

Pierre Colombo

Born the 15.07.1994 at Corneilles en Parisis (95240)

📞 (+33) 650118518 • ✉ colombo.pierre@gmail.com

Websites: 🌐 [in](#) [in](#) [in](#) [in](#)

I am a Associate Professor at Université Paris Saclay (CentraleSupélec). My research focuses on Natural Language Processing with a particular focus on safety and finetuning of language models.

Education

- ✓ **PhD in Computer Science** **Telecom Paris, Institut Polytechnique de Paris**
Paris, France 2018 – 2021
PhD entitled: Learning to represent and generate text using information measures
- ✓ **MSC in Computer Science** **EPFL**
Lausanne, Switzerland 2016 – 2018
- ✓ **MSC in Applied Mathematics** **CentraleSupélec**
Paris, France 2014–2018
- ✓ **Preparatory School MPSI - MP*** **Lycee Condorcet**
Paris, France 2012–2014

Experience

- ✓ **Associate Professor** **Paris France**
Université Paris Saclay (CentraleSupélec) Sept 2022
Main subject: Safe AI, Large Language Model Finetuning, Robust Evaluation of LLM.
- **Invited Professor** **Montreal Canada**
LIVIA, hosted by Prof. Ismail Ben Ayed July - November 2022
Main subject: Safe AI
- **Post Doctoral Researcher** **Paris France**
CNRS-L2S (CentraleSupélec) Nov. 2022 - July. 2023
Main subject: Safe AI
- **Phd Student** **Paris France**
Telecom Paris and IBM Nov 2018 – Nov 2021
 - Subjects: Representation Learning for NLP and NLG, Evaluation of NLG, Information Theory
 - Funding: 3 years grant from IBM
- **Lab Associate** **Los Angeles, USA**
Disney Research 2018 (6 months)
 - Research Subject: Controlled text generation and text classification
- **Research Intern** **Zürich, Switzerland**
IBM Research 2017/2018 (6 months)
 - Research Subject: Deep Learning on Biological Graphs to find drugs efficient against prostate cancer
- **Research Student** **Lausanne, Switzerland**
Swisscom AI Lab 2017 (6 months)
 - Research Subject: Developing new technics for chatbots using information retrieval
- **Lab Assistant** **Lausanne, Switzerland**
Quantum Physic Department of EPFL 2017 (6 months)
 - Developing various sensors algorithms (Python/C++) for a quantum experiment on fermi gas

- Research Intern**

Procter & Gamble

 - Development of a secure application: Android & IOs
 - Developing low level functions to update a toothbrush via Bluetooth LTE

Frankfurt, Germany

2016 (3 months)
- Software Consultant**

Supelec & EPFL Junior Enterprise

 - Debugging web applications developed with Symphony 2 for the French Army
 - Developing several websites 📄

France, Switzerland

2014 - 2018

Publications

Awards.....










- ★ **Best Student Paper Award (Conference Acceptance Rate: 15%)**

Award for: *InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation.*






AAAI 2022

2022

Safe AI (Out Of Distribution | Adversarial Attacks | Hallucinations).....

- ♥  M Darrin, G Staerman, EDC Gomes, JCK Cheung, P Piantanida, **P. Colombo**. *Unsupervised Layer-wise Score Aggregation for Textual OOD Detection.*
- ♥  M Darrin, P Piantanida, **P. Colombo**. *Rainproof: An Umbrella To Shield Text Generators From Out-Of-Distribution Data.*
- ♥  NM Guerreiro, **P. Colombo**, P Piantanida, AFT Martins. *Optimal Transport for Unsupervised Hallucination Detection in Neural Machine Translation.*
- ♥  M Picot, N Noiry, P Piantanida, **P. Colombo**. *Adversarial Attack Detection Under Realistic Constraints*
- ♥  M Picot, G Staerman, F Granese, N Noiry, F Messina, P Piantanida, **P. Colombo**. *A Simple Unsupervised Data Depth-based Method to Detect Adversarial Images.*
- ♥  M. Picot, F. Granese, G. Staerman, F. Messina, M. Romanelli, P. Piantanida, **P. Colombo**. *A Halfspace-Mass Depth-Based Method for Adversarial Attack Detection.* **TMLR 2023**
- ♥  **P. Colombo**, G. Staerman, N. Noiry, P. Piantanida. *Beyond Mahalanobis Distance for Textual OOD Detection.* **NeurIPS 2022**
- ♣  **P. Colombo**, M. Picot, F. Granese, N. Noiry, G. Staerman, P. Piantanida. *Toward Stronger Textual Attack Detectors.*
- ♥  E. Gomes, **P. Colombo**, N. Noiry, G. Staerman, P. Piantanida. *A Functional Perspective on Multi-Layer Out-of-Distribution Detection.*





Large Model Finetuning (Fairness | Few-shot).....

- ♣  **P. Colombo***, V. Pellegrin*, M. Boudiaf, I. Ben Ayed, M. Tami, C. Hudelot P. Piantanida. *Textual Few-Shot For API-based Models*
- ♣  G. Pichler*, **P. Colombo***, M. Boudiaf*, G. Koliander, P. Piantanida. *KNIFE: Kernelized-Neural Differential Entropy Estimation.* **(Oral) ICML 2022**
- ♣  **P. Colombo**, G. Staerman, N. Noiry, P. Piantanida. *Learning Disentangled Textual Representations via Statistical Measures of Similarity.* **(Oral) ACL 2022**
- ♣  **P. Colombo**, C. Clavel and P. Piantanida. *A Novel Estimator of Mutual Information for Learning to Disentangle Textual Representations* **(Oral) ACL 2021**
- ♣  H. Jalalzai*, **P. Colombo***, C. Clavel, E. Gaussier, G. Varni, E. Vignon, and A. Sabourin. *Heavy-tailed representations, text polarity classification & data augmentation.* **NeurIPS 2020**




Evaluation of AI Systems.....

- ♣  **P. Colombo**, G. Staerman, N. Noiry, P. Piantanida. *What are the best Systems? New perspectives*










on NLP benchmarking. **NeurIPS 2022**

- ♥  C. Chung **P. Colombo**, F Suchanek , C. Clavel. *Of Human Criteria and Automatic Metrics: A Survey and Benchmark of the Evaluation of Story Generation*. **(Oral) COLLING 2022**
- ♣  **P. Colombo**, C. Clavel and P. Piantanida. *InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation*. **(Oral) AAAI 2022** **Granted the Outstanding Student Paper Award**
- ♣  **P. Colombo**, G. Staerman, C. Clavel and P. Piantanida. *Automatic Text Evaluation through the Lens of Wasserstein Barycenters*. **(Oral) EMNLP 2021**
- ♦  G. Staerman, P. Mozharovskiy, **P. Colombo**, S. Cl  men  on, F. d'Alch  -Buc, . *A Pseudo-Metric between Probability Distributions based on Depth-Trimmed Regions*.


Large Language Model.....

- ♥  NM Guerreiro, D Alves, J Waldendorf, B Haddow, A Birch, **P. Colombo**, A. Martins. *Hallucinations in Large Multilingual Translation Models*
- ♦  **P. Colombo** among 20+ authors. *The BigScience Corpus A 1.6 TB Composite Multilingual Dataset*. **(Oral) NeurIPS 2022**
- ♦  Over 400 authors with **P. Colombo**. *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model*

Conversational AI.....

- ♦  **P. Colombo** with over 80+ authors. *NL-augmenter: A framework for task-sensitive natural language augmentation*
- ♣  **P. Colombo***, E. Chapuis*, M. Labeau, and C. Clavel. *Code-switched inspired losses for generic spoken dialog representations*. **EMNLP 2021**
- ♣  **P. Colombo**, E. Chapuis, M. Labeau, and C. Clavel. *Improving multimodal fusion via mutual dependency maximisation*. **(Oral) EMNLP 2021**
- ♣  **P. Colombo**, C Yang, G. Varni, and C. Clavel. *Beam search with bidirectional strategies*. **ICNLSP 2021**
- ♣  E. Chapuis*, **P. Colombo***, M. Manica, M. Labeau, and C. Clavel. *Hierarchical pre-training for sequence labelling in spoken dialog*. **Findings of EMNLP 2020**
- ♣  T. Dinkar*, **P. Colombo***, M. Labeau, and C. Clavel. *The importance of fillers for text representations of speech transcripts*. **EMNLP 2020**
- ♣  **P. Colombo***, E. Chapuis*, M. Manica, E. Vignon, G. Varni, and C. Clavel. *Guiding attention in sequence-to-sequence models for dialogue act prediction*. **(Oral) AAAI 2020**
- ♣  A. Garcia*, **P. Colombo***, S. Essid, F. d'Alch  -Buc, and C. Clavel. *From the token to the review: A hierarchical multimodal approach to opinion mining*. **EMNLP 2019**
- ♣  **P. Colombo***, W. Witon*, A. Modi, J. Kennedy, and M. Kapadia. *Affect-driven dialog generation*. **NAACL 2019**

Others.....

- ♣ (PhD Thesis) **P. Colombo** *Learning to represent and generate text using information measures*
- ♣  W. Witon*, **P. Colombo***, A. Modi, and M. Kapadia. *Disney at IEST 2018: Predicting emotions using an ensemble*. In Proceedings of the 9th Workshop WASSA@EMNLP 2018

Patent.....

- ♣ Affect Driven Dialog Generation, A. Modi, M. Kapadia, Douglas A. Fidaleo, J. Kennedy, W. Witon and **P. Colombo**, US Patent 16,226,166, *A framework for Affective Conversational System*

Academic Duties

Teaching

- * **EPFL: Teaching Assistant** **30 hours**
Introduction to Information Theory 2016
- * **Telecom Paris: Teaching Assistant** **30 hours**
Telecom Paris: Introduction to Information Theory, Introduction To Machine Learning 2018 - 2021
- * **IP Paris: Associate Professor** **30 hours**
Advanced Natural Language Processing 2023
- * **Paris Saclay: Associate Professor** **200 hours**
Advanced NLP, Introduction to NLP, Student Project Mentoring 2023

Regularly Reviewing for:

✍ ARR, ACL, EMNLP, NAACL, AAAI, NeurIPS, ICML, ICLR

Academic Duties




- ♦ Nov 2019 - Telecom Paris - Workshop Organizer - Workshop on Machine Learning and Artificial Intelligence
- ♦ (Planned) Organization of ACL 2023 - Book Chair

Talks

- ☐ 🤖 **DCU Dublin April 2022:** What are the best systems? Revisiting evaluation of pretrained neural networks
- ☐ 🤖 **NMT Reading Group March 2022:** InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation
- ☐ 🤖 **Datacraft Paris March 2022:** Quel est le meilleur système? Evaluer les réseaux pré-entraînés.
- ☐ 🤖 **Telecom Paris Jan. 2021:** Learning generic dialog embeddings for sequence labeling tasks
- ☐ 🤖 **LEYA Research Dec. 2021:** Evaluation of Natural Language Generation using Measures of Similarity
- ☐ 🤖 **PhD Defense Nov. 2021:** Learning to Represent and Generate Text using Information Measures
- ☐ 🤖 **EMNLP (oral) Nov. 2021:** Automatic Text Evaluation through the Lens of Wasserstein Barycenters
- ☐ 🤖 **EMNLP (oral) Nov. 2021:** Improving multimodal fusion via mutual dependency maximisation
- ☐ 🤖 **WACAI (oral) Oct. 2021:** Hierarchical Pre-training for Sequence Labelling in Spoken Dialog
- ☐ 🤖 **WACAI (oral) Oct. 2021:** The importance of fillers for text representations of speech transcripts
- ☐ 🤖 **ACL (oral) July. 2021:** A Novel Estimator of Mutual Information for Learning to Disentangle Textual Representations
- ☐ **ALPS Jan. 2021:** Learning to disentangle textual representations and attributes via mutual information
- ☐ 🤖 **NeurIPS Dec. 2020:** Heavy-tailed Representations, Text Polarity Classification & Data Augmentation
- ☐ 🤖 **EMNLP Dec. 2020:** The importance of fillers for text representations of speech transcripts
- ☐ 🤖 **Mid-Phd Defense Oct. 2020:** Methods to represent natural language with application to conversational AI
- ☐ **MLSS July. 2020:** Heavy-tailed Representations, Text Polarity Classification & Data Augmentation
- ☐ **AAAI (oral) Feb. 2020:** Guiding attention in Sequence-to-sequence models for Dialogue Act prediction
- ☐ **EMNLP Dec. 2019:** From the Token to the Review: A Hierarchical Multimodal approach to Opinion Mining
- ☐ **Journée de la chaire DSAIDIS Oct. 2019:** From the Token to the Review: A Hierarchical Multimodal approach to Opinion Mining

(Some) Open Source Contributions

- Contribution to HuggingFace's Dataset: 🤖 SILICONE (10 datasets), 🤖 MIAM (5 datasets)
- 🤖 NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation
- 🤖 TORNADO: TOWards faiR aNd debiAseD representatiOns for NLP

-  NLGEval: Evaluation of NLG
-  RankEval: Evaluation of NLP Systems
-  KNIFE: Estimation of Mutual Information

Miscellaneous

Languages: English (fluent), French (native), German (upper intermediate)

Sport: Athletics (marathons, ultra-trails), Handball, Tennis

Volunteer Work: Board Member from the running association Les Étoiles du 8ème (Paris 8)
: In charge of a Student Group which helps youths living in disadvantaged areas.

Music: Studied music theory and trumpet for 12 years at conservatoire