Philippe Bergna

Philippe-Bergna | pbergna753@gmail.com | philippe753 | +44 7858362489 | ★ British Citizen

EDUCATION

University of Bristol — Bristol, UK
MEng Engineering Mathematics with Study Abroad — *09/2018 - 06/2022*
BEng Engineering Mathematics — *09/2018 - 06/2021*

- First Two Years: Achieved an average of **74% (Ranking 3/∼100)**.
- Final Two Years: Impacted by COVID-19.

Katholieke Universiteit Leuven — Leuven, Belgium
MS Artificial Intelligence — *09/2020 - 07/2021*
Took an AI master's degree in my third year at University.

Relevant Courses: Fundamentals of AI, Artificial Neural Networks, Support Vector Machines, Philosophy Of Mind and AI, Computer Vision, Data Mining, Privacy and Big Data, Neural Computing, Robotics Systems, Stochastic Optimisation, Applied Statistics, and more.

PUBLICATIONS

- **Philippe Bergna**, Jake Thomas. *"Out of Distribution Detection Enhancer with Adversarial Reprogramming"* In preparation for submission to the International Conference on Machine Learning (ICML) in 2025.

EXPERIENCE

Advai (Selected from **1/∼500** applicants) — London, UK
Research Scientist for AI Safety and Security — *11/2023 - Present*
Research Engineer — *09/2022 - 11/2023*
One of the few Research Scientists *without a PhD*.

◇ **3D Physical Adversarial Attacks**

- Collaborated with the UK Ministry of Defense to develop 3D adversarial camouflages that deceive object detection models.
- Employed the Projected Gradient Descent (PGD) algorithm within a Differentiable Transferable Network (DTN) to train adversarial attacks, effectively causing YOLOv7 to misclassify vehicles, and utilized CARLA and Unreal Engine 5 to render photorealistic training data, enhancing the robustness of adversarial attacks across diverse environments and camera projections.
- Achieved **state-of-the-art at 95% undetected rate**, resulting on securing multimillion-pound contracts for the company.

◇ **Adversarial Attacks for Facial Verification Systems**

- Developed transferable and targeted black-box adversarial attacks against iProov, the UK's leading facial verification system.
- Combined an Ensemble of diverse facial verification models and other optimization techniques for maximizing adversarial transferability, achieving a **55% success rate**.

◇ **Active learning for Object Detection Models**

- Developed active learning with adversarial attacks to quantify image uncertainty, improving data selection for object detection.
- Extended DeepFool optimization for YOLOv7, achieving an **8% MaP improvement and reducing training data required by 50%**.

◇ **Out of Distribution Detection and Model Analysis Consultant**

- Often selected as the lead expert of the company as the model evaluator for other computer vision companies (VIZGARD and Brimstone) to guide and improve their model performance and safety.
- This includes out-of-distribution detection, model bias, suggesting better evaluation metrics and possible changes to improve model performance.

University of Bristol — Bristol, UK
Teaching Assistant — *09/2021 – 06/2022*

Courses: Artificial Intelligence, Introduction to Data Science, Introduction to Computer Programming, Engineering Physics 2.

- **Lecturing & Tutoring:** Taught machine learning algorithms (CNNs, LSTMs, Transformers, ResNets, PCA) and programming concepts.
- **Project Supervision:** In charge of guiding and helping final-year students grounds in computer vision projects.
- **Academic Support:** Assisted students with worksheet exercises, debugging, and concept comprehension.
- **Unique Appointment:** Selected as the sole TA without a PhD, demonstrating exceptional expertise and commitment.

ACCADEMIC EXPERIENCE

Dissertation: A Multi-modal Explainable Framework for Detecting Fake News on Twitter — University of Bristol

- Developed a multimodal fake news detection system combining text and image inputs, achieving a **2.8% performance improvement** over uni-modal methods by integrating models like XLM-RoBERTa, BERT, and VGG with XGBoost.
- Enhanced model explainability and robustness using LIME for text and image interpretations and forensic image analysis to detect manipulations.