PROFILE-INT

International Extension of the PROFILE project

Online profiling at the intersection of Laws, Computer Science, and Sociology

2019-2020

From PROFILE to PROFILE-INT

Cominlabs Profile project (2016-2018):

- **Goal:** Analyzing and mitigating the risks of online profiling: building a global perspective at the intersection of law, computer science and sociology
- **Theme 1: the privacy paradox (social science and computer science).** Most people care about their privacy but agree to give private data against online services.
- **Theme 2: profiling regulation (laws and computer science).** the computer and legal control instruments enabling users to understand what the operator does with their data.

Cominlabs International Extension PROFILE-INT (2019-2020):

- **Goal:** Focus on privacy and algorithmic fairness in high-stake decision systems (applications to justice and education).
- **Methodology:** Grounded in real-life use cases, rigorous privacy and fairness models, experimental approach.

Resources

- **Partners:** UR1 (France) and UQÀM (Canada).
- **Cominlabs funding:** 75K€ (1 year PhD student, laptop, travels, internships).
- **Additional funding:** from UQÀM (3+ years PhD student) and UR1 (travels).

Publications

[P1] Tristan Allard, Louis Béziaud, and Sébastien Gambs. Online publication of court records: circumventing the privacy-transparency trade-off. In Workshop on Law & Machine Learning, 2020.

[P2] Tristan Allard, Louis Béziaud, and Sébastien Gambs. Publication of court records: circumventing the privacy-transparency trade-off. In AI Approaches to the Complexity of Legal Systems XI-XII, pages 298–312. Springer, 2020.

[P3] Tristan Allard, Louis Béziaud, and Sébastien Gambs. Simulating socioeconomic-based affirmative action. *Accepted for publication in ReScience, 2022.*

Invited talks and Dissemination

[I1] Louis Béziaud. Publication of court records: circumventing the privacy-transparency trade-off, Webinar on the use of AI in the justice field: Anonymisation and pseudonymisation of judicial decisions, European Commission, 2021.

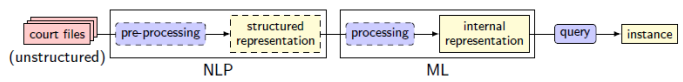
[I2] Tristan Allard, Sébastien Gambs, and Louis Béziaud. La confidentialité différentielle, garante de l'anonymat. Hors-série Pour la Science, n°112, 2021.

[I3] Louis Béziaud. La recherche montre en main, "Intelligence artificielle : par-delà le bien et le mal ?", La Méthode scientifique, 2022.

Axis 1.1: Privacy-Preserving Data Publishing of Court Decision

Opening court decisions :

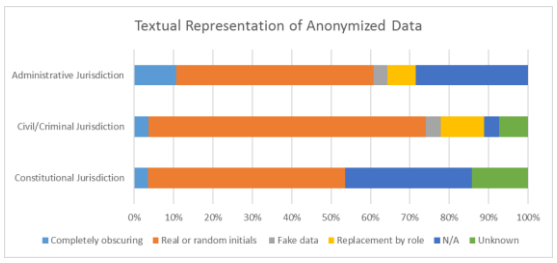
- **Historically:** for transparency (trust, bias inspection, ...) and accessibility (case law).
- **Today's novelty:** from paper-based and in-person court hearings to electronic records to allow **massive, computerized, processing** (Legaltechs!).
- A **current trend** and a **change in scale** (e.g., in France "Arrêté du 28 avril 2021": orders of magnitude more decisions published each year, full scale in 2025)



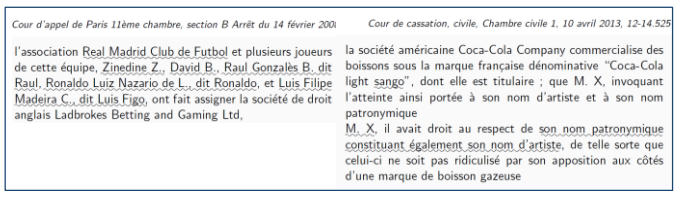
High-level pipeline of court files processing for Legal Techs

Preserving privacy: a mismatch between real world practices and modern approaches

- **Real-world practices:** redaction (common, complex, unsecure)



From: Opijnen, Marc, et al. "On-Line Publication of Court Decisions in the EU : Report of the Policy Group of the Project `Building on the European Case Law Identi_x000c_er'." (2017).



Cour d'appel de Paris 11ème chambre, section B Arrêt du 14 février 200

l'association Real Madrid Club de Futbol et plusieurs joueurs de cette équipe, Zinedine Z., David B., Raul Gonzalès B. dit Raul, Ronaldo Luiz Nazario de L., dit Ronaldo, et Luis Filipe Madeira C., dit Luis Figo, ont fait assigner la société de droit anglais Ladbrokes Betting and Gaming Ltd,

Cour de cassation, civile, Chambre civile 1, 10 avril 2013, 12-14.525

la société américaine Coca-Cola Company commercialise des boissons sous la marque française dénominative "Coca-Cola light sango", dont elle est titulaire ; que M. X, invoquant l'atteinte ainsi portée à son nom d'artiste et à son nom patronymique

M. X, il avait droit au respect de son nom patronymique constituant également son nom d'artiste, de telle sorte que celui-ci ne soit pas ridiculisé par son apposition aux côtés d'une marque de boisson gazeuse

Real-life examples of redaction.

- **Modern privacy approaches :** sound privacy models and algorithms (e.g., differential privacy)

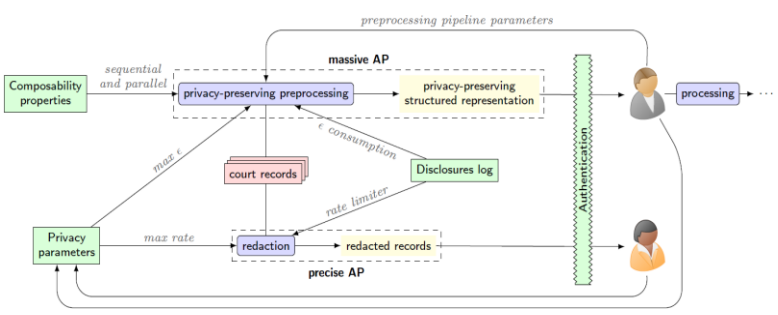
Definition 2. *A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(\mathcal{K})$,*

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{K}(D_2) \in S] \quad (1)$$

From: *Cynthia Dwork. Differential Privacy. ICALP (2) 2006: 1-12.*

⇒ **The real-life approach is necessary** for small scale use cases (e.g., case law) while **the modern approach is sufficient for large scale accesses** (e.g., analytics, LegalTech).

Our proposal [P1, P2, I1] : a multi-modal publication architecture.

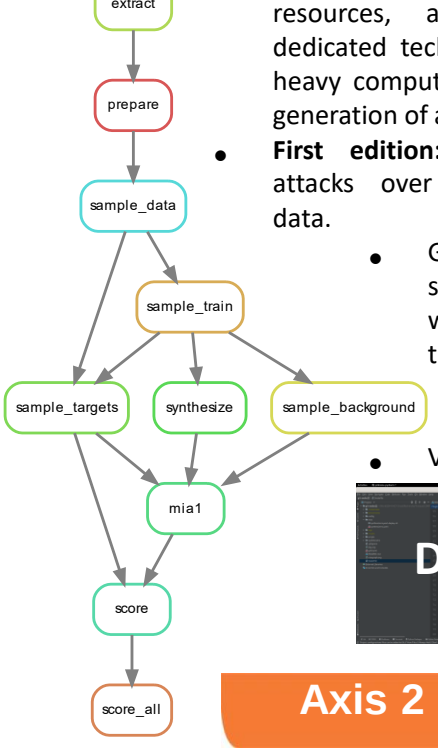


Axis 1.2: Challenging Privacy-Preserving Data Publishing Schemes

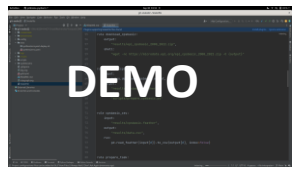
Need for competitions dedicated to attacks over PPDP schemes:

- **Strong competitions exist on the defense side** (e.g., 2018 Nist Differential Privacy Challenge): stimulate research, open source implementations.
- **Weaknesses on the attack side** (e.g., often a neglected phase of existing competitions).

The Snake challenges:



- **Goal:** a **general, secure, and efficient framework** for competitions dedicated to attacks.
- **How:** a dedicated structure (concentrate resources, allocate time), and a dedicated technical infrastructure (e.g., heavy computations offline, automatic generation of attack kits)
- **First edition:** membership inference attacks over synthetically generated data.
  - Given a target and a synthetic dataset, guess whether the target was in the training dataset.
  - Venue to be announced



Axis 2 : Fairness

Context:

- Machine learning models (e.g., classifiers) are increasingly used in **high-stake decisions** (e.g., justice, college enrollment).
- Real-life cases have shown that models might treat individuals unfairly (e.g., COMPAS).
- Fairness metrics aim at defining formally the notion of fairness. But they are numerous, contradicting, and ignore the long term.

Our ongoing work:

- **Goal:** Study systems of fairness metrics on the long run.
- **Challenges:** Data is scarce, performing real-life experiments is infeasible, systems are complex.
- **Current state:** students-colleges simulator implemented [P3], formalization ongoing (e.g., attribute-shift effect, red flag indicator), implementation of metrics ongoing.