TYSON JOHNSON

tysonjohnsondev@gmail.com | tysonjohnson.dev | linkedin.com/in/tyson-d-johnson | github.com/tysonjohnsondev

Computational & Data Sciences graduate (*summa cum laude*, May 2026) who builds and deploys end-to-end ML systems including calibrated clinical risk models, transformer-based NLP, and computer vision. Native English speaker with 2+ years of professional experience in South Korea and intermediate Korean (former U.S. Marine, DLIFLC-trained linguist). Seeking data roles in Korea, focused on healthcare and pharmaceutical analytics.

Education

George Mason University — Incheon, South Korea
B.S. Computational & Data Sciences | *GPA 4.0/4.0, summa cum laude* — *Sep 2023 – May 2026*

- Outstanding Academic Achievement Award in Computational & Data Sciences

Defense Language Institute Foreign Language Center — Monterey, CA
A.A. Korean Studies | *GPA 3.84/4.0, High Honors* — *Sep 2017 – Feb 2019*

Technical Skills

Programming	Python, R, SQL, JavaScript
ML & Data Science	PyTorch, Hugging Face Transformers, scikit-learn, CatBoost, pandas, NumPy
Data Engineering	ETL pipelines, SQLite, Scrapy, Pydantic, GitHub Actions CI/CD
Web & Deployment	FastAPI, Flask, WordPress, Gunicorn, Apache, Cloudflare
Visualization	Plotly, matplotlib, seaborn, GeoPandas, NetworkX
Tools	Git, Linux/CLI, Jupyter, RStudio, LaTeX
Languages	English (Native); Korean (Business Proficient)

Projects

Type 2 Diabetes (T2D) Risk Screener — GitHub • Spring 2026

- Built an ML screening tool for undiagnosed Type 2 Diabetes risk on Korean NHIS health-examination data (∼1M patient records, 40 engineered features).
- Achieved 0.816 ROC-AUC with CatBoost; tuned the threshold to ≥85% sensitivity, added Platt-scaling calibration and per-prediction SHAP explanations for clinical interpretability.
- Deployed a FastAPI backend (Pydantic v2) with an interactive single-page frontend rendering SHAP waterfall charts.

Pharmacovigilance ADE Classifier — GitHub • Spring 2026

- Fine-tuned and compared BERT, BioBERT, and PubMedBERT on adverse-drug-event sentence classification (ADE Corpus v2) to quantify the value of biomedical-domain pretraining.
- Best model (PubMedBERT) reached 0.94 macro-F1 / 0.96 accuracy; reported pairwise gaps with 95% paired-bootstrap confidence intervals across 10 seeds — a fully reproducible study that regenerates every result from committed predictions.

CountryGuesser — GitHub • Spring 2026

- Fine-tuned EfficientNet-B0 on a 500K-image subset of OpenStreetView-5M for country-level image geolocation across 180 countries; reached 57.6% held-out test top-1 accuracy (random baseline 0.6%).
- On a 1,000-image human-vs-model benchmark, the model scored 64.7% — roughly 6× the human average, a gap confirmed by McNemar's test; deployed the webapp on a VPS with rate limiting and CSP headers.

netviz_tools — PyPI v0.2.0 • 2025

- Published a pip-installable Python package (PyPI) extending NetworkX for network visualization — interactive Plotly/Sankey graphs, GeoPandas trade-flow maps, and FAOSTAT trade data (551+ commodities).

Experience

George Mason University — Incheon, South Korea
Student Intern / Tutor — *Feb 2024 – May 2025*

- Supported instruction and grading across three semesters for up to five Computational & Data Sciences courses under three faculty, with class sizes growing from ∼60 to 90+ students per term.
- Authored supplementary student materials (environment/dependency guides) and re-engineered the course's per-semester GitHub setup, replacing manual repo recreation with a templated, CLI-driven workflow.

United States Marine Corps — South Korea & Camp Pendleton, CA
Korean Translator / Liaison → Administrative Specialist — *Jun 2019 – Jun 2022*

- Delivered professional Korean–English translation and authored official reports under a U.S. national-security clearance.
- As company clerk, automated training-records reporting — replaced a manual Excel workbook with an Excel/VBA system generating per-person currency, platoon-level 30/60/90-day due lists, and status dashboards.