

# CLÉMENT REBUFFEL

## CONTACT INFORMATION

---

ADDRESS: 12 rue Anatole de la Forge, 75017 Paris - France

EMAIL: [clement.rebuffel@lip6.fr](mailto:clement.rebuffel@lip6.fr)

## PERSONAL STATEMENT

---

I am a third year **PhD** student at **Sorbonne Université**, and **Research Engineer** at **BNP Paribas**. My research includes all things **Natural Language Generation**, especially from structured data (e.g. tables, graphs, etc.). My work has focused both on encoding different modalities in order to learn better representations of complex data, as well as ensuring factualness in system outputs (constraining for factual outputs and evaluation of factualness).

I am a good thinker with a curious mind and a strong work ethic. In the last year of my PhD I have focused on working with other PhD students, with notably fruitful collaborations with University of Turin (Italy) and University of Aberdeen (UK). All projects (solos & duos) are available on github: <https://github.com/KaijuML>

On a personal note, I am a climbing enthusiast and used to run/swim at least once per week before COVID. I greatly enjoy storytelling, both reading and going to the movies (used to go twice a week w/ movie pass). I'm also a fan of cooking: meals, deserts, as well as **cocktails**!

## PRACTICAL SKILLS

---

Spoken languages	<b>French</b> (fluent); <b>English</b> (fluent); <b>Spanish</b> (I can read books and understand movies).
Machine Learning	<b>Python</b> 🐍 <b>Pytorch</b> 🔥 <b>HuggingFace</b> 😊 <b>Pytorch-Lightning</b> ⚡
Web design	<b>Flask</b> ; <b>Javascript</b> (Vue.js); <b>Netlify</b> ; <b>Azure Functions</b>
Misc.	<b>Git</b> ; <b>LaTeX</b> ; <b>Linux</b> ; <b>ElasticSearch</b> ; <b>MongoDB</b>

## EDUCATION

---

2018-2021	<b>Sorbonne Université</b> PhD on <i>Deep Learning for Data-to-Text Generation</i> . I work on learning better representations of complex structured data, and disentangling factual/incorrect statements in reference texts, in order to produce reliable NLG systems. Supervisors: <a href="#">Laure Soulier</a> and <a href="#">Patrick Gallinari</a>
2016-2017	<b>École Normale Supérieure</b> MSc (final year): Mathematics, Statistics and Machine Learning. Convex Optimization, Kernel Methods, Probabilistic Graphical Models, Monte Carlo Markov Chains, Statistical Learning, etc.
2012-2016	<b>Université Paris-Dauphine</b> MSc (1 <sup>st</sup> year): Mathematics and Statistics. Bachelor's degree: Mathematics, Finance and Programming.

## WORK EXPERIENCE

---

2018-2021	<b>BNP Paribas - Paris, France</b> Data Scientist and Researcher. Currently working with the AI research team of Data&AI Lab, as part of a CIFRE contract. Areas of research include Natural Language Generation, all things NLP and Information Retrieval. Advisory capacity to help my team select the best technology for a task, and do internal reviews of papers.
2017	<b>BNP Paribas - Paris, France</b> Research intern. Part of the team who launched the internal machine learning-powered search engine. I worked on diverse NLP tasks, including Grammatical Error Correction, and Information Retrieval.
2016	<b>R2E - Paris, France</b> Actuary intern; R2E is a subsidiary of AXA and Humanis. Supervised the creation of a platform aimed at accompanying the subscription process and centralising contractual commitments.
2013-2015	<b>Nice, France</b> Waiter/Salesman for three summers; First full time work experiences.

## PUBLICATIONS & ACADEMIA

---

C. Rebuffel, M. Roberti, L. Soulier, R. Cancelliere, G. Scoutheeten, and P. Gallinari. [Controlling Hallucinations at Word Level in Data-to-Text Generation](#). *In submission, Data Mining and Knowledge Discovery (DMKD) & European Conference on Machine Learning (ECML) 2021*.

C. Rebuffel, L. Soulier, G. Scoutheeten, and P. Gallinari. [PARENTing via Model-Agnostic Reinforcement Learning to Correct Pathological Behaviors in Data-to-Text Generation](#). *International Conference on Natural Language Generation (INLG) 2020*.

C. Rebuffel, L. Soulier, G. Scoutheeten, and P. Gallinari. [A Hierarchical Model for Data-to-Text Generation](#). *European Conference on Information Retrieval (ECIR) 2020*.

C. Rebuffel, L. Soulier, G. Scoutheeten, and P. Gallinari. [Capturing Entity Hierarchy in Data-to-Text Generative Models](#). *Joint Conference of the Information Retrieval Communities in Europe (CIRCLE) 2020*.

**Reviewer:** NeurIPS (top 10% of high-scoring reviewers), ACL, SIGIR