Gustave Cortal

gustavecortal.com

gustavecortal@gmail.com
+33 (0)6 52 58 00 56

Research Scientist at Callyope, building evals for frontier models. PhD in natural language processing from ENS Paris-Saclay. Training and evaluating language models in research labs and startups since 2020. 8 peer-reviewed publications, with first-author papers at EMNLP and LREC-COLING. Open models and datasets released on 🤗.

Education

ENS Paris-Saclay — **Paris, France**
PhD in natural language processing. Advisors: A. Finkel, P. Paroubek, and L. Ye — 2022 – 2026

- Thesis, defended Jan 2026: *Natural language processing for subjectivity analysis in personal narratives*.

Université de Technologie de Compiègne — **Compiègne, France**
Engineering degree (MSc equivalent) in data science and artificial intelligence — 2017 – 2022

- Exchange semesters at Taiwan Tech (Taipei, 2020, ranked 1st in computer vision) and Shanghai University (2019).

Selected work experience

Research Scientist, Callyope (startup) — **Paris, France**
NLP for psychiatry — Mar 2025 – Present

- Building a mental health safety benchmark for frontier models from patient speech and psychiatrist ratings.
- Designed an LLM request scheduler that cut processing of 15,000 clinical pages from 3 hours to 30 minutes.
- Optimizing agentic workflows on long clinical documents with recursion and prefix caching.
- Built clustering methods with LLM-generated descriptions to find speech topics linked to clinical scales.

Research Scientist, ENS Paris-Saclay — **Paris, France**
NLP for emotion recognition — Feb – Jul 2022

- Trained French language models to recognize emotions from their psychological components.
- Released a dataset of 812 emotional narratives annotated with these components.

Research Engineer, Dashlane — **Paris, France**
NLP for code generation (part-time) — Sep 2021 – Feb 2022

- Designed transformer and graph-based methods for HTML generation to improve form automation.

Software Engineer, Makezu (startup) — **Paris, France**
NLP for lead generation on social media — Sep 2020 – Feb 2021

- Fine-tuned language models for trend analysis and user profiling on social media.

Selected open-source projects

Open models, datasets and code released on 🤗 Hugging Face and GitHub.

PhiloLM, a language model with 112M parameters, trained from scratch on open philosophy data (PhiloText).
Language models fine-tuned on psychological reasoning, aligned on psychotherapeutic preferences.
Language models that generate dream narratives and annotate their characters and emotions.
thought steering, contrastive activation steering that makes a language model reframe negative thoughts.
lexical sophistication, a sampler that steers a language model toward common or rare English words.
philosophical search, semantic search over passages from philosophy resources.
dérive, a recommender of rare concepts using concept embeddings and Rocchio relevance feedback.
animacy, a tiny autoregressive transformer trained on trajectories of fish and flies.
pylos, a tiny transformer that learned the board game Pylos by self-play reinforcement learning.
order flow, reinforcement learning in a market simulated by a transformer trained on Nasdaq order flow.
ganlive, real-time latent space exploration of GANs trained on my photographs, driven by MIDI.
imagevolve, a CLIP-guided evolutionary search over deterministic image transformations.

Publications

Interpretable Topic Modeling of Spontaneous Speech in Depression Using Large Language Models. **G. Cortal** et al. In *JMIR Mental Health, 13(1), 2026*.

Formalizing Style in Personal Narratives. **G. Cortal** and A. Finkel. In *EMNLP 2025 (main)*. [poster] [slides]

Textualized and Feature-based Models for Compound Multimodal Emotion Recognition in the Wild. N. Richet, S. Belharbi, H. Aslam, M. Schadt, M. González-González, **G. Cortal** et al. In *ABAW at ECCV 2024*.

Improving Language Models for Emotion Analysis: Insights from Cognitive Science. **G. Cortal** and C. Bonard. In *CMCL at ACL 2024*. [French version]

Sequence-to-Sequence Language Models for Character and Emotion Detection in Dream Narratives. **G. Cortal**. In *LREC-COLING 2024*. [slides] [French version]

Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation. **G. Cortal** et al. In *LaTeCH-CLfL at EACL 2023*. [slides] [video]

Teaching experience

As instructor, designed slides, exercises, and exams from scratch.

Natural language processing, EPITA. Instructor, 90h, 150 students. [materials] 2024 – 2025

Statistical learning with R, ENS Paris-Saclay. Instructor, 31h, 25 students. [materials] 2023

Algorithms and complexity, CentraleSupélec. Teaching assistant, 21h, 30 students. 2022 – 2023

Talks

Natural language processing for subjectivity analysis in personal narratives. *Psychiatrie de l'Enfant et de l'Adolescent (Pitié-Salpêtrière, AP-HP), 2025*. [slides]

Statistical language modeling: from n-grams to transformers. *LMF (ENS Paris-Saclay), 2024*. [slides] [video]

Natural language processing: perspectives on emotion analysis. *LMF (ENS Paris-Saclay), 2024*. [slides]

Large language models and emotional understanding. *Institut Jean-Nicod (EHESS, ENS Ulm), 2024*. [slides]

Automatic identification of emotions and their causes. *DATAIA, 2022*. [slides]

Student supervision

R. Faure. *Formalizing subjective experience with language models*. MSc thesis co-supervised with A. Finkel. 2025

N. Richet. *Multimodal emotion analysis*. MSc thesis co-supervised with E. Granger and H. Aslam. 2024

U. Pant. *NLP and formal methods*. Bachelor's project co-supervised with A. Finkel. 2024

R. Hachem. *Imageability prediction in emotional narratives*. MSc thesis co-supervised with A. Finkel. 2023

Additional activities and honors

Reviewed for ACL Rolling Review and the Computational Affective Science and WASSA workshops. 2022 – 2026

Awarded 35,000 GPU hours on the Jean Zay supercomputer to train language models. 2023 – 2026

Built emotion recognition and spoken replies for Egon, an art installation in Paris. 2022

Awarded PhD scholarships from Paris-Saclay University and Artificial Intelligence for Île-de-France. 2022

Skills and languages

Programming Python (PyTorch, scikit-learn, DSPy, LangGraph), TypeScript, R, C/C++.
Training DeepSpeed, PyTorch FSDP, Accelerate, PEFT, TRL.
Serving vLLM, llm-d, Kubernetes.
Languages French (native), English (fluent), Spanish (basic), Mandarin Chinese (basic).