



# Louis Martin

NLP RESEARCH SCIENTIST, PH.D.

☎ (+33) 6 45 82 43 31 | ✉ louismartincs@gmail.com | 🏠 louismartin.eu | 📺 louismartin  
| 📄 louismartincs | 🎓 Louis Martin

Passionate Research Scientist in Natural Language Processing.

## Experience

### Meta AI

RESEARCH SCIENTIST

*London, UK*

*Aug. 2021 - Present*

- **Applied Research** on text understanding to make **recommender systems more interpretable**.
- Lead project on **large scale weakly-supervised pretraining** of NLP models on social media data.
- Explore how **retrieval augmented models** can lead to richer general purpose text embeddings.

### Facebook AI Research & Inria Paris

PH.D. STUDENT

*Paris, France*

*May. 2018 - Aug. 2021*

- Created a demo based on my research in Text Simplification that was **presented to Mark Zuckerberg**. It was also **showcased by Facebook's CTO** at the company-wide *All Hands* conference, in front of 15k employees.
- Presented my research in person to the **French Minister for People with disabilities** (my research is core to the Cap'FALC [↗](#) project).
- My latest text simplification method achieves **state-of-the-art results on all simplification benchmarks** while being unsupervised.
- My research involved **challenging large-scale engineering skills**: I developed and trained CamemBERT on hundreds of GPUs. I scaled my own dense mining pipeline to 1 billion sentences by optimizing disk IOs, caching, all on 1000 preemptable GPUs.
- Major success of the release of our open-source **CamemBERT model** [↗](#) (>300k downloads/month, multiple industry applications).
- My work was **featured in 9 news articles/podcasts**, I gave 3 interviews (VentureBeat [↗](#), Le Monde [↗](#), France Culture [↗](#), ActuaIA [↗](#), ...).

### Facebook AI Research

NLP RESEARCH INTERN

*Paris, France*

*May. 2017 - Nov. 2017*

- Integrated a neural conversational model in Facebook Messenger (PHP and Python) as an internal chatbot demo.
- Trained a neural model for question generation on 100k question/answer pairs as an additional module to the chatbot.

### GoPro

COMPUTER VISION R&D INTERN

*Paris, France*

*Feb. 2016 - Jul. 2016*

- Developed and trained a compact on-device CNN model (3MB) with state-of-the-art performance for face recognition.
- My work led to the submission and acceptance of a **US patent (sole inventor)** on compact neural models.

## Education

### Facebook AI Research, Inria Paris, Sorbonne Université

PH.D. IN COMPUTER SCIENCE

*Paris, France*

*May. 2018 - Aug. 2021*

- Subject: *Deep Learning for Text Simplification using Unsupervised Multilingual Approaches*. Ph.D. at Facebook AI Research and Inria Paris (ALMAnaCH team) supervised by Antoine Bordes (Facebook), Benoît Sagot (Inria) and Éric de la Clergerie (Inria).

### École Normale Supérieure Paris-Saclay

MASTER OF SCIENCE - **MVA** (MATHEMATICS, VISION AND MACHINE LEARNING)

*Cachan, France*

*Oct. 2016 - Apr. 2017*

- Computer Vision (C. Schmid, J. Ponce), Deep Learning (I. Kokkinos), Reinforcement Learning (A. Lazaric), Sparsity and Compressed Sensing (G. Peyre), Kernel Methods (J. Mairal, J.-P. Vert).

### University of New South Wales

POSTGRADUATE EXCHANGE - **MACHINE LEARNING**

*Sydney, Australia*

*March. 2015 - Jul. 2015*

- Artificial Intelligence, Machine Learning, Bayesian Inference.

### École Centrale Paris (now CentraleSupélec)

ENGINEERING DEGREE (MS EQUIVALENT) - **APPLIED MATHEMATICS**

*Châtenay-Malabry, France*

*Sept. 2013 - May. 2017*

- Machine Learning, Natural Language Processing, Distributed Optimization, Stochastic Processes, Non-linear PDEs.

## Teaching

---

- 2019 - 2022 **African Master's in Machine Intelligence - Deep NLP**. Teaching in person for 3 years in Rwanda and Senegal.  
2020 **MVA Master - Deep Learning** - Lecture hall class and NLP final class project.

## Honors & Awards

---

- 2020 **Outstanding Reviewer (EMNLP 2020)**  
2015 **3rd position out of 3242 in Kaggle Bike Sharing Demand Competition** ↗: Entered the competition one week before the end, and finished 3rd using neural networks for the prediction of shared bikes demand.  
2014 **Merit Scholarship**, Granted by École Centrale Paris, for top 1% of students with very good academic results with a significant positive impact on society (thanks to volunteering).

## Other Interests

---

- 2017 - Present **[Effective Altruism (EA)] Active member of the EA community**. Promote altruistic ideas and events on social media, participate in conferences, signed the Giving Pledge, followed 80k hours career coaching.  
2019 **[Side Projects] Creation of Greenhouse Gas Dashboard**. Built an internal dashboard at Facebook to monitor and raise awareness on realtime CO2 emissions of our GPU cluster, with very positive internal reception.  
2019 - 2021 **[Volunteering] Active Involvement in Social Life at Meta**. Organized various virtual social events during the pandemic to improve wellbeing. Elected as an employee representative.  
2013 - 2014 **[Volunteering] President of International Aid Student Association**. Led 12 students, gathered funds (50k€), and built a refuge to help homeless children of Madagascar onsite. Rewarded with merit scholarship.

## Skills and tools

---

- **Programming:** Python (master - NumPy, PyTorch, pandas...), C/C++, Java, R, Matlab, JavaScript, PHP, HTML5, SQL.
- **Open-Source** (🔗 /louismartin): EASSE (maintainer), fairseq (contributor), HuggingFace Transformers (contributor), submitit (early contributor), nevergrad (early contributor).
- **Tools:** Vim (master), Git (master), Linux, web servers, AWS, OVH, tmux, LaTeX, Slurm, VS Code, macOS.
- **Languages Spoken:** French (Native), English (Fluent), German (Written and Spoken), Spanish (Written and Spoken).

## Publications & Patents

---

- [1] Mikel Artetxe, Shruti Bhosale, Naman Goyal, Todor Mihaylov, Myle Ott, Sam Shleifer, Xi Victoria Lin, Jingfei Du, Srinivasan Iyer, Ramakanth Pasunuru, Giri Anantharaman, Xian Li, Shuohui Chen, Halil Akin, Mandeep Baines, **Louis Martin**, Xing Zhou, Punit Singh Koura, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Mona Diab, Zornitsa Kozareva, Ves Stoyanov *Efficient large scale language modeling with mixtures of experts*. **EMNLP 2022**
- [2] **Louis Martin**, Angela Fan, Éric de la Clergerie, Antoine Bordes (Facebook), Benoît Sagot. *MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases*. **LREC 2022**
- [4] Fernando Alva-Manchego\*, **Louis Martin\***, Antoine Bordes, Carolina Scarton, Benoît Sagot, Lucia Specia. *ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations*. **ACL 2020**
- [5] **Louis Martin**, Benoît Sagot, Eric de la Clergerie, Antoine Bordes. *Controllable Sentence Simplification*. **LREC 2020**
- [6] **Louis Martin\***, Benjamin Muller\*, Pedro Javier Ortiz Suárez\*, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah, Benoît Sagot. *Les modèles de langue contextuels Camembert pour le français: impact de la taille et de l'hétérogénéité des données d'entraînement*. **TALN 2020 (nominated for best paper award)**
- [7] **Louis Martin\***, Benjamin Muller\*, Pedro Javier Ortiz Suárez\*, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah, Benoît Sagot. *CamemBERT: a Tasty French Language Model*. **ACL 2020**
- [8] Fernando Alva-Manchego\*, **Louis Martin\***, Carolina Scarton, Lucia Specia. *EASSE: Easier Automatic Sentence Simplification Evaluation*. **EMNLP 2019 - System Demonstrations**
- [9] **Louis Martin**. *Systems and Methods for Compact Convolutional Neural Networks*. **US Patent** 10,185,891 (2019)

\*Equal contribution.