

# SANDRA BURLAUD

NLP Data Scientist

I am a data scientist specialised in NLP applications looking for a remote or hybrid role in London.

## INFO

✉ [burlaudsandra@gmail.com](mailto:burlaudsandra@gmail.com)

📍 London

🌐 **Citizenship: France / USA**

**Pre-settled status in UK**

**(no need for visa)**

## EDUCATION

**University of Edinburgh**  
**College of Arts, Humanities, and Social Sciences**

MSc. Speech and Language Processing

*Graduated with Merit*

**New York University**  
**Courant Institute of Mathematical Sciences**

BA in Computer Science  
Minors: Creative Writing, Psychology

*Graduated Magna Cum Laude*

## SKILLS & LANGUAGES

- Fluent in English, French. Basic Spanish.
- Experienced with Python, shell scripting/bash, and standard NLP/ML tools & libraries
  - numpy, pandas, sklearn, pydantic, spacy, MLFlow, huggingface, OpenAI
- Experienced with large language models
  - **Databricks' LLM Professional Certificate (2023)**
  - BERT based models, GPT models
  - LLM fine-tuning, prompt engineering, risk/bias assessment and mitigation
- Experienced with other ML architectures (CNNs, RNNs, LSTMs, GANs)
- Experienced with git, AWS, Docker, Flyte
- Familiar with:
  - ML model optimisation techniques (onnx quantisation and optimisation)
  - ML model deployment (MLServer)
  - Azure, Elasticsearch, GraphQL, Streamlit
- Basics in Java, Go, R, C, SQL

## WORK EXPERIENCE

DATA SCIENTIST

**Adarga**

*October 2021 - Present*

- Developed efficient stateless **REST APIs** for various NLP tasks (extraction of events, claims, properties, entities, relationships, etc.)
- Developed **LLM-based** information extraction workflows (Flyte)
- Prototyped, developed, trained, deployed NLP machine learning models as part of product team
- Worked with **unstructured text data** (data cleaning, formatting)
- Collaborated with product, customer success, and engineering teams to translate customer requests into technical solutions
- Prioritised iterative development over premature optimisation
- Supported MLOps team with model deployment
- **Unit testing**
- **Notable achievements:**
  - improved BERT based event extraction model speed by 60%
  - improved tense detection model from .31 to .72 F1 score
  - introduced new process for synthesis dataset generation using OpenAI models
  - managed team of six interns
  - mentored UEdinburgh Masters student
  - Mental Health First Aiders training

MASTERS DISSERTATION: SYNTHESIS OF CHILD SPEECH  
**in collaboration with SpeakUnique**

*June 2021 - August 2021*

- Among the first projects to apply deep neural networks to synthesis of child speech
- Used Tacotron2 to train child speed models, both from scratch, and fine-tuned on pre-trained adult models; single and multi-speaker
- Dataset preprocessing, resegmenting recordings

RESEARCH INTERN: TERMINOLOGY EXTRACTION

**New York University Computer Science Department**

*June 2019 - August 2020*

- Project originally in collaboration with US Dept. of Defense
- Adapted open-source terminology extraction program (Termolator, Meyers et al 2015) to work on French corpora
- Used a combination of knowledge-based and statistical techniques to measure term frequency and rank terms
- 85% accuracy against manually annotated documents