Ethics of Artificial Intelligence
&
Large Scale Machine Learning

Jean-Gabriel Ganascia

28 November 2022

Sorbonne University, Head of ACASA/LIP6 team (computer science lab)

Ex-Chairman of the COMETS (CNRS Ethics Committee)

Member of CNPEN – Chairman of Pôle Emploi Ethics Committee

Jean-Gabriel.Ganascia@lip6.fr

SCIENCES SORBONNE UNIVERSITÉ

LIP6

cnrs

Synoptic

1. Two Characters from Mythology

2. Why is There a Problem?

3. How to Solve the Problem?

4. Regulation is not the Solution!

5. Trustworthiness of AI Systems

6. Ethics and Chaos

SORBONNE
UNIVERSITÉ

1

TWO CHARACTERS FROM MYTHOLOGY

Ethics of AI & Large Scale ML - Jean-Gabriel Ganascia

SCIENCES
SORBONNE
UNIVERSITÉ

LIP6

cnrs
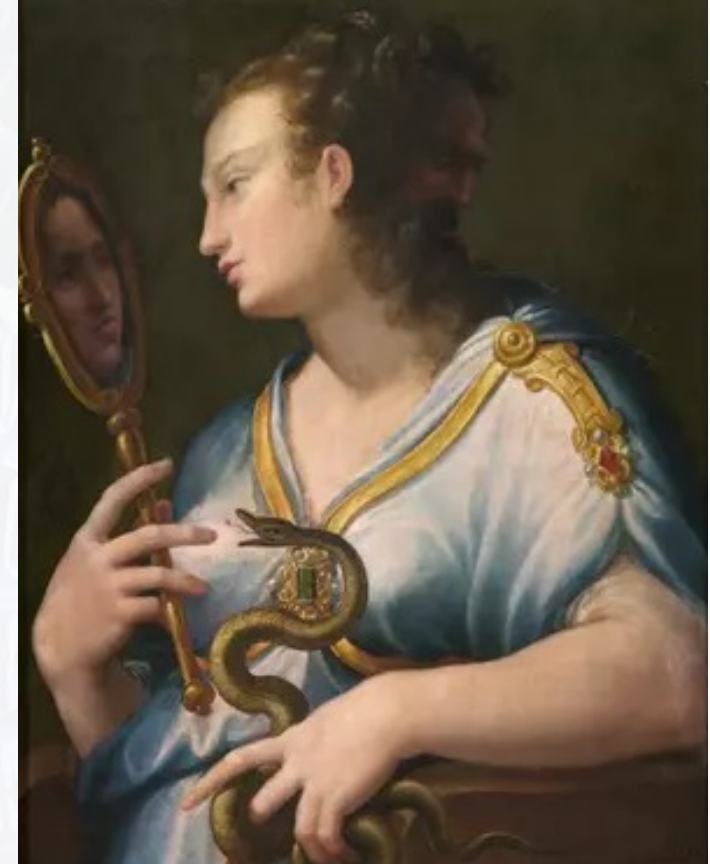
Two Characters from Mythology: Prudence & Themis

Themis

Allegory of the **Justice**

Two symbols:

- she Carries a **Balance**, Sign of Equity
- she wears a **Blindfold** as a Token of Impartiality

Blind Application of <u>Laws</u>

Prudence

Personification of the **Virtue of Prudence**

Two symbols:

- she looks both the past in **a mirror** and the future ahead
- the **Snake** represent the knowledge

Anticipation and wisdom are drawn from **<u>lessons of experience</u>**
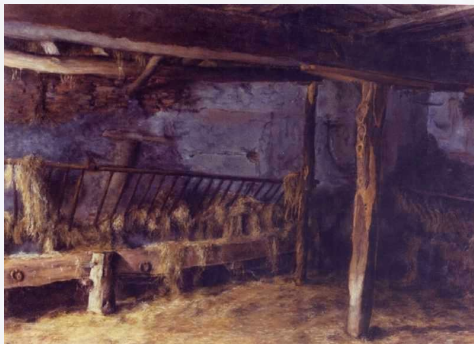
Moral and Ethics

Ethics: Latin ethica; Greek *êthikos*, *êthikê*, from *êthos*, 'custom', 'mores'

Originally, in Greek, *êthos* meant a **place familiar to animals**, e.g. a stable.

With Aristotle, means the **rational deliberation** necessary to act well.
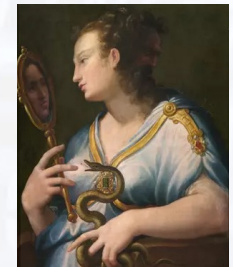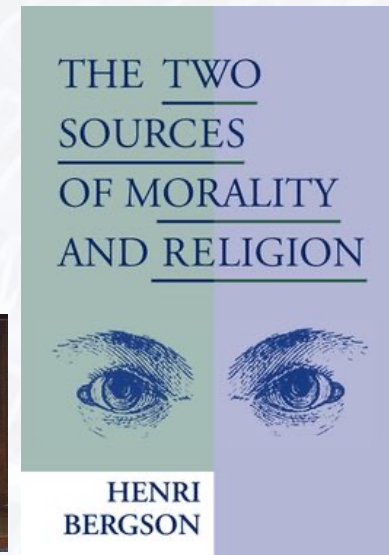
Moral : Latin *moralis* from *mores* → Mores

Approaches:
- *Traditions, precepts*
- *Principles*:
  - **Deontism**: Themis
  - **Utilitarianism**: prudence

The **art (or the science?) of directing one's conduct**
- **Experience**, **traditions**
- **Rationality**, **laws** and **rules**

THE TWO SOURCES OF MORALITY AND RELIGION

HENRI BERGSON

Two Characters from Mythology: Prudence & Themis

Themis

Blind Application of Laws

Symbolic AI Approach

Prudence

Anticipation and wisdom are drawn from **lessons of experience**

Machine Learning Approach

SORBONNE
UNIVERSITÉ

2

WHY IS THERE A PROBLEM

LIP

cnrs

Ethics of AI & Large Scale ML - Jean-Gabriel Ganascia

Examples of Laudable Applications of AI

Example: Health Sciences

Processing Huge Masses of Medical Data

- Extracting medical knowledge from patient data (X-rays, clinical signs, etc.)
- Extracting biological information (e.g. genetic factors explaining the evolution of the disease, etc.)

Bioinformatics

- Modeling biological processes (e.g. mechanism of introduction of the virus into cells, genetic factors explaining the evolution of the disease, etc.)

Extraction of Knowledge from the Scientific Literature

- More than 87,000 papers on CoViD-19 were produced between March and October 2020!

Robotics

- Robotics for the Elderlies
- Surgical Robots and Prothesis

Irresponsible, Unjust and Unfair uses of AI & ML

Use of AI that could infringe human dignity and autonomy

- Surveillance systems that would track every move — social credit in China…

  … and Border Surveillance in US and EU

- Biased AI systems that are discriminatory — facial recognition

- Cast public opprobrium on those who disobey the rules

- Video, speech and/or image synthesis that could produce fake

- AI-based targeting dissemination techniques of these fake news.
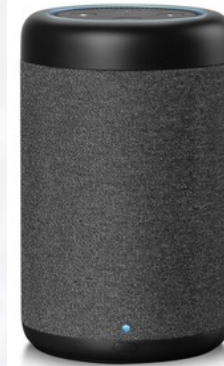
- …
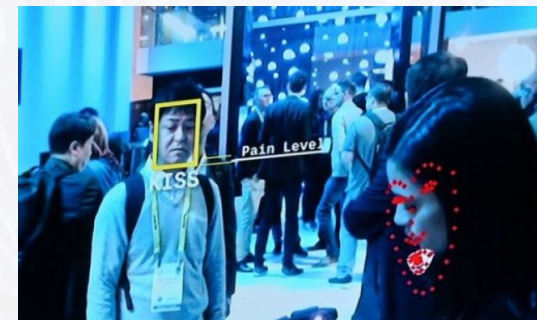
Misuses of Language Generation

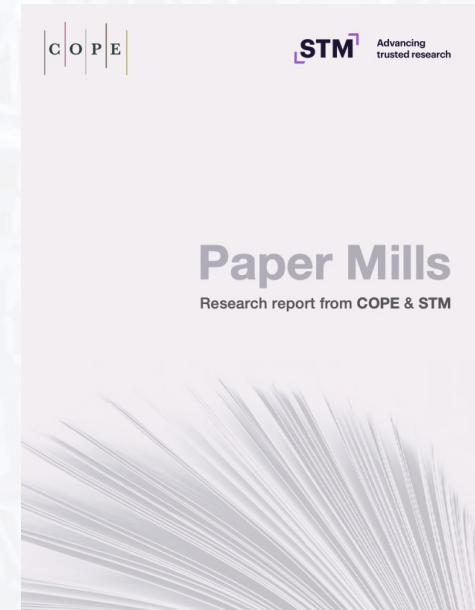- "Chatbots" (dialogues)
  - Domestic spies

  - Industrial disruption

- Stochastic Parrots (*Text Generation based on Language Models (GPT3)*)
  - Paper Mills

  - Bias

SORBONNE
UNIVERSITÉ

Current Application of AI in Robotics

Robots (i.e. "artificial workers"),
bots (i.e. virtual robots)
and automata are extensively used in a great
multiplicities of activities, where they are useful.

But some uses of these technologies
are more debatable as

- Autonomous weapons

- Sexbots

- Chatbots

- Fight Bots

- Influencers

- Politicians

SORBONNE
UNIVERSITÉ

3

HOW TO SOLVE THE PROBLEM?

SCIENCES
SORBONNE
UNIVERSITÉ

LIP

cnrs

W^5 – What Went Wrong and Why Workshop: *Lessons from AI Research and Applications*

Location: Stanford

Participants: John McCarthy, Ed Feigenbaum, Carl Hewitt, Doug Lenat, Manuella Veloso, Mike Pazzani, Craig Knoblock …

Example: "electric elves" are useful and efficient intelligent agents

But, it happened that they awake their owner at 3am to advise him that his 8am plane should have to be 2 hours delayed…

Lesson:
AI system are not just **technical systems,**
that also **socio-technical systems.**
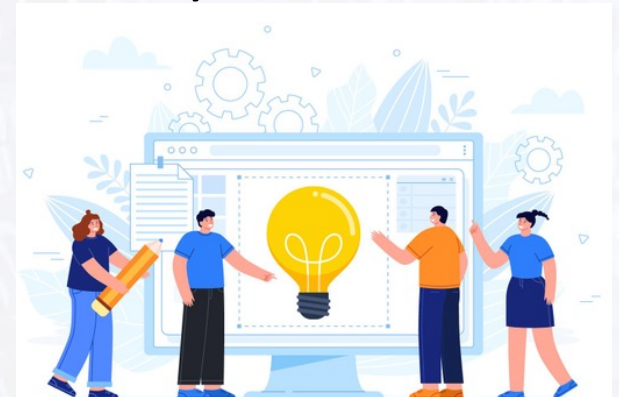
Technical and Sociotechnical Dimensions

Technical devices are there to relieve and help us.

They can be considered in isolation, as such...

And yet, this is not enough to inspire confidence!

Why?

The machines make errors

- How to make them sound?
- How to prove their soundness?

The machines contribute to transform the society

- How to restrict their use?
- How to anticipate and to preclude their misuses?

SORBONNE UNIVERSITÉ

4

REGULATION IS NOT THE SOLUTION!

Ethics of AI & Large Scale ML - Jean-Gabriel Ganascia

Regulation of Artificial Intelligence

The European Commission's

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

DRAFT ETHICS GUIDELINES FOR TRUSTWORTHY AI

Working Document for stakeholders' consultation

Brussels, 18 December 2018

GDPR

1- WELL-BEING PRINCIPLE

The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings.

2- RESPECT FOR AUTONOMY PRINCIPLE

AIS must be developed and used while respecting people's autonomy, and with the goal of increasing people's control over their lives and their surroundings.

3- PROTECTION OF PRIVACY AND INTIMACY PRINCIPLE

SORBONNE UNIVERSITÉ

Based on Human Rights & Bioethics Principles

Trustworthy AI

1. Lawful
2. Ethical
3. Robust

Three layers

1. Principles:
   - Respect for Human Autonomy
   - Prevention of Harms
   - Fairness
   - Explicability
2. Realizing Trustworthy AI
   - **Seven Requirements**: human agency, technical robustness, privacy, transparency, non-discrimination and fairness, societal and environmental well-being, accountability
   - **Technical and non technical methods**
3. Assessing Trustworthy AI

The European Commission's

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

DRAFT ETHICS GUIDELINES FOR TRUSTWORTHY AI

Working Document for stakeholders' consultation

Brussels, 18 December 2018

"Ethics" Committees *(>84 in 2019...)**
More than 67 Principles & recommendations

- The **Asilomar AI Principles**, developed under the auspices of the Future of Life Institute, 2017

- The **Montreal Declaration for Responsible AI**, developed under the auspices of the University of Montreal, 2017

- The **General Principles of *Ethically Aligned Design**: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, IEEE, 2017

- The **Ethical Principles offered in the *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems**,* European Commission's European Group on Ethics in Science and New Technologies, 2018

- The **"five overarching principles for an AI code"** offered in paragraph 417 of the UK House of Lords Artificial Intelligence Committee's report, *AI in the UK: ready, willing and able?*, 2018

- The **Tenets of the Partnership on AI**, a multi-stakeholder organization consisting of academics, researchers, civil society organizations, companies building and utilizing AI technology, 2018

- **20 recommendations** of the AI4People, an Atomium—EISMD initiative designed to lay the foundations for a "Good AI Society", 2018

- **Ethical Guidance for a Trustworthy AI**, High Level Expert Group on Artificial Intelligence, European Commission, 2019

These principles are not only numerous, but also contradictory!

- Privacy vs. Transparency
- Privacy vs. Security
- Lack of discrimination vs. Inclusive policy
- ...

*Brent Mittelstadt, "Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, vol. 1, November 2019, pp. 501–507

SORBONNE UNIVERSITÉ

LIP6

"There are too many principles!"

"It is the plural of good principles that is evil. It is the plural of absolute-relatives that makes all the drama. There are too many principles! Too many people, too many values limiting the infinite value of each value… no doubt, there is nothing wrong with that? But such is however, on this *land of Cockaigne* of the intelligible world, the only conceivable misfortune."

Vladimir Jankélévitch,
Traité des vertus III,
L'innocence et la méchanceté,
p. 103

SORBONNE
UNIVERSITÉ

LIP6

Ethics don't have to be confused with Laws, Regulation and Norms!

THE TWO
SOURCES
OF MORALITY
AND RELIGION

HENRI
BERGSON
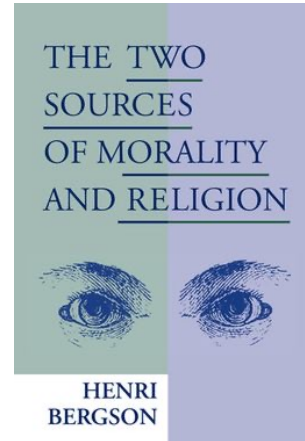
- **Ethics**
  - Art (or science?) of **directing one's conduct**
  - **Based on rational deliberation and traditions**

- **Laws**
  - **Right**: set of human laws
  - **Laws are voted** (Parliaments)
  - **Authority of the law**: sanction
  - **Law enforcement**: what is allowed and what is not

- **Regulation**
  - **Administrative rules** that clarify laws

- **Norms**:
  - **Mandatory rules** that do not necessarily come from the law (e.g. industrial standards, environmental rules)

ISO 22301 CERTIFIED

Norms, Politics and Power

Presence of GAFAMINATUBATX lobbyists in standardization institutions

ETHICS, BUSINESS INDUSTRY, INTEGRITY

SOURCES

SOCIAL MEDIA

FACEBOOK & ETHICS

DATA

REGULATION

Léa M. Sara T.

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

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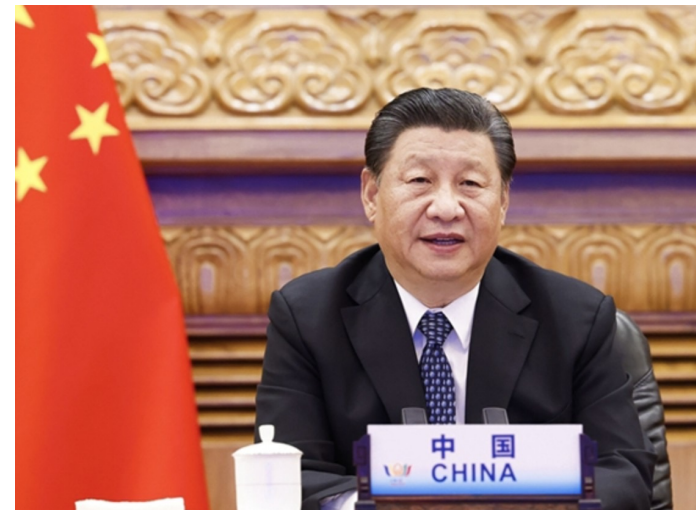
Artificial intelligence has arrived as a multi-purpose tool. The United States and its allies need to do more to establish norms and ensure AI is used in a way that does not harm human rights.

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

Appearance of China in norm building

- XI Jinping

- 5G first evolutionary standard announced **completion of Chinese wisdom into international standards**

- *5G首个演进标准宣布完成 中国智慧融入国际标准),* People's Daily Online, Author: Zhao Chao *(人民网), 4 July 2020*

中国 CHINA

5

TRUSTWORTHINESS OF AI SYSTEMS

SCIENCES
SORBONNE
UNIVERSITÉ

LIP

cnrs

Ethics of AI & Large Scale ML - Jean-Gabriel Ganascia

Making Machines Virtuous – Requirements

- **Respect justice**:
  - Without confusing it to judiciary institution
  - "the just, between the legal and the good" Paul Ricœur

- **Protect privacy**

- **Preserve security**

- **Be transparent**
  - But, without violating privacy and human intimacy

- **Human autonomy**
  - Humans need to remain responsible of their actions

SORBONNE
UNIVERSITÉ

How to Make Machines Virtuous?

- **Make them more sound, robust and "trustworthy"**

- **Prove** their soundness

Add prescriptions to machines in order they:

- Respect justice

- Protect Privacy

- Preserve Security

- Be Transparent

- Preserve Huma Autonomy

Computational Ethics

How to ensure that a program is correct?
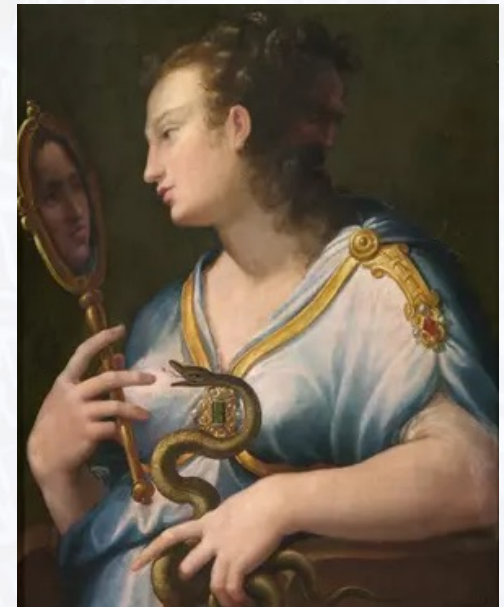
Classical computer science: proof of program

Formal specifications: mathematical description of inputs E, outputs S, and the relationship R(E, S) between inputs and outputs.

Program proof:

- **Prove that for any input E, the output S verifies R(E, S)**
- **Prove that for any input E, the program ends**

However, supervised Machine Learning is inductive and induction is conjectural

What we can do is to anticipate!

SORBONNE UNIVERSITÉ

The Logic of induction

- Induction = Reasoning from the Particular to the General

- Introduction of a "principle of uniformity"

- Construction of an inductive syllogism

What is true for A, B, C, D, … is true for all swans

A, B, C, D, … are white
_____

All swans are white

SORBONNE UNIVERSITÉ

How to ensure that a program is correct?

Classical computer science: proof of program

Formal specifications: mathematical description of inputs E, outputs S, and the relationship R(E, S) between inputs and outputs.
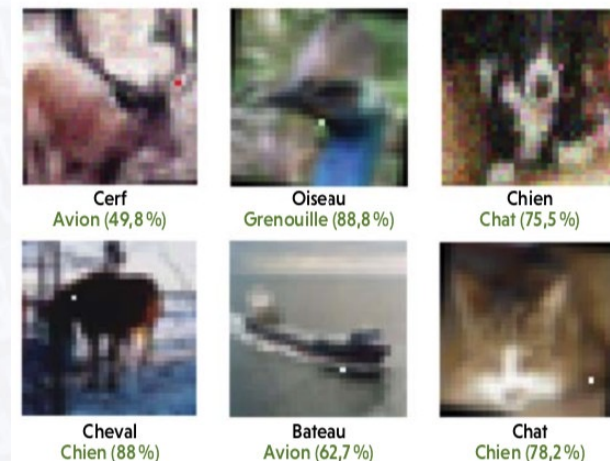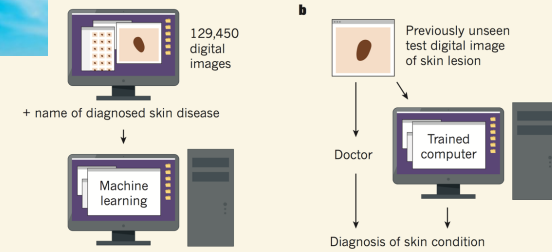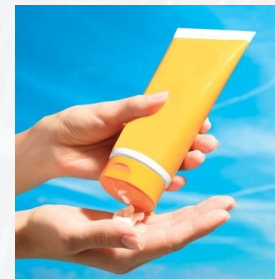
Program proof:

- **Prove that for any input E, the output S verifies R(E, S)**

- **Prove that for any input E, the program ends**

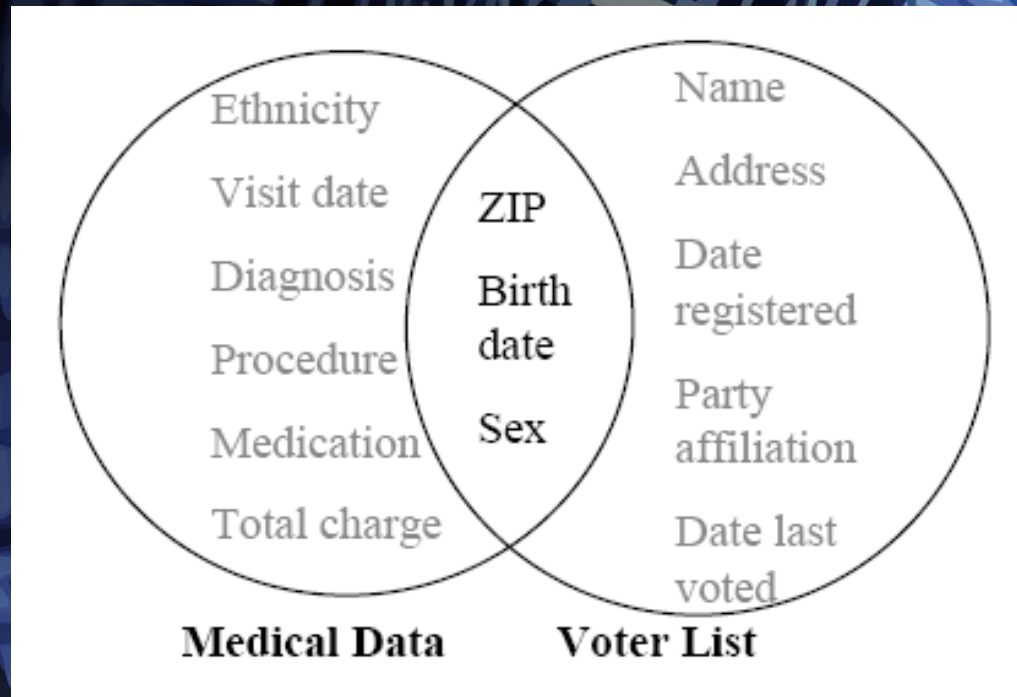## Artificial intelligence based on Inductive Machine Learning: test

No formal specifications, so no proof, but labeled examples!   Test

- Make sure the examples are **well distributed** - no bias – "uniformity principle"

- Ensure the stability of the system by introducing **perturbations** on the examples, and by checking the output

- **Semantics, explanation**:
  - i.e. study, in the images, the areas that affect the conclusion

Protect Privacy

Anonymization ≠ Pseudonymization

Medical Data: Ethnicity, Visit date, Diagnosis, Procedure, Medication, Total charge

Voter List: Name, Address, Date registered, Party affiliation, Date last voted

ZIP, Birth date, Sex

BL_Institut

Principles

- Camouflage
- Adding noise - i.e. false information
- Concealment in a crowd
- K-anonymization

- Cryptography

Other technical issues

- Enhancing accuracy

- Be transparent: explainability (XAI)
- Respect justice: fairness

Building "Ethical" Artificial Agent

Classical Kantian distinction between

- Acting *from duty*

and

- Acting *in accordance with duty*

- "Ethical" Artificial Agents are only acting *in accordance with duty*, because they have no proper motivation

SORBONNE
UNIVERSITÉ

Checking compliance to norms
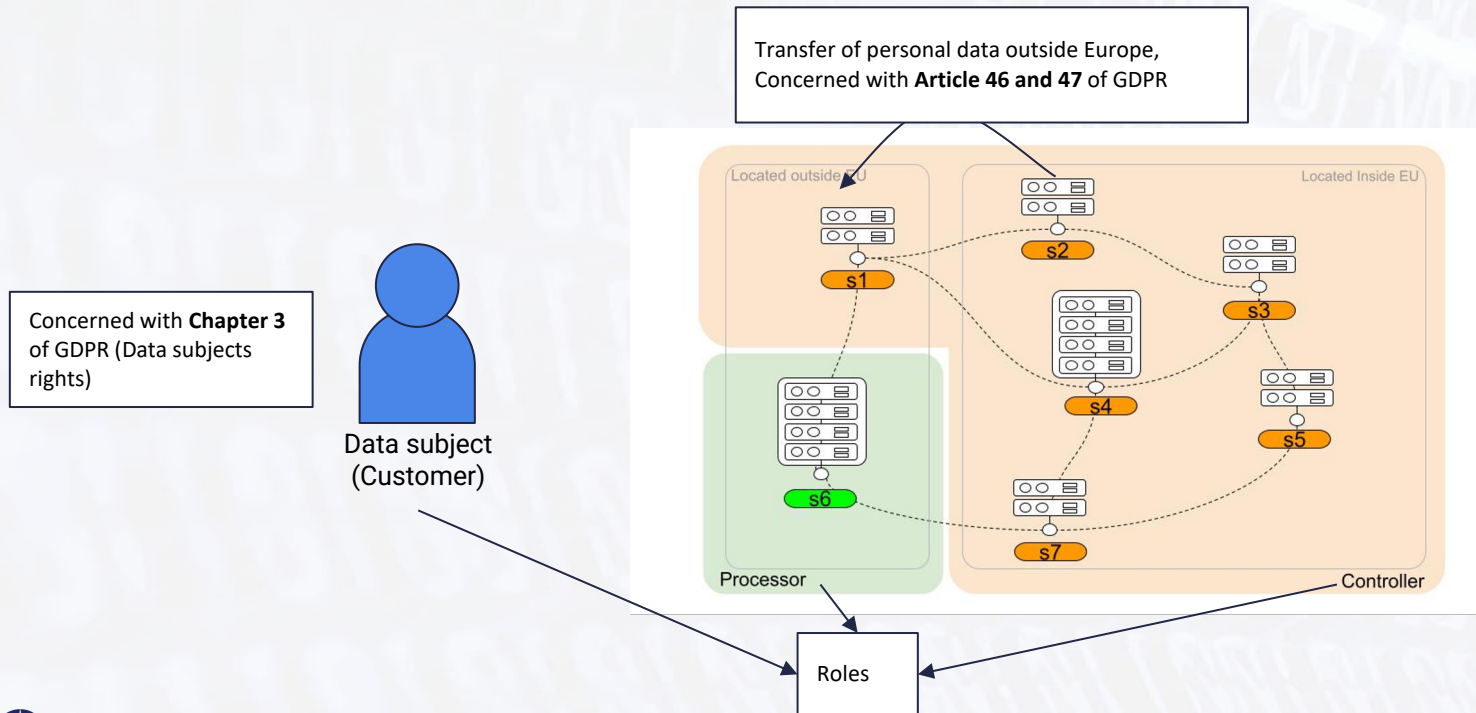
Domain: data manipulation – GDPR

An **international European company** operates in multiple **EU** countries and as well as **US**.

Each sector owns a **server** only for **storing personal data** e.g. s1, s3,...

The servers are connected through an internal network and can transmit data among each other.

Two of the servers are **data processors** (**S4, S6**) in which the company analyses customers data.

A customer is **data subject** who has given her consent for a series of processing

Transfer of personal data outside Europe, Concerned with **Article 46 and 47** of GDPR

Analyzing personal data concerns with **Article 6** lawfulness

Concerned with **Chapter 3** of GDPR (Data subjects rights)

Data subject (Customer)

Roles

SORBONNE UNIVERSITÉ

Data Manipulation Planning & Legal and Ethical Compliance Checking

Automatic data manipulations

+

GDPR compliance checking

Requires:
A logical formalism for representing **data manipulation operators**, their **effects** and **preconditions**

Requires:
An **ontology** or a **policy language** for representing various GDPR requirements, e.g. data subject's consent, regulatory norms etc.

Challenges:
Find the proper formalism to **handle automatic data processing**

Challenges:
Integrate the two so that we achieve automatic data processing and compliance checking

Challenges:
Choose the proper **representation of GDPR** norms that supports automatic compliance checking in our domain

Research on Realtime Compliance Mechanism for AI (RECOMP)
an International Project (France – Germany - Japan)

SORBONNE UNIVERSITÉ

LIP6

Adding prescription to program
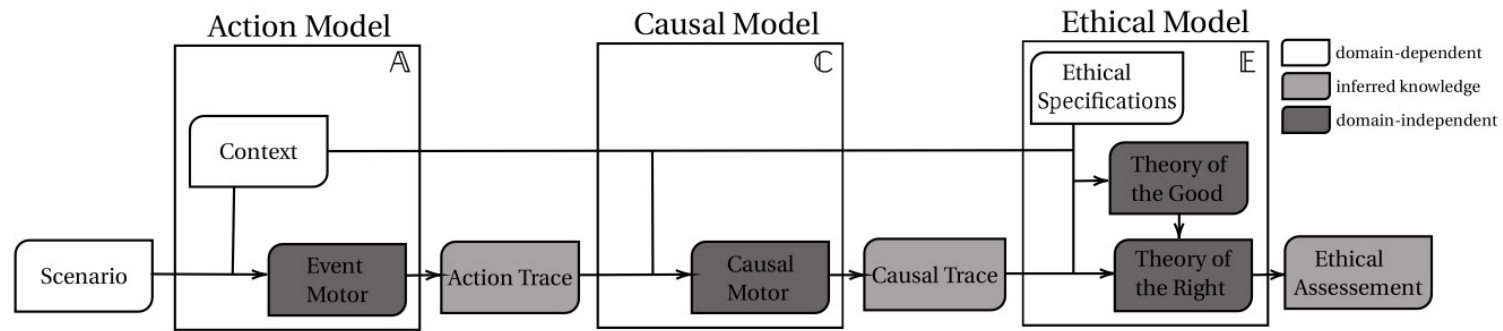
Automating deliberation: three challenges

1. Need to consider the **consequences of actions** that are reasonable to anticipate.
   - Introduce a <u>causal model</u> of the consequences of actions

2. Taking into consideration the **rules of duty**, i.e. *obligations, permissions, prohibitions* etc.
   - Use of <u>modal logics</u> with deontic modalities.

3. Overcoming **ethical dilemmas**, i.e. conflicts of norms
   - Use of <u>non-monotonic formalisms</u> designed to get through logical contradictions

A Sad Example

Was Alec Baldwin responsible?

- *Difference between causality and imputability*

*act(P,S,**G**,A)   person(P),*

situation(S), goal(G), action(A),

will(P, S, G), ← "Prudence": pragmatic imperative

solve_goal(P, S, G, A), ← Sagacity: problematic imperative

maxim(P, S, A). ← Morality: moral imperative

act(P, S, G, A), act(P, S, G, B), A   B.

ACASA team pursues this goal in different directions:
- **causality**
- **explanations**
- **…**

6

ETHICS AND CHAOS

Ethics of AI & Large Scale ML - Jean-Gabriel Ganascia

Prudence: anticipation of the future

Allegory of prudence: the woman with the mirror who looks at both the future and the past… "Ethics by design" - *a priori* ethics
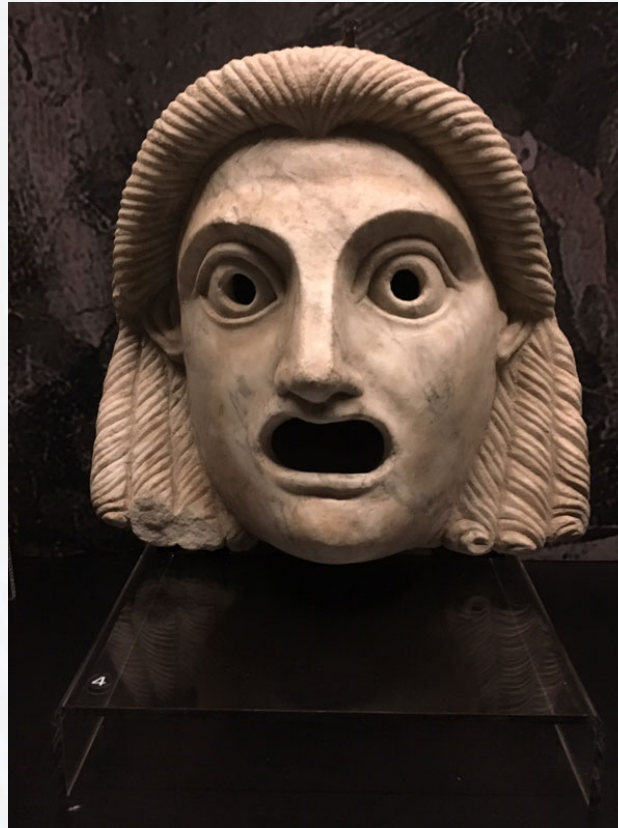
Ethics is not only *prudence*, it is also *openness* to "what (or who) is coming"...

Chaos:

Mathematical Theory   Unpredictible Dynamic Systems

Jacques Derrida: "*The chaos,
it is the form of all future
as such, of all that comes*"

"*The open abyss of Khaos,
it is also the open and gaping
form of my mouth (Khainô)
when I do not know what to
say*"

Think What Happens
Jacques Derrida

Derrida
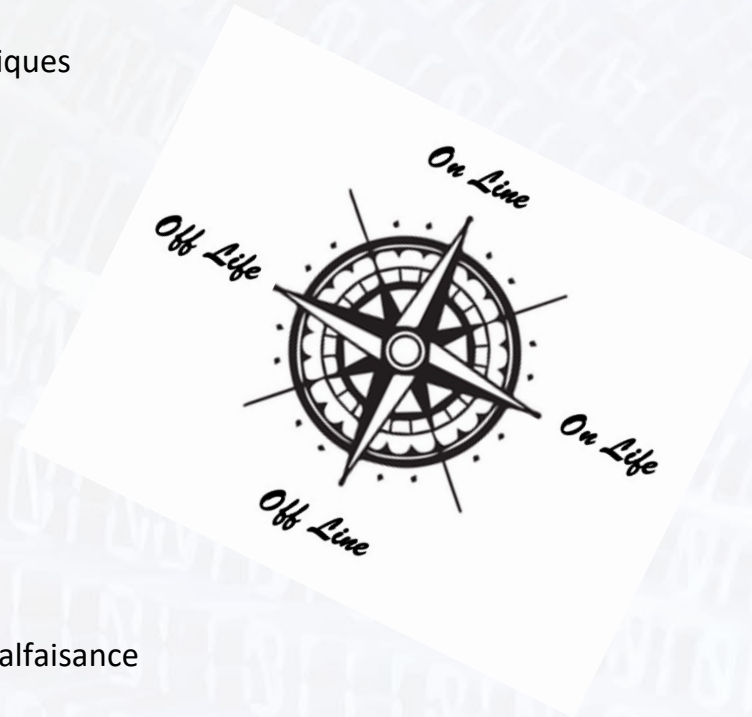pour les temps
à venir

sous la direction de René Major

l'autre pensée **Stock**

Virtual Servitudes (in French)

1. La vie en ligne, mode d'emploi
2. Rose des vents numériques
3. Hors vie
4. En ligne
5. En vie
6. Hors Ligne
7. Coups de marteau
8. Autonomie
9. Bienfaisance et non malfaisance
10. Justice
11. Transparence - explication
12. Contre la servitude

Thank You!