Georges Dupret

26 Playstead Rd. #2
Newton, MA-02458
+1 (617) 678-1443
georges.dupret@gmail.com

EXECUTIVE SUMMARY

Accomplished research and data scientist with a strong focus on organizational and managerial capabilities. Skilled in leading and coordinating complex projects, with extensive experience in machine learning (statistical modeling, decision trees, LLM, DNN, etc.) and statistics (AB experiments, causality, testing, etc.). Demonstrated proficiency in managing project lifecycles, fostering collaboration, and driving strategic initiatives, supported by robust analytical skills in mathematics and programming proficiency.

PROFESSIONAL EXPERIENCE

Spotify, Boston, Massachusetts, USA November 2021 – Present
Principal Scientist

- I spearheaded a ground-up initiative to develop innovative model-free methods for strategic decision-making using historical data. My efforts involved evangelizing the concept and actively seeking out collaborators to drive adoption and implementation, resulting in its increasing use at the highest levels of the company [1].
- Promote and coordinate efforts to implement advanced Large Language Model methods to assess the quality of music playlists (LLM as a Judge) before their release, ensuring high production standards. Develop a Minimum Viable Product to showcase the technology, and coordinate various production teams to scale the application for use in Tableau reporting and as a support tool for editors and annotators.
- Established a research team to investigate the application of Heterogeneous Treatment Effect (HTE) studies in A/B experiments. The analyses generated by HTE are typically unstable and challenging to interpret. We overcame these obstacles, significantly enhancing the tool's usability and accessibility for non-specialists.

Apple, Cupertino, California, USA September 2017 – November 2021
Principal Scientist in Maps Organization working on Ranking, Machine Learning and Data Analysis.

I was part of the team responsible for the search engine powering address and business searches in Apple Maps. While collaborating on ongoing maintenance and enhancements with the team, I also lead several individual initiatives:

- Debugging issues within a large and complex system posed significant challenges. A notable difficulty with the learning algorithm was determining the contribution of different features in ranking items. We leveraged Shapley values as a principled solution, which we extended to effectively compare result pairs.
- With the rise of deep learning at that time, we investigated its application to improve search rankings. We developed the initial MVP for a search system that efficiently utilized embeddings, making it suitable for production.
- Considering the critical role of physical distance in map search, along with user implicit feedback, I developed innovative features and data structures tailored to address map topology and the high dimensionality inherent in user feedback.

Fullpower, Santa Cruz, California, USA August 2016 – September 2017
Data Science and Machine Learning Director

Fullpower is the technology leader for IoT Cloud-Connected Digital Sports, Sleep Monitoring, Smart Home and Connected Objects, powered by AI, Machine Learning and Data Science.

- Led transition from rule-based systems to machine learning in sleep monitoring solutions.
- Organized data collection and label generation in collaboration with sleep specialists, while overseeing the machine learning pipeline development.

AliveCor, San Francisco, California, USA December 2014 – May 2016
Principal Scientist

- Developed advanced algorithms for automated analysis and diagnosis of heart conditions using AliveCor's FDA-approved electrocardiogram device.
- Reorganized the entire codebase in C++ to prioritize testing and reproducibility, ensuring compliance with FDA standards.
- Created tools using R Shiny for efficient triaging of misdiagnosed ECG cases.
- Engineered specialized algorithms for signal filtering, beat classification, and pathology detection.
- Defined priorities and set the agenda for the Science Team in collaboration with Project Managers to align with organizational goals.

Yahoo! Labs, Sunnyvale, California, USA October 2008 – December 2014
Senior Scientist, Machine Learned Ranking Group

- Developed and optimized Web and Vertical Search ranking algorithms utilizing click-through data analysis, leading to new interpretations that enhanced search engine results globally. This work earned a Best Paper Award at WSDM2010.
- Innovated user engagement metrics across Yahoo properties, balancing revenue generation with improved user experience through localized and global metrics.
- Designed computationally efficient methods for query recommendations, improving the relevance and efficiency of search results.
- Advanced the ranking algorithm's sensitivity to document recency, boosting the freshness and relevance of search outputs.

Yahoo! Research Latin America, Santiago, Chile March 2006 – October 2008
Researcher

- Developed novel evaluation metrics for structured documents using principled user model assumptions.
- Managed and supervised multiple dissertations and master's theses, contributing to academic and professional advancements in the field of web research.

Center for Web Research, Universidad de Chile, Santiago, Chile March 2004 – March 2006
Research in Information Retrieval and Text Mining:

- Pioneered statistical models for the explicit modeling of user click behavior within search engine log data, laying foundational work in understanding search engine interactions.
- Conducted advanced Principal Component Analysis to optimize document representation dimensionality within concept space, enhancing text mining capabilities.

- Taught graduate courses including Experimental Design and Analysis, Statistics for Engineers, Data Mining of web click-through data, and Information Retrieval, contributing to the academic growth of engineering students.

IBM, Zürich Research Laboratory, Switzerland January 2001 – March 2003
Research Staff Member.
Research and application in Information Retrieval and Text Mining:

- Applied machine learning techniques to high-dimensional spaces, focusing on the extraction and automatic clustering of issues within Quality Assurance reports for various IBM external projects.
- Developed an innovative method to determine the optimal number of singular values in Latent Semantic Analysis, applied to corpus visualization and query formulation.

IBM, Tokyo Research Institute, Tokyo, Japan 1998 – 2000
Internship.

- Identified optimal data storage methods and implemented them using C++, enhancing data retrieval efficiency for large databases.
- Automated thesaurus construction and keyword clustering processes, and devised novel Singular Value Decomposition approximations using Artificial Neural Networks.

EDUCATION

Ph.D., Policy and Planning Sciences, Tsukuba University
Thesis - 'Constrained Architecture Neural Network and its Application to Data Analysis.'

Master Degree, Policy and Planning Sciences, Tsukuba University
Dissertation - 'Density of Population Analysis Using Artificial Neural Networks.'

Engineer in Applied Mathematics, - Economics Oriented, Catholic University of Louvain, Belgium Dissertation - 'Central Place Theory and Multipurpose Trips.' at Technical University of Lisbon.

LANGUAGES

Fluent in French, English, Spanish, and Portuguese; proficient in daily German and Japanese conversation.

SELECTED PUBLICATIONS

For a comprehensive list, visit my Google Scholar page (more than two thousand citations).

Publication [2] is a formal statistical model of user behavior on a SERP applied to metrics. Publication [3] won best paper award[1] and [4] an Honorable Mention at the SIGIR Test of Time Awards[2]. Absence Time [5] shows how survival analysis can solve complex problem in estimating user satisfaction. Finally, Publication [1] is a novel data driven method to take decision applicable when AB testing is not feasible.

[1] `http://www.wsdm-conference.org/2010/`
[2] `https://sigir.org/awards/test-of-time-awards/`

[1] G. Dupret, K. Sozinov, C. B. Gonzalez, Z. Zacks, A. Yuan, B. Carterette, M. Mai, S. Bansal, G. L. L. Lien, A. Gatash, *et al.*, "Fortune: Running offline scenarios to estimate impact on business metrics," *arXiv preprint arXiv:2403.00133*, 2025. accepted to KDD (ADS) 2025.

[2] G. Dupret and B. Piwowarski, "A user behavior model for average precision and its generalization to graded judgments," in *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pp. 531–538, 2010.

[3] G. Dupret and C. Liao, "A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine," in *Proceedings of the third ACM international conference on Web search and data mining*, pp. 181–190, 2010.

[4] G. E. Dupret and B. Piwowarski, "A user browsing model to predict search engine click data from past observations.," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 331–338, 2008.

[5] G. Dupret and M. Lalmas, "Absence time and user engagement: evaluating ranking functions," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 173–182, 2013.