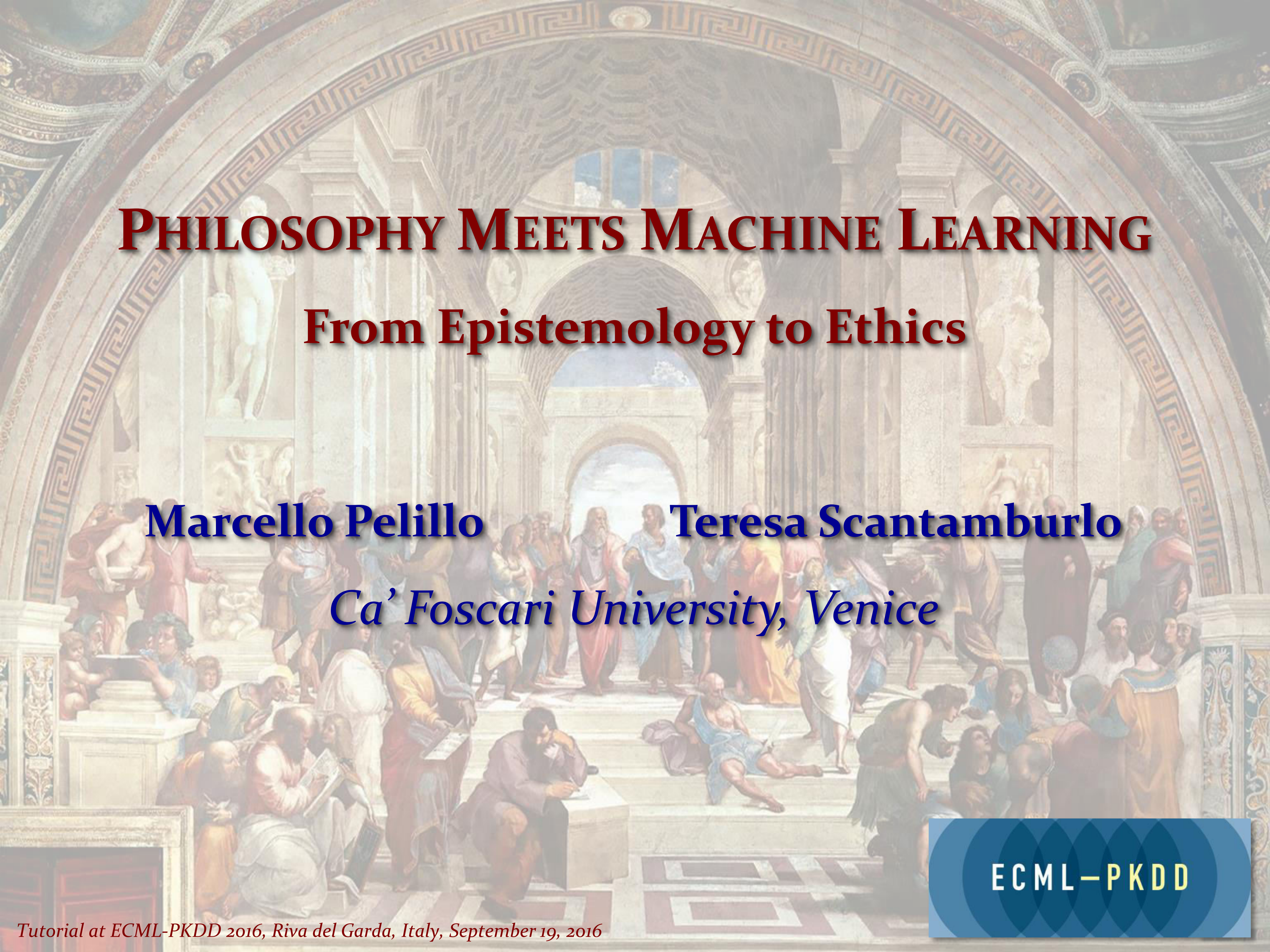


Philosophy Meets Machine Learning

From Epistemology to Ethics

Marcello Pelillo Teresa Scantamburlo

Ca' Foscari University, Venice

Tutorial at ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016

Outline

1. **Introduction to ethics**
   - Motivations
   - Socio-technical systems

2. **Ethics in Data Driven Machine Learning**
   - Privacy
   - Fairness

3. **Conclusions and provoking questions**

Introduction to Ethics

A. Mantegna, *Triumph of the virtues*, 1502

Human flourishing

“Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good; and for this reason the good has rightly been declared to be that at which all things aim.

“Now fine and just actions, which political science investigates, admit of much variety and fluctuation of opinion, so that they may be thought to exist only by convention, and not by nature[…]. We must be content, then, in speaking of such subjects and with such premises to indicate the truth roughly and in outline.”

Aristotle,
The Nicomachean Ethics

Methodological hint: Ethics is not a theoretical discipline!

From Ethics to Applied Ethics…

BIOETHICS

BUSINESS ETHICS

COMPUTER ETHICS

MEDIA ETHICS

SOCIAL ETHICS

ENVIRONMENTAL ETHICS

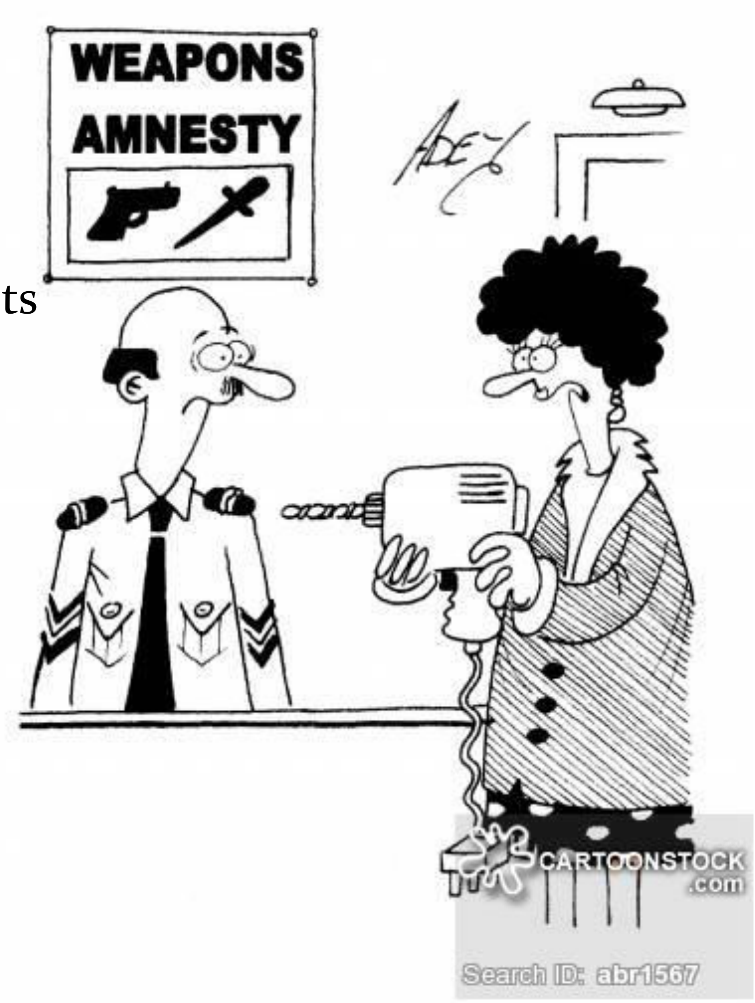
ENGINEERING ETHICS

PROFESSIONAL ETHICS

Ethics and technology

A common sense reasoning:

- Technology produces artifacts
- Ethics deals with the good / wrong use of artifacts
- Ethics pertains people (how they should use artifacts)

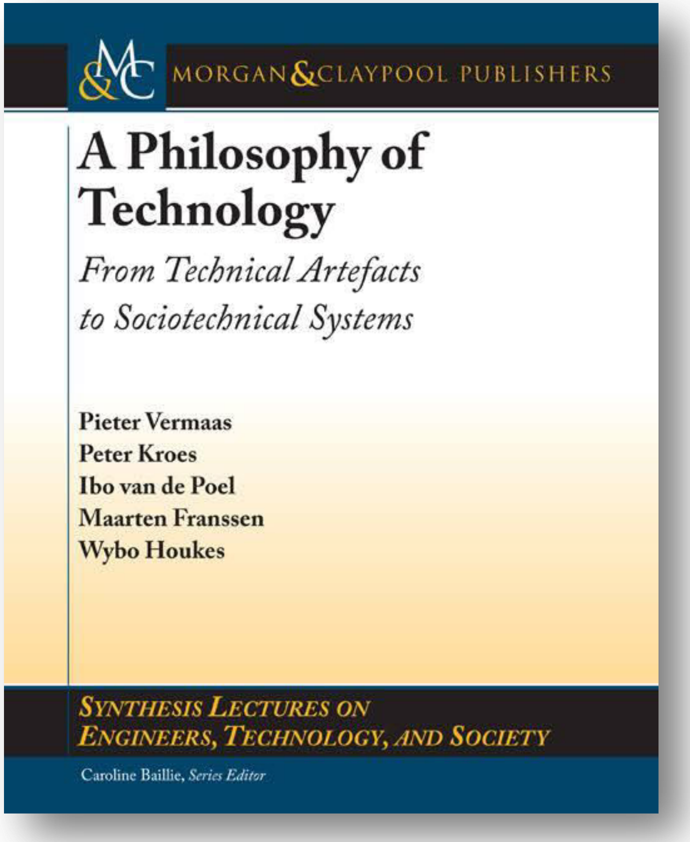


"Well, in the hands of my husband it's almost as lethal as a loaded gun."

Assumption: Technical artifacts are neutral

Question: Are artifacts really neutral?

Technical artefacts



MORGAN&CLAYPOOL PUBLISHERS

A Philosophy of Technology

From Technical Artefacts to Sociotechnical Systems

Pieter Vermaas
Peter Kroes
Ibo van de Poel
Maarten Franssen
Wybo Houkes

SYNTHESIS LECTURES ON ENGINEERS, TECHNOLOGY, AND SOCIETY

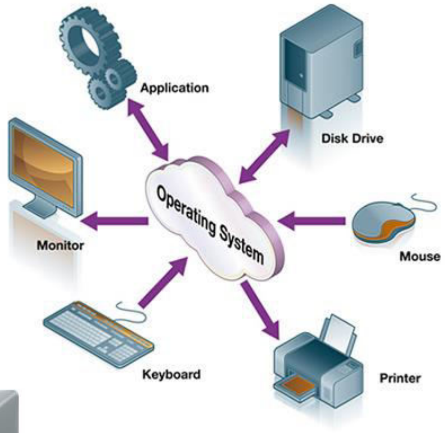
Caroline Baillie, *Series Editor*

- In philosophy of technology a technical artifact is a “physical object with a **technical function** and use plan designed and made by human beings”. Technical artifact is made to serve a **purpose**

- Similarly, in computer science we find **computational artifact** which typically have a dual nature: they have both an abstract guise and a physical one.

Vermaas et. al,
A Philosophy of technology. *From Technical Artefacts to Sociotechnical Systems* (2011)

Some examples of technical artifacts



Application

Disk Drive

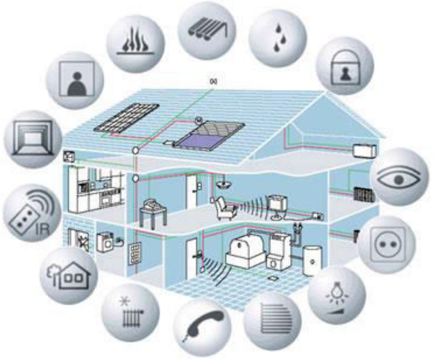
Monitor

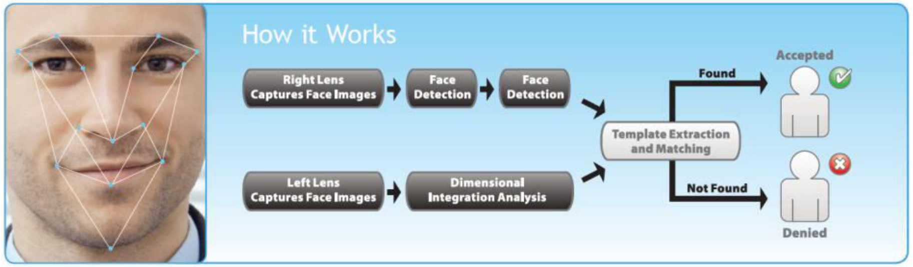
Operating System

Mouse

Keyboard

Printer





How it Works

Right Lens Captures Face Images → Face Detection → Face Detection

Left Lens Captures Face Images → Dimensional Integration Analysis

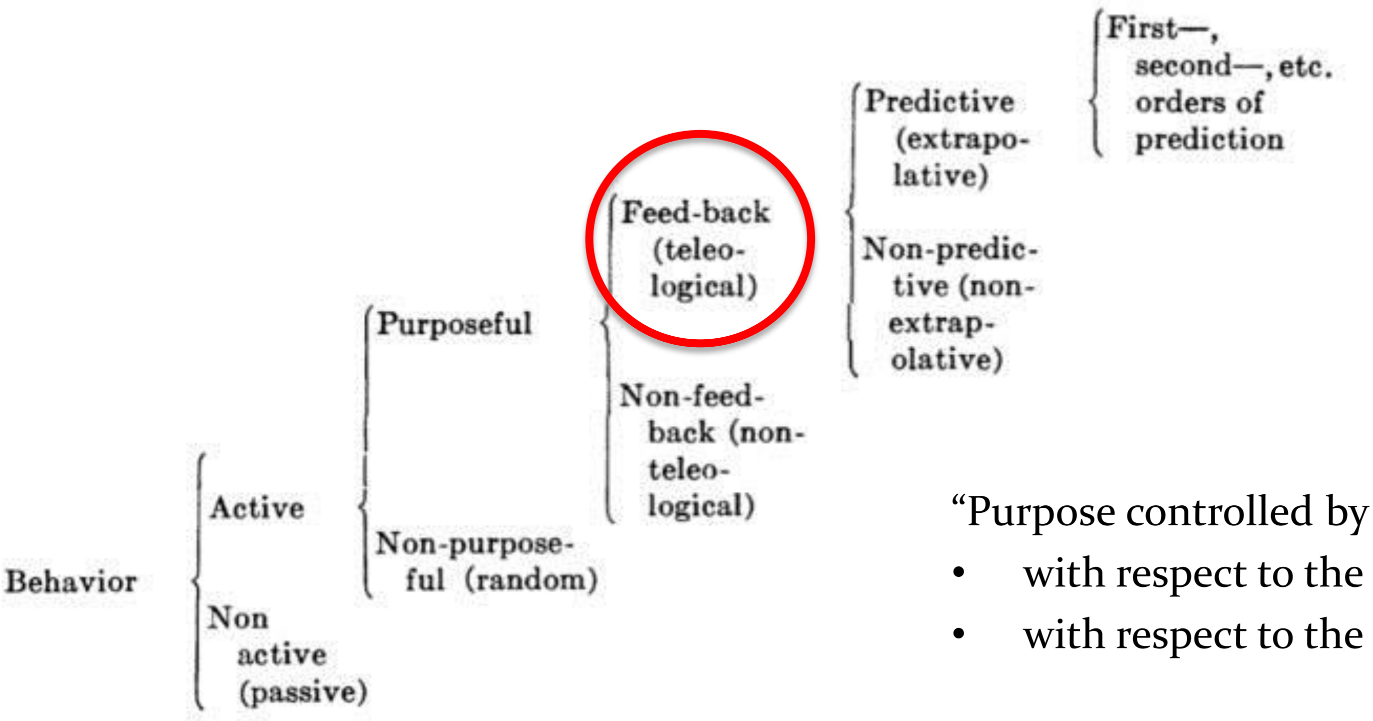
Template Extraction and Matching

Found → Accepted

Not Found → Denied

Behavior, Purpose and Teleology

"Although the definition of purposeful behavior is relatively vague [...] the **concept of purpose** is useful and should, therefore, be retained."



Behavior

Active

Non active (passive)

Purposeful

Non-purposeful (random)

Feed-back (teleological)

Non-feed-back (non-teleological)

Predictive (extrapolative)

Non-predictive (non-extrapolative)

First—, second—, etc. orders of prediction

"Purpose controlled by feed-back":

- with respect to the goal
- with respect to the environment

A. Rosenblueth, N. Wiener and J. Bigelow, *Behavior, Purpose and Teleology* (1943)

Two essential ingredients

- **Function / Purpose**: a technical artifacts is the result of a purposeful human action.

- **Context / Environment**: technical artifacts do not live in a vacuum, to fulfill their function they have to interact with other external components

- Both the purposes and and the environment include **social** and **ethical factors**

Socio-technical systems

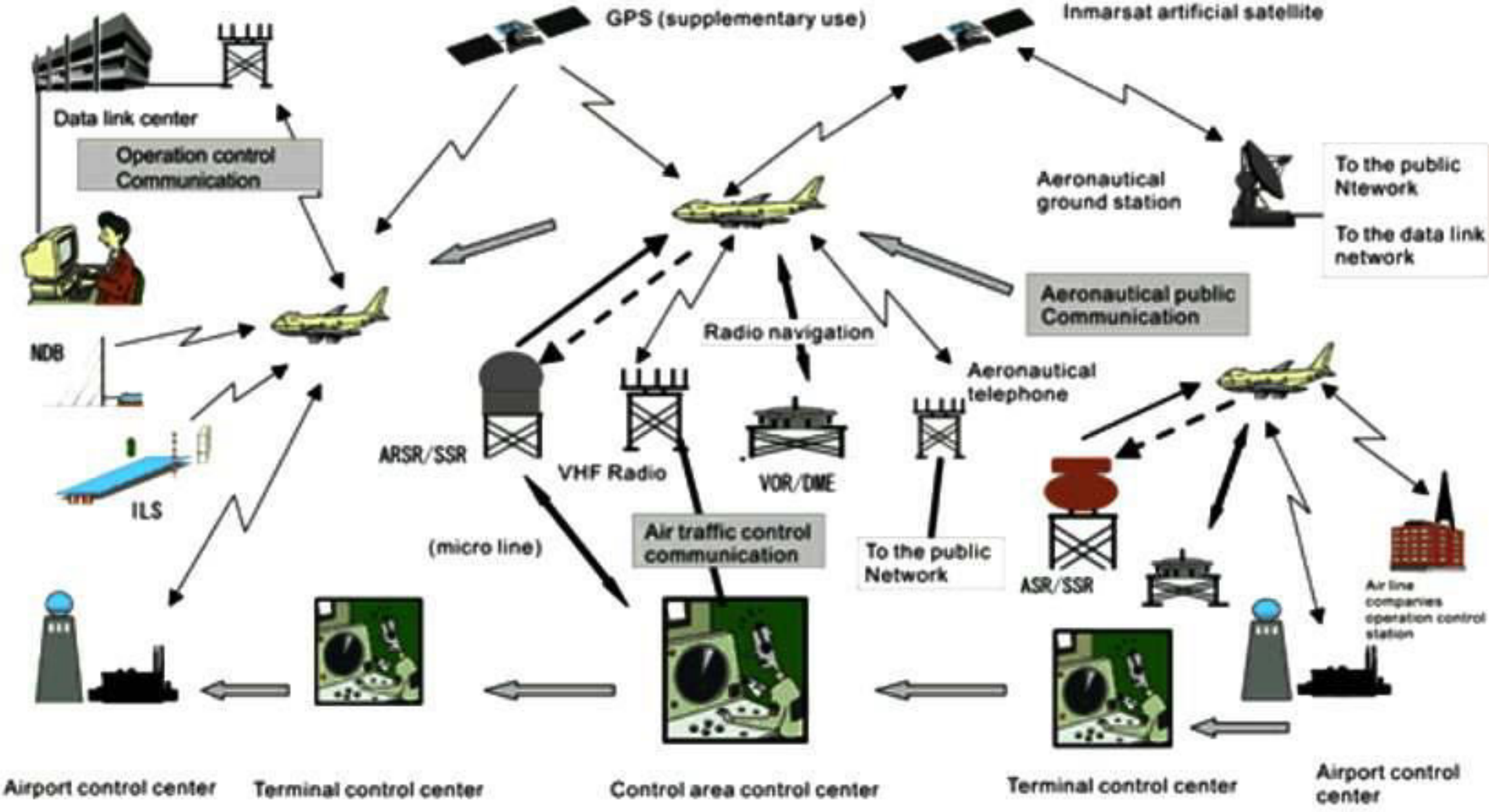
A **Socio-technical system** is "an entity that can be separated into parts which are all simultaneously linked to each other in a specific way." Some components are "hard" things which are governed by various natural law. Others, such as organizations, conditions, rules must be described by drawing on social science.

Vermaas et. al,
A Philosophy of technology.
From Technical Artefacts to Sociotechnical Systems (2011)

Socio-technical systems depend on a host of social, cultural, political and economic arrangements. "They affect us not purely by dint of physical or material properties but by properties they acquire as systems and devices embedded in larger material and **social networks** and **webs of meaning**"

Nissenbaum H.,
Privacy In Context . Technology, Policy and the Integrity of Social Life (2010)

Civil Aviation System



Machine learning system

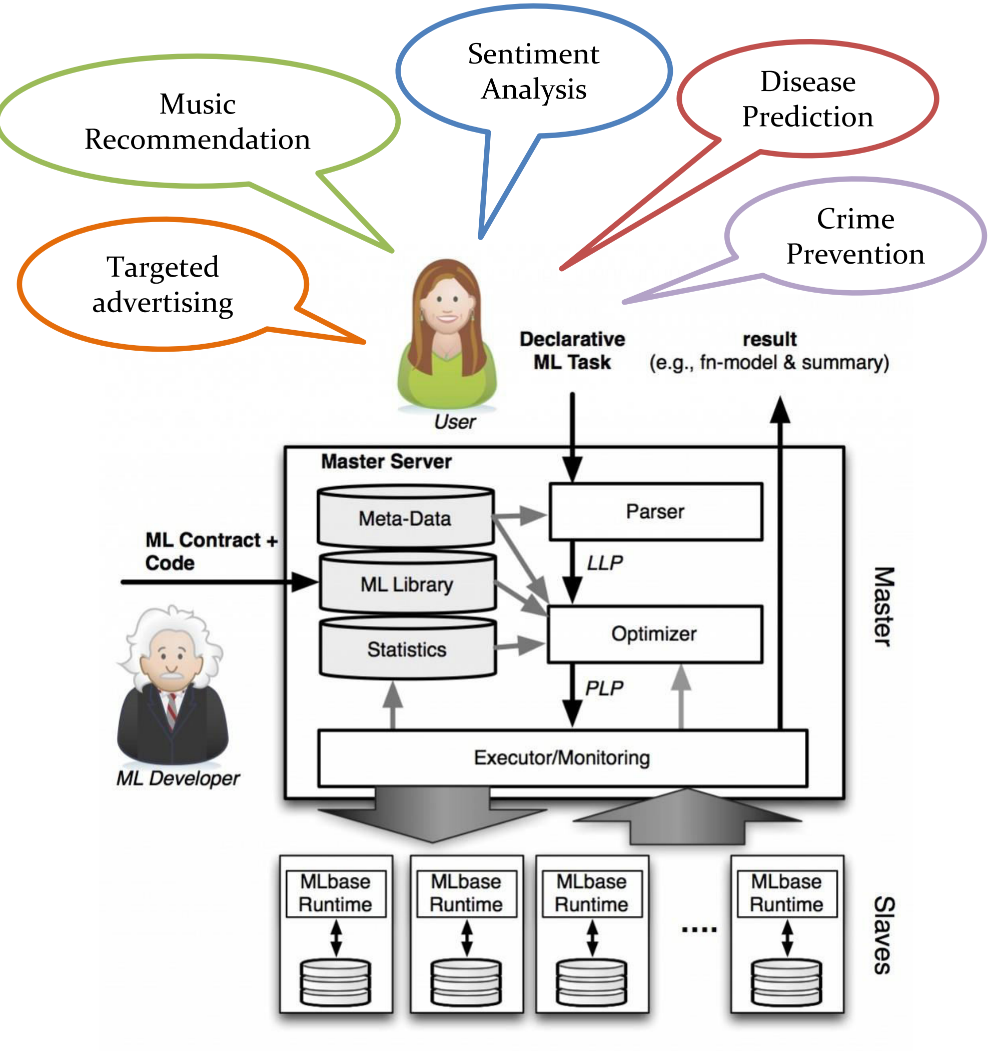
"One major trend driving this expansion is a growing concern with the **environment** in which a machine-learning algorithm operates […]. Broadly speaking, environments provide **various resources** to a learning algorithm and place **constraints** on those resources. Increasingly, machine-learning researchers are formalizing these **relationships**, aiming to design algorithms that are provably effective in various environments and explicitly allow users to express and control trade-offs among resources."

The environment may refer to:

- Large-scale parallel and distributed computing platforms
- Various sources of data (with privacy and ownership concerns)
- The activities associated to data (functions, human factors)
- Other machine learning systems or agents (cooperative or adversarial).

M.I. Jordan and T.M. Mitchel,
Machine learning: Trends, perspectives, and prospects (2015)

Machine Learning in the Big Data Era



See project webpage: http://mlbase.org/

Consider **MLbase project**: to make ML accessible to a broad audience of users and applicable to various data sets and application. Two key **challenges** are:

- To provide a platform for ML researchers and professionals:
  - To allow researchers to inspect execution and experiments;
  - to explain ML task in a simple declarative way / visualize results.
- To build distributed ML algorithms without knowing the details about data partitioning, message passing, etc

Other examples: Google Prediction, Apache Mahout

T. Kraska et al. *Mlbase: A distributed Machine Learning System* (2013)

ML for Science and Society

"Because machine learning is primarily influencing the broader world through its implementation in a wide range of applications, rather than through its novel specialized algorithms or theory, **aspects beyond algorithms** and theory can be (and often are) the most important for knowledge discovery"

C. Rudin, K. Wagstuff,
Machine Learning for Science and Society (2014)

Relevant aspects:

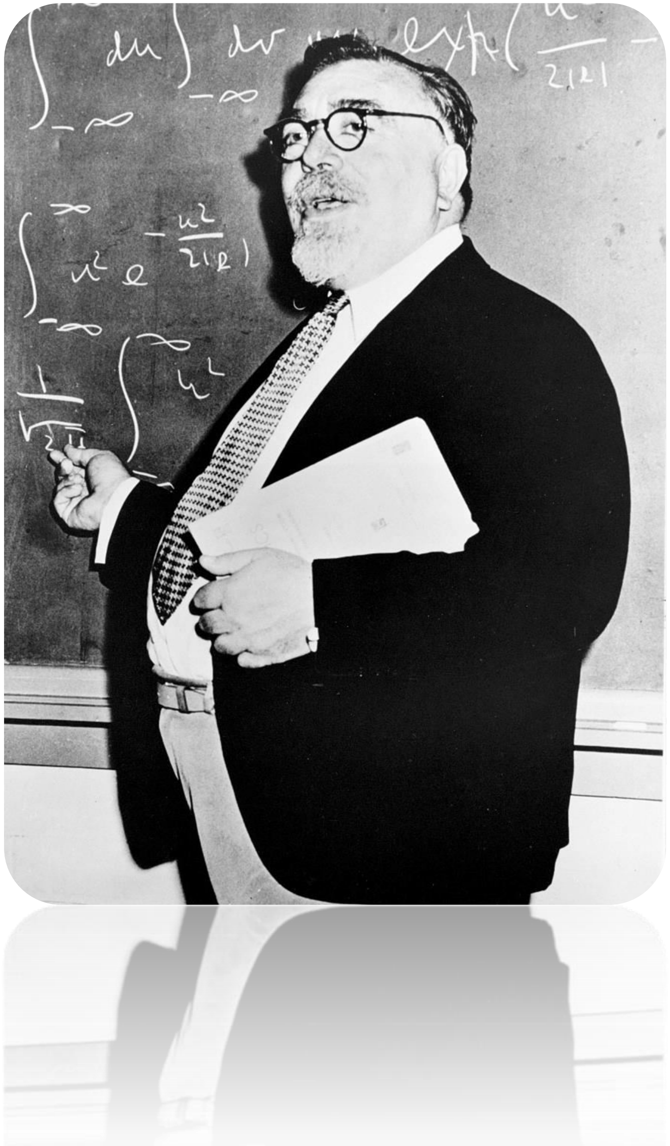
- Domain knowledge/experts
- New performance measures (beyond benchmarks)
- Human-interaction/feedbacks
- Privacy
- Fairness
- Accountability

Living Effectively

“The process of receiving and of using information is the process of our adjusting to the contingencies of the outer environment, and of our **living effectively within** that **environment**.

The needs and the complexity of modern life make **greater demands on this process of information** than even before, and our press, our museums, our scientific laboratories, our universities, our libraries and textbooks, are obliged to meet the needs of this process or fail in their **purpose**.

To live effectively is to live with adequate information. Thus, **communication** and **control** belong to the essence of **man’s inner life**, even as they belong to his live in **society**.”



N. Wiener, *The Human Use of Human Being* (1954)

Neutral artifacts?

- In the development of a socio-technical system many social and ethical aspects are involved

- Not only is it hard to develop “neutral” artifacts, but also it is almost impractical to separate technical artifacts from their (social) context.

- As for Machine learning these considerations suggest that:

  ML systems deal with human life → data on personal and social life (**privacy**, identity, property)

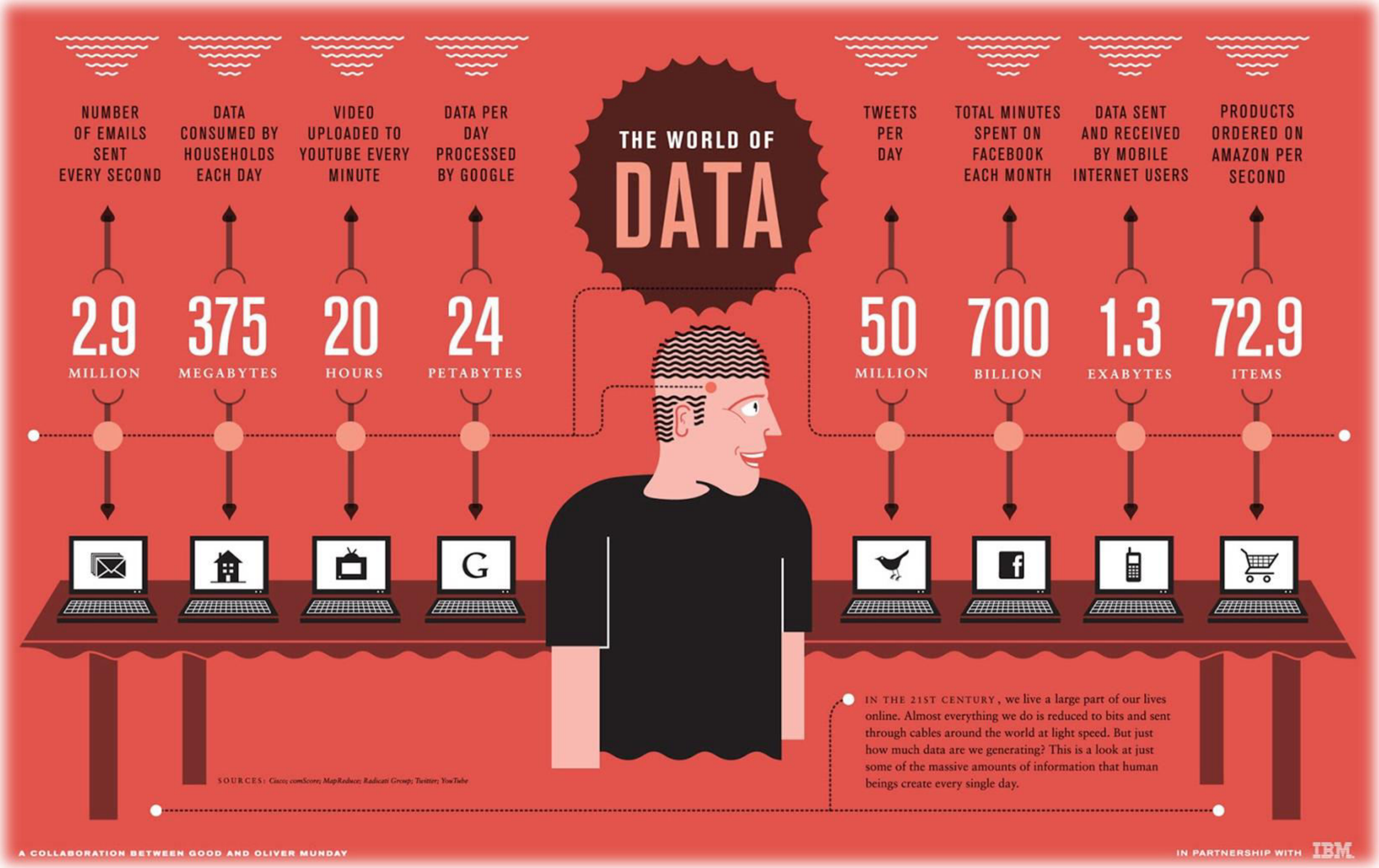
  ML systems impact human life → decision-making (**fairness**, policy making); evaluation criteria in real life application (knowledge experts, human-in-the-loop)

  ML systems interact with humans → human-machine interaction, accountability (black-box testing), transparency (ML accessible to large audience)

Privacy

B. Luini, *An Allegory of Modesty and Vanity*, 1520

Privacy in the big data era


THE WORLD OF
DATA
NUMBER OF EMAILS SENT EVERY SECOND
2.9 MILLION
DATA CONSUMED BY HOUSEHOLDS EACH DAY
375 MEGABYTES
VIDEO UPLOADED TO YOUTUBE EVERY MINUTE
20 HOURS
DATA PER DAY PROCESSED BY GOOGLE
24 PETABYTES
G
TWEETS PER DAY
50 MILLION
TOTAL MINUTES SPENT ON FACEBOOK EACH MONTH
700 BILLION
f
DATA SENT AND RECEIVED BY MOBILE INTERNET USERS
1.3 EXABYTES
PRODUCTS ORDERED ON AMAZON PER SECOND
72.9 ITEMS
A COLLABORATION BETWEEN GOOD AND OLIVER MUNDAY
IN PARTNERSHIP WITH IBM

IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human beings create every single day.

SOURCES: *Cisco; comScore; MapReduce; Radicati Group; Twitter; YouTube*

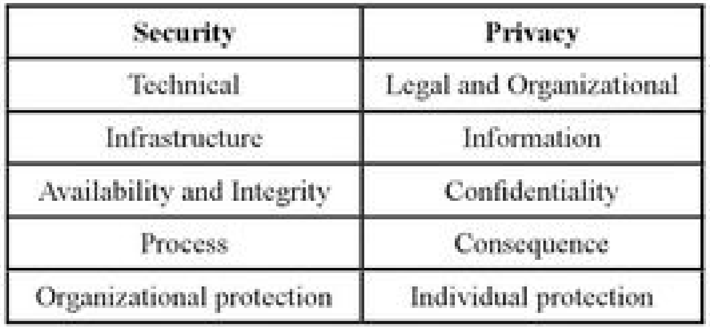
http://exploringbigdata.blogspot.it/

Privacy *vs.* Security

Common sense distinction:

- **Security** is concerned with technical issues (Engineering, Computer science)

- **Privacy** is concerned with legal and organizational issues (Business, Management, Law)



Security	Privacy
Technical	Legal and Organizational
Infrastructure	Information
Availability and Integrity	Confidentiality
Process	Consequence
Organizational protection	Individual protection

R. Subramanian, *Computer Security, Privacy and Politics* (2008)

More specifically:

- Security → access control and authentication

- Privacy → to release all the information **protecting** the **identities** of the people who are the subjects of the data (**anonymity**)

L. Sweeney, *K-Anonymity: A model for protecting privacy* (2002)

Privacy and Personal Data

- **Privacy** (Warren and Brandeis, *The Right to Privacy*, 1890): the value of privacy "is found not in the right to take the profits arising from publication, but in the peace of mind or the relief afforded by the ability to prevent any publication at all".

- **Personal Data** (EU Directive 95, art. 2): They "mean any information relating to an identified or identifiable natural person ('data subject'); an **identifiable person** is one who can be identified , directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological , mental , economic, cultural or social identity"

Some notions of privacy

- The right to be let alone, a way to highlight that privacy involves the respect to live one's life free from intrusion
- **Limited access to the self**: right to decide how much knowledge of personal thought /feeling/private doings and affairs…the public at large shall have (**control**)
- Secrecy (self-interested economic behavior) / Intimacy: from self-protection to the creation of intimate relationships
- Personhood: protecting the integrity of personality (links to other moral values, e.g. personal dignity, autonomy, etc.)

D. Solove, *Conceptualizing Privacy* (2002)

Emerging points:

- Definitions are based on some distinct features and are either too restrictive or too vague
- contraposition between **private** and **public** spheres

Policy Level

Often privacy policy are formulated in terms of **access** and **control**

> Privacy is “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others”

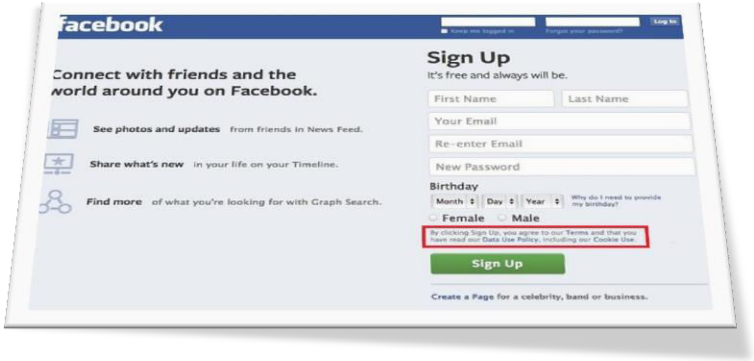
A. Westin, *Privacy and Freedom* (1967)

The underlying perspective

> Privacy Self-Management: “It attempts to be **neutral about substance** — whether certain forms of collecting, using, or disclosing personal data are good or bad — and instead focuses on whether people consent to various privacy practices.”

D. Solove, *Privacy Self-Management and the Consent Dielmma* (2012)

Notice & Consent



facebook

Keep me logged in · Forgot your password? · Log In

Connect with friends and the world around you on Facebook.

See photos and updates from friends in News Feed.

Share what's new in your life on your Timeline.

Find more of what you're looking for with Graph Search.

Sign Up

It's free and always will be.

First Name · Last Name

Your Email

Re-enter Email

New Password

Birthday

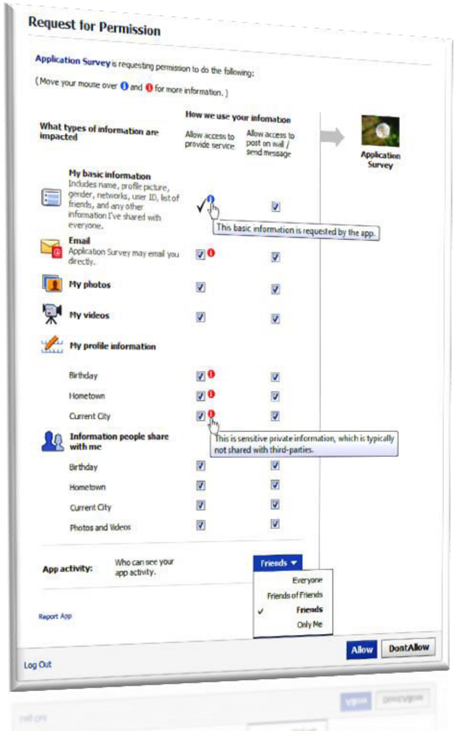
Month · Day · Year · Why do I need to provide my birthday?

○ Female ○ Male

By clicking Sign Up, you agree to our Terms and that you have read our Data Use Policy, including our Cookie Use.

Sign Up

Create a Page for a celebrity, band or business.



Request for Permission

Application Survey is requesting permission to do the following:

(Move your mouse over ⓘ and ⓘ for more information.)

What types of information are impacted	How we use your infomation Allow access to provide service	 Allow access to post on wall / send message
My basic information Includes name, profile picture, gender, networks, user ID, list of friends, and any other information I've shared with everyone.	✓ ⓘ	☑
Email Application Survey may email you directly.	☑ ⓘ	☑
My photos	☑	☑
My videos	☑	☑
My profile information		
Birthday	☑ ⓘ	☑
Hometown	☑ ⓘ	☑
Current City	☑ ⓘ	☑
Information people share with me		
Birthday	☑	☑
Hometown	☑	☑
Current City	☑	☑
Photos and Videos	☑	☑

This basic information is requested by the app.

This is sensitive private information, which is typically not shared with third-parties.

App activity: Who can see your app activity. Friends ▾

Everyone

Friends of Friends

✓ **Friends**

Only Me

Report App

Allow · DontAllow

Log Out

Cognitive and **structural limitations**:

- Most people do not read privacy policy in a regular way
- Many privacy policies are too vague on future uses
- Many people would need more details to make informed choices but additional information increases notice complexity
- Often people lack background knowledge and expertise to evaluate the consequence of their consent
- There are too many entities that collect and use personal data (“it would cost $ 781 billion in lost productivity if everyone were to read every privacy policy at websites they visited ina one-year period”)
- Little bits of innocuous data can say a lot in combination
- …

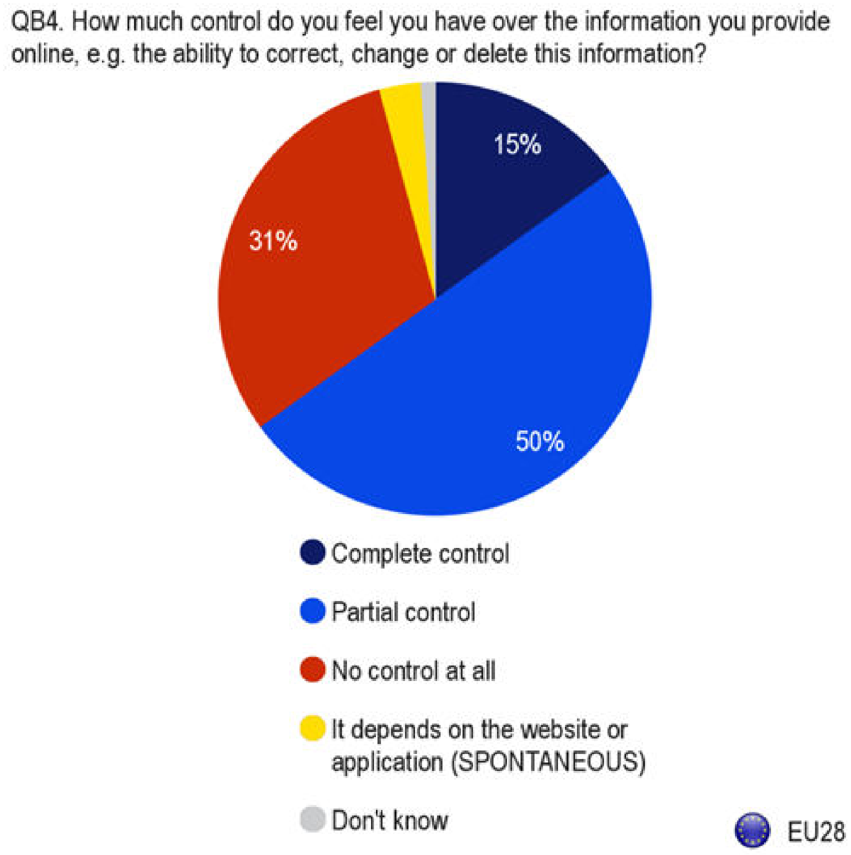
D. Solove,
Privacy Self-Management and the Consent Dielmma (2012)

Data Protection Eurobarometer

The Eurobarometer survey, conducted in March 2015, asked 28 000 EU citizens what they think about the protection of their personal data.

Some **results**:

- Only a minority (15%) feel they have complete control over the information they provide online; 31% think they have no control over it at all.
- Two-thirds of respondents (67%) are concerned about not having complete control over the information they provide online.
- A majority of respondents are concerned about the recording of their activities via payment cards and via mobile phones (55% in both cases).


QB4. How much control do you feel you have over the information you provide online, e.g. the ability to correct, change or delete this information?

http://ec.europa.eu/justice/newsroom/data-protection/news/240615_en.htm

Technical level

With the increasing availability of large-scale data and large-scale data processing privacy protection has become a cause of concern in several areas:

- Database
- Statistics
- Network science
- Machine learning

Problem formulation: To allow the release of private data preventing the re-identification of the data subject.

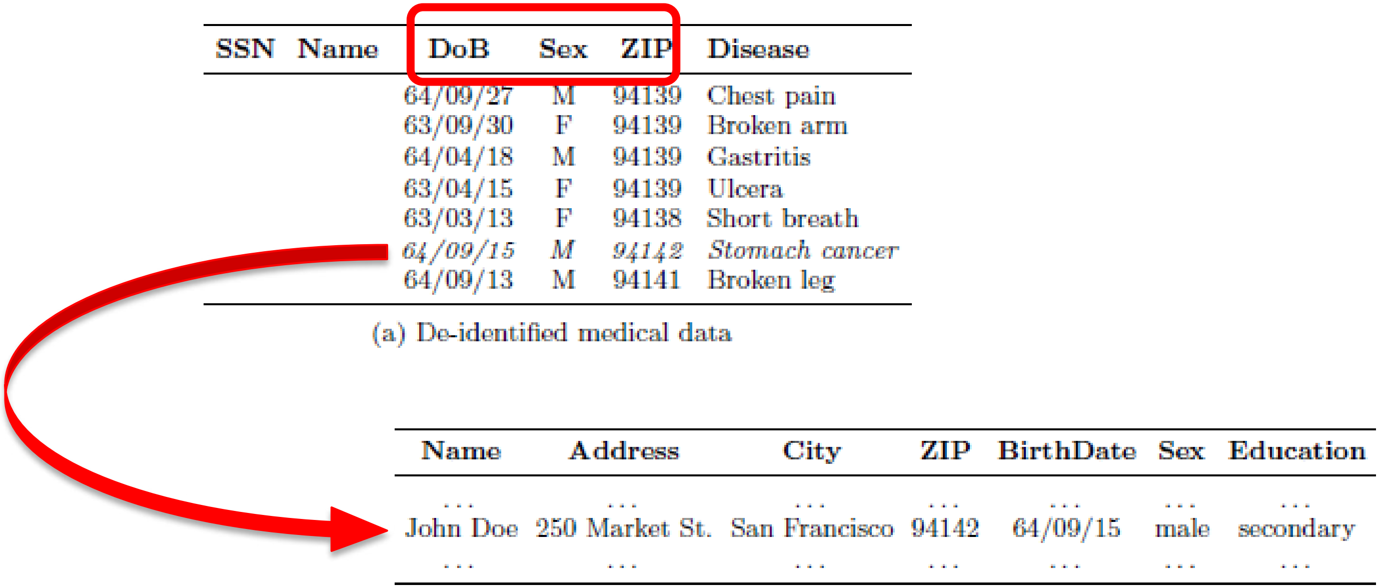
Privacy is associated to **anonymization** which is often obtained through data de-identification.

De-identified data: data in which “all explicit identifiers, such as SSN, name, address, and telephone number, are removed, generalized, or replace with made-up alternative”

L. Sweeney, *Weaving technology and policy together to maintain confidentiality* (1997)

Failing anonymity

The data holder may remove some identifiers but the table could contain sets of attributes (**quasi-identifiers**) that, in combination, can be linked to external available information (auxiliary information) and used to re-identify data subjects.



SSN	Name	DoB	Sex	ZIP	Disease
		64/09/27	M	94139	Chest pain
		63/09/30	F	94139	Broken arm
		64/04/18	M	94139	Gastritis
		63/04/15	F	94139	Ulcera
		63/03/13	F	94138	Short breath
		64/09/15	*M*	*94142*	*Stomach cancer*
		64/09/13	M	94141	Broken leg

(a) De-identified medical data

Name	Address	City	ZIP	BirthDate	Sex	Education
...	...	...	...	...	...	...
John Doe	250 Market St.	San Francisco	94142	64/09/15	male	secondary
...	...	...	...	...	...	...

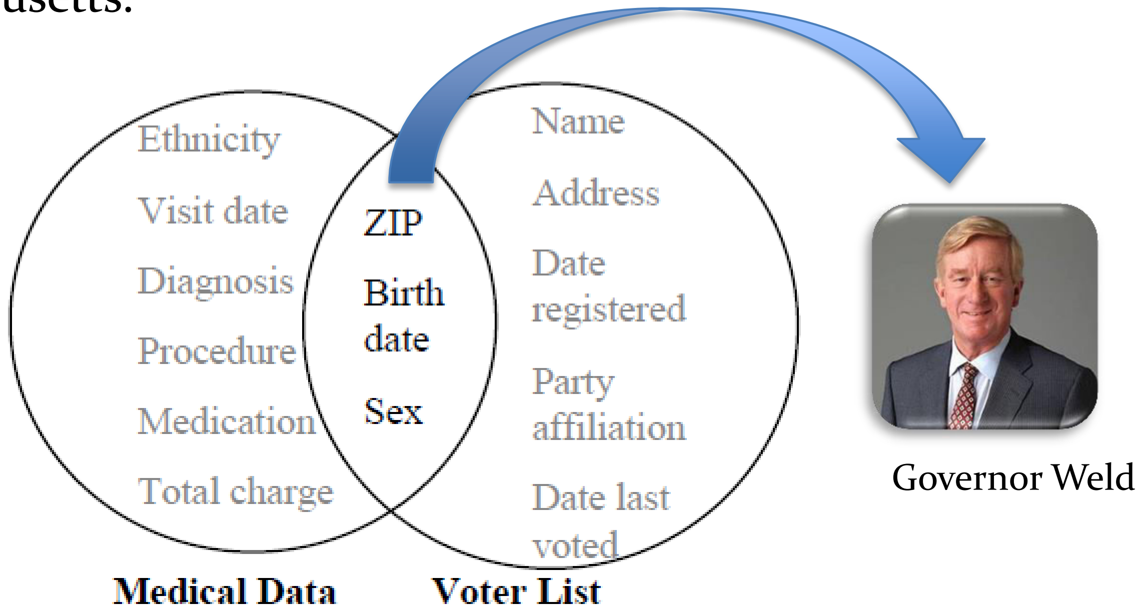
(b) Municipality register

S. De Capitani di Vimercati *et al.*, *Anonymization of statistical data*, 2011

A famous story of re-identification

The Massachusetts Group Insurances Commission (GIC) collected patient-specific data and gave a copy of its database to researches and industries. Data had been anonymized by removing attributes containing patients' name, address, and social security number (SSN). Sweeney purchased the complete voter registration list for Cambridge Massachusetts and…

….combining this data with the GIC records re-identified William Weld, the Governor of Massachusetts.



L. Sweeney, *K-Anonymity: A model for protecting privacy* (2002)

New York Times article on Ms. Arnold

In 2006 AOL Research released internet records amounting to 20 million search queries (650,000 users in a 3-month period) with the aim of “making its content and products freely available to all consumers […] creating opportunities for researchers in academia and industry alike”

AOL released deidentified data, e.g. it removed usernames, IP addresses and used a unique identifier (such as 4417749) to link all of a user’s queries.

“And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern.” (*New York Times*, August 9, 2006)

A Face Is Exposed for AOL Searcher No. 4417749

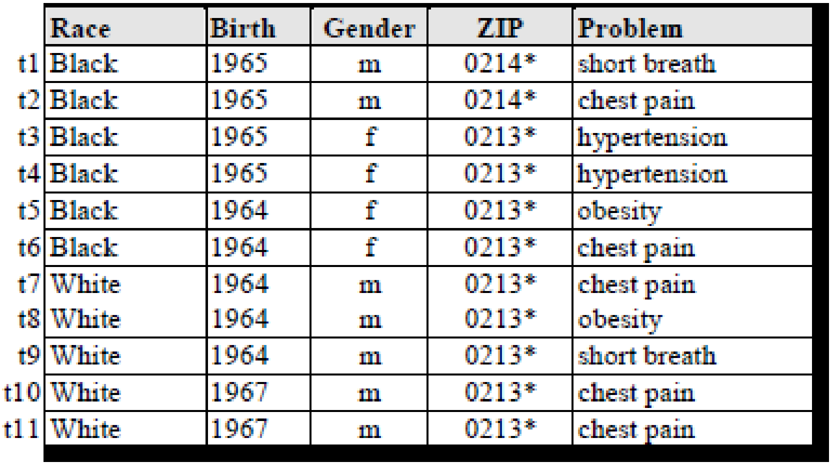
By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

O. Heffetz and K. Ligett, *Privacy and Data-Based Research* (2013)

K- Anonymity

K-anonymity has been introduced to reduce the weakness of classical de-identified data (e.g. GIC, AOL). It characterizes the degree of data protection with respect to inference by linking.

Def: Let T $(A_1,\ldots,A_n)$ be a table and QI_T be the quasi-identifier associated with it. T is said to satisfy *k*-anonymity if and only if each sequence of values in $T[QI_T]$ appears with at least *k* occurrences in $T[QI_T]$.



	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

Figure 2 Example of *k*-anonymity, where *k*=2 and QI={*Race, Birth, Gender, ZIP*}

L. Sweeney, *K-Anonymity: A model for protecting privacy* (2002)

K- Anonymity (example)

A release of data is said to satisfy k-anonymity if each released record has at least (k-1) other records also visible in the release whose values are indistinct over the quasi-identifier.

Example: How to reach k-anonymity by generalizing and suppressing values.

Name	Gender	Age	State	Disease
Mary	Female	21	Ohio	Viral Infection
John	Male	65	Illinois	Cancer
Susan	Female	27	Ohio	No illness
Jack	Male	68	Illinois	Hypertension
Tom	Male	62	Illinois	Heart attack
James	Male	18	New York	No illness
Lucas	Male	17	New York	Viral infection
Emily	Female	29	Ohio	Cancer

Non-anonymized table

Name	Gender	Age	State	Disease
*	Female	$20 < Age \leq 30$	Ohio	Viral Infection
*	Male	$60 < Age \leq 70$	Illinois	Cancer
*	Female	$20 < Age \leq 30$	Ohio	No illness
*	Male	$60 < Age \leq 70$	Illinois	Hypertension
*	Male	$60 < Age \leq 70$	Illinois	No illness
*	Male	$Age < 20$	New York	Heart attack
*	Male	$Age < 20$	New York	Viral infection
*	Female	$20 < Age \leq 30$	Ohio	Cancer

Anonymized table
(2-anonymity with respect to gender, age and state)

De-anonymizing data

In 2008 Narayanan and Shmatikov pointed out that k-anonymity has several weaknesses:

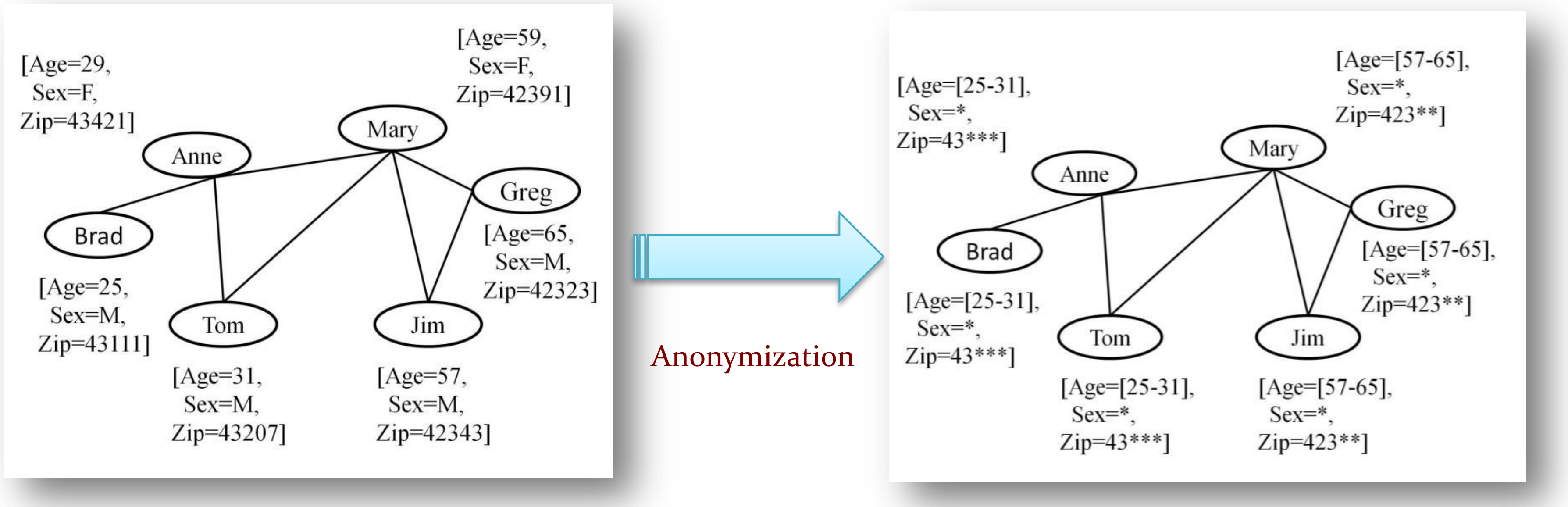
- The adversary may have much auxiliary knowledge (more than just the quasi-identifiers);
- K-anonymization fails on high-dimensional datasets.

Analyzing **Netflix competition** (for improving the company's algorithm that predicts user ratings of films) they claimed that:

> "an adversary who knows a little bit about some subscriber can easily identify her record if it is present in the dataset, or, at the very least, identify a small set of records which include the subscriber's record."

A. Narayanan and V. Shamatikov, *Robust De-anonymization of Large Sparse Datasets* (2008)

What about social networks?



Intuitively, revealing only connections and masking nodes identity should safeguard individual privacy and research purposes (study of graph properties)…

De-anonymizing social networks

In 2007 Backstrom, Dwork and Kleinberg suggested that "**anonymous social network data almost never exists in the absence of outside context**, and an adversary can potentially combine this knowledge with the observed structure to begin compromising privacy [...]Moreover, such an adversary may in fact be a user (or set of users) of the system that is being anonymized."

L. Backstrom, C. Dwork and J. Kleinberg, *Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography* (2007)

In 2008 Narayanan and Shmatikov demonstrated that: "a third of the users who are verifiable members of both Flickr and Twitter1 can be recognized in the completely anonymous Twitter graph with only 12% error rate, even though the overlap in the relationships for these members is less than 15%."

A. Narayanan and V. Shamatikov, *De-anonymization Social Newtorks* (2008)

Differential privacy

Differential privacy tries to provide privacy guarantee free from assumptions about auxiliary information.

Differential privacy provides privacy by process introducing randomness. Now, consider:

- Two **neighbouring databases**, **D** and **D'**: they are identical except that one of them has an additional row/record (each record refers to one individual);
- A computation (function) **K** on such databases

A function **K** provides ε–differential privacy if:

$$\mathrm{Prob}[K(D) = S] \leq e^{\epsilon} \cdot \mathrm{Prob}[K(D') = S]$$

O. Heffetz and K. Ligett, *Privacy and Data-Based Research* (2013)

Privacy mechanism

Intuitively, differential privacy guarantees that a randomized algorithm behaves similarly on similar input databases.

Bayesian interpretation: “an observer with access to the output of a differentially private function should draw almost the same conclusions whether or not one individual's data are included in the analyzed database, regardless of the observer's prior.”

Differential privacy is a property of the **function** not of the outcome.

O. Heffetz and K. Ligett, *Privacy and Data-Based Research* (2013)

Privacy-by-design

A methodological approach to engineering systems stating that privacy should be embedded into the whole lifecycle of IT systems, from the early stages to their ultimate deployment (analogous to the **value sensitive design** approach).

Foundational Principles:

1. Proactive not reactive; Preventative not remedial
2. Privacy as the default setting
3. Privacy embedded into design
4. Full functionality – positive-sum, not zero-sum
5. End-to-end security – full lifecycle protection
6. Visibility and transparency – keep it open
7. Respect for user privacy – keep it user-centric

A. Cavoukian *Privacy by Design* (2009)

Debate on privacy

General perspective

- How Big Data challenges privacy
- Reflecting on the social and cultural/contextual value of privacy

Policy level

- From the "right to control over information" to the "right to be forgotten" (EU Directive 2016)
- Privacy as Contextual Integrity

Technical level

- Attempts which go beyond anonymization (differentially private machine learning, privacy in topic-dependent social networks, privacy via exposure, etc.)
- Privacy-by-design in data mining (e.g. in different application scenarios)

Fairness

Giotto, *Justice*, 1306

Fairness

Fairness / to be fair (dictionary):

- Something in accordance with rules or standards;
- Being appropriate in the circumstance;
- To **behave without** cheating or **trying to achieve an unjust advantage**

J. Rawls argues that justice emerges from a fair cooperation among free and equal moral agents (Justice as Fairness):

"The principles of justice for the basic structure of society are the object of the original agreements. They are principles that free and rational persons concerned to further their own interest would accept in an initial position of equality as defining the fundamental terms of their association. [...] This way of regarding the principles of justice I shall call justice as fairness."

J. Rawls, *A Theory of Justice* (1971)

Fairness → equality / freedom / conditions for building a society

Discrimination

“Discrimination refers to an **unjustified difference** in treatment on the basis of any physical or cultural trait, such as sex, ethnic origin, religion or political opinions”

Human rights laws prohibit discrimination based on sex, gender, ethnicity, skin colour, social origin, language, religion or belief, political or other personal opinion, disability, illness, marital status or age…

In national and international legislations discrimination is often associated to the protection of vulnerable groups:

- **Protected group:** a category of people that could be discriminated on the ground of sex (pregnant women), creed (religious minorities) etc.

- Discrimination is framed as “an unjustified distinction of individuals based on their membership, or perceived membership, in a certain group or category”

A. Romei, S. Ruggieri. *A multidisciplinary survey on discrimination analysis. (2014)*

How big data is unfair

“As we’re on the cusp of using machine learning for rendering basically all kinds of consequential decisions about human beings in domains such as education, employment, advertising, health care and policing, it is important to understand why *machine learning is not, by default, fair or just in any meaningful way.*

This runs counter to **the widespread misbelief that algorithmic decisions tend to be fair**, because math is about equations and not skin color.”

Example of this oversimplification:

> “After all, as the former CPD [Chicago Police Department] computer experts point out, the algorithms in themselves are neutral. ‘This program had absolutely nothing to do with race… but multi-variable equations,’ argues Goldstein. Meanwhile, the potential benefits of predictive policing are profound.” (G. Tett, Financial Times, 2014)

M. Hardt, *How big data is unfair.*
Understanding unintended sources of unfairness in data driven decision making (2014)

Political concerns

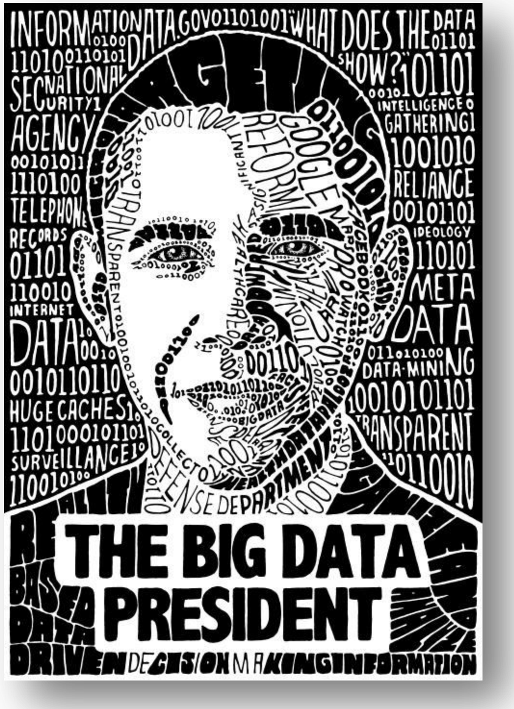
In 2014, President President Obama called for a 90-day review of Big Data.

One of the main finding of this report was that:

- “big data technologies can cause societal harms beyond damages to privacy”;

- big data analytics can create an “opaque decision-making environment”;

- “big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace.”



Big Data: Seizing Opportunities, Preserving Value (2014)

Discrimination in machine learning

Machine learning and data mining represent a form of statistical discrimination. Basically they aim to end up with classification/groupings which make sense.

In the machine learning procedures there are several mechanisms/steps which can play a role in the the production of discriminatory results:

- Defining the “Target Variable” and “Class Labels”
- Training Data
- Feature selection
- Proxies
- Masking

Barocas S. and Selbst A., Big *Data's Disparate Impact* (2014)

Discrimination in online ad delivery

"A Google search for a person's name, such as "Trevon Jones", may yield a personalized ad for public records about Trevon that may be neutral, such as "Looking for Trevon Jones?…", or may be suggestive of an arrest record, such as "Trevon Jones, Arrested?"…"

Latanya Sweeney, *Discrimination in Online Ad Delivery* (2013)

Using a sample of racially associated names Sweeney found out that:

- First names primarily assigned to black babies generated ads suggestive of an arrest in 81 to 86 % of name searches

- Those assigned at birth primarily to whites generated more neutral copy: "arrest" appeared in 23 to 29 %

AdFisher

- Adfisher is a program that simulates browsing behavior and collects information about the ads returned after Google searches

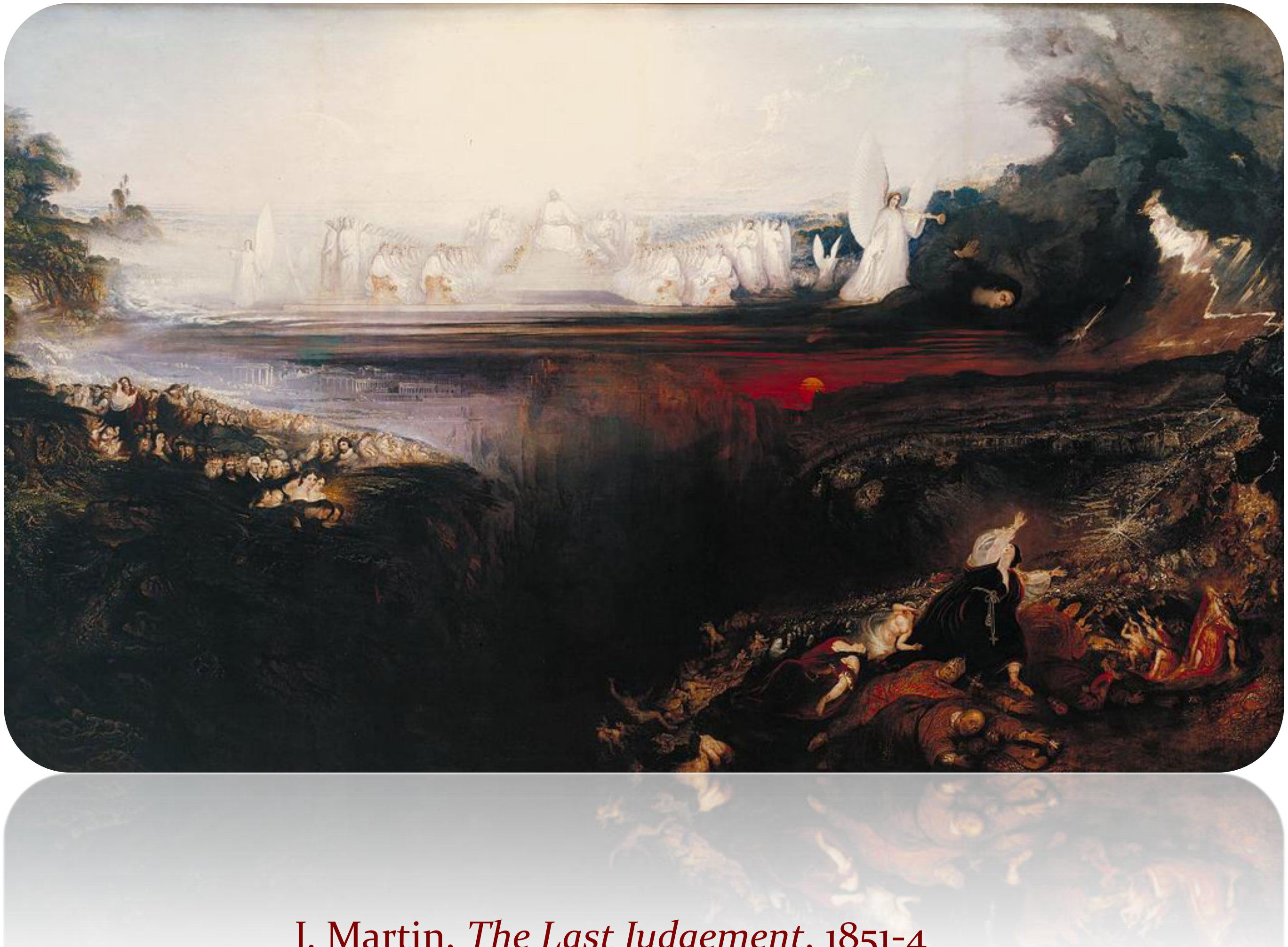
- Experiments:

  - 1.000 fakes users (half women and half men)
  - Researchers simulated users visited websites concerning employment and collected data about which ads they were shown subsequently
  - Finding: more ads related to higher paying jobs were served to men

For instance an executive position paid upward of $ 200,000 per year was shown 1,852 times to the male group and only 318 times to the female group.

Anupam Datta's homepage: http://possibility.cylab.cmu.edu/adfisher/

Concluding Remarks



J. Martin, *The Last Judgement*, 1851-4

Technology and social progress

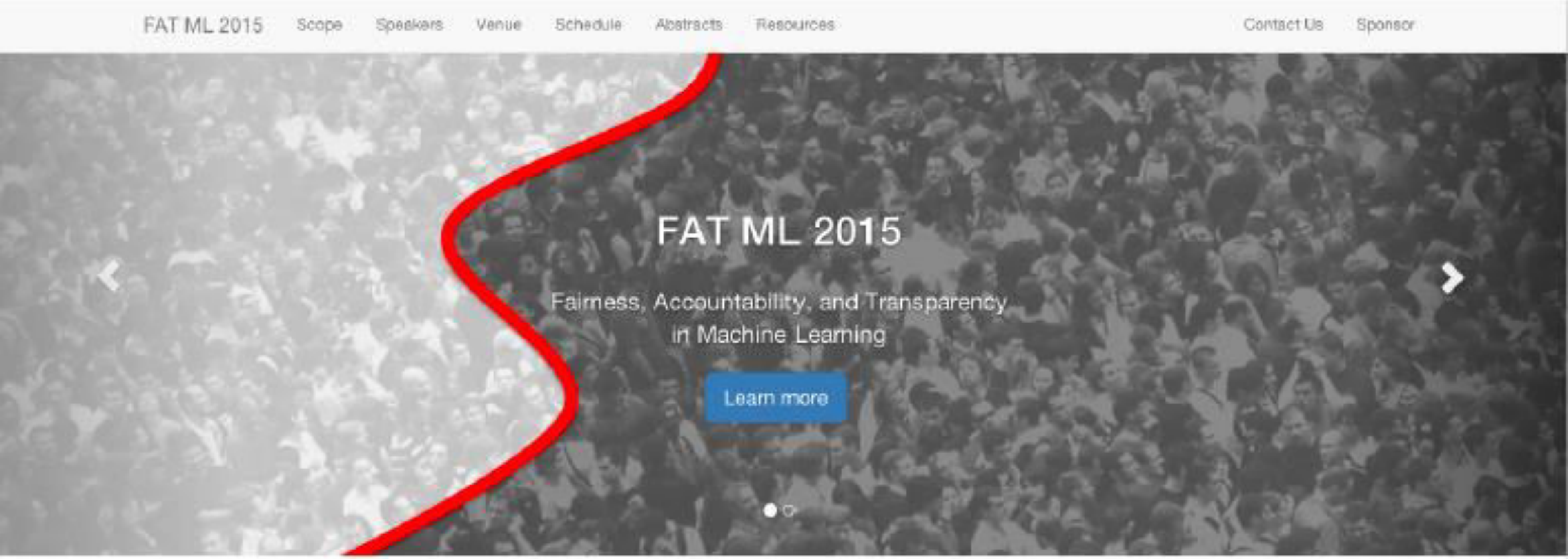
“The glorification of ‘pure’ science…is a rationalization of an escape; it marks a construction of an asylum of refuge, a shirking of responsibility.”

J. Dewey, *The public and its problems* (1927)

To construct a sense of scientific progress that sounds genuinely like progress, with all its positive connotations, we are going to have to embed science even more fully in society. We will need to ask which science will provide us with a better society, and, which science will perhaps undermine it.

H. Douglas, *Pure science and the problem of progress* (2014)

FAT - ML



FAT ML 2015

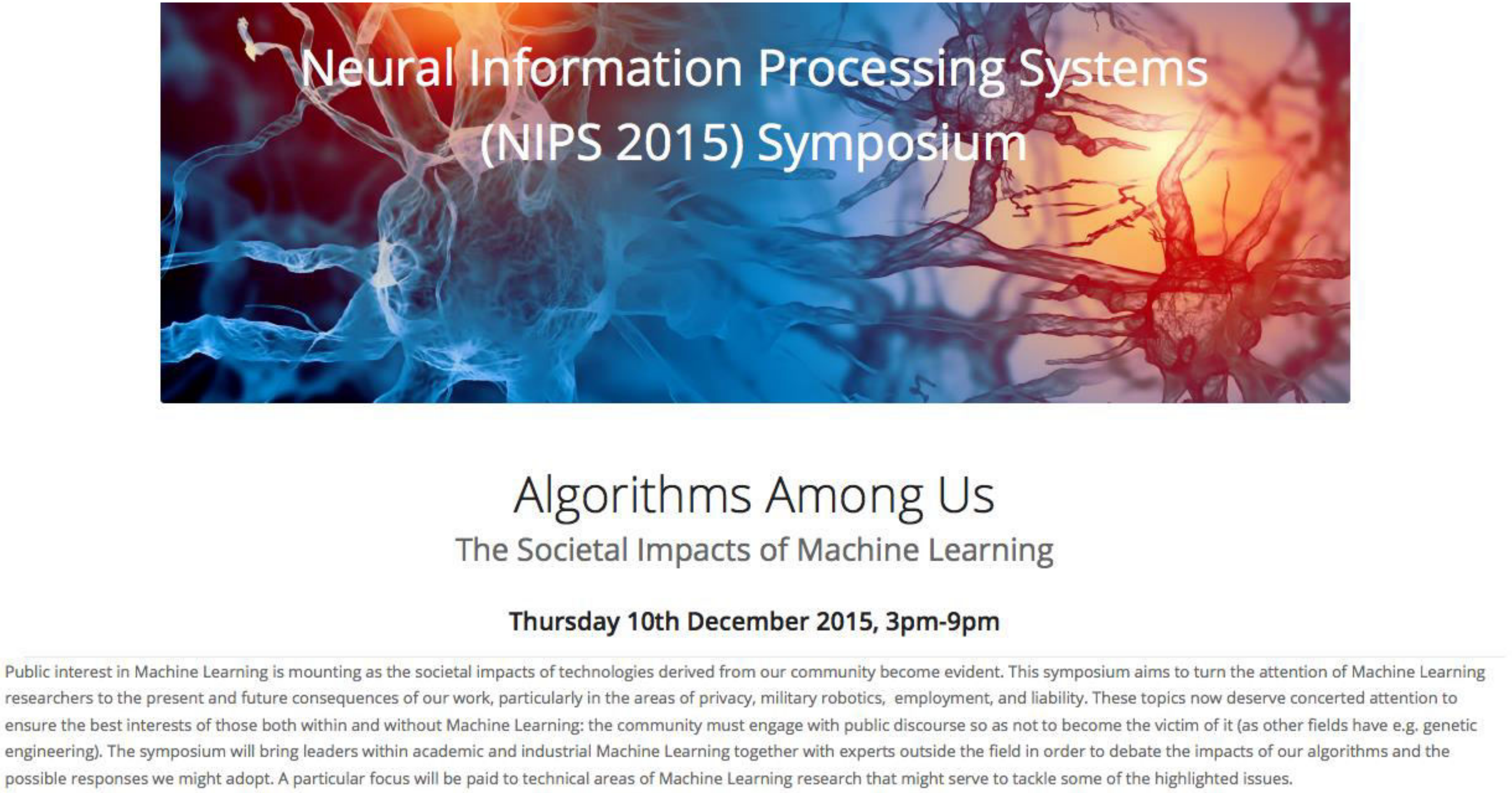
Fairness, Accountability, and Transparency in Machine Learning

FAT ML = Fairness, Accountability and Transparency in Machine Learning
Present at NIPS 2014 and ICML 2015

Organizers: S. Barocas, S. Friedler, M. Hardt, J. Kroll, S. Venkatasubramanian, H. Whallach

http://www.fatml.org/

Algorithms among us



Neural Information Processing Systems (NIPS 2015) Symposium

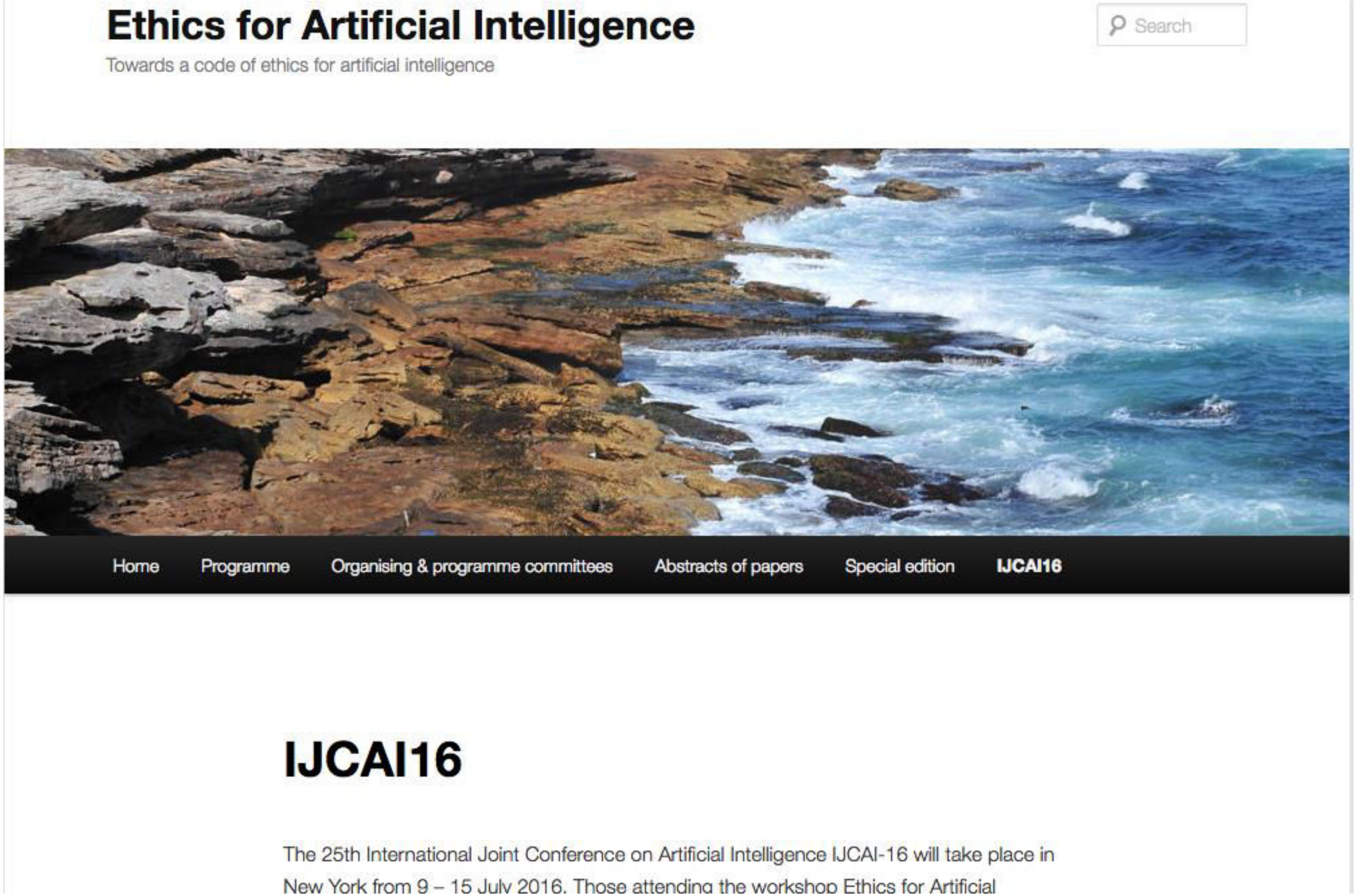
Algorithms Among Us

The Societal Impacts of Machine Learning

Thursday 10th December 2015, 3pm-9pm

Public interest in Machine Learning is mounting as the societal impacts of technologies derived from our community become evident. This symposium aims to turn the attention of Machine Learning researchers to the present and future consequences of our work, particularly in the areas of privacy, military robotics, employment, and liability. These topics now deserve concerted attention to ensure the best interests of those both within and without Machine Learning: the community must engage with public discourse so as not to become the victim of it (as other fields have e.g. genetic engineering). The symposium will bring leaders within academic and industrial Machine Learning together with experts outside the field in order to debate the impacts of our algorithms and the possible responses we might adopt. A particular focus will be paid to technical areas of Machine Learning research that might serve to tackle some of the highlighted issues.

Ethics and AI



Ethics for Artificial Intelligence

Towards a code of ethics for artificial intelligence

IJCAI16

The 25th International Joint Conference on Artificial Intelligence IJCAI-16 will take place in New York from 9 – 15 July 2016. Those attending the workshop Ethics for Artificial

Artificial Intelligence and Life in 2030

The One Hundred Year Study on Artificial Intelligence, launched in the fall of 2014

2016 Report

- Preface
- Executive Summary
- Overview
- Section I: What is Artificial Intelligence?
- Section II: AI by Domain
- Section III: Prospects and Recommendations for Public Policy
- Appendix I: A Short History of AI

Related Documents

- 2015 Study Panel Charge
- June 2016 Interim Summary
- Download Full Report

2016 Report

Welcome to the 2016 Report

The Human Use of Machine Learning

December 16, 2016 – ECLT, Venice

European Centre for Living Technology, Ca' Minich, S. Marco 2940 30124 Venezia, ITALY

Organization

The European Centre for Living Technology (ECLT) is organized as an interuniversity consortium, currently with 17 institutional affiliates.
The Centre is an international meeting point for the development both of basic research and the creation of new technologies for the productive sector.

Research

Seminars

Revisiting Invention in Pottery Making

Speaker: Sander van der Leeuw, Arizona State University
Date: 3 June 2016, ... »

Geometry and Painting in Europe and East Asia

Speaker: Felipe Cucker, City University of Hong Kong
Date: 27 May 2016, ... »

The Decoupling Assumption in Large

www.unive.it/eclt

Short Bibliography

- Vermaas, Kroes, van de Poel, Franssen, Houkes, 2011
- A. Rosenblueth, N. Wiener and J. Bigelow, *Behavior, Purpose and Teleology,* 1943
- L. Sweeney, *K-Anonymity: A model for protecting privacy*, 2002
- O. Heffetz and K. Ligett, *Privacy and Data-Based Research,* 2013
- S. De Capitani di Vimercati *et al.*, *Anonymization of statistical data*, 2011
- Latanya Sweeney, *Discrimination in Online Ad Delivery* (2013)
- Barocas S. and Selbst A., Big *Data's Disparate Impact* (2014)
- H. Douglas, *Pure science and the problem of progress* (2014)
- A. Romei, S. Ruggieri. *A multidisciplinary survey on discrimination analysis. (2014*)

Thanks!

M. Chagall, *The acrobat*, 1930