+972-53-827-2753
idanhorowitz8@gmail.com

Idan Horowitz

Data Scientist

https://www.linkedin.com/in/idanho/
https://github.com/idanh8
Portfolio: https://idanhorowitz.me

Data Scientist who owns problems end to end, from messy raw data to working systems, with industry and research experience across the full data science pipeline. Combines experimental design and creative problem solving with hands-on implementation. Comfortable leading projects, learning unfamiliar tools, delivering under pressure and explaining complex ideas to technical and non-technical audiences.

TECHNICAL SKILLS

Languages	**Proficient:** Python, SQL. **Experience:** R, Java, C++. **Spoken:** English (native), Hebrew (native).
Tech Stack	PyTorch, PyTorch Geometric, HuggingFace Transformers, Azure Databricks, PySpark, LangChain, Qdrant, Pinecone, Git, Pandas, GeoPandas, NumPy, scikit-learn, SciPy, XGBoost, statsmodels, matplotlib, seaborn, Streamlit.

EXPERIENCE

Data Scientist | Caps for Sale (CFS) 2022 – 2024

- Owned end-to-end data pipelines for commercial real-estate risk models, turning 15–20 messy datasets per chain into model-ready features through API ingestion, cleaning, address matching, and spatial-data integration.
- Reduced geographic feature-extraction runtime by 75%, through spatial search filtering and precomputation, preserving outputs while accelerating model iteration.
- Built the full feature engineering suite for the real estate risk model, including integrating macroeconomic and spatial data; learned GeoPandas from scratch to engineer novel spatial features, improving model prediction accuracy by 4%.
- Trained and evaluated XGBoost models to assess engineered features; used model comparisons and SHAP analysis to communicate findings and trade-offs to guide model development and business decisions.
- Communicated model findings, limitations, and proposed solutions to technical and non-technical stakeholders, including executives and colleagues. Documented workflows in a 50-page handover and trained the successor.
- Built the data pipeline and features powering investment reports and product demos that supported a $2M seed raise.

Teaching Assistant | Faculty of Data & Decision Science, Technion 2024 – 2026

- Taught Python data analysis and machine learning tutorials in a 250-student course; developed assignments and mentored project teams, earning a teaching evaluations of 4.46/5.
- Redesigned and tested PySpark assignments and a project for a 100-student course on Azure Databricks, covering data cleaning, ETL, exploratory analysis, and modeling with Spark ML.

EDUCATION

M.Sc. Data Science | Technion | GPA 97.7 2024 – 2026

- Developed NeSCal, a novel train-time calibration loss over embedding-space graphs, advised by Prof. Avigdor Gal. Derived theoretical bounds and connections to manifold regularization. Explored links to GNNs and locally adaptive calibration.
- Built the PyTorch evaluation pipeline across 11 image, text, and tabular settings, including controlled comparisons and ablations. On CIFAR-100 reduced NLL by 8.6% while improving accuracy by 2% versus cross-entropy, and outperforming other established methods.
- Designed and ran matched human–LLM decision-making experiments, LLM agents run through APIs and participant data collected through Prolific. Identified distinct learning patterns masked by similar aggregate performance.

Coursework: Sparse Optimization, ML for Sequential Decision Making, Time Series & Forecasting, Causal Inference, Advanced Econometrics.

Publication: "*Neighborhood Smoothing for Calibration.*" Horowitz, I. & Gal, A. (Under review, ICLR 2027). arXiv:2610.09020.

Publication: *"LLM Agents Display Human Biases but Exhibit Distinct Learning Patterns."* Horowitz, I. & Plonsky, O. arXiv:2503.10248.

B.Sc. Data Science & Engineering | Technion | GPA: 87.3 2020 – 2024

Academic Honours: President's List & Dean's List (x2)

Coursework: Machine Learning 1 & 2, Deep Learning, Language, Computation & Cognition (NLP), Foundations & Applications of AI, Database Management, Data Structures & Algorithms, Data Analysis & Visualization, Distributed Data Management, Software Engineering.

HACKATHONS

- 1st place, Technion Biomedical Engineering: pitched an AI pulse-oximetry concept addressing skin-tone bias.
- 2nd place, Technion Data for Good: built and presented a computer-vision (YOLO-based) and LLM therapist-support demo.

PROJECTS

CLARA: Multi-Agent Loan Risk Assessment System (GitHub)

- Built a 7-agent LangChain pipeline (behavioral profiling, interest-rate prediction, risk categorization, RAG-grounded decisioning, RAG validation, and report generation) automating loan-application risk assessment, wrapped in a Streamlit app.
- Integrated interest-rate and risk-prediction models with Qdrant retrieval over historical loans forrecommendations; implemented structured-output checks and evaluator-driven revision loops across the LangChain workflow.

VR Predictive-Text Typing Interface (GitHub)

- Built a webcam-based typing prototype for VR interaction using MediaPipe gaze/gesture tracking, Gemini predictive text, and Whisper dictation; refined prompts and context controls to improve suggestion relevance.

EXTRACURRICULAR

- Paamonim (2021 – Present): Volunteer financial advisor, research analyst, team lead mentoring 8 volunteers assisting over 20 families.
- Combat commander & platoon sergeant, 101st Paratrooper Battalion, outstanding soldier/commander awards; active reservist.