



**Modèles et apprentissages en sciences humaines et sociales**

MASHS 2024

14 oct. 2024 PARIS (France)



# Recueil des résumés

Modèles et apprentissages en sciences humaines et sociales

à Paris, le 14 octobre 2024

# Table des matières

<b>Sessions invités</b>	<b>3</b>
Utilisation des Modèles de Langage de Grande Taille (LLM) pour le Raisonnement	4
Traitement automatique de textes par apprentissage automatique : de la détection de position ou de thèmes à la recherche de sensorialité dans des collections de documents . . . . .	5
Banquier d'affaires, tout un parcours ! . . . . .	6
Biais Humain, Biais Machine: les LLMs génératifs sont-ils WEIRD? . . . . .	7
Maintenir la confiance dans une réalité synthétique . . . . .	8
<b>Sessions posters</b>	<b>9</b>
Convolution, Self-Attention et discours politique. Que faire des couches intermédiaires des réseaux de neurones profonds ?, Laurent Vanni <i>et al.</i> . . . . .	10
Étude des déclarations de violences de genre sur Twitter en France dans le cadre de #MeToo, Clara Le Gallic-Ach . . . . .	11
Evaluation of BERT-Based Models on Patient Data from French Social Media, Emma Le Priol <i>et al.</i> . . . . .	12
Innovation et Crédit d'Impôt Recherche : Analyse Textuelle des Narratifs d'Entreprises (2017-2022), Rim Bahroun <i>et al.</i> . . . . .	14
Interpréter une variable latente avec un modèle de langage, une application du package NaileR à la compréhension des systèmes alimentaires durables, Nel Hervé <i>et al.</i> . . . . .	15
Modélisation explicative des achats responsables des clients d'une enseigne de distribution : le rôle des valeurs du consommateur, Stéphanie Leprieur <i>et al.</i> . . . . .	16
MRS Unplugged: Effects of explaining algorithmic recommendations on music discovery, an experimental approach, Julien M'barki <i>et al.</i> . . . . .	18
Outils d'IA et enseignement supérieur : quels recours aux IA dans les universités françaises ?, Stéphanie Marty . . . . .	19
Par-delà la prestidigitation : une comparaison entre analyse humaine et analyse par IA d'entretiens en SEF, Marlène Martin <i>et al.</i> . . . . .	20
Post-traitement de la transcription automatique des manuscrits latins médiévaux grâce à un tokeniseur spécialisé et à des modèles de transformers, Svetlana Yatsyk	21
Représenter le champ des économistes en France dans les années 2010 avec une carte auto-organisatrice, Adrien Rougier . . . . .	22
Trajectoires professionnelles de docteur.es au sein de l'enseignement supérieur : l'intérêt des algorithmes de classification de prénom, Matthieu Cisel <i>et al.</i> . . . . .	23
<b>Liste des auteurs</b>	<b>23</b>

## Sessions invitées

# Utilisation des Modèles de Langage de Grande Taille (LLM) pour le Raisonnement

Joseph Rynkiewicz <sup>1</sup>

<sup>1</sup> SAMM, Université Paris 1 Panthéon-Sorbonne, France

Les modèles de langage de grande taille (LLM) tels que GPT-4 ont révolutionné le domaine de l'intelligence artificielle en démontrant une capacité impressionnante à effectuer des tâches de raisonnement. Cependant, bien que les LLM puissent simuler un raisonnement semblable à celui des humains, leur fonctionnement repose principalement sur des corrélations statistiques plutôt que sur une compréhension profonde du contexte ou des concepts sous-jacents et les LLM peuvent parfois produire des conclusions logiquement incorrectes ou incohérentes. Le but de cet exposé est d'étudier les méthodes disponibles pour réduire au maximum les erreurs de raisonnement, plus particulièrement pour de "petits" modèles qui peuvent être plus facilement spécialisés à des problématiques spécifiques.

# Traitement automatique de textes par apprentissage automatique : de la détection de position ou de thèmes à la recherche de sensorialité dans des collections de documents

Christine Largeron <sup>1</sup>

<sup>1</sup> Laboratoire Hubert Curien, Université Jean Monnet, France

Si la science des données s'est d'abord focalisée sur des données de type numérique, elle s'est intéressée ensuite aux données textuelles, au fur à mesure de leur production mais surtout de leur stockage sous forme numérique et, aujourd'hui, les progrès réalisés en matière de traitement automatique de documents figurent parmi les avancées les plus importantes réalisées récemment en IA, notamment grâce aux méthodes de fouille de données et d'apprentissage automatique. Au cours de cette présentation, nous l'illustrerons à travers des tâches variées telles que la détection de position d'un texte ou d'un auteur vis-à-vis d'un sujet donné (stance detection), l'extraction d'information sensorielle ou encore la recherche de thèmes (topic detection) dans des collections de documents. Nous verrons aussi que si l'apprentissage profond, via notamment les grands modèles de langage (LLM), permet de résoudre très efficacement ces tâches, il présente aussi des limites sérieuses comme les biais ou les coûts induits par l'apprentissage des modèles ou encore l'explicabilité et l'interprétabilité des résultats produits.

# Banquier d'affaires, tout un parcours !

Arthur Miguel <sup>1</sup>, Madalina Olteanu <sup>2</sup>

<sup>1</sup> IRISSO, Université Paris Dauphine PSL, Paris, France

<sup>2</sup> CEREMADE, Université Paris Dauphine PSL, Paris, France

Durant les quarante dernières années, les banques d'affaires françaises ont connu de profonds changements institutionnels, qui ont transformé les parcours et les métiers. Lors de cette présentation nous nous pencherons sur le cas de deux établissements, Rothschild & Cie et Lazard Frères, afin de mettre en évidence et de discuter ces transformations. Plus précisément, nous présenterons les résultats d'une étude prosopographique exhaustive, menant à la constitution d'une base de données riche et complète. Nous verrons ensuite si et comment l'analyse séquentielle – via appariement optimal ou modèles probabilistes – permet d'isoler des typologies spécifiques et caractériser les parcours.

# Biais Humain, Biais Machine: les LLMs génératifs sont-ils WEIRD?

Étienne Ollion <sup>1</sup>

<sup>1</sup> CREST – ENSAE, CNRS, École Polytechnique, Institut Polytechnique de Paris – France

L’irruption de chatGPT fin 2022 a donné une visibilité inédite aux Large Language Models (LLMs) génératifs. Depuis, le recours à ces modèles a suscité nombre d’interrogations. L’une d’entre elles a trait aux opinions implicites et aux biais que pourrait avoir ces modèles. Entraînés principalement sur des corpus de textes issus de certains pays et de certains groupes sociaux, ils risqueraient de n’être représentatifs que d’une partie de la population. Plus précisément encore, ils auraient de forte chance d’être WEIRD (Western, Educated, Industrialized, Rich and Democratic). L’interrogation, importante, est devenue essentielle à mesure que ces modèles étaient employés pour remplacer des humains – dans des enquêtes d’opinions, dans des expériences, pour de la connaissance client. Cette présentation présente les résultats d’une enquête où on a sondé des LLMs sur une batterie de questions, avant de les comparer à celles de différentes populations. Elle explore les biais humains, et ce qu’on appelle les biais machine des LLMs, et on propose de réfléchir aux implications de ce résultat.

# Maintenir la confiance dans une réalité synthétique

Florian Barbaro <sup>1</sup>

<sup>1</sup> UncovAI – France

Depuis l'avènement des modèles génératifs d'intelligence artificielle et leur large diffusion auprès du grand public, nous avons constaté de nombreux cas d'utilisation positifs pour nos entreprises et nos sociétés. Cependant, comme toute technologie, elle peut être utilisée à mauvais escient pour influencer les choix des citoyens, par exemple. Dans ce contexte, nous aborderons l'importance de la détection de contenus générés. Puis, nous présenterons la manière dont nous détectons les contenus générés par Intelligence Artificielle chez UncovAI. Enfin, nous terminerons sur des exemples de mauvaises utilisations de modèles génératifs.

# Sessions posters

# Convolution, Self-Attention et discours politique. Que faire des couches intermédiaires des réseaux de neurones profonds ?

Laurent Vanni \* <sup>1</sup>, Damon Mayaffre \* † <sup>2</sup>

<sup>1</sup> Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – France

<sup>2</sup> Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – Centre National de la Recherche Scientifique, Université Côte d’Azur – Laboratoire BCL - UMR 7320MSHS Sud-Est Pôle Universitaire Saint Jean d’Angely Bâtiment SJA 3 24 avenue des Diabls bleus, 06357 Nice Cedex 4, France

Les grands modèles de langage (LLM) ont repoussé ces dernières années les limites du Traitement Automatique des Langues (TAL). Les architectures de type Transformer entraînées sur la quasi-totalité des productions écrites des humanités semblent aujourd’hui prêtes à répondre à toutes les questions envisageables des internautes les plus chevronnés. Cependant, si les sorties machines donnent l’illusion d’une compréhension des requêtes des utilisateurs, l’interprétation des résultats ne peut et ne doit être considérée par les SHS qu’au regard des programmes informatiques utilisés. Les travaux que nous proposons ont pour but d’éclairer la boîte noire de l’IA en utilisant l’ensemble de la chaîne de traitement des réseaux de neurones profonds, du corpus d’apprentissage à la prédiction (classification ou génération de textes). Chaque couche intermédiaire (ou couche cachée) est mobilisée pour extraire l’information apprise par les modèles qui conduit à une prédiction. Les représentations statistiques classiques des mots et/ou des textes se dotent ainsi de plus-values heuristiques nouvelles qui mêlent la vectorisation à grande échelle de l’information (*Embedding*), la détection locale d’association (syntagme) par *convolution* (Vanni et al. 2018) et la détection d’associations plus complexes et à distance (paradigme) par *Self-Attention* (Vanni et al. 2024). Appliqués à l’analyse de discours politique, ces travaux interdisciplinaires interrogent d’une part les linguistes ou les politologues sur le sens des textes et d’autre part les informaticiens sur le sens des hyperparamètres du *deep learning*. Avec une approche dirigée par le corpus – *corpus-driven* – le retour au texte permet une interprétation des marqueurs repérés par le modèle et conduit à une herméneutique des sorties logicielles (Rastier 2011). La prédiction de l’I.A. n’est alors plus une finalité mais un point d’étape nécessaire pour une exploration nouvelle des corpus numériques par les SHS.

**Mots-Clés:** Deep Learning, Linguistique, Texte, Convolution, Self, Attention, Transformer, Explicabilité

---

\*Intervenant

†Auteur correspondant: damon.mayaffre@univ-cotedazur.fr

# Étude des déclarations de violences de genre sur Twitter en France dans le cadre de #MeToo

Clara Le Gallic-Ach \* <sup>1</sup>

<sup>1</sup> Sciences Po (Sciences Po) – Sciences Po – 27, rue Saint-Guillaume - 75337 Paris cedex 07, France

L'étude présentée se concentre sur les déclarations de violences de genre sur Twitter en France dans le cadre de #MeToo, entre octobre 2017 et décembre 2022. Plus de 300 000 tweets ont été analysés, provenant de deux bases de données : une collecte en continu et une collecte rétrospective fin 2022. Les hashtags ciblés incluent #Metoo, #MeTooInceste, #MeTooGay, #IWas, et #BalanceTonPorc.

Les enjeux méthodologiques principaux étaient l'identification des témoignages de violences parmi les tweets collectés et l'analyse détaillée de ces témoignages en termes de contenu et d'intention. Deux modèles de classification automatique basés sur CamemBERT (Martin et al, 2019) ont été développés. Utilisant une approche supervisée avec des annotations manuelles, les modèles ont été entraînés pour des classifications à 4 et 7 classes. Le second modèle, qui classe les témoignages (identifiés par le premier modèle) selon le contexte des violences déclarées (inceste, couple, espaces publics, école, milieu professionnel, connaissances, autres), applique une classification systématique par champs lexicaux pour améliorer la précision. Les modèles atteignent des taux d'accuracy de 87% et 95%.

L'analyse révèle plusieurs points clés. D'abord, il y a une sur-déclaration des violences au travail et à l'école et une sous-déclaration des violences conjugales par rapport aux enquêtes nationales (Enquête Virage, Ined, 2015), mettant en évidence des disparités dans les témoignages #MeToo. Ensuite, une analyse des métadonnées des comptes et tweets (date de création du compte, nombre d'abonnés, d'abonnements, de tweets publiés, nombre de likes, retweets, etc.) a permis de mesurer l'anonymat, la visibilité et l'interaction des utilisateurs. La comparaison entre tweets issus de la collecte journalière et ceux de la collecte rétrospective a permis d'étudier la suppression des tweets. Les tweets identifiés comme témoignages sont surreprésentés parmi les tweets supprimés, anonymes et peu visibles, soulignant l'usage spécifique de Twitter pour témoigner par rapport aux autres types de contenus.

**Mots-Clés:** violences de genre, twitter, #metoo, analyse du langage, témoignages

---

\*Intervenant

# Evaluation of BERT-Based Models on Patient Data from French Social Media

Emma Le Priol \*<sup>1,2</sup>, Manissa Talmatkadi<sup>3</sup>, Stéphane Schuck<sup>3</sup>, Nathalie Texier<sup>3</sup>, Anita Burgun<sup>1,2,4,5</sup>

<sup>1</sup> Université Sorbonne Paris Cité (USPC) – Université Paris Cité – Université Sorbonne Paris Cité | 100-104, avenue de France | 75013 Paris, France

<sup>2</sup> Health data- and model- driven Knowledge Acquisition (HeKA) – Inserm U1138, Centre de Recherche des Cordeliers, L’Institut National de Recherche en Informatique et en Automatique (INRIA) – France

<sup>3</sup> Kap Code – Kap Code – France

<sup>4</sup> Imagine - Institut des maladies génétiques (IHU) (Imagine - U1163) – Imagine – IHU Imagine, 156 rue de Vaugirard, 75015 PARIS et 24 Boulevard du Montparnasse, 75015 Paris, France

<sup>5</sup> Plateforme Bioinformatique [SFR Necker] (BIP-D) – AP-HP Hôpital Necker - Enfants Malades [Paris] – Hôpital Necker - Enfants Malades - Université Paris Descartes 149 rue de Sèvres 75743 PARIS cedex 15 Tour Lavoisier - 7ème étage, France

## Background

Our study aimed to extract new insights on Developmental and Epileptic Encephalopathies (DEE), a group of rare and severe diseases, from French social media messages by evaluating three BERT-based models on a Named Entity Recognition (NER) task. This task involved identifying phenotypes and treatments from social media posts. The pre-trained models evaluated were CamemBERT, CamemBERT-bio, and DrBERT.

## Materials & Methods

They were trained on a dataset containing social media messages discussing DEE and more common diseases, in which treatments and phenotypes entities were annotated. The precision, recall and F1-score were calculated in strict mode for each entity, and for micro, macro and weighted averages. They were averaged on 10 runs.

## Results

Our results revealed that CamemBERT and CamemBERT-bio performed similarly, slightly outperforming DrBERT. Both models showed a balanced performance in precision and recall, though their performance was lower on social media data compared to more "traditional" health datasets.

## Discussion

Our study highlighted the promise of deep learning methods, particularly transformer-based models, in analyzing medical content from social media. However, the limitations of this study include a narrow focus on NER performance and a dataset-specific evaluation, which calls for further research to assess the models on larger and more diverse datasets. Another limitation is the annotation process, which was not the same for the DEE messages and the others, and with a low inter-annotator agreement. Other pre-trained models (SocBERT or BERTweet) could also be explored, as well as generative models. Being able to extract phenotypes and treatment entities from social media messages could significantly enhance the understanding of the challenges faced by patients and caregivers, especially in the context of DEE. Future research will include a new annotation process, as well as exploration of the use other pre-trained and of generative

---

\*Intervenant

models. The long-term objective being to analyze posts from a French patients' organization Facebook group.

**Mots-Clés:** rare disease, developmental and epileptic encephalopathies, bert, transformers, deep learning, health, social media

# Innovation et Crédit d'Impôt Recherche : Analyse Textuelle des Narratifs d'Entreprises (2017-2022)

Rim Bahroun <sup>\*† 1</sup>, Nadine Levratto <sup>\* ‡ 1</sup>, Mathieu Bernard <sup>1</sup>

<sup>1</sup> Economix UMR 235, UPL, Univ Paris Nanterre, CNRS – Economix UMR 235, UPL, Univ Paris Nanterre, CNRS – France

Le Crédit d'Impôt Recherche (CIR) en France est un dispositif fiscal conçu pour encourager les investissements en recherche et développement. Depuis sa refonte en 2008, il représente un coût considérable pour les finances publiques, mais son efficacité reste débattue. Dans cette recherche, nous proposons de dépasser les limites des indicateurs de performances à l'innovation (Dépenses de RD et brevets) en extrayant des signaux pertinents à partir des narratifs d'entreprises bénéficiaires du CIR. Ces derniers nous permettent d'analyser l'évolution des déclarations sur l'innovation en fonction du temps et du destinataire (grand public ou financier). Pour ce faire, nous appliquons des techniques de traitement automatique du texte sur deux corpus : les rapports financiers des entreprises et les articles de journaux spécialisés, couvrant les années 2017 à 2022. Nous introduisons une nouvelle méthode d'incorporation des informations textuelles qui combine le comptage des termes avec des techniques sémantiques basées sur des modèles de langage pré-entraînés, tels que BERT. Cette méthode permet de calculer, pour chaque entreprise du panel, un " indice d'innovation " en fonction de l'année et du type de discours. Les résultats de l'analyse textuelle seront comparés, dans un futur travail, à une typologie obtenue à partir d'une analyse quantitative basée sur des indicateurs traditionnels de l'innovation, tels que les dépenses en R&D ou le nombre de brevets, afin de produire une vision complète de la manière dont les entreprises utilisent le CIR et des effets de ce dispositif sur l'innovation au sein du tissu productif.

**Mots-Clés:** Crédit Impôt Recherche, indice d'innovation, analyse textuelle, technique sémantique, modèles de langage préentraînés

---

\*Intervenant

†Auteur correspondant: rim.bahroun@cnrs.fr

‡Auteur correspondant: nadine.levratto@cnrs.fr

# Interpréter une variable latente avec un modèle de langage, une application du package NaileR à la compréhension des systèmes alimentaires durables

Nel Hervé \* <sup>1</sup>, Sylvie Thoby \* <sup>1</sup>, Sébastien Lê \* <sup>2,3</sup>

<sup>1</sup> L’Institut Agro Rennes-Angers – L’Institut Agro Agrocampus Ouest – 65, rue de Saint-Brieuc, 35000 Rennes Cedex, France

<sup>2</sup> Institut de recherche mathématique de Rennes (IRMAR) – CNRS – France

<sup>3</sup> L’Institut Agro Rennes-Angers – L’Institut Agro Agrocampus Ouest – 65, rue de Saint-Brieuc, 35000 Rennes Cedex, France

La Q-méthode ou Q-méthodologie est une méthode de recherche utilisée en sciences humaines et sociales pour étudier des concepts complexes à travers les points de vue subjectifs d’individus interrogés. Pour ce faire, on identifie un ensemble de stimuli, généralement des assertions, que l’on propose aux interrogés de trier dans des groupes définis a priori par leur sens et par le nombre d’assertions qu’ils peuvent contenir. Dans notre recherche, nous nous intéressons aux systèmes alimentaires durables et en particulier à la perception que l’on peut en avoir, suivant que l’on privilégie la dimension économique, sociale, environnementale...

Pour comprendre cette variabilité de perception, nous avons soumis un ensemble de 45 assertions, structurées selon 5 dimensions, à plus de 500 personnes : un système alimentaire qui répondrait à mes valeurs et besoins, sans compromettre la capacité des générations futures à répondre aux leurs, vendrait beaucoup plus cher les produits qui ne sont pas de saison, imposerait une redistribution équitable des bénéfices tout au long de la chaîne alimentaire, limiterait la viande à deux jours par semaine dans la restauration collective,...

Les données recueillies ont permis d’obtenir une typologie d’interrogés, telle que deux personnes interrogées appartiennent à la même classe si elles partagent la même perception d’un système alimentaire durable. L’interprétation des classes issues de la typologie a été faite “manuellement” d’une part, à l’aide du package NaileR d’autre part. NaileR est un package R conçu pour interpréter des variables latentes continues ou catégorielles : typiquement, des dimensions issues d’une analyse multivariée exploratoire, ou une variable de classe issue d’un algorithme de classification non supervisée. Dans cette recherche, nous montrerons comment tester la similarité entre l’interprétation “manuelle” et les interprétations fournies par un modèle de langage tel que Llama3. Nous proposerons également une représentation graphique de ces différentes interprétations.

**Mots-Clés:** Llama3, Q méthode, système alimentaire, similarité textuelle

---

\*Intervenant

# Modélisation explicative des achats responsables des clients d'une enseigne de distribution : le rôle des valeurs du consommateur

Stéphanie Leprieur \*<sup>†</sup> <sup>1</sup>, Stéphane Magne \*<sup>‡</sup> <sup>2</sup>, Patricia Coutelle \*<sup>§</sup> <sup>1</sup>

<sup>1</sup> Institut d'Administration des Entreprises (IAE) - Tours (VALLOREM) – VALLOREM (E.A. 6296)  
IAE de l'Université de Tours – France

<sup>2</sup> PRISM Sorbonne Paris 1 et LAREQUOI Université Versailles Saint-Quentin – Université Paris 1 -  
Panthéon-Sorbonne – France

34% des émissions de gaz à effet de serre, imputables aux activités humaines, sont liés à l'alimentation (Crippa et al., 2021). Les entreprises agroalimentaires, conscientes de l'impact de leur activité sur les enjeux écologiques, sociologiques et économiques actuels, innovent avec leurs parties prenantes pour développer des gammes de produits alimentaires " durables " qui se définissent au travers de modes de production, de transformation, d'emballage, de distribution voire d'usage, de préparation et de tri plus durables (Innocent et al., 2023). La gamme de légumes Bio Française Équitable de Picard (BFE), terrain d'étude de cette recherche, en est un exemple concret.

Ce sujet de l'alimentation durable mobilise l'attention des chercheurs qui visent, par exemple, à comprendre les pratiques alimentaires des consommateurs ayant un impact sur la transition écologique (Verain et al., 2015 ; Whittall et al., 2023 ; Innocent et al., 2023) : types de produits achetés, lieux d'achats, manière de préparer et de cuisiner des aliments, de les conserver et de gérer les déchets produits.

Aucune recherche n'a, à ce jour, fait le lien entre les valeurs individuelles et les achats réels du consommateur. Pourtant les valeurs constituent " *des buts désirables, trans-situationnels, variant en importance, qui servent de principes guidant les vies des gens* " (Schwartz, 1996, p. 2). Elles influencent les perceptions, choix, attitudes et comportements des individus et des groupes (Rokeach, 1973 ; Schwartz, 1996). Cette recherche propose de modéliser ce lien entre les valeurs, issues de l'inventaire de Schwartz (1992), et les achats réels de la gamme de haricots verts BFE (résultat d'un processus de collaboration innovant et tripartite entre agriculteurs, Picard, industriels). Cette modélisation permettra de comprendre quelles sont les valeurs personnelles qui agissent (ou pas) sur les comportements d'achat de produits alimentaires durables.

La méthodologie utilisée consiste à traiter les données issues d'un questionnaire administré aux clients du programme de fidélité de Picard associées aux données comportementales CRM (*Customer Relationship Management*). Seront réalisées des analyses typologiques, de variance et régressions multiples pour lier valeurs et achats réels de la gamme BFE. L'impact prédictif des

---

\*Intervenant

<sup>†</sup> Auteur correspondant: stephanie.leprieur@etu.univ-tours.fr

<sup>‡</sup> Auteur correspondant: magne.stephane@yahoo.fr

<sup>§</sup> Auteur correspondant: patricia.coutelle@univ-tours.fr

valeurs individuelles sera mesuré sur les achats finaux en se fondant sur des variables comportementales objectives (quantités achetées, montant en euros...)  
Cette recherche permettra de mieux comprendre les déterminants à l'achat de produits équitables (de Ferran, 2006) ou de produits biologiques (Sirieix et al., 2006) et de manière plus globale encore les ressorts des achats de produits alimentaires durables.

**Mots-Clés:** valeurs individuelles, valeur durable, marque, alimentation, produit durable, modélisation prédictive

# MRS Unplugged: Effects of explaining algorithmic recommendations on music discovery, an experimental approach

Julien M'barki <sup>\*†</sup> <sup>1</sup>, Mehdi Louafi <sup>\*</sup>

2

<sup>1</sup> Centre d'économie de la Sorbonne – Paris 1, Pantheon Sorbonne, Centre d'Economie de la Sorbonne  
UMR 8174 – France

<sup>2</sup> Université d'Orléans – Laboratoire d'Economie d'Orléans (LEO) – France

Music streaming services make massive use of algorithms in their music recommender systems (MRS) to guide users to tracks they are likely to enjoy. However, the black-box nature of these algorithms makes them difficult for users to understand, both in terms of how they work and the music they predict. The field of explainable AI (XAI), and in particular its "explanation" side, has emerged to make the uses of AI (including MRSs) more comprehensible to users. This paper aims to observe, using an experimental method, whether the explanation of an MRS algorithm induces a change in discovery behavior on music streaming services. In a theoretical framework, we model two types of users' discovery behaviors, namely "study" and "browse" behaviors. We then test in the lab the explanation effects on these behaviors by explaining a simplified MRS, based on the music genre criterion. Our experimental design consists in a musical listening session lasting 20 minutes, during which participants have to listen to a sample of tracks, generated using our algorithm and based on a preference questionnaire, while being able to "like", "dislike" or "skip" them. Our main outcome measurement is the relative time spent listening to the tracks. Beyond a certain threshold, the track is considered as "studied". This measure will enable us to observe any differences in discovery behaviors between our control groups, who receive no explanation of the MRS, and our treatment group, who receive an explanation of how the tracks are recommended. Temporary results show that explaining recommendations increases the time of listening, reinforcing the study behavior as we predict in the model. Moreover, this effect is wider depending on the treatment level of participants. Additional measures such as the jumps (navigating inside the tracks by "jumping" from the beginning to the middle of the track) provide additional insight on the effect of explanations, showing that people tend to jump less on the tracks.

**Mots-Clés:** Explainable AI, Music recommender systems, experimental economics, music industry

---

\*Intervenant

†Auteur correspondant: julien.m-barki@univ-paris1.fr

# Outils d'IA et enseignement supérieur : quels recours aux IA dans les universités françaises ?

Stéphanie Marty \* <sup>1</sup>

<sup>1</sup> LERASS-CERIC-Université Paul Valéry-Montpellier – LERASS-CERIC-Université Paul Valéry-Montpellier – France

Durant les dernières décennies, l'arrivée des outils d'IA (Intelligence Artificielle) a introduit de nouvelles opportunités dans l'enseignement universitaire. En effet, l'arrivée de ces nouveaux outils - et avec eux la possibilité de traiter des données particulièrement massives - a bousculé différentes sphères structurant les universités, en reconfigurant notamment les process administratifs et pédagogiques.

Dans le présent propos, nous souhaitons explorer et interroger ces transformations. Afin d'étayer notre propos, nous prenons appui sur un corpus documentaire, centré sur six universités françaises ayant recouru à des outils d'IA pour traiter des données massives. *In fine*, l'analyse de ce corpus documentaire nous permet de faire émerger différents recours – administratif et pédagogique – aux outils d'IA dans l'enseignement supérieur.

En effet, sur le plan administratif, ces outils sont aujourd'hui mobilisés pour simplifier, voire automatiser, certains process et certaines tâches (inscriptions des étudiants, transmission de documents, réponses aux étudiants *via* des *chatbots*, automatisation des emplois du temps...).

Sur le plan pédagogique, les outils d'IA sont utilisés pour cerner et connaître les étudiants (parcours, forces, difficultés...), mais également pour concevoir des ingénieries et des supports pédagogiques engageants (innovants, ludiques), accessibles et inclusifs, adaptés notamment aux besoins, aux difficultés et aux handicaps des apprenants (transcription automatisée des cours, production de sous-titres, traduction en langue des signes...). En outre, les outils d'IA sont utilisés pour faire évoluer, d'une part, l'évaluation des étudiants (conception d'évaluations 'sur mesure', corrections automatiques, analyse des résultats d'examens, retours et commentaires - personnalisés et instantanés - aux étudiants, identification des étudiants en situation de décrochage, détection de cas de fraudes aux examens...), et, d'autre part, l'évaluation des enseignements (collecte de *feedbacks* auprès des étudiants et analyse de ces *feedbacks* pour améliorer les dispositifs pédagogiques...).

**Mots-Clés:** Intelligence Artificielle, universités françaises, enseignement supérieur, pédagogie universitaire

---

\*Intervenant

# Par-delà la prestidigitation : une comparaison entre analyse humaine et analyse par IA d'entretiens en SEF

Marlène Martin \* <sup>1</sup>, Romain Jouin \* † <sup>2</sup>

<sup>1</sup> Université de Caen – Université de Caen, laboratoire cirnef – France

<sup>2</sup> Google France – Google France – 8 Rue de Londres, 75009 Paris, France

Le développement spectaculaire des possibilités de l'IA ces toutes dernières années ouvre des perspectives inédites pour le traitement de tous les types de données collectées par les chercheurs et chercheuses - recueil, mise en forme, analyses, interprétations. Les résultats obtenus suscitent souvent fascination, voire sidération, d'autant plus que la manière dont ils sont générés est peu ou mal comprise au niveau conceptuel et technique. Se pose de plus la question récurrente à chaque avancée technologique : en tant qu'humain-e-s, servirons-nous encore à quelque chose, et si oui, à quoi ?

Sur le plan théorique, ce poster rappelle la manière dont sont construits les outils d'IA, et comment une question posée est traitée.

Sur le plan empirique, il propose un exemple de traitement comparé de retranscriptions d'entretiens de recherche, de type compréhensif, menés dans le cadre d'une thèse en science de l'éducation. Dans un premier temps nous avons fait une analyse de contenu thématique assistée par Nvivo, dans un second temps une analyse via le LLM Gemini de Google. Nous présentons cette démarche, notamment les tests de différents prompts et différents degrés de température, de manière à modifier les réponses apportées par l'IA.

Nous identifions les similarités et les différences entre l'analyse humaine et l'analyse par IA et les questions que nous nous sommes posées quant à la nature, la justesse et la scientificité de ces réponses. Finalement, des perspectives questionnant les méthodologies actuelles de recherche et les usages possibles de ces outils, envisagés davantage comme des supports et des points de départ que comme des substitutions, sont proposées.

**Mots-Clés:** Entretiens, analyse de contenu, IA, LLM

---

\*Intervenant

†Auteur correspondant: romain.jouin@gmail.com

# Post-traitement de la transcription automatique des manuscrits latins médiévaux grâce à un tokeniseur spécialisé et à des modèles de transformers

Svetlana Yatsyk \* <sup>1</sup>

<sup>1</sup> CIHAM – CIHAM (UMR 5648) – France

Ce projet établit un pont entre l’informatique et les sciences humaines pour aborder les défis complexes associés à la transcription automatique des manuscrits latins médiévaux. Les modèles de transcription automatique atteignent généralement un plateau de précision, entre 94% et 97% au niveau des caractères. L’affinement de ces modèles exige alors des améliorations incrémentielles, où chaque fraction de pourcentage supplémentaire nécessite des efforts disproportionnés. Les complexités sont accrues par les sauts de ligne inhérents aux manuscrits médiévaux, où l’absence de césure pose des problèmes. Déterminer précisément si un mot doit être coupé entre deux lignes introduit une ambiguïté considérable. Une solution à ce problème est le post-traitement.

Bien que les méthodes algorithmiques et l’apprentissage automatique aient été utilisés avec succès pour cette tâche, chaque approche présente des limitations. Les méthodes algorithmiques (telles que Pie) sont limitées dans leur adaptabilité, tandis que les méthodologies d’apprentissage automatique exigent des ensembles de données étendus et des ressources computationnelles pour l’entraînement. Compte tenu de ces facteurs, nous proposons une solution alternative offrant des avantages économiques par rapport à l’affinement de modèles comme RoBERTa pour la correction d’erreurs.

La solution proposée consiste à entraîner le transformer sur un vocabulaire assez restreint mais spécifique de trigrammes, que le tokeniseur spécialisé est formé à extraire. Plus précisément :

- Développement d’un tokeniseur spécialisé.
- Entraînement du modèle de transformer : Notre modèle est spécifiquement conçu pour reconnaître et corriger les erreurs couramment trouvées dans les transcriptions automatiques de manuscrits latins médiévaux datant du 13ème au 16ème siècle.
- Analyse et simulation des erreurs : En incorporant des erreurs simulées reflétant les erreurs de transcription courantes, nous facilitons une formation robuste du modèle.

**Mots-Clés:** Post, traitement, ATR, RoBERTa

---

\*Intervenant

# Représenter le champ des économistes en France dans les années 2010 avec une carte auto-organisatrice

Adrien Rougier \* <sup>1,2</sup>

<sup>1</sup> Université Libre de Bruxelles – Belgique

<sup>2</sup> Triangle : action, discours, pensée politique et économique – Université Lumière - Lyon 2, Université Lumière - Lyon 2, Université Lumière - Lyon 2 : UMR5206 – France

Notre objectif est de représenter graphiquement l'espace des économistes en France dans les années 2010. Pour cela, des données ont été rassemblées ou inférées sur 140 000 individus. Les variables retenues au niveau agrégé sont le capital scientifique, le capital académique, le capital médiatique, le capital de notoriété, le capital éditorial, le capital économique, et le domaine d'activité (secteur privé, administration, sphère politique, journalisme, enseignement secondaire, enseignants et chercheurs dans le supérieur). Au niveau désagrégé, un ensemble de 43 variables sont distinguées (e.g., le capital médiatique se décompose en un score de télévision, de radio, de presse écrite nationale et de centralité sur twitter).

Une carte auto-organisatrice a été utilisée à des fins de réduction de la dimensionnalité et de regroupement des individus en classes de similarité. Une telle carte est composée de neurones formant une grille, dans notre cas en deux dimensions. Une fois le poids des neurones initialisés (de manière aléatoire ici), la phase d'apprentissage consiste à sélectionner un économiste de la base de données (un vecteur d'entrée) et à trouver le neurone qui lui ressemble le plus (dont le vecteur de poids est le plus proche). Les poids du neurone gagnant, et à un moindre degré les poids de ses voisins, sont ajustés de telle sorte à se rapprocher du vecteur d'entrée. L'ampleur du rapprochement et du voisinage sont définis par une fonction qui diminue au fil des répétitions pour permettre une convergence.

Plusieurs représentations de l'espace des économistes sont ainsi obtenues en faisant varier la forme et la taille de la carte, en fixant différentes pondérations aux variables et en construisant des superclasses d'individus.

---

\*Intervenant

# Trajectoires professionnelles de docteur.es au sein de l'enseignement supérieur : l'intérêt des algorithmes de classification de prénom

Matthieu Cisel \* <sup>1</sup>, Pauline Antognelli \* <sup>2</sup>, Sami Nakab \* <sup>2</sup>

<sup>1</sup> Laboratoire AGORA – CY Cergy Paris Université : EA7392, CY Cergy Paris Université – France

<sup>2</sup> Institut des Humanités numériques – CY Cergy Paris Université – France

Au cours des dix dernières années, sont apparus des outils informatiques visant à associer à des prénoms d'origines variées, d'une part un genre, et d'autre part une origine ethnique. Malgré les limites consubstantielles aux modèles d'apprentissage machine sous-jacents, un certain nombre de recherches en sciences humaines se sont saisies de ces techniques pour produire des résultats prometteurs. Dans la lignée de ces travaux, nous proposons une étude fondée sur la base de données theses.fr et visant à analyser les trajectoires de centaines de milliers de docteur.es au sein de l'enseignement supérieur français. En mobilisant des régressions logistiques, nous cherchons en particulier à identifier des liens statistiques qui existent entre certaines caractéristiques des docteur.es – disciplines d'appartenance, genre et origine ethnique prédite, et la probabilité de devenir soi-même un jour encadrant.e, et présenter ainsi son/sa première doctorant.e. Eu égard à cette progression dans la carrière académique, nous constatons des écarts significatifs entre prénoms prédits comme masculins et ceux prédits comme féminins, écarts amplifiés lorsque l'on croise le genre avec l'origine ethnique du prénom. Laissant de côté l'interprétation des contrastes ainsi révélés, nous nous basons sur ces résultats pour mener une réflexion méthodologique sur les limites et opportunités des librairies visant à classer les prénoms, et sur celles des techniques d'apprentissage machine en général. Si celles-ci ne sont pas suffisamment précises pour produire des statistiques précises, elles ont néanmoins permis ici de dégager rapidement des tendances qu'il convient de confirmer par la suite, tout en respectant les contraintes associées à l'interdiction de la production de statistiques ethniques.

---

\*Intervenant

# Liste des auteurs

Antognelli, Pauline, 23

Bahroun, Rim, 14

Barbaro, Florian, 8

Bernard, Mathieu, 14

Burgun, Anita, 12

Cisel, Matthieu, 23

Coutelle, Patricia, 16

Hervé, Nel, 15

Jouin, Romain, 20

Largerion, Christine, 5

Le Gallic-Ach, Clara, 11

Le Priol, Emma, 12

Leprieur, Stéphanie, 16

Levratto, Nadine, 14

Louafi, Mehdi, 18

Lê, Sébastien, 15

M'Barki, Julien, 18

Magne, Stephane, 16

Martin, Marlène, 20

Marty, Stéphanie, 19

Mayaffre, Damon, 10

Miguel, Arthur, 6

Nakab, Sami, 23

Ollion, Étienne, 7

Olteanu, Madalina, 6

Rougier, Adrien, 22

Rynkiewicz, Joseph, 4

Schuck, Stéphane, 12

Talmatkadi, Manissa, 12

Texier, Nathalie, 12

Thoby, Sylvie, 15

Vanni, Laurent, 10

Yatsyk, Svetlana, 21

