raphaelolivier.github.io

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

# Raphaël Olivier

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

## Education

2019–Ongoing **Carnegie Mellon University**,
*Ph.D in Language Technologies*, Language Technologies Institute
Thesis on the security and robustness of Speech representations, advised by prof. Bhiksha Raj

2017–2019 **Carnegie Mellon University**,
*M.S. in Language Technologies*, Language Technologies Institute
GPA 4.0

2014–2017 **École Polytechnique of Paris**,
*Applied Mathematics and Computer Science*, Ingénieur Polytechnicien Program
This program is the most selective Science and Technology degree in France, with core courses in Mathematics, Sciences and Humanities

2012–2014 **Classes Preparatoires**,
*Math, Physics and Computer Science*, Lycée Pasteur

- Two years of intensive training for nationwide entrance exams to French Grandes Écoles
- Ranked 1st/40 all two years
- Ranked 30th to 60th nationwide at 4 competitive entrance exams

## Experience

June-Aug 2021 **Applied Scientist Intern**, Amazon Alexa, Pittsburgh, PA
I evaluated the Alexa Speech Recognition models against data poisoning attacks and found ways to make them more robust

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

Apr–Aug 2017 **Research Intern**, AgroParisTech, Paris, France
Research project on Transfer Learning for time series using boosting methods, advised by prof. Antoine Cornuejols. The work was presented at the Symposium on Intelligent Data Analysis

June-Aug 2016 **Data Scientist Intern**, DataScienTest, Paris, France
Participated in the creation of DataScienTest, which grew up to become a leading online training platform for Data Scientists. I designed the first generation of Machine Learning exercises and automatic evaluation code.

## Talks

Nov 2022 **Invited research talk**, Technion University, Haifa, Israel

Aug 2022 **Invited research talk**, Security and Privacy for Speech Communication (SPSC) group, online

Sep 2019 **Tutorial**, InterSpeech conference, Graz, Austria
With Prof. Bhiksha Raj and Joseph Keshet, I gave a tutorial on generating adversarial examples for speech and speaker recognition

Sep 2018–May 2022 **Introduction to Deep Learning**, *Prof. Bhiksha Raj*, Guest Lecturer

- I gave several guest lectures for this 200+ students course at Carnegie Mellon University
- I covered topics like Deep Learning History, GANs, Transformers and GNNs
- Previously, I was twice a Teaching Assistant with tasks including Recitations, Homework design and grading, Office Hours and Project mentoring.

## Highlighted Projects

Mar 2022–Ongoing **Attacks and Defenses for Self-Supervised Speech Representations**

- I studied the vulnerabilities of modern Transformer ASR models pretrained with Self-Supervised Learning (Wav2Vec2, WavLM, Data2Vec, etc).
- I show that learning features from a very general pretext training task makes these models more at risk to black-box threat models. Work under review **[Paper][Code]**.
- Currently, I am trying to turn this vulnerability into an asset, by augmenting SSL pretaining with adversarial examples to improve robustness and model transferability.

Sep 2022–Ongoing **Indirect impacts of adversarial vulnerabilities for ASR security**

- Releasing non-robust ASR models has indirect consequences that can be more problematic than the adversarial examples themselves. We study several in our recent research direction.
- We apply *data poisoning* to insert a backdoor in ASR models trained with semi-supervised learning, using targeted adversarial perturbations to fool the teacher during pseudo-labeling. Project in collaboration with Shinji Watanabe's research group.
- We use *membership inference* to predict whether the speaker of a given utterance is in the training data of a model. Our method relies on adversarial perturbations to probe the shape of the local decision boundary.
- We fooled Whisper, a *multilingual* ASR model, to mispredict the language of a sentence. This degrades recognition performance with a very low perturbation budget. Work under review **[Paper][Code]**.

Jan 2020–Dec 2022 **Adversarial defenses for Speech and audio with Smoothing and Speech Processing**

- As part of the GARD Darpa project, I defended Speech Recognition and Audio classification models against adversarial attacks, by combining the Gaussian Smoothing defense with traditional Speech Processing methods.
- First I applied high-frequency filters to the smoothing gaussian noise, to better target adversarial patterns in the High frequency spectrum. Work published at *ICASSP 2021* **[Paper]**
- Then I combined it Speech Enhancement methods and Ttext voting schemes to improve its performance on CTC and Attention models. Work published at *EMNLP 2021* **[Paper][Code]**
- More recently I replaced the enhancement module with a Denoising Diffusion model, applied it to Wav2Vec2 and considerably improved our previous defense. Work scheduled for submission at InterSpeech 2023.

Sep 2021–Sep 2022 **robust_speech: a Speech Robustness package**

- I released a package for evaluating the robustness of ASR models. This package and our results were presented at InterSpeech 2022 **[Paper][Code]**
- robust_speech is currently used by several industry research teams worldwide. I keep maintaining it and releasing new features.

Jan 2021–Dec 2021 **Evaluating robustness beyond adversarial accuracy**

- Using accuracy for evaluating adversarial robustness has limits. We outline them and propose to instead approximate the amount of adversarial perturbations, using angle-based metrics as a proxy. **[Paper][Code]**

Jan 2018 - Nov 2018 **Retrieval-based neural code generation**, *with Prof. Graham Neubig*

- We trained a Code generation encoder-decoder model, with a decoder constrained by the syntax tree and using subtree retrieval in the training set at inference to improve performance.
- Our project was the 2018 State-of-the Art on two coding tasks and was presented at EMNLP 2018 **[Paper][Code]**

## Skills

| | |
|---|---|
| Languages | Python, C/C++, Java, SQL, Bash |
| Frameworks | PyTorch, Tensorflow, DyNet, Fairseq, SpeechBrain |
| Utilities | Anaconda, Git, Jupyter Notebook, Alexa Skills, AWS |

## Courses

| | |
|---|---|
| Machine Learning | Natural Language Processing, Deep Learning, Advanced Machine Learning, Multimodal Machine Learning, Neural Language Translation |
| CS | Algorithms, Advanced Programming, Data Management, Computational Geometry |
| Math | Logic, Algebra, Number Theory, Analysis, Optimization, Differential Equations, Sequences/Series |