



Antoine GOURRU

[antoine.gourru@univ-st-etienne.fr](mailto:antoine.gourru@univ-st-etienne.fr)  
<http://antoinegourru.com>

05/04/91

10 rue Claude Delaroa  
42000, Saint-Etienne



## Activités professionnelles et Formation Universitaire

---

**Depuis Septembre 2022** : Maître de Conférences en Informatique, Télécom Saint-Etienne/Laboratoire Hubert Curien

**2021 – 2022** : ATER, Faculté des Sciences, Université Jean Monnet, Saint-Étienne et équipe Data Intelligence, Laboratoire Hubert Curien

**2018 – 2021** : Doctorat en Informatique sous contrat doctoral, Laboratoire ERIC, Université Lumière Lyon 2, sous la supervision de Pr.Julien Velcin et de Pr.Julien Jacques (Laboratoire ERIC/Université Lumière Lyon 2) « Apprentissage de représentations d'auteurs et de documents : approches probabilistes à partir de représentations pré-entraînées. ». Thèse soutenue publiquement le 23/11/2021 devant un jury composé de :

- Marianne Clausel, Professeure des Universités, Université de Lorraine (Rapportrice)
- Benjamin Piwowarski, Chercheur Première classe, Sorbonne Université (Rapporteur)
- Christophe Gravier, Professeur des Universités, Université Jean Monnet (Examineur)
- Lynda Tamine-Lechani, Professeure des Universités, Université Paul Sabatier, Toulouse (Examinatrice)
- Julien Jacques, Professeur des Universités, Université Lumière Lyon 2 (Directeur de Thèse)
- Julien Velcin, Professeur des Universités, Université Lumière Lyon 2 (Directeur de Thèse)
- Ian Davidson, Full Professor, University of California, Davis (invité)

**2017-2018** : Master 2 Informatique, mention Data Mining, Université Lumière Lyon 2

**Stage** : Stage de recherche sous la supervision de Pr Ian Davidson (UC Davis) et Pr Julien Velcin, Laboratoire ERIC, Lyon : Détection de communautés et modélisation des flux d'informations dans les réseaux sociaux

**2016-2017** : Master 1 Informatique, Université Lumière Lyon 2

**Stage** : Stage de recherche sous la supervision de Pr Julien Velcin, Laboratoire ERIC, Lyon : Titrage automatique de modèles thématiques

**2013-2016** : Licence bidisciplinaire, Mathématiques et Informatique appliquées aux sciences humaines et sociales, mention Sciences Cognitives (MIASHS), Université Lumière Lyon 2

**2009** : Baccalauréat général, mention SVT, spécialité Mathématiques

## Thématiques de recherche

---

**Apprentissage automatique**, Deep Learning, Natural Language Processing, Large Language Models, Fairness, Explainability, Model Compression

## Publications principales :

Leteno, T., Gourru, A., Laclau, C., EMonet, R., & Gravier, C. (2023, April). Fair Text Classification with Wasserstein Independence

Dans *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*  
<https://arxiv.org/abs/2311.12689>

Dans cet article, nous nous intéressons à améliorer l'équité des modèles de langage en classification, c'est-à-dire leur capacité à ne pas discriminer une donnée en fonction de son groupe d'appartenance, notamment le genre ou le groupe ethnique. Le modèle s'inspire de l'approche adversaire pour créer une indépendance entre les représentations apprises pour prédire notre étiquette cible et celles apprises pour prédire l'attribut sensible. L'intérêt principal de l'approche est qu'elle ne nécessite pas d'annotations d'attributs sensibles dans les données d'entraînement et de test, ce qui la rend plus adaptée aux scénarios réels que les méthodes existantes.

Gourru, A., Velcin, J., Gravier, C., Jacques, J. (2022). Dynamic Gaussian Embedding of Authors  
Dans *Proceedings of the ACM Web Conference 2022* (pp. 2109-2119).

<https://openreview.net/attachment?id=ecsoxXEbxL&name=pdf>

Les auteurs publient des documents de manière dynamique. Leur sujet d'intérêt et leur style d'écriture peuvent changer au fil du temps. Nous proposons un nouveau modèle d'apprentissage de représentation, DGEA (Dynamic Gaussian Embedding of Authors), qui capture cette évolution. La représentation de l'auteur au temps  $t$  est une distribution gaussienne qui exploite des vecteurs de documents pré-entraînés, et qui dépend des publications observées jusqu'à  $t$ . Nous intégrons des contraintes telles que les représentations évoluent de manière lisse au cours du temps.

Gourru, A., Velcin, J., Jacques, J. (2020). Gaussian Embedding of Linked Documents from a Pretrained Semantic Space.

Dans *Proceedings of the 29th International Joint Conference on Artificial Intelligence, Main track (short papers)*, pages 3912-3918

<https://www.ijcai.org/proceedings/2020/541>.

GELD est une nouvelle méthode qui construit des représentations des documents liés (par exemple, des réseaux de citations) dans un espace sémantique pré-entraîné (par exemple, un ensemble d'embedding de mots). Nous formulons le problème de telle sorte que nous modélisons chaque document comme une distribution gaussienne dans l'espace vectoriel des mots. Nous concevons un modèle génératif qui combine à la fois les mots et les liens du graphe de document.

Davidson, I., Gourru, A., Ravi, S., (2018). The Cluster Description Problem - Complexity Results, Formulations and Approximations.

Dans *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pages 6193–6203,

<https://proceedings.neurips.cc/paper/2018/file/3fd60983292458bf7dee75f12d5e9e05-Paper.pdf>

Un défi pour l'IA explicable est de trouver une explication compacte de partitions préconstruites en utilisant un ensemble descripteurs/tags issus d'un dictionnaire commun. Nous montrons que le problème de faisabilité consistant à tester l'existence d'une description distincte est généralement difficile à résoudre, même pour deux clusters. Par conséquent, nous explorons des formulations ILP pour des problèmes plus petits et un cadre relaxé, mais restreint qui mène à un algorithme en temps polynomial pour des problèmes plus grands.

## Projets de recherche

---

MOLE : Model Optimization via Learning Examples, projet FIL, avec Jairo Cugliari

ANR Famous : équité des modèles multimodaux de machine learning

Projet région (R&D Booster) : chatbots intégrant du QA, Avec QACompany et Wikit

ANR DIKé : biais et équité des modèles de langue distillés

ANR Lifranum : Analyse de la littérature nativement numérique

## Séminaires

---

Frugalité et IA, journée ALLYS, Lyon, Octobre 2023 (Table ronde)

Fairness for NLP, groupe de lecture du GDR ARIA, Juin 2023

LEAVE: An End-to-End Variational Model for Fair Edge Prediction, Ghent, AI for HR and Public Employment Services, Février 2023

Learning Gaussian representations for graphs and textual data in a multi-objective context, UCL - London's Global University, Workshop on Functional Data (led by Nicolas Hernandez), Novembre 2022

Fairness for Machine Learning, Laboratoire Hubert Curien, Mai 2023

Les biais de genre des générateurs automatiques du langage. Démonstration à partir du chat GPT et autres dispositifs, Journée d'étude *Discours publicitaires, discours genrés : vers une société post-genre ?* Mars 2023

## Encadrement

---

### Thèses

**Octobre 2022-**: Thibaud Leteno: Biais et Equité des modèles de langue comprimés , co-encadré à 50% avec Christophe Gravier (LabHC) et Charlotte Laclau (Télécom Paris).

**Octobre 2021-** : Kunpeng Guo, CIFRE avec The QACompany: Improving the Performance of Question-Answering System in Domain-Specific Settings, co-encadré à 33%, avec Christophe Gravier (LabHC) et Dennis Diefenbach (Wikit)

### Chargé.e.s de recherche/Post doctorant.e.s

**2023-**: Emma Genthon : Impact of parameter fine tuning on model fairness.

**2023-**: Raphael Chevasson: Impact of parameter fine tuning on catastrophic forgetting

### Stages

2024

Nour Bouchouchi, M2, Multimodal Summarization of Scientific Documents, avec Laure Soulier (ISIR), Christophe Gravier (LabHC), Leo Hemamou (Sanofi) et Maali Mnasri (Sanofi)

2023

Ben Kabongo, M2, Data2Text personnalisé, avec Laure Soulier (ISIR) et Christophe Gravier (LabHC)

Emma Genthon, M2, Compression extrême de Transformers avec Virginie Fresse (LabHC)

Maia Sutter, M2, Unsupervised Stance Detection, avec Christine Largeron (LabHC) et Amine Trabelsi (Sherbrook University)

Enora Rivoal et Hadrien Haag, 1<sup>ère</sup> année ingénieur, Equité des systèmes de recommandation, avec Bissan Audeh (InaSoft) et Christine Largeron (LabHC)

## 2022

Thibaud Leteno, M2, Does fine-tuning make compressed language models sexist or racist ?, co-encadré avec Christophe Gravier (LabHC) et Charlotte Laclau (Télécom Paris)

## 2021

Mathilde Rodrigues Da Rocha, L2 : Application des Graph Neural Networks aux données IRMf, co-encadré avec Guillaume Metzler (ERIC), puis mémoire de L3

## 2020

Rohit Yadav, M2 : Variational Information Bottleneck for author embedding from pretrained language models, co-encadré avec Julien Velcin (ERIC)

Encadrement de 3 lycéens, « Initier un jeune à la recherche », avec l'association Un Peu de Bon Science pendant une semaine. Ils ont travaillé sur des algorithmes de classification de documents.

## 2019

Thomas Cambon, L3 : Développement d'un package python de titrage automatique de topic model, co-encadré avec Julien Velcin (ERIC)

## Animation de la communauté scientifique

---

### Membre de jury

« Qualifying exam » de Micheal Livanos, supervisé par Prof Ian Davidson, UC Davis, California

Comité de sélection, poste MCF section 27, Université Jean Monnet, Saint-Etienne

Comité de sélection, poste MCF section 26, Université Lumière Lyon 2

Soutenance de thèse de Laura Nguyen, 30 Avril 2024 (thèse dirigée par Benjamin Piwowarski)

### Organisation de groupes de lectures

Organisation et gestion du groupe de lecture Quality in Machine Learning, Laboratoire Hubert Curien

Participation à l'organisation du groupe de lecture Apprentissage et modèles pour le TAL (GDR TAL, bi-mensuel)

### Organisation de conférences

Septembre 2022 : Participation à l'organisation du workshop Graph-Quality@ECML-PKDD 2022

Juillet 2022 : Organisation du workshop Diké - Bias in (compressed) language Models

Octobre 2021 : Participation à l'organisation de la conférence internationale CCS 2021

Juil. 2019 : Participation à l'organisation de la conférence internationale SIGIR 2019

Mai 2019 : Participation à l'organisation du "Workshop on Representation Learning For Complex Data", Lyon 2019

Nov. -Déc. 2017 : Participation à l'organisation de la conférence internationale Complex Networks 2017

### Program Committee

NEURIPS 2023, EMNLP 2023, ECML 2023, CAP 2023, ACL 2023, TAL 2023, ICML 2023, IDA 2023, EMNLP 2022, ECML-PKDD 2022, IJCAI 2022, ACL 2022, AAAI 2022, ECML-PKDD 2021, AAAI 2021, IJCAI 2020, ECML-PKDD 2019, CORIA 2019, ACM SAC 2018

### Activités de médiation

---

07/12/23 : Club Gier, Impact de l'IA sur le tissu industriel

29/09/2023 : Instant Cult : l'IA dans la pop culture, Nuit Européenne des chercheur.euse.s

Novembre 2021 : bénévole dans le cadre de l'événement Super Demain, organisé par fréquence-écoles, dont le but est d'accompagner les enfants (et leurs parents) dans leurs pratiques numériques.

Novembre 2020 : Participation à l'organisation d'une journée IA et SHS dans le cadre de la Fête de la Science 2020 ([Vidéo](#)). Développement d'une activité de médiation destinée aux enfants : algorithmes de résolution de labyrinthe avec un robot mBOT.

2020 : Développement d'un logiciel en Scratch en collaboration avec Jairo Cugliari (ERIC), destiné aux plus jeunes pour découvrir l'IA à travers le jeu de NIM (<https://scratch.mit.edu/projects/367938703/>).

2021 – 2022 : Membre de l'association Aider à Apprendre, qui organise des séances de soutien scolaire pour les enfants en difficultés du 2ème arrondissement de Lyon (cours bénévoles d'1h tous les jeudis soir).

## Liste complète des publications

### Conférences internationales avec comité de lecture

Gourru, A., Laclau, C., Choudhary, M. & Largeron, C. (2024). Variational Perspective on Fair Edge Prediction

A paraître dans *International Symposium on Intelligent Data Analysis*

Rang B

Sutter, M., Gourru, A., Trabelsi, A., Largeron, C. (2024) Unsupervised stance detection for social media discussions: A generic baseline

A paraître dans *Conference of the European Chapter of the Association for Computational Linguistics*

Rang A

Davidson, I., Livanos, M., Gourru, A., Walker, P., Velcin, J., Ravi, S. S. (2024) An Exemplars-Based Approach for Explainable Clustering: Complexity and Efficient Approximation Algorithms  
A paraître dans *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*  
Rang A

Leteno, T., Gourru, A., Laclau, C., Emonet, R., & Gravier, C. (2023). Fair Text Classification with Wasserstein Independence  
*Dans Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*  
<https://arxiv.org/pdf/2311.12689.pdf>  
Rang A\*

Guo, K., Diefenbach, D., Gourru, A., & Gravier, C. (2023) Fine-tuning Strategies for Domain Specific Question Answering under Low Annotation Budget Constraints  
*Dans IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 166-171)*  
<https://arxiv.org/pdf/2401.09168.pdf>  
Rang B

Guo, K., Diefenbach, D., Gourru, A., & Gravier, C. (2023). Wikidata as a seed for Web Extraction.  
*Dans Proceedings of the ACM Web Conference 2023 (pp. 2402-2411).*  
<https://arxiv.org/pdf/2401.09175.pdf>  
Rang A\*

Leteno, T., Gourru, A., Laclau, C., & Gravier, C. (2023). An Investigation of Structures Responsible for Gender Bias in BERT and DistilBERT.  
*Dans International Symposium on Intelligent Data Analysis (pp. 249-261). Cham: Springer Nature*  
<https://arxiv.org/pdf/2401.06495.pdf>  
Rang B

Gourru, A., Velcin, J., Gravier, C., & Jacques, J. (2022). Dynamic Gaussian Embedding of Authors. *Dans Proceedings of the ACM Web Conference 2022 (pp. 2109-2119).*  
<https://openreview.net/attachment?id=ecsoxXEbXL&name=pdf>  
Rang A\*

Gourru, A., Velcin, J., Jacques, J. (2020). Gaussian Embedding of Linked Documents from a Pretrained Semantic Space.  
*Dans Proceedings of the 29th International Joint Conference on Artificial Intelligence, Main track (short papers), pages 3912-3918*  
<https://www.ijcai.org/proceedings/2020/541>.  
Rang A\*

Gourru, A., Guille, A., Velcin, J., Jacques, J. (2020). Document network projection in pretrained word embedding space.  
*Dans Advances in Information Retrieval. ECIR 2020. Lecture Notes in Computer Science (short papers), vol 12036.* Springer, pages 150-157  
[https://doi.org/10.1007/978-3-030-45442-5\\_19](https://doi.org/10.1007/978-3-030-45442-5_19).  
Rang A

Davidson, I., Gourru, A., Ravi, S., (2018). The Cluster Description Problem - Complexity Results, Formulations and Approximations.  
*Dans Advances in Neural Information Processing Systems 31 (NeurIPS 2018), pages 6193–6203,*  
<https://proceedings.neurips.cc/paper/2018/file/3fd60983292458bf7dee75f12d5e9e05-Paper.pdf>  
Rang A\*

Gourru, A., Velcin, J., Roche, M., Gravier, C., Poncelet, P. (2018). United We Stand: Using Multiple Strategies for Topic Labeling.  
*Dans Natural Language Processing and Information Systems. NLDB 2018. Lecture Notes in Computer Science, vol 10859. Springer, Cham. pages 352-363*  
[https://doi.org/10.1007/978-3-319-91947-8\\_37](https://doi.org/10.1007/978-3-319-91947-8_37)  
Rang C

### Conférences internationales avec comité de lecture, papier démo

Velcin, J., Gourru, A., Giry-Fouquet, E., Gravier, C., Roche, M., & Poncelet, P. (2018). Readitopics: Make your topic models readable via labeling and browsing.  
*Dans Demo track, IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pages 5874–5876*  
<https://doi.org/10.24963/ijcai.2018/867>  
Rang A\*

GUO, K., Defretiere, C., Diefenbach, D., Gravier, C., Gourru, A. (2022) QAnswer: Towards question answering search over websites,  
*Dans Companion Proceedings of the Web Conference 2022*  
<https://arxiv.org/abs/2401.09175>  
Rang A\*

### Reuves internationales avec comité de lecture

Davidson, I., Gourru, A., Velcin, J., Wu, Y. (2020). Behavioral differences: insights, explanations, and comparisons of French and US Twitter usage during elections.  
*Dans Social Network Analysis and Mining 10, Article numéro 6 (2020).*  
<https://doi.org/10.1007/s13278-019-0611-9>  
H-Index : 36

### Conférences nationales avec comité de lecture

Choudhary, M., Gourru, A., Laclau, C., & Largeron, C. (2023, February). Comment rendre les GNN plus équitables pour la prédiction.  
<https://hal.science/hal-04016156/document>  
*Dans Extraction et Gestion des Connaissances : Actes de la conférence EGC'2023 (Vol. 39).*

Gourru, A., Yadav, R., Velcin, J., (2021). Apprentissage de Représentations d'Auteurs et de Documents,  
*Dans Extraction et Gestion des Connaissances, EGC 2021, vol. RNTI-E-37, pages 11-23*  
<https://editions-rnti.fr/?inprocid=1002633>

### Workshops/ateliers internationaux avec comité de lecture

Terreau, E., Gourru, A., Velcin, J. (2021) Writing style author embedding evaluation.  
*Dans proceedings de The 2nd Workshop on "Evaluation & Comparison of NLP Systems" (@EMNLP 2021)*  
[https://eval4nlp.github.io/accepted\\_papers.html](https://eval4nlp.github.io/accepted_papers.html)



Brochier, R., Gourru, A., Guille, A., Velcin, J. (2020). Scientific Expert Finding in Document Networks: an Evaluation Framework.

*Dans proceeding de BIR 2020 Workshop on Bibliometric-enhanced Information Retrieval (@ECIR 2020)*

<http://ceur-ws.org/Vol-2591/paper-03.pdf>

Choudhary, M., Gourru, A., Laclau, C., Largeron, C. (2022) Learning Fair Variational Embedding with Graph Neural Networks,

*Graph-Quality@ECML-PKDD 2022*

<https://graphquality.github.io/program.html>

### **Conférences nationales avec comité de lecture sans acte**

Leteno, T., Gourru, A., Laclau, C., Emonet, R., & Gravier, C. (2023). Fair Bert using Wasserstein Disentanglement

Conférence francophone sur l'Apprentissage Automatique (CAp-23)

Gourru, A., Velcin, J., Jacques, J. (2021). Représentation d'auteurs par des distributions gaussiennes dans un contexte dynamique

Conférence francophone sur l'Apprentissage Automatique (CAp-21)

<https://cap2021.sciencesconf.org/resource/page/id/8>

Selosse, M., Gourru, A., Jacques, J., Velcin, J. (2019). Tri-clustering pour données de comptage. *Journées Des Statistiques*

<http://ids2019.sfds.asso.fr/program/>

*(Vendredi 07 Juin, track « Fouille de données : méthodologie »)*

### **Posters**

Gourru, A., Velcin J., E., Jacques J., Guille A. (2019) Document Network Projection in pretrained word embedding space

*Workshop on Representation Learning for Complex Data, Lyon*

Gourru, A., Giry-Fouquet, E., Davidson, I., & Velcin, J. (2018) Community structure and information diffusion in social networks.

*StatLearn 2018, Nice*