

Protection, Processing & Mining of Data in the EU: A Copyright Perspective

Brussels School of Artificial Intelligence, 17 February 2023

Julien Cabay

Professor Université Libre de Bruxelles (ULB – JurisLab)

Associate Professor Université de Liège (ULiège)

Invited Professor KULeuven

julien.cabay@ulb.be



Outline

- Introduction
- Part 1: Data Protection
- Part 2: Data Processing
- Part 3: Data Mining
- Conclusion



Part 1: Introduction



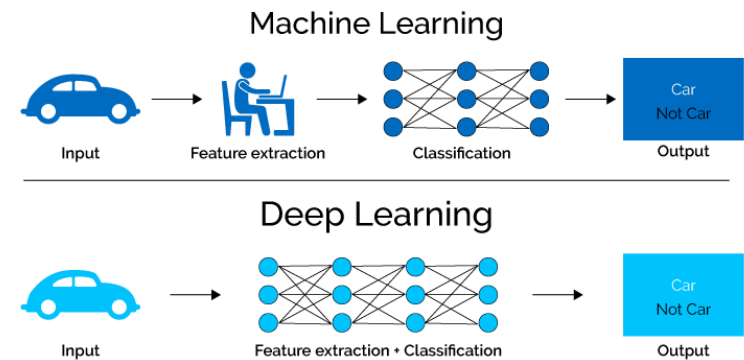
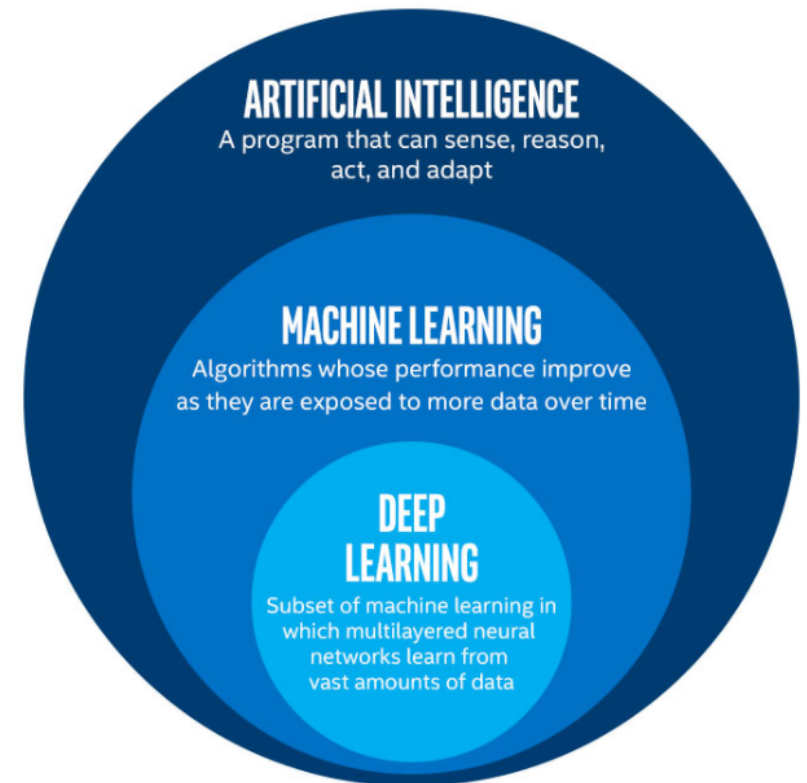
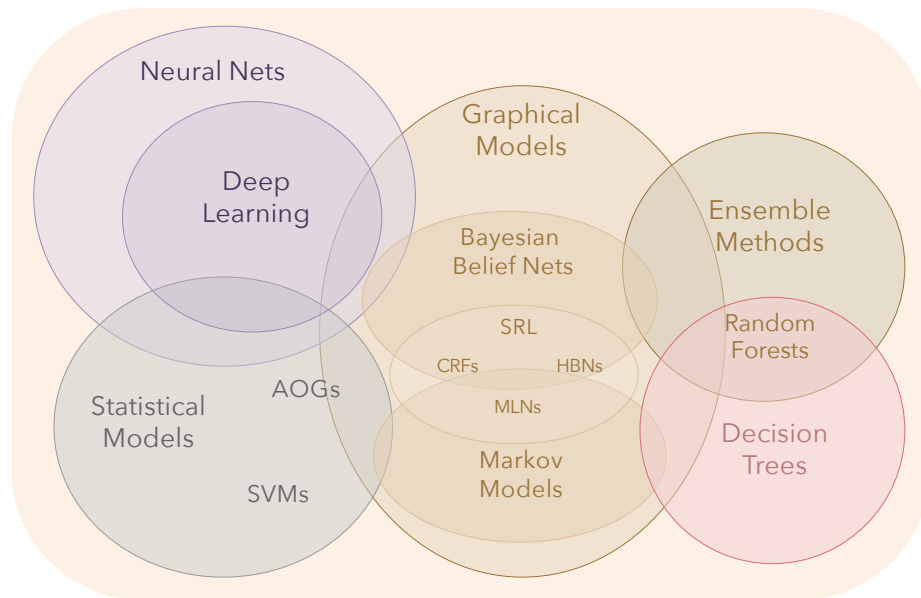
AI Promises

*« (...) 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 (...) »***

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

AI Promises

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

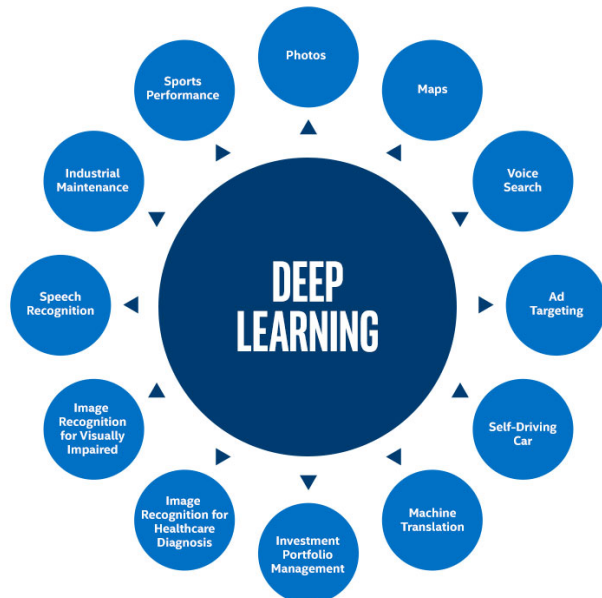




AI Promises

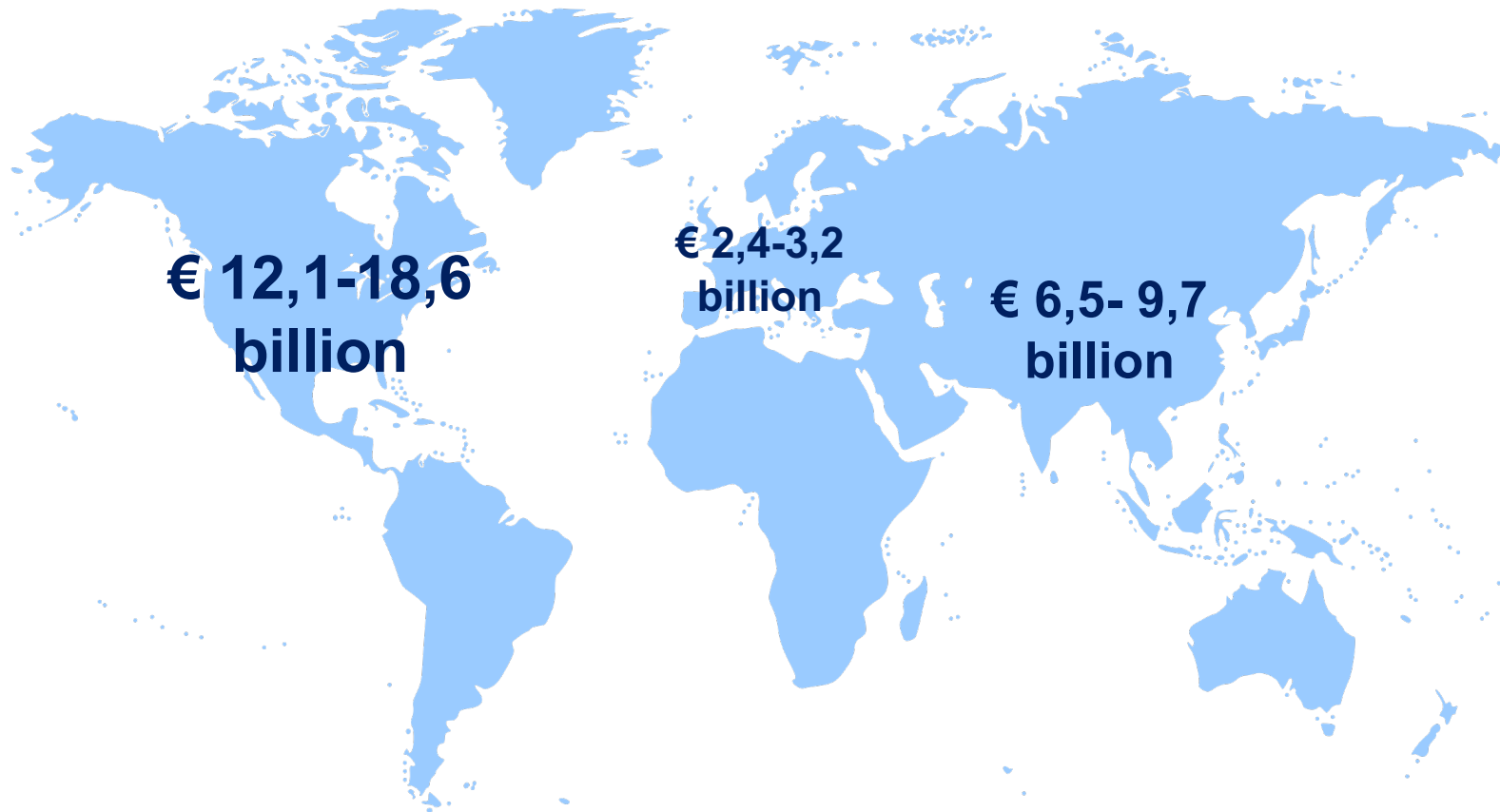
- Myriad of applications

Deep Learning Opportunities



AI Promises

- Overall private investments in AI in 2016





AI Promises

*« (...) **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 (...) »***

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



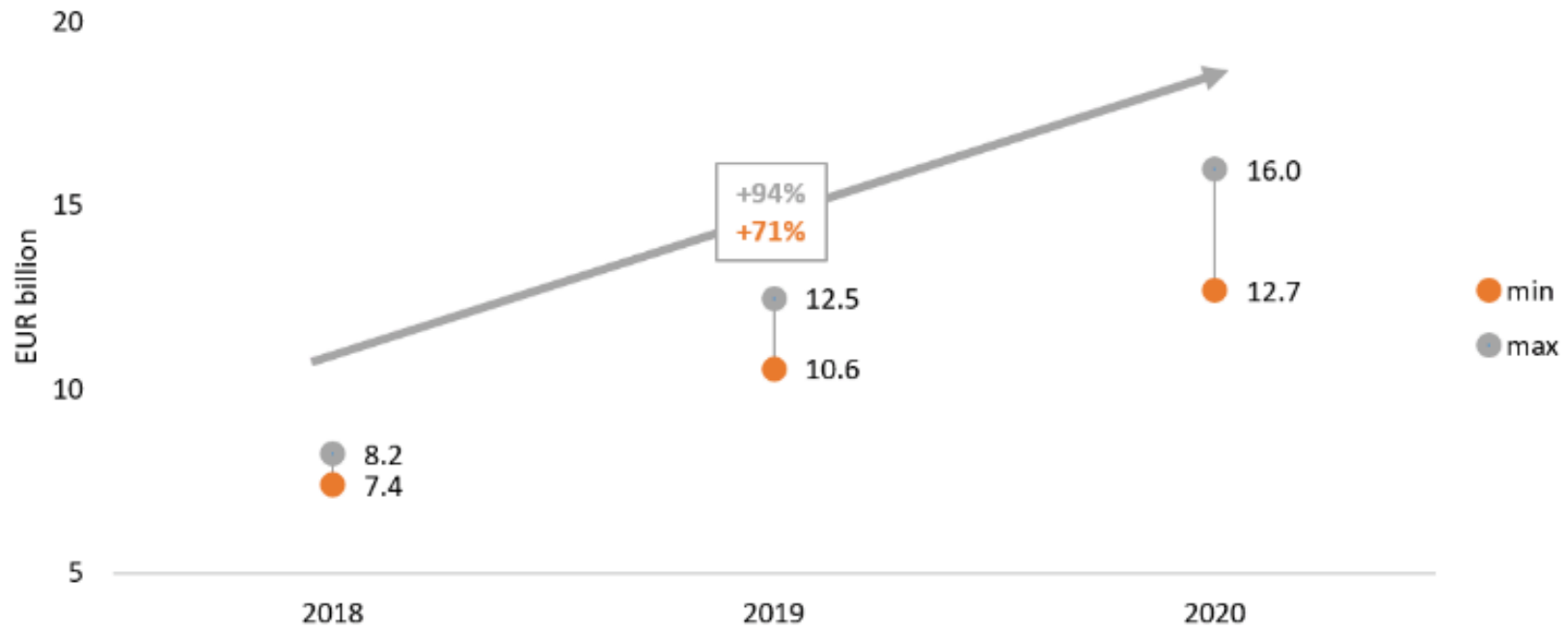
AI Promises

« (...) Over the **past three years, EU funding** for research and innovation for AI has **risen to €1.5 billion**, i.e. a 70% increase compared to the previous period (...) »

White Paper *On Artificial Intelligence – A European Approach to Excellence and Trust*,
Brussels, 19.2.2020, COM(2020) 65 final

AI Promises

• EU AI Investment 2018-2020

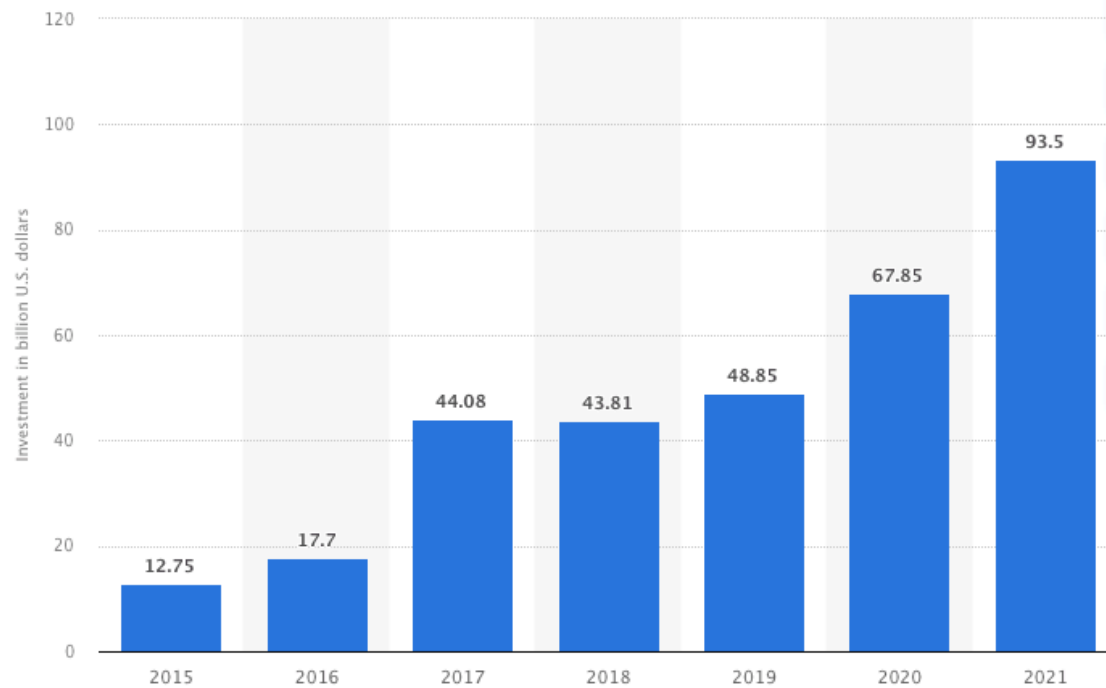


Evas, T., Sipinen, M., Ulbrich, M., Dalla Benetta, A., Sobolewski, M. and Nepelski, D., AI Watch: Estimating AI Investments in the European Union, EUR 31114 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53433-4, doi:10.2760/702029, JRC129174 :

<https://publications.jrc.ec.europa.eu/repository/handle/JRC129174>

AI Promises

- **US AI Investment 2015-2021**



<https://www.statista.com/statistics/941137/ai-investment-and-funding-worldwide/>


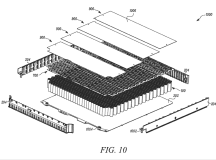




AI & IP

‘(...) 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 (...)’

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

IP basics

	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 character	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	



AI & IP

‘ ... ’

EU Commission White Paper, *On Artificial Intelligence – A European Approach to Excellence and Trust*, Brussels, 19.2.2020, COM(2020) 65 final ;
Communication, *A European strategy for data*, Brussels, 19.2.2020, COM(2020) 66 final ;
Communication, *Shaping Europe’s digital future*, Brussels, 19.2.2020, COM(2020) 67 final ;
Communication, *Fostering a European Approach to Artificial Intelligence*, Brussels, 21.4.2021, COM(2021) 205 final

AI & IP

*‘(...) the Commission will, together with stakeholders and IP offices, **explore the use of new technologies such as AI and blockchain to further improve the effectiveness of our IP systems.** In fact, new technologies can help facilitate the protection of IP, improve transparency, allow for a smoother distribution of license fees, and more effectively tackle counterfeiting and piracy (...)*’

*‘(...) In addition, the digital revolution requires **reflection on how and what is to be protected. AI technologies are creating new works and inventions.** In some cases, for instance in the cultural sector, the use of inventive machines may become the norm. These developments raise the question of what protection should be given to products created with the help of AI technologies (...)*’

‘Article 70 – Confidentiality

1. National competent authorities and notified bodies involved in the application of this Regulation shall respect the **confidentiality** of information and data obtained in carrying out their tasks and activities in such a manner as to protect, in particular:

(a) **intellectual property rights**, and confidential business information or trade secrets of a natural or legal person, **including source code**, except the cases referred to in Article 5 of Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure apply. (...)

IP as a hurdle to the development of AI?

*‘(...) the Commission will explore the need for legislative action on issues that affect relations between actors in the data-agile economy to provide incentives for horizontal data sharing across sectors (...). One or more of the following issues could be taken forward in a Data Act (2021): (...) **evaluating the IPR framework with a view to further enhance data access and use** (including a possible revision of the **Database Directive** and a possible clarification of the application of the **Trade Secrets Protection Directive** as an enabling framework) (...)’*

IP as a hurdle to the development of AI?

*‘(...) In order not to hinder the exercise of the right of users to access and use such data in accordance with Article 4 of this Regulation or of the right to share such data with third parties in accordance with Article 5 of this Regulation, the **sui generis right** provided for in Article 7 of Directive 96/9/EC **does not apply to databases containing data obtained from or generated by the use of a product or a related service (...)**’*

'IP' as a hurdle to the development of AI?

*'(...) the Commission will explore the need for legislative action on issues that affect relations between actors in the data-agile economy to provide incentives for horizontal data sharing across sectors (...). One or more of the following issues could be taken forward in a Data Act (2021): (...) **evaluating the IPR framework with a view to further enhance data access and use** (including a possible revision of the Database Directive and a possible clarification of the application of the **Trade Secrets Protection Directive** as an enabling framework) (...).'*

'IP' as a hurdle to the development of AI?

*'(...) It is true that the legal protection of **trade secrets facilitates the exchange of information.** (...) for more data sharing to happen, companies should feel confident that it takes place in a secure environment, with authorised companies and under the right conditions. Companies should benefit of tools that would help them **fight misappropriation or misuse of their data.** (...)'*

Call for tenders, Study on the legal protection of trade secrets in the context of the data economy (GRO/SME/20/F/206), 23.6.2020, Ares(2020)3269941

*'(...) **Trade secrets shall only be disclosed provided that all specific necessary measures are taken to preserve the confidentiality of trade secrets in particular with respect to third parties (...)**'*

Art. 4(3) Data Act Proposal (COM(2022) 68 final)

See also art. 5(8), 8(6), 17(2)(c), 19(2)

AI & Copyright

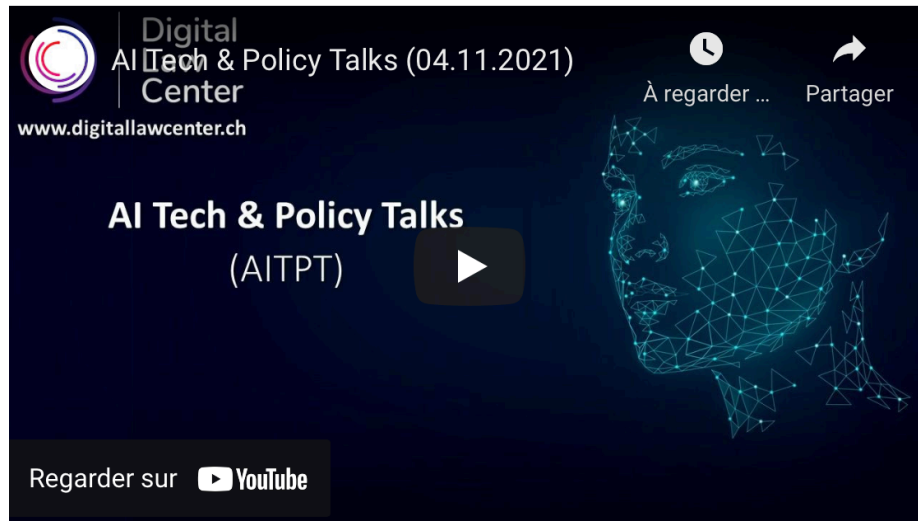
• Interaction between AI & Copyright ?

- 2 main (but not only) questions:
 - What if INPUT = Copyrighted works?
 - Could OUTPUT = Copyrighted works?

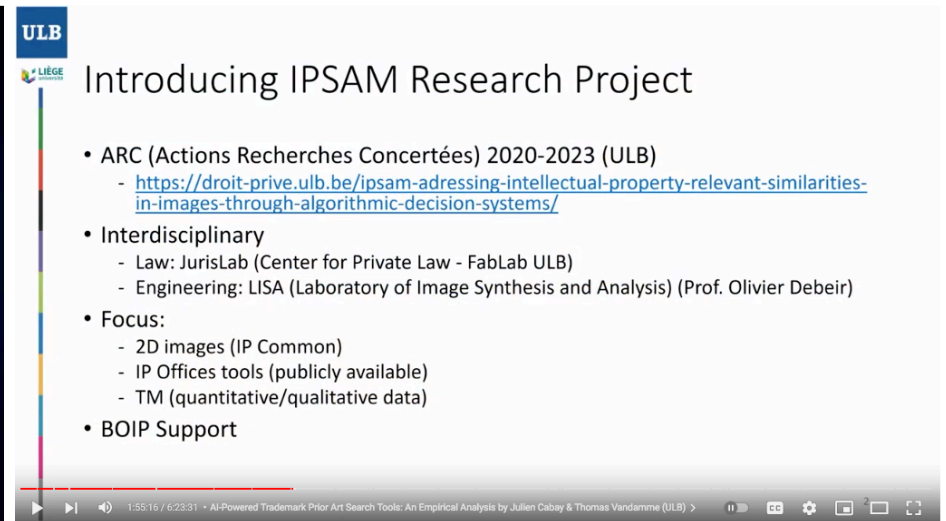


AI & IP in practice: IPSAM Research Project

- <https://droit-prive.ulb.be/ipsam-adressing-intellectual-property-relevant-similarities-in-images-through-algorithmic-decision-systems/>



<https://www.youtube.com/watch?v=dy2BNJGjF6o>



<https://www.youtube.com/watch?v=RP703UAKDRY&t=6858s>



Part 1: Data Protection

Data as inputs

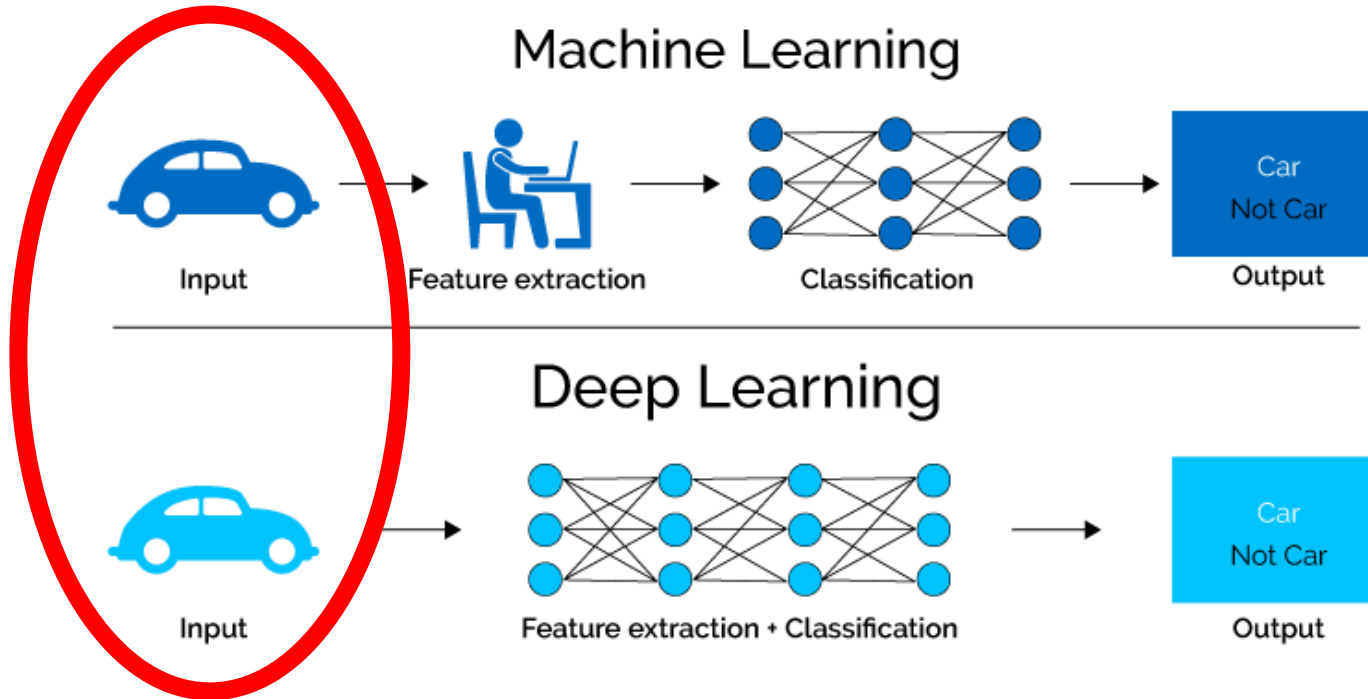


« (...) **AI needs vast amounts of data to be developed.** Machine learning, a type of AI, works by identifying patterns in available data and then applying the knowledge to new data. The larger a data set, the better even subtle relations in the data can be discovered. When it comes to using AI, data-rich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

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

The Economist, May 6th, 2017

Data as inputs



- Legal issues? Depending on the input/data

Data protection

• Protection available for data through IP

- Copyright/Neighbouring rights (InfoSoc Dir.; Software Dir.; Database Dir.; DSM Dir.)
 - Works (not ideas/facts)
 - Other subject-matters
- Database *sui generis* right (Database Dir.)
 - Substantial part of the content of a database
 - Not data itself
 - Investments in the collection, not the creation of data as a by-product of another economic activity (see *Fixtures marketing* cases [C-46/02; C-338/02; C-444/02] and *British Horseracing Board* [C-203/02])
 - If repeated and systematic, also unsubstantial parts

Data protection

- **Protection available for data outside of IP**

- 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
- Contract law (CJEU, *Ryanair*, C-30/14 (2015))
 - Data (restricted access/use, through terms and conditions)

Data protection

- **Allowing use/Fostering sharing of data**

- Directive (EU) 2019/1024 on open data and the re-use of public sector information (Open Data/PSI Dir.)
 - Documents held by public sector bodies; Research data
 - Scope:
 - *'minimum rules governing the **re-use** and the practical arrangements for facilitating the re-use of'*
- Regulation (EU) 2022/868 (Data Governance Act)
 - Data held by public sector bodies; Personal and non-personal data held by data subject (according to GDPR)
 - Scope:
 - *'conditions for the **re-use**, within the Union, of certain categories of **data** held by **public** sector bodies'*
 - *'framework for the provision of data intermediation services' (i.e. 'service which aims to establish commercial relationships for the purposes of **data sharing**)'*
 - *'framework for voluntary registration of entities which collect and process **data** made available for **altruistic** purposes'*
- (Data Act proposal [(COM)2022 68 final])
 - Data generated by the use of a product or related service
 - Scope :
 - *'making available to the user of that product or service'*
 - *'making available by data holders to data recipients' ('data sharing')*
 - *'making available by data holders to public sector bodies or Union institutions, agencies or bodies, where there is an exceptional need'*

Data copyright protection

- **Copyright and related rights**

- 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 Single Market [DSM Dir.]
 - Rights on press publications (Publishers)

Data copyright 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**

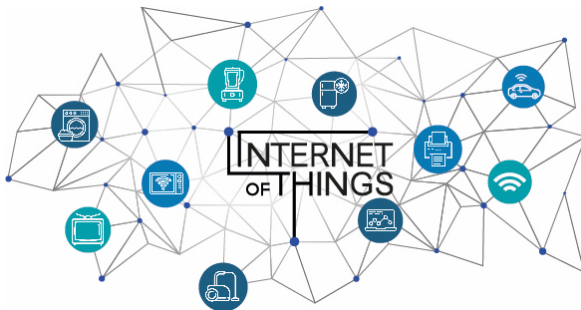
- = ‘Author’s own intellectual creation’ (CJEU, *Infopaq*, C-5/08 (2009))
- Criteria:
 - ‘Through the choice, sequence and combination’ (*idem*)
 - ‘Personal touch’ (CJEU, *Painer*, C-145/10 (2011))
 - ‘Free and creative choices’, not fulfilled when ‘dictated by technical considerations, rules or constraints which leave no room for creative freedom’ (CJEU, *Football Dataco*, C-604/10 (2012))

Data copyright protection

- **No © data**

- Topical example : raw machine-generated data and applications

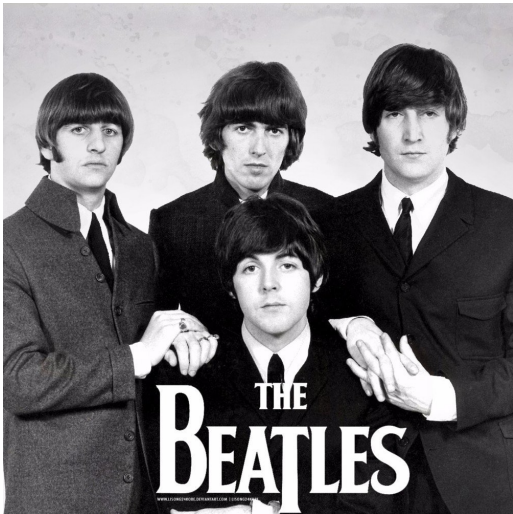
- See EU Commission, Communication *Building A European Data Economy*, Brussels, 10.1.2017, COM(2017) 9



Data copyright protection

- © data

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



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

Data copyright protection

- © data?

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





Part 2: Data Processing

Data processing & copyright?

© DATA/INPUT →



- **© questions to be addressed :**
 - (1) Are data reproduced in the © sense?
 - (2) If yes, is there any authorization?
 - (3) If not, is there any justification?

(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)



(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

(1) Are data reproduced in the © sense?

- CJEU, *Infopaq*, C-5/08 (2009)

- Data capture process

- 1) Selection of relevant publications
 - 2) Scanning 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

(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 »

(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 »*

(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/curation
 - Step 3: Data analysis, machine learning (training)

(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

(2) If yes, is there any authorization?

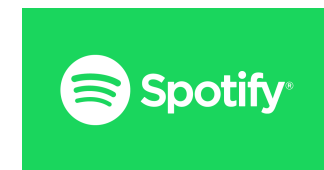
- Consent may be implicit, but not hypothetical
 - In particular, works made freely available on the Internet by its author are not free of © and cannot be reproduced
 - See CJEU, *Soulier & Doke*, C-301/15 (2016)

*« (...) 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** »*

(2) If yes, is there any authorization?

- It depends on the source:
 - Own generated data: © ownership
 - Ex: Sony (The Beatles' music)
 - Data transferred by customers : © 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)



(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 observed (≠ restrictive)
 - See 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

(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**

(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 (...) »

⇒ **NO**

(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 »

(3) If not, is there any justification?

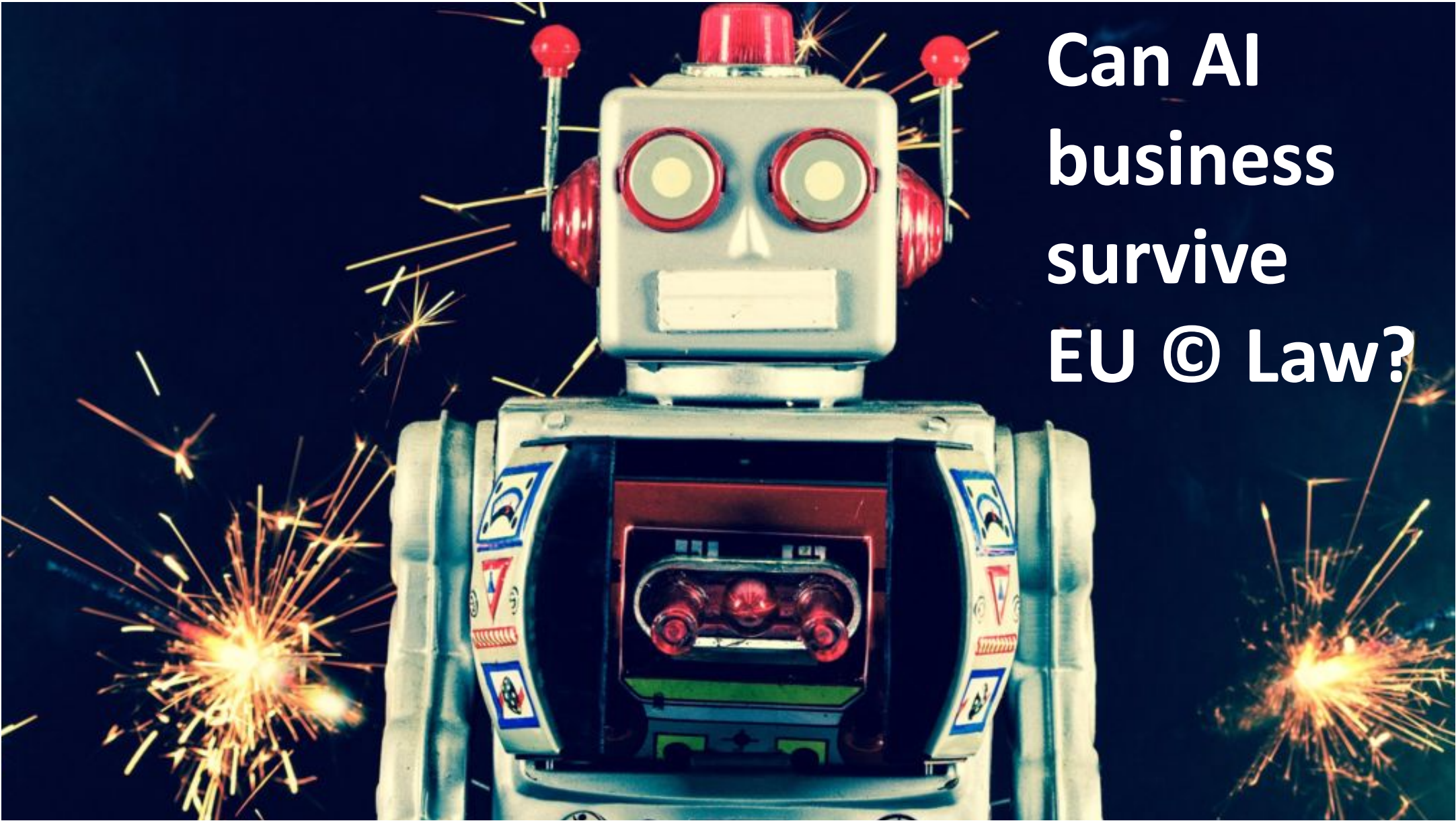
- **Exception for transient copies for AI business?**

- 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 3: Data Mining

Text and Data Mining Allowed

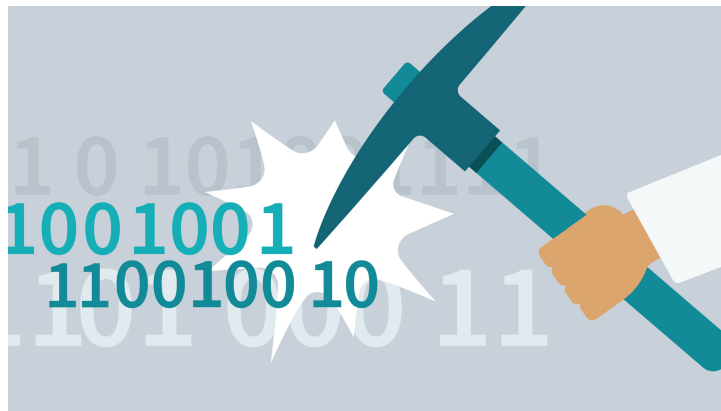
« **New technologies enable the automated computational analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining. Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering 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 »
- Applications in art. 3 (scientific research) and 4 (general)



Text and Data Mining Allowed

- **Much more than copyright...**

*« (...) It may be argued that under the **misleading** label Text and data mining (TDM) what has been regulated at the EU level in Arts. 3 and 4 goes far beyond a **mere copyright exception**. In fact, it should be **reclassified as the legal regulation of AI via the allocation of property rights in its building blocks**, or in other words, as a property-right approach to the regulation of AI (...) »*

(Margoni/Kretschmer 2021)

Non-commercial TDM

• TDM exceptions for non-commercial research

- Prior 2019: 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]
 - Germany (2017): § 60d Urheberrechtsgesetz
- Since 2019 (implementation deadline, 7 june 2021)
 - Art. 3 DSM Dir.
 - Mandatory exception
 - Belgium: Code de droit économique [CDE]
 - XI.191/1, §1er, 7° CDE (general)
 - XI.191/1, §1er, 7° CDE (database copyright)
 - XI.310, § 3, 1° CDE (database *sui generis* right)
 - XI.217/1, 6° CDE (neighbouring right)
 - No exception applicable to softwares

Non-commercial TDM

- 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. (...) »

Non-commercial TDM

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

- 'Research organisation', 'Cultural heritage institution' : 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)

Commercial TDM

- **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)

- Since 2019 (implementation deadline, 7 june 2021)

- Art. 4 DSM Dir.

- Mandatory exception

- Belgium: Code de droit économique [CDE]

- XI.190, 20° CDE (general)

- XI.190, 20° CDE (database copyright)

- XI.299, § 5 CDE (software)

- XI.310, § 3, 2° CDE (database *sui generis* right)

- XI.217, 19° CDE (neighbouring right)

Commercial TDM (EU)

- **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. »*

Commercial TDM (EU)

- **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

⇒ **AI business opportunity, but...**

Commercial TDM (EU)

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

- 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 (...) »
- 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)

⇒... **limitations, and...**

Commercial TDM (EU)

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

- Scope?

- Application to works made available prior adoption/entry into force of the directive?
 - See CJEU, *Soulier & Dore*, C-301/15 (2016) (no prior information on future use = hypothetical consent)
 - Applications to works with 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 (...)* »
 - Application to works not lawfully made available online?
 - See in particular CJEU, *Renckhoff*, C-161/17 (2018) (unauthorized making available online of work already lawfully made available online)

⇒... **legal uncertainty**



Commercial TDM (EU)

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

- Practice?

- Standard ('TDM.txt' akin to 'robots.txt')?
 - Distinguishing ©/not © by machines?
 - Evidence (TDM restriction/violation)?
 - Overlap with art. 3 DSM Dir.?
 - Licensing schemes?

⇒... **practical uncertainty**

Commercial TDM (USA)

- 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 multiple copies 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)

Commercial TDM (USA)

- **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* (...) »

Commercial TDM (USA)

- **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 (...)** »*



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

Commercial TDM (USA)

- A quick look at the USA exception

*« (...) When AI algorithms or processes “learn” their functions by ingesting copyrighted works, reproductions of those works are made in the process as the works are digitized and/or “read” by the AI algorithms or processes. **Some mass digitization scenarios may be a fair use, whereas others may be infringements.** Although mass digitization for purposes of machine learning (ML) “ingestion” processes —and large-scale ingestion of already-digitized works—has not yet been tested by the courts, some rights holders argue that AI trainers should be required to **compensate** the authors and rights holders whose copyrighted works their machines are ingesting as a simple matter of doing business »*

(USPTO 2020)

Commercial TDM (Japan)

- 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)

⇒« **Japan as a paradise for machine learning** » (Ueno, 2021)

TDM exceptions

	EU	China	USA	Japan
TDM	Art. 3 (scientific research) and 4 (general) Dir. 2019/790	Closed list, no TDM but law under revision and possible « fair use/fair dealing »	17 USC 107 (<i>fair use</i> , after <i>Authors Guild v. Google Inc.</i> , 804 F.3d 202 (2d Cir. 2015))	Art. 30-4 Japanese Copyright Act (exception 'not for enjoying the idea or emotions expressed in a work')
Reach	Limited	« Chinese courts have realised the disadvantage of limiting fair use to enumerated circumstances and brought in a more flexible approach for finding fair use by learning from their US counterparts » (Wang, He, p. 23)	« Uses involving robotic readers are fast-tracked for fair use » (Grimmelmann, p. 667) ; but doubts as to « expressive machine learning » (Sobel, pp. 66-79)	« Japan as a paradise for machine learning » (Ueno)

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



Selected sources

- Grimmelman, J., “Copyright for Literate Robots”, *101 Iowa Law Review* 657 (2016)
- Margoni, T., Kretschmer, M., “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology” (July 14, 2021). Available at SSRN: <https://ssrn.com/abstract=3886695> or <http://dx.doi.org/10.2139/ssrn.3886695>
- Sobel, B. L. W., “Artificial Intelligence’s Fair Use Crisis”, *41 Columbia Journal of Law & the Arts* 45 (2017)
- Ueno, T., “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication”, *GRUR International*, 70(2), 2021, 145–152
- USPTO, “Public Views on Artificial Intelligence and Intellectual Property Policy” (October 2020). Available at https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-05.pdf?utm_campaign=subscriptioncenter&utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=
- Wang, J., He, T., “To share is fair: The changing face of China’s fair use doctrine in the sharing economy and beyond”, *Computer Law & Security Review*, 35(1), 2019, 15-28



Thank you for your attention!

julien.cabay@ulb.be



Back up slides





(Part 4: Data Production)



REMBRANDT



THE NEXT REMBRANDT

Copyright for AI-generated « works »?

- Hot debate, based on a fallacy
 - *Pro* : similarities (qualitative)
 - Comp. « As humans beings recede from direct participation in the creation of many works, continued insistence on human authorship as a prerequisite to copyright threatens the protection – and, ultimately, the production- of works that are **indistinguishable in merit and value** from protected works created by human beings » (Denicola (2016), p. 269)
 - *Contra* : differences (quantitative)
 - Comp. « **AI-generation of new creation** based on a training set **can be unleashed** with little marginal costs, and can explore any kind of combinations and variations » (Sartor, Lagioia, Contissa (2018), p. 12)

⇒ **Difference in scale is critical**

⇒ **AI production « as such » should remain in the public domain**

Copyright for AI-assisted works?

- Hotter debate, based on a second fallacy
 - *Pro*: distinction computer-assisted works/computer-generated works
 - *Contra*: no clear-cut dichotomy
 - Comp. « There is a **continuum** between, at one extreme, 'computer-assisted' works, and at the other extreme, autonomously-generated works » (McCutcheon (2013), p. 929)

⇒ **In theory, AI production « implemented » in a human creation might be protected**

⇒ **In practice, most of the concerns for AI production « as such » remain and legal uncertainty since the distinction computer-assisted works/computer generated works is more of a *continuum* rather than a dichotomy**

My conclusion

- **AI production is (quantitatively) different from Human production and if IP protection is evidenced as necessary, then less protection than copyright might prove better to avoid undesirable consequences**
 - See J. CABAY, “Copyright protection For AI Production? An EU Law Perspective”, <https://euipo.europa.eu/knowledge/enrol/index.php?id=4035>
 - See J. CABAY, “Droit d’auteur et intelligence artificielle : comparaison n’est pas raison”, *Entertainment & Law*, 2019, pp. 307-325
 - See J. CABAY, “Mort ou résurrection de l’auteur ? A propos de l’intelligence artificielle et de la propriété intellectuelle”, *Rev. Dr. ULg*, 2019, pp. 179-190





Conclusion