Idan Schwartz

Senior Lecturer (Assistant Professor)
Department of Computer Science and Artificial Intelligence, Bar-Ilan University, Ramat Gan, Israel

idanschwartz@gmail.com · idansc.github.io · Google Scholar · idansc · 0000-0002-4379-1987

Research Interests

Multimodal learning, mostly generative: aligning vision, language, audio, and video. Attention as the recurring theme, from high-order and factor-graph attention modules for multimodal fusion to attention guidance in diffusion models. Cognitive-science questions about deep learning, including perception and comprehension.

Academic Positions

2025–	**Senior Lecturer (Assistant Professor)**, Department of Computer Science and Artificial Intelligence, Bar-Ilan University. Founded and lead the Multimodal Lab.
2023–2024	**Lecturer** (part-time), School of Computer Science, Tel Aviv University.
2022–2023	**Postdoctoral Researcher**, School of Computer Science, Tel Aviv University. Multimodal generative models, with Prof. Lior Wolf.
2018–2019	**Visiting Scholar**, University of Illinois at Urbana-Champaign, with Prof. Alexander G. Schwing.

Education

2015–2022	**Ph.D. in Computer Science**, Technion – Israel Institute of Technology (direct track). Thesis: *Cognitive Models in Deep Learning*. Advisors: Prof. Tamir Hazan (Technion) and Prof. Alexander G. Schwing (UIUC).
2011–2015	**B.Sc. in Computer Science**, Technion – Israel Institute of Technology.

Industry Experience

2025–	**Chief Scientist**, Aigency.
2019–2025	**Head of Research**, Spot by NetApp. Led ML research for cloud-workload optimization: data-driven algorithms from the cloud behavior of thousands of customers; six patents (below).
2019	**Researcher**, Microsoft, Search, Assistant and Intelligence group. NLP for meeting insights: deep learning models that extract action items from transcripts.
2016–2018	**Researcher**, eBay, Catalog group. Led product-deduplication research with computer vision and NLP, reducing eBay's duplicates from 30% to 12%; with Dr. Ido Guy and Dr. Kira Radinsky.
2011–2015	**Software Developer**, Intel, Haifa. A C++ framework for chip testing.
2008–2011	**Full-Stack Web Developer**, IDF service. Systems processing millions of records.

Publications

Google Scholar, September 2026: 1,353 citations, h-index 15. Full list on Google Scholar.

2026	LaMI: Augmenting Large Language Models via Late Multi-Image Fusion G. Yariv, **I. Schwartz**, Y. Adi, S. Benaim · **ACL 2026**
	TempoControl: Temporal Attention Guidance for Text-to-Video Models S. Schiber, O. Lindenbaum, **I. Schwartz** · **CVPR 2026**
	Detection-Driven Object Count Optimization for Text-to-Image Diffusion Models O. Zafar, Y. Cohen, L. Wolf, **I. Schwartz** · **WACV 2026**
	Single Image Iterative Subject-driven Generation and Editing Y. Shpitzer, A. Weidenfeld, G. Chechik, **I. Schwartz** · **CVPR 2026, P13N Workshop**
2024	Diverse and Aligned Audio-to-Video Generation via Text-to-Video Model Adaptation G. Yariv, I. Gat, S. Benaim, L. Wolf, **I. Schwartz**[*], Y. Adi[*] · **AAAI 2024**

2023 Discriminative Class Tokens for Text-to-Image Diffusion Models
I. Schwartz, V. Snæbjarnarson, H. Chefer, S. Belongie, L. Wolf, S. Benaim · **ICCV 2023**

Zero-Shot Video Captioning with Evolving Pseudo-Tokens
Y. Tewel, Y. Shalev, R. Nadler, **I. Schwartz**, L. Wolf · **BMVC 2023**

AudioToken: Adaptation of Text-Conditioned Diffusion Models for Audio-to-Image Generation
G. Yariv, I. Gat, L. Wolf, Y. Adi[*], **I. Schwartz**[*] · **INTERSPEECH 2023**

2022 Optimizing Relevance Maps of Vision Transformers Improves Robustness
H. Chefer, **I. Schwartz**, L. Wolf · **NeurIPS 2022**

ZeroCap: Zero-Shot Image-to-Text Generation for Visual-Semantic Arithmetic
Y. Tewel, Y. Shalev, **I. Schwartz**, L. Wolf · **CVPR 2022**

Describing Sets of Images with Textual-PCA
O. Hupert, **I. Schwartz**, L. Wolf · **Findings of EMNLP 2022**

Ordered Attention for Coherent Visual Storytelling
T. Braude, **I. Schwartz**, A. G. Schwing, A. Shamir · **ACM Multimedia 2022**

Latent Space Explanation by Intervention
I. Gat, G. Lorberbom, **I. Schwartz**, T. Hazan · **AAAI 2022**

Video and Text Matching with Conditioned Embeddings
A. Ali, **I. Schwartz**, T. Hazan, L. Wolf · **WACV 2022**

2021 Perceptual Score: What Data Modalities Does Your Model Perceive?
I. Gat, **I. Schwartz**, A. G. Schwing · **NeurIPS 2021**

Ensemble of MRR and NDCG Models for Visual Dialog
I. Schwartz · **NAACL 2021**
Winner of the Visual Dialog Challenge 2020.

2020 Removing Bias in Multi-Modal Classifiers: Regularization by Maximizing Functional Entropies
I. Gat, **I. Schwartz**, A. G. Schwing, T. Hazan · **NeurIPS 2020**

2019 Factor Graph Attention
I. Schwartz, S. Yu, T. Hazan, A. G. Schwing · **CVPR 2019**
First place in the Visual Dialog Challenge 2019 on MRR, R@1, R@5, R@10, and Mean Rank.

A Simple Baseline for Audio-Visual Scene-Aware Dialog
I. Schwartz, A. G. Schwing, T. Hazan · **CVPR 2019**

2017 High-Order Attention Models for Visual Question Answering
I. Schwartz, A. G. Schwing, T. Hazan · **NeurIPS 2017**

[*] Equal contribution.

Patents

2026 Adaptive Request Grouping Based on Node Pool Impact
I. Schwartz, R. Kramer, S. Jacobson, I. Haskel, T. S. Shafir, T. Ohayon, et al. · US application 18/924,118

2025 Cloud Instance Type Recommendations
I. Schwartz, G. Yariv, T. Ohayon · US application 18/924,802

Resource Distribution Engine(s) for Allocating and Securing Reclaimable Resources Within a Cloud Environment
I. Schwartz, T. Ohayon, G. Yariv, O. Gurfinkel, M. Goldberg, R. Vladimirsky, et al. · US application 18/634,377

2024 Interruption Predictions for Cloud Compute Instances
I. Schwartz, O. Muchnik, J. Cohen, K. McGrath, A. Shachar · US patent 11,915,053 (granted)

Identifying Anomalous Activities in a Cloud Computing Environment
Y. Shen, A. Benameur, A. X. Ough, I. Schwartz · US application 18/344,664

Spare Resource Availability Prediction with Limited Historical Data
T. Ohayon, I. Schwartz · US application 18/046,970

2023 Search System for Providing Web Crawling Query Prioritization Based on Classification Operation Performance
I. Guy, I. Schwartz, K. Radinsky · US patent 11,636,164 (granted)

Teaching

2025–	**Bar-Ilan University.** Developed three new courses in the Department of Computer Science: *Deep Learning* (89-6877), *Multimodal Deep Learning* (89-604), and *Multimodal Generative AI* (89-4018), bringing multimodal fusion, diffusion and autoregressive generation, and inference-time control into the curriculum with extensive hands-on programming.
2023–2024	**Tel Aviv University.** Lecturer: *Deep Learning* (0368-4488), *Workshop in Multimodal* (0368-3562).
2015–2022	**Technion.** Lecturer, Autumn Data Science School (with Dr. Kira Radinsky). Guest lecturer: Deep Learning (097200, 236606), Natural Language Processing (097215), Deep Learning for NLP (232601). Teaching assistant in charge: Advanced Data Science (236605), Theory of Compilation (236360). Teaching assistant: Introduction to Systems Programming (234122).

Professional Service

Area Chair	ICML 2025, 2026; NeurIPS 2025, 2026; CVPR 2026.
Organization	Co-organizer, *The Multimodal Day*, Bar-Ilan University, May 28, 2025: a full-day meeting of the Israeli multimodal-learning community with nine invited speakers. Co-organizer, 15th Israel Machine Vision Conference (IMVC 2026).

Selected Talks

2024	Multimodal Attention, Perception, Comprehension.
2022	Zero-Shot Vision-to-Text Methods, seminar, UC Berkeley.
2019	Attention Models for Vision and Language, Tel Aviv University.