Yonatan Bitton, Curriculum Vitae, July 2023

Contact Information

yonatanbitton.github.io yonatanbitton1@gmail.com linkedin.com/in/yonatanbitton

Current Positions

Research Scientist, Consistency Team, 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 Shwartz
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] **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
arXiv preprint

[2] **Transferring Visual Attributes from Natural Language to Verified Image Generation**
Valerio. R, Bordalo. J, Yarom. M, **Bitton. Y**, Szpektor. I, Magalhaes. J
arXiv preprint

[3] **WYSIWYR: 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
arXiv preprint

[4] **q2d: Turning Question into Dialogs to Teach Models How to Search**
Bitton. Y, Cohen. S, Hakimi. I, Lewenberg. Y, Aharoni. R, Weinreb. E,
arXiv preprint

[†] Parallel to studies.

[5] **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
arXiv preprint

[6] **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/

[7] **IRFL: Image Recognition of Figurative Language**
Yosef. R, **Bitton. Y**, Shahaf. D
arXiv preprint

[8] **WHOOPS! A Vision-and-Language Commonsense Benchmark of Heterogeneous Objects and Situations**
Guetta. N*, **Bitton. Y***, Hessel. J, Schmidt. L, Elovici. Y, Stanovsky. G, Shwartz. R,
Under Review

[9] **VASR: Visual Analogies of Situation Recognition**
Bitton. Y, Yosef. R, Strugo. E, Shahaf D, Shwartz. R, Stanovsky. G
Association for the Advancement of Artificial Intelligence (**AAAI 2023**)
Selected as an **Oral Presentation**

[10] **WinoGAViL: Gamified Association Benchmark to Challenge Vision-and-Language Models**
Bitton. Y*, Guetta. N*, Yosef. R, Bansal. M, Stanovsky. G, Shwartz. R,
Neural Information Processing Systems Datasets and Benchmarks Track (**NeurIPS 2022**)
Selected as a **Featured Presentation** (Updated version of "Oral Presentation")

[11] **Data Efficient Masked Language Modeling For Vision and Language**
Bitton. Y, Stanovsky. G, Elhadad. M, Shwartz. R,
Findings of the Association for Computational Linguistics: **EMNLP 2021**

[12] **Automatic Generation of Contrast Sets from Scene Graphs: Probing the Compositional Consistency of GQA**
Bitton. Y, Stanovsky. G, Shwartz. R, Elhadad. M,
North American Chapter of the Association of Computational Linguistics (**NAACL 2021**)

[13] **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

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

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