

Pierre Colombo

Born the 15.07.1994 at Corneilles en Parisis (95240)

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

Websites: [!\[\]\(666e09182d4cd268646ea700ea60dcdf_img.jpg\)](#) [!\[\]\(1ef1ef0bf9af6c6996401964cf280f2d_img.jpg\)](#) [!\[\]\(e9a80c8557f9285916925bd4ac40fff5_img.jpg\)](#) [!\[\]\(88e2edecff3400e68a80dd08c57d2f9c_img.jpg\)](#)

I am a postdoctoral student working with [Pablo Piantanida](#) on Safe AI. My research focuses on Natural Language Processing and its applications.

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

- **Post Doctoral Researcher** **Paris France**
CNRS-L2S (CentraleSupélec), Main subject: Safe AI Nov 2021
- Main Advisors: Pablo Piantanida
- **Phd Student** **Paris France**
Telecom Paris and IBM Nov 2018 – Nov 2021
- Main Advisors: Chloé Clavel, Pablo Piantanida, Mathieu Labeau.
- 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.
- Master Thesis entitled : Deciphering gene deregulation in cancer development using deep learning.
- **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** **Frankfurt, Germany**
Procter & Gamble 2016 (3 months)
- Development of a secure application: Android & IOs.
- Developing low level functions to update a toothbrush via Bluetooth LTE.


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

Publications







Conference.....

1.  **P. Colombo**, C. Clavel and P. Piantanida. InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation. **AAAI 2021**
2.  **P. Colombo**, C Yang, G. Varni, and C. Clavel. Beam search with bidirectional strategies. **ICNLSP 2021**
3.  **P. Colombo***, E. Chapuis*, M. Labeau, and C. Clavel. Code-switched inspired losses for generic spoken dialog representations. **EMNLP 2021**
4.  **P. Colombo**, G. Staerman, C. Clavel and P. Piantanida. Automatic Text Evaluation through the Lens of Wasserstein Barycenters. **(Oral) EMNLP 2021**
5.  **P. Colombo**, E. Chapuis, M. Labeau, and C. Clavel. Improving multimodal fusion via mutual dependency maximisation. **(Oral) EMNLP 2021**
6.  **P. Colombo**, C. Clavel and P. Piantanida. A Novel Estimator of Mutual Information for Learning to Disentangle Textual Representations **(Oral) ACL 2021**
7.  E. Chapuis*, **P. Colombo***, M. Manica, M. Labeau, and C. Clavel. Hierarchical pre-training for sequence labelling in spoken dialog. **Findings of EMNLP 2020**
8.  T. Dinkar*, **P. Colombo***, M. Labeau, and C. Clavel. The importance of fillers for text representations of speech transcripts. **EMNLP 2020**
9.  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**
10.  **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**
11.  A. Garcia*, **P. Colombo***, S. Essid, F. d'Alché-Buc, and C. Clavel. From the token to the review: A hierarchical multimodal approach to opinion mining. **EMNLP 2019**
12.  **P. Colombo***, W. Witon*, A. Modi, J. Kennedy, and M. Kapadia. Affect-driven dialog generation. **NAACL 2019**

Workshop.....

1.  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

Preprint.....

1.  **P. Colombo** and over 30 co-authors. NL-Augmenter. Submitted at **NAACL 2022**
2.  G. Staerman, P. Mozharovskiy, **P. Colombo**, S. Cléménçon, F. d'Alché-Buc, . A Pseudo-Metric between Probability Distributions based on Depth-Trimmed Regions. Submitted at **ICML 2022**
3.  G. Pichler*, **P. Colombo***, M. Boudiaf*, G. Koliander, P. Piantanida. KNIFE: Kernelized-Neural Differential Entropy Estimation. Submitted at **ICML 2022**
4.  **P. Colombo**, N. Noiry, Ekhine Irrozki, Stephan Cléménçon, What are the best systems? New perspectives on NLP Benchmarking. Submitted at **ICML 2022**
5.  **P. Colombo**, N. Noiry, G. Staerman, P. Piantanida. Learning Disentangled Textual Representations via Statistical Measures of Similarity. Submitted at **ACL 2022**
6.  **P. Colombo**, N. Noiry, G. Staerman, P. Piantanida. Learning to Disentangle Representations via Mutual Information Maximization. Submitted at **NAACL 2022**

Patent

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

Other Research Activities

Talks

Speaker

2019 -

- 🗣️ (Planned) DCU Dublin April 2022: To be defined.
- 🗣️ (Planned) Datacraft Paris March 2022: To be defined.
- 🗣️ 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 (oral) Nov. 2021:
- 🗣️ 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
- 🗣️ AACL (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

Open source

Contributor

2021 -

- Contribution to 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.

PhD Mentoring

- 👤 Marine Picot, 👤 Federica Granese, 👤 Eduardo Gomes, 👤 Malik Boudiaf, 👤 Cyril Chung.

2021 -

Intern Mentoring

- Leila Bassou, IBM France: Research Internship on Email Classification using BERT.

2018 - 2019

Teaching Assistant

- EPFL - Telecom Paris: MDI341, SI221/MICAS911, SI221, CS-110

2016 -

Summer & Winter School

Participant

2020 - 2021

- Advanced Language Processing School (ALPS 2021)
- Machine Learning Summer School (MLSS 2020)

Services

- Reviewing, ACL Rolling Review, ACL, EMNLP, NAACL, AACL.

2020 -

Organizer

- *Telecom Paris, Workshop on Machine Learning and Artificial Intelligence*

Nov 2019

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