

Pierre Colombo

Paris 75013 - France

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

Websites: 🌐   

I am a last year PhD student in the SSA group at Telecom Paris and IBM. I am looking for a **researcher position on the field of natural language processing**. I am open to relocation.

Education

- **PhD in Deep Learning** **Telecom Paris, Institut Polytechnique de Paris**
Paris, France 2018 – 2021
- **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

- **Phd Student** **Paris France**
Telecom Paris and IBM Nov 2018 – Nov 2021
 - Main Advisors: Chloé Clavel, Pablo Piantanida, Mathieu Labeau.
 - Keywords: Representation Learning for Natural Language, Natural Language Generation (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.
- **Software Consultant** **France, Switzerland**
Supelec & EPFL Junior Enterprise 2014 - 2018


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

Publications




Conference.....

1.  **P. Colombo***, E. Chapuis*, M. Labeau, and C. Clavel. Code-switched inspired losses for generic spoken dialog representations. **EMNLP 2021**
2.  **P. Colombo**, G. Staerman, C. Clavel and P. Piantanida. Automatic Text Evaluation through the Lens of Wasserstein Barycenters. **EMNLP 2021**
3.  **P. Colombo**, E. Chapuis, M. Labeau, and C. Clavel. Improving multimodal fusion via mutual dependency maximisation. **EMNLP 2021**
4.  **P. Colombo**, C. Clavel and P. Piantanida. A Novel Estimator of Mutual Information for Learning to Disentangle Textual Representations (**Oral**) **ACL 2021**
5.  E. Chapuis*, **P. Colombo***, M. Manica, M. Labeau, and C. Clavel. Hierarchical pre-training for sequence labelling in spoken dialog. **Findings of EMNLP 2020**
6.  T. Dinkar*, **P. Colombo***, M. Labeau, and C. Clavel. The importance of fillers for text representations of speech transcripts. **EMNLP 2020**
7.  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**
8.  **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**
9.  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**
10.  **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.  Georg Pichler*, **P. Colombo***, Malik Boudiaf*, Günther Koliander, Pablo Piantanida. KNIFE: Kernelized-Neural Differential Entropy Estimation. Submitted at **NeurIPS 2021**
2.  **P. Colombo**, C. Clavel and P. Piantanida. InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation Submitted at **TACL**
3.  **P. Colombo**, C Yang, G. Varni, and C. Clavel. Beam search with bidirectional strategies. 2020

Patent.....

1. 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

Other Research Activities

- **Workshop on Machine Learning and Artificial Intelligence** **Telecom Paris**
Organizer *Nov 2019*
- **Summer & Winter School** **Virtual**
Participant *2020 - 2021*
 - Advanced Language Processing School (ALPS)(2021)
 - Machine Learning Summer School (MLSS) (2020)

- **Talks** **Various Locations**
2019 - 2021
 - **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 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** **Hugging Face Library**
2021
 - *Contributor*
 - Contribution to Datasets: SILICONE (10 datasets), MIAM (5 datasets)
- **Intern Mentoring** **IBM France**
2018 - 2019
 - *Leila Bassou, Research Internship on Email Classification using BERT.*
- **Teaching Assistant** **EPFL - Telecom Paris**
2016 - Present
 - *MDI341, SI221/MICAS911, SI221, CS-110*

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