



United Nations  
Economic Commission for Africa

# Artificial Intelligence and Tax Fraud Detection

Ad Hoc Expert Group Meeting

Enhancing Domestic Resource Mobilization through Digital Technologies in  
North Africa

---

CHRISTOPHE HURLIN

UNIVERSITY OF ORLEANS AND INSTITUT UNIVERSITAIRE DE FRANCE (IUF)

Rabat, 12 November, 2025

---

## Christophe HURLIN

- Professeur d'économie à l'Université d'Orléans.
- Membre senior de l'Institut Universitaire de France (IUF).
- Responsable du Master Econométrie et Statistique Appliquée (ESA).
- Consultant pour WB, UNECA, Banque de France, ACPR, European Commission, PwC, SG, Mobilize, MSA, etc.
- Co-Responsable de la formation “**Artificial Intelligence and Machine Learning for Economists and Statisticians**”, ECA-HR
- Domaine de recherche : Econométrie, Machine Learning, Risque de crédit.
- Publications dans *Management Science*, *Journal of Financial Econometrics*, *Journal of Business and Economic Statistics*, *European Journal of Operational Research*, *Journal of Financial and Quantitative Analysis*, *Review of Finance*, *Journal of Banking and Finance*, *Journal of Empirical Finance*, etc.

[Google Scholar](#)   [Site personnel](#)

# Intelligence artificielle et détection de la fraude : les grandes questions

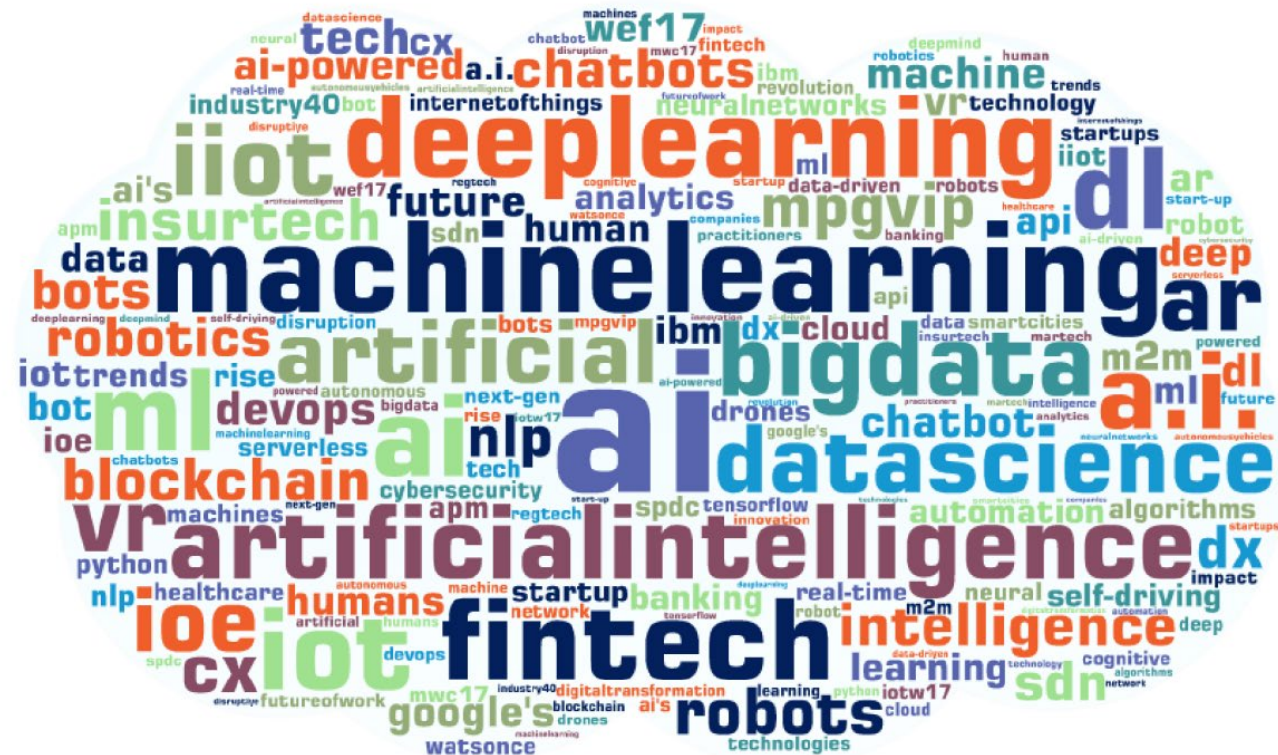


**Combien de recettes fiscales sont perdues chaque année à cause de la fraude dans les pays d'Afrique du Nord ?**

**Combien pourrait-on en récupérer grâce l'Intelligence Artificielle ?**

1. Typologie des IA mobilisés pour détecter la fraude.
2. Quelques cas d'usage observés dans les administrations et les entreprises en France.
  - « Foncier innovant » ou détection par IA des piscines non déclarées.
  - IA et détection de la fraude à la TVA.
  - ML et estimation du manque à gagner sur des fraudes au travail dissimulé.
3. Pourquoi ces approches fonctionnent-elles ?
4. Quelles en sont les principales limites ?

# Clarifions les concepts..



# Intelligence artificielle vs Intelligence prédictive

L'intelligence artificielle (IA) désigne l'ensemble des techniques permettant à une machine d'imiter certaines fonctions cognitives humaines : percevoir, raisonner, apprendre et décider.

## IA prédictive

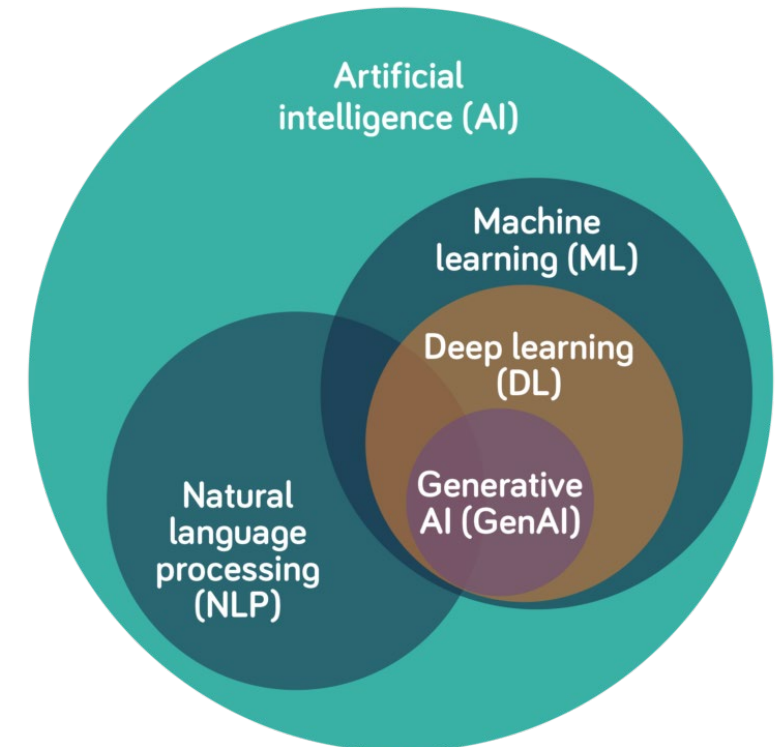
- Objectif : apprendre à partir de données historiques pour prédire un comportement futur (fraude, défaut, évitement fiscal).
- Fondée sur le **Machine Learning** (ML) : modèles statistiques capables d'apprendre à partir des données sans programmation explicite.

## IA générative

- Objectif : créer de nouvelles données (textes, images, code).
- Moins utilisée dans la détection de fraude, mais utile pour la simulation ou la génération de données synthétiques.

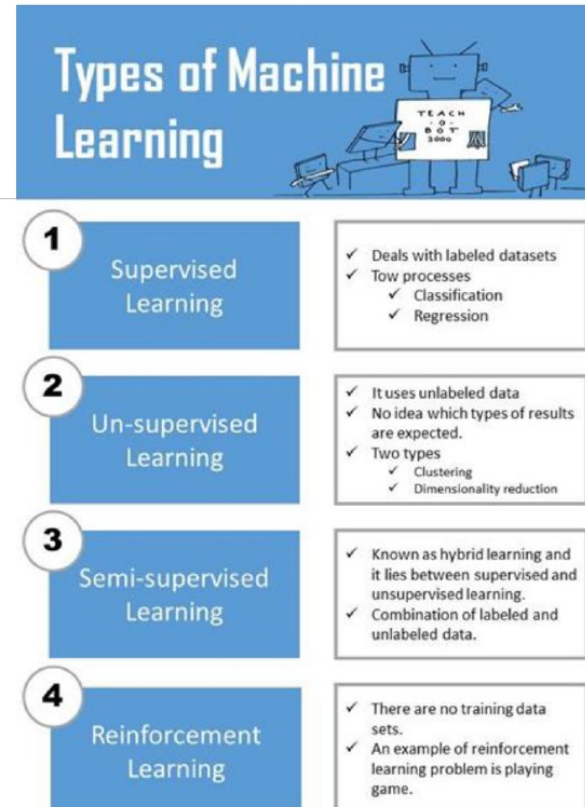
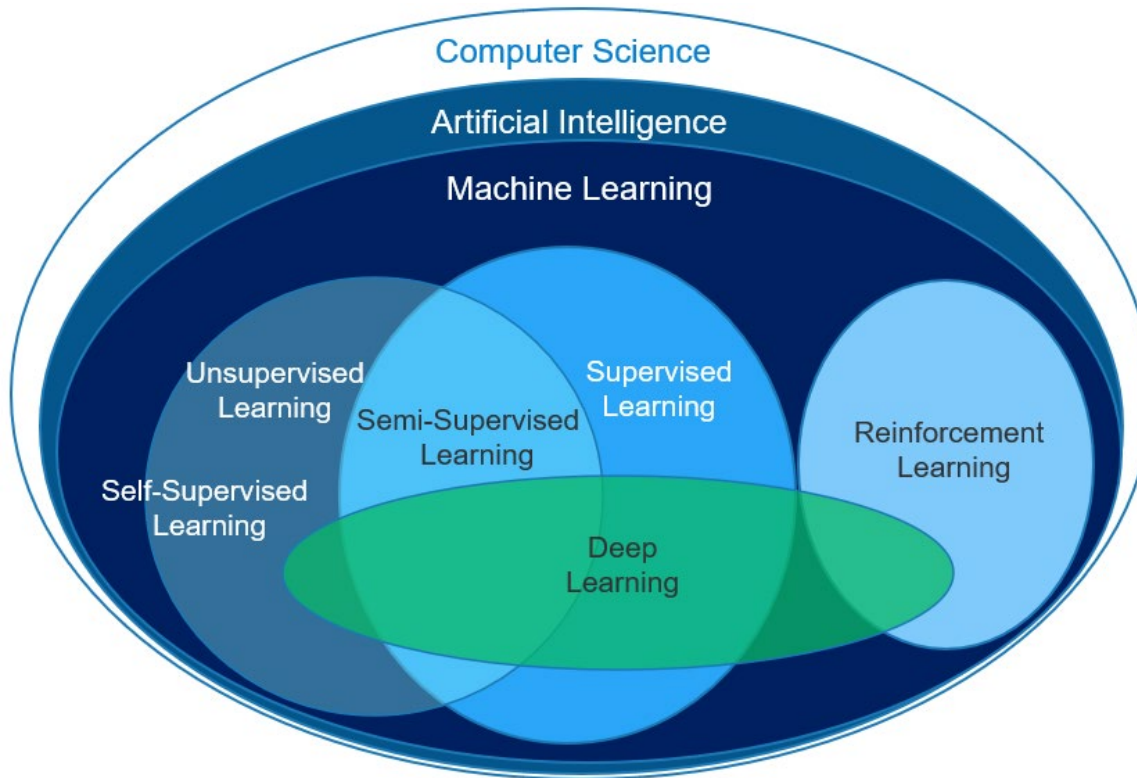
## Exemples

- Un modèle prédictif estime la probabilité de fraude à partir de milliers de déclarations, tandis qu'un modèle génératif pourrait produire des cas simulés de fraude pour tester la robustesse du dispositif.





# Types de Machine Learning



# Apprentissage supervisé et non supervisé pour la détection de la fraude

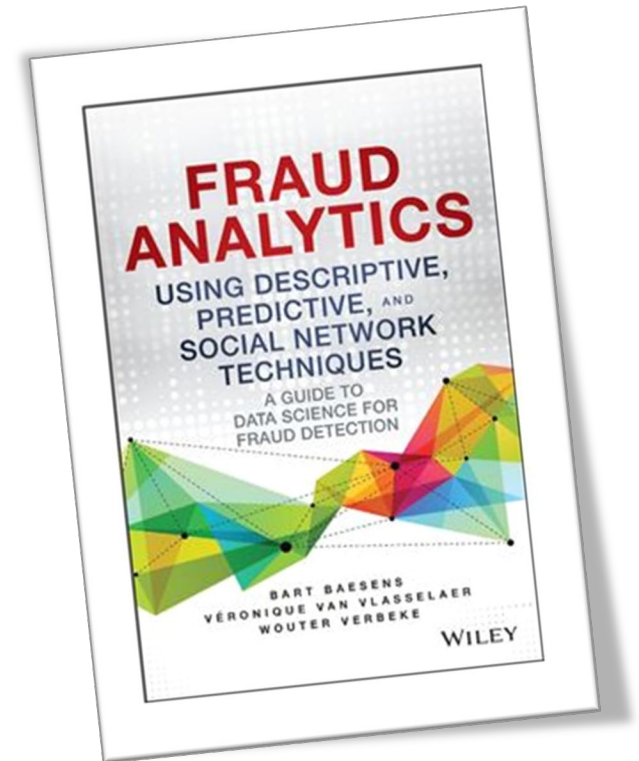
**Objectif :** Identifier automatiquement les comportements frauduleux à partir de données administratives ou financières.

## Apprentissage supervisé

- Données **labellisées** (fraude / non-fraude).
- L'algorithme apprend les régularités entre variables observées et présence de fraude.
- **Méthodes** : Arbres de décision, Random Forest, XGBoost.
- **Avantages** : Bonne performance prédictive, gestion d'un grand nombre de variables.
- **Limites** : Besoin de données labellisées issues d'enquêtes ou de contrôles, biais de sélection, coûts élevés, difficulté à détecter les fraudes nouvelles.

## Apprentissage non supervisé

- Données **non labellisées**.
- Recherche d'**anomalies** ou de **structures inhabituelles** dans les données.
- **Méthodes** : Clustering (K-means, DBSCAN), Isolation Forest, Local Outlier Factor.
- **Exemples** : Détection de réseaux de sociétés fictives (fraude à la TVA), segmentation d'entreprises, anomalies dans les transactions financières.



Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. Wiley.

# Numérisation et IA : leviers de la mobilisation fiscale en Afrique du Nord

---

## Contexte

- Poids élevé de l'économie informelle ( $\approx 40\%$  de l'emploi) et ratios impôts/PIB souvent inférieurs à  $15\%$ .
- Objectif : renforcer la mobilisation des ressources nationales (MRN) pour financer le développement.

## Le numérique, un préalable à l'utilisation de l'IA

- La transformation numérique modernise les systèmes fiscaux, améliore la conformité, et permet de nourrir l'utilisation de l'IA.
- Les systèmes d'identification numérique renforcent la traçabilité et la transparence des transactions.

## Exemples nationaux

- **Égypte** : plateforme de facturation électronique; hausse prévue des recettes fiscales de  $+3\%$  du PIB d'ici 2027.
- **Maroc** : plus de 500 services fiscaux numérisés, facturation électronique, intégration progressive du secteur informel.
- **Algérie** : portail fiscal numérique *Jibayatic* et systèmes d'identification fiscale électronique.
- **Mauritanie** : plateformes de paiement numérique et e-déclaration fiscale ; ratio impôts/PIB passé de  $12,2\%$  à  $14,1\%$  entre 2020 et 2022.
- **Tunisie** : solutions fiscales mobiles et suivi numérique des dépenses publiques.

Source : UNECA, Note conceptuelle, Réunion du groupe d'experts, Rabat, novembre 2025.



## Cas 1 : « Foncier innovant » ou détection par IA des piscines non déclarées



### Taxe foncière : Grâce à l'IA, le fisc a repéré 140.000 piscines non déclarées en 2023

FRAUDE FISCALE - Le nombre de piscines que leurs propriétaires n'avaient pas déclarées à l'administration fiscale et qui ont été détectées par l'intelligence artificielle utilisée par le fisc a été multiplié par sept entre 2022 et 2023



# Cas 1 : « Foncier innovant » ou détection par IA des piscines non déclarées

---



## Méthode

Analyse d'images satellites ou aériennes à haute résolution pour repérer la présence de piscines sur les propriétés.

Traitement par algorithmes d'IA pour classifier la présence ou non d'un bassin (classification supervisée).

Croisement automatique avec les fichiers cadastraux et fonciers pour identifier les piscines qui n'ont pas été déclarées à l'administration fiscale.



## Données mobilisées

Images satellites / photographies aériennes (Google).

Données cadastrales et de valeurs foncières.

Historique des déclarations et taxes foncières des bâtiments résidentiels.

Données géolocalisées des constructions et modifications du bâti.



## Résultats et gains

Plus de **140 000 piscines non déclarées** identifiées en 2023.

Recettes fiscales supplémentaires estimées à **10 millions €** sur neuf départements lors de l'expérimentation en 2022.

Gains de près de **40 millions €** en 2023 suite à la généralisation.

Coût : **23 millions €** sur 2021-2022



## Acteurs impliqués

DGFIP (Direction Générale des Finances Publiques).


Google (technologie de traitement des images et IA).

Capgemini (intégration, prestation de services).

Collectivités locales (bénéficiaires des recettes foncières).

## Cas 2 : un levier stratégique contre la fraude à la TVA

### AI, a weapon against tax fraud

On October 3rd, 2023  4 min reading time



**Christophe Gale**

Head of the Engineering and Digital Innovation Division at the Prime Minister's Office

#### Key takeaways

- Tax fraud is a major issue, accounting for between 4% and 15% of the tax gap in various OECD countries.
- In France, there is a desire to step up the fight against fraud, in particular by using artificial intelligence tools.
- The CISIRH has developed an operational and theoretical framework for comparing different fraud detection algorithms around the world.
- To combat tax fraud effectively, AI and algorithms will not be enough; this fight must be part of a collective and human approach.

Source: Polytechnique Insights, 2023

## Cas 2 : un levier stratégique contre la fraude à la TVA

---



### Contexte

La fraude fiscale représenterait entre **4 % et 15 % du « tax-gap »** dans les pays de l'Organisation de coopération et de développement économiques (OCDE).

En France, la seule fraude à la TVA est estimée à ~ **20-25 milliards €**.

Objectif de l'étude : Déployer un cadre opérationnel permettant de comparer des algorithmes de détection de fraude fiscale développés par différents chercheurs.



### Données mobilisées

Constitution d'une **base de données fictive** (pour raisons de confidentialité) incluant : catégorie socio-professionnelle, revenus, dépenses, patrimoine.

Algorithmes testés : forêt aléatoire (Random Forest), réseaux de neurones, apprentissage supervisé optimisé (AUPRC jusqu'à 0,851)



### Résultats et gains

Un modèle d'IA ne suffit pas : le succès dépend d'un **pilotage humain**, d'une intégration dans les processus de contrôle, et d'une logique d'équipe pluridisciplinaire.

Les données accessibles, la qualité de l'étiquetage (labelling) et la représentativité sont déterminantes pour la performance.

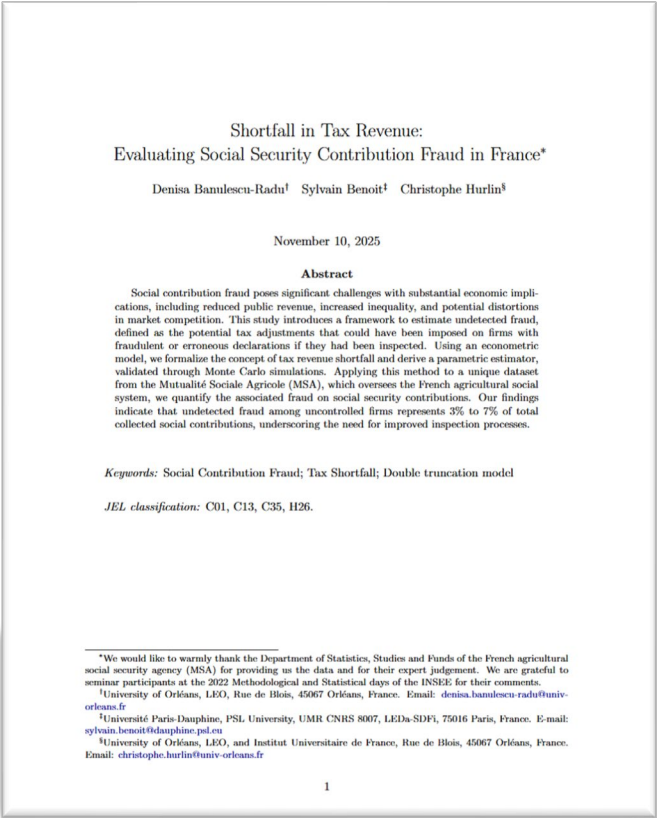


### Acteurs impliqués

Direction générale des finances publiques (DGFi)

Chercheurs de Polytechnique

# Cas 3 : Estimation d'un manque à gagner lié à la fraude sociale



 Download This Paper

[Open PDF in Browser](#)

## Shortfall in Tax Revenue: Evaluating Social Security Contribution Fraud in France

*Université Paris-Dauphine Research Paper No. 5054436*

47 Pages • Posted: 13 Dec 2024

[Sylvain Benoit](#)

Université Paris Dauphine - LEDa-SDFi

[Christophe Hurlin](#)

University of Orléans

[Denisa Banulescu Radu](#)

University of Orléans; Maastricht School of Business and Economics

Date Written: February 11, 2023

### Abstract

Social contribution fraud poses significant challenges with substantial economic implications, including reduced public revenue, increased inequality, and potential distortions in market competition. This study introduces a framework to estimate undetected fraud, defined as the potential tax adjustments that could have been imposed on firms with fraudulent or erroneous declarations if they had been inspected. Using an econometric model, we formalize the concept of tax revenue shortfall and derive a parametric estimator, validated through Monte Carlo simulations. Applying this method to a unique dataset from the Mutualité Sociale Agricole (MSA), which oversees the French agricultural social system, we quantify the associated fraud on social security contributions. Our findings indicate that undetected fraud among uncontrolled firms represents 3% to 7% of total collected social contributions, underscoring the need for improved inspection processes.

**Keywords:** Social Contribution Fraud, Tax Shortfall, Double truncation model JEL classification: C01, C13, C35, H26

**JEL Classification:** C01, C13, C35, H26

## Cas 3 : Estimation d'un manque à gagner lié à la fraude sociale



### Méthode

Modélisation du manque à gagner lié au travail dissimulé en deux temps : pré-traitement des données et modélisation.

Deux étages : estimation de la probabilité de fraude, puis estimation du montant du redressement conditionnellement à la fraude.

Modèles de classification : forêts aléatoires, XGBoost, apprentissage semi-supervisé (self-training, label spreading) pour intégrer les entreprises non contrôlées.

Estimation du montant de redressement par tableaux de contingence par secteur ou arbres de régression.



### Données mobilisées

Données administratives issues des déclarations sociales et des cotisations.

Bases de suivi des contrôles et des redressements.

Appariement des deux bases pour constituer une base de modélisation (entreprises contrôlées) et une base de projection (entreprises non contrôlées).

Ensemble de 40 à 45 variables explicatives, réduites à une vingtaine de variables discrétisées pour la modélisation.



### Résultats et gains

Probabilité de fraude et montant potentiel de redressement estimés pour chaque entreprise et trimestre.

Manque à gagner agrégé obtenu par sommation des redressements potentiels.

Estimation structurée du manque à gagner lié au travail dissimulé.

Comparaison des contributions sectorielles et simulation de l'effet d'un meilleur ciblage des contrôles sur les recettes sociales.



### Acteur impliqué

Mutualité Sociale Agricole (MSA) : organisme de sécurité sociale chargé de la protection sociale et du recouvrement des cotisations des exploitants et salariés agricoles.



## Conclusion : pourquoi l'IA dans la lutte contre la fraude ?

---



Explosion du volume et de la variété des données administratives.

Complexité croissante des montages frauduleux.

Besoin d'outils prédictifs plutôt que purement déclaratifs ou aléatoires.

Capacité d'intégrer des « nouvelles données » non structurées (image, texte, etc.)

**Message clé :** L'IA ne remplace pas le contrôle fiscal, elle oriente et priorise mieux les enquêtes.

# Artificial Intelligence and Machine Learning for Economists and Statisticians, ECA-HR

---



# Contact

---

Christophe HURLIN  
University of Orleans  
University Institute of France  
[christophe.hurlin@univ-orleans.fr](mailto:christophe.hurlin@univ-orleans.fr)



**institut  
universitaire  
de France**