idanshenfeld@gmail.com \ idanshen@mit.edu
857-3332981
www.idanshenfeld.com

# IDAN SHENFELD

## PROFILE

EECS PhD candidate at MIT, working on reinforcement learning algorithms and their applications in NLP and robotics. Over three years of industrial research experience in leading global companies.

## EDUCATION

**PH.D. IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, MIT**

- September 2022 – present. (GPA 5.00/5.00)
- Awarded the Qualcomm Innovation Fellowship 2024 and the Google-MIT Collaboration Grant.
- Advised by **Professor Pulkit Agrawal**.

**B.S. IN COMPUTER ENGINEERING, TECHNION**

- Class of 2021, Summa Cum Laude (GPA 94.8/100).
- Expedited course of study (3 years instead of 4); Rothschild scholarship; Apple Award for Excellence BSc Students; Dean's or Rector's List, all semesters.
- Conducted research in Reinforcement Learning under the supervision of **Professor Aviv Tamar**, and in Geometrical Learning under the supervision of **Professor Ron Kimmel**.

## PROFESSIONAL EXPERIENCE

**STUDENT RESEARCHER, GOOGLE DEEPMIND**
June 2024- January 2025

- Student researcher at the Gemini Safety team where I worked with **Ahmad Beirami** and **Preethi Lahoti**.
- Conduct research on reasoning capabilities and knowledge distillation.

**APPLIED RESEARCHER SCIENTIST, GENERAL MOTORS AV PROJECT**
August 2021- September 2022

- Conducted research as part of the Perception group in GM's autonomous vehicle project.
- Focused on problems such as General Obstacle Detection, Road Segmentation, Sensor Fusion and more.

**MACHINE LEARNING ALGORITHM ENGINEER, SAMSUNG FLASH SOLUTIONS RESEARCH LAB (AFSL)**
October 2017- October 2018

- Researched ML algorithms for storage systems.
- Led the development of innovative error correction modules integrating classical ECC techniques with deep learning algorithms.

**DATA AND ML DEPARTMENT LEADER, UNIT 8200, ISRAELI DEFENCE FORCE**
July 2015- October 2017

- Military service at Unit 8200, the Israeli equivalent of the NSA. Finished the service as an officer at the rank of First Lieutenant.
- Started as a data analyst; at the end managed 4 teams with a total of over 40 employees.
- Received an Award of Excellence for exemplary departmental performance.

## SELECTED PUBLICATIONS

(FOR A FULL LIST CHECK MY GOOGLE SCHOLAR)

- **Self-Distillation Enables Continual Learning.**
  *Idan Shenfeld*, Mehul Damani, Jonas Hübotter, Pulkit Agrawal. ICML 2026 (under review)

- **RL's Razor: Why Online Reinforcement Learning Forgets Less.**
  *Idan Shenfeld*, Jyo Pari, Pulkit Agrawal. ICLR 2026.

- **Online Language Model Personalization via Reward Factorization.**
  *Idan Shenfeld*, Felix Faltings, Pulkit Agrawal, Aldo Pacchiano. COLM 2025.

- **Value Augmented Sampling for Language Model Alignment and Personalization.**
  *Idan Shenfeld*, Seungwook Han, Akash Srivastava, Yoon Kim, Pulkit Agrawal. ICLR 2024 Workshop on Reliable and Responsible Foundation Models (*Oral*).

- **Curiosity-driven Red-teaming for Large Language Models.**
  Zhang-Wei Hong, *Idan Shenfeld*, Tsun-Hsuan Wang, Yung-Sung Chuang, Aldo Pareja, James R. Glass, Akash Srivastava, Pulkit Agrawal. ICLR 2024.