

# PhD Student, Computer Science / Maths

Research at the intersection of CS and biology

## EXPERIENCE

- 1. 2023-Present** **PhD — Fresnel Inst., Maths Inst. of Marseille, Aix Marseille Univ., CNRS**
  - Increased scores by **2-fold for studying virus trafficking** by designing an LLM + Bayesian filtering framework to improve particle tracking in noisy & cluttered scenes
  - Proved that transformers scale effectively to multiple particle tracking problems **achieving up to 70% accuracy improvement**, where the motion itself is simple but the challenge arises from the super-exponential hypothetical paths that a particle could take
- 2. 2023 (6 mon)** **Research Intern — Maths Inst. of Marseille**
  - Validated the superior computational efficiency of LLM-based tracking, **saving compute by >2x**, relative to track ambiguity through empirical testing and theoretical complexity analysis
  - Established **evidence of flow-aligned glandular growth** in liver organoids, achieving a first-of-its-kind automated quantification of morphogenesis by early-stage prototyping of an image-segmentation framework and **3D reconstruction**
- 3. 2022 (3 mon)** **Student Research Consultant — Stony Brook Univ., NY**

Developed an LLM for **human DNA analysis, pretrained on 500,000 sequences**, improving sequence representations by integrating learning from both nearby and distant DNA patterns
- 4. 2021 (6 mon)** **Software Dev. Intern — Quantiphi Inc., Mumbai**

Optimised internal legal operations by reducing processing latency by **building 15+ major features** of a custom software framework for automated contract handling
- 5. 2019 (3 mon)** **Research Intern — IIT Patna**

Achieved **10% higher accuracy and +0.08 normalized mutual information** for gene clustering across 3 disease types by designing a multi-omics gene clustering framework
- 6. 2018 (6 mon)** **Research Intern — Calcutta Univ.**

Improved **human activity recognition accuracy by 7%**, designing a deep learning strategy that surpassed traditional machine learning benchmarks

---

## GRANTS

- Secured funding of **EUR 22,000** for 6 month-doctoral research by **Turing Centre of Living Systems**
- Helped secure funding of **EUR 419,000** by producing preliminary data for the 2025 **PEPR Math-Vives Grant** "Hybrid Hidden Markov Chain Modelling for the spatial mapping of intracellular trafficking in 3D developing tissues", by **Philippe Roudot** (Fresnel Inst.) & **Claudio Collinet** (Devt. Biology Inst. of Marseille)

---

## EDUCATION

- PhD, CS / Maths, Aix-Marseille Univ. and CNRS (FR) 2023-Present
- MS, Computational and Mathematical Biology, Aix-Marseille Univ. (18.13 / 20) (FR) 2021-2023
- BTech, Computer Science and Engineering, International Inst. of IT, Bhubaneswar (8.51 / 10) (IN) 2017-2021

## **PEER RECOGNITION**

- Talk at Paris Cité University; invited by **Garance Martin** (IUT de Paris - Rives de Seine), 04.2026
  - French round table on informatics oriented theses in Marseille; invited by **Véronique Buat** (Director of the Doctorate School of Physics and Material Sciences), Marseille, 03.2025
  - Workshop on AI and Physical Sciences; invited by **Sandrine Anthoine** (Maths Inst. of Marseille) & **Valentin Emiya** (Laboratory of Informatics & Systems), Marseille, 11.2024
- 

## **SCIENTIFIC OUTREACH**

- French colloquium on AI for biomedical imaging (IABM); Lyon, 03.2026
  - French autumn school on AI and Planetary Boundaries; organised by **Sylvain Bouveret** (Grenoble Informatics Laboratory), Aussois (FR), 10.2025
  - Centuri scientific day; organised by the **Turing Centre for Living Systems**, Palais du Pharo, Marseille, 09.2025
  - Climathiques, French conference on mathematics, climate and ethics; organised by **Indira Chatterjee** (Univ. Côte d'Azur) & the **Centre International de Rencontres Mathématiques (CIRM)**, Marseille, 01.2025
  - 3 Minute-Thesis competition: secured the **13th position**; organised by the **European Association for Signal Processing, Barbara Pascal** (Laboratory of Digital Science of Nantes) & **Juliàn Tachella** (Laboratory of Signals, Systems and Physics), Lyon, 07.2024
- 

## **SELECT PUBLICATIONS**

- **Mishra P. and Roudot P., (2024). Comparative study of transformer robustness for multiple particle tracking without clutter. In EUSIPCO.** This paper offers a proof-of-concept study of the benefit of the transformer architecture in tracking scenarios where the prediction challenge does not come from the complexity of the sequence, but from the number of hypothetical paths the sequence can take. We show that while the transformer requires very little training to significantly outperform conventional tracking approaches on long sequences, it cannot match the theoretically optimal performances of MHT on short sequences even with extensive training. Hence, our work motivates the broader application of transformers in sequences and opens the way to the development of frugal methods thanks to the combination of both statistical and neural network frameworks for particle tracking.
  - **Dutta P., Mishra P., Saha S. (2020). Incomplete multi-view gene clustering with data regeneration using Shape Boltzmann Machine. Computers in Biology and Medicine, Elsevier.** We addressed the problem of clustering gene data when different types of data (views) are missing. Our proposed approach used Shape Boltzmann Machines to regenerate missing data and subsequently perform clustering. This method aims to enhance the accuracy and reliability of gene clustering by leveraging information from multiple data sources, even in the presence of incomplete data. We evaluated our method on several real-world gene datasets and demonstrated its superior performance compared to existing approaches.
- 

**Technical stack:** ML for science, Neural networks (Transformer, CNN, LSTM), Probabilistic models (Bayesian filtering, HMM, Kalman filtering), Representation learning, Interpretability and opening the black box, Algorithm design, Complexity analysis, Population-based heuristics

**Languages:** fluent in English, French, Hindi, Odia