

SERGIO ARNAUD

✉ sergio@sergioarnaud.com 🌐 sergioarnaud.com

Research Focus

Senior Research Engineer at Meta FAIR working on world models for robotics. My research focuses on self-supervised video understanding, visual representations for perception and control, and vision-language grounding for embodied AI.

Experience

Senior Research Engineer

January 2025 – Present

Meta Fundamental AI Research (FAIR)

Menlo Park, CA

- Physical World Modeling.

Research Engineer

March 2024 – December 2024

Meta Fundamental AI Research (FAIR)

Menlo Park, CA

- 3D vision-language grounding for robotic manipulation.

AI Resident

September 2022 – February 2024

Meta Fundamental AI Research (FAIR)

Menlo Park, CA

- Visual representations for robot control and language models for planning.

Tech Lead (AI)

March 2020 – July 2022

deep_dive (dive.ai)

Mexico City, Mexico

- Computer Vision and Natural Language Processing.

Data Scientist

January 2018 – August 2019

deep_dive (dive.ai)

Mexico City, Mexico

- Anomaly detection, Fraud Prevention

Research Collaboration

August 2021 – August 2022

MIT Brain and Cognitive Science

Remote

- Human-Level RL through Theory-Based Modeling with CoCoSci Group.

Education

BSc Applied Mathematics

2015 – 2020

Instituto Tecnológico Autónomo de México (ITAM)

Mexico City

- Graduated with highest honors (Magna Cum Laude), top 3% of students
- GPA: 94/100
- Academic Excellence Scholarship recipient
- ACM-ICPC Regional Finalist (2016, 2017)

Exchange Program

2019 – 2020

Stockholm University

Stockholm, Sweden

- Mathematical optimization and Statistical Learning

Selected Publications

- [1] *V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning*
M. Assran*, A. Bardes*, D. Fan*, Q. Garrido*, R. Howes*, M. Komeili*, M. Muckley*, A. Rizvi*, C. Roberts*, K. Sinha*, A. Zholus*, **S. Arnaud***, et al.
arXiv 2025.
- [2] *Locate 3D: Real-World Object Localization via Self-Supervised Learning in 3D*
S. Arnaud*, P. McVay*, A. Martin*, A. Majumdar, K.M. Jatavallabhula, P. Thomas, R. Partsey, D. Dugas, A. Gejji, A. Sax, et al.
ICML 2025 Spotlight (top 2.5%).
- [3] *Unifying 2D and 3D Vision-Language Understanding*
A. Jain, A. Swerdlow, Y. Wang, **S. Arnaud**, A. Martin, A. Sax, F. Meier, K. Fragkiadaki
ICML 2025.
- [4] *From Thousands to Billions: 3D Visual Language Grounding via Render-Supervised Distillation*
A. Cao, **S. Arnaud**, O. Maksymets, J. Yang, A. Jain, S. Yenamandra, A. Martin, V.-P. Berges, P. McVay, R. Partsey, et al.
ICML 2025.
- [5] *OpenEQA: Embodied Question Answering in the Era of Foundation Models*
A. Majumdar*, A. Ajay*, X. Zhang*, P. Putta, S. Yenamandra, M. Henaff, S. Silwal, P. McVay, O. Maksymets, **S. Arnaud**, K. Yadav, Q. Li, B. Newman, et al.
CVPR 2024.
- [6] *What do we learn from a large-scale study of pre-trained visual representations in sim and real environments?*
S. Silwal*, K. Yadav*, T. Wu*, J. Vakil*, A. Majumdar*, **S. Arnaud***, C. Chen, V.-P. Berges, D. Batra, A. Rajeswaran, et al.
ICRA 2024.
- [7] *Where are we in the search for an Artificial Visual Cortex for Embodied Intelligence?*
A. Majumdar*, K. Yadav*, **S. Arnaud***, Y.J. Ma, C. Chen, S. Silwal, A. Jain, V.-P. Berges, T. Wu, J. Vakil, et al.
NeurIPS 2023.
- [8] *ASC: Adaptive Skill Coordination for Robotic Mobile Manipulation*
N. Yokoyama, A. Clegg, J. Truong, E. Undersander, T.-Y. Yang, **S. Arnaud**, S. Ha, D. Batra, A. Rai
IEEE RA-L 2023.
- [9] *Human-level reinforcement learning through theory-based modeling, exploration, and planning*
Pedro A. Tsividis, Joao Loula, Jake Burga, Juan Pablo Rodriguez, **Sergio Arnaud**, Nathan Foss, Andres Campero, Thomas Pouncy, Samuel J. Gershman, Joshua B. Tenenbaum
arXiv 2021.

Technical Skills

Core Expertise: World Models, Self-Supervised Learning, 3D Vision & Perception, Visual Representations for Control, Foundation Models for Robotics

Deep Learning: PyTorch, Large-Scale Distributed Training, Multimodal Transformer Architectures.

Robotics & Control: Model Predictive Control, Sim-to-Real Transfer, Real-World Robot Learning, Vision-Language-Action Models.

General: Python, CUDA, High Performance Computing, AWS Infrastructure, Production ML Systems.