AI4EU

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Trustworthy AI
The AI4EU approach

Ulises Cortés

2019

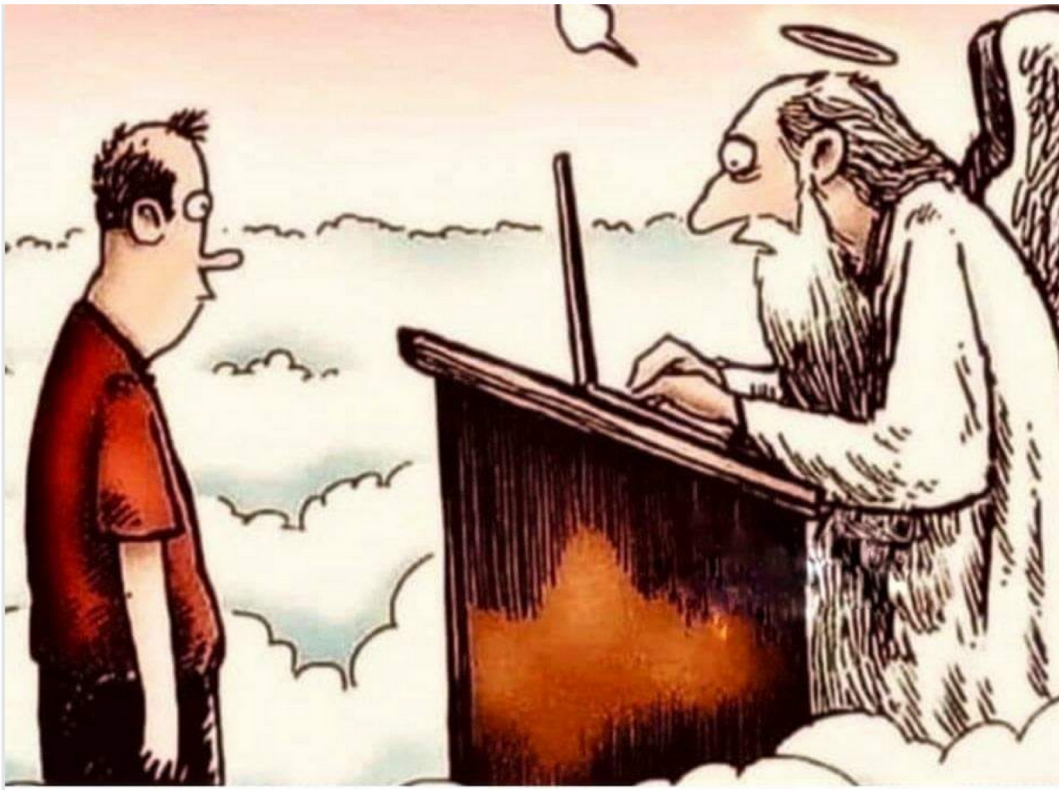
Who am I

- **Ulises Cortés**
  - **AI researcher since (1982)**
  - **Professor of Artificial Intelligence (2006)**
  - **Coordinator of the Masters program on AI (2005)**
  - **Head of the HPAI research group at Barcelona Supercomputing Center (2017)**
  - **AI4EU ELSEC WP5 Coordinator (2019)**

- ia@cs.upc.edu
- http://www.cs.upc.edu/~ia

What I am not

- **I am not**
  - **A Philosopher,**
  - **An Ethicist,**
  - **A Futurist,**



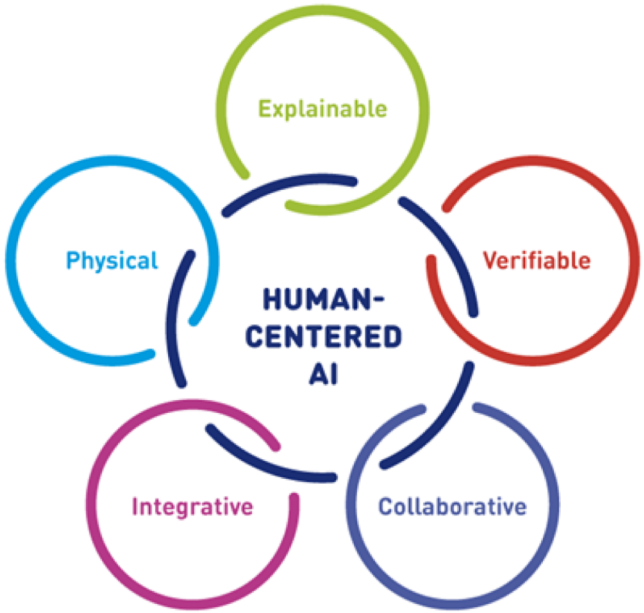
ia@cs.upc.edu

The AI On-Demand Platform and Ecosystem

AI4EU is a collaborative H2020 Project that aims to

- Mobilize the entire European AI community to make AI promises real for the European Society and Economy
- Create a leading collaborative AI European Platform to nurture economic growth.
- **Key figures**
  - 82 members (~60 leading research institutes)
  - 21 partnering countries
  - €3m Cascade Funding

Based on 5 Research Areas



Ethical Observatory

Strategic Research and Innovation agenda

ia@cs.upc.edu

Growing questioning about AI

Le Monde

Intelligence artificielle : 3 100 employés pressent Google d'arrêter d'aider le Pentagone

L'existence d'un partenariat entre la firme et le ministère de la défense américain a été révélée en mars. Une initiative qui déplaît à de nombreux salariés de l'entreprise.

***Reuters Science News**, June 21, 2016*

"A draft European parliament motion suggests that the growing intelligence, pervasiveness and autonomy of the growing army of European robot workers requires rethinking of everything from **taxation to legal liability** and classifying them as **"electronic persons"** making their owners liable to **pay social security for robot workers**."

LE FIGARO

L'Intelligence artificielle peut-elle être éthique?

Par Laetitia Pouliquen | Mis à jour le 25/02/2019 à 09:31 / Publié le 21/02/2019 à 15:25

Elon Musk leads 116 experts calling for outright ban of killer robots

Open letter signed by Tesla chief and Alphabet's Mustafa Suleyman urges UN to block use of lethal autonomous weapons to prevent third age of war

***Art Bilger**, Venture Capitalist & Board Member of the **Wharton Business School** cited an **Oxford study**:*

"All developed nations will see a loss of **47% in the next 25 years** in blue and white collar jobs beginning with the manufacturing industry"

Jobs at risk: Accountants, doctors, lawyers, teachers, bureaucrats, financial analysts, production line workers, drivers, most routine support jobs, middle management jobs that merely interpret data, restaurant waiters.....

The Economist reports:

NO GOVERNMENT IS PREPARED

The person as product (let's take an example)

- What can we do with just one person's data?
- What can we do with that X's data?
  - G can look at X's financial records.
  - G can tell if X pay her bills on time.
  - G know if X is good to give a loan to.
  - G can look at X's medical records; G can see if your pump is still pumping -- see if he is good to offer insurance to.
  - G can look at X's clicking patterns.

Do the right thing

- The question is not **What G can do with with X's data?** But **Which is the right thing to do?**
- These are some selected choices:
  - Should G be collecting it, gathering it, so G can make X's online experience better?
  - So G can make money?
  - So we (China, Europe, USA) can protect ourselves if X was up to no good?
  - Or should we respect X's privacy, protect his dignity and leave him alone?

Do the right thing

- Hawei vs iPhone?
-
-

Do the right thing

- Hawei vs iPhone?
- Should we be collecting all of that X's data to make his experiences better and to protect ourselves in case he's up to no good? Or should we leave him alone?
-

Do the right thing

- Hawei vs iPhone?
- Should we be collecting all of that X's data to make his experiences better and to protect ourselves in case he's up to no good? Or should we leave him alone?
- When trying to evaluate what we should do in this case, should we use a Kantian deontological moral framework, or should we use a Millian consequentialist one?

Dilemmas, Values & Videotapes

- Security and Privacy
- Safety and Efficiency
- Accountability and Confidentiality
- Prosperity and Sustainability
- **Moral overload: you cannot have it all**

Unethical uses of AI

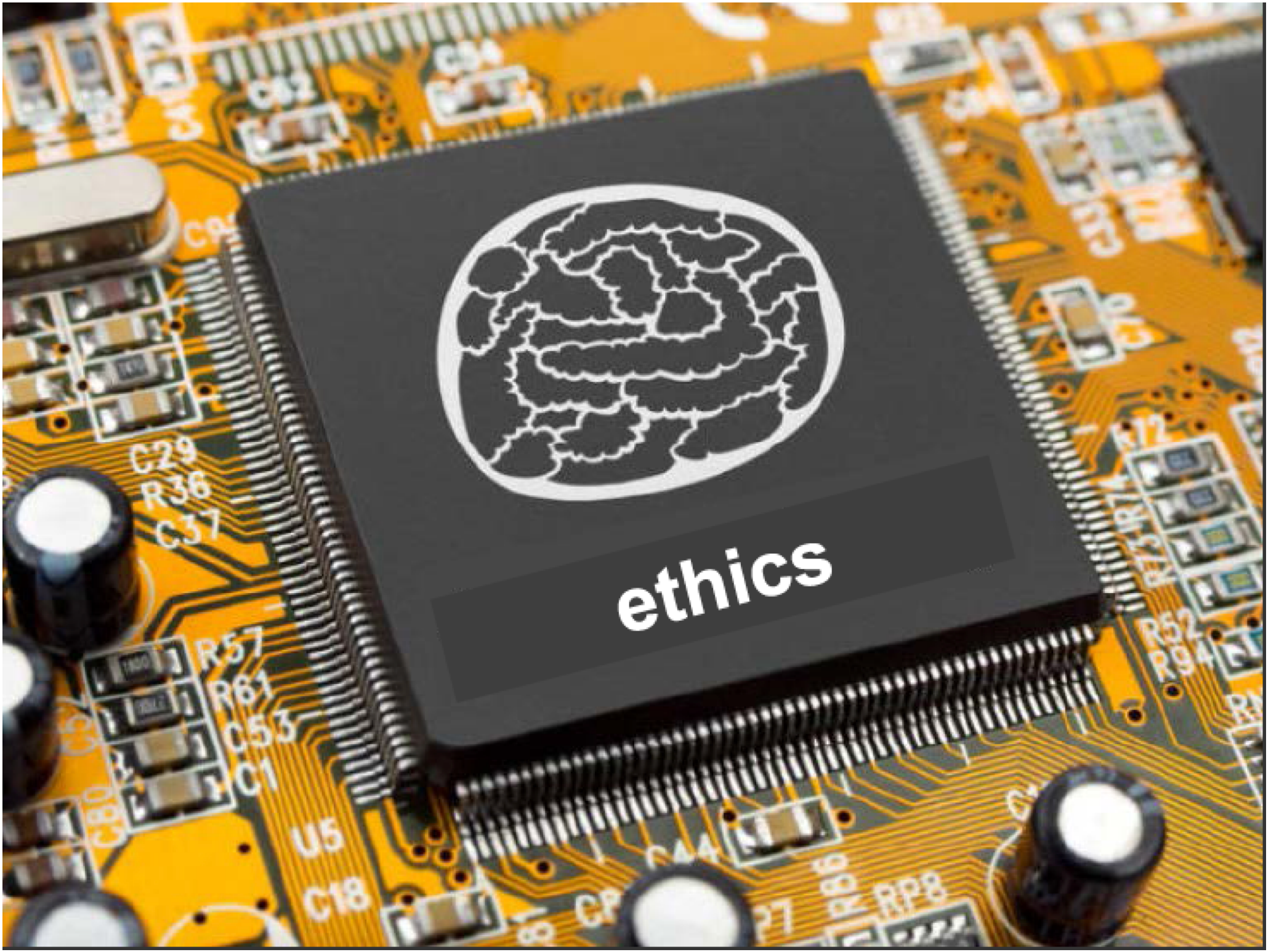
- Algorithmic Bias
- Autonomous guns (aka Killer Robots)
- Cambridge Analytics
- Cyber-snooping
- Job displacement
- Misinformation and Fake News

Examples of unethical research

- **Tuskegee Experiment** (1932-1972) American researchers purposely withheld treatment for 399 African-American people with syphilis for the sole purpose of studying the long term effects of the disease.
- **Willowbrook Study** (1963-1966) Children with developmental disabilities were deliberately infected with Hepatitis (some were even fed fecal matter). Purpose of the study was to examine the course of the disease and to test a potential immunization
- **Human radiation experiments by the US Department of Defense & Atomic Energy Commission**.
- Milgram's Obedience Study-Researchers asked participants to *Pseudo-shocking* confederates in order to examine obedience.
- **Zimbardo's Stanford Prison Experiment** (1971). Study had to be ended prematurely because of abusive behaviors generated participants who where assigned as guards over those subjects that were assigned as prisoners.

Responses to unethical research

- Nuremberg created as a result of cruel experiments the Nazis conducted on humans during WWII.
- NIH Ethics Committee (1964)
- Declaration of Helsinki (1964, ’75, ’83, ’89, ’00)
- Beecher “Ethics & Clinical Research” (1966) [NEJM, 274, 1354-60]. Available at http://sladen.hfhs.org/IRB/images/nejm-beecher.pdf
- 1973 Congressional Hearings on Quality of Heath Care and Human Experimentation.
- National Research Act of 1974
- Established the IRB system.


ethics

which leads to a multiplication of "*reference sources*"

Parlement européen
2014-2019

TEXTES ADOPTÉS

P8_TA(2017)0051

Règles de droit civil sur la robotique

Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règle

The European Commission's
HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

DRAFT
ETHICS GUIDELINES
FOR TRUSTWORTHY AI

stakeholders' consultation
December 2018

CÉDRIC VILLANI
Mathématicien et député de l'Essonne

DONNER UN SENS À L'INTELLIGENCE ARTIFICIELLE

POUR UNE STRATÉGIE NATIONALE ET EUROPÉENNE

ASILOMAR AI PRINCIPLES

OECD Principles on Artificial Intelligence

On 22 May 2019 the OECD adopted its **Principles on Artificial Intelligence**, the first international standards agreed by governments for the responsible stewardship of trustworthy AI.

The OECD Principles on AI include concrete recommendations for public policy and strategy. The general scope of the Principles ensures they can be applied to AI developments around the world.

We are also planning to launch a **policy observatory** to ensure the beneficial use of AI later in the year.

Montréal Declaration Responsible AI

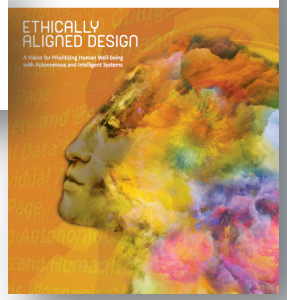
MONTREAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE 2018

Software Engineering Code of Ethics

IEEE-CS/ACM Joint Task Force on
Software Engineering Ethics and Professional Practices

Overview - Version 2

IEEE
Advancing Technology for Humanity



CNIL.

La CNIL appelle à la tenue d'un débat démocratique sur les nouveaux usages des caméras vidéo

Parlement europeen
2014-2019

TEXTES ADOPTÉS
Édition provisoire

P8_TA-PROV(2019)0081

ue industrielle européenne globale sur l'intelligence artificielle et

CCW/GGE.1/2017/3

Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects

22 December 2017

Original: English

Geneva, 13–17 November 2017
Item 7 of the agenda
Adoption of the report

The Japanese Society for Artificial Intelligence Ethical Guidelines

Companies also contribute to the noise

Partnership on AI
to benefit people and socie

MAIF
assureur militant

POUR UN MONDE NUMÉRIQUE RÉSOLUMENT HUMAIN ET ÉTHIQUE

Le numérique occupe une place toujours plus importante dans notre société, notre économie et nos vies. Nous oeuvrons à la MAIF pour que nos sociétaires, comme nos équipes, bénéficient des bienfaits de son développement.

Telefónica
Our Artificial Intelligence Principles

DEUTSCHE TELEKOM

DIGITAL ETHICS GUIDELINES ON AI

Facebook Reportedly Has A Dedicated AI Ethics Team

Sam Shead Contributor
I cover artificial intelligence and Google DeepMind.

The Future Computed

Artificial Intelligence and its role in society

By Microsoft

Foreword by Brad Smith and Harry Shum

AI

AI at Google: our principles

Sundar Pichai
CEO
Published Jun 7, 2018

At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and easier to compose, to a digital assistant you can speak to naturally, to photos that pop the fun stuff out for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to predict the risk of wildfires. Farmers

Microsoft AI principles

signing AI to be trustworthy requires creating solutions that reflect ethical principles that are deeply rooted in important and timeless values.

Fairness
AI systems should treat all people fairly

Reliability & Safety
AI systems should perform reliably and safely

Privacy & Security
AI systems should be secure and respect privacy

Inclusiveness
AI systems should empower everyone and engage people

Transparency
AI systems should be understandable

Accountability
AI systems should have algorithmic accountability

Ethical Values

- Autonomy
- Beneficence
- Non-maleficence
- Justice
- Fidelity

- Think for a moment, *how* might these principles relate to research?

What is the use of Ethics?

If ethical theories are to be useful in practice, they need to affect the way human beings behave.

- Is this applicable to a machine?
- To which kind of machines?
- Are there ethical machines?

Why should we care about (AI) Ethics

- **So many ethical situations that we encounter each day that we should care.**
- **Some unethical actions can violate law.**
- **Others, though not illegal, can have drastic consequences for our careers and reputations**
- **We should care about ethics for our own self interest**

Machine Ethics and *Regular* Ethics

- Is machine ethics different from regular ethics?
- Is there an ethical difference in browsing someone else's computer/device and browsing their desk drawer?
  - No!
- What we have are ethical situations where computers and/or intelligent systems are involved.
- **Machines allow people to perform unethical actions faster than ever before**.
- Or perform actions that were too difficult or impossible using manual methods.

Online (Internet) privacy

- (pre-Internet) The Privacy Act of 1974 prevents **unauthorized disclosure of personal information** held by the federal government. A person has the right to review their own personal information, ask for corrections and be informed of any disclosures.

- The Financial Monetization Act of 1999 requires financial institutions to provide customers with a privacy policy that explains what kind of information is being collected and how it is being used. Financial institutions are also required to have safeguards that protect the information they collect from customers.

Privacy

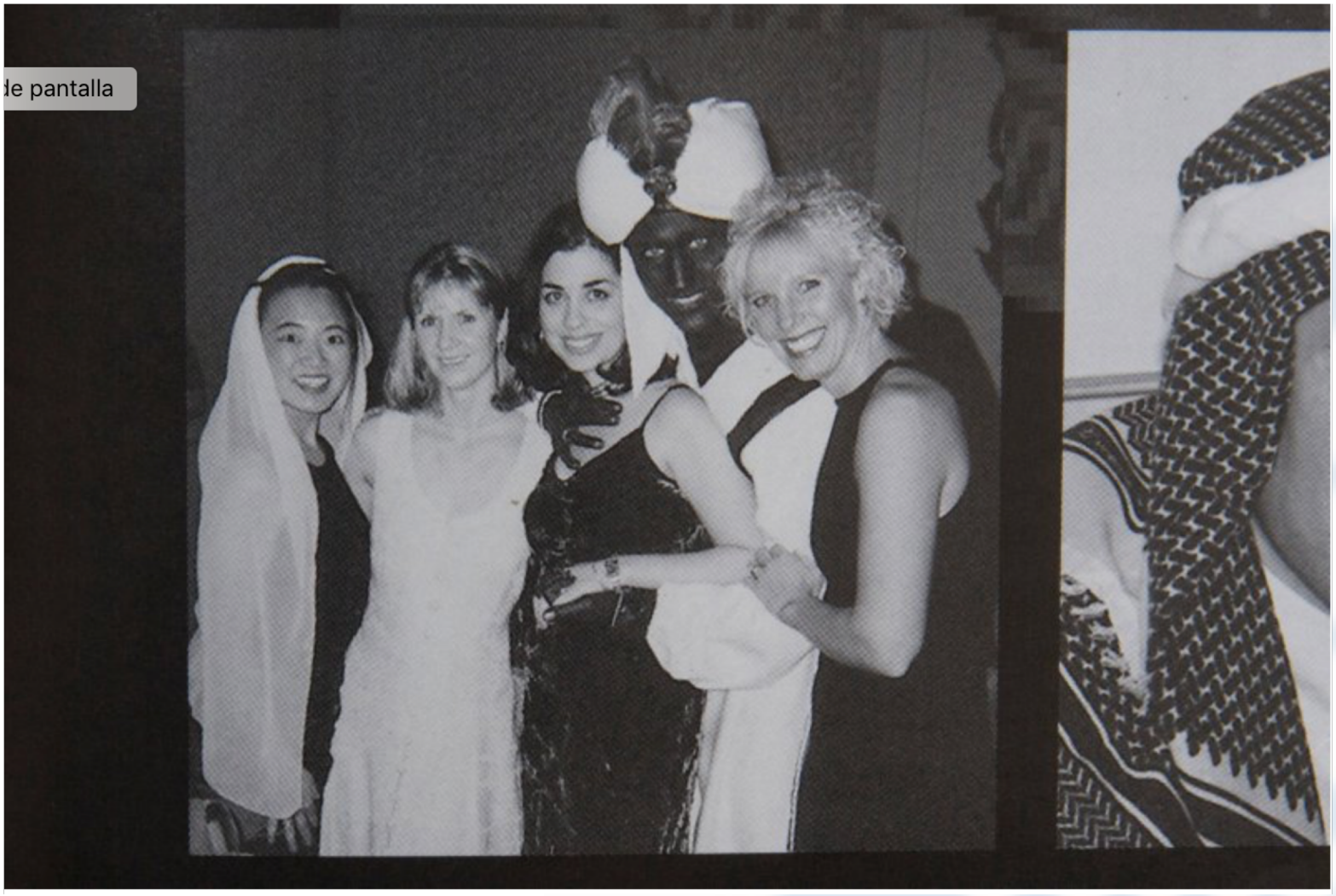
- There's an interesting paradox here, with Internet users being less likely to take action to protect their privacy, while non-users tend to be put off by privacy concerns.

Privacy: Face Match

Google's latest smart display brings with it a controversial new feature that's always watching. Face Match, introduced on the Google Nest Hub Max, uses the smart display's front-facing camera as a security feature and a way to participate in video calls. It also shows you your photos, texts, calendar details and so on when it recognizes your face.

This mode of facial recognition sounds simple enough at first. But the way companies like Google collect, store and process face data has become a top concern for privacy-minded consumers..

Fake memory (Memoria fingida)



European High-Level Expert Group on AI
Ethics Guidelines for Trustworthy AI
(April 2019)

Lawful AI: Legal compliance with Primary law (treaties, charter of fundamental rights), secondary law (GDPR, product liability directive), CoE conventions, State laws, Sector-specific regulations (*e.g.*, healthcare).

Ethical AI: alignment with ethical principles and norms.

Robust AI: safety, security by design (technical robustness), appropriate application operational contexts and limitation of unintended consequences (non-technical robustness).

https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Ethical Principles for Trustworthy AI

Ethical imperatives

Principle of Autonomy: "*Preserve Human Agency and control*"

Principle of Non maleficence: "*Do no Harm*" - Neither cause nor exacerbate harm or otherwise adversely affect human beings. safety and security, technical robustness.

Principle of Justice: "*Be Fair*". Equal and just distribution of benefits and costs, free from unfair bias, increase social fairness

Principle of Explicability: "*Operate transparently*". Traceability, auditability, transparent system capabilities, …

Foundations of Trustworthy AI: A Human-Centric Approach

- **Respect for human dignity.** Humans are moral subjects, not objects to be scored, herded or manipulated.
- **Freedom of the individual**. Fundamental rights, control over one's own life and choices, protection from sovereign intrusion
- **Respect for democracy and justice**. Protection of democratic processes and human deliberation
- **Equality, non-discrimination and solidarity.** No bias, no exclusion
- **Citizens' rights**. Access to administration and services (including non-citizens).

Relationship between Ethics and Law

The relationship between ethics and law leads to four possible states

	Legal	Not Legal
Ethical	**I**	**I I**
Not Ethical	**III**	**IV**

Relationship between Ethics and Law

The relationship between ethics and law leads to four possible states

	Legal	Not Legal
Ethical	**I**	**I I**
Not Ethical	**III**	**IV**

We're living in a world of low government effectiveness, and there the prevailing neo-liberal idea is that companies should be free to do what they want. **Our system is optimized for companies that do everything that is legal to maximize profits, with little nod to morality**.

B. Schneier

AI-Based Socio-Technical Systems

- Need to comply with human values
- Be technically dependable and socially trustworthy
- Need both **technical** and **non-technical** frameworks

Conclusions

- Science and technology − AI included - influence and are influenced by our socio-economic systems
- For humans (and machines) knowing ethics is not being ethical
  - Different contexts, different ethics led to different decisions.
- Artificial Intelligence ought to have ART
  - Accountability, Responsibility, Transparency

A classic dilema

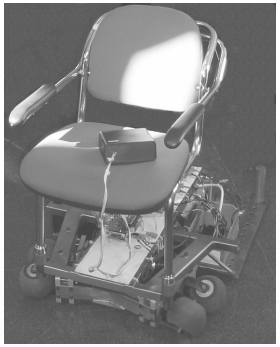
Science: "Can we?"
Ethics: "Should we?"

Navigating the future requires attention, care, and the willingness to make some hard choices.

W. Wallach

“The best way to predict the future is to invent it.”

Alan Kay



http://www.cs.upc.es/~webia/KEMLG/