

Mathias Vast

Last year PhD Student in Information Retrieval

ISIR, Sorbonne Université & Sinequa (CIFRE) · Paris, France

vast@isir.upmc.fr · mathvast.github.io · github.com/MathVast · [Google Scholar](https://scholar.google.com/citations?user=...) · [LinkedIn](https://www.linkedin.com/in/mathvast)

CURRENT POSITION

Mar 2023 –
present

PhD Candidate (CIFRE), ISIR (MLIA team), Sorbonne Université & Sinequa, Paris
Improving the Robustness, Efficiency and Effectiveness of Neural Rankers in Information Retrieval

Advised by Benjamin Piwowarski (CNRS, ISIR), Laure Soulier (ISIR) and Basile Van Cooten (Sinequa). Research on zero-shot domain adaptation, interpretability of neural ranking models (cross-encoders, learned sparse retrieval), and efficient re-ranking architectures.

PUBLICATIONS

Peer-reviewed conference papers

- EMNLP 2026 **M. Vast***, V. Morand*, J. Mothe, B. Piwowarski. *MICE: Minimal Interaction Cross-Encoders for Efficient Re-ranking*. [arXiv](#) · [code](#)
- CIKM 2026 V. Morand*, **M. Vast***, B. Van Cooten, L. Soulier, J. Mothe, B. Piwowarski. *Reproducing and Comparing Distillation Techniques for Cross-Encoders*. [arXiv](#) · [code](#)
- CORIA 2026 **M. Vast***, V. Morand*, B. Van Cooten, L. Soulier, J. Mothe, B. Piwowarski. *Encodeurs croisés légers pour ré-ordonner les résultats de recherche d'information* [HAL](#)
- SIGIR 2026 M. Branga-Peicu, **M. Vast**, B. Van Cooten, L. Soulier, J. Françoise, B. Piwowarski, B. Caramiaux. *IR Lens: A Tool for Interpreting Cross-Encoder Models*. [website](#) · [code](#)
- SIGIR 2026 Y. Zong, **M. Vast**, B. Van Cooten, L. Soulier, B. Piwowarski. *From Tokens to Concepts: Leveraging SAE for SPLADE*. [DOI](#)
- ICTIR 2024 **M. Vast**, B. Van Cooten, L. Soulier, B. Piwowarski. *Which Neurons Matter in IR? Applying Integrated Gradients-based Methods to Understand Cross-Encoders*. [DOI](#) · [code](#)
- ECIR 2024 **M. Vast**, Y. Zong, B. Van Cooten, B. Piwowarski, L. Soulier. *Simple Domain Adaptation for Sparse Retrievers*. [DOI](#) · [arXiv](#)
- CORIA 2025 **M. Vast**, B. Van Cooten, L. Soulier, B. Piwowarski. *Comprendre la Nature des Signaux de Correspondance dans les Modèles Neuronaux pour la RI*. [HAL](#)
- CORIA 2024 **M. Vast**, Y. Zong, B. Van Cooten, B. Piwowarski, L. Soulier. *Adaptation de Domaine Simple pour la Recherche Parcimonieuse*.

Workshop papers

- WExIR @ SIGIR 2025 **M. Vast**, B. Van Cooten, L. Soulier, B. Piwowarski. *Understanding Matching Mechanisms in Cross-Encoders*. [HAL](#)

Under review

- TOIS **M. Vast**, B. Van Cooten, L. Soulier, B. Piwowarski. *How BERT-based Cross-Encoders Model Relevance in Information Retrieval?*

TALKS & PRESENTATIONS

- Jul 2026 Poster presentation, **SIGIR 2026**, Melbourne, France.
- 2026 Paper presentation, **CORIA 2026**, Nantes, France.
- Jul 2025 Paper + Poster presentation, **WExIR workshop @ SIGIR 2025**, Padova, Italy.
- Jul 2024 Paper presentation, **ICTIR 2024**, Washington D.C., US.
- Mar 2024 Poster presentation, **ECIR 2024**, Glasgow, UK.

PROFESSIONAL EXPERIENCE

Sep 2022 – Mar 2023	Machine Learning Engineer , Sinequa, Paris Pre-PhD period. Neural search models for production. PyTorch, PyTorch-Lightning, Hugging Face.
Feb 2022 – Jul 2022	Research Intern – Deep Learning , Sinequa, Paris Implemented a Fastformer -based neural ranker; applied pruning methods (SparseBERT, Fastformers) for efficient inference.
Sep 2020 – Feb 2021	Data Science Intern , Enedis, Paris Explainable AI methods for detecting bias and errors in neural networks; computer vision and NLP applications.

EDUCATION

2017 – 2022	Engineering degree (Diplôme d'ingénieur), Computer Science , Université de Technologie de Compiègne, France. Specialization: Artificial Intelligence & Data Science. GPA: 5/5.
2021 – 2022	MSc, Learning and Optimization of Complex Systems , Université de Technologie de Compiègne, France.
Feb – Jul 2019	Exchange semester , Sino-European School of Technology (UTSEUS), Shanghai, China.

SERVICE & OUTREACH

ongoing	Outreach talks on machine learning, deep learning and explainable AI to high-school students.
2017 – 2022	Member, DataVenture (data science student association), UTC.

SKILLS & LANGUAGES

research	Neural IR (cross-encoders, Late-interaction, SPLADE), interpretability, domain adaptation, efficient inference (distillation, pruning).
programming	Python (PyTorch, PyTorch-Lightning, Hugging Face), C/C++, R, SQL, Git.
languages	French (native) · English (C1, TOEIC 990) · Spanish (B2) · Mandarin (A2).