

Jean-Baptiste LAMARE

Senior Research Scientist | Deep Learning and NLP

[in linkedin.com/in/jblamare](https://www.linkedin.com/in/jblamare) github.com/jblamare [Google Scholar](https://scholar.google.com/citations?user=...) [Personal Website](https://www.jblamare.com)
☎ 412-626-8707 @ jblamare@gmail.com 📍 San Francisco, CA 🇫🇷 French citizen, H1-B visa

PROFESSIONAL EXPERIENCE

Present August 2019	SENIOR RESEARCH SCIENTIST (April 2021 - Present) RESEARCH SCIENTIST (August 2019 - April 2021) ENLITIC, SAN FRANCISCO, CA <ul style="list-style-type: none">> Parsed and analyzed millions of EHR and DICOM data points to extract insightful features.> Developed supervised NLP models for multiclass/multilabel classification, NER, and information retrieval, using various architectures such as transformers (BERT, XLNET...) and RNNs (LSTMs, GRUs...).> Trained specific Language Models for clinical NLP.> Created rule-based and unsupervised NLP models for entity recognition and entity linking.> Implemented several NLP pipelines for free text report structuring.> Trained supervised computer vision models for abnormality detection on radiology images, as well as multimodal models at the intersection of CV and NLP.> Maintained a web-based application for clinical report labeling.> Designed end-to-end labeling pipelines for radiology reports and images, including interface design, case selection, quality assurance, inter-annotator agreement.> Developed and maintained several shared repositories for deep learning model training and deployment, as well as python package creation.> Led a successful blind test of our models on a client site.> Submitted and published several papers at peer-reviewed conferences.
August 2019 August 2017	GRADUATE RESEARCH ASSISTANT CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA <ul style="list-style-type: none">> Collected and labeled a dataset of tweets during public safety events.> Trained Language Models on the collected dataset to understand and capture the meaning and noise of Twitter data.> Developed a Deep Learning useful tweets extraction system for public safety events.> Designed and implemented a web interface for users to visualize extracted tweets in real time and provide feedback on which tweets are actually useful.> Created an active learning paradigm to fine-tune the model according to the user feedback.> Collected a dataset and implemented models to detect accidents on traffic camera videos.> Published two papers at peer-reviewed conferences.
August 2017 April 2017	MACHINE LEARNING INTERN SRI INTERNATIONAL, MENLO PARK, CA <ul style="list-style-type: none">> Collected a dataset of ransomware related paragraphs through web scraping and mechanical turk.> Developed and compared different question answering models on the dataset.

PUBLICATIONS

ON THE DIMINISHING RETURN OF LABELING CLINICAL REPORTS Paper EMNLP presentation Jean-Baptiste Lamare, Tobi Olatunji, Li Yao Deep Learning NLP Clinical Tensorflow Data labeling	EMNLP 2020 - CLINICAL NLP WORKSHOP
ACCIDENT FORECASTING IN CCTV TRAFFIC CAMERA VIDEOS Paper Data github.com/ankitshah009/CarCrash_forecasting_and_detection Ankit Shah, Jean-Baptiste Lamare, Tuan Nguyen Anh, Alexander Hauptmann Deep Learning CV Tensorflow Data collection Data labeling	AVSS 2018 - T4S WORKSHOP
MULTIMODAL FILTERING OF SOCIAL MEDIA FOR TEMPORAL MONITORING AND EVENT ANALYSIS Paper Po-Yao Huang, Junwei Liang, Jean-Baptiste Lamare, Alexander Hauptmann Deep Learning NLP CV Multimodal Social media Tensorflow Data collection Data labeling	ICMR 2018

EDUCATION

- August 2019 **Carnegie Mellon University, School of Computer Science (Pittsburgh, PA)**
Master of Language Technologies - Natural Language, Computer Vision, Audio Processing
Selected Coursework : *Algorithms for NLP, Grammars and Lexicons, Search Engines, Neural Networks for NLP (Teaching Assistantship), Deep Learning, Multimodal ML*
- August 2017 **Ecole Polytechnique, Department of Computer Science (Palaiseau, France)**
Master of Data Science - Bachelor of Science
Selected Coursework : *Machine Learning, Big Data, Databases, Text Mining and NLP Linear Regression, Markov Chains, Probabilities, Mechanics, Economics*

PROJECTS

BLOOM

JANUARY 2019 - MAY 2019


Carnegie Mellon University (Pittsburgh, PA) -  github.com/TevenLeScao/bloom

Created a tool to create aesthetic visualizations of a neural network's training process.

Deep Learning Machine Translation Fractals

KICK THE DRUMMER OUT

JANUARY 2019 - MAY 2019


Carnegie Mellon University (Pittsburgh, PA) -  github.com/jblamare/Beat-generator

Created an LSTM model to generate drum tracks, in a polyphonic variation of a Language Model.

Deep Learning Audio processing LSTM PyTorch

RORSCHACH GENERATION

JANUARY 2019 - MAY 2019

Carnegie Mellon University (Pittsburgh, PA) -  github.com/craigastewart/rorschach

Trained GANs to generate imitation Rorschach ink blots from photographs.

Deep Learning CV GAN PyTorch

URBAN JUNGLE

JANUARY 2019 - MAY 2019

Carnegie Mellon University (Pittsburgh, PA) -  github.com/TevenLeScao/artXml

Developed a tool transforming a picture using style transfer and Deep Dream.

Deep Learning CV Style Transfer Deep Dream PyTorch

EMBODIED QUESTION ANSWERING (EQA)

AUGUST 2018 - DECEMBER 2018

Carnegie Mellon University (Pittsburgh, PA)

Improved the Visual Question Answering (VQA) and navigation metrics on the EQA task.

Deep Learning NLP CV Multimodal VQA PyTorch

A NEURAL MUSIC CRITIC

JANUARY 2018 - MAY 2018

Carnegie Mellon University (Pittsburgh, PA) -  github.com/jblamare/Song-Review-Generation

Created a model that automatically writes a review from the audio file of a song

Deep Learning NLP Audio processing Text generation PyTorch

TURNING DAY INTO NIGHT

JANUARY 2018 - MAY 2018

Carnegie Mellon University (Pittsburgh, PA) -  github.com/AmosLewis/Image2Image_Style_transfer_pytorch

Trained a model converting a daylight picture into the same image at night

Deep Learning CV GAN PyTorch

SKILLS

Fields	Deep Learning, NLP, Computer Vision, Multimodal Learning
Programming Languages	Python, Java, C++, R, SQL, HTML, Javascript, CSS
Machine Learning frameworks	PyTorch, Tensorflow, Spacy, Stanza, Scikit-learn, Pandas, Numpy
Tools/Technology	Git, Mechanical Turk, Lucene, MapReduce, Scilab, Raspberry
Languages	English (fluent), French (native), German (intermediate), Spanish (beginner)

MISCELLANEOUS

Hobbies	Competitive tennis and skiing, guitar, piano, improvisational theater, chess, traveling
Student activities	Ski team (Team captain), Tennis club, Job Fair Organizer (Head of the Industry unit)