Clément Rebuffel

clement.rebuffel@pm.me

I am a **Quantitative Researcher** at **G-Research** (London, UK), and a final year **PhD** student at **Sorbonne Université** (Paris, France). At work, I research systematic trading ideas to predict the future of financial markets using large text corpora. My academic research includes all things **Natural Language Generation**, especially from structured data (e.g. tables, graphs, etc.). I have fruitful collaborations with University of Turin (Italy) and University of Aberdeen (UK). All projects (solos & duos) are available on github: https://github.com/KaijuML.

Skills

Lang. **French** (native) **English** (fluent) **Spanish** (I can read books and understand movies).
Prog. **Python** (5+ years) (Pytorch 🔥 HuggingFace 🤗 Pytorch-Lightning ⚡)
Web **Flask** **Javascript** (Vue.js) **Netlify** **Azure Functions**

Work Experience

2021+ **G-Research - London, UK**
Quantitative Researcher, part of the NLP team. Working on predicting financial markets using large textual corpora.

2017–2021 **BNP Paribas - Paris, France**
Data Scientist and Researcher. Currently working with the AI research team of Data&AI Lab, as part of a CIFRE contract (aka Industry PhD). Part of the team who launched the internal machine learning-powered search engine. Areas of research include Natural Language Generation, all things NLP and Information Retrieval. Advisory capacity: help team select the best technology for a task, internal reviews of papers.

Education

2018-2021 **Sorbonne Université**
PhD on *Deep Learning for Data-to-Text Generation*. Learning better representations of complex structured data, disentangling factual/incorrect statements in reference texts, producing reliable and controllable NLG systems.
Supervisors: Laure Soulier and Patrick Gallinari

2019 **RLSS**
Summer school on Reinforcement Learning with the SequeL team.

2016-2017 **École Normale Supérieure**
MSc (final year): Mathematics, Statistics and Machine Learning.
Convex Optimization, Kernel Methods, Probabilistic Graphical Models, Monte Carlo Markov Chains, Statistical Learning, etc.

2012-2016 **Université Paris-Dauphine**
MSc (1st year): Mathematics and Statistics.
Bachelor's degree: Mathematics, Finance and Programming.

Publications & Academia

2021 **C. Rebuffel**, T. Scialom, L. Soulier, B. Piwowarski, S. Lamprier, J. Staiano, G. Scoutheeten, and P. Gallinari. Data-QuestEval: A Referenceless Metric for Data-to-Text Semantic Evaluation. *2021*.

2021 **C. Rebuffel**, M. Roberti, L. Soulier, R. Canceliere, G. Scoutheeten, and P. Gallinari. Controlling Hallucinations at Word Level in Data-to-Text Generation. *DMKD 2021*.

2020 **C. Rebuffel**, L. Soulier, G. Scoutheeten, and P. Gallinari. PARENTing via Model-Agnostic Reinforcement Learning to Correct Pathological Behaviors in Data-to-Text Generation. *INLG 2020*.

2020 **C. Rebuffel**, L. Soulier, G. Scoutheeten, and P. Gallinari. A Hierarchical Model for Data-to-Text Generation. *ECIR 2020*.

2020 **C. Rebuffel**, L. Soulier, G. Scoutheeten, and P. Gallinari. Capturing Entity Hierarchy in Data-to-Text Generative Models. *CIRCLE 2020*.

Reviews NeurIPS (top 10% of high-scoring reviewers), ACL, SIGIR

Competitions

2021 **Google CodeJam**: Up to round 2
Multiple rounds of online algorithmic coding puzzles.

2020 **Google HashCode**: Top 5% world (30/787 France)
Team programming competition on a complex engineering problem.

2017 **ENS Data Challenge**: 1st
Predicting the age of patients from EEGs.