

# William de Vazelhes

PhD candidate in ML at MBZUAI

Dubai, UAE, 30 years old  
+971 585192388  
wdevazelhes@gmail.com  
wdevazelhes.github.io  


## Summary

- **9 years of experience in ML.**
- **PhD in ML at MBZUAI** (expected 2024) and **MSc in ML at Supélec** (2016).
- **Publications in NeurIPS, ICLR, JMLR, AAAI and IJCAI.**
- **scikit-learn contributor** (implemented `NeighborhoodComponentsAnalysis` + various PRs) and core contributor of `metric-learn` (Python package with **+1.3K stars on Github** and a **JMLR paper with +60 citations**).
- **Skills** : Math for ML (Linear Algebra, Optimization, Stats. & Proba., Real Analysis, etc.), Python : `numpy` `scipy` `scikit-learn` `pandas` `matplotlib` (advanced), `pytorch` `jax` `tensorflow` (medium).

## Education

- Aug. 2021 to May 2024 (expected) **PhD Candidate**, MBZUAI, UAE, Machine Learning Major, *Thesis PDF, Video, Slides.*
- **Supervisor** : Bin Gu
  - **Topic** : Hard-Thresholding algorithms. (Optimization (Zeroth Order, First Order, Deterministic & Stochastic), guarantees for Sparse Recovery).
  - Applications : Medical Imaging, Portfolio Optimization.
  - 1st year courses : Optimization, Advanced ML, Probabilistic ML (Cumulative GPA : **3.9**)
- Aug. 2012 to Aug. 2016 **Master of Science**, SUPELEC (*highly selective "Grande Ecole"*), (⊂ CentraleSupélec ⊂ Paris-Saclay), France, SIR major (Machine Learning and Signal Processing).  
Machine Learning, Numerical Optimization, Image and Speech Processing, Statistics and Probabilities, High Performance Computing, Software Engineering, Electrical Engineering.
- Jan. 2014 to June 2014 **ERASMUS Programme**, UNIVERSIDAD CARLOS III, Madrid, Spain.  
Medical Image Processing, Multimodality Imaging, Telecommunications, Communication Theory.
- 2010 to 2012 **Preparatory School**, STANISLAS, Paris, *PCSI-PSI*.  
Intensive preparation in Mathematics and Physics for the French "*Grandes Ecoles*" competition. Set Theory, Linear Algebra, Real Analysis, Multivariate Analysis, Differential Equations, Physics.

## Professional Experience

- 2023-2024 **Reviewer**, ICML (2024), ICLR (2023), NEURIPS (2023).
- Spring 2023 **Teaching Assistant**, MBZUAI, Abu Dhabi, UAE, *Syllabus here.*  
Optimization course (MTH702) : helping students during labs and grading assignments.
- Dec. 2019 to April 2021 **PhD Student**, HUAWEI & UNIVERSITE GUSTAVE EIFFEL, Paris, France.
- **Supervisor** : Abdellatif Zaidi
  - **Topic** : Variational Autoencoders through the lens of Rate-Distortion Theory, Minimax Rates in Wasserstein distance for Kernel Density Estimation, Wasserstein Barycenter for Data Fusion.
  - Implemented an algorithm for the Wasserstein Barycenter in Mindspore (Huawei's ML framework).
- Sept. 2017 to Aug. 2019 **Research Engineer**, INRIA, Magnet Team, Lille, France.
- **Supervisor** : Aurélien Bellet
  - Developed a `scikit-learn` compatible package for metric learning. (<https://github.com/scikit-learn-contrib/metric-learn>) (**1.3K+☆, 60+ cit.**).
  - Improved the speed, memory cost, and robustness of pre-existing algorithms by vectorizing computations in novel ways, changing the iterative methods for optimizing the algorithms, solving problems due to numerical uncertainties, and check-proofing the algorithms by testing them on mathematical toy problems.
  - Contributed to `scikit-learn`. Main contribution : `NeighborhoodComponentsAnalysis` (NCA).
- Oct. 2016 to July 2017 **Data Scientist**, SIDETRADE, Boulogne-Billancourt, France.
- Improved the number of records successfully fuzzy matched in the company's heterogeneous database by ~ 30%, using `elasticsearch` + query randomization in Python + ML-based results filtering in `scikit-learn`. Developed both the prototype and part of the production code (pickling, and efficient batch prediction API).
  - Prototyped mail classification algorithms to help send the appropriate invoice reminder or warning to consumers.
  - Prepared the team's GPU server to integrate it with Dataiku's built-in machine learning interface.

- Apr. 2016 **Research Intern**, ORANGE LABS, Chatillon, France, *Report here*.  
to Sept. 2016 - **Supervisor** : Romain Laroche  
- Implemented Deep RL algorithms (DQN, DDQN, A3C) for Human-Machine Dialogue, in Python and Java.
- Jan. 2015 **Research Intern**, POLYTECHNIQUE MONTREAL, Montreal, Canada, *Report here*.  
to July 2015 - **Supervisor** : Yves Goussard  
- Derived a MAP cost function for CT reconstruction (and its gradient in closed form), optimized it with l-bfgs-b, and integrated the code within the Matlab CT simulation library of the lab.
- 2014 **Physics and Maths Tutor**, Paris, France.

## Publications (\* : equal contribution)

- 2024 **IJCAI**, Hard-Thresholding Meets Evolution Strategies in Reinforcement Learning, *Chengqian Gao\**, *William de Vazelhes\**, *Hualin Zhang*, *Bin Gu*, *Zhiqiang Xu*.
- 2024 **ICLR**, New Insight of Variance reduce in Zero-Order Hard-Thresholding : Mitigating Gradient Error and Expansivity Contradictions, *Xinzhe Yuan*, *William de Vazelhes*, *Bin Gu*, *Huan Xiong*. [Paper] [Supplementary and Code]
- 2024 **AAAI**, Iterative Regularization with k-support Norm : An Important Complement to Sparse Recovery, *William de Vazelhes*, *Bhaskar Mukhoty*, *Xiao-Tong Yuan*, *Bin Gu*. [Paper] [Code]
- 2024 **AAAI**, Limited Memory Online Gradient Descent for Kernelized Pairwise Learning with Dynamic Averaging, *Hilal AlQuabeh*, *William de Vazelhes*, *Bin Gu*.
- 2023 **NeurIPS**, Direct Training of SNN using Local Zeroth Order Method, *Bhaskar Mukhoty\**, *Velibor Bojković\**, *William de Vazelhes*, *Xiaohan Zhao*, *Giulia De Masi*, *Huan Xiong*, *Bin Gu*. [Paper] [Code]
- 2022 **NeurIPS**, Zeroth-Order Hard-Thresholding : Gradient Error vs. Expansivity, *William de Vazelhes*, *Hualin Zhang*, *Huimin Wu*, *Xiao-Tong Yuan*, *Bin Gu*. [Paper] [Poster] [Video] [Supplemental and Code]
- 2022 **ICDM**, Efficient Semi-Supervised Adversarial Training without Guessing Labels, *Huimin Wu*, *William Vazelhes*, *Bin Gu*. [Paper]
- 2020 **JMLR**, metric-learn : Metric Learning Algorithms in Python, *William de Vazelhes*, *CJ Carey*, *Yuan Tang*, *Nathalie Vauquier*, *Aurélien Bellet*. [Github Repo] [Paper] [Video]

## Talks

- April 2024 **PhD Defense**. Abu Dhabi, UAE. On Iterative Hard-Thresholding : Gradient Estimation and Non-Convex Projections. *Video*, *Slides*.
- May 2023 **Candidacy Exam Presentation**. Abu Dhabi, UAE. Zeroth-Order Hard-Thresholding, Sparsity with k-support Norm, Perspectives. *Slides*.
- Dec. 2022 **NeurIPS**. New Orleans, USA. Zeroth-Order Hard-Thresholding : Gradient Error vs. Expansivity. *Video and Slides*, *Poster*.
- March 2022 **MBZUAI Group Seminar**. Abu Dhabi, UAE. Forward Mode and Directional Derivatives. *Slides*.
- Nov. 2021 **MBZUAI Group Seminar**. Abu Dhabi, UAE. Presenting the literature on sparse optimization to the group. *Slides*.
- Oct. 2018 **PyConFR**. Lille, France. Metric-learn : a scikit-learn compatible package for metric learning. *Video*, *Slides*.

## Awards

- Oct. 2022 NeurIPS 2022 Scholar Award.

## Software

### Python libraries

`scikit-learn`. Contributions : `NeighborhoodComponentsAnalysis` + various PRs.

`metric-learn`. Core contributor. +1.3K stars on Github and JMLR paper with +60 citations.

### Sprints

- Feb. 2019 `scikit-learn` sprint in Paris, France.

July 2018 `scikit-learn` sprint during SciPy in Austin, Texas, USA.

### Languages, Libraries, and Tools

Python: `numpy` `scipy` `scikit-learn` `pandas` `matplotlib` (advanced), `pytorch` `jax` `tensorflow` (medium). Linux, OS X, Windows, Git, Conda, Pip, VSCode, Emacs, Pycharm, Vim.

### Projects

- Spring 2022 **Final project for ML702**, MBZUAI, 3 teammates, Report here.  
Literature review on Bi-level optimization.
- Fall 2021 **Final project for ML701**, MBZUAI, 3 teammates, Code here, Report here.  
"*Attacks4Good*" : Data poisoning to improve fairness, using Bayesian Optimization.
- Oct. 2015 **Robot control with speech recognition**, SUPELEC, 2 teammates, *Code here*, *Report here*.  
to Jan. 2016 Implemented a robot voice control with a KNN with DTW distance and a dataset of commands MFCCs (MATLAB, C++, ROS).
- Sept. 2013 **Game creation**, SUPELEC, 2 teammates.  
to Dec. 2013 Implemented a "connect the dots" video game in JAVA.
- Apr. 2013 **Neural firing rates analysis**, SUPELEC, 2 teammates, *Code and report here* : [goo.gl/XmnQnT](http://goo.gl/XmnQnT).  
to June 2013 Implemented an Izhikevich spiking neural network model and studied its various regimes. (MATLAB)

### Languages

English (TOEFL 112/120), French (Mother Tongue), Spanish (Intermediate), Moroccan (Beginner).

### Volunteering

- 2012 to 2014 CHEER UP! : visiting teenagers suffering from cancer, to help them realize projects (e.g. learning the guitar, meeting celebrities etc). Organized a golf tournament for fundraising.