

WooAh Choi

Machine Learning & Backend Engineer

wooah.choi2@gmail.com | github.com/hncpj | linkedin.com/in/wooah-choi | backenddeveloper.tistory.com

PROFESSIONAL SUMMARY

Machine Learning Engineer with a backend background focused on forecasting, information retrieval, and decision-making under uncertainty. I'm strengthening ML/NLP/IR fundamentals (math-first), including reinforcement learning and LLM-based NLP, and applying them to an energy trading research project built around leakage-resistant evaluation and backtesting.

WORK EXPERIENCE

Konan Technology Backend Software Engineer | Nov 2020 - Jun 2023

Built and improved enterprise search and analysis services centered on information retrieval, indexing pipelines, and ranking logic. Applied NLP/ML approaches including BERT-based NER, SFX, SRL, fastText, and vector search. Leveraged search logs and time-series signals to ship query intelligence features (related terms, recommendations, popular queries, typo suggestions) and support quality monitoring and stabilization.

• Supreme Court of Korea - E-litigation Search and Analysis

Contributed to improving query understanding for a judicial information search service, focusing on intent-driven retrieval for case law and rulings. Worked on a query analysis approach using speech act analysis (SFX) and semantic role labeling (SRL) to derive structured constraints from user queries, such as time (recency), location (court), and case context. Collected and processed search logs, identified patterns, and ran transfer-learning experiments for SFX classification using PyTorch.

For entity extraction, trained and evaluated a BERT-based NER model to detect entities such as case number, law name, and party name, and supported mapping extracted tags to retrieval fields. When baseline quality was not sufficient, fine-tuned the NER model, re-indexed documents with updated tags, and measured a 4%+ improvement in extraction quality in internal evaluation.

Tech: Java, Spring, Oracle, Jenkins, GitLab, Python, PyTorch, BERT, Docker, Linux, C++

• Millie's Library - Book Search and User Pattern Analysis

Built an indexing automation pipeline handling 100M+ data points and designed a dynamic volume system intended to scale beyond 1B records.

Improved sorting quality by 38 percent by introducing a more detailed ranking logic and validating changes via monitoring during stabilization. Conducted user pattern analysis to refine retrieval behavior and ranking priorities.

Tech: C++, Python, MySQL, Linux

• Heungkuk Insurance - Main Website Search Renewal

Renewed the web search experience and shipped query assistance features including auto-completion and popular search terms. Improved usability with AJAX-based continuous scrolling and supported production rollout with log-based quality checks.

Tech: Java, Oracle, SVN, Linux, JavaScript

• Ministry of SMEs and Startups - Venture Confirmation System Search

Implemented search features and APIs, including text extraction and indexing from documents stored in AWS S3. Delivered supporting backend utilities such as recursive file and folder deletion logic to handle operational workflows.

Tech: Java, PostgreSQL, SVN, Linux, AWS S3

TECHNICAL SKILLS

Languages: TypeScript, Java, JavaScript, SQL

Frameworks: NestJS, Spring, Spring Boot, Express.js

Databases: PostgreSQL, MySQL, Redis, MongoDB

Cloud: AWS (EC2, S3, RDS, Lambda), Docker, Jenkins, GitHub Actions

API: REST API, GraphQL, OAuth2, JWT

Tools: Git, Linux, Nginx, TypeORM, JPA

PROJECTS

- **Grid Flow Trade - Battery Energy Storage Trading (Research)**

Research and design work for a battery energy storage trading system in the UK electricity market. Defining market scope, data requirements, and a leakage-resistant evaluation setup (rolling or walk-forward validation) before implementation. Next milestones include building a baseline forecasting and backtesting loop, then iterating toward probabilistic forecasting for risk-aware bidding and position sizing.

Tech: Time Series Forecasting, Backtesting, Probabilistic Forecasting, Rolling Validation, Python, PostgreSQL

Status: Research

- **Generative Storytelling and Illustration AI Agent**

Prototype pipeline that turns keyword inputs into stories using generative AI APIs, then generates picture-book style illustrations from the story prompts. Iterating on prompt templates and orchestration to improve consistency, with a planned extension toward short-form animation workflows.

Tech: LLM APIs, Prompting, Automation, Pipeline Design

Status: Prototype

- **[Kaggle] Google AI4Code - Notebook Cell Ordering**

Transformer-based approach (CodeBERT, DistilBERT) for predicting notebook cell order, including EDA, feature engineering, and iterative experiments. Result: Top ~17%.

Tech: Transformers, CodeBERT, DistilBERT, NLP, Experimentation

Status: Completed

- **[Kaggle] RSNA 2022 Cervical Spine Fracture Detection**

Medical imaging pipeline with augmentation, normalization, and ROI extraction, followed by EfficientNetV2-based classification experiments for slice-level targets. Result: Top ~28%.

Tech: Computer Vision, EfficientNetV2, Preprocessing, Model Training

Status: Completed

- **[ATE] Recipe and Ingredients Management**

Full-stack application for managing recipes and tracking ingredients, including recipe CRUD, inventory tracking, and shopping list generation.

Tech: NestJS, TypeORM, PostgreSQL, TypeScript, REST API

Status: In Development

- **SO Easy - Share Office Platform**

Platform connecting shared office providers with businesses seeking flexible workspaces. Includes availability, booking management, and payment integration.

Tech: NestJS, TypeScript, PostgreSQL, GraphQL, AWS

Status: In Development