

L'intelligence artificielle

Partie 3: questions de société

Jean-Gabriel Ganascia

13 janvier 2025

Professeur Sorbonne Université,

Chercheur intelligence artificielle au LIP6, EurAI Fellow

ex-président du COMETS (comité d'éthique du CNRS)

Président du comité d'éthique de France Travail

Président du comité d'orientation du CHEC (Cycle des Hautes Études de la Culture)

Jean-Gabriel.Ganascia@lip6.fr

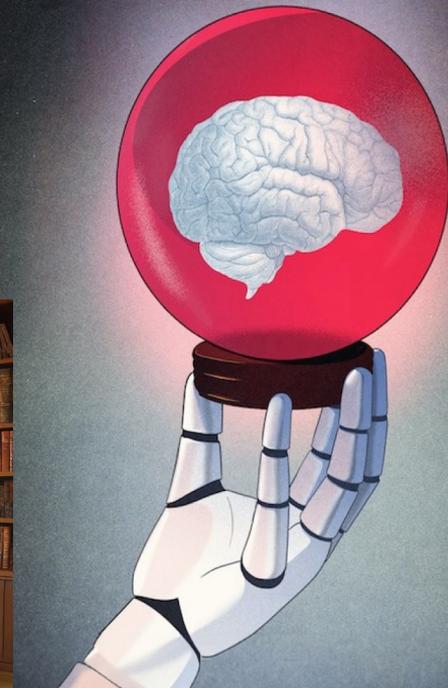
Synoptique

1. La langue d'Ésope
2. La régulation suffit-elle?
3. Visions d'apocalypse
4. « Charité ensorcelée »



JEAN-GABRIEL GANASCIA

L'IA. expliquée aux humains



Seuil

B.



1

LA LANGUE D'ÉSOPE...

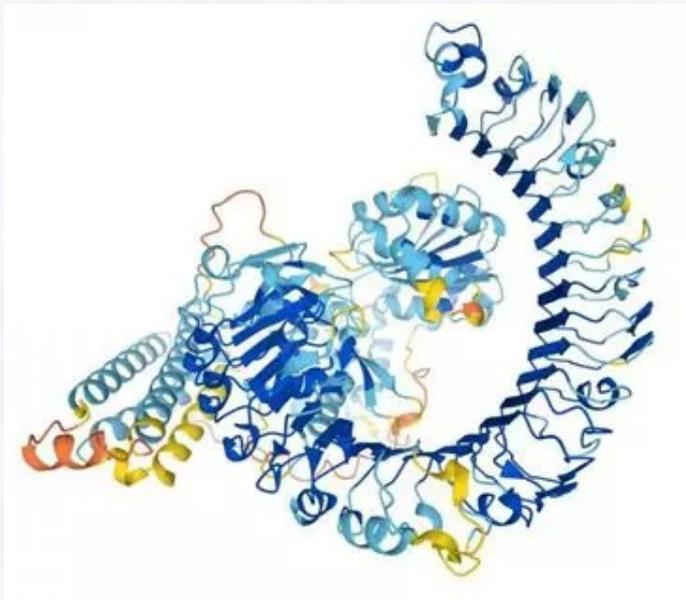
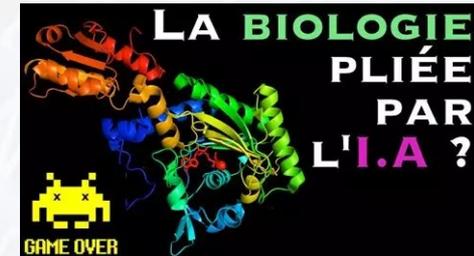
Prix Nobel de chimie 2024

Demis Hassabis (Société Google DeepMind)

John Jumper (Programme AlphaFold2)

David Boulanger (Univ. Washington)

Résolution du problème du repliement des protéines



Dystopie éducative: détection de l'attention – interfaces cerveau ordinateur



Responsabilité juridique des agents conversationnelles

Anecdote

Jake Moffatt — novembre 2022

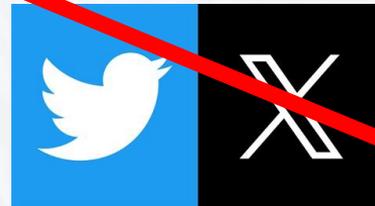
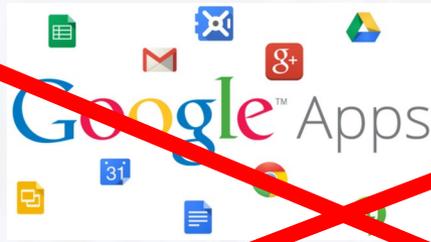
- Jake Moffatt visite le site d'Air Canada afin d'acheter un billet pour assister aux obsèques de sa grand-mère
- Le chatbot de la compagnie prétendit qu'il y avait des réductions pour les personnes achetant un billet au dernier moment lorsqu'elles justifient d'un deuil, **à condition de se manifester dans les 9 semaines qui suivent l'achat**
- Lorsque Jake Moffatt souhaita se faire rembourser, la compagnie lui expliqua qu'**il aurait fallu se manifester au moment de l'achat**
- L'avocat de la compagnie prétendit que « *le chatbot est une entité légale à part, responsable de ses propres actions.* »
- Les magistrats ne l'ont pas suivi... La compagnie a dû rembourser. Elle a retiré le chatbot de son site!



AIR CANADA

Risque pour les libertés — Chine 2024!

- Plus de billets de banque ni de cartes de crédit
- Pas d'accès au Monde
- Plus de Google, plus de moteur, sauf Baidu
- Plus de WhatsApp, X, LinkedIn, ... sauf WeChat, mais il faut être parrainé



Allemagne mars 2024 — PimEyes

- Utilisation de PimEyes + logiciel de vieillissement pour retrouver une terroriste Fraction Armée Rouge (Daniela Klette) 30 ans après...



« Deep fake » (ou vidéotox) Macron



Infox Question ancienne — Dimensions nouvelles

Question ancienne

- « Art de la guerre », *Sun Tzu*
- Libelles
- Canards
- Propagande
- Désinformation
- ...

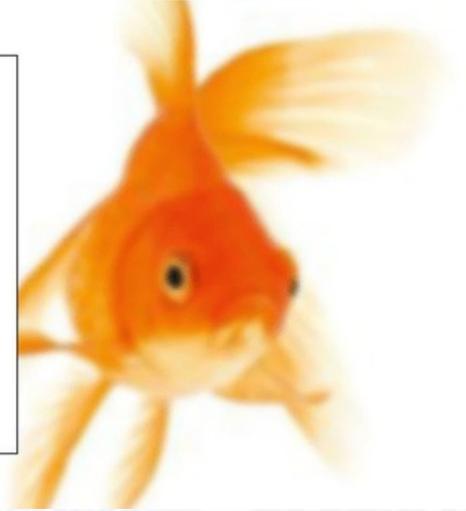
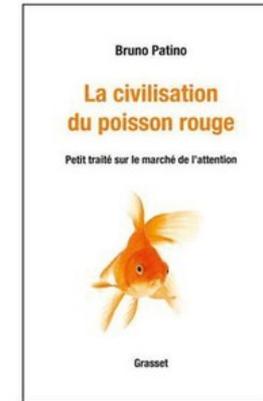


Dimensions nouvelles

- Production massive, à la fois industrielle et accessible à tous.
- **Sémiotique de l'infox: « indisigne »**
 - Typologie des signes par C. Peirce (1839-1914): symbole, légisigne, indisigne
- **Profilage: criblage du public avec l'IA et adressage des infox**



La civilisation des poissons rouges!



Les vieux médias tendaient à rendre les sociétés plus monolithiques
Les nouveaux médias sociaux accroissent la fragmentation de société déjà fragmentées: chacun dans son bocal...
→ disparition de l'espace public!



2

LA RÉGULATION SUFFIT-ELLE?

Régulation frénétique du numérique!

Accélération depuis décembre 2019

Opposition de la commission aux États membres



Régulation européenne sur l'IA et anticipation

Risk level	Examples of AI systems	Allowed in the EU?
Unacceptable risk	Social scoring used by governments; toys using voice assistance which encourages dangerous behaviour	No
High risk	Scoring of exams; AI application in robot-assisted surgery; verification of authenticity of travel documents	Yes, subject to mandatory requirements, ex-ante and ex-post enforcement
Limited risk	Chatbots; "deep fake" videos	Yes, subject to transparency requirements
Minimal risk	AI-enabled video games; spam filters	Yes

Risque inadmissible:

- i. techniques subliminales!
- ii. systèmes de notation sociale
 - Exception pour les banques et les sociétés privées
 - Ex. permis à point!
- iii. biométrie en temps réel

Risque faible

- iv. filtre à spam
- v. jeux vidéo



Chatbot:

- vi. risque modéré à faible (en 2022)
- vii. introduction d'une réglementation spécifique et très stricte pour les LLMs en 2023
- viii. risque dépendant des applications

Va-t-on vraiment mettre en œuvre le AI-Act?



James Vicary admitted making false claims about subliminal advertising in 1957 forcing many people to think subliminal messages didn't work!



Ne pas confondre éthique, droit et normes

• Droit

• Ensemble de lois

- Vote loi dans les démocraties (assemblées)

(distinction lois naturelles et lois humaines)

• Application de la loi: le permis et le défendu

• Autorité de la loi: sanction



Attributs: le glaive, la balance et le bandeau...

from Net Politics, Digital and Cyberspace Policy Program, and Renewing America

The Importance of International Norms in Artificial Intelligence Ethics

Artificial intelligence has arrived as a multi-purpose tool. The United States and its allies need to do more to establish norms and ensure AI is used in a way that does not harm human rights.



The leaders of the Group of Seven countries stand for a photo during the G7 leaders summit in Germany in June 2022 Jonathan Ernst/Reuters

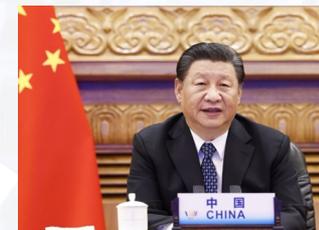
• Règlementation (« régulation »)

- Règles administratives qui clarifient les lois

• Normes

- Normes: règles obligatoires qui ne proviennent pas nécessairement de la loi

- (ex. règles industrielles, environnementales)
- En 2020, Annonce par Xi Jinping de la première norme évolutive liée à la 5G qui souligne l'intégration de la sagesse chinoise dans les normes internationales

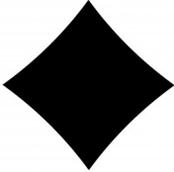


3

VISIONS D'APOCALYPSE

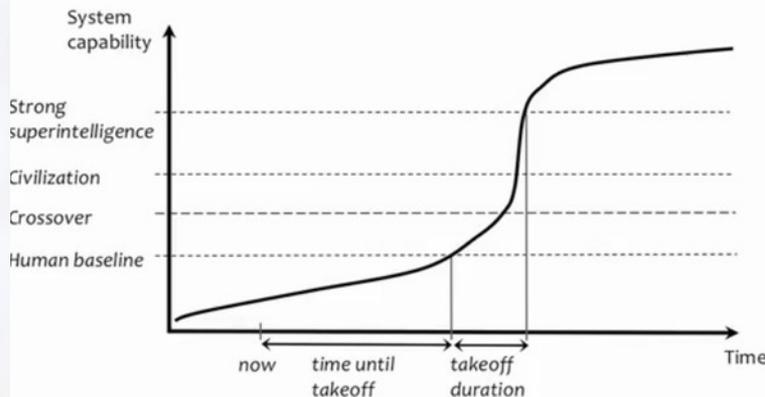
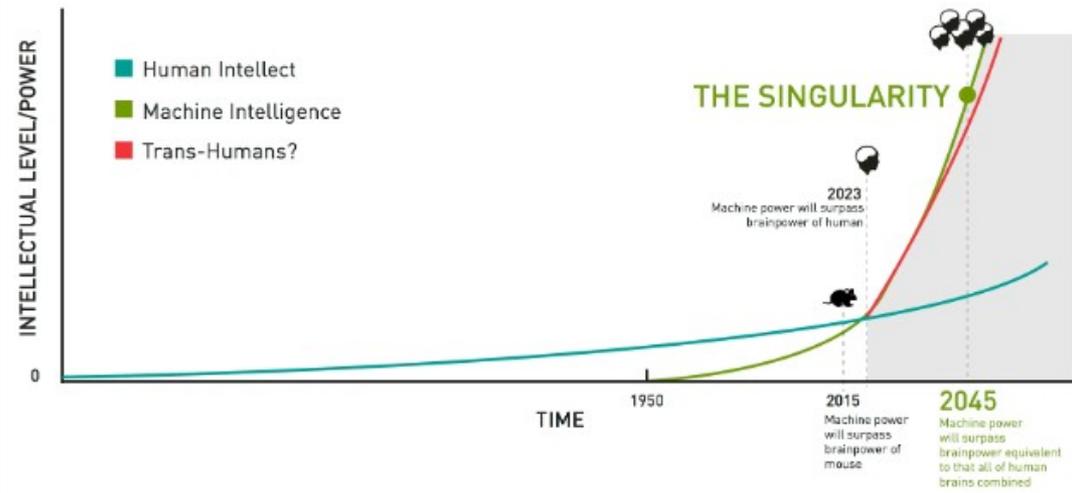
Super-intelligence et Singularité

- Advenue d'une intelligence supérieure

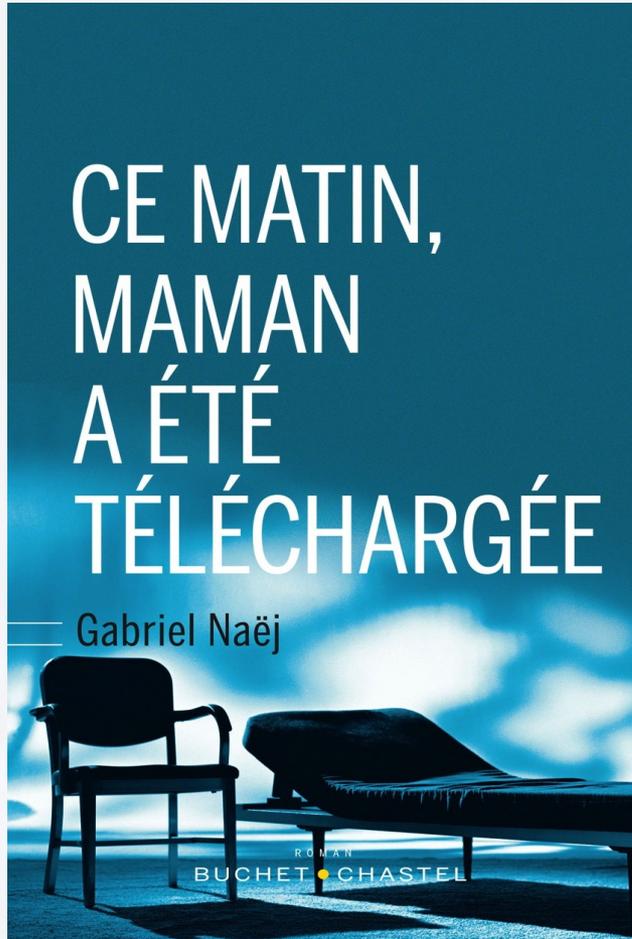

Future
of Humanity
Institute
UNIVERSITY OF OXFORD



SINGULARITY TIMELINE
Rise in human intellect could be driven by integrating with machines in the future



Autobiographie de Gabriel Naëj



La Singularité est advenue

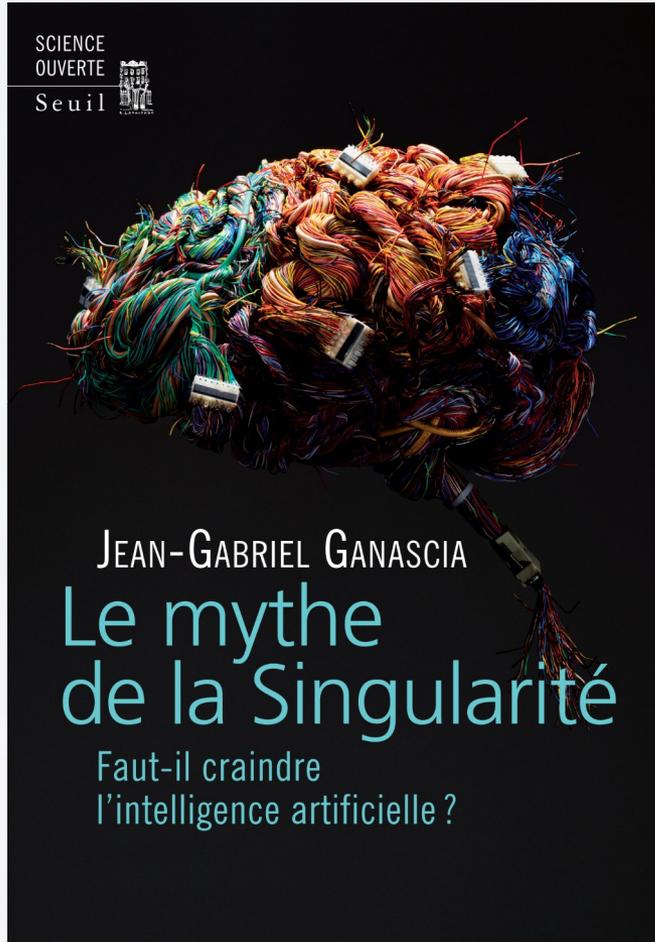
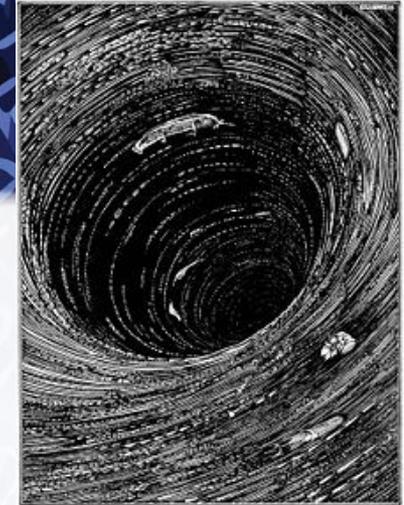
Et, comble de malheur pour lui, sa maman, sa vraie mère décide de procéder à une métempsychose* numérique.

Et en plus, elle s'est réincarnée dans un robot sexuel ce qui le trouble beaucoup

Et, elle fait irruption à tout moment dans sa vie...

* passage de l'âme d'un corps dans un autre

La Singularité technologique



Réponses:

- **La loi de Moore**
- Limitations de la loi de Moore
- Puissance de calcul \neq intelligence (au sens d'esprit...)



4

« CHARITÉ ENSORCELÉE »

Une saison en enfer, Arthur Rimbaud
Délires, I, Vierge folle, L'époux infernal

« Il a peut-être des secrets pour *changer la vie* ? Non, il ne fait qu'en chercher, me répliquais-je. Enfin sa charité est ensorcelée, et j'en suis la prisonnière. »

Pétition!

fin mai 2023

Geoffrey Hinton

Emeritus Professor of Computer Science, University of Toronto

Yoshua Bengio

Professor of Computer Science, U. Montreal / Mila

Demis Hassabis

CEO, Google DeepMind

Sam Altman

CEO, OpenAI

Dario Amodei

CEO, Anthropic

Dawn Song

Professor of Computer Science, UC Berkeley

Ted Lieu

Congressman, US House of Representatives

Bill Gates

Gates Ventures

Ya-Qin Zhang

Professor and Dean, AIR, Tsinghua University

Ilya Sutskever

Co-Founder and Chief Scientist, OpenAI

Shane Legg

Chief AGI Scientist and Co-Founder, Google DeepMind

Martin Hellman

Professor Emeritus of Electrical Engineering, Stanford

James Manyika

SVP, Research, Technology & Society, Google Alphabet

Yi Zeng

Professor and Director of Brain-Inspired Cognitive AI Lab, Institute of Automation, Chinese Academy of Sciences

Xianyuan Zhan

Assistant Professor, Tsinghua University

Albert Efimov

Chief of Research, Russian Association of Artificial Intelligence

Premiers signataires

Les pompiers pyromanes!

Quelle est leur finalité?



Center for
AI Safety



AI experts, journalists, policymakers, and the public are increasingly discussing a broad spectrum of important and urgent risks from AI. Even so, it can be difficult to voice concerns about some of advanced AI's most severe risks. The succinct statement below aims to overcome this obstacle and open up discussion. It is also meant to create common knowledge of the growing number of experts and public figures who also take some of advanced AI's most severe risks seriously.

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

La « charité ensorcelée » des géants du web



Diversity and Inclusion

IBM CEO's Letter to Congress on Racial Justice Reform



IBM annonce qu'il se retire du marché de la reconnaissance faciale étant donné qu'elle sert au profilage racial et appelle à une réforme de la police *Le 9 juin 2020*

Selon des documents examinés par Liberty Investigations et The Verge, IBM a signé un contrat de 69,8 millions de dollars avec le gouvernement britannique pour créer une plateforme biométrique dotée d'une fonctionnalité de reconnaissance faciale le mois dernier. *Le 1^{er} septembre 2023*

“Over time I think we will probably see a closer merger of biological intelligence and digital intelligence.”

1. Annonce révélatrice d'un projet démoniaque!
2. Mensonge sur sa réalisation
3. Pouvoir d'influence exercé avec la complicité des médias



Elon Musk Neuralink & l'esprit de la ruche

- “Over time I think we will probably see a **closer merger of biological intelligence and digital intelligence.**”
- “it's mostly about the bandwidth, the speed of the connection between your brain and **the digital version of yourself,** particularly output. ”

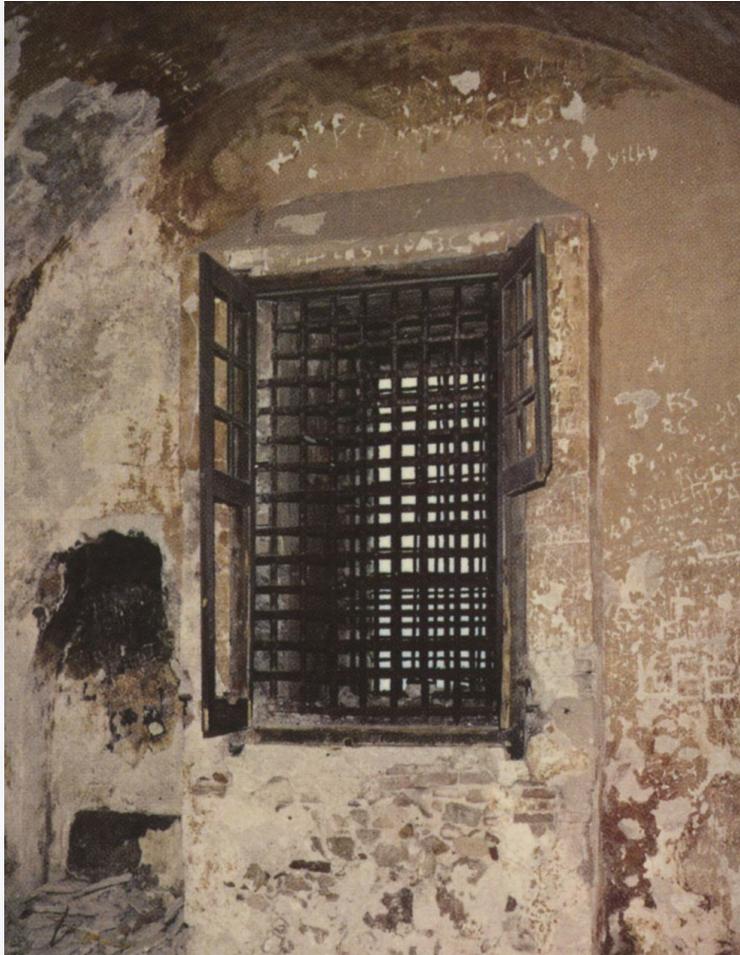
1. **Annonce révélatrice d'un projet proprement démoniaque!**
2. **Mensonge sur sa réalisation**
3. **Pouvoir d'influence exercé avec la complicité des médias**



L'esprit de la ruche: Collaboration des esprits qui partagent les connaissances et les opinions les uns avec les autres

La "dentelle neurale" (I. Bank)

« Servitude volontaire » E. de La Boétie



« Le théâtre, les jeux, les farces, les spectacles, les gladiateurs, les bêtes curieuses, les médailles, les tableaux et autres drogues de cette espèce étaient pour les peuples anciens les appâts de la servitude, le prix de leur liberté ravie, les outils de la tyrannie. Ce moyen, cette pratique, ces allèchements étaient ceux qu'employaient les anciens tyrans pour endormir leurs sujets sous le joug. [...] **Mais ils ne font guère mieux ceux d'aujourd'hui qui, avant de commettre leurs crimes les plus graves, les font toujours précéder de quelques jolis discours sur le bien public et le soulagement des malheureux** »

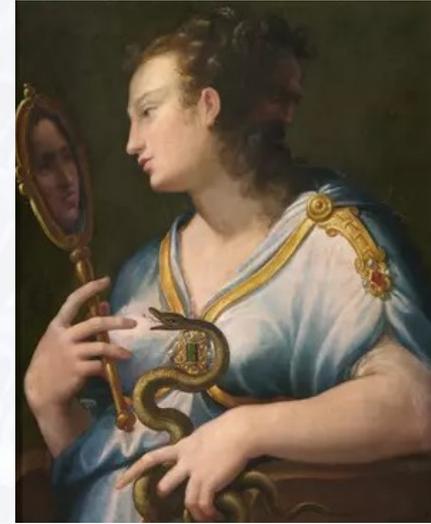
« Don't be Evil! » Google

Rachat de Twitter par Elon Musk pour garantir la liberté d'expression!

Appel au moratoire sur l'intelligence artificielle par ceux qui la produisent!...

L'éthique, c'est accueillir ce (et celui) qui vient!

L'éthique, ce n'est pas la seule prudence (connaissance et expérience), ce n'est pas la seule loi (Thémis), c'est aussi et d'abord l'ouverture au présent

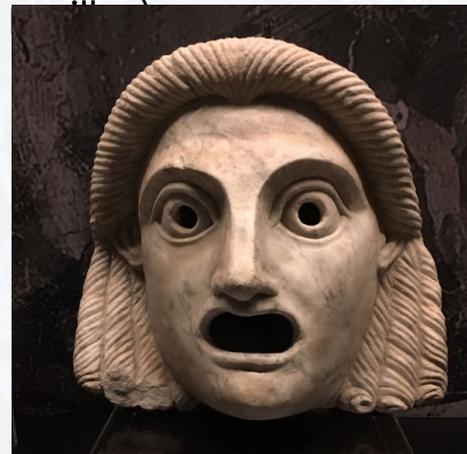


L'éthique se pense face au chaos, au présent, à ce qui arrive inopinément:

Théorie mathématique du chaos → étude des systèmes dynamiques instables, et, par là imprévisibles parce très sensible aux conditions initiales (effet papillon)

Jacques Derrida: le chaos, c'est la
« *forme de tout avenir comme tel, de tout ce qui vient* »

« *l'abîme ouvert du Khaos, c'est aussi la forme ouverte et béante de ma bouche (Khainô) quand je ne sais plus quoi dire* ».



PENSER CE QUI VIENT

Jacques Derrida

Derrida
pour les temps
à venir



sous la direction de René Major

SCIENCE
OUVERTE



Seuil

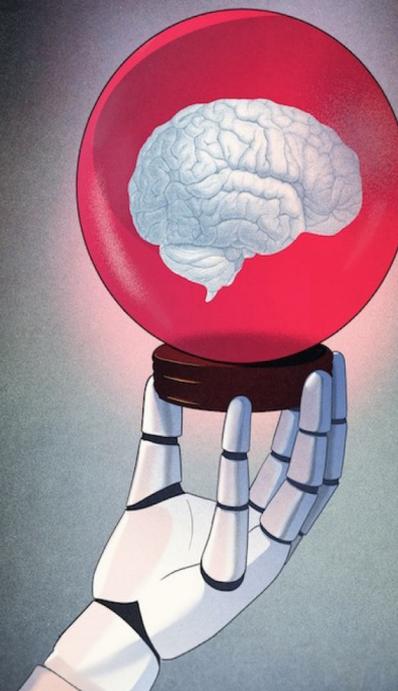


JEAN-GABRIEL
GANASCIA
Servitudes
virtuelles

Merci!

JEAN-GABRIEL GANASCIA

L'IA expliquée
aux humains



Seuil