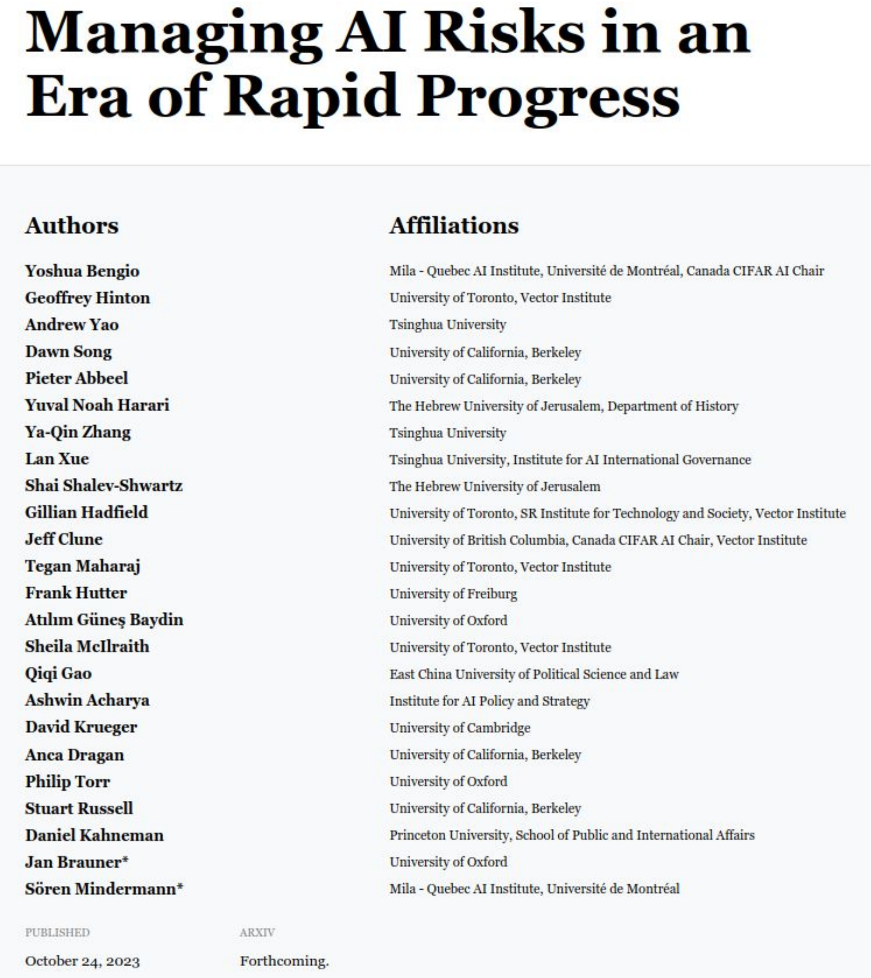
Some introductory ideas for the panel “AI: across innovations and ethics”

Luca Mari
lmari@liuc.it
https://lmari.github.io

2023 IEEE International Conference on
Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering
Milano, 27 October 2023

A *timely* subject…

… three days ago:
https://managing-ai-risks.com

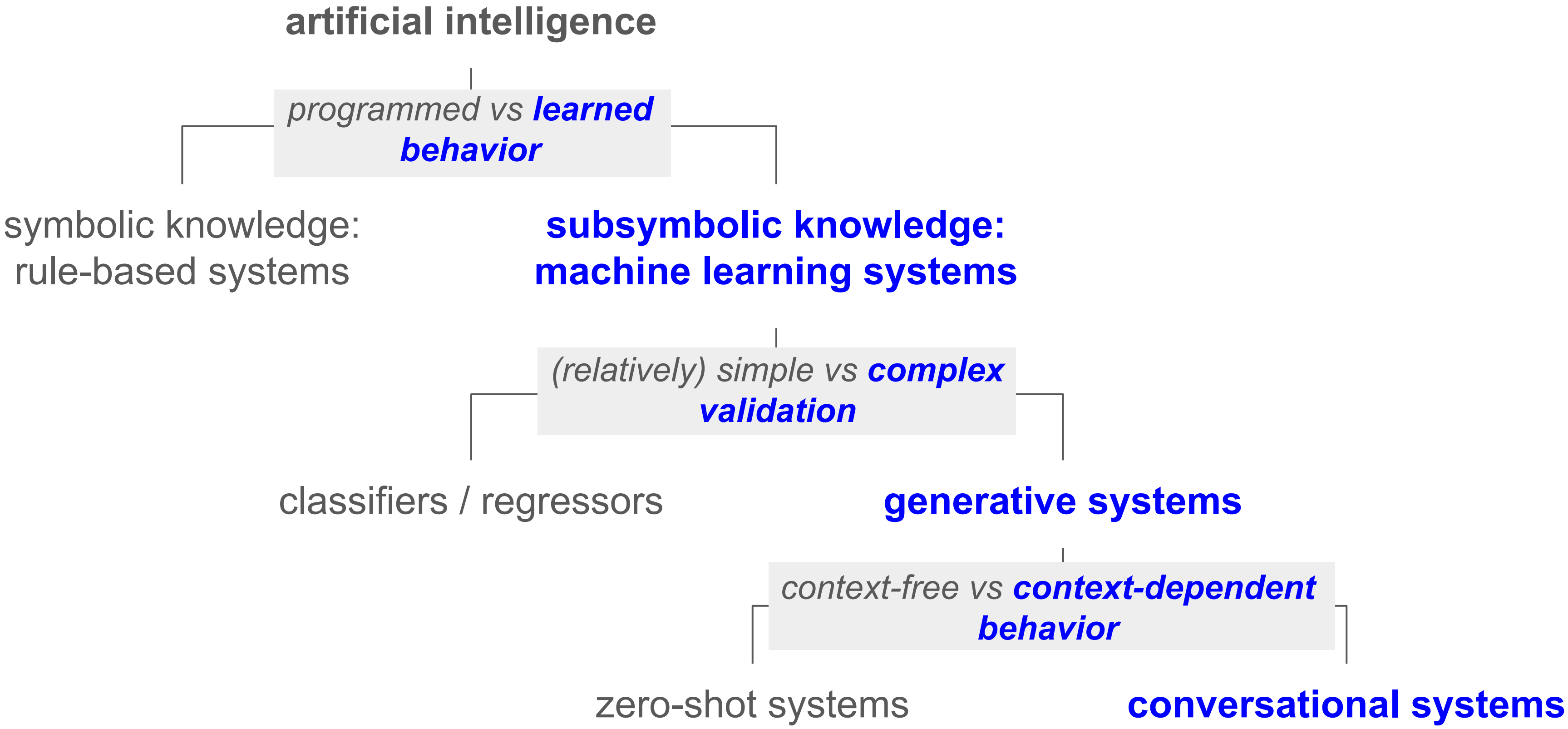


Managing AI Risks in an Era of Rapid Progress

Authors	Affiliations
Yoshua Bengio	Mila - Quebec AI Institute, Université de Montréal, Canada CIFAR AI Chair
Geoffrey Hinton	University of Toronto, Vector Institute
Andrew Yao	Tsinghua University
Dawn Song	University of California, Berkeley
Pieter Abbeel	University of California, Berkeley
Yuval Noah Harari	The Hebrew University of Jerusalem, Department of History
Ya-Qin Zhang	Tsinghua University
Lan Xue	Tsinghua University, Institute for AI International Governance
Shai Shalev-Shwartz	The Hebrew University of Jerusalem
Gillian Hadfield	University of Toronto, SR Institute for Technology and Society, Vector Institute
Jeff Clune	University of British Columbia, Canada CIFAR AI Chair, Vector Institute
Tegan Maharaj	University of Toronto, Vector Institute
Frank Hutter	University of Freiburg
Atılım Güneş Baydin	University of Oxford
Sheila McIlraith	University of Toronto, Vector Institute
Qiqi Gao	East China University of Political Science and Law
Ashwin Acharya	Institute for AI Policy and Strategy
David Krueger	University of Cambridge
Anca Dragan	University of California, Berkeley
Philip Torr	University of Oxford
Stuart Russell	University of California, Berkeley
Daniel Kahneman	Princeton University, School of Public and International Affairs
Jan Brauner*	University of Oxford
Sören Mindermann*	Mila - Quebec AI Institute, Université de Montréal

PUBLISHED	ARXIV
October 24, 2023	Forthcoming.

The context (a candidate classification)



artificial intelligence

programmed vs ***learned behavior***

symbolic knowledge: rule-based systems

subsymbolic knowledge: machine learning systems

(relatively) simple vs ***complex validation***

classifiers / regressors

generative systems

context-free vs ***context-dependent behavior***

zero-shot systems

conversational systems

The example of a conversation

(ChatGPT, with GPT4, October 2023)

Luca: Do you have any knowledge of the book "Measurement across the sciences", by Luca Mari, Mark Wilson, and Andrew Maul, published by Springer Nature?

ChatGPT: As of my last knowledge update in September 2023, I don't have access to specific information on a book titled "Measurement across the sciences" by Luca Mari, Mark Wilson, and Andrew Maul published by Springer Nature. It's possible that the book is a recent publication or that it hasn't been widely covered in publicly available sources. I recommend checking Springer Nature's website or other book retailers for the most up-to-date information.

The example of another conversation

Chatting with an AI… (*not edited*)
A conversation about
the contents of a scientific book

The entity with which we had this conversation:

- writes a good English, and other languages
- produces original texts
- fulfills complex requests
- adapts its arguments to the context
- proposes creative contents
- analyzes and summarizes long texts
- shows sophisticated linguistic skills
- …

The novelty is not in **what** it knows, but in **how** it (knows and) interacts

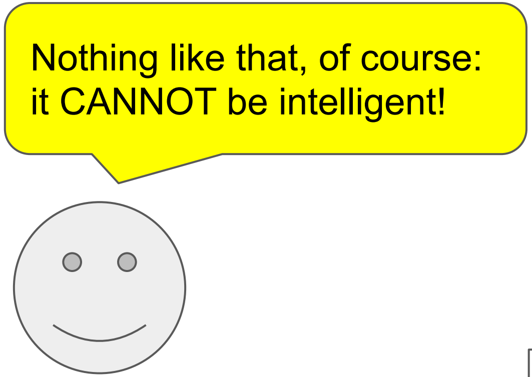
A position

It is the first time that we can have conversations in natural languages
with an entity which does not belong to our species

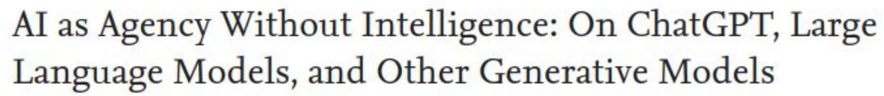
This new scenario is generating and will generate
both opportunities and risks

Hypothesis: what is happening around ChatGPT & its siblings
will be the third “***cultural revolution***” in the Western world:
– Copernicus showed us our **cosmological** non-centrality
– Darwin showed us our **biological** non-originality
– chatbots are showing us our **cognitive** non-uniqueness

Oh c’mon…

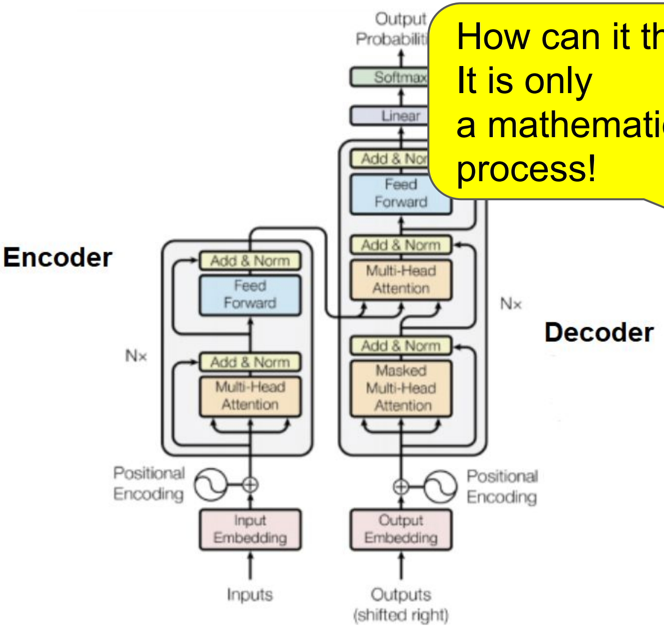


For example:


AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models

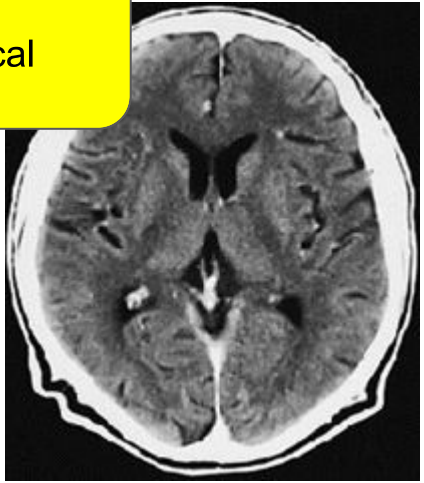
10 March, Philosophy & Technology, https://link.springer.com/article/10.1007/s13347-023-00621-y

An interpretation…



How can it think?
It is only
a mathematical
process!

How can it think?
It is only
an electrochemical
process!



(be either always or never reductionist!)

https://arxiv.org/abs/1706.03762

Issues

Current chatbots produce texts that are the outcome of **autonomous processing**, from a large amount of texts, not of searches / queries in databases

This makes them novel entities, able to operate in original and sophisticated ways but:
- **not always *trustable*** in the factual information they report
- **usually not *explainable*** in their behavior
- **never *accountable*** for what they produce

In my (preliminary) opinion,
- *while their* ***trustability*** will be significantly improved
- *and their* ***explainability*** will be only marginally improved
- *their* ***accountability*** will remain out of scope

Moral responsibility in front of the human society cannot be delegated to non-human entities

Thanks for your attention

Luca Mari

lmari@liuc.it
https://lmari.github.io