

Fairness, Justice, Artificial Intelligence and Decision Making



Jean-Gabriel Ganascia

ACASA team - LIP6 – CNRS (UMR 7606) - Sorbonne University

Institut Universitaire de France

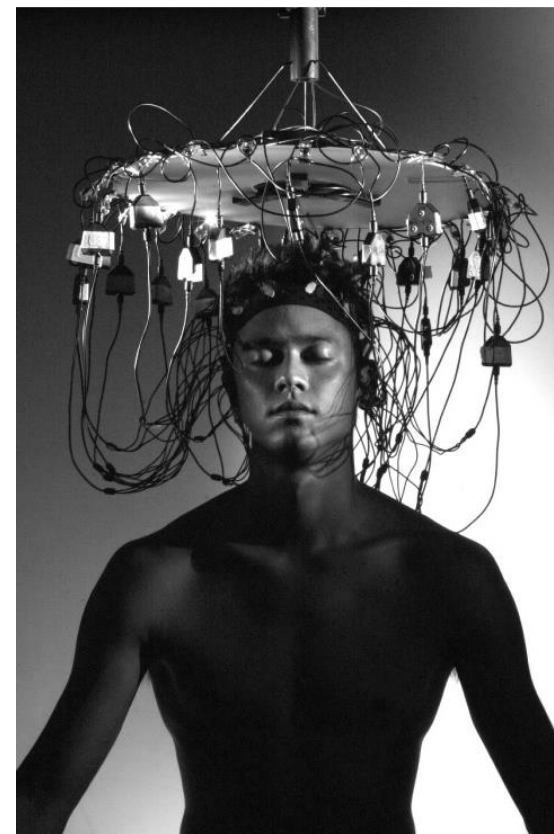
Chairman of the CNRS Ethical Committee

4, place Jussieu, 75252 Paris Cedex 05, FRANCE

Jean-Gabriel.Ganascia@lip6.fr

Synoptic

1. Artificial Intelligence and Artificial Agents
2. Decision Making: Autonomous vs. Automatic Agents
3. Fairness and Justice
4. Making Fair Agents
5. Making Fair Artificial Intelligence
6. An Ethical Trilemma



Agent = Anima (Aristotelian meaning) — *soul*
Aristotle ... *De anima*

▶ **Sensation**: presence of things



Perception

▶ **Representation**: apparition of absent things...

▶ **Thought**: computation (*calculus*) on representations

AI = Artificial Agents Programmation

1. *Information acquisition and perception*: allows information acquisition, treatment, classification and integration.
2. *Memory* and *Learning* allow storage and recall of information. Allows also abstraction and knowledge extraction.
3. *Reasoning, Thought*. concerns deduction and other reasoning. Concerns also the reorganization of memory
4. The *executive functions* of decision-making and action.
5. *Expressive functions* make possible communication between agents.

Chatbots – interactive agents



INTRODUCING
amazon echo
Always ready, connected,
and fast. **Just ask.**

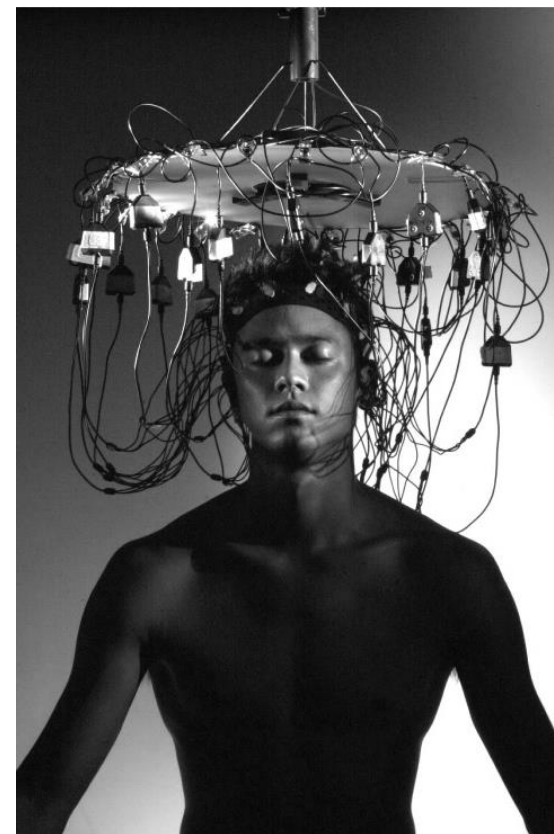


Your Google Assistant,
always ready to help.

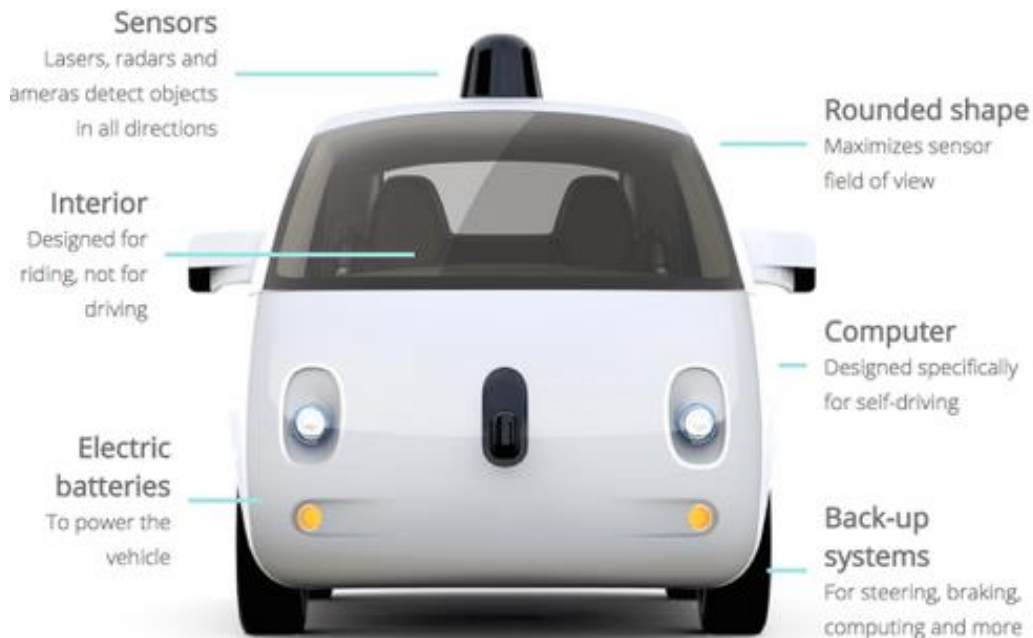


Synoptic

1. Artificial Intelligence and Artificial Agents
- 2. Decision Making: Autonomous vs. Automatic Agents**
3. Fairness and Justice
4. Making Fair Autonomous Agents
5. An Ethical Trilemma



Autonomous Agent: *autonomous car*



- **Autonomy** (*in a technical meaning*): sequence of causalities between information acquisition, decision and action without any human agent intervention

Degrees of Autonomy

There are different degrees of automation: the Society of Automotive Engineers (SAE) has defined six.

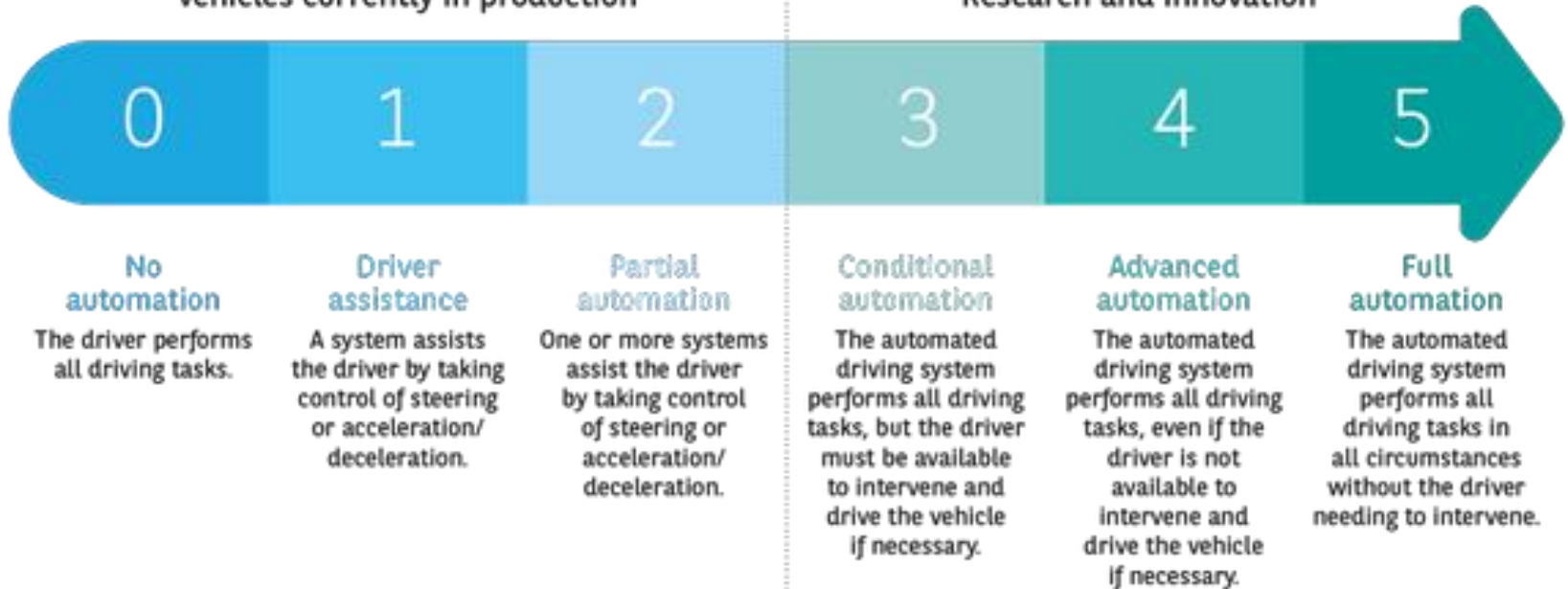
Source: BIPE, according to the SAE.



Vehicles currently in production

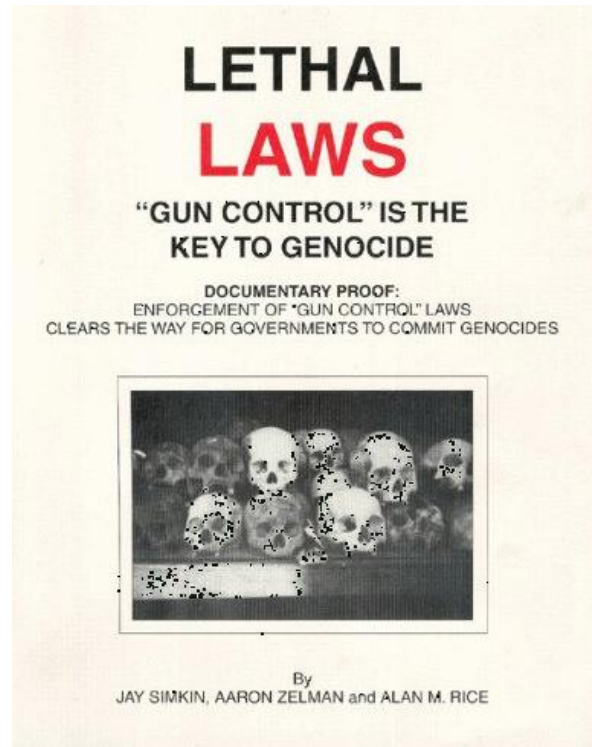


Research and innovation



Lethal Autonomous Weapon Systems

- LAWS (« killer robots »)



Are Lethal Autonomous Weapons Autonomous?

Are Autonomous Agents Autonomous? Agents?

- **Autonomous agents** are not **agents** in philosophical terms

- *Agent*: entity that acts → intention, will, responsibility...
- An *artificial agent* has no intention, he is not responsible



A Knife
without
Blade,
which
has no
handle

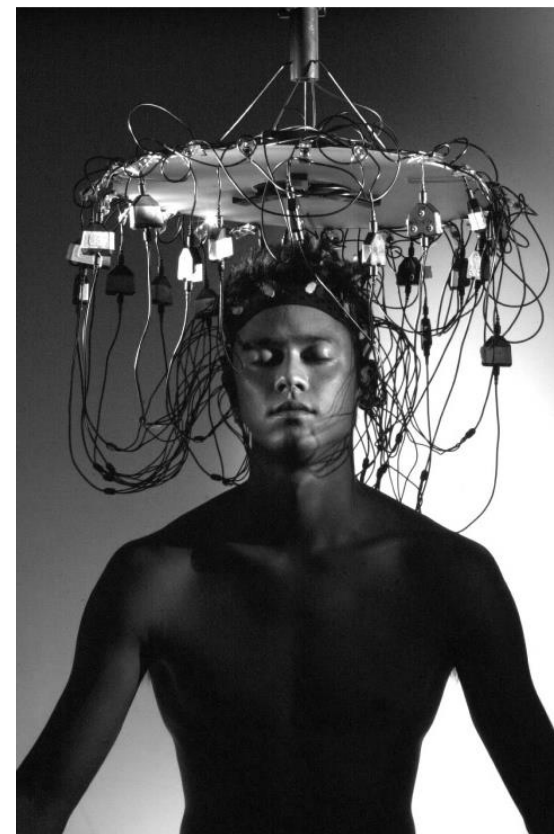
- **Autonomous robots** are not **autonomous** in philosophical terms

- An *autonomous agent* define its own law itself
- The goal are given from outside: heteronomous agent

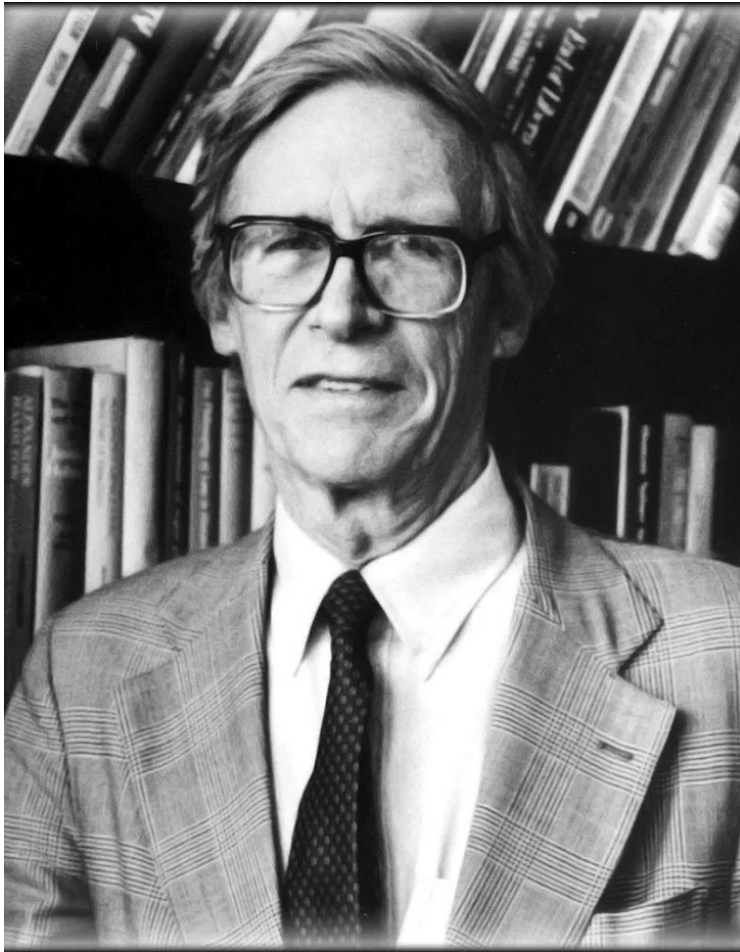
- **Autonomy ≠ automaticity**

Synoptic

1. Artificial Intelligence and Artificial Agents
2. Decision Making: Autonomous vs. Automatic Agents
- 3. Fairness and Justice**
4. Making Fair Agents
5. Making Fair Artificial Intelligence
6. An Ethical Trilemma



Justice as Fairness — John Rawls



- Theory of Justice
- Two Principles
 1. Equal claim to a fully adequate scheme of equal basic rights and liberties
 2. Equal Distribution of Opportunities
 1. “fair equality of opportunity.”
 2. “Difference Principle.”

Principle of Justice: “Be Fair”

Fairness: Developers and implementers need to ensure that individuals and minority groups maintain freedom from ***bias***, stigmatization and ***discrimination***

- ***bias:*** prejudice for or against something or somebody, that may result in unfair decisions.
- ***discrimination:*** concerns the variability of AI results between individuals or groups of people based on the exploitation of differences in their characteristics that can be considered either intentionally or unintentionally (such as ethnicity, gender, sexual orientation or age), which may negatively impact such individuals or groups



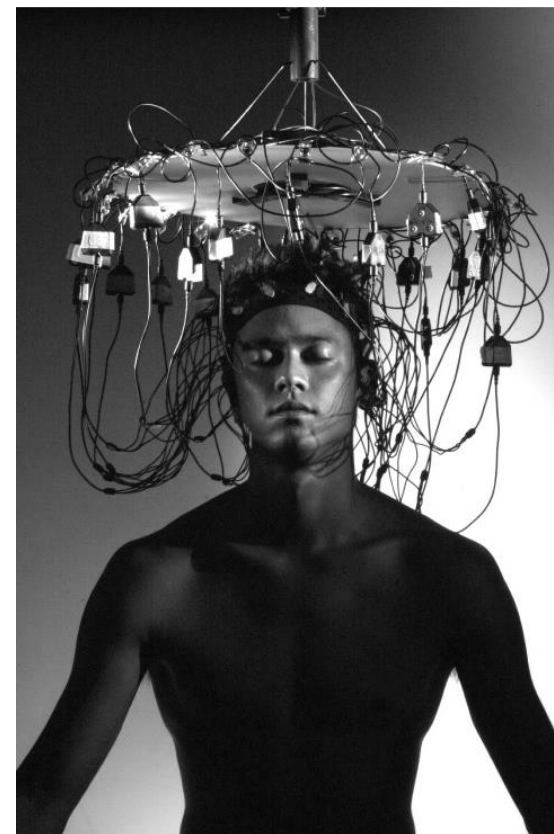
Lady Justice Wears a Blindfold



- Allegory of the Impartiality of Justice
- Impartiality of Data and Algorithms?
- Are Machine Free of Dogmas and Bias?

Synoptic

1. Artificial Intelligence and Artificial Agents
2. Decision Making: Autonomous vs. Automatic Agents
3. Fairness and Justice
4. **Making Fair Agents**
5. Making Fair Artificial Intelligence
6. An Ethical Trilemma



How an Autonomous Agent can be Unfair?

- Racism, sexism, ...
- Cars
- Car accidents



- Bots that assist diabetic people
- Allocation of jobs, organs, etc.
- Pacemaker
- Lethal Autonomous Weapon Systems
- ...



Programming an Ethical Governor

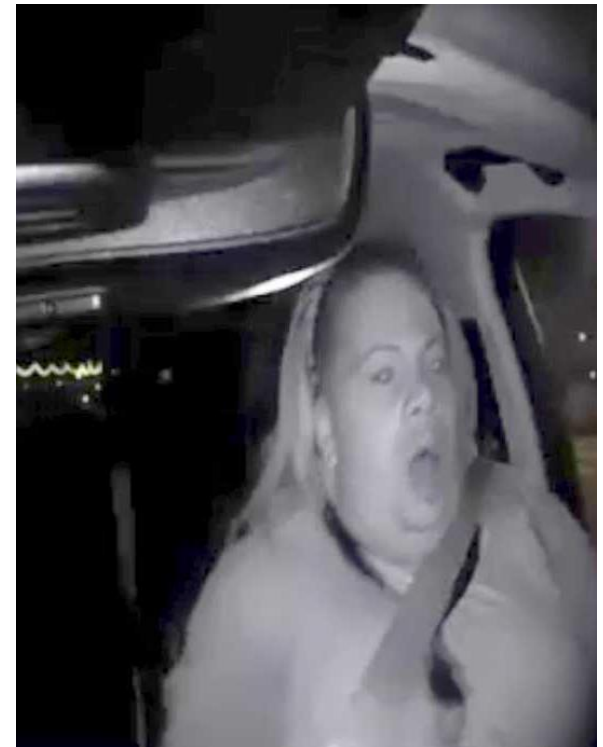
ETHICAL DISTINCTIONS

- **Judgment**: operation of the mind that derive concepts from information (*≠ from judiciary judgment*)

Uber car accident March 2018

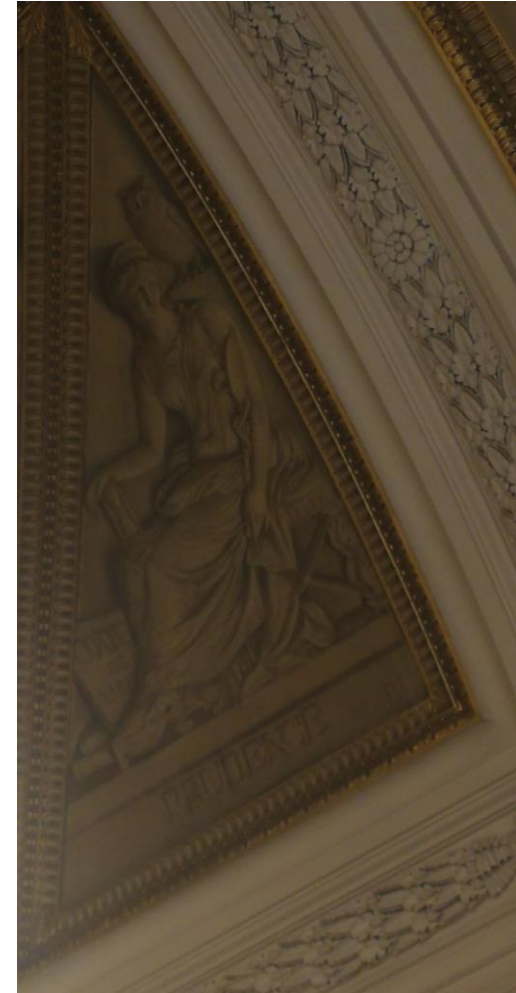
Inquiry: May 2018

- The sensors have correctly detected
- The UBER company had decided to ignore some details (*plastic bags, leaves, etc.*) that generate chaotic drive and discomfort for passengers!



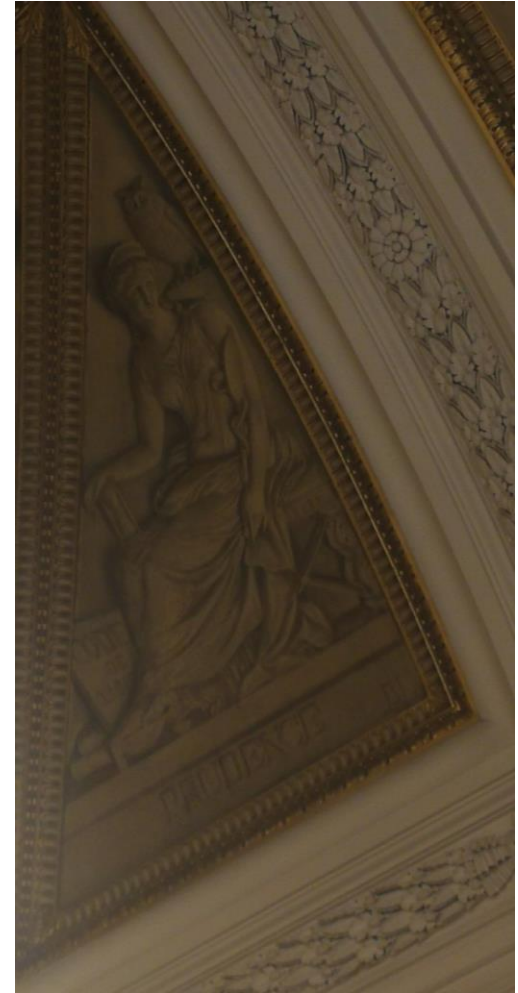
Ethical Distinctions

- **Judgment**: operation of the mind that derive concepts from information
- **Prudence**: choice of objectives



Ethical Distinctions

- **Judgment**: operation of the mind that derive concepts from information
- **Prudence**: choice of objectives
- **Sagacity**: determination of actions that reach the objectives (intelligence?)
- **Deliberation**: choice between different actions



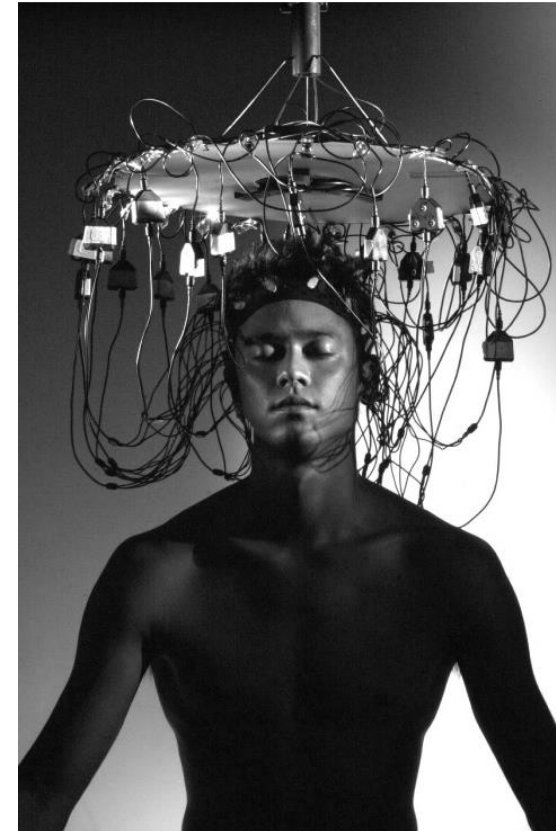
Traditional example of conflict

- Lying is prohibited
- Murderers want you to say where is the friend to which you gave the hospitality
- What to have to do?

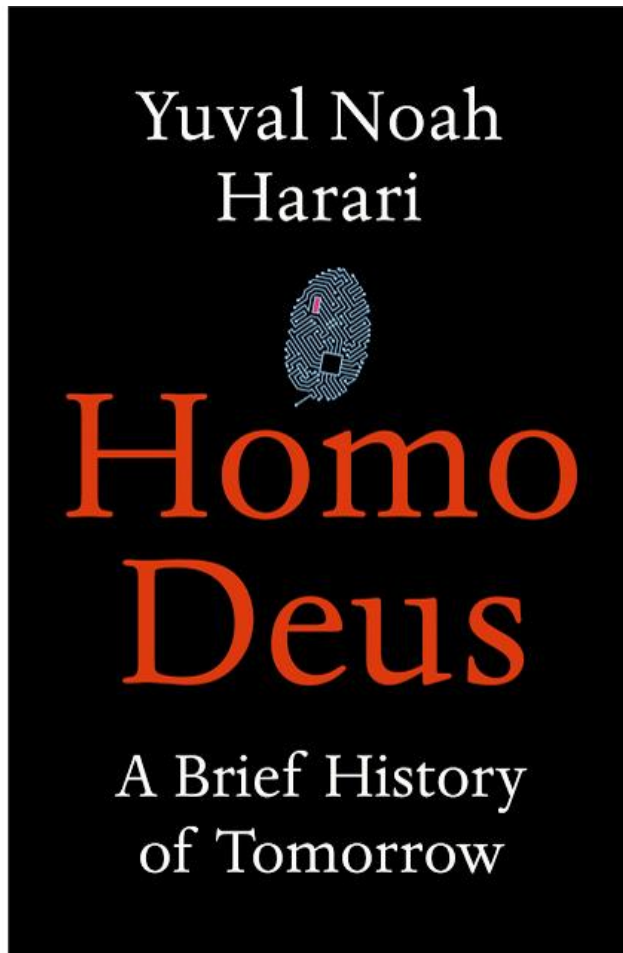


Synoptic

1. Artificial Intelligence and Artificial Agents
2. Decision Making: Autonomous vs. Automatic Agents
3. Fairness and Justice
4. Making Fair Agents
5. **Making Fair Artificial Intelligence**
6. An Ethical Trilemma



Homo Deus — Pop Science



- Three stages in the history of Humanity
 - Religions
 - Humanism
 - “**Dataism**”: end of humanism
- Man’s subjugation to the machines, which **will decide** for us

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

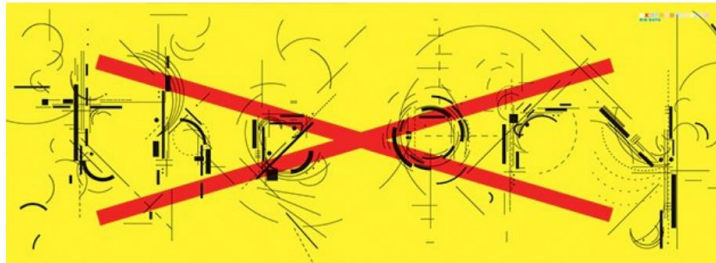


Illustration: Marian Bantjes "All models are wrong, but some are useful."

The Peta Byte Age!

- No sampling
- Gathering everything, without prior model
- No semantic or causal analysis is required.
- “Correlation is enough”

Epistemology vs. Ethics of Big Data



- Not compliant with GDPR
- Not compliant with HLEG on AI

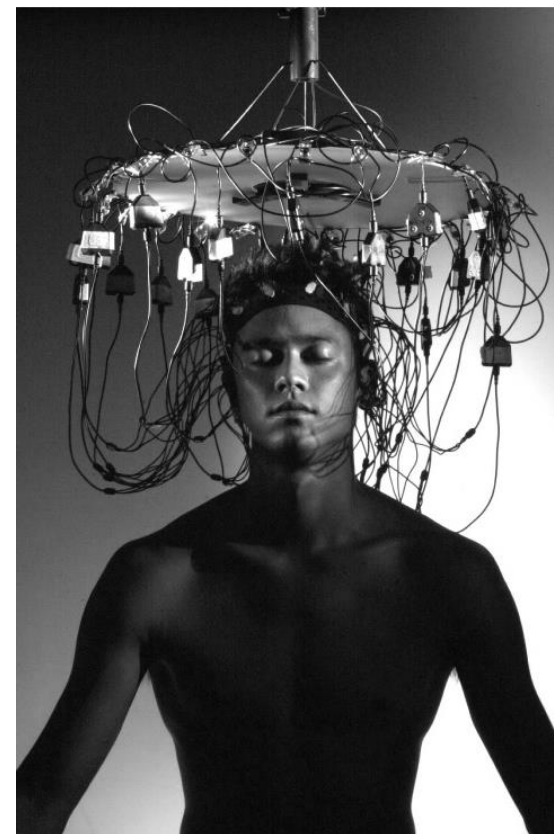


The Peta Byte Age!

- No sampling
- Gathering everything, without prior model
- No semantic or causal analysis is required.
- “Correlation is enough”

Synoptic

1. Artificial Intelligence and Artificial Agents
2. Decision Making: Autonomous vs. Automatic Agents
3. Fairness and Justice
4. Making Fair Agents
5. Making Fair Artificial Intelligence
6. **An Ethical Trilemma**



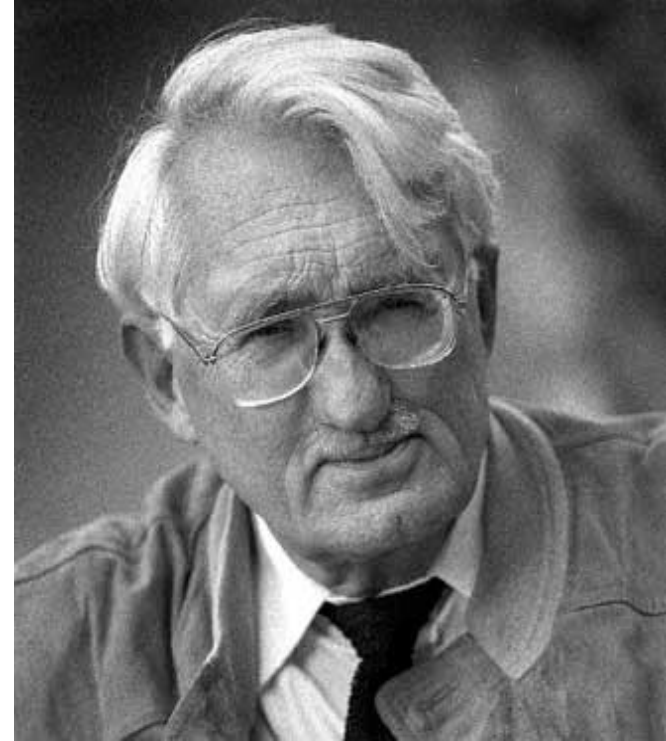
An Ethical Trilemma

- **Diversity (Inclusion)**
- **Non-discrimination**
- **Fairness (Justice)**



From Machine Ethics to Ethics Machine

- **Ethical Machines:** building machines that take decisions conform to ethical rules
- **Machines for Ethics :** facilitate or allow machine-enhanced normative deliberation — i.e. ethical decision made by humans but aided by machine —
- **Communicative Action:** cooperative action undertaken by agents based upon mutual deliberation and argumentation



Jürgen Habermas