

# Giannis Karamanolakis

Senior Applied Scientist at Amazon AGI  
New York, NY 10001

WWW: <https://gkaramanolakis.github.io>  
Scholar: [link](#) (1400+ citations, 10+ h-index)  
E-mail: [gkaraman@cs.columbia.edu](mailto:gkaraman@cs.columbia.edu)

## Professional and Research Experience

---

6. **Amazon AGI**, New York, NY, United States. 05/2025 - now  
*Senior Applied Scientist*  
*Applied Scientist* 09/2022 - 05/2025  
Group: AGI team under Rohit Prasad (2023-now), Before: NLU team at Alexa AI (2022-2023).
  - Training Amazon Nova foundation models from inception to production (Amazon blog, Forbes).
  - Developed novel methods for large-scale LLM pre-training, fine-tuning, evaluation, and data processing.
  - Implemented API grounding methods for the first LLM-powered Alexa+ version (Amazon blog, WSJ, CNET).
  - Mentored Ph.D. interns on LLM training, distillation, and merging (ACL '24, NeurIPS ENLSP '24, NAACL '25).
5. **Columbia University**, Department of Computer Science, New York, NY, United States. 09/2017 - 08/2022  
*Graduate Research Assistant*  
Advisors: Luis Gravano, Daniel Hsu.  
Groups: Machine Learning group, NLP group, Database Research group.
  - Developed minimally-supervised machine teaching frameworks (EMNLP'19/20, LOUHI'20, SocialNLP'20, TACL'24).
  - Deployed machine learning systems in health departments to analyze social media for public health (NSF).
  - Increased the recall of food poisoning events by 48.6% via a new weakly-supervised neural network (WNUT'19).
4. **Microsoft Research AI**, Redmond, WA, United States. 05/2020 - 08/2020  
*Research Scientist Intern*  
Supervisors: Ahmed Awadallah, Subho Mukherjee, Guoqing Zheng (Language & Information Technologies team).
  - Developed ASTRA, a semi-supervised learning method for self-training with expert-defined rules (NAACL '21).
  - Designed WALNUT, a benchmark on semi/weakly-supervised learning for natural language tasks (NAACL '22).
3. **Amazon**, Seattle, WA, United States. 05/2019 - 12/2019  
*Applied Scientist / Machine Learning Scientist Intern*, Supervisor: Xin Luna Dong (Product Graph team).
  - Developed TXtract, a neural network for large-scale knowledge extraction from product descriptions (ACL '20 paper).
  - Improved product understanding through 15% higher coverage in knowledge graph construction (KDD '20 paper).
2. **Behavioral Signals**, Los Angeles, United States (worked remotely from Athens, Greece). 06/2016 - 06/2017  
*Machine Learning Engineer*, Supervisors: Shri Narayanan, Alexandros Potamianos.
  - Trained statistical models for recognizing core emotions from conversational speech.
  - Used speaker interaction patterns to improve the prediction of key performance indicators.
1. **NCSR "Demokritos"**, Computational Intelligence Laboratory, Athens, Greece. 11/2015 - 2/2016  
*Research Intern*, Supervisors: Theodoros Giannakopoulos, Stavros Perantonis.
  - Developed audio-based classifiers for acoustic event detection and soundscape quality estimation.

## Education

---

2. **Columbia University**, New York, NY, United States. 09/2017 - 08/2022  
*Ph.D. in Computer Science, M.Phil. in Computer Science, M.Sc. in Computer Science* GPA: 4.26/4.00  
 Advisors: Prof. Luis Gravano, Prof. Daniel Hsu.  
 Ph.D. Dissertation: “Efficient Machine Teaching Frameworks for Natural Language Processing”.
1. **National Technical University of Athens**, Athens, Greece. 09/2011 - 07/2017  
*M.Eng. in Electrical & Computer Engineering (5-year study program, MSc equivalent)* GPA: 9.26/10.00  
 Advisor: Prof. Alexandros Potamianos.  
 Dissertation: “Grounding Word Semantic Representations to Acoustic and Visual Perceptual Modalities”.

## Selected Publications

---

5. Amazon Artificial General Intelligence. **The Amazon Nova family of models: Technical report and model card.** *Amazon Technical Reports*, 2024.
4. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Interactive Machine Teaching by Labeling Rules and Instances.** *Transactions of the Association for Computational Linguistics (TACL)*, 2024.
3. **Giannis Karamanolakis**, Subhabrata Mukherjee, Guoqing Zheng, and Ahmed Hassan Awadallah. **Self-Training with Weak Supervision.** In *NAACL*, 2021 (Oral).
2. **Giannis Karamanolakis**, Jun Ma, and Xin Luna Dong. **TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories.** In *ACL*, 2020 (Oral).
1. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training.** In *EMNLP*, 2019 (Oral).

## Selected Fellowships, Honors, and Awards

---

8. **Columbia PhD Award**, Davide Giri Memorial Prize, New York, US 05/2022
7. **Leventis Foundation Scholarship**, New York, US 08/2019, 08/2020, 08/2021
6. **Gerondelis Foundation Scholarship**, New York, US 06/2018
5. **Joseph F. Traub Fellowship**, Columbia University, New York, US 04/2018
4. **Addressing the Opioid Epidemic, Design Challenge Award**, Columbia University, New York, US 12/2017
3. **Columbia Engineering Presidential Fellowship**, New York, US 08/2017 - 05/2022
2. **Eurobank EFG Scholarship**, Athens, Greece 06/2012
1. **Hellenic Mathematic Society “Euclides” & “Thales” Awards**, Sitia, Greece 02/2009, 03/2010

## Technical Skills, Talks, and Academic Service

---

- **Programming:** Python (PyTorch, Tensorflow, Keras, Pyro, Huggingface, Spacy, Gensim), C++, Java, MATLAB, C.
- **Agile Development and Project Management:** Git, JIRA, Jenkins, Scrum, Kanban.
- **Invited Industry Talks:** Two Sigma (Dec. ’20), RelationalAI (May ’20), NYAS (Mar. ’20), AT&T Labs (Dec. ’19).
- **Conference/Workshop Talks:** NAACL ’21, DeeLIO ’20, ACL ’20, EMNLP ’19, RecSys ’18, InterSpeech ’16.
- **Reviewer Activity:** NeurIPS ’19-24 (top 10% reviewer), ICML ’21-24 (expert reviewer), ICLR ’24, ACL ’19-24, EMNLP ’21-24, ACL Rolling Review ’21, Machine Learning Journal ’21, TrueFact ’20, TKDD journal ’19, WWW ’19.
- **Mentoring:** Designed research projects for 14 students (8 undergraduate, 6 graduate); 7 published papers.

## All Publications

---

21. Amazon Artificial General Intelligence. **The Amazon Nova family of models: Technical report and model card.** *Amazon Technical Reports*, 2024.
20. Yuhang Zhu, **Giannis Karamanolakis**, Victor Soto, Mayank Kulkarni, Anna Rumshisky, Furong Huang, Wei Ai, and Jianhua Lu. **MergeME: Model Merging Techniques for Homogeneous and Heterogeneous MoEs.** *NAACL*, 2025.
19. Namrata Shivagunde, Mayank Kulkarni, **Giannis Karamanolakis**, Jack GM FitzGerald, Yannick Versley, Saleh Soltan, Volkan Cevher, Jianhua Lu, and Anna Rumshisky. **Approximations may be all you need: Towards pre-training LLMs with low-rank decomposition and optimizers.** *NeurIPS 2024 Workshop on Efficient Natural Language and Speech Processing (ENLSP-IV)*, 2024.
18. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Interactive Machine Teaching by Labeling Rules and Instances.** *Transactions of the Association for Computational Linguistics (TACL)*, 2024.
17. Guoqing Zheng, **Giannis Karamanolakis**, Kai Shu, and Ahmed Hassan Awadallah. **WALNUT: A Benchmark on Semi-weakly Supervised Learning for Natural Language Understanding.** In *NAACL*, 2022 (Oral).
16. Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Anjana Arunkumar, Arjun Ashok, Arut Selvan Dhanasekaran, Atharva Naik, David Stap, Eshaan Pathak, **Giannis Karamanolakis**, Haizhi Gary Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Maitreya Patel, Kuntal Kumar Pal, Mehrad Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Shailaja Keyur Sampat, Savan Doshi, Siddhartha Mishra, Sujan Reddy, Sumanta Patro, Tanay Dixit, Xudong Shen, Chitta Baral, Yejin Choi, Hannaneh Hajishirzi, Noah A. Smith, and Daniel Khashabi. **Benchmarking Generalization via In-Context Instructions on 1,600+ Language Tasks .** In *EMNLP*, 2022.
15. **Giannis Karamanolakis**, Subhabrata Mukherjee, Guoqing Zheng, and Ahmed Hassan Awadallah. **Self-Training with Weak Supervision.** In *NAACL*, 2021 (Oral).
14. Ivy Cao, Zizhou Liu, **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Quantifying the Effects of COVID-19 on Restaurant Reviews.** In *EMNLP 9th International Workshop on Natural Language Processing for Social Media (SocialNLP)*, 2021 (Oral).
13. Tejit Pabari, **Giannis Karamanolakis**, Max Mauerman, Beth Tellman, Mitchell Thomas, and Eugene Wu. **Flood Event Extraction from Newspaper Media in Bangladesh.** In *Book of Abstracts of the 8th International Conference on Water and Flood Management (ICWFM)*, 2021.
12. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Cross-Lingual Text Classification with Minimal Resources By Transferring a Sparse Teacher.** In *Findings of EMNLP*, 2020.
11. Ziyi Liu, **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Detecting Foodborne Illness Complaints in Multiple Languages Using English Annotations Only.** In *EMNLP Workshop on Health Text Mining and Information Analysis*, 2020 (Oral).
10. **Giannis Karamanolakis**, Jun Ma, and Xin Luna Dong. **TXtract: Taxonomy-Aware Knowledge Extraction for Thousands of Product Categories.** In *ACL*, 2020 (Oral).
9. Xin Luna Dong, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, Chenwei Zhang, Tong Zhao, Gabriel Blanco Saldana, Saurabh Deshpande, Alexandre Michetti Manduca, Jay Ren, Surender Pal Singh, Fan Xiao, Haw-Shiuan Chang, **Giannis Karamanolakis**, Yuning Mao, Yaqing Wang, Christos Faloutsos, Andrew McCallum, and Jiawei Han. **AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types.** In *KDD*, 2020 (Oral).
8. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training.** In *EMNLP*, 2019 (Oral).
7. **Giannis Karamanolakis**, Daniel Hsu, and Luis Gravano. **Weakly Supervised Attention Networks for Fine-Grained Opinion Mining and Public Health.** In *EMNLP Workshop on Noisy User-generated Text (WNUT)*, 2019 (Oral).

6. **Giannis Karamanolakis, Daniel Hsu, and Luis Gravano. Training Neural Networks for Aspect Extraction Using Descriptive Keywords Only.** In *ICLR 2nd Workshop on Learning from Limited Labeled Data (LLD)*, 2019.
5. Katy Ilonka Gero, **Giannis Karamanolakis**, and Lydia Chilton. **Transfer Learning for Style-Specific Text Generation.** In *NeurIPS Workshop on Machine Learning for Creativity and Design*, 2018.
4. **Giannis Karamanolakis, Kevin Raji Cherian, Ananth Ravi Narayan, Jie Yuan, Da Tang, and Tony Jebara. Item Recommendation with Variational Autoencoders and Heterogenous Priors.** In *RecSys Workshop on Deep Learning for Recommender Systems*, 2018 (Oral).
3. Georgios Paraskevopoulos, **Giannis Karamanolakis**, Elias Iosif, Aggelos Pikrakis, and Alexandros Potamianos. **Sensory-Aware Multimodal Fusion for Word Semantic Similarity Estimation.** In *EUSIPCO Multimodal Processing, Modeling, and Learning (Multi-Learn) Workshop*, 2017 (Oral).
2. **Giannis Karamanolakis, Elias Iosif, Athanasia Zlatintsi, Aggelos Pikrakis, and Alexandros Potamianos. Audio-based Distributional Semantic Models for Music Auto-tagging and Similarity Measurement.** In *EUSIPCO Multimodal Processing, Modeling, and Learning Workshop*, 2017 (Oral).
1. **Giannis Karamanolakis, Elias Iosif, Athanasia Zlatintsi, Aggelos Pikrakis, and Alexandros Potamianos. Audio-Based Distributional Representations of Meaning Using a Fusion of Feature Encodings.** In *INTERSPEECH*, 2016 (Oral).