AI-Powered Trade Mark Search Engines
&
Computation of the Likelihood of Confusion Test

Beneluxvereniging voor Merken- en Modellen recht (BMM) Conference, Antwerp, 26 March 2026

Julien Cabay

Associate Professor ULB
Associate Professor ULiège
Invited Professor KULeuven
julien.cabay@ulb.be

Outline

- Introduction
- The IPSAM Project
- Pilot Study
- Main Studies – Key Findings
- Further Research

INTRODUCTION


YouTube
Broadcast Yourself™
amazon.com
IKEA
adidas
Nikon
Google
facebook
twitter
Unilever
SAMSUNG
Coca-Cola
YAHOO!
MasterCard

TM LoC (I)

*"1. Upon opposition by the proprietor of an earlier trade mark, the **trade mark** applied for **shall not be registered**:*
(…)
*(b) if, because of its **identity with, or similarity** to, the earlier trade **mark** and the identity or similarity of the **goods or services** covered by the trade marks there exists a **likelihood of confusion** on the part of the public in the territory in which the earlier trade mark is protected; the likelihood of confusion includes the likelihood of association with the earlier trade mark."*

Art. 8(1)(b) EU Trade Mark Regulation 2017/1001 [EUTMR]

// Art. 5(1)(b) EU Trade Mark Directive 2015/2436 [EUTMD], *implemented in* Art. 2.2*ter*(1)(b) Benelux Convention on Intellectual Property [BCIP]

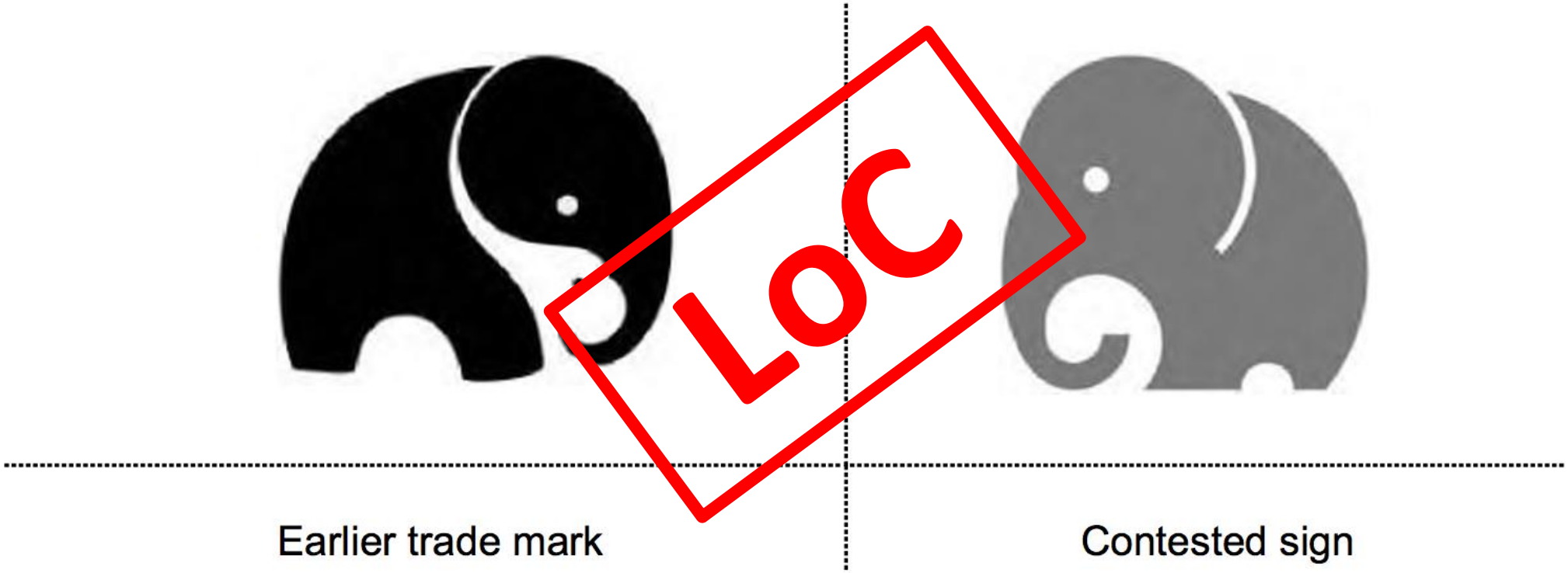
TM LoC (II)

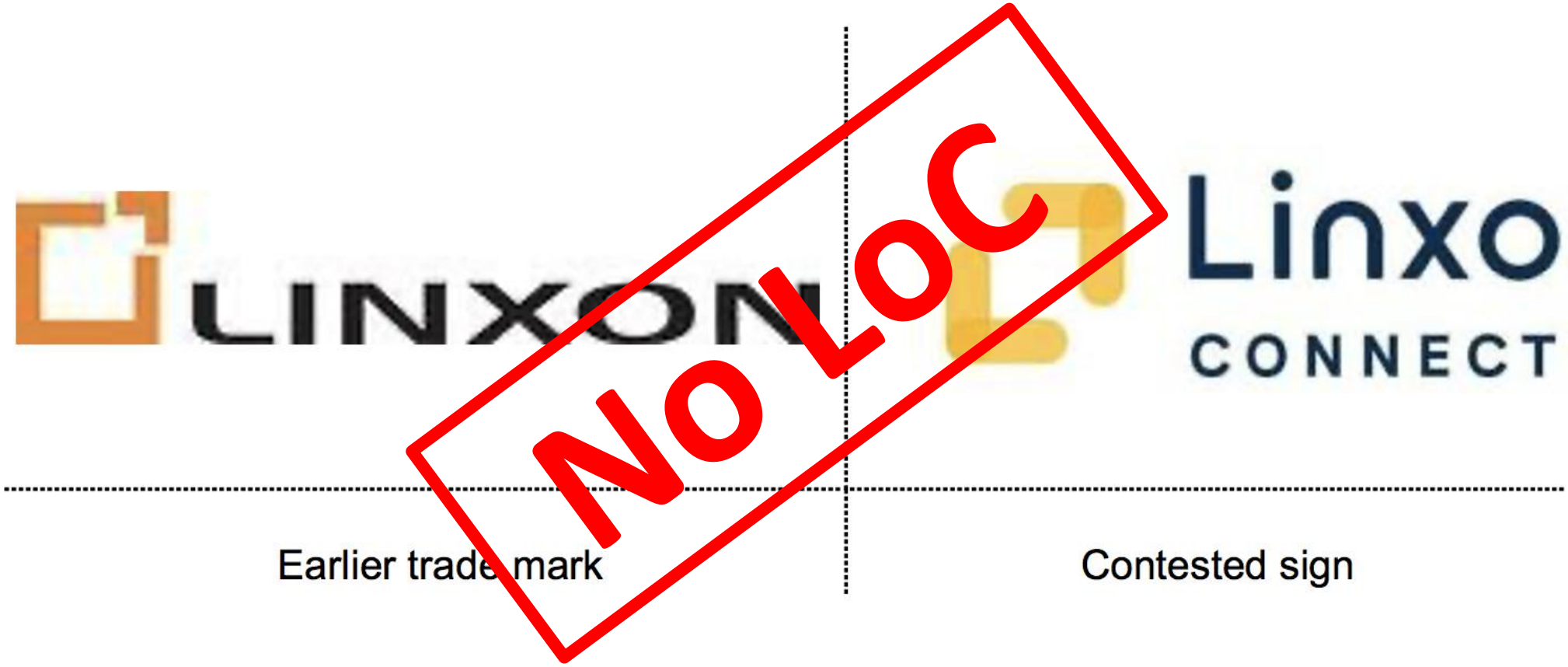
*"That global appreciation of the **visual, aural or conceptual similarity** of the marks in question, must be based on the overall impression given by the marks, bearing in mind, in particular, their distinctive and dominant components. The wording of Article 4(1)(b) of the Directive - '... there exists a **likelihood of confusion** on the part of the public ...' - shows that the perception of marks in the mind of the average consumer of the type of goods or services in question plays a decisive role in the global appreciation of the likelihood of confusion. The **average consumer** normally perceives a mark as a whole and does not proceed to analyse its various details."* (point 23)

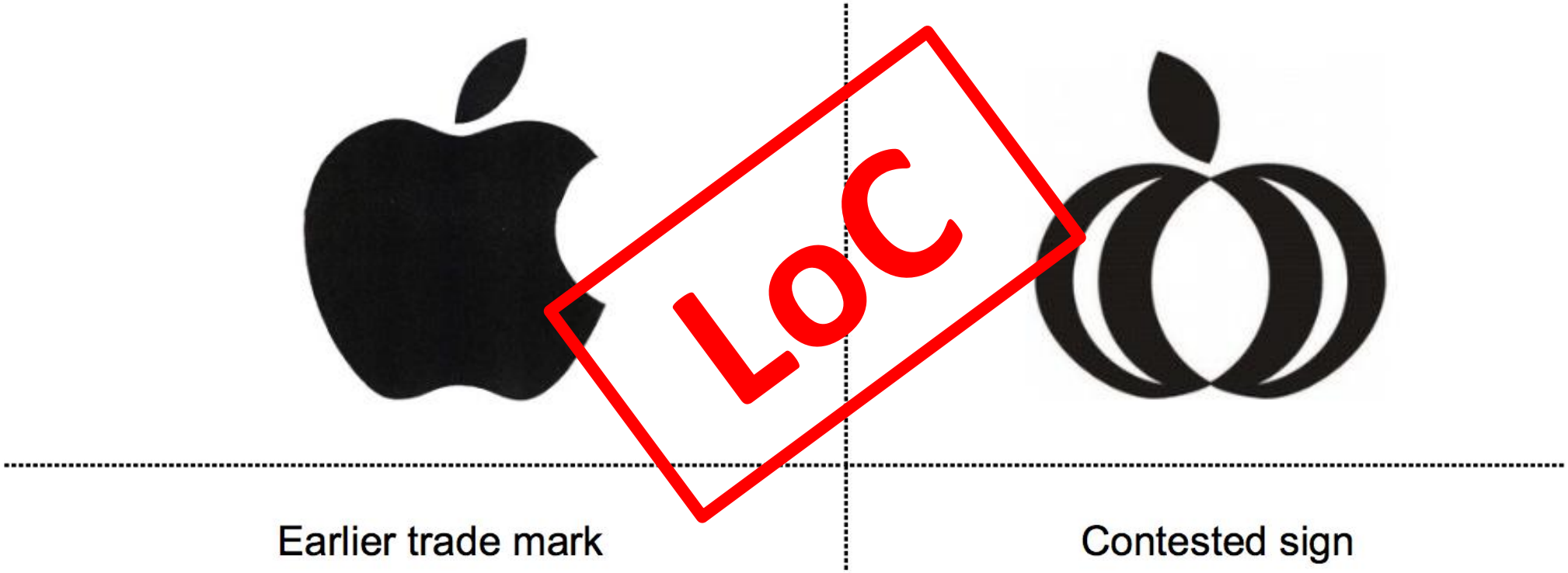
CJEU, *Sabel v Puma*, C-251/95 (1997)

TRADEMARK

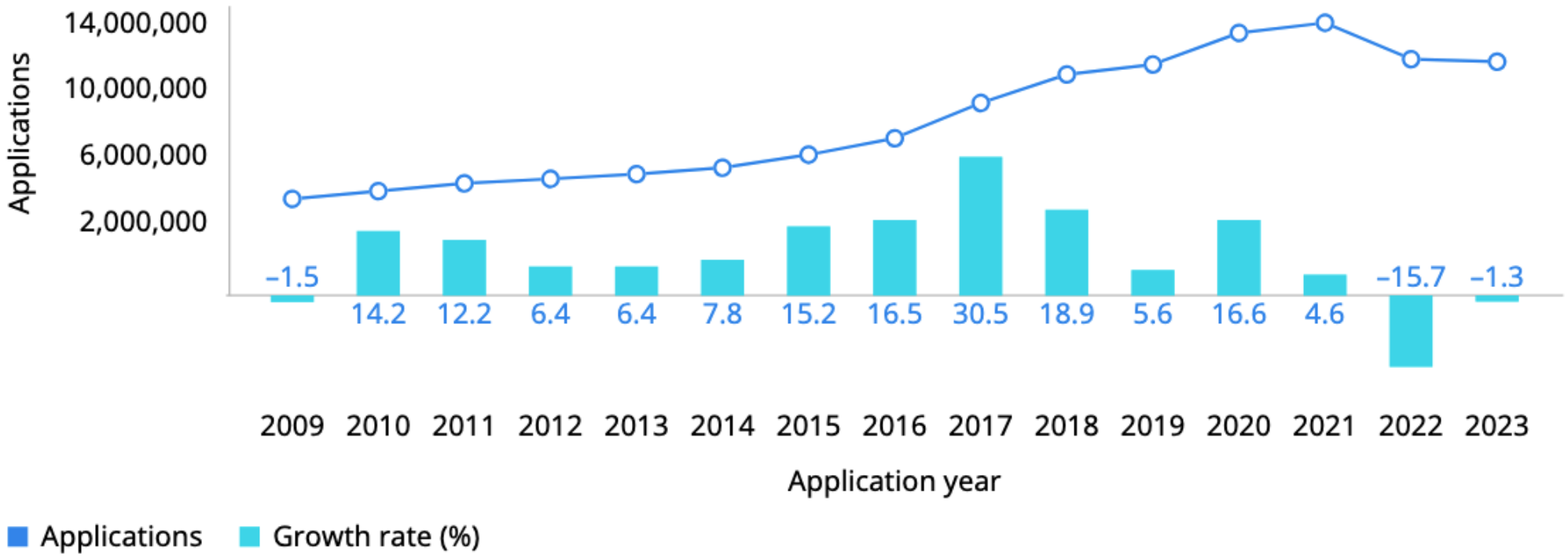
QUIZ


LoC
Earlier trade mark
Contested sign


LINXON
Linxo
CONNECT
No LoC
Earlier trade mark
Contested sign


LoC
Earlier trade mark
Contested sign

B1. **Trend in trademark applications worldwide, 2009–2023**



Source: World Intellectual Property Indicators 2024, p. 82

What do you see?

- *Although the figurative element of the contested sign does not clearly represent a specific fruit and may therefore be perceived in different ways by consumers, in view of its overall shape and the dip at the top and bottom, at least a non-negligible part of the relevant public in the European Union will perceive it, together with the detached leaf, as a* ***stylised representation of an apple***

  [Opposition No B 3 059 743]

English AA Login

EUIPO EUROPEAN UNION INTELLECTUAL PROPERTY OFFICE

Protect your intellectual property in the European Union

The EUIPO's database *access*

Search trade marks, designs, owners, representatives, Bulletins and Office decisions in one single application. Learn how

Advanced search

Search

Drag 1 image to search for trade marks and up to 7 for designs

JPG, PNG, GIF and TIF are allowed

To obtain better results upload high quality images

Terms and Conditions

5.9.0-RC1.1

About EUIPO

European Union Intellectual Property Office

Public procurement

Jobs

Press

Transparency portal

An agency of the European Union

EU Agencies Network

Help & FAQs

Find the answers to all your questions on trade marks, designs and our online services

Contact us

Our Information Centre staff can guide you before and during the application process

Avenida de Europa, 4, 03008 Alicante, Spain
Information Centre: +34 965139100

Twitter

LinkedIn

Facebook

YouTube

Instagram

eSearch plus

The EUIPO's database *access*

Search trade marks, designs, owners, representatives, Bulletins and Office decisions in one single application. Learn how

Advanced search

Search

Drag 1 image to search for trade marks and up to 7 for designs

JPG, PNG, GIF and TIF are allowed

To obtain better results upload high quality images

Terms and Conditions

What do you see?

- *The contested trade mark, also appears, consists of the **depiction of a stylised** heart or fruit, such as a **apple**, a red colour that depicts a morso [bite] on the upper right hand side and a black leafette when inclined on its top.*

  [Opposition No B 2 042 755]

What do you see?

- *(…) the contested sign, being a very abstract representation, might also be perceived, because of its colour, form and leaf, as an orange or anyway as a citrus fruit, but* ***definitely not*** *as* ***an apple****.*

  [Opposition No B 2 354 911]

What do you see?

- *(…) the opponent asserts that the contested application will be perceived as a **stylized apple** device. The Opposition Division cannot concur with this finding and considers that **such perception would be too unrealistic and inventive**.*

  [Opposition No B 3 144 739]

Here is what the machine sees…

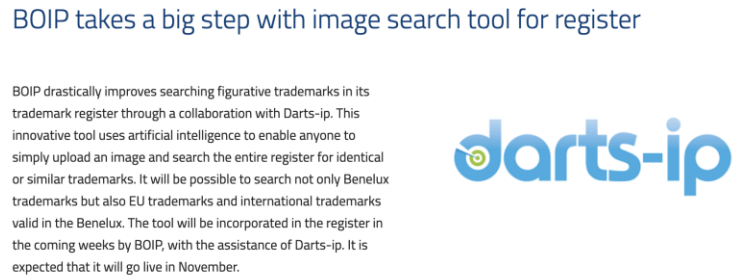
THE IPSAM PROJECT

Intellectual Property Similarities Assessment Model (IPSAM)

- ARC (Actions Recherches Concertées) 2020-2025 (ULB)
  - https://droit-prive.ulb.be/ipsam-adressing-intellectual-property-relevant-similarities-in-images-through-algorithmic-decision-systems/
- Interdisciplinary
  - Law: JurisLab (Center for Private Law - FabLab ULB)
    - Pr. Julien Cabay
  - Engineering: LISA (Language and Image, Synthesis and Analysis)
    - Pr. Ir. Olivier Debeir ; Dr. Ir. Thomas Vandamme
- BOIP Support
- 1 Ph.D. Thesis, 2 papers & several conferences

IP Similarity and Technology (back in 2021)

- Solutions “at scale”
  - Private Companies
    - Monitoring and enforcement
    - Prior art and clearance search
  - Regulator
    - Intermediaries content moderation
  - **Public IP Offices**
    - Registration
- Solutions “at scale” are here to stay => IPSAM
  - Focus:
    - 2D images (IP Common)
    - **IP Offices tools (publicly available)**
    - TM (quantitative/qualitative data)



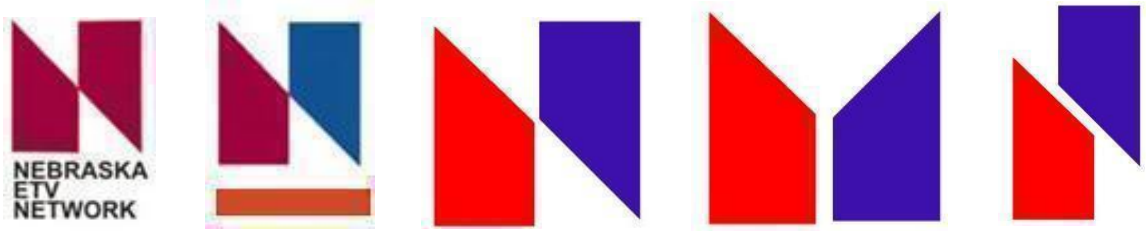
SOTA AI-Powered TM tools (back in 2021)

- Technical
  - Tursun *e.a.*
    - ‘METU’ TM Dataset (Tursun *e.a.* 2017)
    - Similarities identified by ‘expert’
    - Text removal (Tursun *e.a.* 2019)
  - Trappey *e.a.* 2020
    - Logos Dataset
    - 300 US infringement case law for evaluation purpose
- Involving legal scholars
  - Katyal & Kesari 2020
    - 115 Word TMs
    - USPTO rejections based on 15 U.S.C. § 1052(d)
  - Moerland & Freitas 2021
    - Seven type of searches, including figurative TMs

Component-based Attention for Large-scale Trademark Retrieval

Osman Tursun *[1], Simon Denman[1], Sabesan Sivapalan[1], Sridha Sridharan[1], Clinton Fookes[1], and Sandra Mau[2]

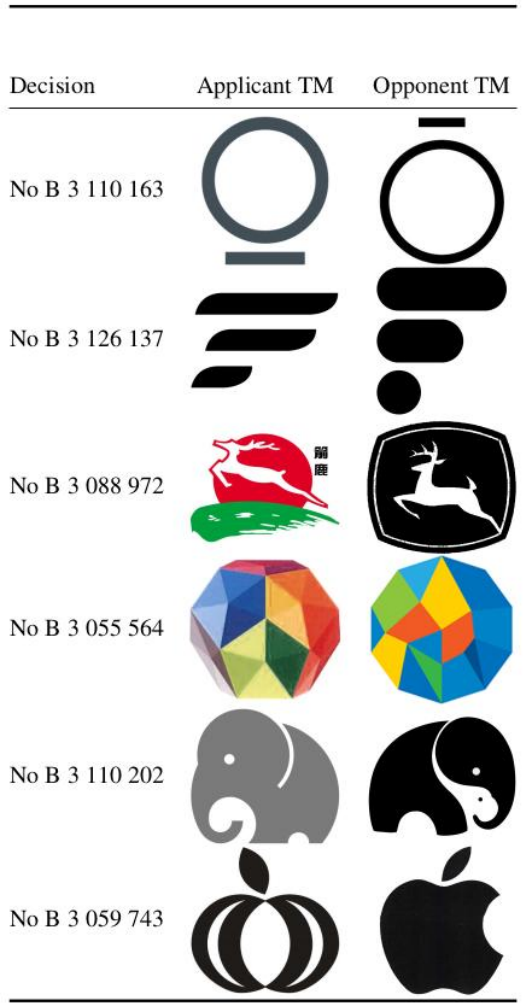
[1]Image and Video Research Laboratory, SAIVT, Queensland University of Technology
[2]TrademarkVision



PILOT STUDY

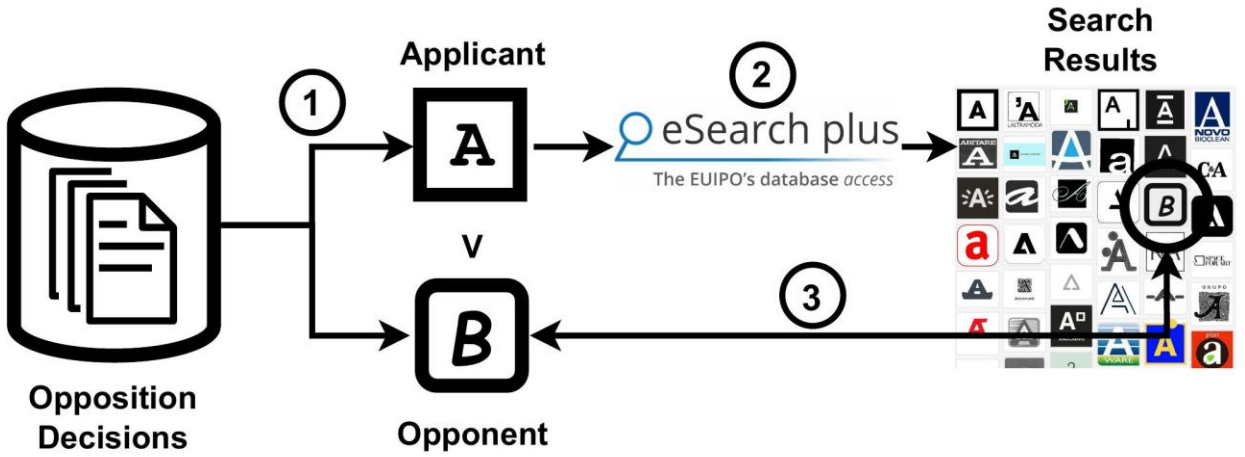
Ground Truth

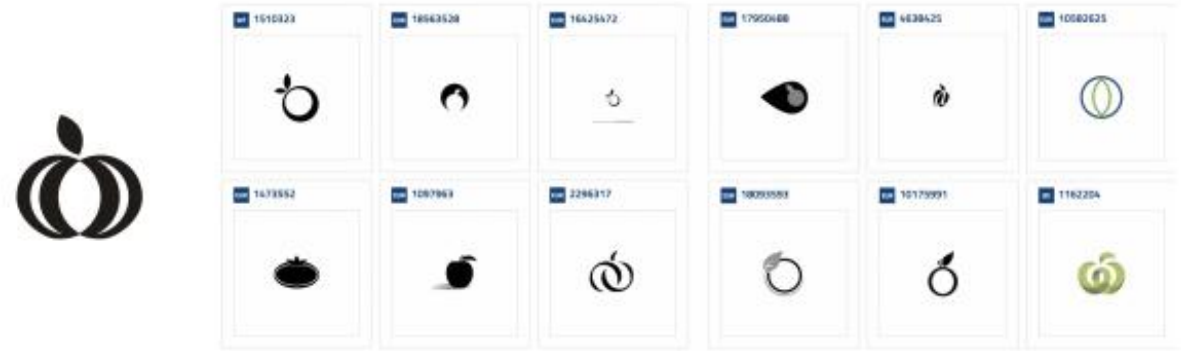
- Methodology
  - **LoC according to EUIPO (art. 8(1)(b) EUTMR)**
    - // Katyal & Kesari 2021 (USPTO rejections based on 15 U.S.C. § 1052(d))
    - ≠ METU dataset (Tursun *e.a.* 2017: similarities identified by 'expert')
  - **Figurative EUTM**
    - ≠ Katyal & Kesari 2021 (word TM)
    - // METU dataset (Tursun *e.a.* 2017: text removal)
  - Testing BOIP, EUIPO and WIPO tools
    - // Moerland & Freitas 2021



Method

- Methodology
  - LoC according to EUIPO (art. 8(1)(b) EUTMR)
    - // Katyal & Kesari 2021 (USPTO rejections based on 15 U.S.C. § 1052(d))
    - ≠ METU dataset (Tursun & Aker & Kalkan 2017: similarities identified by 'expert')
  - Figurative EUTM
    - ≠ Katyal & Kesari 2021 (word TM)
    - // METU dataset (Tursun & Aker & Kalkan 2017: text removal)
  - **Match/Rank in BOIP, EUIPO and WIPO tools**
    - // Moerland & Freitas 2021





(a) Query Image (b) Retrieval Results

Figure 3.2.: Retrieval results from the BOIP system.

(a) Query Image (b) Retrieval Results

Figure 3.3.: Retrieval results from the EUIPO system.

Conclusions

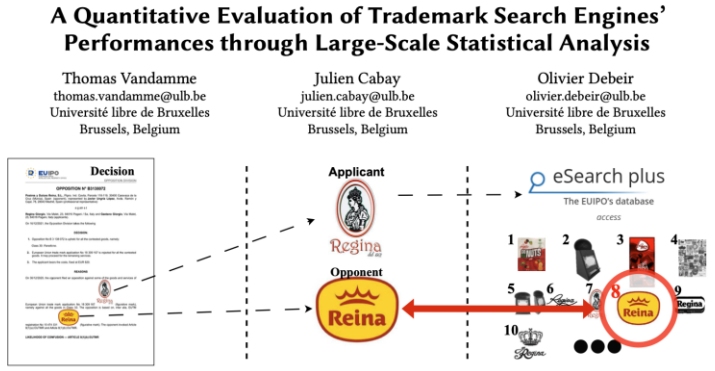
- Underwhelming performances:
  - Opponent TM rarely found by the systems
  - Retrievals occur late, when they occur (e.g. the Apple TM found in position 562 at the EUIPO and not found by the BOIP)
- Surprising results
  - Ex. Apple TM correct matches in WIPO's tool: 238 (Spain), 264 (Corea), 472 (North Macedonia) !

- Necessity to **scale the experiment up**, increase testing dataset size
  - Quantitatively objectify the tools performances

MAIN STUDIES – KEY FINDINGS

Scaling Up

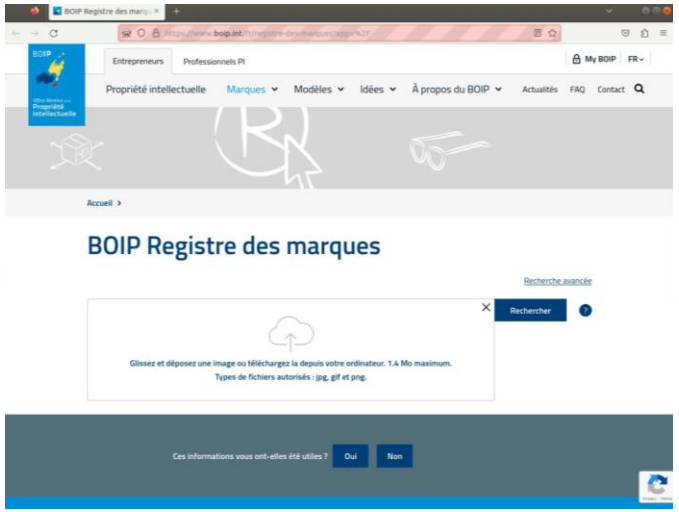
- Test Sets: Figurative TM Opposition Decisions
  - Study 1 : **5852 EUIPO**
  - Study 2: **553 BOIP**
- Automated Data Mining Process
  - **TDM** exception for scientific research purposes (art. 3. CDSM Directive 2019/790)



Study 1: T. Vandamme, J. Cabay, O. Debeir, "A Quantitative Evaluation of Trademark Search Engines' Performances through Large-Scale Statistical Analysis", *in* X., *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023), June 19–23, 2023, Braga, Portugal*, ACM, New York (NY, USA), 2023, pp. 343-350

Study 2: J. Cabay, T. Vandamme, O. Debeir, "Looking through the Crack in the Black Box: A Comparative Case Law Benchmark for Auditing AI-Powered Trade Mark Search Engines", *Computer Law & Security Review,* 2025, Vol. 59, 106167

TDM



DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 17 April 2019

on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

Article 3

Text and data mining for the purposes of scientific research

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. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively.

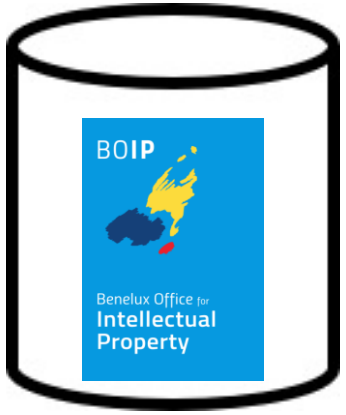
Performances

2023

System	Decisions	Matches	Hit Rate	Average Rank
BOIP (EU Fig. Only)	4 024	372	9.24 %	10.42
EUIPO-nBOIP	4 024	3 338	82.95 % *	2.27

* Test set contamination

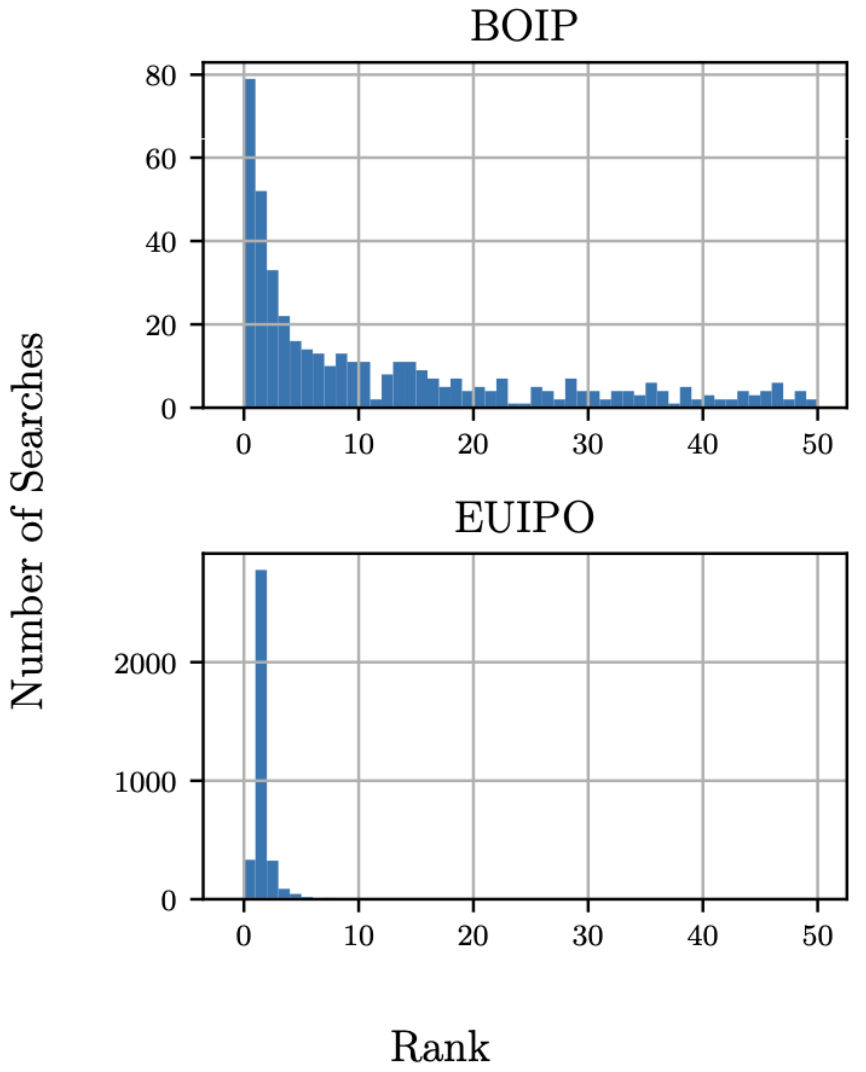
2025



System	Decisions	Matches	Hit Rate	Average Rank
BOIP (Fig. Only)	183	16	8.74 %	11.38
EUIPO-nBOIP	183	12	6.56 %	4.42

Latest numbers according to refined framework *in*
T. Vandamme, *Algorithmic Confusion: A Transversal Study of Computational Trade Mark*, Ph.D. in Engineering and Technology, ULB, 2025, 217 p.

Study 1 Key Finding



> 2.91. As a conclusion, the EUIPO's system is significantly more performant than BOIP's, for both metrics. However, we question the validity of the EUIPO's numbers, since the almost systematic presence of the match in second position. Our main hypothesis is that EUIPO's system was indeed trained on the very same set (or at least a major subset) that we used. Further research will need to verify this, for example by using another set of decisions.

Study 2 Key Findings

- The EUIPO's vs the BOIP's system: low performances in general
- The BOIP's system: no noticeable influence of the opposition decisions (BOIP or EUIPO) and opacity of the system
- **The EUIPO's system: the EUIPO opposition decisions as training dataset**
  - **Test set contamination** (= data used for evaluation appear in the training set)
  - **Overfitting** (= the model has "memorised" the training dataset)

Overfitting (I)

- **Results – EUIPO Overfitting**
  - Considering a different testing dataset reduces the HR from 82.95% to 6.56% (> 12-fold reduction)
  - New, uncorrupted performances, are slightly below those of the BOIP

⇒ Test-Set of EUIPO decisions was **contaminated**

⇒ The model is **overfitted**

With decisions from the EUIPO

System	Decisions	Matches	Hit Rate	Average Rank
BOIP (EU Fig. Only)	4 024	372	9.24 %	10.42
EUIPO-nBOIP	4 024	3 338	82.95 %	2.27

Table 4.10.: Performances under the Model Comparis framework. The six matches "lost" to the BOIP between the sults under the *Systems Benchmark* framework (378, see T e 4.9) and these results (372) are due to the removal of deci ns concerning opponent TM which were absent from the EU O database. These six matches concerned decisions which were the 82 (4 106 minus 4 024) decisions removed from the set of nsidered decisions.

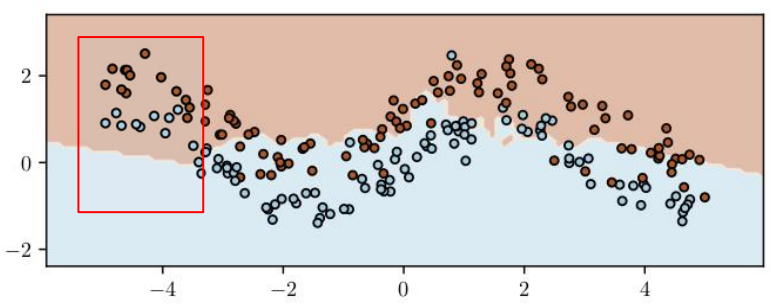
With decisions from the BOIP

System	Decisions	Matches	Hit te	Average Rank
BOIP (Fig. Only)	183	16	8.74 %	11.38
EUIPO-nBOIP	183	12	6.56 %	4.42

Table 5.9.: Performances under the Model Comparison framework, without considering the decisions that have an equivalent at the EUIPO.

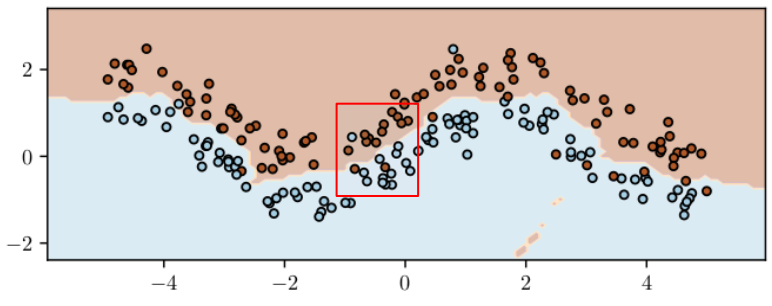
Overfitting (II)

Underfitting



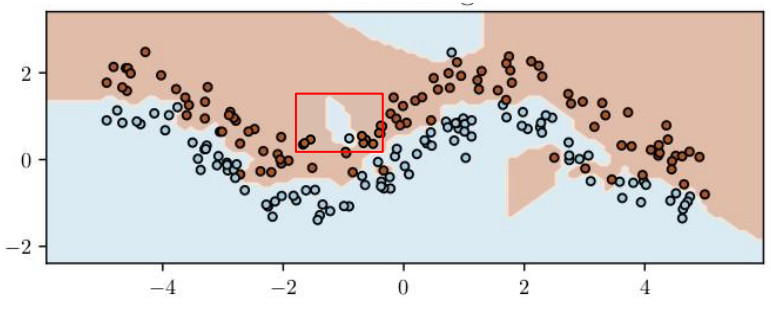
Major Classification Errors

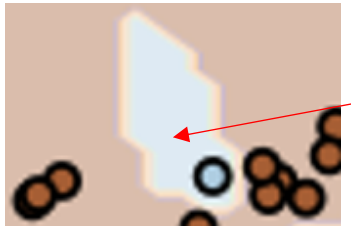
Good fit



Minor Classification Errors

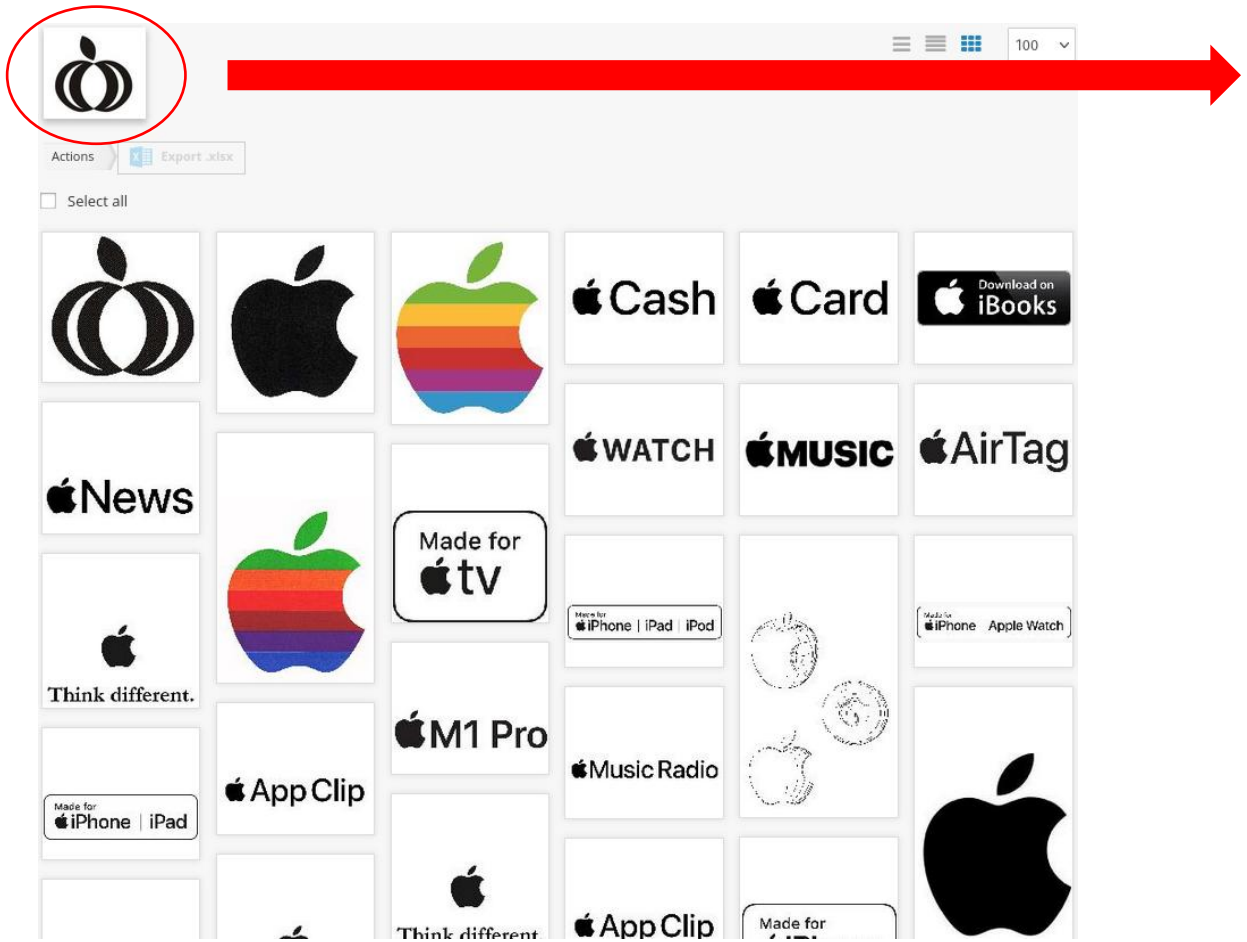
Overfitting

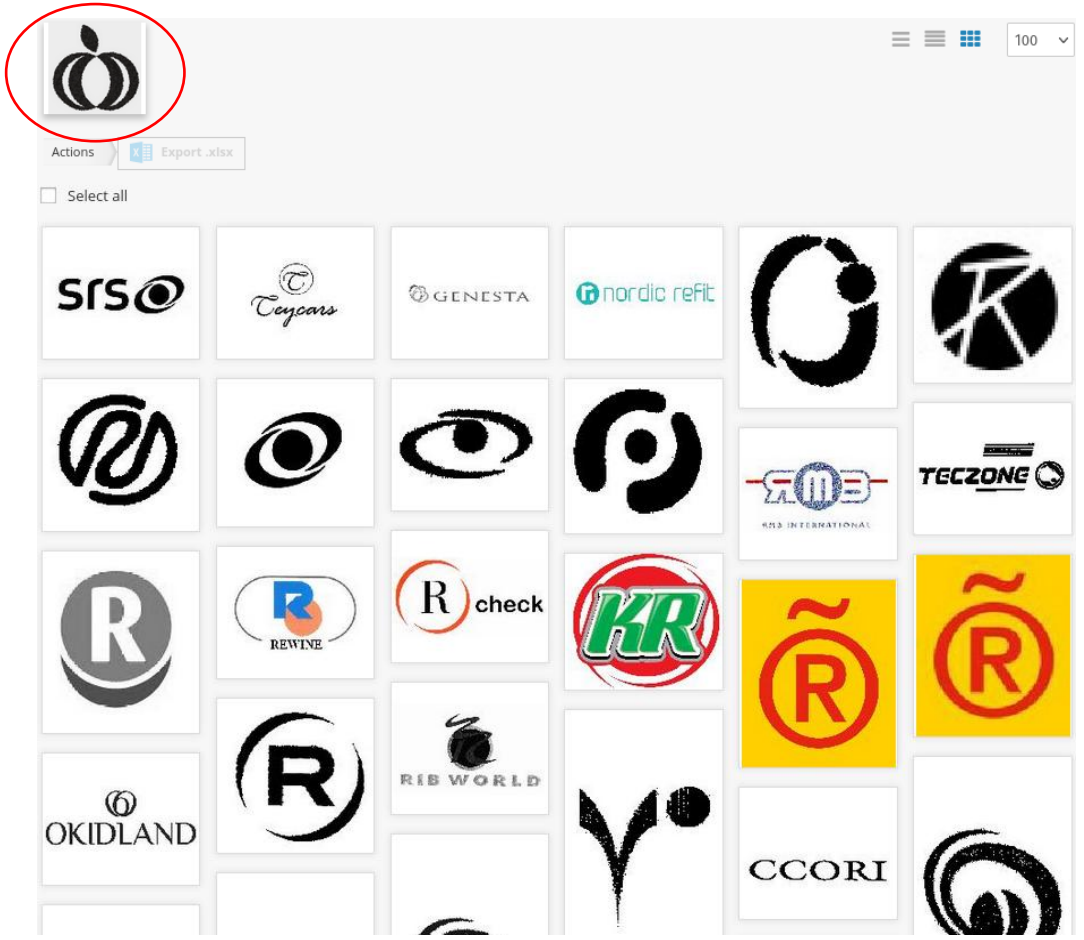




Dramatic Generalization Errors

Overfitting (III)



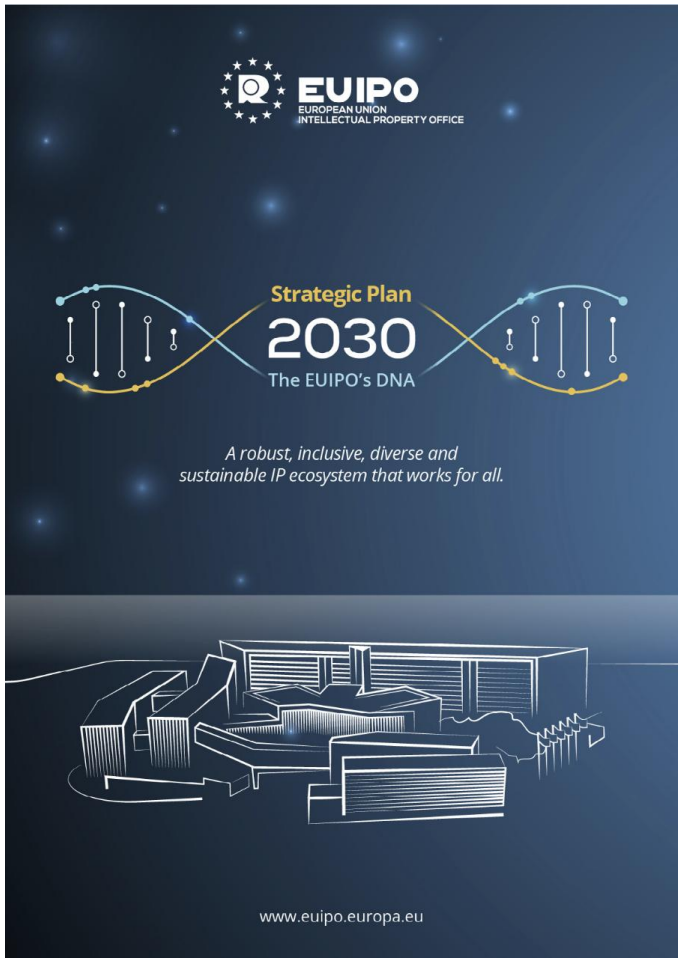


Key Take Aways…

- Resorting to technology for TM (more generally IP) similarities?
  - **Low performances**
    - Hit rate < 10% = business case not met
  - **Benchmarking biased**
    - Study 1: Overfitting hypothesis was an « accident » (*see* histograms)
    - Study 2: Evidencing overfitting subject to
      - Understanding training rationale and data (LoC)
      - Alternative datasets (BOIP decisions vs. EUIPO decisions)
      - Accessible (open data)
      - Identical principles of law (EUTMR = EUTMD/BCIP)
      - Not used as training set (EUIPO ≠ BOIP)
      - TDM exception for scientific research purposes (no opt-out)
- Solutions “at scale” are here to stay (IPSAM assumption)
  => further research needed

FURTHER RESEARCH

Further Development: AI @ IP Offices (I)



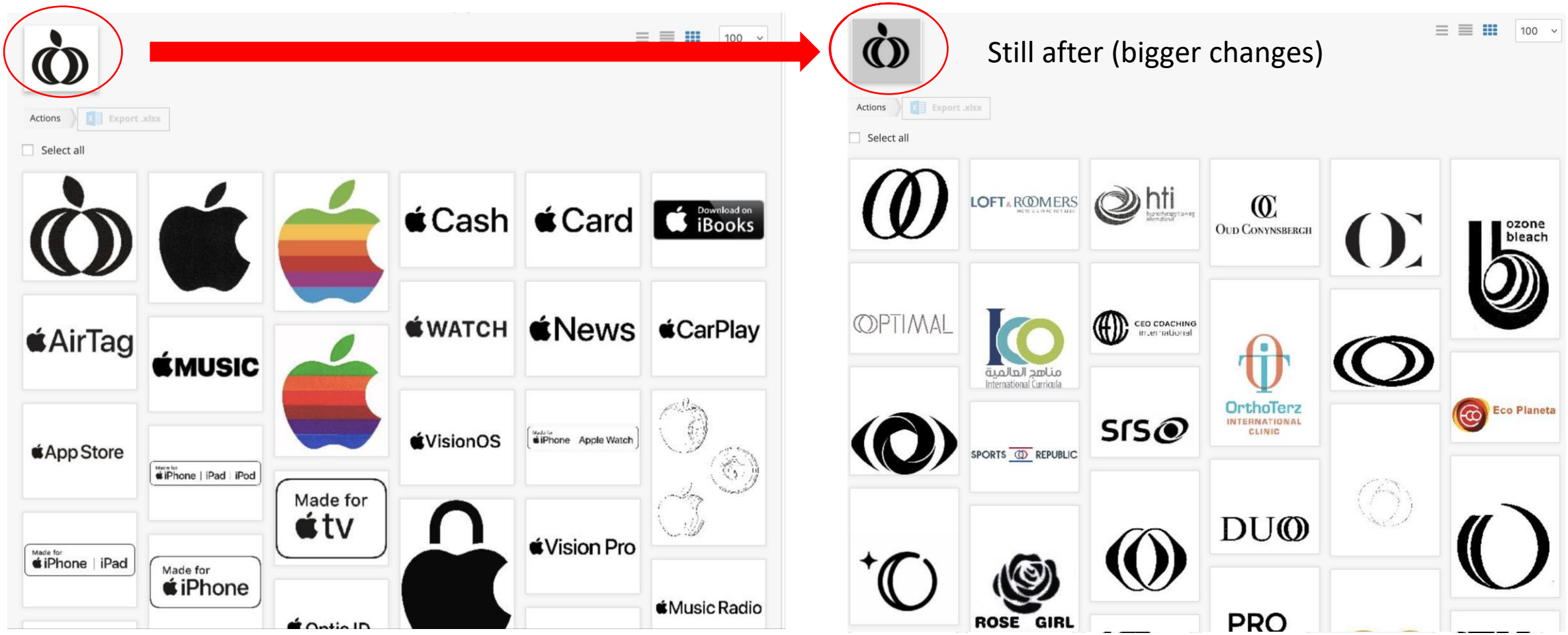
- **Key Initiative 3**

'Talent, technology, together' – take full advantage of new technologies

AI is not a rival to human intelligence; rather, it is to be used in partnership with it, with the final decision always taken by the human. We will develop AI-powered solutions to increase consistency and predictability by reducing the risks of refusals and deficiencies. We will increase efficiency and productivity by expanding automation for routine and time-consuming tasks. Over time, with the adoption of AI and machine learning, an important aspect of the role of examiners at the EUIPO will evolve towards becoming the trainers of the AI systems in the initial stages of the use of such tools, working to enhance their accuracy and scalability. When implementing AI solutions, we will ensure that the conversation between human and machine will always remain human-centric. The final evaluation of the output of the AI tool and the final decision will always be the responsibility of the examiner. We will also investigate the adoption of low-code development, which has the potential to enable staff to create applications with minimal coding effort.

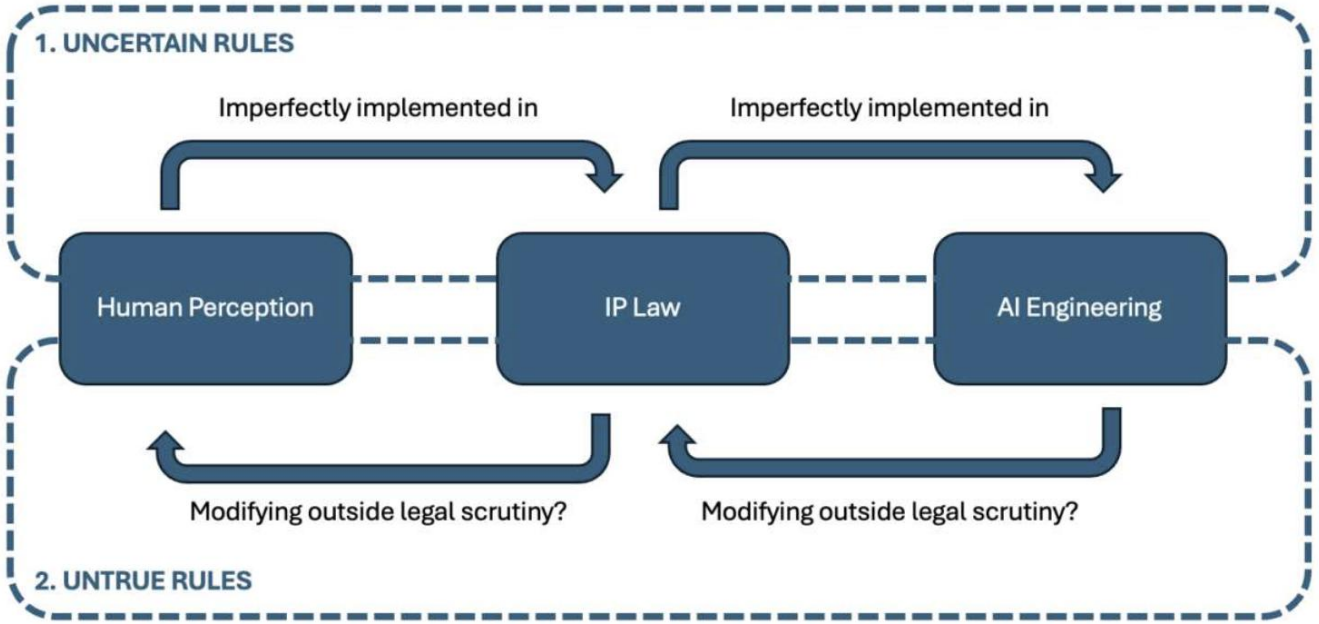
European Union Intellectual Propery Office, EUIPO Strategic Plan 2030, December 2024, p. 22

Further Development: AI @ IP Offices (II)



Further Research: AnthroPI @ JurisLab

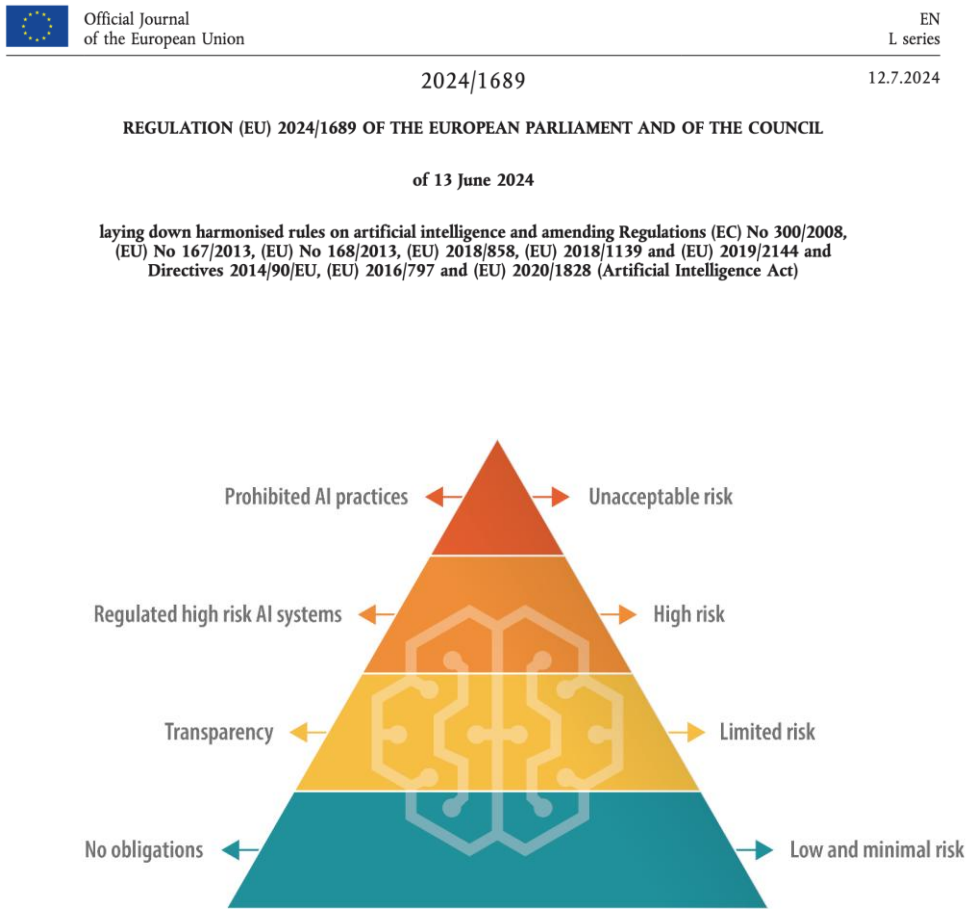
- ***AI-Driven Harmonisation of Similarity Assessment under the Intellectual Property Anthropocentric Framework: From Uncertain to Untrue***
  [AnthroPI]
  - FNRS research project (MIS 2026-2028)
  - Interdisciplinary
  - https://droit-prive.ulb.be/anthropi/



- WP1: engineering of AI-powered IP similarity assessment tools
- WP2: **transparency and explainability** of AI-powered IP similarity assessment tools
- WP3: **automation biases** in the use of AI-powered IP similarity assessment tools
- WP4: impact of AI-powered IP similarity assessment tools on the anthropocentric approach of IP law

Transparency & Explainability (I)

- In particular: *'(…) **we do not know the exact mechanics** of how each search engine defines similarity or the the treshold that each chooses when optimizing information retrieval'* (Katyal & Kesari 2020)
- In general
  - Issue: conditions for black-box evaluation of 'Legal AIs' cannot be met in most cases (no alternative test set)
  - Paradox: if they can be met (alternative test set), training set does not include all relevant data -> representative of the status of the law?
  - Avenue: enhanced transparency requirements on 'Legal AIs' are necessary



⇒**What about the AI Act?**

Transparency & Explainability (II)

Article 6

Classification rules for high-risk AI systems

2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall be considered to be high-risk.

ANNEX III

High-risk AI systems referred to in Article 6(2)

8. Administration of justice and democratic processes:

(a) AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution;

Transparency & Explainability (III)

Article 6

Classification rules for high-risk AI systems

3. By derogation from paragraph 2, an AI system referred to in Annex III shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.

The first subparagraph shall apply where any of the following conditions is fulfilled:

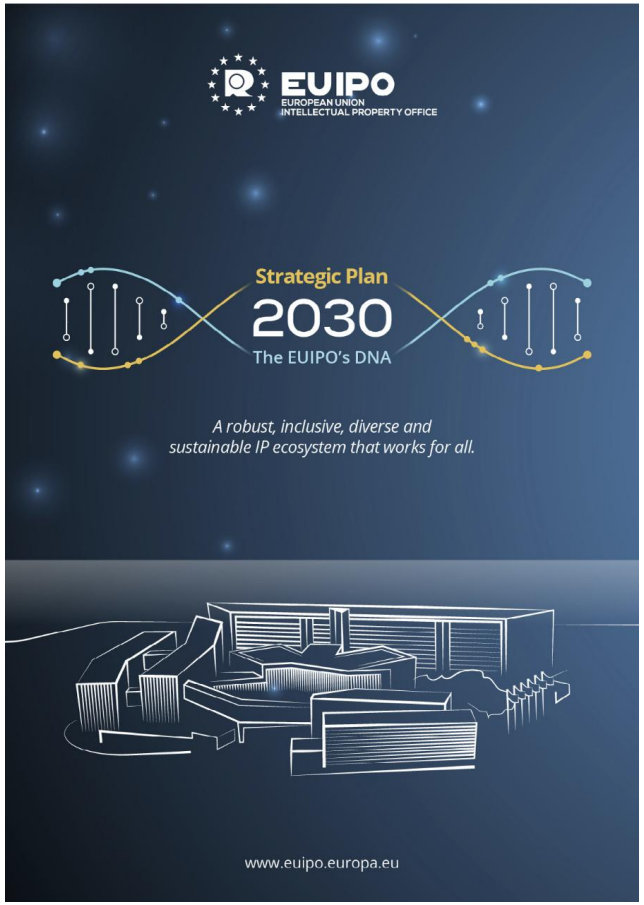
(c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or

Article 14

Human oversight

4. For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

(b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;



Transparency & Explainability (IV)

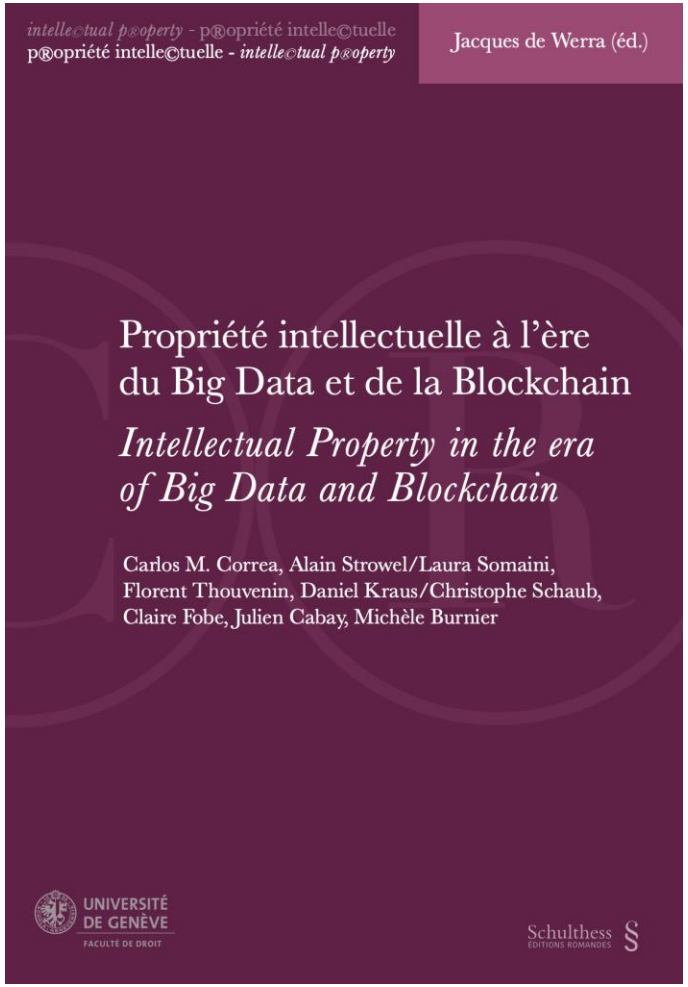
- Starting the conversation in Academia

Transparantieverplichtingen?

- TM search = "high risk" volgens art. 6 lid 2 AI Act?
- Annex III: "Administration of justice and democratic processes"
- Cabay et al terughoudend (p. 7), ik niet
- Dus: transparantieverplichting bestaat al

  "AI-systemen met een hoog risico worden op zodanige wijze ontworpen en ontwikkeld dat de werking ervan voldoende transparant is om gebruiks-verantwoordelijken in staat te stellen de output van een systeem te interpreteren en op passende wijze te gebruiken." (art. 13 lid 1 AI Act)
- Gebruiksverantwoordelijken = alle users, niet alleen bureau

Automation Biases (I)



"Personally, it is not the artificial intelligence that I fear, but the use that practitioners could make of the tools. The abandonment of critical thinking when faced with a statistics bar, for example"

(Fobe 2020, p. 166)

(former head of legal content at Darts-IP, now part of Clarivate)

Automation Biases (II)

XAI

=

PROBLEM ?

RESEARCH ARTICLE

JCP JOURNAL OF CONSUMER PSYCHOLOGY SCP SOCIETY FOR CONSUMER PSYCHOLOGY

Unintended effects of algorithmic transparency: The mere prospect of an explanation can foster the illusion of understanding how an algorithm works

Massimiliano Ostinelli[1] | **Andrea Bonezzi**[2] | **Monika Lisjak**[3]

Automation Biases (III)

- Starting the conversation in Academia

The role of explainable AI in the context of the AI Act

Cecilia Panigutti*
cecilia.panigutti@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Ronan Hamon*
ronan.hamon@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Isabelle Hupont*
isabelle.hupont-torres@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

David Fernandez Llorca
david.fernandez-llorca@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Delia Fano Yela
delia.fano-yela@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Henrik Junklewitz
henrik.junklewitz@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Salvatore Scalzo
salvatore.scalzo@ec.europa.eu
European Commission
Brussels, Belgium

Gabriele Mazzini
gabriele.mazzini@ec.europa.eu
European Commission
Brussels, Belgium

Ignacio Sanchez
ignacio.sanchez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Josep Soler Garrido
josep.soler-garrido@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Emilia Gomez
emilia.gomez-gutierrez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

> making them the original target users of AI explanations. Therefore, it might be considered surprising to discover that, even for this stakeholder group, AI explanations might increase the risk of automation bias. Indeed, it has been shown that AI experts tends to over-trust the AI system if presented with explanations containing numerical representations [32] and visualizations [62]. These findings suggest that further scientific research is needed to better understand the relationship between AI explanations and trust.

“There is a crack in everything, that’s how the light gets in.”

Thanks ! julien.cabay@ulb.be

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

Vandamme, T. (2025). **Algorithmic Confusion: A Transversal Study of Computational Trade Mark Similarity**. https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/390049/Details

Quoted sources

- IPSAM Project
  - Study 1: T. Vandamme, J. Cabay, O. Debeir, "A Quantitative Evaluation of Trademark Search Engines' Performances through Large-Scale Statistical Analysis", *in* X., *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023), June 19–23, 2023, Braga, Portugal*, ACM, New York (NY, USA), 2023, pp. 343-350
  - Study 2: J. Cabay, T. Vandamme, O. Debeir, "Looking through the Crack in the Black Box: A Comparative Case Law Benchmark for Auditing AI-Powered Trade Mark Search Engines", *Computer Law & Security Review,* 2025, Vol. 59, 106167
  - Ph.D. thesis: T. Vandamme, *Algorithmic Confusion: A Transversal Study of Computational Trade Mark,* Ph.D. in Engineering and Technology, ULB, 2025, 217 p.
- Other
  - C. Fobe, "Les outils des recherche générés par les 'Big Data' et l'intelligence artificielle », *in* J. de Werra (éd.), *Propriété intellectuelle à l'ère du Big Data et de la Blockchain – Intellectual Property in the era of Big Data and Blockchain*, Genève - Zürich, Schulthess Editions Romandes, 2020
  - S. K. Katyal, A. Kesari, "Trademark Search, Artificial Intelligence, and the Role of the Private Sector", *35 Berkeley Tech. L.J. 501* (2020)
  - M. Ostinelli, A., Bonezzi, M., Lisjak, "Unintended Effects of Algorithmic Transparency: The Mere Prospect of an Explanation Can Foster the Illusion of Understanding How an Algorithm Works", *Journal of Consumer Psychology*, 2024
  - C. Panigutti, e.a., "The role of explainable AI in the context of the AI Act", in *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023, pp. 1139-1150
  - O. Tursun, C. Aker, S. Kalkan, "A Large-scale Dataset and Benchmark for Similar Trademark Retrieval", 14 Oct. 2017, https://arXiv.org/abs/1701.05766
  - O. Tursun, S. Denman, S. Sivapalan, S. Sridharan, C. Fookes, S. Mau, "Component-based Attention for Large-scale Trademark Retrieval," *IEEE Trans. Inform. Forensic Secur.*, 2019, pp. 1–14, doi: 10.1109/TIFS.2019.2959921
  - A. Moerland, C. Freitas, "Artificial Intelligence and Trademark Assessment", in J.-A. Lee, R. Hilty, K. C. Liu (eds.), *Artificial Intelligence and Intellectual Property*, Oxford University Press, 2021, pp. 266-291
  - C. V. Trappey, A. J.C. Trappey, S. C.-C. Lin. "Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies". In: Advanced Engineering Informatics 45 (Aug. 2020), p. 101120, URL: https://linkinghub.elsevier.com/retrieve/pii/S1474034620300896