

Yonatan Bitton, Curriculum Vitae, December 2023

Contact Information

yonatanbitton.github.io

yonatanbitton1@gmail.com

linkedin.com/in/yonatanbitton

Current Positions

Research Scientist, Multimodal Consistency, Google Research 2023-Present
Vision-and-language. Recent works include image-text alignment, improving text-to-image models, and visual instruction tuning.

Education

PhD in Computer Science, The Hebrew University of Jerusalem 2020–2023
Advisors: Prof. Gabriel Stanovsky and Prof. Roy Schwartz
Thesis: Bridging Vision and Language with Data.

MSc in Computer Science, *magna cum laude*, Ben Gurion University 2019–2020
Advisors: Prof. Michael Elhadad and Prof. Eitan Bachmat
Thesis: Cross-lingual entity linking and visual question answering. GPA 97

BSc in Computer Science, Ben Gurion University, 2015-2019 2015–2019

Work Experience[†]

Research Intern, Google 2022-2023
Cerebra team: focusing on conversational AI, engaged with leading language models (LaMDA, PaLM, BARD); leveraged synthetic data for [query generation](#), crafted personalized agents, and augmented LLM memory capabilities.

Applied Scientist, Amazon Lab126 2019-2022
[Visual Fitness Halo Team](#) - Developed a virtual fitness trainer, specializing in 2D/3D pose estimation, action recognition, error correction, on-device deployment and more.

Researcher, IBM Research 2017-2019
Developing machine-learning methods to detect frauds

Peer-Reviewed Publications

* indicates equal contribution. For abstracts and more information, see [Google Scholar](#).

- [1] **Mismatch Quest: Visual and Textual Feedback for Image-Text Misalignment**
Gordon. G*, **Bitton. Y***, Shafrir. Y, Garg. R, Chen. X, Lischinski. D, Cohen-Or D, Szpektor. I
arXiv preprint
- [2] **VideoCon: Robust Video-Language Alignment via Contrast Captions**
Bansal. H, **Bitton. Y**, Szpektor. I, Kai-Wei. C, Grover. A
arXiv preprint
- [3] **VisIT-Bench: A Benchmark for Vision-Language Instruction Following Inspired by Real-World Use**
Bitton. Y*, Bansal. H*, Hessel. J*, Shao. R, Zhu. W, Awadalla. A, Gardner. J, Taori. R, Schimdt. L
Neural Information Processing Systems Datasets and Benchmarks Track (**NeurIPS 2023**)
- [4] **ParallelPARC: A Scalable Pipeline for Generating Natural-Language Analogies**
Sultan. O, Yosef. R, **Bitton. Y**, Shahaf. D
arXiv preprint
- [5] **VisIT-Bench: A Benchmark for Vision-Language Instruction Following Inspired by Real-World Use**

[†] Parallel to studies.

Bitton. Y*, Bansal. H*, Hessel. J*, Shao. R, Zhu. W, Awadalla. A, Gardner. J, Taori. R, Schimdt. L
Neural Information Processing Systems Datasets and Benchmarks Track (**NeurIPS 2023**)

- [6] **Read, Look or Listen? What’s Needed for Solving a Multimodal Dataset**
Madvil. N, **Bitton. Y**, Schwartz. R
arXiv preprint
- [7] **Transferring Visual Attributes from Natural Language to Verified Image Generation**
Valerio. R, Bordalo. J, Yarom. M, **Bitton. Y**, Szpektor. I, Magalhaes. J
arXiv preprint
- [8] **What You See is What You Read? Improving Text-Image Alignment Evaluation**
Bitton. Y*, Yarom. M*, Changpinyo. S, Aharoni. R, Herzig. J, Lang. O, Ofek. E, Szpektor. I
Neural Information Processing Systems (**NeurIPS 2023**)
- [9] **q2d: Turning Question into Dialogs to Teach Models How to Search**
Bitton. Y, Cohen. S, Hakimi. I, Lewenberg. Y, Aharoni. R, Weinreb. E,
Conference on Empirical Methods in Natural Language Processing: **EMNLP 2023**
- [10] **DataComp: In search of the next generation of multimodal datasets via data scaling**
Yitzhak. S, Ilharco. G, Fang. A, Hayase. J, Smyrnis. G, Nguyen. T, Marten. R, Wortsman. M,
Ghosh. D, Zhang. J, Orgad. E, Entezari. R, Daras. G, Pratt. S, Ramanujan. V, **Bitton. Y**,
Mussmann. S, Vencu. R, Cherti. M, Krishna. R, Wei. P, Saukh. O, Ratner. A, Song. S, Hajishirzi.
H, Farhadi. A, Beaumont. R, Oh. S, Dimakis. A, Jitsev. J, Carmon. Y, Shankar. V, Schmidt. L
Neural Information Processing Systems Datasets and Benchmarks Track (**NeurIPS 2023**)
- [11] **OpenFlamingo: An open-source framework for training vision-language models with in-context learning**
Awadalla. A, Gao. I, Gardner. J, Hessel. J, Hafany. Y, Zhu. W, Gedre. S, **Bitton. Y**, Kalyani.
M, Kornblith. S, Koh. P, Ilharco. G, Wortsman. M, Schmidt. L
Blog release: <https://laion.ai/blog/open-flamingo/>
- [12] **IRFL: Image Recognition of Figurative Language**
Yosef. R, **Bitton. Y**, Shahaf. D
Findings of the Conference on Empirical Methods in Natural Language Processing: **EMNLP 2023**
- [13] **WHOOOPS! A Vision-and-Language Commonsense Benchmark of Heterogeneous Objects and Situations**
Guetta. N*, **Bitton. Y***, Hessel. J, Schmidt. L, Elovici. Y, Stanovsky. G, Schwartz. R,
International Conference on Computer Vision (**ICCV 2023**)
Neural Information Processing Systems Creative AI Track (**NeurIPS 2023**) - Gallery
- [14] **VASR: Visual Analogies of Situation Recognition**
Bitton. Y, Yosef. R, Strugo. E, Shahaf D, Schwartz. R, Stanovsky. G
Association for the Advancement of Artificial Intelligence (**AAAI 2023**)
Selected as an **Oral Presentation**
- [15] **WinoGAViL: Gamified Association Benchmark to Challenge Vision-and-Language Models**
Bitton. Y*, Guetta. N*, Yosef. R, Bansal. M, Stanovsky. G, Schwartz. R,
Neural Information Processing Systems Datasets and Benchmarks Track (**NeurIPS 2022**)
Selected as a **Featured Presentation** (Updated version of “Oral Presentation”)
- [16] **Data Efficient Masked Language Modeling For Vision and Language**
Bitton. Y, Stanovsky. G, Elhadad. M, Schwartz. R,
Findings of the Conference on Empirical Methods in Natural Language Processing: **EMNLP 2021**
- [17] **Automatic Generation of Contrast Sets from Scene Graphs: Probing the Compositional Consistency of GQA**
Bitton. Y, Stanovsky. G, Schwartz. R, Elhadad. M,
North American Chapter of the Association of Computational Linguistics (**NAACL 2021**)

[18] **Cross-lingual Unified Medical Language System entity linking in online health communities**

Bitton. Y, Cohen. R, Schifter. T, Bachmat. E, Elhadad. M, Elhadad. N
Journal of the American Medical Informatics Association (**JAMIA 2020**)

Selected Awards and Scholarships

PHD AWARDS

KLA Scholarship for Outstanding Graduate Students 2022

MSC AWARDS

Dean's Award for Excellence 2020

Graduated with honors (*magna cum laude*) 2020

Computer Science Department Research Excellence Award for journal publication 2020

Advising

STUDENTS MENTORED AT THE HEBREW UNIVERSITY

Oren Sultan 2023

Nitzan Bitton-Guetta 2022-2023

Netta Madvil 2022-2023

Ron Yosef 2022-2023

Professional Activities

CONFERENCE REVIEWER

Conference on Empirical Methods in Natural Language Processing (EMNLP), Industry Track 2023

Annual Meeting of the Association of Computational Linguistics (ACL) 2023

North American Chapter of the Association of Computational Linguistics (NAACL) 2022

NeurIPS Datasets and Benchmarks 2022

Annual Meeting of the Association of Computational Linguistics (ACL) 2021

Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021

Invited Talks

Bridging Vision and Language with Data: From Perception to Understanding April-June 2023

Hebrew University of Jerusalem, NLP-IL Reading Group, Microsoft Israel (MSAI-HIVE team), Meta AI Research Tel-Aviv, Technion, Ben Gurion University, Google Tel-Aviv, Bar-Ilan University, IBM Research (Israel NLP team), Tel Aviv University

Talk record is available in [YouTube](#)

Commonsense Benchmarks for Vision and Language November 2022

NLP Seminar at Cornell Tech, Google Research Israel, the Hebrew University of Jerusalem

q2d: Turning Questions into Dialogs to Teach Models How to Search September 2022

Conversational applications with LLMs - Summit in Google Zurich

WinoGAViL: Gamified Association Benchmark to Challenge Vision-and-Language Models June 2022

IBM Research Israel

VASR: Visual Analogies of Situation Recognition May 2022

Computer Vision Seminar at the Hebrew University of Jerusalem

Open Source

Breaking Common Sense: WHOOPS! A Vision-and-Language Benchmark of Synthetic and Compositional Images

Project website: <https://whoops-benchmark.github.io/>

Huggingface dataset: <https://huggingface.co/datasets/nlphuji/whoops>

WinoGAViL: Gamified Association Benchmark To Challenge Vision-And-Language Models

Project website: <https://winogavil.github.io/>

Software: <https://github.com/WinoGAViL/WinoGAViL-experiments>

VASR: Visual Analogies of Situation Recognition

Project website: <https://vasr-dataset.github.io/>

Software: <https://github.com/vasr-dataset/vasr>

Data Efficient Masked Language Modeling for Vision and Language

Software: https://github.com/yonatanbitton/data_efficient_masked_language_modeling_for_vision_and_language

Automatic Generation of Contrast Sets from Scene Graphs

Software: https://github.com/yonatanbitton/automatic_generation_of_contrast_sets_from_scene_graphs

Cross-lingual unified medical language system entity linking in online health communities

Software: <https://github.com/yonatanbitton/mdtel>