

Mathieu Riviere

Email: mathieu.riviere@chartes.psl.eu

GitHub: <https://github.com/icimathieu>

Website: <https://icimathieu.github.io/>

Profile

First-year MA student in Digital Humanities at Ecole des Chartes, interested in web infrastructure in a broad sense, data processing workflows, and opportunities enabled by generative AI, I am looking for a research internship in digital humanities.

My master's thesis in the history of science focuses on interactions between Pasteurians, scientific journals, the press, and parliamentarians under the Third Republic. It relies on computational methods: Gallica scraping, OCR, corpus structuring, annotation, RAG, and more broadly natural language processing.

Outlook: History agrégation, then a PhD in history with a strong digital humanities component.

Skills

- **Languages:** English C1 (TOEIC 860, 2020); German B1-B2
- **Programming / data & web:** Python (numpy, pandas, networkx, scikit-learn, PyTorch, OpenAI API, folium); HTML, JavaScript and RSS (basic web development); C (basics), R (basics).
- **Scraping:** BeautifulSoup (HTML parsing), Selenium (anti-bot bypass), Scrapy (proxy and VPN management): scraping of press and archive sources (Gallica, *Le Monde diplomatique*, Archives du Vaucluse)
- **OCR, vision & annotation:** local models (PaddleOCR, Tesseract, YOLO) or API models (Qwen-VL, Gemini), and Label Studio.
- **Structuring / editing:** XML-TEI, XPath, LaTeX, digital publishing workflows, basic database knowledge.
- **Textometry / NLP:** TXM, Iramuteq, Python processing (n-grams, function words, distances, etc.), RAG, NER, stylometry, emotion detection.
- **Tools:** GitHub (version control and publishing), HuggingFace, LLMs via API or locally.

Selected Projects

Archives du Vaucluse - OCR pipeline, geolocation, and postcard mapping (ENC Hackathon 2026)

- **Repository:** <https://github.com/icimathieu/vaucluse>
- **Scope:** OCR of postcards with VLMs, metadata extraction and cleaning, georeferencing, JSON/CSV structuring, and production of an interactive map (HTML). Database querying both locally and via API.

Oral/Written stylometry - Jordan Bardella corpus

- **Repository:** https://github.com/icimathieu/stylometrie_bardella_v1
- **Scope:** stylometric comparison between a written corpus, a transcribed oral corpus, and a control corpus; feature extraction (n-grams, function words), distance measures (cosine, Burrows' Delta), preprocessing scripts, and visualizations.

Master's thesis - currently: scraping, scientific corpus structuring, and annotation

- **Repositories:** https://github.com/icimathieu/scraping_pdf & <https://github.com/icimathieu/transcription>
- **Scope:** scraping and metadata extraction from Gallica; OCR of scientific journals (late 19th to early 20th century) with comparison of different models; corpus structuring; file and database management; first experiments in annotation with Label Studio and RAG (unpublished).

Experience

Video creator - [Histosef](#)

June 2023 - present

Production of popular history video content on YouTube based on scientific readings and, at times, essays.

Writer - L'Ouvreuse (Sorbonne webzine)

July 2024 - present

Research internship - Digital editing and textometry (CACTUS group, ENS de Lyon)

January 2025 - April 2025

Participation in the digital editing of medieval texts from the [Base de Francais Medieval](#):

- metadata research and integration
- manuscript/text alignment work
- proofreading editions
- zone annotation (bounding boxes) on illuminations

Front desk and security agent - Musee d'art et d'histoire Baron Gerard (Bayeux)

June 2023; June-July 2024

Private tutoring

September 2022 - December 2023

Tutoring high-school students in mathematics and scientific subjects.

Education

- **Ecole nationale des Chartes, Ecole Normale Superieure and Universite PSL** - MA in Digital Humanities (Sept. 2025 - present)

Research seminars in history and philosophy at EHESS and ENS; practical digital humanities classes at ENC; geography lectures at Sorbonne.

- **Universite Paris 1 Pantheon-Sorbonne** - BA in History, highest honors (Sept. 2022 - June 2025)

Additional courses in history and computer science; research seminars in history and philosophy; auditing political science undergraduate courses.

- **Lycee Saint-Louis (Paris 6e)** - CPGE PCSI (Sept. 2021 - July 2022) - 23rd / 46