

Dor Bernsohn

AI ENGINEER · DATA SCIENTIST

Ramat Gan, Israel · 054-680-7054 · dor.bernsohn@gmail.com

[LinkedIn](#) · [GitHub](#) · [Hugging Face](#)

AI Engineer and Data Scientist with 10 years of experience building and shipping NLP and LLM systems end-to-end, with a track record of taking novel research methods into production.

EXPERIENCE

AI Engineer · *Stealth Startup* 2025 – Present

- Fine-tuning small language models on proprietary company data to solve complex, domain-specific tasks.

Senior Data Scientist · *Darrow.ai* 2021 – 2025

Promoted from Data Scientist (2021 – 2023) to Senior Data Scientist (2023 – 2025).

- Designed and built the data pipelines and research workflows that powered the company's original case-discovery engine.
- Developed the first automated underwriting system in the legal-claims industry, reducing case evaluation from days of manual analysis to minutes.
- Initiated and built an LLM routing and distillation pipeline: routed production traffic through teacher APIs (OpenAI, Gemini, Anthropic) and automatically fine-tuned open-source student models (Qwen, Llama) once enough data accumulated — enabling faster, privacy-preserving on-prem deployment.

Data Scientist · *Verint* 2018 – 2021

- Replaced legacy rule-based systems with transformer-based ML/DL models, measurably improving accuracy and robustness.
- Adapted methods from published research to customer-specific problems and data.
- Built demos and POCs to showcase product capabilities to prospective customers.

Data Analyst · *Bank of Israel* 2016 – 2018

- Built predictive models and statistical analyses, using hypothesis testing to validate assumptions and inform division strategy.
- Presented findings to the division to guide strategy and operations.

PUBLICATIONS

LeggLens: Leveraging LLMs for Legal Violation Identification in Unstructured Text · *EACL 2024 (Oral)* ·

arxiv.org/abs/2402.04335

ClassActionPrediction: A Challenging Benchmark for Legal Judgment Prediction of Class Action Cases in the US · *NLLP*

Workshop @ EMNLP 2022 · arxiv.org/abs/2211.00582

SKILLS

LLMs	Fine-tuning open-source models, distillation, routing, vLLM inference; OpenAI, Anthropic, Gemini APIs
ML & NLP	HuggingFace (Transformers, PEFT, TRL), deep learning, applied NLP research, statistical modeling
Engineering	Python, Docker, Git, Linux
Data & Infra	SQL, PostgreSQL, Elasticsearch, FAISS, AWS
Languages	Hebrew (native) · English (fluent)

EDUCATION

M.A. in Statistics · *Hebrew University of Jerusalem* 2017 – 2019

B.A. in Statistics · *Hebrew University of Jerusalem* 2014 – 2017

MILITARY SERVICE

Combat Soldier · *Orev Nahal, IDF* 2010 – 2013