5000 Forbes Avenue
Pittsburgh PA, USA 15213
+1 412-626-8660
rolivier@cs.cmu.edu
Github: RaphaelOlivier
Website: raphaelolivier.github.io

Raphaël Olivier

I am a PhD candidate at CMU working with prof. Bhiksha Raj on Robust Speech Representations, Secure and Trustworthy Machine Learning, Adversarial threats and defenses and Data Privacy.

Education

2019-2023 **Carnegie Mellon University**, *Ph.D in Language Technologies*, Language Technologies Institute

2017–2019 **Carnegie Mellon University**, *M.S. in Language Technologies*, Language Technologies Institute

2014–2017 **École Polytechnique**, *Ingénieur Program*, Math & CS, Paris. I ranked 45th at national entrance exam

Highlighted Research Projects

2023 **Applications and Security of Large Language Models and Diffusion models**

- Recent project on finetuning LLMs (T5, LLaMA) for spoken tasks using PEFT and audio feature extraction
- I am also studying the vulnerabilities of LLMs (closed and open-source) to adversarial attacks
- I am investigating applications of adversarial attacks to watermark the outputs of diffusion models

2019-2023 **Thesis project: Attacks and Defenses on Speech Recognition**

- I designed white-box and black-box attacks that can fool Speech recognition models (Whisper, Wav2vec2, HuBERT, WavLM, etc.) into transcribing any target, or leak information about their training data. Work published at **InterSpeech 2022** and **InterSpeech 2023**, two more articles under review.
- I proposed smoothing with speech enhancement and adversarial training-based defenses for ASR against adversarial attacks. Work published at **ICASSP 2021** and **EMNLP 2021**, one more article under review.
- I proposed *adversarial sparsity*, a novel metric to evaluate adversarial robustness. Accepted at **ICML 2023**
- I released **robust_speech**, an open-source framework for evaluating the robustness of speech models.
- I gave invited talks on my thesis work at the SPSC webinar and the Technion Machine Learning seminar (2022).
- Tutorial on adversarial attacks for speech at **InterSpeech 2019** with profs. Bhiksha Raj and Yossi Keshet.

Jan 2018 - Nov 2018 **Neural code generation**, with Prof. Graham Neubig

- We trained a then-state-of-the-art LSTM encoder-decoder model for code generation, using machine-translation inspired retrieval methods. Work published at **EMNLP 2018**.

Experience

June-Aug 2021 **Applied Scientist Intern**, Amazon Alexa, Pittsburgh, PA
I designed mitigation techniques against backdoor poisoning attacks for Alexa's Speech Recognition models.

June-Aug 2020 **Applied Scientist Intern**, Amazon Alexa, Pittsburgh, PA
I worked on privacy and membership inference attacks and defenses on Alexa's Speech Recognition models

Apr–Aug 2017 **Research Intern**, AgroParisTech, Paris, France, mentored by prof. Antoine Cornuejols
Transfer Learning for time series using AdaBoost. Work published at the Symposium on Intelligent Data Analysis

June-Aug 2016 **Data Scientist Intern**, DataScienTest, Paris, France
Participated in the creation of DataScienTest, a leading online training platform for Data Scientists.

Skills and Coursework

Languages: Python, C/C++, Java, SQL, Bash

Frameworks: PyTorch, Tensorflow, Numpy, Pandas, Scikit-Learn, HuggingFace, ESPNet, Fairseq, SpeechBrain

Models: Whisper, Wav2Vec2, Data2Vec, Encodec, RNN-T, MLMs, T5, LongT5, ViT, StableDiffusion

ML: NLP, Deep Learning (**TA** in 2018/2019), Advanced ML, Multimodal ML, Speech Recognition

CS: Algorithms, Advanced Programming, Data Management, Computational Geometry

Math: Logic, Algebra, Number Theory, Analysis, Optimization, Differential Equations, Sequences/Series