

# Jérémie Dentan

*PhD candidate in Machine Learning  
Security and Privacy*

📞 (+33) 6 95 20 04 42  
✉️ jeremie.dentan@polytechnique.edu  
🌐 <https://dentanjeremie.github.io>  
Paris, France

## About me

I am currently pursuing a PhD in [machine learning security and privacy](#) at École Polytechnique, Paris. My research focuses on emerging challenges in large language models, including data memorization, uncertainty quantification and adversarial robustness. I am seeking a [PhD internship](#) in machine learning.

## Selected Publications

- 2025 *Guess or Recall? Training CNNs to Classify and Localize Memorization in LLMs.*  
**Jérémie Dentan**, Davide Buscaldi, Sonia Vanier. ArXiv Preprint, 2025.  
<https://arxiv.org/abs/2508.02573>
- 2025 *Unveiling Decision-Making in LLMs for Text Classification : Extraction of influential and interpretable concepts with Sparse Autoencoders.*  
Mathis Le Bail, **Jérémie Dentan**, Davide Buscaldi, Sonia Vanier. ArXiv Preprint, 2025.  
<https://arxiv.org/abs/2506.23951>
- 2025 *Predicting memorization within Large Language Models fine-tuned for classification.*  
**Jérémie Dentan**, Davide Buscaldi, Aymen Shabou Sonia Vanier. ECAI 2025.  
<https://arxiv.org/abs/2409.18858>
- 2024 *Reconstructing Training Data From Document Understanding Models.*  
**Jérémie Dentan**, Arnaud Paran, Aymen Shabou. Usenix Security 24.  
<https://www.usenix.org/conference/usenixsecurity24/presentation/dentan>

## Education

- Nov 2023– **PhD Candidate, École Polytechnique , Palaiseau, France .**
- Nov 2026 Measuring and Improving the Trustworthiness of Language Models.
- What amount of training data can we extract from language model?
  - How to audit models during training and protect samples from memorization?
  - Why does memorization happens and where is it encoded in the LLM?
- Sep 2022– **Master MVA: "Mathematics, Vision, Learning", ENS Paris-Saclay , France .**
- Aug 2023 A one year graduating program to specialize in machine learning and applied mathematics.
- ML courses: Convex Optimization, Deep Learning, Kernel Methods.
  - Theoretical courses: Model & Information Theory, Geometric Data Analysis.
  - Applications courses: ML for time series, ML for text and graph data, Detection theory.
  - Seminars: Reproducible research and free software, Responsible ML.
- Final grade: 17.86/20 (highest honours).
- Master thesis: *Reconstructing Training Data from Document Understanding Models.*
- Sep 2019– **Master of Engineering, École Polytechnique Paris , France .**
- Aug 2022 "Cycle ingénieur polytechnicien, promotion X2019"
- **Computer Science and Machine Learning:** Algorithmics, Deep Learning, Reinforcement Learning, Systems for Big Data, NLP, Database Management, Data Visualization, Federated Learning, Web Development, Object-oriented programming, Theoretical Computer Science.
  - **Applied Mathematics:** Statistics, Optimization and Control, Statistics.
  - **Business:** Microeconomics, Macroeconomics, Corporate Governance and CSR.
- Cumulative GPA: 3.89 (top 20%).

Sep 2017–	<b>"Classe Préparatoire MPSI/MP*", Sainte Geneviève High School , France .</b>
Aug 2019	Similar to the beginning of a BSc in Mathematics and Physics, it is a 2-year intensive program preparing for the national competitive exam for entry to engineering schools. <b>Cumulative GPA:</b> 3.99 (ranked 5 out of 150).
Jun 2017	<b>French Baccalaureate, Military High School of Aix-en-Provence , France .</b> Mathematics: 18th at "Concours Général" (most selective A-level competition in France). <b>Ranking:</b> 1st of Aix-Marseille Academy (>25k students). Grade: 20.68/20.

## Work experience

Apr 2023–	<b>Research Assistant, Crédit Agricole DataLab Group , Paris, France .</b>
Sept 2023	Worked on the first <a href="#">model inversion attacks</a> that target document processing models. <ul style="list-style-type: none"> <li>○ Adapted existing methods coming from computer vision and NLP to develop the <a href="#">first existing attack</a> against document AI.</li> <li>○ Designed an evaluation protocol and conducted extensive empirical evaluation of this attack.</li> <li>○ <a href="#">Paper under review</a> for Usenix Security Symposium 2024.</li> </ul>
Apr 2022–	<b>Research Assistant, Oracle Labs , Zürich, Switzerland .</b>
Aug 2022	Worked on an internal tool for risk prediction, with a focus on its <a href="#">entity resolution</a> pipeline. <ul style="list-style-type: none"> <li>○ Used entity resolution techniques and applied them to real-world data pertaining to companies, including deduplication, record linkage and canonicalization.</li> <li>○ Initiated and implemented a cross-project general-purpose entity resolution <a href="#">API to centralize work</a> on that field.</li> </ul>
Jun 2021–	<b>Business Analyst, Public Sector team, BearingPoint , Paris, France .</b>
Aug 2021	Various missions for the <a href="#">digital transformation of French administrations</a> , including: diagnostic phase; project scope; funding plan; optimization of a financial audit software; software validation phase; project owner support.
Jan 2021–	<b>Project Manager and Quality Manager, XProjets , Palaiseau, France .</b>
Dec 2021	A small consulting firm (100–150k€ turnover) run by a dozen students of Polytechnique. <ul style="list-style-type: none"> <li>○ As quality manager, <a href="#">optimized the internal processes</a> for our missions.</li> <li>○ As project manager, ran several missions in technology for firms, startups and public administration, mainly in web development.</li> </ul> <p><a href="#">Main accomplishment in 2021:</a> XProjets ranked in L30 (top 30 junior companies in France).</p>
Nov 2019–	<b>Deputy Project Manager, French Embassy , Antananarivo, Madagascar .</b>
Mar 2020	Internship as a French armed-forces officer. Placement in the <a href="#">diplomatic community</a> , and especially in the team of French-Malagasy cooperation for civil protection. Led Malagasy firemen in a fire station, and organized formations for Malagasy firemen and policemen.

## Languages and IT

Languages	French (native), English (fluent, TOEFL 100/120), German (elementary level, B1 certified), Spanish (beginner).
IT Languages	Preferred language: Python. Good knowledges in JS, HTML. Knowledges in Java, C++, D3, VBA.

## Interests and activities

Passions	Very strong interest for classical and romantic music, played the piano for 12 years.
Sports	Running, climbing, and mountaineering. I have also practiced tennis, judo, and badminton.
Army	Merit Medal of the Malagasy Army, 2020, in recognition of my work for the military cooperation.