Rémi Godet

Intern at INRIA research center on privacy and federated learning with an engineering background and high interest in novel deeplearning techniques. Currently looking for PhD position in a dynamic and diverse environment !
remgodet@outlook.fr — +33 7 50 68 79 71 — linkedin.com/in/remi-godet — github.com/remigodet

RESEARCH INTERESTS

Adversarial Attacks, Machine Learning Intellectual Property, AI privacy, LLM agents

ACADEMIC EXPERIENCE

INRIA Sophia-Antipolis, France
Research Intern August 2024 — Present
Tutored research on improving federated learning state-of-the-art frameworks by sharing in a 1-shot fashion compressed models or datasets from clients to the server. Distillation is performed server-side to train a global model to be shared. Exploration of the impact of plug-in privacy methods such as Differential Privacy on accuracy and stability.
# Federated Learning, Distillation, Meta-learning, Differential Privacy

EDUCATION

Cornell University, Ithaca, USA Aug 2023 — May 2024
Master of Engineering in Computer Science Cumulative GPA: 3.820/4.00

CentraleSupelec, Gif-sur-Yvette, France Sept 2021 — Present
Diplôme d'Ingénieur (Master of Science *eq.*) Cumulative GPA: 3.99/4.33

PROJECTS

LLM agent for web tasks Jan 2024 — May 2024
Participating in a research project aimed at designing a code-producing LLM in an HTLM environment (Webarena). Using chain-of-thought and ReAct prompting, the LLM outputs python solutions to user-defined queries. To guide its generation, it is provided with a dynamically built library of functions. I took part in using an LLM to compress and extract program functions from previous solutions in order to build up the library.
# LLM agent, Program Synthesis, CoT, ReAct CentraleSupelec

LLM detection through interaction Jan 2024 — May 2024
Designed a protocol where an LLM is tasked with judging the responses of its LLM peers, one of which is a different model (ie. GPT 3 instead of 4), an effort at sketching out an "LLM-to-LLM Turing test". This work is an attempt to pave the way for more sophisticated methods of language model interactions and contributing to a broader understanding of AI agents identity.
# LLM self-awareness, Security, LLM agentic protocol CentraleSupelec

Minority adversarial attacks in low dimensional cooperative tasks Sept 2023 — Dec 2023
Examination of the effectiveness of minority adversarial agents in a low-dimensional environment by training a set of cooperative agents then freezing their policies, training an adversary as a single agent reinforcement learning problem and comparing it to random and simple rogue agents. We found evidence of the agent exploiting local behavior of the victims but no emergence of out-of-distribution behavior able to disrupt the entire population.
# cMARL, MAPPO, PPO Adversarial Attack Cornell University

Comparing watermark trigger sets generation in NLP models Sept 2023 — Dec 2023
Real use case of protecting a natural-language-based machine learning application by exploring choices in terms of different trigger sets while retaining performance.
# Watermarking, NLP Cornell University

Real-world bench-marking of adversarial attacks on CV models Jan 2023 — June 2023
Exploration of attacks and defenses around ML models, and bench-marking on real-world attacks on stop signs, to constitute a dataset for Thales light ThereSIS lab.
# Poisoning, Evasion, Extraction attacks, Real-world robustness, Detectability CentraleSupelec

Robust digital signature of ML models by watermarking Sept 2022 — Jan 2023
Research with IRT-SystemX on watermarking an image classifier based on a hidden set of key images to assert the intellectual property of the algorithm.

# Watermarking of models, Encryption, Robustness, Adversarial attack CentraleSupelec

Robustification of ML models to out-of-distribution outliers Jan 2022 — June 2022
Normalization of an autoencoder to single out isolated outliers, and exploration of the latent space with MCMC methods to find the manifold of errorless reconstructions. Project made in association with the French CEA institution.
# CV, Manifold Sampling, MCMC, Energy-based Model CentraleSupelec

Fake news detection algorithm Sept 2021
Implementation of NLP methods (from SVM to LSTM) to gauge the efficiency of Twitter fake-news detection based on a labeled corpus from the government.
# NLP, LSTM, Labeling CentraleSupelec

Automatic soil cover classification Jan 2023 — June 2023
Work with Preligens R&D scientists on Sentinel2 large images to implement and access different algorithms to perform classification and segmentation of satellite data.
# SVM, Random Forest, CNN CentraleSupelec

SELECTED COURSES

Master's Courses

- Cornell University CS 6700 Advanced Artificial Intelligence
- Cornell University CS 5306 Crowdsourcing & Human Computation
- Cornell University CS 6888 Deeplearning
- Cornell University CS 6756 Learning for Robot Decision Making
- Cornell University CS 5780 Intro to Machine Learning
- CentraleSupelec 2EL1580 Artificial Intelligence
- CentraleSupelec 2CC1000 Control theory

Bachelor's Courses

- CentraleSupelec 1CC5000 Statistics and Learning
- CentraleSupelec 1SL1500 Partial Differential Equations
- CentraleSupelec 1SL1000 Convergence, Integration, Probability
- CentraleSupelec 1CC200 Algorithmics and Complexity

OTHER EXPERIENCES

Automatants Gif-sur-Yvette, France
Part of CentraleSupelec's AI association. Sept 2021 — June 2022
Pics Gif-sur-Yvette, France
Part of CentraleSupelec's photography association, as project lead. Sept 2021 — Jan 2023

ENGLISH & GRE TESTS

IELTS (Academic): 7.5 (overall score)

Listening: 8.5 — Reading: 8.0
Speaking: 7.5 — Writing: 6.5
Test date: Nov 2022

GRE General Test:

Quant: 168 — Verbal: 166
Analytical writing: 4.0
Test date: Nov 2022

SKILLS

- **Programming:** Python, Java, C#, VHDL
- **Software:** Docker, Azure
- **Soft Skills:** Management, Planning, Leadership