



Nguyễn Bảo Lê

Date of birth: Dec 14, 1996

Phone: +84-918496934

Email: baoln1412@gmail.com

Address: District Phu Nhuan, Ho Chi Minh City, Vietnam

Portfolio: portfolio.desoul.dev

LinkedIn: linkedin.com/in/baoln1412

PROFESSIONAL SUMMARY

Data analyst with seven years across Business Intelligence, Data Analytics and applied AI, and a finance background. I build the full path from raw data to decision: ETL pipelines and warehouse models, the dashboards business teams run on, and the machine-learning and generative-AI systems that turn unstructured market and player data into answers. I also design and ship production LLM systems end to end — retrieval, grounding gates, governed agent execution and automated evaluation — and I decide with measurements rather than vendor claims: I raised retrieval accuracy from 79% → 93% on a corpus of 1,000+ documents, and I have killed features I had already built when the measurement said they would cost recall. Domain depth in mobile game publishing for the SEA market: user acquisition, monetization, retention and competitive benchmarking. **Available now for contract work, full time.**

TECHNICAL SKILLS

AI & Machine Learning	LLM pipeline design and orchestration (Claude Agent SDK, Gemini, ChatGPT); retrieval and RAG; grounding and evaluation gates, LLM-as-judge; prompt and context engineering; generative media (image, text-to-speech, video); regression models for churn and LTV prediction; RFM customer segmentation; A/B testing
Data Engineering	Airflow, Spark, Hadoop, Trino, ETL pipeline design, API integration, web crawling, relational data modelling
Databases & Storage	PostgreSQL, Hadoop, Trino, Firestore, BigQuery-style warehouses, Google Sheets / SharePoint distribution
BI & Visualization	Tableau, Looker, Power BI, Google App Sheet, custom dashboards, cohort and funnel analysis, Sankey and waterfall user-journey visualization
Languages & Tools	SQL, Python, Google Apps Script, GitLab, Google Cloud APIs
Product & Market Data	AppsFlyer, Facebook Marketing & Page API, YouTube API, Steam, Sensor Tower, Data.ai
Languages	Vietnamese (native), English (professional working proficiency)

WORK EXPERIENCE

VNG Corporation

March 2019 – May 2026

Data Analyst Specialist of Game Publishing Studio - SEA Market.

- Data Pipelines & Warehouse:

- + A layered warehouse on AWS Athena — raw extracts standardised into per-user and per-campaign tables, an incrementally rebuilt player profile, and a single reporting table the dashboards read from. Terabytes a day across 30+ published titles.
- + Ingestion from attribution (AppsFlyer pull and push APIs: installs, downloads, registrations, postbacks, ad events) and from game servers (purchases, logins, account creation, new users), reconciled into one daily grain.
- + A written data catalogue with canonical metric definitions — presence-based daily actives and retention, lifetime spend for VIP tiers, completed payments only for revenue — so two teams asking the same question got the same number.
- + Scalable ETL on Hadoop, Airflow and Spark; PostgreSQL warehouse models and query performance; raw-input validation and discrepancy resolution designed with the data engineering and DevOps teams.
- + API integrations against Facebook Marketing and Page, YouTube, Steam, Sensor Tower and Data.ai, consolidating ad performance, engagement and market data into one place.
- + Automated distribution of the same standardised data to wherever each team already worked — Google Sheets, SharePoint, PostgreSQL, Hadoop, Trino, Firestore — and version control and documentation of the whole thing in GitLab.

- Reporting & BI:

- + Dashboards in Tableau, Looker and Power BI, with a generated registry built by probing the live dashboard files rather than the checked-in SQL, verifying each date field's timezone against the data. It exists because the two had silently diverged.
- + Automated weekly briefings per title and an executive user-acquisition report joining attribution data to warehouse daily summaries, delivered as a written document rather than a dashboard someone has to interpret.
- + Channel diagnostics for paid media — spend headroom, funnel timing, per-network performance — and first-session funnel analysis from download through tutorial, used to find and fix drop-off caused by download and loading failures.
- + User journeys from acquisition through conversion and churn, visualised as Sankey and waterfall charts.
- + Internal tools where a dashboard was the wrong shape — a project-management and workload-distribution portal on Google Apps Script, AppSheet and Looker, and proof-of-concept web apps on Google Cloud APIs.

Achievement:

- + **months → 1–2 weeks market review cycle.** Competitor benchmarks for an acquisition review, assembled by hand until a watcher over the product teams' own request file pulled each set from Sensor Tower and returned a workbook.

- AI & Machine Learning Application:

- + Crawling and analysis pipelines over social channel data, producing sentiment and competitor analysis for published games.
- + Generative AI categorising user feedback and market research at a volume manual review could not reach.
- + Regression models for churn and LTV; RFM segmentation aimed at targeted marketing.

Achievement:

- + **–80% manual workload, service and product teams.** AI-driven market research analysis replacing manual review.
- + **+30% monthly game revenue.** From product decisions informed by predictive modelling.

- Market Research & Publishing Decisions:

- + Benchmarks combining internal performance data with Sensor Tower and Data.ai, segmented by genre, sub-genre, IP, theme and developer to qualify acquisition candidates.
- + Business cases and P&L projections for game acquisitions, and pre-launch readiness across payments, servers, localisation, tracking and reporting.
- + A/B tests on campaign performance and in-game event planning.

Achievement:

- + **−80% average cost per install.** Marketing optimisation at VNG, measured against spend. Proven inside one month on a \$10,000 launch budget split 20/80: two weeks testing creative, channel and audience on a fifth of it, two weeks confirming the winner held at volume on the rest.
- + **\$5 → \$0.80–1.00 per install at launch, before and after.** The same figure as a before and after rather than a percentage, because a percentage with no starting point is a claim and a pair of prices is a result.

Vietbank

September 2018 - February 2019

Trader of Vietbank's Foreign exchange investing department

- Data Analyzing:

- + Update news on the foreign exchange market and find good investment opportunities using fundamental analytics and technical analytics.
- + Make in-day transactions for G7 currencies.
- + Daily report investment results for stakeholders.

SELECTED RESULTS

- Cost per install cut by four fifths, proven in a month (VNG Corporation, one month):

- + **Situation.** A game launch was acquiring players at about \$5 per install. At that price the launch budget buys an audience too small to judge the title on, and every downstream number — retention, revenue per user, payback — is read off a cohort that is not worth reading.
- + **Action.** Ran it as an experiment with a budget split rather than a campaign with an opinion. Two weeks on 20% of a \$10,000 budget to test creative, channel and audience against cost per install; two weeks on the remaining 80% to confirm the winner held at volume rather than at sample size.
- + **Result.** Cost per install settled between \$0.80 and \$1.00 for the launch phase, roughly a fifth of where it started — the same money buying about five times the players, on a figure that had already survived a confirmation round. **−80% average cost per install · \$5 → \$0.80–1.00 per install at launch, before and after**

- A market review that took months, answered in a fortnight (VNG Corporation):

- + **Situation.** Evaluating a game for acquisition meant assembling competitor benchmarks by hand. The review cycle ran for months, which is longer than the window in which a title is available to bid on.
- + **Action.** Built a watcher over the shared request file the product teams already used: it picks up each new request, pulls the benchmark set for that title from Sensor Tower, and writes a finished workbook back. Debounced against sync lag, locked to one machine, with per-row state so a failed row is retried and a done row is never rebuilt.
- + **Result.** The review cycle fell from months to one or two weeks, and the benchmark set stopped varying with whoever assembled it. Product teams requested reviews directly instead of queueing. **months → 1–2 weeks market review cycle**

INDEPENDENT AI PROJECTS

AI Console

2026 – present

Research, orchestration and evaluation platform — Python, FastAPI, Claude Agent SDK

- + Built an end-to-end pipeline that crawls trending technical sources, reads each repository with an LLM, scores it against a portfolio of active projects and publishes a daily triage report — **1,000+ repositories evaluated** over 50+ daily reports.
- + **140+ of those sources applied** — an improvement actually built from the source, not merely a bookmark — across 7 projects: agent orchestration and LLM tooling, developer and build tooling, interface and design systems, data pipelines and crawling, and security scanning.
- + Designed the grounding and evaluation layer: schema-constrained model output with retries, an automated judge that blocks ungrounded claims before they reach a report, and a per-project rule system that gates which actions an agent may take.
- + Raised retrieval accuracy from **79% → 93%** by replacing keyword search with a hybrid keyword-plus-embedding approach, measured against a hand-labelled evaluation set.
- + Diagnosed and fixed a silent data-quality defect in which four of five ingestion sources supplied a fabricated popularity metric that the model then reasoned from, affecting scoring across the corpus — root-caused, regression-tested and back-filled.
- + 30k+ lines of Python under test-driven development across 100+ test modules.

AI Content Generator

2026 – present

Bilingual multimodal content engine — Python, FastAPI, generative media

- + Built a seven-tier applied-AI content library delivered in English and Vietnamese, with an HTML/SVG-to-video renderer that produces narrated explainer videos at zero marginal generation cost.
- + Implemented illustrated comic generation using reference-image conditioning for character consistency across panels, with automated quality gates for layout, lettering and teaching order that block a failing page rather than shipping it.
- + Benchmarked four Vietnamese text-to-speech engines on word error rate and real-time factor, selecting on measured results (**8.8% WER**) rather than vendor claims.

Facebook Content Generator

2026 – present

Automated social content pipeline — Python, DuckDB

- + Medallion-architecture data pipeline (bronze / silver / gold in DuckDB) with scheduled LLM classification and generation, grounded on retrieved article bodies rather than headlines.
- + Ran a controlled evaluation of a proposed keyword pre-filter over 3,600 labelled decisions and rejected it on the measured result, avoiding a change that would have degraded recall.

EDUCATION

University of Economics HCMC

2014 - 2018

Major: Finance

GPA: 3.5/4

CERTIFICATIONS

Cloud, Machine Learning, and Security Academy — Coursera

2024

Advanced Business Analytics Certificate — University of Colorado Boulder

2022

IBM Data Analyst Professional Certificate

2021

Data Analysis in Practice — Data Analysis School

2020