EURA NOVA

INTERNSHIP OFFERS 2026

FACULTIES OF SCIENCE AND ENGINEERING

EURANOVA

Welcome to Euranova's Internship Program

Euranova is an international consulting company that specialises in data-driven business solutions, guided by a strong company culture and a passion for innovation. Founded in September 2008 and located in Brussels, Marseille, and Tunis, our purpose is to bring life to our customers' great ideas by offering best-in-class services in data science, software engineering, and data architecture. To do this, we offer counsel, R&D, and solutions.

Inside and out, we care for collaboration, impact, and excellence, and in line with this course of action, we offer academic programs in partnership with universities. See below for details regarding our internship offers.

Explore Our Internship Opportunities

This document presents internship topics supervised by our consultancy and innovation department. Each project is an opportunity to be actively involved in the development of solutions to address tomorrow's challenges in ICTs and to implement them today, and can be developed into end-of-study projects or master's theses. The students will work in a dedicated international team of engineers with diverse expertise in machine learning, computer vision, Large Language Models (LLMs), data privacy, AI governance, high-performance computing, etc.

Our company culture is built on mutual enrichment, continuous learning, empathy in coding, and knowledge sharing. Though we consider that attitudes take precedence over technical prowess, we usually work with solution-oriented engineers in computeré science and applied sciences, who thrive in a collaborative environment and embrace challenges with humility, curiosity, and a scientific mindset. To learn more about our R&D activities or company culture, visit our website at https://euranova.eu.

How to apply

Interested in being a part of our story? Here's how to move forward:

- When you have gone through our internship offers, pick your favourite three.
- Draft a short text for each one, stating why you find it interesting and what you would do about it.
- Send us this statement, along with your CV, to career@euranova.eu.

Please note that the locations and dates are indicative; do not hesitate to contact us to find an arrangement. Although previous experience with the technologies mentioned in the offers is appreciated, it won't be amongst the main criteria for intern selection.

Submission deadline: we encourage you to apply early, as we will review applications as they come in to find the best matches. Should you have any questions regarding our internship offers or the selection process, do not hesitate to reach out to career@euranova.eu.

Debiasing AI Models via Data Attribution

1. Context

As AI regulations such as the EU AI Act require organisations to demonstrate that their models are fair, transparent, and auditable, there is a growing need for fair and unbiased data science. This internship tackles this challenge in a data-centric way by focusing on the root causes of bias in training data.

Data attribution techniques trace model predictions back to the specific training examples that most influence them. This allows us to identify the exact sources of bias and take targeted corrective actions, providing a more actionable and effective way to improve fairness.

Key challenges this project will address:

- Combining sample-level and group-level attribution methods to detect both individual and cohort-level sources of bias.
- Exploring debiasing strategies such as targeted data corrections, reweighting, or augmentation.
- Validating causal links between influential data and biased outcomes via counterfactual tests.
- Evaluating trade-offs between fairness and model performance, and benchmarking multiple attribution and debiasing methods to identify the most effective combinations.

2. Technologies

- Data attribution methods
- Fairness evaluation metrics
- Python, PyTorch/TensorFlow
- Data-centric ML pipelines and benchmarking

3. Objectives

- Study the state of the art in data attribution and bias mitigation, including both sample-level and group-level attribution techniques.
- Design and implement a debiasing pipeline that integrates multiple attribution methods to identify the most influential and bias-inducing training samples.
- Experiment with debiasing strategies such as data correction, reweighting, and augmentation, guided by the attribution results.
- Develop and apply counterfactual testing procedures to validate causal links between influential data and biased predictions.
- Benchmark attribution and debiasing methods on public datasets, evaluating trade-offs between fairness improvements and model performance.
- Deliver a reproducible pipeline and a set of best practices for data-centric bias mitigation.

4. Where and when

Belgium (Mont-Saint-Guibert), from February to June 2026.

5. References

[1] P. W. Koh and P. Liang, "Understanding Black-box Predictions via Influence Functions," *International Conference on Machine Learning (ICML)*, 2017, https://arxiv.org/pdf/1703.04730.

[2] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys (CSUR)*, 2021, https://arxiv.org/pdf/1908.09635.

[3] A. Arnaiz-Rodriguez and N. Oliver, "Towards Algorithmic Fairness by means of Instance-level Data Re-weighting based on Shapley Values," 2024, https://arxiv.org/pdf/2303.01928.

Connectionism Meets Symbolism: Let LLMs Build Constrained Programming Models

1. Context

Do you want to know the optimal path from A to B? Do you want to optimally choose which machines to turn on and use to reach your production target at a minimal cost? Do you need to plan your activity so that you and your staff deliver as much value as possible in a minimum of time? Data science will likely not help there, or can it?

Many real-world tasks belong to the realm of optimisation: you know the constraints around your business problem, and you would like to know what to do to minimise/maximise a certain objective. This can be modelled through mathematics, by creating variables, and linking them amongst themselves/with constant parameters which define the problem.

How can we build these programs automatically? Well, some machine learning models have become quite good at turning natural language into code recently...

On top of helping humans formulate such programs, having robust ways to design such programs from natural text descriptions could make LLMs a lot better where they are currently bad: tasks which require planning/optimisation.

2. Technologies

- Large Language Models (LLMs)
- Constrained Optimisation using Pyomo
- Python

3. Objectives

- Study the literature around the use of LLMs to generate constrained programs.
- Set up a simple benchmark to evaluate how good LLMs are at modelling constrained problems.
- Use this benchmark to evaluate how good the current best coding models are at solving those using the Pyomo optimisation library.
- Based on the analysis of failure modes, some automatic prompt optimisation/context engineering using DSPy, and workflow creation using LangGraph, you will increase the robustness of generations.
- The final goal will be to solve a variant of a complex real-world optimisation problem encountered in the past within Euranova.

4. Where and when

Belgium (Mont-Saint-Guibert), from March/April to August/September 2025.

5. References

[1] Yuliang Song and Eldan Cohen, “Do LLMs Understand Constraint Programming? Zero-Shot Constraint Programming Model Generation Using LLMs”, 2025, THE 19TH LEARNING AND INTELLIGENT OPTIMIZATION CONFERENCE, https://openreview.net/forum?id=6zlpzSKzqj

Scalable and Stable Training Methods for High-Resolution 3D Gaussian Splatting

1. Context

Gaussian Splatting (GS) is a recent method for 3D reconstruction and real-time rendering that represents scenes with millions of 3D Gaussians. While powerful, current GS implementations face a significant limitation with very high-resolution imagery (e.g., 100MP+ aerial photos). The standard approach is to drastically downsample these images, which discards the high-frequency details that are crucial for accurate, fine-grained reconstruction.

A potential workaround is to create a multi-scale dataset by combining downsampled overview images with high-resolution patches of key areas of interest. However, this approach often leads to training instability, causing visual artefacts and inconsistent levels of detail. The core optimisation and adaptive densification mechanisms in GS are not inherently designed to handle the extreme variance in scale and detail present in such datasets.

This research internship will investigate and develop novel training methodologies to enable Gaussian Splatting to effectively leverage high-resolution, multi-scale data. The goal is to produce globally coherent 3D models with high-fidelity details, directly from the source imagery, without destructive downsampling.

2. Technologies

- 3D Reconstruction / Photogrammetry
- 3D Gaussian Splatting
- Deep Learning (Pytorch, nerfstudio)
- Optimisation Methods

3. Objectives

- Conduct a literature review of 3D Gaussian Splatting, focusing on state-of-the-art optimisation methods and adaptive densification mechanisms, and explore existing techniques for handling very high-resolution imagery for 3D reconstruction.
- Establish a baseline by implementing a standard GS pipeline on a custom multi-scale dataset to reproduce, visualize, analyse, and quantify the known training instabilities and visual artefacts.
- Develop and implement novel training strategies to stabilize the learning process. This will involve experimenting with coarse-to-fine training, scale-aware loss functions, modified densification criteria (e.g., splitting and pruning), and multi-stage optimisation schedules.
- Systematically evaluate the proposed methods against the baseline, using both qualitative visual inspection and quantitative metrics to measure improvements in detail preservation, artifact reduction, and overall model coherence.
- The final goal is to deliver a robust training methodology that successfully generates a high-fidelity 3D model from high-resolution imagery.

4. Where and when

Belgium (Mont-Saint-Guibert), from February to June 2026.

5. References

[1] Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.

[2] Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., ... & Kanazawa, A. (2023, July). Nerfstudio: A modular framework for neural radiance field development. In ACM SIGGRAPH 2023 conference proceedings (pp. 1-12).

[3] Cheng, Z., Sun, M., Liu Y., Ge, Z., Tang, L., Xu, M., Li, Y. & Pan, P. (2025) CLoD-GS: Continuous Level-of-Detail via 3D Gaussian Splatting. arXiv preprint arXiv:2510.09997.

[4] Xia, J., & Liu, L. (2025). Close-up-GS: Enhancing Close-Up View Synthesis in 3D Gaussian Splatting with Progressive Self-Training. arXiv preprint arXiv:2503.09396.

[5] Jiang, L., Ren, K., Yu, M., Xu, L., Dong, J., Lu, T., ... & Dai, B. (2025). Horizon-GS: Unified 3D Gaussian Splatting for Large-Scale Aerial-to-Ground Scenes. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 26789-26799).

Design and Implementation of a Reusable AI Governance Dashboard for Model Risk & Compliance

1. Context

Organisations developing AI solutions must increasingly ensure compliance with regulatory and governance frameworks.

However, most teams lack tools to measure and visualise governance maturity, making it difficult to operationalise Responsible AI principles.

This internship aims to design and implement a Power BI–based AI Governance Dashboard that provides real-time visibility into the governance and risk status of AI projects. It can be seen as a transformation of the AIG manual checklist to a dynamic, automated system.

The goal is to transform qualitative governance practices into quantitative metrics, enabling both internal teams and client projects to monitor and improve AI model compliance and risk management in a scalable and repeatable way.

The intern will define the underlying data schema and KPIs, develop interactive visualisations (e.g., compliance scores, risk heatmaps, progress tracking…), and test the dashboard on multiple representative AI use cases. A final report will summarise findings, lessons learned, and recommendations for integration into Euranova's AI Governance Framework.

2. Technologies

- Data Visualisation: Power BI
- AI Governance Frameworks: Euranova AIG Framework, EU AI Act, ISO/IEC 42001, NIST AI RMF
- Data Management & Analytics: Python, Pandas

3. Objectives

The internship aims to build a reusable AI Governance Dashboard that can be applied across multiple AI use cases. The objectives are the following:

- AIG Framework understanding
- Metric & Data Design
    - Define measurable KPIs and metadata (e.g., model lifecycle status, audit artifacts, bias tests, explainability reports).
    - Develop a generic data schema, enabling reusability across projects.
- Dashboard Development
    - Implement the dashboard using Power BI or Streamlit/Python.
    - Build interactive visualisations (compliance scores, risk heatmaps, documentation progress..).
    - Ensure configurability for different AI project types.
- Validation & Reporting
    - Test the dashboard on 2–3 representative AI use cases.
    - Collect feedback from data scientists and governance officers.

- Produce a short report with recommendations for future integration into Euranova's AI Governance Framework.

4. Where and when

Tunis, from February to June 2026

5. References

[1] ISO/IEC 42001:2023 – *Artificial Intelligence Management System – Requirements.* International Organization for Standardization (ISO), 2023.

[2] Regulation (EU) 2024/1689 – *European Artificial Intelligence Act.* Official Journal of the European Union, 2024.

[3] National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* U.S. Department of Commerce, 2023.

[4] European Commission – *Ethics Guidelines for Trustworthy AI.* High-Level Expert Group on Artificial Intelligence, 2019.

Deep Learning for Organ-Aware Tumour Segmentation in PET/CT Imaging

1. Context

In the recent era, deep learning has emerged as a pivotal tool for the diagnosis and follow-up of various medical conditions, particularly in the sector of cancerous tumour detection and tracking [1, 2]. At Euranova, we have developed a cutting-edge deep learning model that automatically segments tumours and biomarkers from PET/CT scans. While highly effective, this initial version operates without anatomical context: it identifies a lesion but not the organ in which it resides.

However, the clinical significance of a tumour is deeply tied to its location. International response criteria (such as RECIST or PERCIST) are often organ-specific, meaning that understanding the anatomical context is crucial for accurate patient response evaluation. This internship aims to bridge this gap by integrating state-of-the-art organ segmentation [3] into our existing pipeline. By providing precise organ labels for each detected tumour, we will not only enrich our current product but also explore novel methods to use this anatomical information to improve the core performance of our tumour segmentation model.

2. Knowledge & Technologies

Knowledge:

- Solid foundation in Deep Learning, particularly for Computer Vision (e.g., CNNs, Transformers).
- Strong understanding of semantic segmentation concepts and architectures (e.g., U-Net).
- Previous experiments with medical imaging (specifically CT and PET) is a plus.

Technologies:

- Excellent programming skills in Python.
- Hands-on experience with a major deep learning framework, preferably PyTorch.
- Proficiency with the standard data science stack: NumPy, Pandas, Scikit-learn.
- Experience with version control using Git is a plus.
- Knowledge of medical imaging libraries (e.g., MONAI, SimpleITK) and containerization (Docker) is a plus.

3. Objectives

The main goals of this internship are to develop and integrate an organ-labelling module and leverage it to improve our core technology. The selected candidate will:

- Conduct a state-of-the-art review of deep learning models for whole-body organ segmentation on CT images to identify the most promising approaches.
- Select, adapt, and integrate a high-performance organ segmentation model into our existing AI pipeline.
- Develop a robust methodology to accurately assign an organ label to each tumour/biomarker detected by our primary segmentation model.
- Investigate and develop novel techniques to leverage organ-specific information as an input or a prior to improve the accuracy and robustness of our core tumour segmentation model.

- Evaluate the performance of the integrated pipeline and the improved segmentation model on our private, in-house datasets.
- Collaborate with our team of AI scientists and medical experts to ensure the clinical relevance and technical excellence of the developed solution.

By the end of the internship, the candidate will have produced:

- A functional module integrated into our pipeline that provides accurate organ labels for all detected biomarkers.
- A new version of the deep learning tumour segmentation model that demonstrates superior performance by incorporating organ-aware context.
- A document detailing the state-of-the-art review, methodologies developed, experimental setup, and final results.
- A contribution to a potential scientific publication or internal presentation summarising the project's innovative findings.

4. Where and when

The internship will begin in February/April 2025 and will last for 5 to 6 months, in the R&D Department in the Marseille office. It will be supervised by experts from France and Belgium.

5. References

[1] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[2] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 18(2), 203-211.

[3] https://github.com/wasserth/TotalSegmentator

Bias Identification and Mitigation in Pretrained Medical Image Representations

1. Context

Medical image segmentation, which involves making per-pixel predictions, is a crucial process for assisting early disease detection, diagnosis, monitoring treatment, and follow-up. Leveraging pretrained models such as nnU-Net [1], MedSAM [2], STU-Net [3], TransUNet [4], or DInov3 [5] has become essential in medical imaging for enhancing segmentation performance and accelerating the development of new diagnostic tools. However, these models are prone to inheriting and amplifying biases from their training data, which can negatively affect their performance and fairness. Fairness is defined as a model's ability to deliver consistent performance across diverse patient populations and sub-groups otherwise known as bias. Biases can be demographic (e.g., gender, age), technical (e.g., different acquisition parameters or machines), or representation biases (e.g., class imbalance). The objective of this internship is to systematically investigate the impact of biases on pretrained medical image segmentation models and to develop strategies for mitigating these biases to ensure more equitable and reliable model performance. This gives rise to the following research questions:

- What constitutes a good medical image representation?
- How can bias be visualized and quantified using these embeddings?
- How can the impact of a biased embedding on segmentation performance be measured?

At Euranova, a Python package (Famous) has been developed specifically for mitigating bias in deep learning segmentation models. The intern's role will involve actively contributing to the enhancement of its core modules and validating their performance across various real-world applications.

2. Technologies

- Excellent programming skills in Python.
- Hands-on experience with a major deep learning framework, preferably PyTorch.
- Proficiency with the standard data science stack: NumPy, Pandas, Scikit-learn.
- Experience with version control using Git is a plus.
- Knowledge of medical imaging libraries (e.g., MONAI, SimpleITK) and containerization (Docker) is a plus.

3. Objectives

The objectives are the following:

- Define and Evaluate medical image representations: Review state of the art models including Dinov3, TransUnet, nnUnet, and/or other relevant architectures and integrate them into the famous package in order to have meaningful image representation for a medical dataset.
- Visual and Quantify bias: Apply dimensionality reduction techniques (e.g., t-SNE, UMAP) to project embeddings and reveal distinctions between different bias groups. Develop tools within the Famous package—including dimensionality reduction methods and distance metrics—to quantify bias-group diversity or disparity.

- Measure the Impact of bias on segmentation models: Train a segmentation model using these embeddings and evaluate its performance across different bias groups to analyse the correlation between bias-group dissimilarity and segmentation performance.

The work will be part of a work package integrated into the *Famous* package previously developed at Euranova. The intern is expected to deliver:

- A module for Integration of embeddings from medical image segmentation libraries and state-of-the-art models into the Famous package : Monai, torchvision....
- A module that visualizes bias embeddings using feature and dimensionality reduction techniques, highlighting disparities among bias groups.
- A module that reports the explored data or distribution similarity measures and quantifies the disparity of different bias groups.
- A code for training of a segmentation model based on these embeddings and quantification of the impact of bias on segmentation performance.
- A clean, test-driven code with proper documentation, adhering to best coding practices.
- A clear report showcasing the results and the main findings

4. Where and when

The internship will begin in February/April 2025 and will last for 5 to 6 months, in the R&D Department in the Marseille office. It will be supervised by experts from France and Belgium.

5. References

[1] Isensee, F., et al. (2020). "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature Methods

[2] Wang, Y., et al. (2024). "Segment Anything in Medical Images." Nature Communications.

[3] Li, X., et al. (2023). "STU-Net: The largest pre-trained medical image segmentation model." GitHub Repository.

[4] Chen, J., et al. (2021). "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation." Medical Image Analysis

[5] Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., & et al. (2025). DINOv3: Self-supervised learning for vision at unprecedented scale. arXiv. https://arxiv.org/abs/2508.10104

Lightweight Model Deployment for Computer Vision and VLM

1. Context

The rapid advancement in Large Language Models (LLMs), Vision Language Models (VLMs), and classic computer vision deep learning models presents both incredible opportunities and significant challenges. A primary hurdle is the high computational cost and resource requirements of deploying these SOTA (State-of-the-Art) models, which limits their accessibility. A key trend in the industry is to overcome these limitations by making models more lightweight and efficient. This is achieved through techniques such as quantization, distillation, and pruning.

At Euranova, we have established expertise in applying these optimization techniques to computer vision models and possess specialized hardware, including Nvidia Jetson AGX Orin and STMicroelectronics microcontrollers. It's noticed that some large models with billions of parameters have been successfully deployed and tested on the Nvidia Jetson AGX Orin, and recent STMicroelectronics microcontroller boards can also embed deep learning models. This ensures the feasibility of this project and provides a solid foundation for this internship.

2. Technologies

We are looking for a candidate with the following knowledge:

- Prior experience with training and evaluation of deep learning models using either the PyTorch or TensorFlow framework.
- Knowledge of ONNX and TensorRT is a plus.
- Good programming skills in Python and C/C++.

3. Objectives

The primary objective of this internship is to research and explore the possibilities of using model optimization techniques for deploying SOTA models on edge computing boards. You will:

- Gain hands-on experience with optimization & regularization techniques like quantization, pruning and distillation.
- Utilize and become proficient with relevant libraries and frameworks, including PyTorch, NVIDIA and the Hugging Face ecosystem.
- Deploy selected VLM or classic computer vision models—specifically, we are considering models that perform well in physics understanding and common sense to understand the real world like the Nvidia Cosmo model or V-JEPA 2—onto our target hardware, specifically the Nvidia Jetson AGX Orin and, if feasible, microcontroller boards.

By the end of the internship, you will be expected to produce and present the following:

- A detailed report on the research findings and methodologies used.
- The results of the evaluation for each quantization, distillation, and pruning technique explored.
- Demonstrations of the deployed lightweight models running efficiently on the specified hardware.
- Reusable code that can be integrated into future projects.

4. Where and when

The internship will begin in February/April 2025 and will last for 5 to 6 months, in the R&D Department in the Marseille office. It will be supervised by experts from France and Belgium.

5. References

- https://www.jetson-ai-lab.com/index.html
- https://github.com/nvidia-cosmos/cosmos-reason1
- https://github.com/facebookresearch/vjepa2
- https://github.com/pytorch/ao
- https://github.com/vllm-project/vllm
- https://github.com/NVIDIA/TensorRT

Evaluation and Optimisation of Protein Language Models for Function Prediction

1. Context

Proteins are the workhorse of life; they're responsible for many activities in our cells, tissues, organs, and bodies, performing essential and versatile functions that are critical to sustaining human health and the environment [1]. Each protein is composed of 20 types of amino acids, making protein sequences analogous to natural language.

At the intersection of the rapidly growing biological data landscape and advancements in Natural Language Processing (NLP), Protein Language models (PLMs) have emerged as a transformative force in modern research [2]. These models generate distributed embedded representations that encode semantic information about proteins, which can be applied to different downstream tasks such as 3D structure prediction, function prediction, and protein design.

In this internship, we will focus specifically on predicting protein function, as these predictions will assist researchers in understanding how proteins function and could lead to the development of medical treatments and therapies.

2. Technologies

- Protein Language models (ProtBERT, ESM-2, ProGen..[3])
- Deep learning
- Python

3. Objectives

- Review the state of the art in Protein Language Models (PLMs), including existing architectures, the training paradigm, and their applications in protein understanding and prediction tasks.
- Evaluate the performance of selected state-of-the-art PLMs on protein function prediction, using benchmark datasets such as SwissProt and UniProt, and annotation frameworks including Gene Ontology (GO).
- Investigate and evaluate potential improvements of PLMs by exploring strategies such as fine-tuning, adapter-based techniques, and multimodal integration that incorporate not only sequence data but also structural (3D) and functional information, as well as other enhancements that may improve both model performance and computational efficiency.

4. Where and when

Tunis, from February to June 2026

5. References

[1] Nijkamp, Erik, et al. "Progen2: exploring the boundaries of protein language models." Cell systems 14.11

(2023): 968-978.
https://www.cell.com/cell-systems/fulltext/S2405-4712(23)00272-7?uuid=uuid%3A33a13789-008a-44fc-82e4-ef8

83945c62b

[2] Wang, Lei, et al. "A comprehensive review of protein language models." arXiv preprint arXiv:2502.06881

(2025). https://arxiv.org/abs/2502.06881

[3] https://github.com/ISYSLAB-HUST/Protein-Language-Models