Jean-Baptiste LAMARE

Senior Research Scientist | Deep Learning and NLP

in linkedin.com/in/jblamare ⚬ github.com/jblamare 🎓 Google Scholar ↗ Personal Website
📱 412-626-8707 @ jblamare@gmail.com 📍 San Francisco, CA ℹ French citizen, H1-B visa

💼 PROFESSIONAL EXPERIENCE

Present **August 2019**	**Senior Research Scientist** (April 2021 - Present) ENLITIC, SAN FRANCISCO, CA **Research Scientist** (August 2019 - April 2021)

> Parsed and analyzed millions of **EHR and DICOM data** 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

> Developed a **useful tweets extraction** system for public safety events (dataset and models).
> Collected a dataset and implemented models to **detect accidents** on traffic camera videos.

August 2017 **April 2017**	**Machine Learning Intern** SRI INTERNATIONAL, MENLO PARK, CA

> Collected a dataset of ransomware related paragraphs through web scraping and mechanical turk.
> Developed and compared different question answering models on the dataset (including LSTMs).

📖 PUBLICATIONS

ON THE DIMINISHING RETURN OF LABELING CLINICAL REPORTS EMNLP 2020 - CLINICAL NLP WORKSHOP
↗ Paper ↗ EMNLP presentation
Jean-Baptiste Lamare, Tobi Olatunji, Li Yao
`Deep Learning` `NLP` `Clinical` `Tensorflow` `Data labeling`

ACCIDENT FORECASTING IN CCTV TRAFFIC CAMERA VIDEOS AVSS 2018 - T4S WORKSHOP
↗ 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`

MULTIMODAL FILTERING OF SOCIAL MEDIA FOR TEMPORAL MONITORING AND EVENT ANALYSIS ICMR 2018
↗ Paper
Po-Yao Huang, Junwei Liang, Jean-Baptiste Lamare, Alexander Hauptman
`Deep Learning` `NLP` `CV` `Multimodal` `Social media` `Tensorflow` `Data collection` `Data labeling`

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

January 2019 - May 2019

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

`Deep Learning` `Machine Translation` `Fractals`

Kick the drummer out
Carnegie Mellon University (Pittsburgh, PA) - ⭕ github.com/jblamare/Beat-generator

January 2019 - May 2019

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

`Deep Learning` `Audio processing` `LSTM` `PyTorch`

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

January 2019 - May 2019

Trained GANs to generate imitation Rorschach ink blots from photographs.

`Deep Learning` `CV` `GAN` `PyTorch`

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

January 2019 - May 2019

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

`Deep Learning` `CV` `Style Transfer` `Deep Dream` `PyTorch`

Embodied Question Answering (EQA)
Carnegie Mellon University (Pittsburgh, PA)

August 2018 - December 2018

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

`Deep Learning` `NLP` `CV` `Multimodal` `VQA` `PyTorch`

A Neural Music Critic
Carnegie Mellon University (Pittsburgh, PA) - ⭕ github.com/jblamare/Song-Review-Generation

January 2018 - May 2018

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

January 2018 - May 2018

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

`Deep Learning` `CV` `GAN` `PyTorch`

Fact-Checking application in collaboration with IBM
Ecole Polytechnique (Palaiseau, France)

August 2016 - March 2017

Developed a tool determining a statement's veracity by searching through a news database.

`NLP` `Text Mining` `IBM Watson`

Mine-hunting drone : awarded the "Prix des PSC" prize (top 3 out of 81)
Ecole Polytechnique (Palaiseau, France)

August 2015 - May 2016

Designed and constructed a mine detection system, embedded on a flying drone

`Data processing` `Java`

≣ 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)