

Mathieu Rivière

Adresse mail : mathieu.riviere@chartes.psl.eu

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

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

Profil

Étudiant en M1 Humanités Numériques à l'École des Chartes, intéressé par les infrastructures du web au sens large, les workflows de traitement de données et les opportunités permises par l'IA générative, je souhaite réaliser un stage de recherche en humanités numériques.

Mon mémoire de master en histoire des sciences porte sur les interactions entre pastoriens, revues scientifiques, presse et parlementaires sous la IIIe République. Il mobilise des méthodes computationnelles : scraping Gallica, OCR, structuration de corpus, annotation, RAG, classification et plus largement traitement automatique de la langue.

Perspectives : Agrégation d'histoire puis doctorat en histoire avec forte composante en humanités numériques.

Compétences

- **Langues** : Anglais C1 (TOEIC 860, 2020) ; Allemand B1–B2
- **Programmation / data & web** : Python (numpy, pandas, networkx, scikit-learn, PyTorch, API OpenAI, folium) ; HTML, JavaScript et RSS (développement web simple) ; C (bases), R (bases).
- **Scraping** : BeautifulSoup (parsing HTML), Selenium (contournement d'anti-bots), Scrapy (gestion de proxys et VPN) : scraping de presse et d'archives (Gallica, *Le Monde diplomatique*, Archives du Vaucluse)
- **OCR, vision & annotation** : modèles en local (PaddleOCR, Tesseract, YOLO) ou via API (Qwen-VL, Gemini) et Label Studio.
- **Structuration / édition** : XML-TEI, XPath, LaTeX, workflows d'édition numérique, notions en bases de données.
- **Textométrie / TAL** : TXM, Iramuteq, traitements Python (n-grams, mots-fonctions, distances, etc.), RAG, NER, stylométrie, détection d'émotions.

- **Outils** : GitHub (versionnement et publication), HuggingFace, LLMs via API ou en local.
-

Projets (sélection)

Archives du Vaucluse — Pipeline OCR, géolocalisation, et cartographie de cartes postales (Hackathon ENC 2026)

- **Dépôt** : <https://github.com/icimathieu/vaucluse>
- **Objet** : OCR de cartes postales avec VLMs, extraction et nettoyage de métadonnées, géoréférencement, structuration JSON/CSV et production d'une carte interactive (HTML). Interrogations de bases de données en local et via API.

Stylométrie oral/écrit — corpus Jordan Bardella

- **Dépôt** : https://github.com/icimathieu/stylometrie_bardella_v1
- **Objet** : comparaison stylométrique entre corpus écrit, corpus oral transcrit et corpus de contrôle ; extraction de traits (n-grams, mots-fonctions), mesures de distance (cosinus, Burrows' Delta), scripts de prétraitement et visualisations.

Mémoire de master — actuellement : scraping, structuration de corpus scientifiques et annotation

- **Dépôts** : https://github.com/icimathieu/scraping_pdf & <https://github.com/icimathieu/transcription>
 - **Objet** : scraping et extraction de métadonnées depuis Gallica ; OCR de revues scientifiques (fin XIXe – début XXe) avec comparaison de différents modèles ; structuration de corpus ; gestion de fichiers et bases de données ; premières expérimentations d'annotation avec Label Studio et de RAG (non publiées)
-

Expérience

Vidéaste — [Histosef](#)

Juin 2023 – actuel

Production de contenus audiovisuels de popularisation historique sur YouTube à partir de lectures scientifiques voire d'essais.

Rédacteur — L'Ouvreuse (webzine Sorbonne)

Juillet 2024 – actuel

Stage de recherche — Édition numérique et textométrie (groupe CACTUS, ENS de Lyon)

Janvier 2025 – Avril 2025

Participation à l'édition numérique de textes médiévaux de la [Base de Français Médiéval](#) :

- recherche et intégration de métadonnées
- correspondances manuscrits/textes
- relecture d'éditions
- annotation de zones (bounding boxes) sur enluminures

Agent d'accueil et de surveillance — Musée d'art et d'histoire Baron Gérard (Bayeux)

Juin 2023 ; Juin–Juillet 2024

Cours particuliers

Septembre 2022 – Décembre 2023

Accompagnement de lycéens en mathématiques et disciplines scientifiques.

Formation

- **École nationale des Chartes, École Normale Supérieure et Université PSL** — Master Humanités numériques (Sept. 2025 – actuel)

Séminaires de recherche en histoire et philosophie à l'EHESS et à l'ENS ; travaux dirigés d'humanités numériques à l'ENC ; cours magistraux de géographie à la Sorbonne.

- **Université Paris 1 Panthéon-Sorbonne** — Licence d'histoire, mention Très Bien (Sept. 2022 – Juin 2025)

Cours complémentaires en histoire et informatique ; séminaires de recherche en histoire et philosophie ; auditeur libre de la licence de sciences politiques.

- **Lycée Saint-Louis (Paris 6e)** — CPGE PCSI (Sept. 2021 – Juil. 2022) — 23e / 46