# Building an AI-based Business: The Copyright Issue

**Julien Cabay**

22 August 2019

# Who am I?

**Postdoc Researcher FNRS**

**Associate Professor ULB**

**Associate Professor ULiège**

Contact : jcabay@ulb.ac.be

# Course Agenda

- Introduction
- Part 1: Artificial Intelligence
- Part 2: Intellectual Property
- Part 3: Data Protection
- Part 4: The Copyright Issue
- Part 5: Text and Data Mining Allowed
- Conclusion

# Part 1: Introduction

# Introduction

« (…) **Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry**. Growth in computing power, availability of data and progress in algorithms have turned AI into one of the **most strategic technologies of the 21st century**. The stakes could not be higher. **The way we approach AI will define the world we live in** (…) »

# Introduction

- **Overall private investments in AI in 2016**
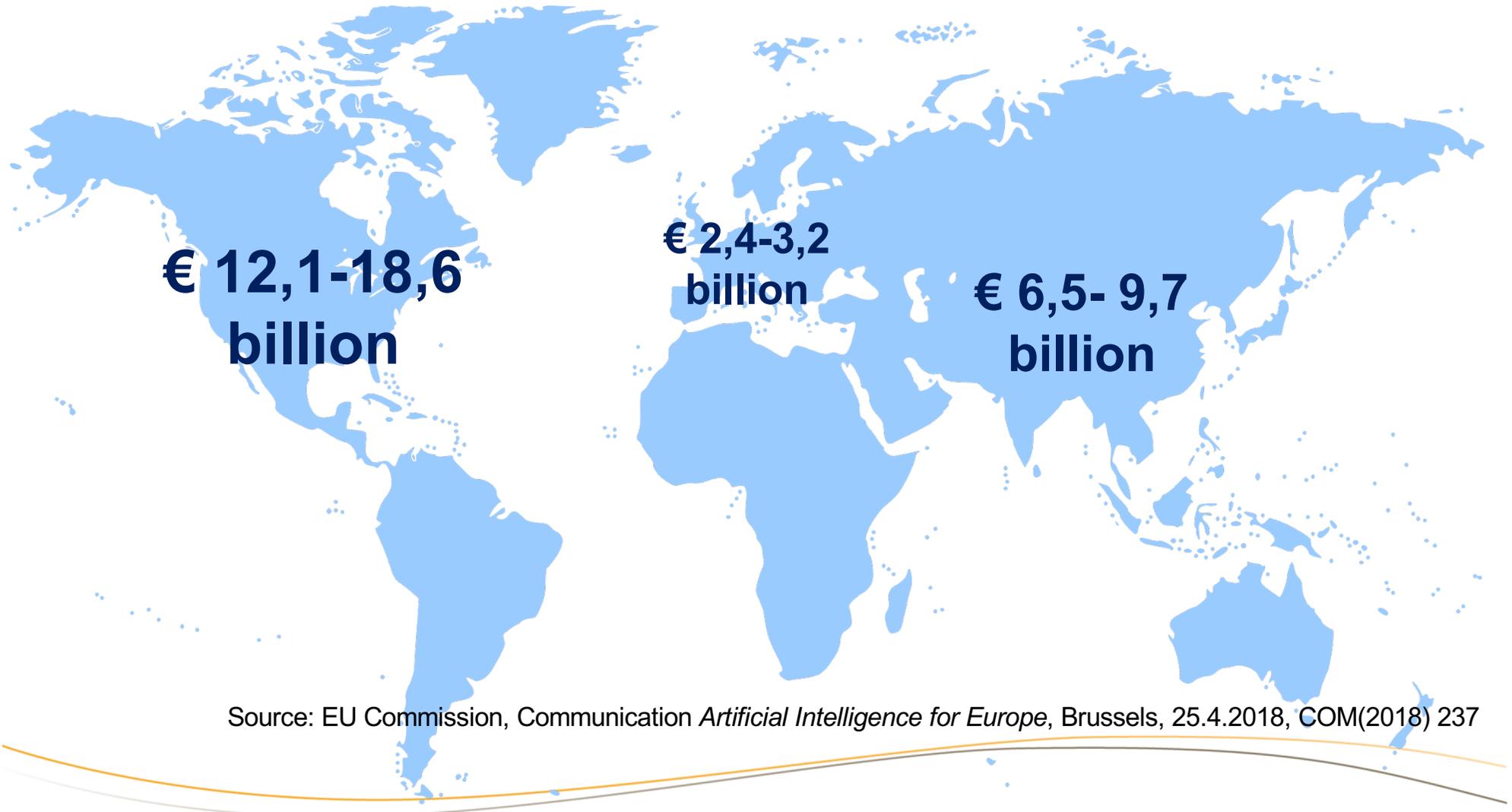
€ 12,1-18,6 billion

€ 2,4-3,2 billion

€ 6,5- 9,7 billion

Source: EU Commission, Communication *Artificial Intelligence for Europe*, Brussels, 25.4.2018, COM(2018) 237

# Introduction

« (...) Public and private **research and development investments in AI** in the EU last year (2017) were estimated to total EUR 4-5 billion. **The EU as a whole (public and private sectors combined)** should aim to increase this investment to **at least EUR 20 billion by the end of 2020**. It should then **aim for more than EUR 20 billion per year over the following decade** (...) »

Source: EU Commission, Communication *Artificial Intelligence for Europe*, Brussels, 25.4.2018, COM(2018) 237

# Introduction

« *(…)* ***notre niveau d'ambition minimal devrait être de EUR 80 millions par an. Cela correspond à au moins 1 milliard à l'horizon 2030.*** *Cependant, si on regarde l'USA et la Chine, qui ont compris l'ambition requise,* ***nous devons être encore plus audacieux.***

*(…)* »

(Belgium)

Source: AI4Belgium Report, 28 March 2019, available on https://www.ai4belgium.be/wp-content/uploads/2019/04/rapport_fr.pdf

# Introduction

« *(...)* ***Reflection will be needed on interactions between AI and intellectual property rights****, from the perspective of both intellectual property offices and users, with a view to fostering innovation and legal certainty in a balanced way (...)* »

Source: EU Commission, Communication *Artificial Intelligence for Europe*, Brussels, 25.4.2018, COM(2018) 237
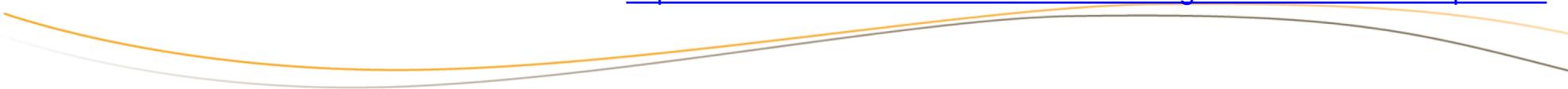
# Part 1: Artificial Intelligence
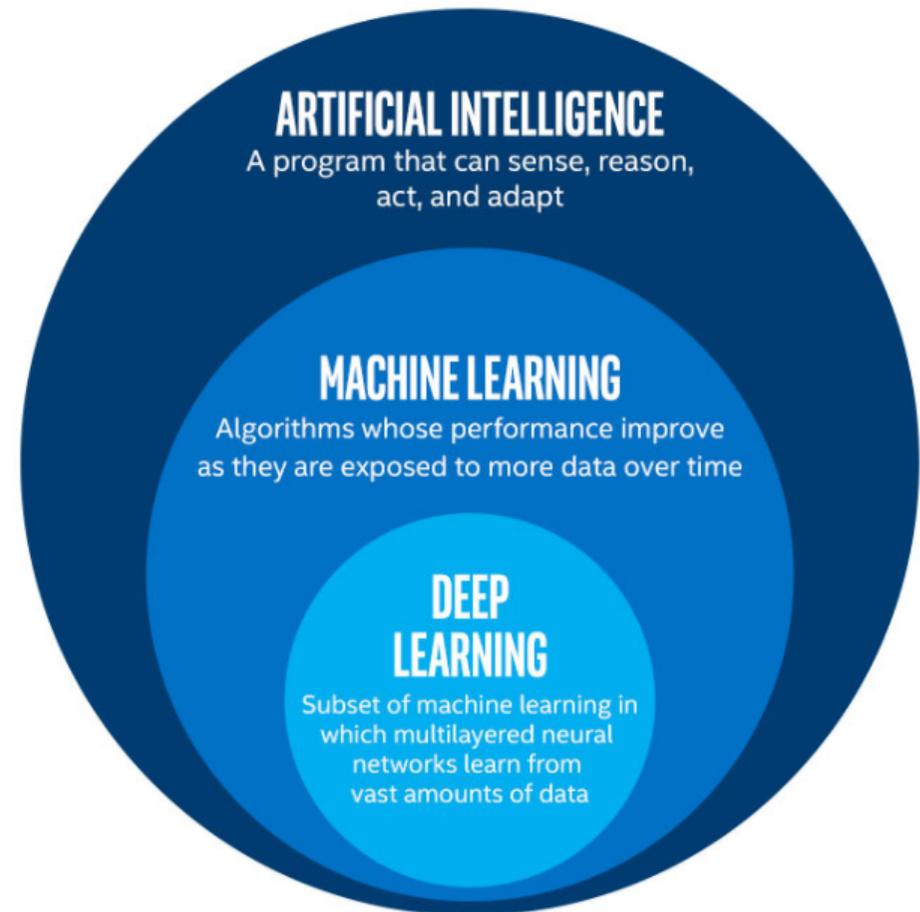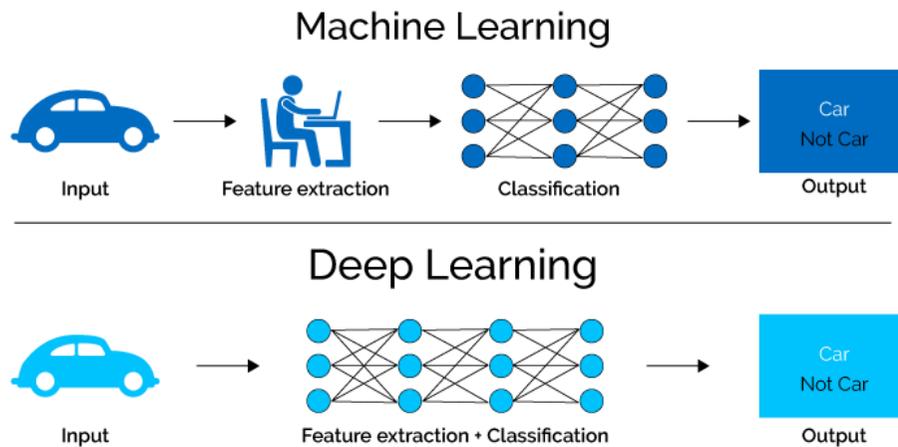
# Artificial Intelligence

- **AI = any program that can**
  - **Sense** = Identify and recognize meaningful objects or concepts in the midst of vast data. Is that a stoplight? Is it a tumor or normal tissue?
  - **Reason** = Understand the larger context, and make a plan to achieve a goal. If the goal is to avoid a collision, the car must calculate the likelihood of a crash based on vehicle behaviors, proximity, speed, and road conditions.
  - **Act** = Either recommend or directly initiate the best course of action. Based on vehicle and traffic analysis, it may brake, accelerate, or prepare safety mechanisms.
  - **Adapt** = We must be able to adapt algorithms at each phase based on experience, retraining them to be ever more intelligent. Autonomous vehicle algorithms should be re-trained to recognize more blind spots, factor new variables into the context, and adjust actions based on previous incidents.

Source: Singh Niven, *How to get started as a Developer in AI*,
https://software.intel.com/en-us/articles/how-to-get-started-as-a-developer-in-ai

# Artificial Intelligence

- **AI as an umbrella term** that includes machine learning and deep learning (as a subset of the latter)



Machine Learning

Input → Feature extraction → Classification → Output (Car / Not Car)

Deep Learning

Input → Feature extraction + Classification → Output (Car / Not Car)



ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt

MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data
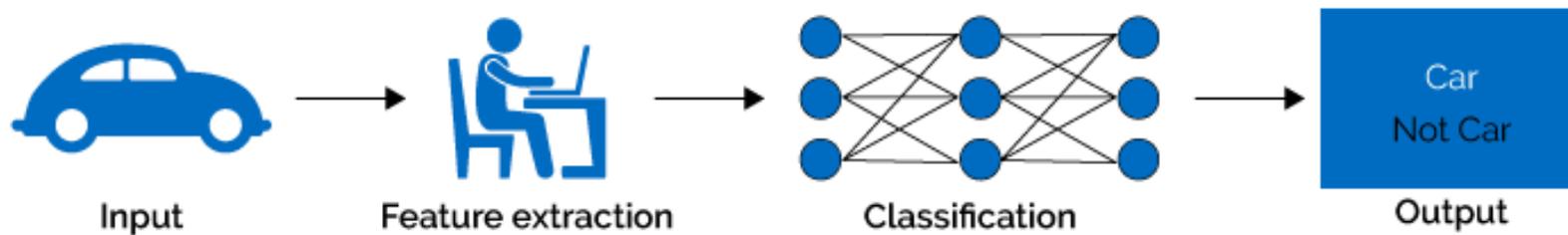
# Artificial Intelligence
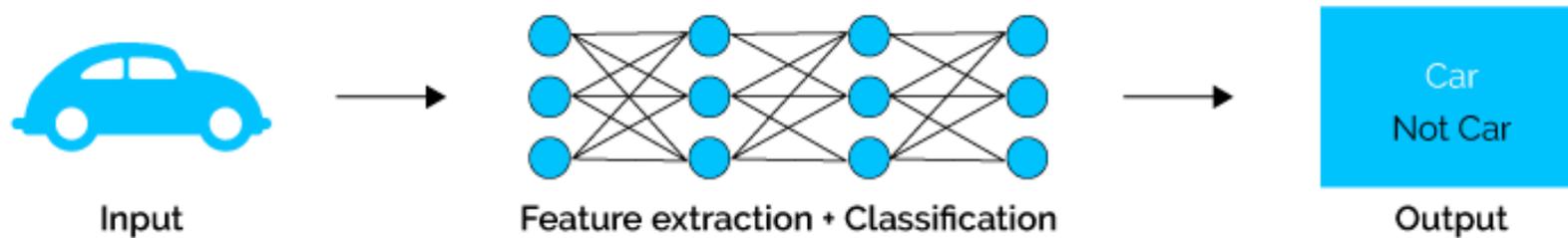
# Artificial Intelligence



Deep Learning Opportunities

AGRICULTURE
TARANIS
BENSON·HILL BIOSYSTEMS

AUTO
drive.ai · nuro · nexar · Mapillary · PERCEPTIVE AUTOMATA · pony.ai
DEEPMAP · MOMENTA · Iris Automation · AEYE · DEEPSCALE

HEALTHCARE
Butterfly Network · PAIGE · IDx
Atomwise · insitro · viz.ai
Niramai · LeanTaaS · ARTERYS
mindstrong · Qventus · gauss
OWKIN · medopad

GOVERNMENT
Face++ 旷视 · one concern · SHIELD AI
FORTEM TECHNOLOGIES · 依图 YITU · 商汤 sensetime

2019

AI100
CB INSIGHTS

SEMICONDUCTOR
cerebras · Horizon Robotics
GRAPHCORE · MYTHIC
habana · thinci
SYNTIANT

TELECOM
Mist

FINANCE & INSURANCE
4Paradigm 第四范式 · DATAVISOR · BEHAVOX
AppZen · BIOCATCH Less Friction · Less Frdud · (h[s]) HYPERSCIENCE

INDUSTRIALS
sense · KEBOTIX · zymergen
falkonry · C3.ai · LANDING AI

RETAIL
twentybn · SIGNIFYD
ABEJA · AiFi · sift

MEDIA
AI FOUNDATION · ARRAIY
New Knowledge

REAL ESTATE
HOVER
SKYLINE

LEGAL, COMPLIANCE, & HR
EigenTechnologies · onfido
LawGeex · textio

ENTERPRISE TECH

TRAINING DATA
AI.REVERIE · defined crowd
Mighty Ai

SOFTWARE DEVELOPMENT
applitools · mabl

DATA MANAGEMENT
dataiku · dremio
MACHINIFY · H2O.ai
TRIFACTA · tamr
DataRobot · SIGOPT

CYBERSECURITY
AGARI · anodot
AREA 1 · perimeterx
JASK · DEMISTO
VECTRA · SHAPE SECURITY

ADS, SALES, & MARKETING
GONG · Unbabel
gamalon · fullstory
bouncex

OTHER APPLICATIONS
AUTOMATION ANYWHERE · UiPath
ELEMENT AI · Descartes Labs
sparkcognition · Orbital Insight
PROWLER.io the decision company

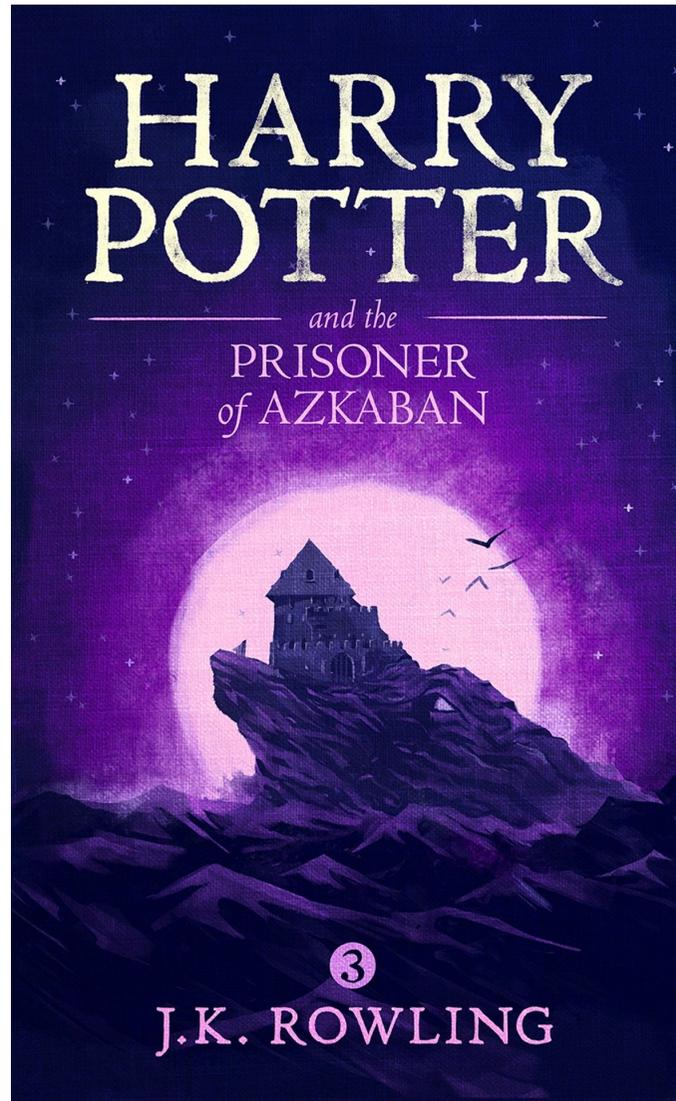Source: https://s3.amazonaws.com/cbi-research-portal-uploads/2019/02/01125325/AI-100-2019-2.png

# Part 2: Intellectual Property
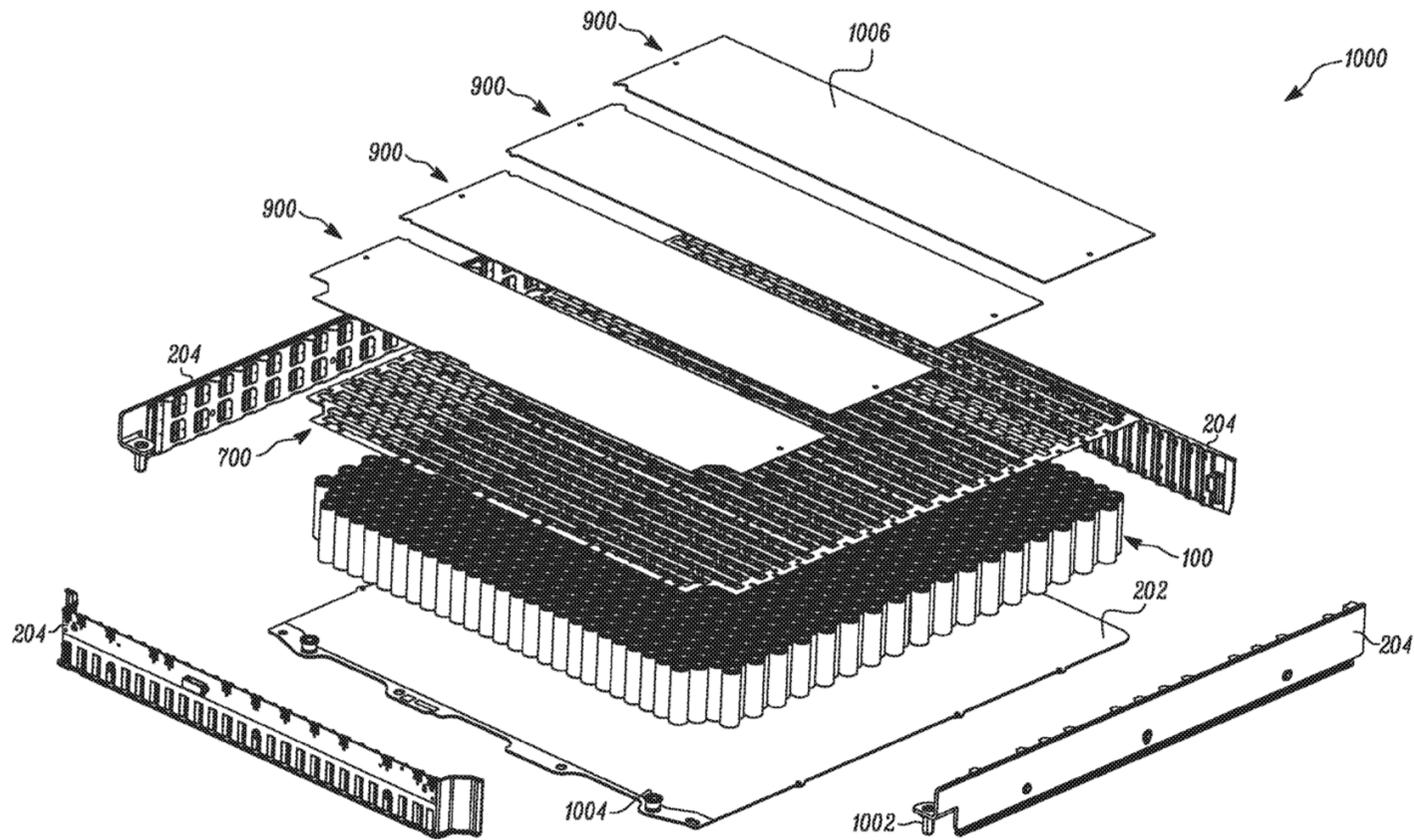
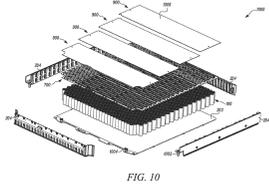# Intellectual Property

FIG. 10

# Intellectual Property

# Intellectual Property

# Intellectual Property

# Intellectual Property

| | Subject matter | Protection requirements | Registration | Duration | Scope of protection | Example |
|---|---|---|---|---|---|---|
| **Copyright** | Work | Originality, Expression | No | Life of the author + 70 years | Exploitation when copying |  |
| **Patent** | Invention | Novelty, Inventive Step, Industrial Applicability | Yes | 20 years | Exploitation even without copying |  |
| **Design** | Appearance of a product | Novelty, Individual chararacter | Yes, except unregistered Community design | 5 years renewable, max 25 years (reg.); 3 years (unreg.) | Exploitation when copying (unreg.) and even without copying (reg.) |  |
| **Trademark** | Sign guaranteeing consumers the origin of good/service | Distinctiveness, Representation | Yes, except well known mark | 10 years renewable indefinitely | Exploitation even without copying when adverse effect on function |  |

# Intellectual Property

- **Copyright and related rights** (general legal framework in the EU)
  - Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society [InfoSoc Dir.]
    - Copyright (Authors)
    - Neighbouring rights (Performers, Phonogram/Film Producers, Broadcasting Organisations)
  - Directive 2009/24/EC on the legal protection of computer programs [Software Dir.]
    - Copyright (Authors)
  - Directive 96/9/EC on the legal protection of databases [Database Dir.]
    - Copyright (Authors)
    - Database *sui generis* right (Maker)
  - Directive (EU) 2019/790 on copyright and related rights in the Digital [DSM Dir.]
    - Single Market
    - Rights on press publications (Publishers)

# Intellectual Property

- **Copyright** (basic features)
  - Territoriality: EU harmonization but national protection (Belgium, *see* art. XI.165 et seq. Code de droit économique)
  - Subject matter: Works, *i.e.* literar, artistic, musical, software (code), database (structure)
  - Requirements for protection:
    - Expression (not ideas)
    - Originality, *i.e.* author's own intellectual creation, personal touch, free and creative choices
  - Ownership: Author (principle, but exceptions; contracts)
  - Rights:
    - Moral: Divulgation, Attribution, Integrity
    - Economic: Reproduction, Communication to the public, Distribution, Rental, Lending
  - Exceptions : exhaustive list (ex: private copying, quotation, parody, etc.)
  - Duration: life of the author + 70 years

# Intellectual Property

- **Database *sui generis* right** (basic features)
    - Territoriality: National (Belgium, *see* art. XI.305 et seq. Code de droit économique)
    - Subject matter: Database, *i.e. « collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means »*
    - Requirements for protection: substantial investment (quantitative/qualitative)
    - Ownership: Maker (investor)
    - Rights (economic) : Extraction, Re-Utilization
    - Exceptions: exhaustive list (ex: private purposes, research, etc.)
    - Duration: 15 years after completion

# Part 3: Data Protection
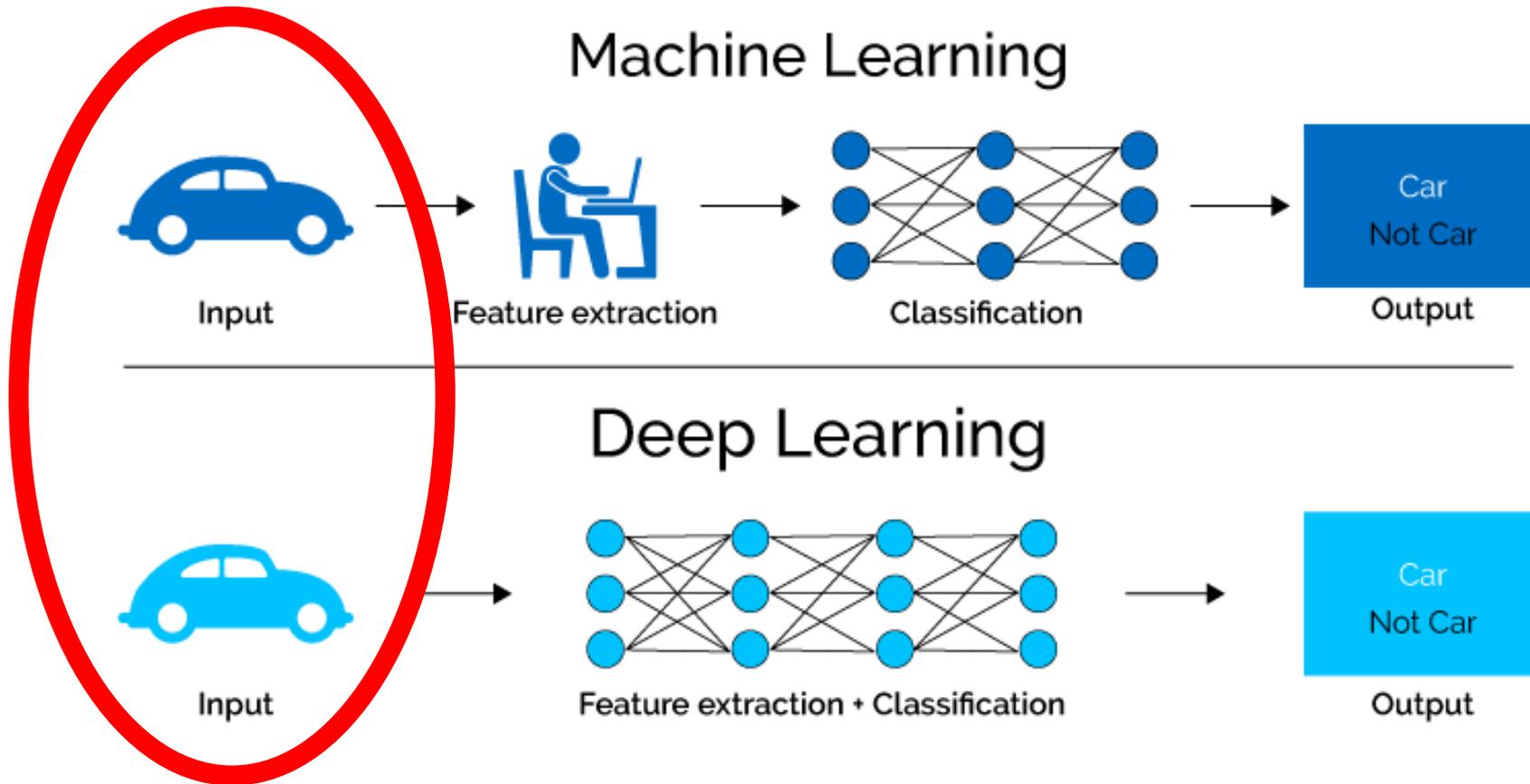
# Data Protection



« (...) **AI needs vast amounts of data to be developed**. Machine learning, a type of AI, works by identifying patterns in available data and then applying the knowledge to new data. The larger a data set, the better even subtle relations in the data can be discovered. When it comes to using AI, data-rich environments also provide for more opportunities. This is because data is the way the algorithm learns about and interacts with its environment (...) »

Source: EU Commission, Communication *Artificial Intelligence for Europe*, Brussels, 25.4.2018, COM(2018) 237

Credit: *The Economist*, May 6th, 2017

# Data Protection

## Machine Learning



Input → Feature extraction → Classification → Output (Car / Not Car)

## Deep Learning



Input → Feature extraction + Classification → Output (Car / Not Car)

- **Legal issues? Depending on the input/data**

# Data Protection

- **Protection available for data** (general legal framework in the EU)
  - Regulation (EU) 2016/679 on the protection of personal data and on the free movement of such data [GDPR]
    - Personal Data
  - Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their lawful acquisition, use and disclosure [Trade secrets Dir.]
    - Trade Secrets
  - Database *sui generis* right (Database Dir.)
    - Substantial part of the content of a database
    - If repeated and systematic, also unsubstantial parts
  - Copyright/Neighbouring rights (InfoSoc Dir.; Software Dir.; Database Dir.; DSM Dir.)
    - Works
    - Other subject-matters
  - Contract law (CJEU, *Ryanair*, C-30/14 (2015))
    - Data (restricted access/use, through terms and conditions)

# Data Protection

- **Copyright protection for data?**
  - Art. 2(1) Berne Convention for the protection of literary and artistic works: « The expression "literary and artistic works" shall include every production in the literary, scientific and artistic domain, **whatever may be the mode or form of its expression** (…) »
    - But Copyright do not protect ideas or facts
    - But Copyright do not protect non original expressions

  - ⇒ Where data is the expression of a work, copyright protection is available provided it is **original**
    - = 'Author's own intellectual creation' (CJEU, *Infopaq*, C-5/08 (2009))
    - Criteria:
      - 'Through the choice, sequence and combination' (*idem*)
      - 'Personal touch' (CJEU, *Painer*, C-145/10 (2011))
      - 'Free and creative choices', not fulfiled when 'dictated by technical considerations, rules or constraints which leave no room for creative freedom' (CJEU, *Football Dataco*, C-604/10 (2012))
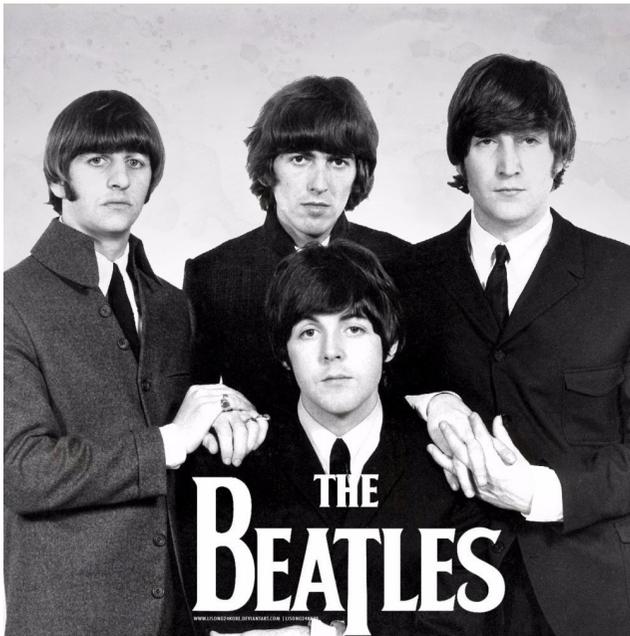
# Data Protection

- **No © data**
  - Topical example : raw machine-generated data and applications
    - *See* EU Commission, Communication *Building A European Data Economy*, Brussels, 10.1.2017, COM(2017) 9
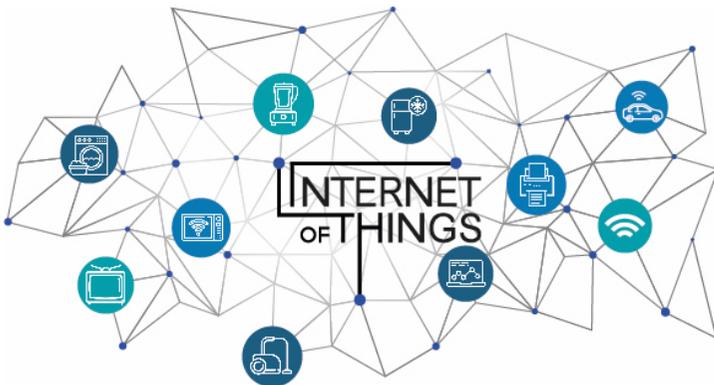
# Data Protection

- **© data**
  - Topical example : (recent) musical works and AI generated music
    - © protection = life of the author + 70 years



Source: Sony CSL Research Lab, *Daddy's Car*, https://www.youtube.com/watch?v=LSHZ_b05W7o

# Data Protection

- **© data?**
  - Topical example : photographs and facial recognition
    - © and portrait photograph? *See* CJEU, *Painer*, C-145/10 (2011)

# Part 4: The Copyright Issue

# The Copyright Issue

# © DATA/INPUT ⟹



- **The copyright issue:**
  - (1) Are data reproduced in the © sense?
  - (2) If yes, is there any authorization?
  - (3) If not, is there any justification?

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - Art. 2(a) InfoSoc Dir.: « *(…) exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part:  for authors, of their works (…)* »
    - Broad interpretation
    - *See* CJEU, *Infopaq*, C-5/08 (2009)

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - CJEU, *Infopaq*, C-5/08 (2009)
    - Facts
      - Infopaq runs a media monitoring and analysis business
      - Principal activity is drawing up summaries of selected articles from Danish daily newspapers and other periodicals, selected on the basis of certain subject criteria agreed with customers
      - Selection is made by means of a 'data capture process'
      - DDF (association of publishers) sues for copyright infringement

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - CJEU, *Infopaq*, C-5/08 (2009)
    - Data capture process
      - 1) Selection of relevant publications
      - 2) Scaning and creation of TIFF file (image file)
      - 3) Translation of TIFF file by OCR server into data readable by text processing program (text file)
        - At the end of this stage, deletion of image file
      - 4) Text file processed to find search word and generation of result data for each match, the result data comprising (among others) 5 prior words / search word / 5 following words
        - At the end of this stage, deletion of text file
      - 5) Printing out result data

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - CJEU, *Infopaq*, C-5/08 (2009)
    - Decision

      « ***An act occurring during a data capture process, which consists of storing an extract of a protected work comprising 11 words*** *and printing out that extract, is such as to **come within the concept of reproduction in part** within the meaning of Article 2 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, **if the elements thus reproduced are the expression of the intellectual creation of their author**; it is for the national court to make this determination* »

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - CJEU, *Football Association Premier League*, C-403/08 and C-429/08 (2011)
    - Decision

      « *Article 2(a) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society must be interpreted as meaning that the* **reproduction right extends to transient fragments of the works within the memory of a satellite decoder and on a television screen**, *provided that those fragments contain elements which are the* **expression of the authors' own intellectual creation**, *and the unit composed of the fragments reproduced simultaneously must be examined in order to determine whether it contains such elements* »

# The Copyright Issue

- **(1) Are data reproduced in the © sense?**
  - Conclusion: reproduction in the © sense likely in AI business
    - At the input stage:
      - Step 1: Data access/gathering/selection/storage
      - Step 2: Data labelling/cleaning
      - Step 3: Data analysis, machine learning (training)

# The Copyright Issue

- **(2) If yes, is there any authorization?**
  - Prior consent as a requisite => OPT-IN and not opt-out
    - *See* CJEU, *Soulier & Doke*, C-301/15 (2016)

      « *(…) the rights guaranteed to authors by Article 2(a) and Article 3(1) of Directive 2001/29 are preventive in nature, in the sense that **any reproduction** or communication to the public of a work by a third party **requires the prior consent of its author** (…)*

  - … even where the author cannot be identified and located
    - Exception : use of Orphan Works by some cultural institutions
      - *See* Directive 2012/28/EU on certain permitted uses of orphan works

# The Copyright Issue

- **(2) If yes, is there any authorization?**
  - Consent may be implicit, but not hypothetical
    - In particular, works made freely available on the Internet by its author are not free of © and cannot be reproduced
      - *See* CJEU, *Soulier & Doke*, C-301/15 (2016)
        « *(…) the circumstances in which **implicit** consent can be admitted must be **strictly defined** in order not to deprive of effect the very principle of the author's prior consent. In particular, every author **must actually be informed of the future use** of his work by a third party and the means at his disposal to prohibit it if he so wishes.*
        ***Failing any actual prior information** relating to that future use, the author is unable to adopt a position on it and, therefore, to prohibit it, if necessary, so that the very existence of his **implicit consent appears purely hypothetical** in that regard.* »

      - *See* CJEU, *Renckhoff*, C-161-17 (2018)
        « *The concept of '**communication to the public'**, within the meaning of Article 3(1) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, must be interpreted as meaning that it **covers the posting on one website of a photograph previously posted, without any restriction preventing it from being downloaded and with the consent of the copyright holder, on another website** »*

# The Copyright Issue

- **(2) If yes, is there any authorization?**
  - It depends on the source:
    - Own generated data: © ownership
      - Ex: Sony (The Beatles' music)
    - Data transfered by consumers : © licenses (through terms and conditions)
      - Ex: Facebook (material posted by users)
    - Data from 1rd party : © licenses
      - Ex: Spotify (music licensed from collecting societies/artists)
    - Data from Internet: 'open source' licenses
      - Ex: IBM (photographs dataset from Flickr under CC licenses)

# The Copyright Issue

- **(3) If not, is there any justification?**
  - Art. 5 InfoSoc Dir. : exhaustive list of exceptions
    - Optional, except for transient copy (art. 5(1))
    - Strict interpretation, effectiveness should be safeguarded and purpose obeserved (≠ restrictive)
      - *See* recently CJEU, *Funke Medien*, C-469/17; *Spiegel Online*, C-516/17 (2019))
    - '3 step-test' (art. 5(5))
      - Certain special cases
      - Do not conflict with a normal exploitation of the work
      - Do not unreasonably prejudice the legitimate interests of the rightholder

# The Copyright Issue

- **(3) If not, is there any justification?**
  - **Private use exception** for AI business?
    - Art. 5(2)(b) InfoSoc Dir.: « *(…) reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial (…)* »
  - ⇒ **NO**

# The Copyright Issue

- **(3) If not, is there any justification?**
  - **Research exception** for AI business?
    - Art. 5(3)(a) InfoSoc Dir.: « *(...) use for the **sole purpose** of illustration for teaching or **scientific research**, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the* **non-commercial purpose** *to be achieved (...)* »
  - $\Rightarrow$ **NO**

# The Copyright Issue

- **(3) If not, is there any justification?**
  - **Exception for transient copies** for AI business?
    - Art. 5(1) InfoSoc. Dir.: « *Temporary acts of reproduction referred to in Article 2, which are **transient or incidental** [and] an **integral and essential part of a technological process** and whose sole purpose is to enable:*
      *(a) **a transmission in a network between third parties by an intermediary**, or*
      *(b) **a lawful use***
      *of a work or other subject-matter to be made, and which have **no independent economic significance**, shall be exempted from the reproduction right provided for in Article 2* »
      - = 5 cumulative conditions (CJUE, *Infopaq*, C-5/08 (2009))
      - *See* CJUE, *Football Association Premier League*, C-403/08 and C-429/08 (2011))

        « *In accordance with its objective, **that exception must allow and ensure the development and operation of new technologies** and safeguard a fair balance between the rights and interests of right holders, on the one hand, and of users of protected works who wish to avail themselves of those new technologies, on the other* »

# The Copyright Issue

- **(3) If not, is there any justification?**
  - **Exception for transient copies** for AI business?
    - CJUE, *Infopaq (II)*, C-302/10 (2012)

    « (…) those acts must not have independent economic significance, in that the **economic advantage** derived from their implementation **must not be either distinct or separable from the economic advantage derived from the lawful use of the work concerned** and it **must not generate an additional economic advantage going beyond that derived from that use of the protected work**. (…)
    an advantage derived from an act of temporary reproduction is distinct and separable if the author of that act is likely to make a **profit due to the economic exploitation of the temporary reproductions themselves**.
    The same is true if the acts of temporary reproduction lead to a **change in the subject matter reproduced**, as it exists when the technological process concerned is initiated, because those acts no longer aim to facilitate its use, but the use of a different subject matter (…) »

  ⇒  **PROBABLY NO**

# Can AI business survive EU © Law?

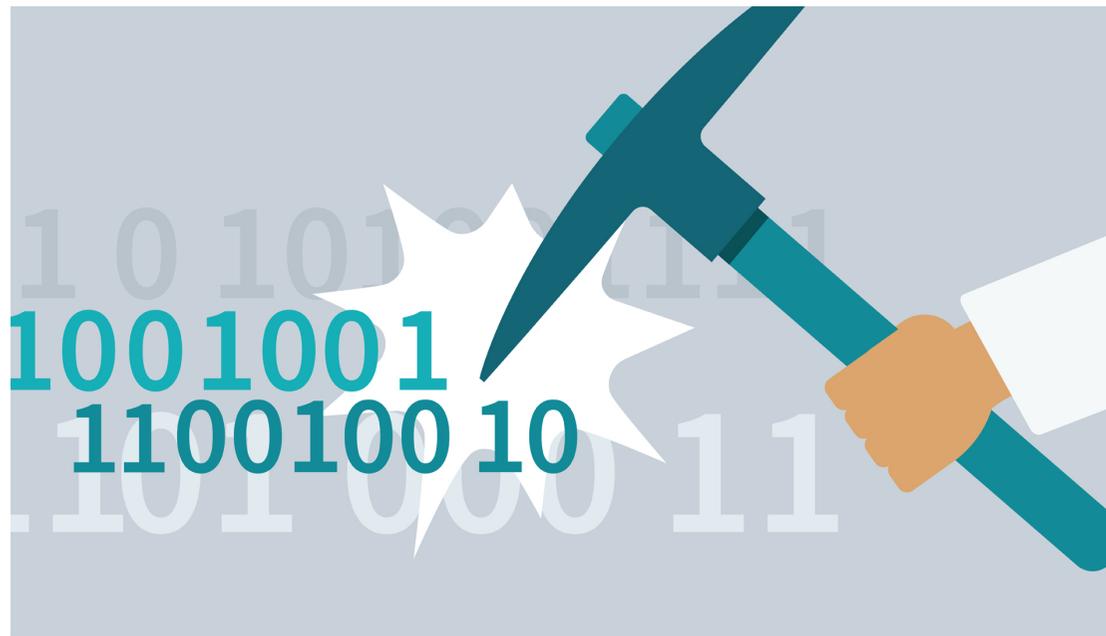# Part 5: Text and Data Mining Allowed

# Text and Data Mining Allowed

« *New technologies enable the automated computational analysis of information in digital form*, such as text, sounds, images or data, generally known as text and data mining. *Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible*. (...) text and data mining can , in particular, benefit the research community and, in so doing, *support innovation*. (...) in the Union, such organisations and institutions are confronted with *legal uncertainty* as to the extent to which they can perform text and data mining of content. In certain instances, *text and data mining can involve acts protected by copyright*, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are normalised in the process of text and data mining. *Where no exception or limitation applies, an authorisation* to undertake such acts *is required* from rightholders »

(Recital 8, DSM Dir.)

# Text and Data Mining Allowed

- **Text and Data Mining (TDM)?**

  – *See* art. 2(2) DSM Dir. : « '**text and data mining**' means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations »

# Text and Data Mining Allowed

- **TDM exceptions for non-commercial research**
  - National EU Member States
    - UK (2014): s. 29A Copyright, Designs and Patents Act [CDPA]
    - France (2016): art. L 122-5 (10) Code de la propriété intellectuelle [CPI]
      - (Implementing Decree not adopted)
    - Germany (2017): § 60d Urheberrechtsgesetz
  - EU (2019)
    - Art. 3 DSM Dir.
      - Mandatory exception
      - To be implemented in national law by June 7, 2021

# Text and Data Mining Allowed

- **A quick look at art. 3 DSM Dir.**

« *1. Member States shall provide for an* **exception** *to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for* **reproductions** *and extractions made* **by research organisations and cultural heritage institutions** *in order to carry out, for the* **purposes of scientific research, text and data mining** *of* **works** *or other subject matter to which they have* **lawful access***.*

*2. Copies of works or other subject matter made in compliance with paragraph 1 shall be* **stored** *with an* **appropriate level of security** *and may be* **retained for the purposes of scientific research***, including for the verification of research results.*

*3. (...) 4. (...)* »

# Text and Data Mining Allowed

- **A quick look at art. 3 DSM Dir.**
  - 'Research organisation', 'Cultural heritage instititution' : not-for-profit or public-interest mission (*see* def. art. 2(1) and (3); recital 12)
    - But TDM exception not excluded when « *(...) research activities are carried out in the framework of public-private partnerships. While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools  (...)* » (recital 11)

# Text and Data Mining Allowed

- **TDM exceptions not limited to non-commercial research**
  - Non EU Member States
    - Japan (2009; modified in 2018): art. 30-4 Copyright Act
    - USA : 17 U.S. Code §107 ('*Fair Use*'), as interpreted by case law
      - *See* in particular *Authors Guild v. Google Inc.*, 804 F.3d 202 (2d Cir. 2015)
  - EU (2019)
    - Art. 4 DSM Dir.
      - Mandatory exception
      - To be implemented in national law by June 7, 2021

# Text and Data Mining Allowed

- **A closer look at art. 4 DSM Dir.**

« *1. Member States shall provide for an **exception** or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for **reproductions** and extractions **of lawfully accessible works** and other subject matter for the **purposes of text and data mining**.*

*2. **Reproductions** and extractions made pursuant to paragraph 1 may be **retained for as long as is necessary for the purposes of text and data mining**.*

*3. The exception or limitation provided for in paragraph 1 shall apply on condition that the **use of works** and other subject matter referred to in that paragraph has **not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online**.*

*4. This Article shall not affect the application of Article 3 of this Directive.* »

# Text and Data Mining Allowed

- **A closer look at art. 4 DSM Dir.**
  - Not limited to specific beneficiaries
    - Comp. art. 3 DSM Dir. which is limited to research organisations and cultural heritage institutions (non profit)
  - Reproduction can be retained for as long as necessary
    - Comp. art. 5(1) InfoSoc Dir. which is limited to transient copies
  - Not limited to non-commercial purposes
    - But does not mean all commercial purposes are valid
      - *See* art. 5(5) InfoSoc Dir. (with reference in art. 7(2) DSM Dir.) ('3 step-test')
        - Do not conflict with a normal exploitation of the work
        - Do not unreasonably prejudice the legitimate interests of the rightholder
  - Subject to lawfull access to the works
    - But can be limited by contract
      - *See* art. 7(1) DSM Dir.: « *Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable* » (no reference to art. 4)
      - *See* recital 18: « (…) *Rightholders should remain able to license the uses of their works or other subject matter falling outside the scope of the mandatory exception provided for in this Directive for text and data mining for the purposes of scientific research and of the existing exceptions and limitations provided for in Directive 2001/29/EC* (…) »

# Text and Data Mining Allowed

- **A closer look at art. 4 DSM Dir.**
  - Works made publicly available online can be subject to TDM, unless it has been expressly reserved by rightholder in appropriate manner (such as machine-readable means)
    - Machine-redable means include 'metadata' and 'terms and conditions of a website or service' (recital 18)
    - Many questions:
      - Validity as such?
        - *See* CJEU, *Soulier & Doke*, C-301/15 (2016) (prior consent)
      - Application to works made available prior adoption/entry into force of the directive?
        - *See* CJEU, *Soulier & Doke*, C-301/15 (2016) (no prior information on future use = hypothetical consent)
      - Applications to works wich country of origin is not EU Member States?
        - *See* art. 5(2) Berne Convention: « *The enjoyment and the exercise of these rights shall not be subject to any formality (…)* »
      - Exclusion in practice?
        - 'TDM.txt' akin to 'robots.txt'

    ⇒ **NO LEGAL CERTAINTY**

# Text and Data Mining Allowed

- ## A quick look at the USA exception

*« Notwithstanding the provisions of sections 106 and 106A, the **fair use** of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiplecopies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—*

*(1) the **purpose and character of the use**, including whether such use is of a commercial nature or is for nonprofit educational purposes;*

*(2) the **nature of the copyrighted work**;*

*(3) the **amount and substantiality of the portion used** in relation to the copyrighted work as a whole; and*

*(4) the **effect of the use upon the potential market** for or value of the copyrighted work.*

*The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors »*

(17 U.S.C. §107)

# Text and Data Mining Allowed

- **A quick look at the USA exception**
  - *See* in general *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994)
    - *« (…) case-by-case analysis (…) »*
    - 4 factors are not to *« (…) be treated in isolation, one from another. All are to be explored, and the results weighed together, in light of the purposes of copyright (…) »*
    - *« (…) the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use (…) »*

# Text and Data Mining Allowed

- **A quick look at the USA exception**
  - *Authors Guild v. Google Inc.*, 804 F.3d 202 (2d Cir. 2015)
    - Facts: Google, scanning of more than 20 million books, search engine allowing new forms of research (including text and data mining), scanned image of books serving for improvement of search engine
    - Decision:

    *« (…) Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses. The purpose of the copying is highly **transformative**, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. **Google's commercial nature and profit motivation do not justify denial of fair use** (…) »*

$\Rightarrow$ *« **Uses involving robotic readers are fast-tracked for fair use** »*
(Grimmelmann, 2016)

# Text and Data Mining Allowed

- **A quick look at the Japanese exception**

« *It is permissible to **exploit work**, in any way and to the extent considered necessary, in any of the following cases or other **cases where such exploitation is not for enjoying or causing another person to enjoy the ideas or emotions expressed in such work**; provided, however that this does not apply if the exploitation would unreasonably prejudice the interests of the copyright owner in light of the natures and purposes of such work, as well as the circumstances of such exploitation:*
*(i) … (for experiment)*

*(ii)  … (for **data analysis**)*
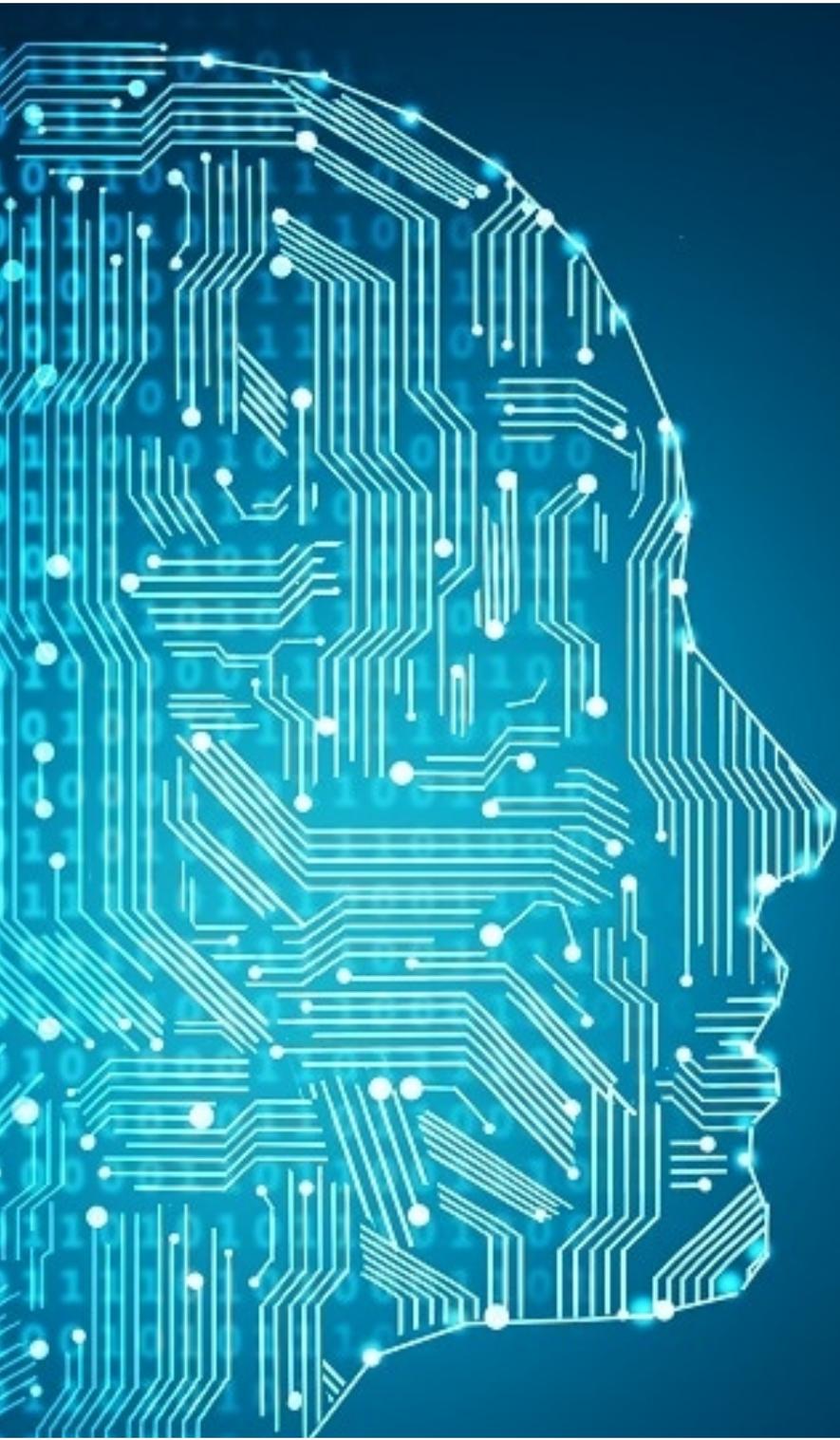
*(iii) … (without perceiving)* »

(art. 30-4 Japanese Copyright Act)

$\Rightarrow$ « ***Japan as a paradise for machine learning*** » (Ueno, 2019)

# Conclusion

# CONCLUSION

- **The Copyright issue is one issue among others**

- **The DSM Dir. will make EU Copyright law more AI friendly**

- **The new EU TDM (not limited to research) exception is affected with legal and pracitical uncertainties**

- **The exceptions allowing TDM are broader in other juridictions (outside of EU) and may offer greater perspective for AI-based businesses**

Thank you for your attention!

jcabay@ulb.ac.be