Big Data and its implication for regulation

Roberto V. Zicari
Z-Inspection® Initiative
http://z-inspection.org

Workshop on Regulatory Requirements in the context of Industry 4.0
Hanoi Province , November 12, 2021

Session Content

ം

1. **A gentle introduction to Big Data**

2. **European Commission: Proposed Legal Framework on AI.**

3. **European Commission: The General Data Protection Regulation (GDPR)**

.

BIG DATA

The term "**Big Data**" refers to large amounts of different types of data produced with high velocity from a high number of various types of sources.

Big Data

Every day, 2.5 quintillion bytes of data are created.

This data comes from:

digital pictures, videos, posts to social media sites, intelligent sensors, purchase transaction records, cell phone GPS signals to name a few.

In 2013, estimates reached 4 *zettabytes* of data generated worldwide

ᴄʒ   Mary Meeker and Liang Yu, Internet Trends, Kleiner Perkins Caulfield Byers, 2013, http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013.

How Big is Big Data?

1 *petabyte* is
1,000 terabytes (TB)= 10^{15} bytes

1 *zettabyte* is
1,000 000,000,000,000,000,000 bytes
== 10^{21} bytes

How Big is Big Data?

ℰℬ

"Imagine that every person in the United States (ca. 320,590,000) took a digital photo every second of every day for over a month.
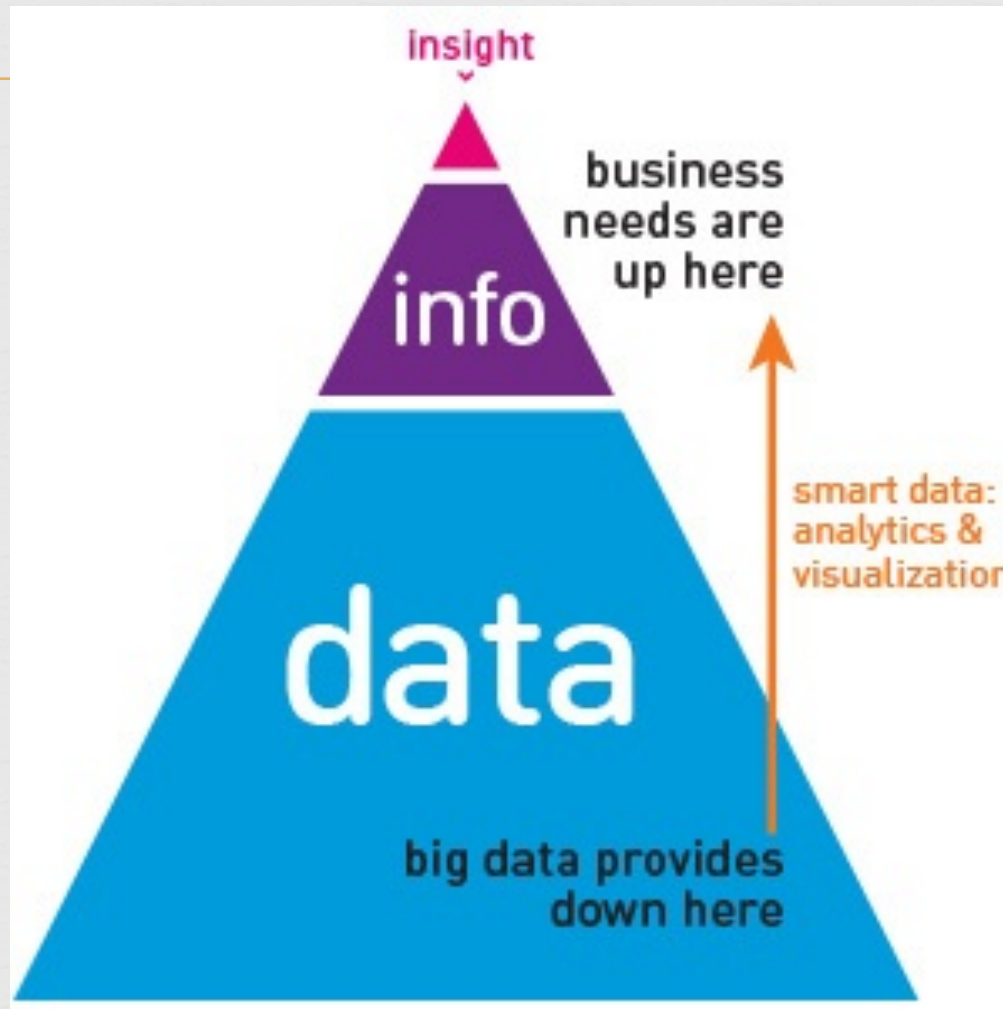
All of those photos put together would equal about one zettabyte" (*)

(*) Source: *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES* Executive Office of the President, MAY 2014 -The White House, Washington.

What to do with Data?
From Data to Insight

7

AI and Data

 "AI is akin to building a rocket ship. You need a huge engine and a lot of fuel. The rocket engine is the learning algorithms but *the fuel is the huge amounts of data we can feed to these algorithms."*

-- Andrew Ng

Examples of
BIG DATA USE CASES

ॐ

- ॐ Log Analytics
- ॐ Fraud Detection
- ॐ Social Media and Sentiment Analysis
- ॐ Risk Modeling and Management
- ॐ Energy sector
- ॐ Transportation
- ॐ Healthcare
- ॐ Security
- ॐ Public sector administration
- ॐ Global personal location data
- ॐ Retail
- ॐ Manufacturing

Layers of (Legal and Ethical) Regulations

ℭ European Data Law
   GDPR

   Data Governance Act

ℭ **(Proposed) AI Act**

ℭ more…

EU Proposed Legal Framework on AI

ଓଃ

ଓଃ " The Commission proposes new rules and actions aiming to turn Europe into the global hub for trustworthy Artificial Intelligence (AI).

Press release, 21 April 2021, Brussels

Source EU

Motivation

ୠ **Margrethe Vestager**, Executive Vice-President for a Europe fit for the Digital Age, said:

"On Artificial Intelligence, **trust** is a must, not a nice to have.

With these landmark rules, the EU is spearheading the development of new global norms *to make sure AI can be trusted.*

By setting the standards, we can pave the way to *ethical technology worldwide and ensure that the EU remains competitive along the way.* Future-proof and innovation-friendly, our rules will intervene where strictly needed: when the safety and fundamental rights of EU citizens are at stake."

EU Proposed Legal Framework on AI

❧ The combination of a **legal framework on AI** and a new Coordinated Plan with Member States will guarantee the safety and fundamental rights of people and businesses, while strengthening AI uptake, investment and innovation across the EU.

❧ New rules on Machinery will complement this approach by adapting safety rules to increase users' trust in the new, versatile generation of products."

Material scope of application, cf. Art. 3 No. 1

"Artificial Intelligence System" (AI System):

software developed using one or more of the techniques and concepts listed in Annex I and capable of producing results such as content, predictions, recommendations or decisions that influence the environment with which it interacts, in relation to a set of objectives defined by humans;

Annex I

— Techniques and concepts according to the Annex (power of amendment of the Commission!):

- **Machine learning** concepts, with supervised, unsupervised, and reinforcement learning using a wide range of methods, including deep learning;

- **Logic and knowledge-based concepts**, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deduction engines, (symbolic) reasoning and expert systems;

- **Statistical approaches**, Bayesian estimation, search and optimization methods.

The EU Follows a Risk-adapted Regulatory Approach

ଔ The German Data Ethics Commission recommended adopting a **risk-adapted regulatory approach** to algorithmic systems. The principle underlying this approach should be as follows: **the greater the potential for harm, the more stringent the requirements and the more far-reaching the intervention by means of regulatory instruments.**

ଔ https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html

Risk-adapted Regulatory Approach
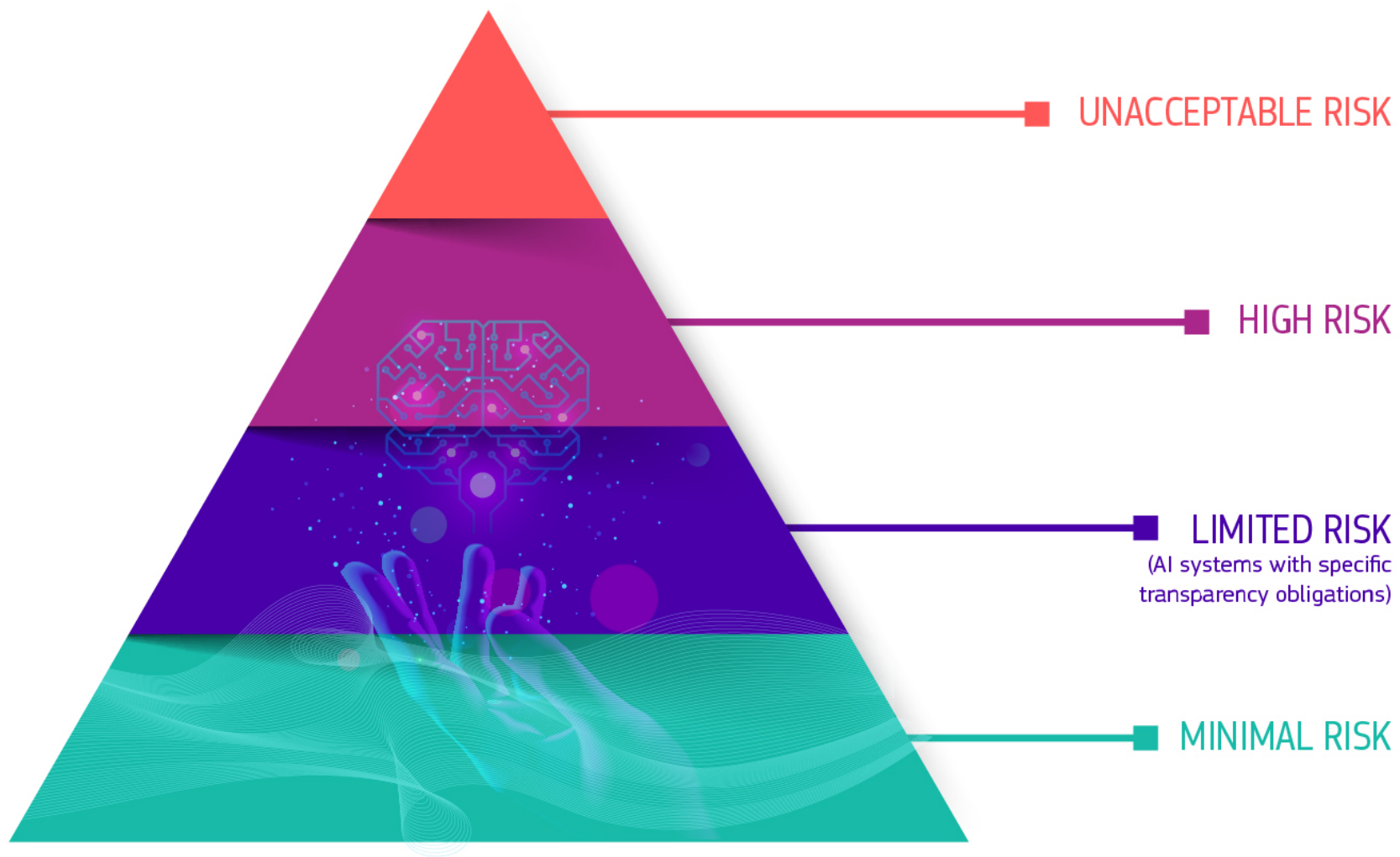
When assessing this potential for harm, the **sociotechnical system as a whole must be considered**, or in other words all the components of an algorithmic application, including all the people involved, from the development phase – for example the training data used – right through to its implementation in an application environment and any evaluation and adjustment measures.

Risk-adapted regulatory system

ल

ल **HIGH RISK** complete or partial ban of an algorithmic system

ल additional measures such as live interface for "always on" oversight by supervisory institutions

ल additional measures such as ex-ante approval procedures

ल measures such as formal and substantive requirements (e. g. transparency obligations, publication of a risk assessment) or monitoring procedures (e.g. disclosure obligations towards supervisory bodies, ex-post controls, audit procedures)

ल **LOW RISK** no special measures

UNACCEPTABLE RISK

HIGH RISK

LIMITED RISK
(AI systems with specific
transparency obligations)

MINIMAL RISK

EU: AI systems identified as **high-risk**

include AI technology used in:

- **Critical infrastructures** (e.g. transport), that could put the life and health of citizens at risk;
- **Educational or vocational training**, that may determine the access to education and professional course of someone's life (e.g. scoring of exams);
- **Safety components of products** (e.g. AI application in robot-assisted surgery);

EU: AI systems identified as **high-risk**

❧

- **Employment, workers management and access to self-employment** (e.g. CV-sorting software for recruitment procedures);

- **Essential private and public services** (e.g. credit scoring denying citizens opportunity to obtain a loan);

- **Law enforcement** that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);

EU: AI systems identified as **high-risk**

❧

❧ **Migration, asylum and border control management** (e.g. verification of authenticity of travel documents);

❧ **Administration of justice** and **democratic processes** (e.g. applying the law to a concrete set of facts).

EU: AI systems identified as **high-risks**

❧

ଔ High-risk AI systems will be subject to **strict obligations** before they can be put on the market

Remote biometric identification

 All remote biometric identification systems are considered high risk and subject to strict requirements.

High-risk AI systems **obligations**

ଜ **Adequate risk assessment and mitigation systems;**

ଜ **High quality of the datasets** feeding the system to minimise risks and discriminatory outcomes;

High-risk AI systems **obligations**

ം **Logging of activity to ensure traceability of results**;

ം **Detailed documentation** providing all information necessary on the system and its purpose for authorities to assess its compliance;

High-risk AI systems **obligations**

ೞ **Clear and adequate information** to the user;

ೞ **Appropriate human oversight** measures to minimise risk;

ೞ High level of **robustness**, **security** and **accuracy**.

Limited Risks

ର **Limited risk**, i.e. AI systems with specific transparency obligations:

When using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.

Minimal Risks

ଔ **Minimal risk:** The legal proposal allows the free use of applications such as AI-enabled video games or spam filters.

The vast majority of AI systems fall into this category. The draft Regulation does not intervene here, as these AI systems represent only minimal or no risk for citizens' rights or safety.

AI governance

In terms of governance, the Commission proposes that national competent market surveillance authorities supervise the new rules, while the creation of a **European Artificial Intelligence Board** will facilitate their implementation, as well as drive the development of standards for AI.

AI governance

Additionally, *voluntary codes of conduct* are proposed for *non-high-risk AI,* as well as *regulatory sandboxes* to facilitate responsible innovation.

Regulatory sandboxes

ა

- **Regulatory sandboxes** enable in a real-life environment the testing of innovative technologies, products, services or approaches, *which are not fully compliant with the existing legal and regulatory framework.*

- They are operated for *a limited time and in a limited part* of a sector or area.

- The purpose of regulatory sandboxes is *to learn about the opportunities and risk*s that a particular innovation carries and to develop the right regulatory environment to accommodate it.

- Experimentation clauses are often the legal basis for regulatory sandboxes.

© **2021** FEDERAL MINISTRY FOR ECONOMIC AFFAIRS AND ENERGY

Source: https://www.bmwi.de/Redaktion/EN/Dossier/regulatory-sandboxes.html

Conformity assessments

ℰ According to the EU proposal, for **healthcare products** (generally medical devices) **conformity assessment** would see the involvement of professional certification bodies already designated under the Regulations on medical devices.

ℰ They are called "**Notified Bodies**".

Conformity assessments

For **healthcare-related AI systems, which are not medical devices, but are in the "high-risk"** list of Annex III (e.g. AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid)**, for now the EU foresee a conformity assessment done directly by the manufacturer** (who certainly can foresee an audit mechanism, even external, but which remains entirely under **the manufacturer's responsibility**).

The General Data Protection Regulation

(GDPR)

ﾃ EU Law which addresses the subject of safeguarding **privacy of personal data** of the citizens of the EU and EEA and

ﾃ How data collected data might be transported out of the EU/EEA.

GDPR: Data Privacy

ം **Personal Data**

 Any information relating to an person who can be *identified, directly or **indirectly**,* in particular by *reference to an identifier* such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

GDPR: Data Privacy

Consent

Any freely given, specific, informed and unambiguous indication of his or her wishes by which the data subject, either by a statement or by a clear affirmative action, signifies *agreement to personal data relating to them being processed.*

GDPR: Data Privacy

☧

Profiling

Any *automated processing of personal data to determine certain criteria about a person*, in particular to **analyse or predict** aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

GDPR: Collecting Data

☙

The information that must be provided to the subject when the data is being collected, is defined in the GDPR:

- the identity and the contact details of the controller and DPO.

- the *purposes* of the processing for which the personal data are intended the legal basis of the processing.

GDPR: Collecting Data

- where applicable the legitimate interests pursued by the controller or by a *third party.*

- where applicable, the recipients or categories of recipients of the personal data.

- where applicable, that *the controller intends to transfer personal data internationally*

-the period for which the personal data will be stored, or if this is not possible, *the criteria used to determine this period.*

Source: https://arxiv.org/pdf/1806.03253.pdf

GDPR: Collecting Data

☞

- the existence of *the right to* access, rectify or *erase* the personal data.

- *the right to data portability.*

- *the right to withdraw consent at any time.*

- *the right to lodge a compla*int to a supervisory authority.

- Source: https://arxiv.org/pdf/1806.03253.pdf