



# ALESSIO DEVOTO

[devoto.alessio@gmail.com](mailto:devoto.alessio@gmail.com)

[alessiodevoto.io](http://alessiodevoto.io)

[X/devoto\\_alessio](https://x.com/devoto_alessio)

## EDUCATION

---

<b>PhD in Data Science</b> La Sapienza, University of Rome. Focus on Efficient and Adaptive neural networks and Explainability for AI models. Supervisor: Prof. Simone Scardapane.	Nov 2022 – Present
<b>Visiting Researcher</b> The University of Edinburgh. Focus on NLP with emphasis on efficient inference and explainability.	Mar 2024 – Jul 2024
<b>Master's Degree in Computer Engineering</b> La Sapienza, University of Rome – Final mark: 110/110 cum Laude.	Sep 2019 – Jan 2022
<b>Visiting Student</b> Universidad Politecnica de Valencia, Spain.	Feb 2021 – Jul 2021
<b>Bachelor's Degree in Control and Computer Engineering</b> La Sapienza University of Rome – Final mark: 110/110 cum Laude.	Sep 2016 – Oct 2019

## EXPERIENCE

---

<b>Teaching Assistant</b> Teaching assistant for Neural Networks for Data Science Applications. Led hands-on PyTorch tutorials and project supervision for 120+ MSc students.	Sep 2023 – Present
<b>ICF trainee Coach</b> Training to become a life & business coach (30+ hours experience as individual coach).	Feb 2020 – Present
<b>Tutor</b> Tutor for 40+ university/high school students (Maths, Latin, Ancient Greek).	Sep 2016 – Present
<b>Research Internship – ISPAMM Lab</b> Development of models for explainable High Energy Physics.	Jan 2022 – Nov 2022

## BLOG

---

I maintain a small blog where I share code tutorials and insights on various deep learning topics, like implementing a *"ViT from scratch in pure JAX"* or *"Logitlens from scratch without interpretability libraries"*.  
Visit my blog here: <https://alessiodevoto.github.io/blog>.

## PROJECTS

---

<b>Explainability for High Energy Physics</b> Partnership with CERN and University of Liverpool. Developed explainability methods for AI models (mainly GNNs) for Science Discovery. <a href="#">MUCCA Project Website</a>	Feb 2023 – Present
<b>Next Generation 6G communications.</b> EU funded project with academic and industrial partners. Designed adaptive neural networks to make communication pipelines efficient and goal-oriented, for next-generation 6G. <a href="#">6G-GOALS Website</a>	Mar 2023 – Present

## SELECTED PUBLICATIONS

---

- A Simple and Effective  $L_2$  Norm-Based Strategy for KV Cache Compression.** A. Devoto\* , Y. Zhao\*, S. Scardapane, and P. Minervini. *Empirical Findings in Natural Language Processing (EMNLP)*, 2024. ([arXiv:2406.11430](#))
- Steering Knowledge Selection Behaviours in LLMs via SAE-Based Representation Engineering.** Y. Zhao, A. Devoto , G. Hong, X. Du, A. P. Gema, H. Wang, K.-F. Wong, and P. Minervini. *arXiv preprint arXiv:2410.15999*, 2024.
- Adaptive Computation Modules: Granular Conditional Computation For Efficient Inference.** B. Wójcik, A. Devoto , K. Pustelnik, P. Minervini, and S. Scardapane. *arXiv preprint arXiv:2312.10193*, 2023.
- Adaptive Layer Selection for Efficient Vision Transformer Fine-Tuning.** A. Devoto , F. Alvetreti, J. Pomponi, P. Di Lorenzo, P. Minervini, and S. Scardapane. *arXiv preprint arXiv:2408.08670*, 2024.
- Analysing the Residual Stream of Language Models Under Knowledge Conflicts.** Y. Zhao, X. Du, G. Hong, A. P. Gema, A. Devoto , H. Wang, X. He, K.-F. Wong, and P. Minervini. *Mechanistic Interpretability Workshop (MINT) NeurIPS 2024*. [arXiv:2410.16090](#), 2024.
- Are We Done with MMLU?** A. P. Gema, J. O. J. Leang, G. Hong, A. Devoto , A. C. M. Mancino, R. Saxena, X. He, Y. Zhao, X. Du, and M. R. G. Madani. *arXiv preprint arXiv:2406.04127*, 2024.
- Adaptive Semantic Token Selection for AI-native Goal-oriented Communications.** A. Devoto , S. Petruzzi, J. Pomponi, P. Di Lorenzo, and S. Scardapane. *Global Communications Conference (GlobeComm 2024)* [arXiv:2405.02330](#), 2024.
- Reidentification of Objects From Aerial Photos With Hybrid Siamese Neural Networks.** A. Devoto , I. Spinelli, F. Murabito, F. Chiovoloni, R. Musmeci, and S. Scardapane. *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2997–3005, 2022. IEEE.
- Enhancing High-Energy Particle Physics Collision Analysis through Graph Data Attribution Techniques.** A. Verdone, A. Devoto , C. Sebastiani, J. Carmignani, M. D’Onofrio, S. Giagu, S. Scardapane, and M. Panella. *WIRN*, [arXiv:2407.14859](#) 2024.
- Conditional computation in neural networks: principles and research trends.** S. Scardapane, A. Baiocchi, A. Devoto , V. Marsocci, P. Minervini, and J. Pomponi. *Artificial Intelligence*, [arXiv:2403.07965](#) 2024.
- Cascaded Scaling Classifier: class incremental learning with probability scaling.** J. Pomponi, A. Devoto , and S. Scardapane. *arXiv preprint arXiv:2402.01262*, 2024.

## TECHNICAL SKILLS

---

- Deep Learning Frameworks:** PyTorch, JAX, Hugging Face Transformers
- Programming Languages:** Python, C, Java
- Web Development:** HTML, JavaScript, CSS
- ML Specialties:** Adaptive Neural Networks, LLM Optimization, KV Cache Compression
- Development Tools:** Git, Docker, Unix/Linux
- Research Areas:** Explainable AI, Efficient Inference

## LANGUAGES

---

- Italian:** Native
- English:** C2
- Spanish:** C1
- Portuguese:** B2 & learning