Georgios BALIKAS

Chief Scientist - PhD in ML, NLP

G / B

✉ geompalik@hotmail.com

☎ +33 7 62 66 19 06

📍 Grenoble, France

in Georgios Balikas

GitHub balikasg [Personal Site]

KEY ACHIEVEMENTS AND EXPERTISE

- Delivered significant commercial impact through data science and analytics: 12% gross margin improvement with pricing algorithms for eCommerce, 8% sales growth in seasonal ads campaigns, 10% engagement increase with AI systems developing pricing solutions for 41 countries.

- Collaborated effectively across executive, product, and engineering functions (Manager Survey Score: 4.62/5.0). Led teams of 12 data scientists with +10% improvements in productivity and career development (2 promotions).

- Published research at top-tier venues with 1,176 citations (CIKM, SIGIR, ACL,..), with focus on applications of deep learning and LLMs. Expert in LLMs, Retrieval Augmented Generation, dense/sparse retrieval, Named Entity Recognition, Ranking, Classification, Regression

PROFESSIONAL EXPERIENCE

Chief Scientist at Kelkoo Group (May 2024 - today)   Grenoble, France

Leading ML/NLP innovation and pricing strategy for a global e-commerce platform generating €1.2M monthly gross margin (GM) across 41 countries. Directing a team of 12 data scientists and software engineers to deliver cutting-edge search and pricing solutions. Key Achievements:

- Architected new traffic optimization system using Mixed Integer Linear Programming, delivering **12% group-wide GM improvement** through improved pricing strategies. The system automatically decides the daily budget allocation across online advertisement campaigns by modeling the problem as resource allocation and marking mix modeling.

- Identified and led development of seasonal pricing intelligence system accounting: Achieved **6% YoY GM** improvement, 8% sales growth during the 2024 Black Friday period to target eCommerce promotions.

- Identified and led NLP initiatives for e-commerce content enhancement (Title optimization for Google product listing ads, GTIN classification)

- Drive cross-functional collaboration with executive, product, and engineering teams, ensuring smooth project execution and high team satisfaction and retention. Annual Manager Survey Score: 4.62/5.0

- Mentored and guided 3+ junior data scientists in ML/NLP: +10% team velocity and 2 promotions to Senior levels

LEAD DATA SCIENTIST at Salesforce (Mar 2019 - May 2024)   Grenoble, France

Design production systems (ranking, generative AI, NLP), anticipate of future needs and extensions, evaluate algorithms and analyze of production metrics. On Generative AI:

- Introduced a Generative AI version of Salesforce's Question Answering product based on Retrieval Augmented Generation (now in production). **Increased positive engagement by 10%**

- Prototyped an agent for Text-to-SOQL and CRM action automation on a subset of CRM objects: after demoing, a dedicated team was formed, to deliver it

On Search Question Answering that lets users find information in the form of a snippet:

- Explored dense and sparse passage retrieval strategies to improve results relevance. Optimized the pipeline to achieve **60% positive user engagement**

- Introduced fine-tuning processes of embedding models from Transformer neural networks (30% gains on precision)

Developed and owned deep neural networks for the Search system aiming to improve search user experience by formulating hypotheses, evaluating them offline and validating the results with online by A/B tests.

- Delivered a model that optimizes of the query autocompletion search algorithm. The model improved the system's latency by 15% and **reduced the cost to serve by 18%** [ECIR'22, Slides]

- Delivered a model that orders the search results for user queries. Impact: millions of queries daily-thousands of organisations. **Click-Through-Rate: +2%**, Mean Reciprocal Rank: +2.5%.
- Delivered a model for Named Entity Recognition (NER) as part of a semantic parsing system for user queries to recognize between a dozen of custom concepts. To overcome the training data absence, I designed a probabilistic grammar to massively generate synthetic data. [EPFL]
- Search Award Q1'21: Innovation and Service Ownership for designing and delivering Grafana monitoring and alerting
- **Innovation Prizes** on Fall'20, Winter'21 and Fall'22 hackathons for contributions around fine-tuning transformer models for multilingual search. The work on the Fall'22 hackathon inspired my CIKM'23 industry track submission
- Mentored 3 Software Engineers and 5 Data Scientists on best ML practices: all of them improved the models, 3 promotions to Senior levels

Previous Experience   Grenoble/Paris

Data Scientist as Kelkoo Group (Nov'17-Feb'19): Developed bidding algorithms for second price auctions used in Google's Product Listing Ads for growth margin maximization. Automated the manual bidding, saving >1,000 men days/year

Software Engineer at Pierre and Marie Curie University, Paris 6 (Oct'13-Aug'14): Designed and developed the BioASQ participants area using Python/Django. Collaborated with a consortium of 6 partners

Technical Leadership & Community Engagement

- **EU Expert Evaluator**: Assess machine learning project submissions to be funded by the EU and evaluate the progress of the funded projects. Recent Horizon project: HORIZON-CL4-2024-HUMAN-03-01 for Multi-Modal LLMs
- University of Grenoble Alps **Industry expert**: Validate the standards of Multidisciplinary Institute in AI (MIAI) certification by assessing the data science skills of individuals in career-transition programs by evaluating of their internships
- Grenoble Data Science Meetup co-organizer: Lead a thriving community of 900+ data scientists, fostering collaboration and knowledge sharing through monthly events
- Applied ML Days@EPFL 2024 Track Chair on "Search & AI" track: Focused on industry applications of LLMs and Search
- **Demo chair** at European Conference of Machine Learning (ECML 2022)
- Academic conference reviewer: Review submissions for leading ML conferences: ECML, ECIR, SIGIR, CIKM, and more

CONFERENCE PUBLICATIONS (NON-EXHAUSTIVE)

- Comparative Analysis of Open Source and Commercial Embedding Models for Question Answering, CIKM' 23, Industry Track
- Accelerating Autocomplete Suggestions at Salesforce Search, ECIR'22, Industry Track
- John-Arthur at emEval-2023 Task 4: Fine-Tuning Large Language Models for Arguments Classification, SemEval@ACL 2023, System ranked 2/40
- Query Understanding for Natural Language Enterprise Search, DeepNLP@SIGIR'20

Publications: google.scholar/dblp. Patents: google patents. **References available** upon request

EDUCATION

- PhD Thesis: "Mining and Learning from Multilingual Text Collections using Topic Models and Word Embeddings". University of Grenoble Alps (Sep 2017)
- MSc, Artificial Intelligence and Decision Making, Université Pierre et Marie Curie, Paris (Sep 2014)
- MSc, Electrical and Comp. Engineering Dept. Aristotle University, Thessaloniki (Jun 2011)

LANGUAGES

Greek	Native
English	Excellent Knowledge
French	Professional Knowledge