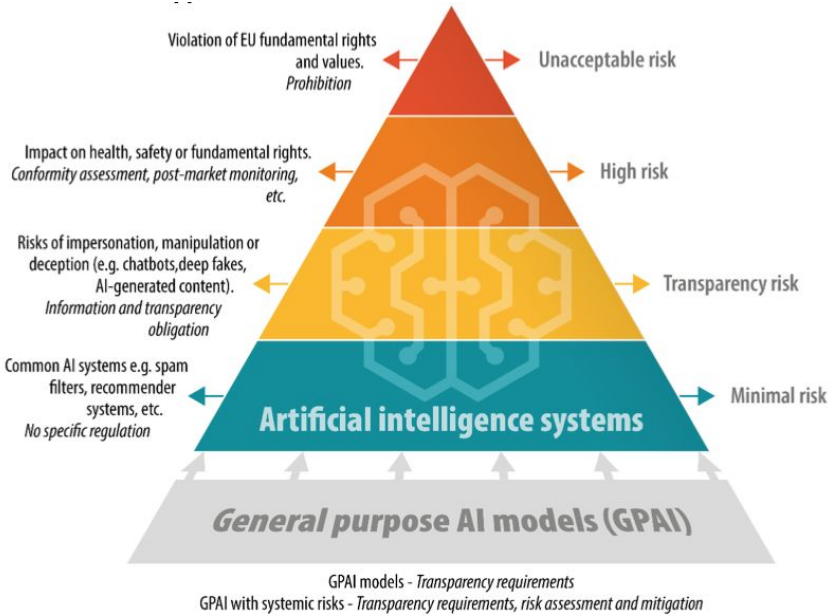
Regulating today's AI



TRANSPARENCY RISK

- *Can impersonate people*
- ***Must** inform users that there is a chatbot in the other end.*
- ***Must** (in some cases) mark large scale synthetic content as such (e.g. with watermarks)*
- *AI systems in workplace must **Must** inform workers and their representatives.*

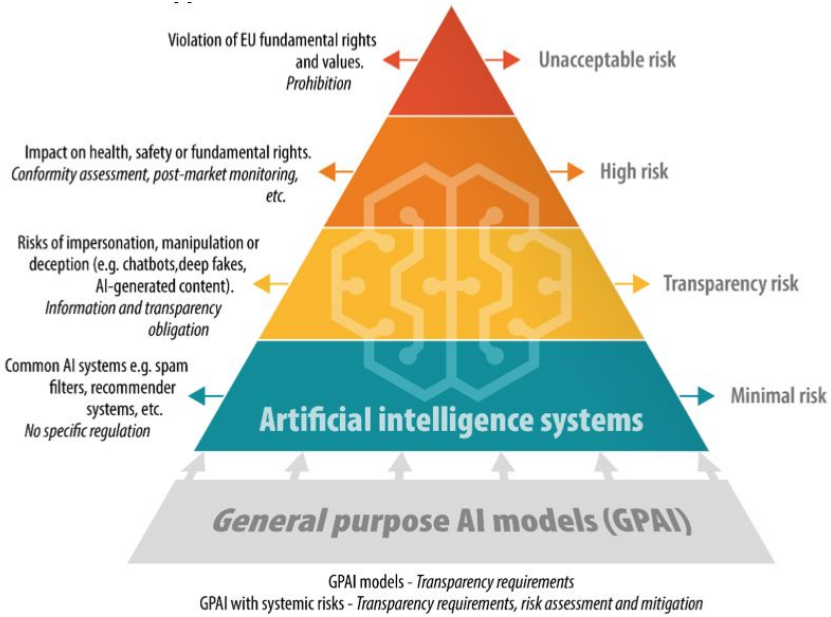
Regulating today's AI



HIGH RISK

- ***All** systems that profile **natural persons** will be considered high risk.*
- *Some sectoral law can classify systems as high risk (e.g. medical devices).*
- *The commission maintains a list of specific high risk areas, and high risk systems.*
- *All providers of such systems **must** run a* "conformity assessment procedure" before products can be sold in the EU.

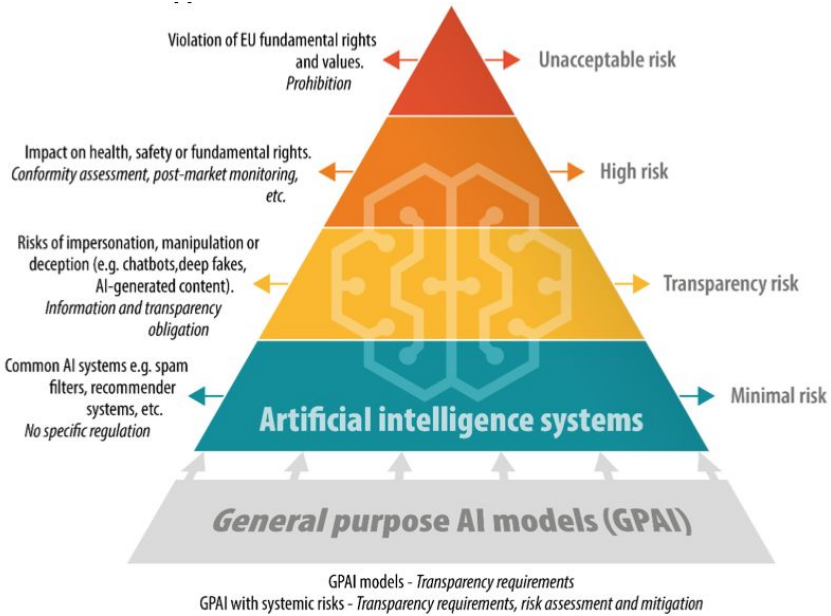
Regulating today's AI



PROHIBITED AI

- *Systems that can cause significant harm by ..*
- *Subliminal or manipulative or deceptive techniques that distort decision making*
- *Exploit vulnerabilities (age, disability …)*
- *Biometric categorization (race, political opinions, trade union, sex life,...)*
- *Classify based on behavior or characteristics.*
- *Real time biometrics in public spaces (with exceptions)*
- *Assess probability of committing crimes*
- *Create facial recognition databases*
- *Inferring emotions in workplace or educational institutions (except for medical or safety reasons).*
- ***Prohibited systems have to be phased out within six months after the act enters into force.***

Regulating today's AI



PRACTICES

- *If not compliant*: Fines up to €30 million or 6 % of the total worldwide annual turnover)
- *Sandbox: Environment for testing of innovative AI systems under strict regulatory oversight (for real world testing)*

Regulating today's AI

The European Digital Markets Act (DMA)

- *"Large online platforms" and their "gatekeepers":*
  - *Browsers, Search, Video, Social networks, …*
  - *Alphabet, Amazon, Apple, TikTok (ByteDance), Meta, Microsoft.*
- *Stricter rules for consent across services (single website OK no longer sufficient).*
- *More interoperability (e.g. search engines)*
- *Verification of how&where ads are spread.*
- *Permit businesses to have users link up outside gatekeeper's services.*
- *Users must be permitted to un-install whatever they want*

How to regulate

- *Today*: EU AI Act, **The European Digital Markets Act (DMA)**, GDPR, European Convention of Human Rights, Intellectual property law, market access legislation. Applied to *known* phenomena.

Regulating today's AI

...

- *ECHR article 10: Freedom of expression §1.*

  Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.

How to regulate

- *Today*: EU AI Act, The European Digital Markets Act (DMA), GDPR, **European Convention of Human Rights, Intellectual property law, market access legislation.** Applied to *known* phenomena.

Regulating plausible, near future AI

What

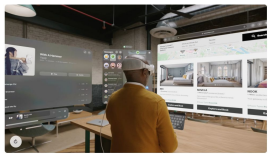
- Robots in “human space” (where humans go).
  - See everything (books,furniture, people, behavior, …)
  - Interact with everyone-.
  - Will do profiling of physical humans
  - Can be anywhere from high risk to prohibited.
  - Classification/certification will be fun :-)

The robots (etc.) are coming:

Deepmind robot cooking

Humanoid robots

AR/VR



Regulating plausible, near future AI

What

- **Small models** will become very capable
  - “Small” is relative
    - Algorithmic and HW improvements make tomorrow’s small models very capable.
    - Maybe hard to train, not hard to run.
  - Can run on modest hardware
    - Inside robots
    - Inside personal compute devices (laptops, tablets, headsets, glasses)
- Distributed ownership
  - Many “small” owners
- Can be modified (“fine tuned”)

How to regulate:

- Define safe subspaces
  - … with limited areas of use
- Maybe tamper-proofing
  - Difficult with SW based learning systems.
- Maybe primarily go for the “transparency” risk type regulation
  - Always know when an AI is involved.
  - Know what’s in the box.
- If widespread ownership, then maybe owner/operator must be responsible for system’s actions?
  - Like with cars?
  - If so: **Who will ensure the AIs?**

Regulating somewhat plausible, more future AI

What

- Don’t know. Crystal ball doesn’t reach that far into the future

How

- Just hope that the current framework holds and can be modified.
-

How to regulate the far future of AI, if all bets are off?

What

- Artificial General Intelligence
- (maybe) as smart capable as humans for just about any task.
- Self aware
- With agency, a.k.a. “Free will”.
- Plausible within 5 months (GPT5) to 50 years (lifetime of my children?) time frame?

How

- Fundamentally unknown
- We will have to learn how to co-exist with in many respects superior species.
- If we want to regulate them, and we probably do, it will have to be a collaborative effort
- … with the AGI(s).
- It will help if their motivations are aligned with ours (AI alignment research)
- **My guess/hope**: If we do our best *today*, then the machines may adopt those practices.
- … so we had better take it seriously

Thank you for your attention