# Tuan Le, Ph.D.

tle618m@gmail.com
tuanle618.github.io
tuanle618

## EDUCATION

**Freie Universität Berlin** *October 2020 - February 2025*
Ph.D. Computer Science Final grade: *summa cum laude*
Work in `AI4Sciences` research group led by Prof. Dr. Frank Noé
Dissertation title: Graph Representation Learning of Molecules for *de novo* Drug Development.
Research in neural network architecture design for supervised learning on small molecules with follow-up on generative modeling for structure-based drug design.

**Ludwig Maximilian University of Munich** *May 2017 - November 2019*
M.Sc. Statistics
Master's thesis: *De novo drug design in continuous space.* Grade: 1.0 (very good)
Course work in classical statistical methods such as survival analysis, stochastic processes, numerical optimization and predictive modeling with machine learning and deep learning.

**Karlsruhe Institute for Technology** *September 2013 - September 2016*
B.Sc. Industrial Engineering
Bachelor's thesis: *Bayesian regression using asymmetric loss functions.* Grade: 1.7 (very good)
Course work in basic and advanced statistics, econometrics, operations research.

## PROFESSIONAL EXPERIENCE

**Senior Research Scientist**, Pfizer in Berlin *February 2023 - present*
Work in *Machine Learning Research Team* led by Dr. Djork-Arné Clevert.
Research on generative models for structure-based drug design, including diffusion and flow matching approaches.
Work spans model development, multi-objective molecular optimization, and deployment to internal drug discovery teams.

**Research Scientist - PhD student**, Bayer AG in Berlin *February 2020 - January 2023*
Work in *Machine Learning Research Team* led by Dr. Djork-Arné Clevert.
Research on graph neural networks for molecular property prediction and equivariant architectures for 3D molecules.

**Master's thesis & Data Science intern**, Bayer AG in Berlin *October 2018 - November 2019*
Master's thesis on generative modeling for de novo drug design.
Internship work on deep learning for text classification.

## SKILLS

| | |
|---|---|
| **Programming Languages** | Python, R, Bash, Linux, LaTeX |
| **ML & Scientific Computing** | PyTorch, PyTorch Lightning, NumPy, SciPy, Pandas, Scikit-Learn |
| **MLOps & Infrastructure** | MLflow, Docker, SLURM, Git |
| **Cheminformatics** | RDKit, PyMol, Biotite, |
| **Languages** | Fluent in German, English, Vietnamese; basic French and Polish |

## RESEARCH PUBLICATIONS (* EQUAL CONTRIBUTION / FIRST AUTHORSHIP)

- Preprints

    1. J. Cremer*, T. Le, M. M. Ghahremanpour, E. Sługocka, F. Menezes*, D-A. Clevert, *FLOWR.root: A flow matching based foundation model for joint multi-purpose structure-aware 3D ligand generation and affinity prediction*, **arXiv 2025**

- Journal Articles

    1. T. Le*, Y. Guan, D-A. Clevert, K. Schütt, *Coupled fragment-based generative modeling with stochastic interpolants*, **Digital Discovery, 2026**

    2. T. Le*, J. Cremer*, D-A. Clevert, K. Schütt, *Equivariant diffusion for structure-based de novo ligand generation with latent-conditioning*, **Journal of Cheminformatics 2025**

    3. K. Elez, T. Hempel, J. H. Shrimp, N. Moor, L. Raich, C. Rocha, R. Winter, T. Le, S. Pöhlmann, M. Hoffmann, M. D. Hall, F. Noé, *Simulations and active learning enable efficient identification of an experimentally-validated broad coronavirus inhibitor*, **Nature Communications, 2025**

    4. J. Cremer*, T. Le*, F. Noé, D-A. Clevert, K. Schütt, *PILOT: equivariant diffusion for pocket-conditioned de novo ligand generation with multi-objective guidance via importance sampling*, **Chemical Science, 2024**

    5. P. A. Marin Zapata*, O. Méndez-Lucio*, T. Le, C. J. Beese, J. Wichard, D. Rouquié and D. Clevert, *Cell morphology-guided de novo hit design by conditioning GANs on phenotypic image features*, **Digital Discovery, 2023**

    6. D-A. Clevert*, T. Le, R. Winter, F. Montanari *Img2Mol – accurate SMILES recognition from molecular graphical depictions*, **Chemical Science, 2021**xw

    7. T. Le*, R. Winter, F. Noé, D-A. Clevert, *Neuraldecipher – Reverse-Engineering Extended-Connectivity Fingerprints (ECFPs) to their Molecular Structures*, **Chemical Science, 2020**

- Conference Articles

    1. M. Bertolini*, T. Le*, D-A. Clevert, *Diffusion Generative Modeling on Lie Group Representations*, **NeurIPS 2025, Spotlight Paper**

    2. T. Le*, J. Cremer*, D-A. Clevert, K. Schütt, *Latent-Guided Equivariant Diffusion for Controlled Structure-Based De Novo Ligand Generation*, **Machine Learning for Life and Material Sciences Workshop at ICML 2024**

    3. T. Le*, J. Cremer*, F. Noé, D-A. Clevert K. Schütt, *Navigating the Design Space of Equivariant Diffusion-Based Generative Models for De Novo 3D Molecule Generation*, **ICLR 2024**

    4. T. Le*, F. Noé, D-A. Clevert, *Representation Learning on Biomolecular Structures using Equivariant Graph Attention,* **Learning on Graphs Conference 2022**

    5. R. Winter*, M. Bertolini*, T. Le, F. Noé, D-A. Clevert, *Unsupervised Learning of Group Invariant and Equivariant Representations,* **NeurIPS 2022**

    6. T. Le*, M. Bertolini, F. Noé, D-A. Clevert, *Parameterized Hypercomplex Graph Neural Networks for Graph Classification*, **Artificial Neural Networks and Machine Learning – ICANN 2021**

    7. T. Le*, M. Bertolini*, M. Boef*, F. Montanari, D-A. Clevert, *Going full hyper: hyperbolic and hypercomplex graph embeddings for ADMET modeling* **Machine Learning for Molecules Workshop at NeurIPS 2020**

## TALKS

1. *Structure-based Drug Design using 3D Equivariant Generative Models*, **Third AiChemist School + AIChemist CECAM Flagship School in Lausanne**, (23 April 2025, invited talk)
2. *Unsupervised Learning of Molecular Conformations*, **Young Researcher's Workshop on Machine Learning for Materials 2022**, (09-13 May 2022, contributed talk)
3. *Neuraldecipher – Reverse-Engineering Extended-Connectivity Fingerprints (ECFPs) to their Molecular Structures*, **9th RDKIT User-Group-Meeting, 2020**, (06 October 2020, contributed talk)

## PARTICIPATION

1. Machine Learning Summer School 2022, Cracow: (27.06.2022 - 02.07.2022)
2. Young Researcher's Workshop on Machine Learning for Materials 2022, Trieste: (09.05.2022 - 13.05.2022)

## ACADEMIC SERVICE

- Reviewer for: RSC Chemical Science, ICANN, JCIM, ML4-Molecules Workshops, ICLR, SPIGM @ NeurIPS
- Member of the program committee for the ML4LM workshop at ICML 2024 in Vienna
- Volunteering work for organizing the RDKit User Group Meeting (UGM) 2022 in Berlin