Yoav Gur Arieh

📞 054-941-2753　✉ yoavgurarieh@gmail.com　in @yoav-gur-arieh

Graduate researcher in NLP and AI with 9 years of software engineering experience, specializing in interpretability and knowledge adaptation in large language models.

EDUCATION

Tel Aviv University — Computer Science MSc Student, *The Adi Lautman Interdisciplinary Program for Outstanding Students*

2022 – 2025

- Third-year student at Tel Aviv University's Lautman honors program, pursuing a computer science degree within a four-year **direct-to-masters** track emphasizing interdisciplinary study, with a focus on computer science.
- Engaging in diverse coursework spanning mathematics, computational biology, economics, and history, fostering a broad skill set and interdisciplinary understanding.
- Taking advantage of the program's flexibility in order to take high-level courses regardless of prerequisites, independently meeting their demands through self-learning.

Current grade average: **93.52**

EXPERIENCE

Graduate Researcher – *NLP and Interpretability (Dr. Mor Geva's Lab)*

August 2024 – Present

- Researching novel methods of knowledge adaptation in LLMs, employing interpretability techniques to analyze inter-module communication and optimize knowledge integration within the model.
- Working on new methods for large scale automatic interpretation of sparse auto-encoder features, emphasizing compute and memory efficiency, while defining rigorous criteria for effective explanations.
- Establishing a large-scale data collection pipeline with hermetically isolated and parsed entities and facts, designed for training LLMs with precise knowledge attribution.

Laminar Security / Rubrik — *Senior Software Developer*

March 2021 - Present

- Joined the startup during its inception and spearheaded the development of crucial front-end, back-end, and agent-based systems from scratch.
- Led the development of a document sensitivity classification engine, involving parsing documents, analyzing their content, and categorizing their sensitivity levels.
- Conducted technical research on extracting decrypted data from encrypted traffic in cloud-based clusters, and successfully implemented an efficient solution.

IDF - 8200 Unit (Subunit 7149) — *Senior Developer & Researcher*

March 2016 - March 2021

- Carried out months-long solo research projects into esoteric and opaque technologies that have little or no publicly available documentation, culminating in the development of technical solutions.
- Led through to completion highly complex, high-risk projects involving multiple teams.
- Worked on smaller sporadic projects entailing deep dives into complex topics.

Checkpoint Software Technologies — *Software Developer*

August 2015 - March 2016

- Contributed to the development of a log aggregator and analyzer product tailored for large enterprises. Designed and implemented automated reports featuring visualization of insights.

SKILLS

Python · PyTorch · HF Transformers · TransformerLens · SAE Lens · NLP Interpretability · Golang · C · Javascript · CSS · eBPF · Linux Internals · Embedded systems · Networks · Cyber Security · Reverse Engineering

AWARDS

Personal award from the Military Intelligence Directorate for excellence (including a rank promotion)

Award for oustanding project (8200)

Worked on two separate projects that received the Israel Defense Prize

First place team in the 2015 *'Skillz'* coding competition

PROJECTS

Avian Neuronal Response Project – Analyzed avian neuronal activity in response to varied bird calls (authentic and artificial), and used machine learning techniques to classify stimuli.

Parkalot – Developed an app and backend infrastructure to display current space availability in Tel Aviv parking lots and predict future trends.

Hoppa – Developed a platformer genre android game using the Unity game engine.

LANGUAGES

Hebrew and **English** - Native proficiency.

SCORES

Psychometric: 750
Bagrut Average: 112