

Fanny Ducel

<https://fannyducel.github.io>
fanny.ducel@lisn.upsaclay.fr

Education

PhD in Computer Science

2023 –

Université Paris-Saclay

Gif-sur-Yvette, France

- Evaluation of stereotypical biases in autoregressive language models.
- Under the supervision of Aurélie Névéol and Karĕn Fort.

Master in Linguistics and Computer Science, with highest honours

2021 – 2023

Sorbonne Université

Paris, France

- M2: Valedictorian. Annual grade : 17.9/20. Thesis on the Evaluation of stereotypical biases in auto-regressive language models.
- M1: Valedictorian. Annual grade : 17.35/20. Thesis on the Analysis of claims in NLP articles by unsupervised learning.

Bachelor in Language Sciences, Computer Science focus, with highest honours

2018 – 2021

Sorbonne Université

Paris, France

Research Experience

Research Internship (M2 - 35h/week)

March 2023 - September 2023

LORLA, supervised by Karĕn Fort and Aurélie Névéol

Nancy, France

- Achieved literature review on stereotypical biases in autoregressive language models with a critical approach
- Experiments with manipulation of masked and autoregressive language models for bias assessment, reproduction and adaptation of bias mitigation and evaluation experiments
- Participation in the international research project MultiCrowsPairs, along with 20 other researchers and students from 8 different countries

Research Assistant (Part-time CDD - 17h/week)

September 2022

CERES, Sorbonne Université, supervised by Gaĕl Lejeune

Paris, France

- Participation in a project to map researchers according to their publications : wrote Python scripts to test new vectorization and clustering techniques, created graphs in Gephi, reorganized the Github repository with proper documentation

Research Internship (M1 - 35h/week)

July 2022 - September 2022

LORLA, supervised by Karĕn Fort and Maxime Amblard

Nancy, France

- Establishing a state of the art on the linguistic concepts of *hedging* and *claims*
- Creation and pre-processing of ACL articles corpus
- Extraction of *claims* in the corpus and classification by using *clustering* methods in Python
- Evaluation of unsupervised learning methods

Research Internship (L3 - 35h/week)

June 2021 - August 2021

LORLA, supervised by Karĕn Fort, Gaĕl Lejeune and Yves Lepage

Nancy, France

- Manual annotation of NLP articles (TALN, LREC, ACL)
- Automatic classification of these articles depending on whether the studied language is mentioned
- Work subsequently accepted at LREC and TALN

Teaching

Tutor for « De l'Écrit à l'information » (Bachelor's level) - 4h January 2025 - February 2025
Université de Lorraine *Nancy, France*

- Introduction to linguistics, for first year students. In French.

Tutor for Ethics in NLP (Master's level) - 10h November 2024 - January 2025
Université de Lorraine *Nancy, France*

- Introduction to ethics concepts, supervising group projects, evaluation and grading of oral presentations for NLP Master 1 and Master 2 students. In English.

Lecturer for a Python Programming class (Master's level) - 13h September 2024 - November 2024
Université de Lorraine *Nancy, France*

- Introduction to Python for beginners in NLP/Cognitive Science masters: data structures, functions, libraries, introduction to object-oriented. In English.

Tutor for Python Programming labs (Master's level) - 16h September 2024 - November 2024
Université de Lorraine *Nancy, France*

- Introducing labs, helping out students to complete them and correcting their assignments. Beginner group. In English.

Introduction to Grid5000 (4h) September 2024
Université de Lorraine *Nancy, France*

- First connection, key notions (site/cluster/job), essential commands (file transfer, resource reservation. In English.)

Tutor for Python Programming labs (Master's level) - 18h September 2023 - December 2023
Université de Lorraine *Nancy, France*

- Introducing labs, helping out students to complete them and correcting their assignments. Advanced group. In French.

Tutor in Computer Science for Language Science Students September 2021 - March 2023
Sorbonne Université *Paris, France*

- Shifts of 3 hours per week with students of the Bachelor in Language Sciences needing help to understand and complete their exercises and projects in Python programming, Web development (HTML and CSS), language engineering (TXM, Antconc, GATE), first order logic, ... In French.

Supervision

Co-supervision of NLP master students' supervised projects November 2024 / June 2025
Université de Lorraine, Nancy, France

- Co-supervision of two groups of 4 M1 students in NLP. Participation in meetings, scientific contribution and input.
- Group 1 : Racist bias detection in language models through code-switching. Léa Bouchet, Benjamin Caissial, Loïc Leclerc, Arthur Pedrini.
- Group 2 : Evaluating stereotypical biases in NLP benchmarks. Ibrahim Al Kotob, Binesh Arakkal Remesh, Lucie Digoin-Caparro, Shayan Ahmed Shariff.

Co-supervision of a bachelor's intern May 2025 / July 2025
LORIA, Nancy, France

- Co-supervision of Vidit Khazanchi for his work on racist bias evaluation through English-Hindi code-switching.

Co-supervision of a bachelor's intern

April 2025 / May 2025

LORIA, Nancy, France

- Co-supervision of Jeffrey André in his work on reproducing, extending and promotion of my gender bias framework.

Co-supervision of a Master's intern

March 2024 - August 2024

LORIA, Nancy, France

- Co-supervised Clémentine Bleuze who worked on "Analysing claims in NLP research". Discussions and weekly meetings with the intern, scientific contribution to the project.

Other Contributions

Peer-reviewing

- Reviewer for TALN 2025 (2 articles)
- Reviewer (4 articles) and ethic reviewer (8 articles) for ACL 2025
- Secondary reviewer for TALN 2024 (1 article)
- Member of the program committee and reviewer for the ATALA Day Ethics and NLP 2024 (2 long abstracts)
- Primary and secondary reviewer for LREC-COLING 2024 (3 articles)
- Secondary reviewer for Undone Computer Science 2024 (1 article)
- Secondary reviewer for GdR-LIFT 2023 (2 long abstracts)
- Secondary reviewer for ACL 2023 (1 article)
- Secondary reviewer for TALN 2023 (1 article)

Participation in conferences organization

- Student volunteer at ACL 2023 (registration desk, poster session)
- Student volunteer at IWCS 2023 (in charge of coffee breaks)

Science popularisation

- "PhD Pizza Time" for PhD students and interns in Computer Science: *How to define, understand and evaluate stereotypical biases in language models ?* – March 2025, LORIA, Nancy, France
- "Girls and Maths" Day (with the "Elles bougent !" organization): half-day event with middle-school girls to discuss about science and PhD – February 2025, Lycée Loritz, Nancy, France
- "CHICHE" talks: 1h-1h30 oral presentations to three groups of high-schoolers / four groups of high-schoolers. December 2023 / January 2024. Lycée Varoquaux, Tomblaine, France / Lycée Saint-Exupéry, Fameck, France

Invited talks

- *Comment définir, comprendre et évaluer les biais stéréotypés des modèles de langue ?* In the context of the work group *Intelligence Artificielle Souventable, Intelligible et Vérifiable (IASIV)* of Université Paris-Saclay. March 2025, Palaiseau, France.
- *Les enjeux éthiques de l'IA par le prisme du TAL et des biais stéréotypés.* Jointly with Karën Fort, on ethics and stereotypical biases in NLP, during a half-day workshop on AI. June 2024, ATILF, Nancy, France.

Awards

- Best Master's Thesis Award from the French Applied Linguistics Organization (AFLA), 2024 edition.

Member of the editorial board of the "Éthique et TAL" blog

- Blog post authoring and reviewer for blog posts.

Publications

International peer-reviewed conferences

"Women do not have heart attacks !" Gender Biases in Automatically Generated Clinical Cases in French. Fanny Ducel, Nicolas Hiebel, Olivier Ferret, Karën Fort, Aurélie Névéol. In Findings of the Association for Computational Linguistics: NAACL 2025, April 2025, Albuquerque, United States. <https://inria.hal.science/hal-04938811>

Participation : Design and implementation of an experiment to adapt French gender detection system, analysis of results for biomedical documents, writing of the article.

Your Stereotypical Mileage May Vary: Practical Challenges of Evaluating Biases in Multiple Languages and Cultural Contexts. Karën Fort, Laura Alonso Alemany, Luciana Benotti, Julien Bezançon, Claudia Borg, Marthese Borg, Yongjian Chen, Fanny Ducel, et al. Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), May 2024, Turin, Italy. <https://aclanthology.org/2024.lrec-main.1545/>

Participation : Meetings and discussions with 22 researchers from 8 different countries. Corrections in the English and French corpora, writing of a section of the article, designing the poster.

The Elephant in the Room: Analyzing the Presence of Big Tech in Natural Language Processing Research. Mohamed Abdalla, Jan Philip Wahle, Terry Ruas, Aurélie Névéol, Fanny Ducel, Saif Mohammad, Karën Fort. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2023, Toronto, Canada. - *CORE A**. <https://arxiv.org/pdf/2305.02797.pdf>

Participation : Frequent meetings and discussions with 6 other scientists from 3 different countries involved in the project. Extraction, analysis and formatting of raw data by Python scripts, manual annotations, proofreading and correction of the article.

Do we Name the Languages we Study? The #BenderRule in LREC and ACL articles. Fanny Ducel, Karën Fort, Gaël Lejeune, Yves Lepage. LREC 2022 - International Conference on Language Resources and Evaluation (LREC), June 2022, Marseille, France. - *CORE C, QUALIS B2*. <https://aclanthology.org/2022.lrec-1.60.pdf>

Participation : Oral plenary presentation at LREC, manual annotation of NLP articles, constitution and pre-processing of a corpus of articles from ACL Anthology, implementation of automatic classifiers in Python, analysis and formatting of results, writing of the article with an ethical approach calling for diversity, realization of the slides for the presentation

National peer-reviewed conferences

« Les femmes ne font pas de crise cardiaque ! » Étude des biais de genre dans les cas cliniques synthétiques en français. Fanny Ducel, Nicolas Hiebel, Olivier Ferret, Karën Fort, Aurélie Névéol. Conférence sur le Traitement Automatique des Langues Naturelles (TALN), July 2025, Marseille, France. Accepted.

Participation : Translation of the NAACL 2025 paper's abstract.

« De nos jours, ce sont les résultats qui comptent » : création et étude diachronique d'un corpus de revendications issues d'articles de TAL. Clémentine Bleuze, Fanny Ducel, Maxime Amblard, Karën Fort. Conférence sur le Traitement Automatique des Langues Naturelles (TALN), July 2025, Toulouse, France. Accepted.

Participation : Took part in the writing process and manual annotation campaigns.

La linguistique appliquée pour une IA plus éthique. Fanny Ducel, Aurélie Névéol, Karën Fort. Naturel et Artificiel en Linguistique Appliquée : une époque de paradoxes (NÉALA), July 2025, Nancy, France. Accepted.

Participation : Writing of a long abstract (position paper style), oral presentation.

Évaluation automatique des biais de genre dans des modèles de langue auto-régressifs. Fanny Ducel, Aurélie Névéol, Karën Fort. Conférence sur le Traitement Automatique des Langues Naturelles (TALN), Juillet 2024, Toulouse, France. <https://hal.science/hal-04621134v1>

Participation : Designing and implementing the experience which includes an automatic gender detection for French and Italian, stereotypical bias metrics, writing and presenting the article at the conference.

Langues « par défaut » ? Analyse contrastive et diachronique des langues non citées dans les articles de TALN et d'ACL. Fanny Ducel, Karën Fort, Gaël Lejeune, Yves Lepage. Conférence sur le Traitement Automatique des Langues Naturelles (TALN), June 2022, Avignon, France. <https://aclanthology.org/2022.jeptalnrecital-taln.14.pdf>

Participation : This is a second part of the work presented at LREC, so the tasks are very similar, but on different data. Production and presentation of the poster at TALN and same tasks as for LREC.

Journals

"You'll be a nurse, my son!" Automatically Assessing Gender Biases in Autoregressive Language Models in French and Italian Fanny Ducel, Aurélie Névéol, Karën Fort. Language Resources and Evaluation, to be published (accepted article).

Participation : Designing and implementing the experience which includes an automatic gender detection for French and Italian, stereotypical bias metrics, writing the article.

La recherche sur les biais dans les modèles de langue est biaisée : état de l'art en abyme. Fanny Ducel, Aurélie Névéol, Karën Fort. Revue TAL, September 2024. https://www.atala.org/content/tal_64_3_5

Participation : Reading and manually annotating a hundred articles on stereotypical biases in Language Models, writing a literature review paper with meta-data analysis.

Tutorials

Navigating Ethical Challenges in NLP: Hands-on strategies for students and researchers. Luciana Benotti, Fanny Ducel, Karën Fort, Guido Ivetta, Min-Yen Kan and Minzhi Li. ACL, August 2025, Vienna, Austria. Accepted.

Participation : Writing sections in the tutorial proposal, updating the tutorial material and resources and presenting the tutorial.

Workshops

Introducing MascuLead: the First Gender Bias Leaderboard. Fanny DuceL, Jeffrey André, Aurélie Névéol, Karën Fort. Ethic and Alignment of (Large) Language Models (@ CORIA-TALN), July 2025, Marseille, France. Accepted.

Participation : Part of the demonstrator's website design process and leaderboards creations, writing and presentation of the article.

"Un corpus qui fait tomber la barrière de la langue" : annotation de revendications dans des articles de TAL. Clémentine Bleuze, Fanny DuceL, Karën Fort, Maxime Amblard. Journées de lancement LIFT 2, November 2024, Orléans, France. <https://inria.hal.science/hal-04880335>

Participation : Proofreading the paper, manual annotations.

Desiderata for Actionable Bias Research Fanny DuceL, Aurélie Névéol and Karën Fort. New Perspectives on Bias and Discrimination in Language Technology, November 2024, Amsterdam, Netherlands. <https://inria.hal.science/hal-04755691>.

Participation : Writing a position paper about current issues in bias research and proposing solutions.

What ChatGPT tells us about ourselves. Aaron Boussidan, Fanny DuceL, Aurélie Névéol and Karën Fort. Journée Éthique et TAL 2024, April 2024, Nancy, France. <https://inria.hal.science/hal-04521121>.

Participation : Designing an experiment on *ChatGPT*, writing a long abstract, creating the poster.

Bias Identification in Language Models is Biased. Fanny DuceL, Aurélie Névéol, Karën Fort. Workshop on Algorithmic Injustice, June 2023, Amsterdam, Netherlands.

Participation : Established a state of the art on biases in language models, annotated above 60 articles, extracted meta-data, wrote the extended abstract of a literature review for a workshop.

Master Theses

Évaluation des biais stéréotypés dans les modèles de langue auto-régressifs : état de l'art et exploration de techniques extrinsèques Fanny DuceL. https://fannyducel.github.io/_pages/MemoireM2Fanny.pdf

Analyse des claims dans les articles de Traitement Automatique des Langues à l'aide d'une méthode par apprentissage non supervisé. Fanny DuceL. <https://tinyurl.com/2p976fx>

Specialized Skills

Languages: Python, SQL, SPARQL, Voice XML, UML, HTML, CSS, Java, C++, \LaTeX , Markdown

NLP tools: TXM, Antconc, Unitex, Praat, GREW-match, GATE, HTK, MBROLA, espeak, audeep

Other tools: Neo4j, GitHub, Apache Solr, GANT, Agile (Scrum), Gephi, Twitter API

Python libraries: Scikit-learn, SpaCy, NLTK, CoreNLP, Gensim, Pandas, LangChain, Pytorch

Language Skills

French (Native), **English** (C1), **German** (A2-B1)

Other Interests

Research interests: NLP, Ethics, Biases, Linguistic diversity, Sociolinguistics, Sociology, Psychology, Philosophy.