Semih Yagli (U.S. Permanent Resident)

Email: semihyagli@gmail.com | Phone: +1 (609) 608-6323

GitHub: github.com/semihyagli ResearchGate: www.researchgate.net/profile/Semih-Yagli

Data Scientist with expertise in data analysis, machine learning, and full stack application development, passionate about delivering creative, data-driven solutions to complex problems and generating actionable insights.

Educational Background

Princeton University, Princeton, New Jersey Sep. 2016 – May 2021
Electrical and Computer Engineering

Ph.D. in Electrical and Computer Engineering Dissertation Advisor: Prof. H. Vincent Poor
M.A. in Electrical Engineering GPA: 3.98/4.00

Orta Doğu Technical University, Ankara, Turkey (Double-Major) Sep. 2009 – May 2014
Mathematics, **Bachelor of Science**, GPA: 3.96/4.00
Electrical and Electronics Engineering, **Bachelor of Science**, GPA: 3.96/4.00

Patents

Automatic Configuration Management in Cellular Networks (under review) 2022
Machine Learning Models and Methods of Use for Reranking (under review) 2024

Coding and Technical

Programming Languages:	Python (XGBoost, Keras, PyTorch, sklearn), Java, JavaScript, React, React Native, bash.
Data & Analytics:	SQL, NoSQL, BigQuery, R, Python (Pandas, PySpark, Matplotlib), Tableau.
Machine Learning & AI:	Neural Networks, Large Language Models, Natural Language Processing (NLP), Classification, Bayesian Inference, Clustering, Regression Analysis.
Development & Platforms:	Google Cloud, BigQuery, Firestore, Azure, Amazon Web Services, Axios HTTP.
Quantitative Skills:	Probability Theory, Optimization, Data Visualization.

Work Experience

Walmart Global Tech, Senior Data Scientist Oct. 2021 – Present

Lead end-to-end machine learning (ML) model development for e-commerce Search and BuyBox ranking systems, integrating engineering and data science capabilities across Python and Java. Conduct advanced experimentation, signal design, model training, and runtime deployment to drive large-scale business outcomes. Inform strategy through rigorous A/B testing, support infrastructure improvements, and mentor junior data scientists on navigating Walmart's complex systems.

Data Analytics & Machine Learning Engineer at Search Ranking:

- Delivered a 24% lift in Gross Merchandise Value (GMV) by unifying the default and personalized search rankers into a scalable, unified, and personalized ranking system.
- Improved query-item relevance by 11% and boosted ranking precision by developing a token-importance signal and implementing it via continuous integration and continuous deployment (CI/CD) Java code changes and new ML models.
- Led model improvements using interleaving and A/B tests to ensure statistical rigor and stakeholder alignment.
- Reduced latency and improved reliability by deploying new ranking features in Walmart's backend infrastructure.

Technologies: Java (Runtime & Unit Testing), Python (XGBoost), Google Cloud, Azure, SQL.

Senior Data Scientist at Buy Box Offer Ranking:

- Increased add-to-cart conversions by 4% by designing a custom Constrained Neural Network (CNN) that enforced price and shipping monotonicity, a critical feature in e-commerce ranking logic.
- Enhanced multi-offer conversion rate by 3.8% by modeling the relationship between shipping speed, pricing, and customer behavior, delivering optimized item-page rankings.
- Improved business performance by engineering tradeoff-aware ranking features and validating model impact through robust experimentation and backtesting.

Technologies: Python (PyTorch, Keras, Pandas), SQL, BigQuery, Azure, Tableau.

Personal Projects and Experiences

Etiqueta – AI Data Label, Full Stack Web and Application Developer   Mar. 2023 – Present

Spearheaded the development and release of Etiqueta, the only mobile-native platform for AI-aided data labeling, crowd-sourcing and image segmentation. Designed and deployed the full technology stack—from frontend UX/UI to backend services—demonstrating creative problem solving, technical agility, and full-cycle product ownership.

Built for iOS, Android as well as ARM-based macOS and Windows devices, Etiqueta enables scalable data annotation via mobile gameplay.

Platform Impact & Innovation
- Labeled 100K previously unlabeled images and segmented 8K images using Etiqueta, supporting AI/ML model developments through gamified crowdsourcing.
- Launched Etiqueta on Apple App Store and Google Play Store—the only public, mobile-native data labeling app.
- Published labeled public datasets on GitHub as well as a tech report, laying the groundwork for open dataset distribution and research adoption.

Frontend Development:
- Optimized UI/UX with React Native to support 30 label interactions per screen, increasing user throughput and reducing screen-switch latency.
- Built interactive AI-aided gameplay and leaderboards in JavaScript and React to boost user engagement.
- Designed mobile-friendly layouts for iOS, Android, and ARM-based macOS to ensure responsiveness and seamless UX.

Technologies: Javascript, React, React-Native, Axios HTTP, OAuth.

Backend Development:
- Built scalable Python (FastAPI) and Firestore backends for real-time task distribution and data collection.
- Secured user data with OAuth authentication, preventing unauthorized access and ensuring end-to-end privacy.
- Implemented backend unit testing to ensure code quality and system reliability across development cycles.

Technologies: Firestore, NoSQL, Python (FastAPI, pytest), OAuth.

Extra Curricular

Languages: Fluent in English, French, Spanish, and Turkish (Native).
Mentorship: Volunteer mentor at Princeton GradFutures program; currently mentoring 4 students.
Leadership: Regional Princeton Graduate Alumni Leader; organizing networking events for alumni.

Awards and Honors

Excellence in Teaching Award at Princeton University	2018
Turkish Presidential Fellowship Award	2009 – 2014
Ranked 78th out of 1.5M+ candidates in the Turkish National University Entrance Exam	2008

Ph.D. Research Highlights

Bounded Gaussian Mean Estimation, 50-year old open problem
- Used non-linear optimization to prove that least favorable distribution has finite support.
- My technique based on oscillation theorem has been extensively used in multi access and wiretap channel setups.

Information Secrecy and User Privacy
- Developed a clever trick to analyze weakly dependent random variables as if they were independent.
- My technique, which I named 'poissonization', has been employed in large deviations research.

Privacy in Machine Learning
- Using supervised and adversarial machine learning tools to remove sensitive information from data.
- Minimal computational burden on the user makes the solution feasible for all sorts of ML prediction tasks.

Professional Services

Technical Program Committees:

IEEE 54th Annual Conference on Information Sciences and Systems (CISS 2020), Princeton, NJ, USA, March 18-20, 2020.

IEEE 52th Annual Conference on Information Sciences and Systems (CISS 2018), Princeton, NJ, USA, March 22-23, 2018.

Publications

Ongoing Work:

S. Yagli, "Etiqueta: AI-Aided, Gamified Data Labeling to Label and Segment Data," *Manuscript in preparation.*

A. Dytso, S. Yagli, and L. Barletta (Eds.), "Collected Works of Hans Sylvain Withenhausen," *Manuscript prepared.*

Journal Papers:

M. Al, S. Yagli, and S. Y. Kung, "Privacy Enhancing Machine Learning via Removal of Unwanted Dependencies," *IEEE Transactions on Neural Networks and Learning Systems*, Sep. 2021.

A. Dytso, S. Yagli, H. V. Poor, and S. Shamai (Shitz), "Capacity Achieving Distribution for the Amplitude Constrained Additive Gaussian Channel: An Upper Bound on the Number of Mass Points," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2006-2022, Apr. 2020.

S. Yagli and P. Cuff, "Exact Exponent for Soft Covering," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6234-6262, Oct. 2019.

S. Yagli, Y. Altuǧ, and S. Verdú, "Minimax Rényi Redundancy," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3715-3733, May 2018.

Conference Proceedings:

S. Yagli, A. Dytso, and H. V. Poor, "Information-Theoretic Bounds on the Generalization Error and Privacy Leakage in Federated Learning," in *Proc. 21st IEEE International Workshop On Signal Processing Advances in Wireless Communications (SPAWC 2020)*, Atlanta, GA, USA, May 2020.

S. Yagli, A. Dytso, and H. V. Poor, "Estimation of Bounded Normal Mean: An Alternative Proof for the Discreteness of the Least Favorable Prior," in *Proc. 2019 IEEE Information Theory Workshop (ITW)*, Visby, Sweden, Aug. 2019.

S. Yagli, A. Dytso, H. V. Poor, and S. Shamai (Shitz), "An Upper Bound on the Number of Mass Points in the Capacity Achieving Distribution for the Amplitude Constrained Additive Gaussian Channel," in *Proc. 2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1907-1911.

S. Yagli, A. Dytso, H. V. Poor, and S. Shamai (Shitz), "Some Aspects of Totally Positive Kernels Useful in Information Theory," in *Proc. 2019 MoTION Workshop of Wireless Communications and Networking Conference (WCNC)*, Marrakech, Morocco, Apr. 2019, pp. 1-6.

S. Yagli and P. Cuff, "Exact Soft-Covering Exponent," in *Proc. 2018 IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, Jun. 2018, pp. 1680-1684.

S. Yagli, Y. Altuǧ, and S. Verdú, "Minimax Rényi Redundancy," in *Proc. 2017 IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2980-2984.

Dissertation:

S. Yagli, "Topics in Information and Estimation Theory: Parameter Estimation, Lossless Compression, Constrained Channels, and Error Exponents," PhD Dissertation, Princeton University, 2021.

Technical Reports:

S. Temel, S. Yagli, and S. Gören, "P, PD, PI, PID Controllers," Middle East Technical University, Ankara, Turkey, Technical Report, 2013.