Junho Hong

hjunho@u.northwestern.edu · (314) 665-5753 · junhohong.dev

EDUCATION

Northwestern University Evanston, IL

B.A., Computer Science & Data Science — GPA 4.0 *Expected June 2028*

Relevant Coursework: Large Language Models, Machine Learning, Matrix Algebra, Data Structures & Algorithms

Activities: Northwestern AI Club, WildHacks

EXPERIENCE

Northwestern × Lawrence Berkeley National Lab JGI Evanston, IL

Undergraduate Researcher, GenomeOcean MoE *Jan 2026 – Present*

- Built sparse-upcycling pipeline converting dense 120M-param foundation model into Mixtral-style MoE (8 experts, top-2 routing; 714M total / 205M active); trained ∼52B tokens on 4× H200 GPUs via NeMo / Megatron-Core.
- Showed MoE matches dense baselines on validation loss and downstream benchmarks with 2.5× fewer active parameters.
- Designed expert-specialization analysis via Jensen–Shannon divergence on routing; MoE experts develop emergent class-conditional routing predictive of ground-truth class under held-out cross-validation; abstract submitted to ISMB 2026.

NAVER Cloud Seoul, Korea

AI Research Intern, Digital Healthcare Lab *Jun 2025 – Aug 2025*

- Fine-tuned Gemma-3-27B with LoRA (TRL + PEFT) for structured drug-target interaction prediction as ordered tuples; designed component-level F1 evaluation across exact-match and three element-wise scores.
- Probed base LLMs on SMILES and protein-sequence comprehension to diagnose failure modes; iterated prompt from zero-shot CoT to 2-shot JSON schema; curated 21-class output taxonomy with domain experts.
- Surveyed biomedical foundation models across text, vision, science, and multimodal modalities; delivered internal briefing to 40+ researchers and executives.

PROJECTS

Muddbites Telemetry *Feb 2026 – Present*

- Shipped inventory-analytics platform used by a Northwestern student-org vending team: warehouse + per-slot stock ledger with atomic transactions, receipt OCR with line-item stitching, slot-level sales.
- Designed two-phase insights engine separating price rules from structural rules, with a blocked-product set preventing pending price changes from polluting allocation math.
- Implemented price-elasticity modeling grounded in food-pricing literature (Andreyeva 2010); slot-day accounting via interval intersection against the assignment timeline.

Lyrics-based Genre Classification *Feb 2026 – Mar 2026*

- Built LoRA fine-tuning benchmark comparing three transformer architectures (RoBERTa, GPT-2, T5) on lyrics-only 5-genre classification; matched 590K adapter budgets isolated architecture effects from capacity.
- Designed four-part interpretability analysis (sentence-embedding centroids, layer probing, UMAP/t-SNE geometry, contrastive Integrated Gradients); confusion rates tracked inter-genre linguistic overlap (Pearson $r = 0.63$), locating the F1 0.61 plateau in task-level signal limits rather than model capacity.

Sync • synccal.vercel.app *Nov 2025 – Feb 2026*

- Built full-stack scheduling app with two-way Google Calendar and Outlook sync; handled recurring events at instance level with stable per-instance identifiers reconciled across platforms.
- Implemented incremental sync with change-token recovery and HTTP-410 full-resync fallback; server-side RRULE expansion for timezone-aware availability.
- Designed group scheduling that intersects free slots across participants using UTC-millisecond slot IDs for timezone-independent overlap; configurable working-hour window, slot granularity, and minimum-duration parameters.

MediaMatch *Aug 2025 – Nov 2025*

- Built cross-media recommender (movies, TV, books) with a router over three models — SVD collaborative filtering (MovieLens 25M), content-based taste-vector embeddings, item-item CF.
- Designed dual recommendation rails ("For You" personalized, "People Love" community-driven) with anchored weighted sampling (top 3 fixed, remainder proportional to score^2) and per-media quality gates.
- Engineered catalog enrichment pipeline: on-the-fly vector-DB embedding, EMA-blended review vectors, two-tier write-through cache over external metadata APIs.

TECHNICAL SKILLS

- **Languages:** Python, JavaScript/TypeScript, Java, C, C++, SQL, R
- **AI/ML:** Deep Learning, NLP, Transformers, LLM Fine-Tuning, LoRA/PEFT, TRL, PyTorch, Hugging Face, scikit-learn, NumPy, Pandas, RDKit, Qdrant, sentence-transformers, NeMo, Megatron-Core, CUDA, Distributed Training
- **Web:** React, Next.js, Node.js, Express, FastAPI, Tailwind CSS, Firebase/Firestore, MongoDB
- **Tools:** Git, GitHub Actions, Docker, Linux, Slurm, Jupyter/Colab