Georgios Smyrnis

gsmyrnis@utexas.edu – georgiossmyrnis.github.io

Education

University of Texas at Austin **Austin TX, USA**
Ph. D. in Electrical and Computer Engineering *08/2020–08/2025*
- Academic Track: Decision, Information and Communications Engineering.
- Dissertation: "*Data Curation for Foundation Model Training*"
- Part of the Wireless Networking and Communications Group.
- Acquired Master of Science in Engineering in ECE on May 2023, while working towards my final degree.

National Technical University of Athens (NTUA) **Athens, Greece**
Diploma in Electrical and Computer Engineering *09/2014–07/2020*
(*Joint bachelor's & master's degree*)
- Concentration: Machine Learning, Computer Science.
- Thesis: "*Tropical Polynomial Division and Neural Network Minimization*".
- Final Grade: 9.94/10 (1st place during graduation).

Professional and Research Experience

Anthropic PBC **San Francisco CA, USA**
Member of Technical Staff *08/2025–current*

Toyota Research Institute **Los Altos CA, USA**
Intern *05/2024–08/2024*
- Research internship over the summer.

Google LLC **Mountain View CA, USA**
Student Researcher *05/2022–08/2022, 09/2022–10/2022*
- Summer student researcher position. Part time student researcher during September & October.

University of Texas at Austin **Austin TX, USA**
Graduate Research Assistant *06/2021–05/2022, 09/2022–01/2024 09/2024–08/2025*
- Performing research on the fields of contrastive learning and self-supervision.

School of ECE, University of Texas at Austin **Austin TX, USA**
Teaching Assistant *09/2020–05/2021 01/2024–05/2024*
- Teaching assistant for the course "Data Science Lab".

National Center for Scientific Research "Demokritos" **Athens, Greece**
Intern *09/2019–10/2019*
- Interned at the Institute of Informatics and Telecommunications, Computational Intelligence Lab.
- Implemented a system for real-time action classification using data from a Kinect camera.
- Evaluated various methods for segmentation of such data into parts containing human actions.

National Technical University of Athens (NTUA) **Athens, Greece**
Laboratory Teaching Assistant *09/2016–01/2018, 09/2018–05/2019, 03/2020–07/2020*
- Assisted with lab exercises for "Computer Programming" and "Programming Techniques" courses, until 2019.
- During 2020, assisted with lab exercises for the course "Computer Vision".

Publications & Preprints

- E. Guha, R. Marten, S. Keh, N. Raoof, G. Smyrnis et al., "OpenThoughts: Data Recipes for Reasoning Models", in *ICLR 2026*, 2026.
- J. Li, A. Fang, G. Smyrnis, M. Ivgi, M. Jordan, S. Y. Gadre, et al. "DataComp-LM: In search of the next generation of training sets for language models." in *NeurIPS 2024, Datasets and Benchmarks Track*, 2024.
- S. Y. Gadre, G. Smyrnis, V. Shankar, S. Gururangan, M. Wortsman, R. Shao, J. Mercat et al. "Language

models scale reliably with over-training and on downstream tasks." in *ICLR 2025*, 2025.
- G. Smyrnis, S. Ravula, S. Sanghavi, and A. Dimakis, "Multimodal Distillation of CLIP Models", in *NeurIPS 2023 Workshop: Self-Supervised Learning - Theory and Practice*, 2023
- E. Tsaprazlis, G. Smyrnis, A. Dimakis, and P. Maragos, "Enhancing CLIP with a Third Modality", in *NeurIPS 2023 Workshop: Self-Supervised Learning - Theory and Practice*, 2023
- S. Y. Gadre, G. Ilharco, A. Fang, J. Hayase, G. Smyrnis, T. Nguyen, R. Marten et al. "DataComp: In search of the next generation of multimodal datasets.", in *NeurIPS 2023 Datasets and Benchmarks Track*, 2023.
- G. Smyrnis, M. Jordan, A. Uppal, G. Daras and A. Dimakis, "Lovasz Theta Contrastive Learning", in *NeurIPS 2022 Workshop: Self-Supervised Learning - Theory and Practice*, 2022
- P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "Neural Network Approximation based on Hausdorff distance of Zonotopes", in *ICLR 2022*, 2022
- S. Ravula, G. Smyrnis, M. Jordan and A. Dimakis, "Inverse Problems Leveraging Pre-trained Contrastive Representations", in *Proc' NeurIPS 2021*, 2021
- G. Smyrnis and P. Maragos, "Multiclass Neural Network Minimization via Tropical Newton Polytope Approximation" in *Proc' ICML 2020*, 2020.
- G. Smyrnis, P. Maragos and G. Retsinas, "Maxpolynomial Division with Application To Neural Network Simplification" in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 2020.
- G. Pikramenos, K. Kechagias, T. Psallidas, G. Smyrnis, E. Spyrou and S. Perantonis, "Dimensionality Reduction and Attention Mechanisms for Extracting Affective State from Sound Spectrograms" in *International Conference on Pattern Recognition Applications and Methods 2020* (*selected papers*), 2020.
- G. Pikramenos, G. Smyrnis, I. Vernikos, T. Konidaris, E. Spyrou and S. Perantonis, "Sentiment Analysis from Sound Spectrograms via Soft BoVW and Temporal Structure Modelling", in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM*, 2020.
- G. Smyrnis and P. Maragos, "Tropical Polynomial Division and Neural Networks", *arXiv:1911.12922*, 2019.

Volunteering Experience

- Organizer for the Data-centric Machine Learning Research Workshop at ICML 2024.
- Reviewer for various iterations of ICML, NeurIPS and ICLR.

Honors & Awards

- **Onassis Scholarship** for doctorate studies.
- **"C. Chrysovergis" & "I. Kondoulis" Prizes**, for 1st place graduation from undergraduate studies (2020).
- **Scholarship**, for undergraduate studies, "Ialemos Kyprianidis" bequest.
- **"Paris Kanellakis" Prize**, for highest grades in Information Technology courses (2016-2017, 2017-2018).
- **"Thomaideion" Award** (1st place) for course grades (2015-2016, 2016-2017, 2017-2018).
- **"KARY" Award** for highest course grades (2015-2016).
- **International Physics Olympiad 2014**, Member of Greek delegation, Honorable Mention.

Skills

- Programming: Python, MATLAB, C/C++.
- Software Tools/Libraries: Pytorch, Keras.
- Skills: Machine Learning, Computer Vision, Natural Language Processing.
- Languages: English (fluent), French, Greek (native).