

Sebastian Cavada

Updated April 20, 2026

Email: sebastian.cavada@mbzuai.ac.ae

GitHub: [Sebo-the-tramp](https://github.com/Sebo-the-tramp)

Languages: Italian (C2), English (C1, IELTS 8), German (B2)

Homepage: sebo-the-tramp.github.io

Scholar: [Google Scholar](#)

Citizenship: Italy (EU)

Research Interests **3D Scene Reconstruction & Generation · Geometry-Grounded Foundation Models**
World Models · Gaussian Splatting

Education **Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)** Abu Dhabi, UAE
Master of Research in Computer Vision, GPA 3.72/4 2023 – 2025
Focused on large-scale 3D reconstruction, 3D foundation models, and Gaussian splatting. Supervised by Prof. [Ian Reid](#).

Free University of Bozen-Bolzano Bolzano, Italy
B.Sc. in Computer Science, 108/110 2019 – 2023
Trilingual program (Italian, English, German). Thesis supervised by Prof. [Oswald Lanz](#).

Universität Osnabrück Osnabrück, Germany
Erasmus Exchange in Cognitive Science 2022
Coursework emphasis on machine learning and computer vision.

Publications **NewtPhys: Do Foundation Models Understand Newtonian Physics?** 2026
[Sebastian Cavada](#), Soumava Paul, Tuan-Hung Vu, Andrei Bursuc, Raoul de Charette
Physics-grounded benchmark for evaluating vision-language and foundation models on low-level Newtonian reasoning (11k image sequences, 141k VQA pairs, 60+ evaluated models).
To appear on arXiv, submitted to ECCV 26 [\[Project Page\]](#)

CAD-Assistant: Tool-Augmented VLLMs as Generic CAD Task Solvers 2025
Dimitrios Mallis, Ahmet Serdar Karadeniz, [Sebastian Cavada](#), Danila Rukhovich, Niki Foteinopoulou, Kseniya Cherenkova, Anis Kacem, Djamila Aouada
General-purpose CAD agent combining VLLM planning with FreeCAD-based tool augmentation for multimodal CAD design workflows.
ICCV 2025 [\[arXiv\]](#)

All Languages Matter: Evaluating LMMs on Culturally Diverse 100 Languages 2025
Ashmal Vayani *et al.* (co-author: [Sebastian Cavada](#))
Benchmark for evaluating large multimodal models on culturally diverse visual reasoning across 100 languages and 13 cultural aspects, with strong low-resource language coverage.
CVPR 2025 [\[Paper\]](#)

Research Experience **Research Intern, CovisionLab - Bressanone** February 2026 – Present
Diffusion-driven synthesis of out-of-distribution visual examples to probe and expand perception models. Latent representation analysis from foundation models, prototype alignment, and robustness evaluation.

Collaborator, Astra-Vision group - Inria Paris (hybrid) August 2025 - January 2026
Designed **NewtPhys**, a physics-grounded 4D benchmark using 3D Gaussian Splatting and mesh-free simulation at scale (11k sequences, 141k VQA pairs). Led large-scale evaluation of 60+ vision-language and foundation models, revealing systematic failures in low-level Newtonian reasoning. This also included a 1 month Inria visit in October 2025. Supervised by Prof. [Raoul de Charette](#).

	Visiting Researcher - MBZUAI	Summer 2025
	Extended 3D foundation models with camera-parameter priors for improved pose-aware depth estimation and 3D reconstruction quality in visual perception pipelines. Supervised by Prof. Ian Reid .	
	Research Intern, CVI2 - University of Luxembourg	Summer 2024
	Explored finetuning strategies and supporting data pipelines for CAD-oriented multimodal models, contributing to CAD-Assistant, a tool-augmented framework for generic CAD task solving. Supervised by Dr. Anis Kacem .	
	Research Intern, E3DA - FBK Trento	Summer 2023
	Formulated and released an end-to-end vision pipeline for low-resource deployment, from GPU prototyping (Python) to MCU inference (C). Supervised by Francesco Paissan .	
	Thesis Intern - UniBZ	2022 – 2023
	Developed and analyzed real-time egocentric/exocentric 3D human pose tracking with an indoor digital twin for consistent scene-level perception. Supervised by Prof. Oswald Lanz .	
Thesis & Research Projects	Master's Research Project, MBZUAI	2025
	Research on 3D foundation models and camera-parameter priors to improve perception-to-3D understanding and pose-aware depth estimation for robust visual intelligence. Thesis: Towards a Unified Pipeline for Scalable 3D Reconstruction via Scene Decomposition and Camera-Aware Fusion .	
	Bachelor's Thesis, Free University of Bozen-Bolzano	2023
	Developed a real-time 3D pose tracking system integrating egocentric and exocentric video streams with an indoor digital twin for consistent 3D scene perception. Thesis: <i>Real-Time Human 3D Pose Tracking by Fusing Egocentric and Exocentric Video Streams</i>	
Selected Awards	Fully Funded M.Res. Scholarship Winner	2023 – 2025
	Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), UAE.	
	Winner, Pioneers 4.0 Hackathon	March 2024
	EDGE Group / MoIAT / LIPTON, UAE.	
Technical Skills	Programming: Python, C++, C, C#, CUDA, Git	
	Machine Learning / Vision: PyTorch, torchrun, Weights & Biases, diffusion-based video generation, vision-language model evaluation pipelines, 3D Gaussian Splatting, 3D reconstruction, pose estimation, benchmark design	
	Simulation / 3D Tooling: COLMAP, Blender, Unreal Engine, Gaussian Splatting	
	Compute / Systems: HPC training pipelines, distributed workflows.	
Open-Source & Ongoing Projects	Open-Source Contributions	2024 – Present
	Contributor to micromind-toolkit/micromind ; developed tinygsplat as a compact Gaussian-splatting implementation and benchmarking sandbox.	
	In Progress: orchestra	2026 – Present
	Building orchestra , a broker-worker dispatch layer for model serving focused on queue routing, resource-aware scheduling, and lightweight worker orchestration.	
Teaching & Service	Student Tutor, Computer Science	2021 – 2023
	Free University of Bozen-Bolzano. Provided one-to-one academic tutoring and mentoring.	
	Student Representative, Computer Science Faculty	2020 – 2022
	Free University of Bozen-Bolzano.	

Referees

[Prof. Ian Reid](#) (MBZUAI), [Prof. Raoul de Charette](#) (Inria Paris), and [Prof. Oswald Lanz](#) (Free University of Bozen-Bolzano).