

Clément Longeac

ML/NLP researcher specializing in hybrid AI pipelines for clinical NLP , balancing **performance, explainability, and energy frugality**. Designed the DEMNE cascade framework (Rules→CRF→Transformer→LLM) for automated Determination of extraction method for named entity, co-authored a paper in redaction, and built production NLP pipelines at **AP-HP** (~11M patient records). **Applying for the Anthropic Fellows Program to contribute empirical research on AI safety, interpretability, and scalable oversight** — motivated by building AI systems that are safe and beneficial for society.

Contact :

Noisy-Le-Grand, France



(FR) +33 6 28 03 85 97



clement.longeac@edu.esiee.fr



www.linkedin.com/in/clèment-longeac/

EDUCATION

09/2024-07/2027

ESIEE - PARIS (France) , Engineering school, Cité Descartes, 2 Bd Blaise Pascal, 93160 Noisy-le-Grand - Currently in fourth of a five year program leading to the equivalent of an MS in Engineering in DataScience & AI -

09/2021-07/2024

Claude Fauriel High School, Saint Étienne (42) (France) – A two-year post-secondary level course in Engineering Science/Physics and Chemistry

- Specialized in physics and applied chemistry

PUBLICATIONS & RESEARCH

- Longeac C. et al. — “DEMNE: Determination of Extraction Method for Named Entity” — JMIR, [IN PREPARATION]

EXPERIENCE

11/25 – 06/26 Research Fellow — Tremplin Recherche Program ESIEE Paris × LBA Aix-Marseille Université × AP-HP (Entrepôt de Données de Santé) ; Supervisor: Dr. A. Redjdal | Co-investigator: Dr. E. Kempf

- Designed **DEMNE**, a formal cascade NLP routing framework (Rules→Transformer→LLM) for extracting 7 oncological biomarkers from 95+ French clinical reports; achieved **83.1% exact ternary concordance** across 65 entities and 3 multi-institutional corpora.

- Developed **5 novel corpus-characterization metrics** (Templateability, Lexical Homogeneity, Contextual Risk, Frequency, Feasibility) computed automatically from BRAT annotations — corpus-agnostic paradigm selection.

- Executed exhaustive **grid search** over **12,636 threshold configurations** with asymmetric ordinal loss , reducing energy cost by $\times 10^4$ vs full-LLM baseline.

- Built and deployed DEMNE at **EDS-NLP pipelines at AP-HP** for oncological signal detection using eds.contextual_matcher.

- Published **open-source** GitHub repository with full metric computation pipeline, Streamlit dashboard.

05/25 – 08/25 Internship — Synchrotron-SOLEIL – Paris-Saclay

About my missions:

- Writing autotest scripts using OpenCL

- Running them on the Debian ROCm CI infrastructure (via Salsa)

- Using results to validate source code modifications

- Goal: provide a clear overview of scientific software compatibility across the various AMD GPUs available in the CI ROCm.

Working with the rocm-test-launcher script to:

- Run installation/deinstallation tests on clean virtual machines

- Automate autotest procedures for ROCm. Aims to improve AMD GPU support across Debian

12/24 – Today Permanent contract — Complétude — Private tutors — Math / Physics — Secondary level

Today Personal commitment and high regard for the protection of bees and beekeeping

IT SKILLS

• Programming: Python (Advanced) / C (Advanced) / C++ / Bash / Shell / VHDL

• ML & NLP: PyTorch / TensorFlow / HuggingFace Transformers / spaCy / scikit-learn / pycrfsuite

• LLM Research: DrBERT / BioMistral / GPT-4 / Claude API / Fine-tuning / Few-shot prompting

• Clinical NLP: EDS-NLP (edsnlp) / BRAT annotation / CoNLL NER / eco2ai

• Research Methods: Grid search optimization / Asymmetric loss / Corpus analysis / Frugal AI / XAI

• OS & Tools: Linux/Debian / Git / GitHub / GitLab / CI-CD / Streamlit / Plotly / Pandas / SQL

• Compliance: EU AI Act (Art. 12/14) / RGPD / HDS / AFNOR SPEC 2314

• GPU: AMD ROCm / OpenCL (Intermediate)

GITHUB/GITLAB PERSONAL PROJECT & CONTRIBUTIONS:

DEMNE Research Repo: github.com/longeacc/DEMNE-

Repo GitHub: <https://github.com/longeacc>

Repo GitLab-Salsa: https://salsa.debian.org/Clement_LONGEAC

Kaggle projects: <https://www.kaggle.com/clmentlongeac/code>

LANGUAGE LEVELS

. French (native language)

. English (advanced)

. Spanish (intermediate)

. Japanese (beginner)

. Italian (beginner)

LICENSES / CERTIFICATIONS

. PIX certification

. TOEIC & TOEFL (in progress)

. IELTS (in progress)

SOFT SKILLS

. Research-driven

. Versatile

. Creative

. Adaptable

. Teamwork

. Determined

. Pedagogical (Private Tutor)

. Leadership (Synchrotron-SOLEIL)

INTERESTS

. AI Safety & Alignment

. Basketball (Advanced)

. Beekeeping (Beekeeper)

. Particle physics (CERN, ITER)

. Carl Sagan, Stephen Hawking

REFERENCES AVAILABLE ON REQUEST

Research Supervisor: Dr. Akram REDJDAL

akram.redjdal@esiee.fr

- Internship Director: Emmanuel FARHI

emmanuel.farhi@synchron-soleil.fr

Frédéric-Emmanuel PICCA

frederic-emmanuel.picca@synchron-soleil.fr