

Truc-Viet “Joe” Le, Ph.D.

vietexob@gmail.com • +65 9420 8417

Senior Data Scientist / AI Engineer

[LinkedIn](#) • [Singapore](#) • [GitHub](#)

Innovative AI expert driving operational efficiency and solving complex business problems.

An accomplished data scientist with proven track record of developing scalable AI products across finance, e-commerce, and technology sectors. Adept at leveraging machine learning (ML), natural language processing (NLP), and generative AI (GenAI) solutions to drive digital transformation and optimise business processes. Skilled in fine-tuning large language models (LLMs) and leveraging graph analytics to detect anomalies in financial data. Ph.D. in Information Systems with patents and publications in AI & data science.

Areas of Expertise: ML & AI | Data Science & Analytics | NLP & LLMs | Enterprise AI Integration | Data Visualisation | Cloud & Big Data

Technical Proficiencies

- **Programming languages:** R, Python, Java, Scala, JavaScript, Bash, SQL
- **Development tools & technologies:** Git, Docker, FastAPI, Streamlit, Flask, CI/CD Pipelines
- **Frameworks & libraries:** Pandas, scikit-learn, NumPy, NLTK, spaCy, Gensim, Hugging Face, TensorFlow, PyTorch, Langchain

Professional Experience

EFG International, Singapore

Jan 2024 — Present

Senior Data Scientist (VP)

Working in a cross-functional team across Europe and Asia to enhance operational efficiency and mitigate banking compliance risks by leveraging GenAI and ML algorithms. Developed AI-driven solutions for internal knowledge search (RAG system), unstructured document processing (OCR), and client risk assessment while driving enterprise AI adoption through training and mentorship.

- Utilised few-shot learning to fine-tune open-source LLMs for effective language translation in financial terms.
- Bolstered productivity by saving ~20 minutes per user per week with in-house RAG system for internal knowledge search.
- Advanced AI literacy and adoption by conducting comprehensive training (e.g., prompt engineering) for non-technical users.
- Refined client risk assessment using investment and trading features, resulting in more efficient periodic review processes.

Credit Suisse AG (Part of UBS Group), Singapore

Dec 2021 — Dec 2023

Senior Data Scientist (AVP)

Led the development of AI solutions for AML (anti-money laundering) and financial crime detection in Wealth Management (WM) Platform global team. Automated entity resolution and semantic extraction from SWIFT messages, significantly cutting investigation time. Used graph analytics to visualize and uncover hidden relationships and money-laundering typologies.

- Developed efficient name matching ML models and pipelines for transaction monitoring, deployed on Spark clusters.
- Achieved over 95% accuracy in first/third-party transaction classification, leading to > 70% reduction in false AML alerts.
- Used NLP and named entity recognition (NER) to extract semantics (e.g., purposes of payment) from SWIFT messages.
- Leveraged knowledge graphs (Neo4j) for entity resolution and revealed hidden money-laundering relationships using Power BI network visualization, reducing investigation time by 2-3 hours per case.

Agoda.com (Part of Booking Holdings Group), Singapore

Sept 2019 — Nov 2021

Data Scientist

Agoda.com is a leading online travel agency (OTA) in the APAC region. In this role, I was part of the global Data Science team and led efforts to tackle the property matching problem in OTA by developing efficient matching algorithms and ML models for mapping duplicate listings across different suppliers, improving search, recommendation, and conversation rate.

- Reduced duplicate listings by approx. 80% by deploying high-precision and scalable mapping pipelines on Spark clusters.
- Utilized deep learning and embeddings to merge property galleries of duplicate listings, significantly improved content and increased user engagement and conversion rates by approx. 20%.
- Used Bayesian networks to model the causal relationships between web performance and bookings on Agoda websites and recommended optimization strategies to performance engineering team to further increase conversion rate.
- All solutions were validated through rigorous experimental design and A/B testing.

Played a pivotal role in the development of the SAP Cash Application module at SAP Labs Singapore. Automated accounts receivable by matching bank statements with invoices. Developed deep learning solutions to solve the multi-match problem in cash application.

- Secured three U.S. patents for innovative algorithmic solutions to the multi-match problem using graphs and embeddings.
- Improved matching accuracy of bank statements to open invoices to > 98% by using NER on unstructured data fields.
- Enhanced automation transparency and trustworthiness by implementing LIME framework for result explanation.
- Deployed the products on AWS and GCP and optimized for memory footprint, while providing end-to-end customer support.

Education

Ph.D. in Information Systems (GPA: 3.7/4), Singapore Management University (SMU), Singapore **Jan 2013 – Jul 2017**

- Dissertation: *An Integrated Framework for Modelling & Predicting Spatiotemporal Phenomena in Urban Environments*
- Completed a 3-month research internship at IBM Research (Singapore), leading to a conference paper at IEEE ICC 2017.

Visiting (Exchange) Ph.D. Scholar (GPA: 3.84/4), Carnegie Mellon University (CMU), Pittsburgh, PA **Aug 2014 – Jul 2015**

- Completed advanced coursework in Machine Learning, Robotics, and Optimisation at CMU's School of Computer Science.
- Served as a teaching assistant for Mobile Intelligence & Business at Heinz College of Information Systems and Public Policy.

M.Sc. in Mathematical Sciences (GPA: 3.3/4), Nanyang Technological University (NTU), Singapore **Jan 2011 – Jul 2012**

- Thesis: *Pareto Stable Matchings: An Empirical Study*
- Relevant coursework: Game Theory, Number Theory, Cryptography

B.Eng. in Computer Engineering (GPA: 3.56/4), Nanyang Technological University (NTU), Singapore **Aug 2005 – Jul 2009**

- Final Year Project: *A Mathematical Model of Hospital Length of Stay*
- Completed internship at Hochschule für Technik Rapperswil (Switzerland), developed an open-source plugin for Eclipse IDE.

Publications & Patents

Publications:

- **Truc Viet Le**, Baoyang Song & Laura Wynter. **Real-time Prediction of Length of Stay Using Passive Wi-Fi Sensing**. IEEE ICC 2017, Internet of Things (IoT) Track, Paris, France. [\[Link\]](#)
- **Truc Viet Le**, Richard Oentaryo, Siyuan Liu & Hoong Chuin Lau. **Local Gaussian Processes for Efficient Fine-Grained Traffic Flow Prediction**. IEEE Transactions on Big Data (TBD), Special Issue on Urban Computing, 3(2), 194–207. [\[Link\]](#)
- **Truc Viet Le**, Siyuan Liu & Hoong Chuin Lau. **A Reinforcement Learning Framework for Trajectory Prediction Under Uncertainty**. ECAI 2016, The Hague, The Netherlands. [\[Link\]](#)
- **Truc Viet Le**, Siyuan Liu, Hoong Chuin Lau & Ramayya Krishnan. **Predicting Bundles of Spatial Locations from Learning Revealed Preference Data**. AAMAS 2015, Istanbul, Turkey. [\[Link\]](#)

Patents:

- **Truc Viet Le**, Sean Saito, Rajalingappaa Shanmugamani & Chaitanya Joshi. **A Graphical Approach to the Multi-match Problem**. Issued Jun 11, 2020. Patent No.: US20200184281A1. [\[Link\]](#)
- Sean Saito, **Truc Viet Le**, Rajalingappaa Shanmugamani & Chaitanya Joshi. **Representing Sets of Entities for Matching Problems**. Issued Jun 4, 2020. Patent No.: US20200175559A1. [\[Link\]](#)
- Rajalingappaa Shanmugamani, Chaitanya Joshi, Rajesh Arumugam, Sean Saito & **Truc Viet Le**. **Utilising Embeddings for Efficient Matching of Entities**. Issued Jun 18, 2020. Patent No.: US20200193511A1. [\[Link\]](#)

Media Appearances

- Gave a talk at **FOSSASIA 2018** – [Machine Learning for Smart Cities](#)
- Gave a talk at **Python User Group Singapore** – [Reinforcement Learning Using Python](#)