

PORTFOLIO

AI Researcher / Engineer

Phone

010-9205-0971

Email

jshim0978@gmail.com

INTRODUCTION

안녕하세요, 심정우입니다.

I started my journey with web development and expanded into AI and data analysis, working on diverse projects. My key interests lie in machine learning and natural language processing (NLP), focusing on data-driven decision-making and model optimization.

Currently, I am researching solutions for hallucination issues in large language models (LLMs) to develop more reliable AI systems. I enjoy learning new technologies and applying them to real-world projects, always striving for continuous growth and innovation.



EDUCATION

학력 사항



Korea University (PRML Lab)

M.S. in Artificial Intelligence, 2023.03 - 2025.02

 Research Topics : Natural Language Processing (NLP), Large Language Models (LLM), Hallucination Mitigation

GPA: 4.21/4.5



Chungnam National University

B.S. in Computer Science, 2015.03 - 2021.02

 Main Courses : Data Science, Artificial Intelligence, Network Structure, Web Programming

GPA: 3.18/4.5

WORK EXPERIENCE

경력 사항

Mobile Entropy

Research Department,
2021-2023



솔루션/SI/ERP/CRM
산업

중소기업
기업형태

38명
직원수

2001.10.16
설립

대표 김도균

매출 40억 원 (2019)

주소 서울 마포구 독막로 275 6층

웹사이트 -

소개 모바일엔트로피(주)는 1999년 사업 초기부터 연구와 개발 전문팀을 운영하여 현재까지 고객의 눈높이에 맞는 맞춤형 솔루션을 공급하고 시스템별 전담 지원조직으로 안정적 운영 서비스를 제공하며, 도시가스사의 특화된 솔루션 서비스를 20년간 공급 및 운영하고 있습니다. 특화된 솔루션으로는 M.E Billing 솔루션, 고객지원 솔루션, GIS 현장지원 솔루션, 관제 솔루션, AI원격검침 솔루션, RPA 솔루션이 도시가스 및 상하수도에 최적화되어 있습니다. 향후에도 도시가스 및 상하수도 솔루션을 연구개발 하여 미래의 시스템에 완벽을 추구하고자 합니다.

Position

Research Department Manager (Industrial Technical Personnel)

Main Tasks

Full Stack Web Development and In-House Project Management

Projects

- CNCITY Customer Privacy Masking Project
- JB Smart Safety System Project
- KakaoPay Smart Sendbox
- Developing OpenGiro Payment System
- B2C Kakao Authentication
- B2C Kakao e-Documents
- NHN Cloud Migration of In-House Projects
- Monitoring System for In-House Projects

핵심 역량 – Soft Skills



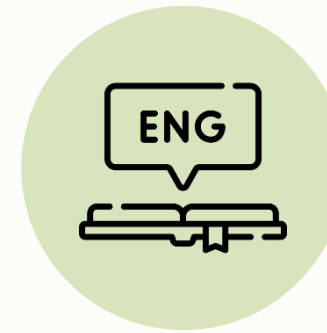
Collaboration

With experience in web development and involvement in AI platform development, I have honed my ability to collaborate effectively within cross-functional teams, ensuring smooth communication and successful project execution.



Adaptation

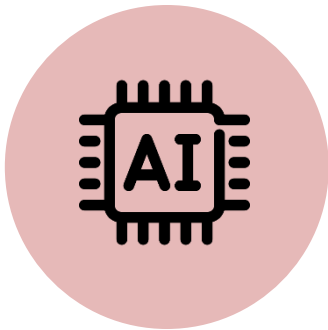
My experience spans from web development to AI research and platform operations, demonstrating my ability to quickly adapt to new technologies, workflows, and industry demands while maintaining high-quality outcomes.



English Skills

With international experience, I have attained a high level of English proficiency, including a TOEIC score of 980, TEPS 556, and an OPIc AL rating. This enables me to communicate effectively in global environments and excel in technical documentation and collaborative projects.

핵심 역량 – Technical Skills



AI & Machine Learning

- | | Confidence |
|---------------------------|------------|
| ▶ TensorFlow, PyTorch | □□□□□□ |
| ▶ Data Preprocessing | □□□□□□ |
| ▶ Data Analysis | □□□□□□ |
| ▶ LLM fine-tuning | □□□□□□ |
| ▶ LLM Prompt Engineering | □□□□□□ |
| ▶ LLM Prompt Optimization | □□□□□□ |
| ▶ Model Evaluation | □□□□□□ |



Web Development

- | | Confidence |
|---------------------------|------------|
| ▶ React.js, JavaScript | □□□□□□ |
| ▶ RESTful API Development | □□□□□□ |
| ▶ Spring Boot, FastAPI | □□□□□□ |
| ▶ Authentication (JWT) | □□□□□□ |
| ▶ RDBMS (MySQL, Oracle) | □□□□□□ |
| ▶ API Load Balancing | □□□□□□ |
| ▶ Security (HTTPS, CORS) | □□□□□□ |



Tools & Platforms

- | | Confidence |
|-----------------------------|------------|
| ▶ Git | □□□□□□ |
| ▶ Linux (Ubuntu, CentOS) | □□□□□□ |
| ▶ Docker | □□□□□□ |
| ▶ Prometheus & Grafana | □□□□□□ |
| ▶ CI/CD Pipelines (Jenkins) | □□□□□□ |
| ▶ NHN Cloud Services | □□□□□□ |
| ▶ Shell Scripting | □□□□□□ |

연구 활동

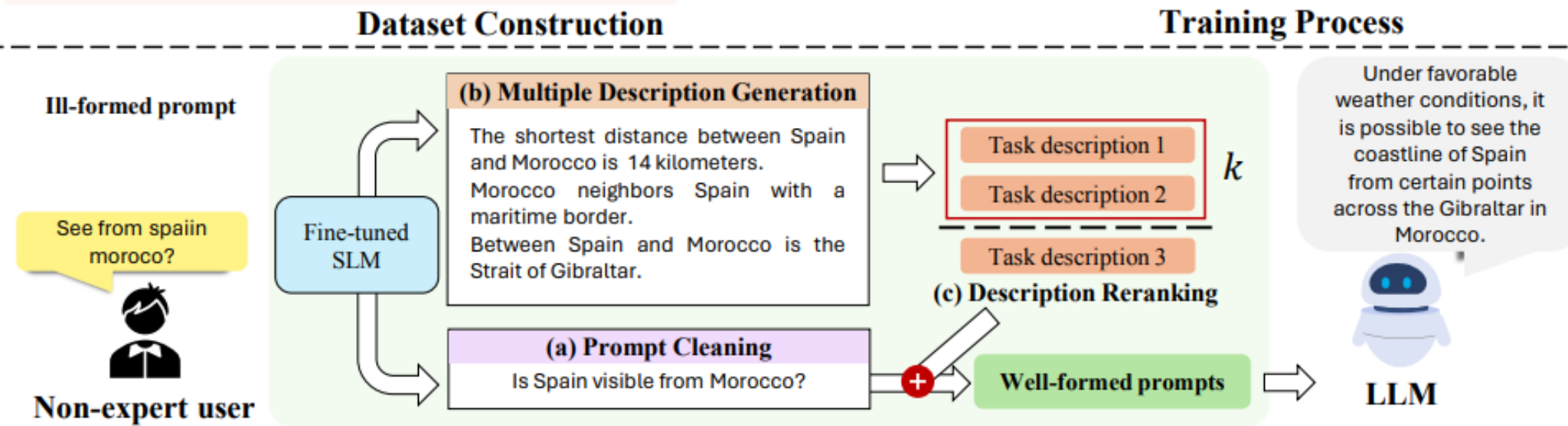
IEEE 2024 System, Man, Cybernetics

Jung-Woo Shim, Yeong-Joon Ju, Ji-Hoon Park, Seong-Whan Lee,
"CPR: Mitigating Large Language Model Hallucinations with Curative Prompt Refinement",
(ACCEPTED), (Sep. 2024)



H-INDEX

63



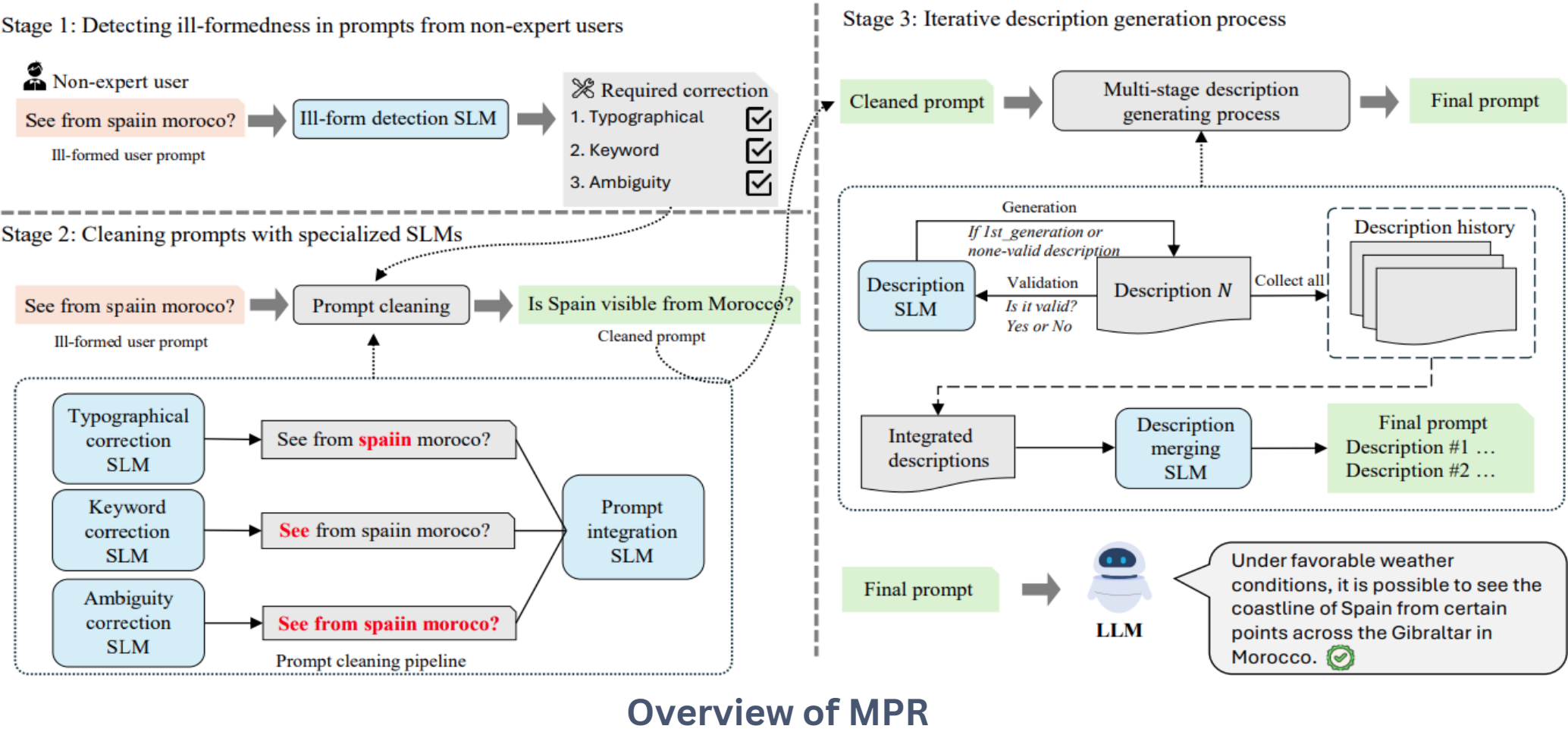
Overview of CPR

We introduce *Curative Prompt Refinement (CPR)*,
a *plug-and-play framework for curative prompt refinement*
that

- 1) *cleans ill-formed prompts, and*
- 2) *generates additional informative task descriptions to align the intention of the user and the prompt using a fine-tuned small language model. When applied to language models, we discover that CPR significantly increases the quality of generation while also mitigating hallucination.*

Elsevier Pattern Recognition

Jung-Woo Shim, Yeong-Joon Ju, Ji-Hoon Park, Seong-Whan Lee,
"Multi-stage Prompt Refinement for Mitigating Hallucinations in Large Language Models",
(UNDER REVIEW), (2025)



We introduce Multi-stage Prompt Refinement (MPR), a framework designed to systematically improve these ill-formed prompts across multiple stages. Each stage addresses specific errors such as punctuation, typographical mistakes, and misuse of key terms, using small language models (SLMs) fine-tuned for these tasks. MPR iteratively enhances the clarity of prompts with additional context and employs a self-reflection mechanism with ranking to prioritize the most relevant input.

PROJECTS

Mobile Entropy

JB Smart Safety System

KakaoPay Smart SendBox

NHN Cloud Migration of In-House Projects

Monitoring System for In-House Projects

Korea University

ElementaryGPT : LLM for Elementary School Teachers

CBRNGPT : LLM for Chemical, Biological, Radioactive, and Nuclear Situations

COMA : Career Optimization and Management Assistant

GenAI Academy : Generating Educational Tools Using Generative AI and Digital Twins

PROJECTS

JB Smart Safety System (2021)

: IoT Device Control System for Gas Pipes

Tech Stack

- ▶ Frontend: jQuery, JavaScript
- ▶ Backend: Spring Boot, OracleDB
- ▶ IoT & Data Processing: MQTT, WebSockets

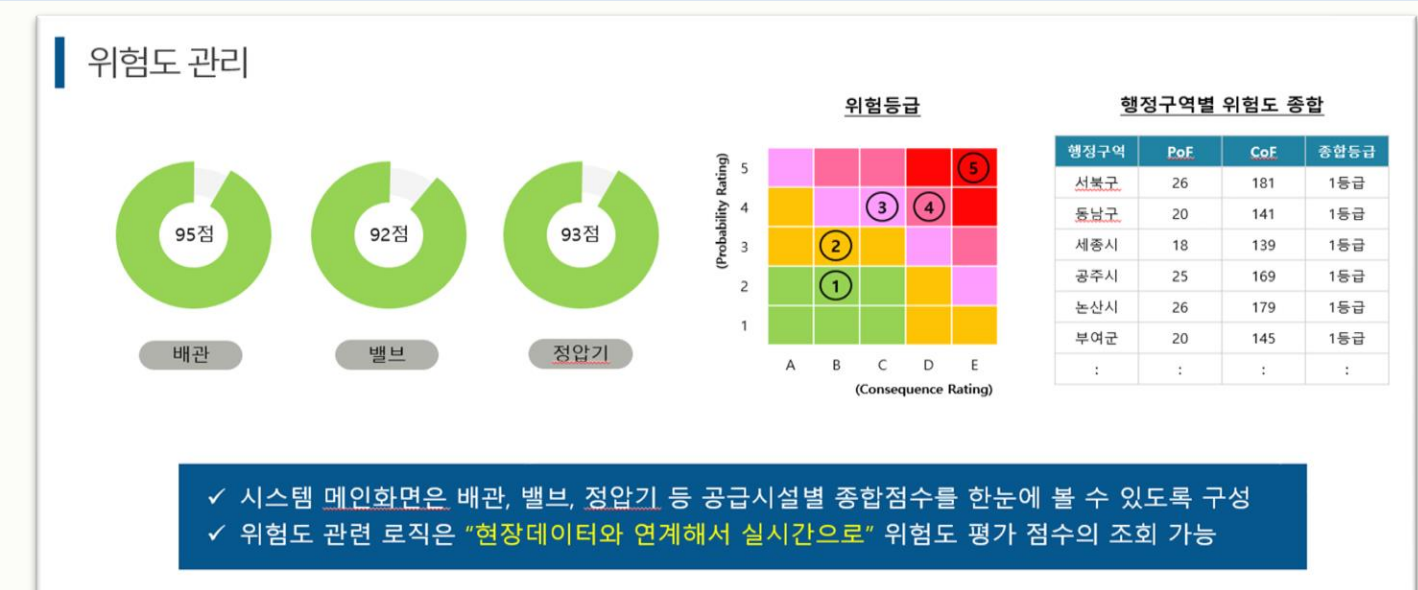
My Role (Team of 2)

- Frontend Development (*Primary Role*)
 - Designed and implemented the dashboard UI for real-time safety monitoring, using jQuery.
 - Developed responsive UI components for managing IoT devices, alarms, and system logs.
 - Integrated the frontend with RESTful APIs, ensuring smooth data communication with the backend.
- Backend Development (*Partial Contribution*)
 - Assisted in developing API endpoints using Spring Boot for safety evaluations, device control, and data retrieval.
 - Contributed to real-time data processing and alarm systems for IoT sensors and pipeline safety monitoring.
 - Helped implement authentication and access control for different user roles in the system.

Contribution ★★☆☆☆



IoT Device Control Page



Requirements for Risk Management

KakaoPay Smart Sandbox

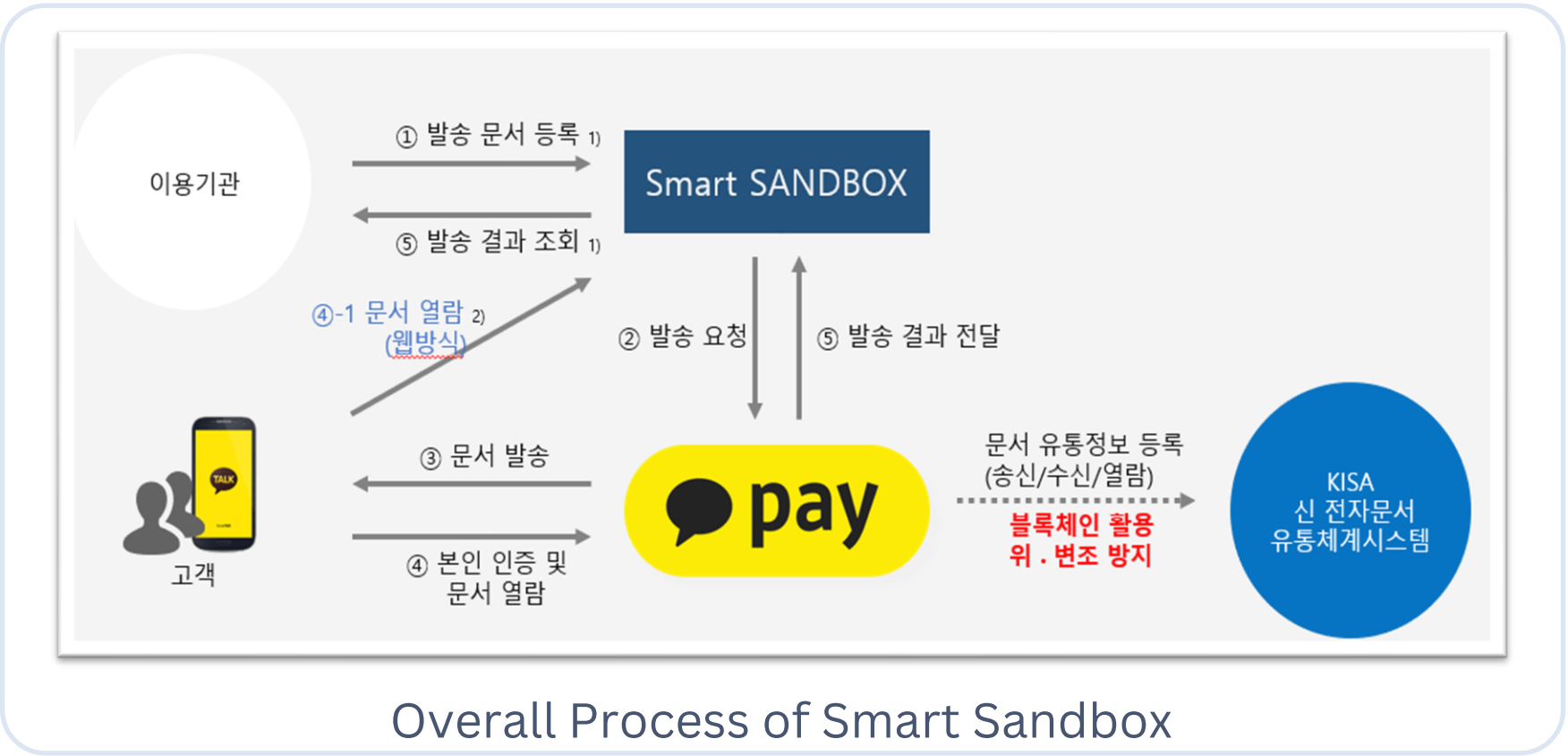
: KakaoPay e-Document Sender for Gas Company Customers

Tech Stack

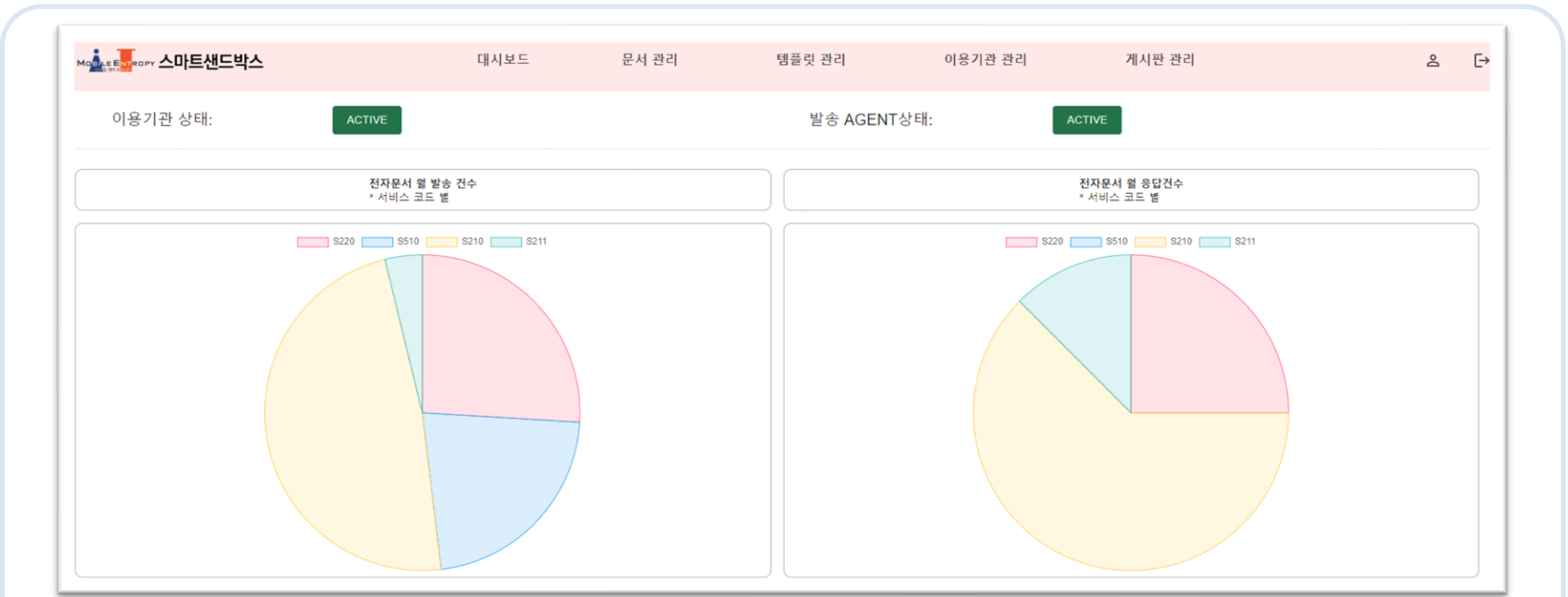
- ▶ Frontend: React.js, Node.js, JavaScript, Redux
- ▶ Backend: Spring Boot, MySQL, Oracle
- ▶ API Integration: KakaoPay API, RESTful APIs

My Role (Team of 3)

- Frontend Development (*Primary Role*)
 - Designed and developed the dashboard using React.js.
 - Created a responsive UI/UX for each user roles (admin, client, general user).
 - Integrated API endpoints with the frontend to handle document transactions, notifications, and user authentication.
 - Developed real-time data visualization components for monitoring system activity and transaction status.
- Backend Development (*Partial Contribution*)
 - Assisted in developing RESTful APIs using Spring Boot for data retrieval and processing.
 - Contributed to the authentication and authorization system using Spring Security + JWT.
 - Helped implement backend logic for document processing, KakaoPay API integration, and notification management.



Overall Process of Smart Sandbox



Dashboard Page of Smart Sandbox

PROJECTS

NHN Cloud Migration (2023)

: of In-House Projects

Tech Stack

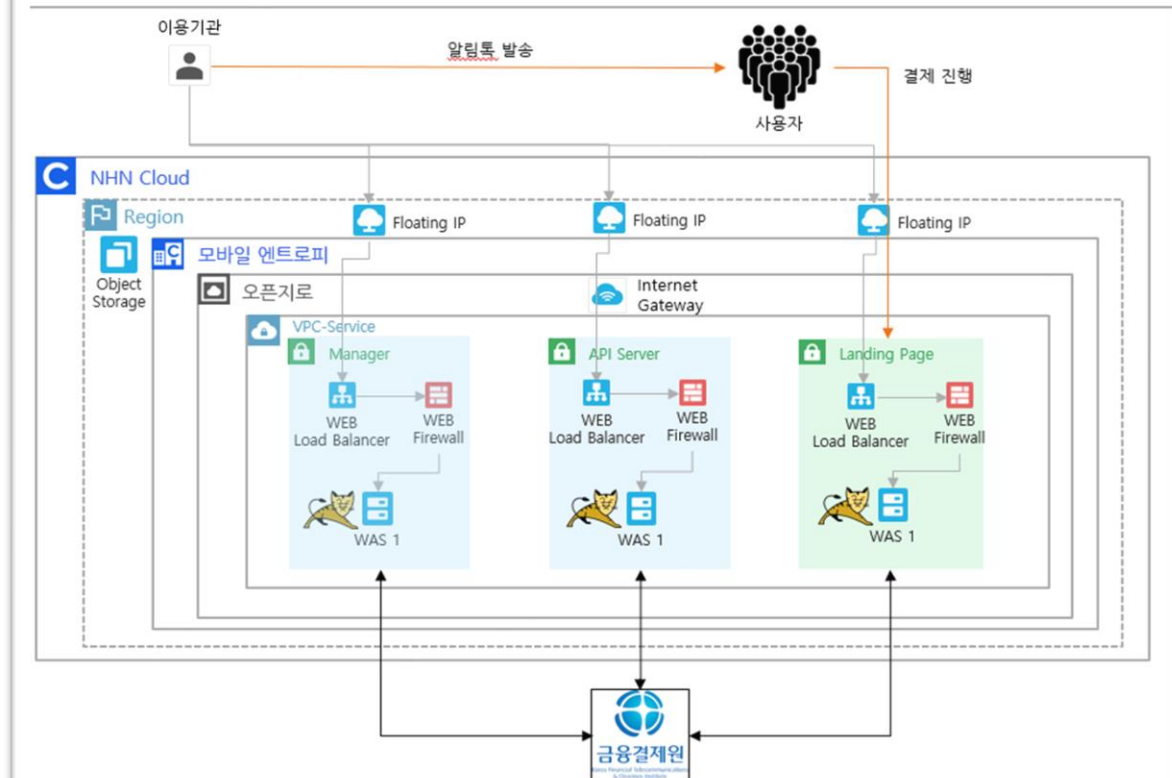
- ▶ Cloud Platform: NHN Cloud
- ▶ Networking: VPC, Floating IP, Load Balancer, Internet Gateway
- ▶ Infrastructure: Linux (CentOS)

My Role (Team of 1)

- Cloud Migration & Infrastructure Setup (*Primary Role*)
 - Led the migration of all company projects from local servers to NHN Cloud.
 - Designed and implemented the network architecture, ensuring secure and efficient cloud connectivity.
 - Configured VPC, subnets, security groups, and access controls to meet project requirements.
- Project Deployment & Cloud Management
 - Migrated multiple enterprise projects including:
 - KakaoPay Payment & Authentication Systems
 - OpenGiro Service
 - Smart Sandbox Platform
 - Set up instance provisioning, OS configurations, and database management for each project.
 - Established load balancers, floating IP assignments, and network routing to ensure seamless traffic flow.

Contribution ★★★★★

오픈지로 - 네트워크 / 하드웨어 / 소프트웨어 구성도



하드웨어/소프트웨어 상세

관리페이지 / 랜딩페이지

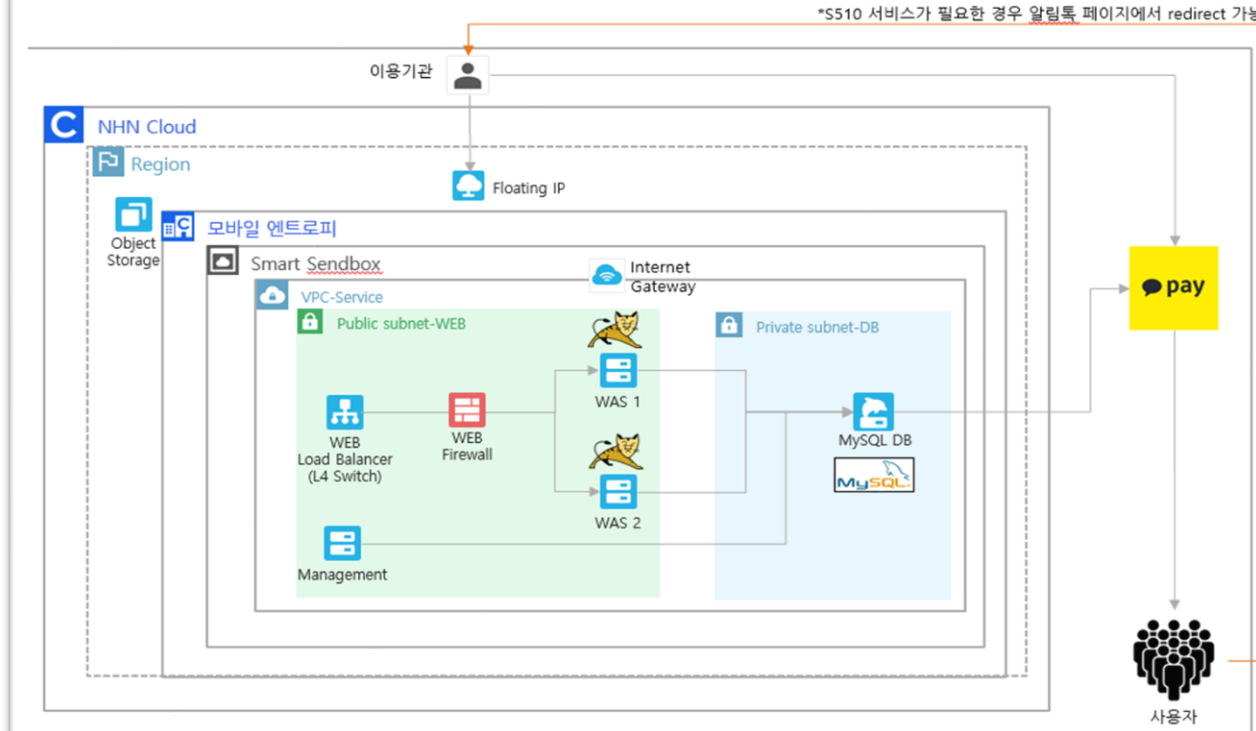
WAS	Tomcat 8.5
Framework	Node.js + React.js + Spring Boot 2.6.2 + JWT
JDK	Open JDK 11
OS	Linux (CentOS 7.8)
H/W	1Core CPU / 2Gb RAM / 40Gb HDD

API & File 서버

WAS	Tomcat 8.5
Framework	Node.js + React.js + Spring Boot 2.6.2 + JWT
JDK	Open JDK 11
OS	Linux (CentOS 7.8)
H/W	4Core CPU / 8Gb RAM / 40Gb HDD

Structure of Project OpenGiro

스마트샌드박스 - 네트워크 / 하드웨어 / 소프트웨어 구성도



하드웨어/소프트웨어 상세

운영 서버

WAS	Tomcat 8.5
Framework	Node.js + React.js + Spring Boot 2.6.2 + JWT
JDK	Open JDK 11
OS	Linux (CentOS 7.8)
H/W	1Core / 2Gb RAM / 60Gb HDD

개발 환경

WAS	Tomcat 8.5
Framework	Node.js + React.js + Spring Boot 2.6.2 + JWT
JDK	Open JDK 11
OS	Linux (CentOS 7.8)
H/W	1Core / 2Gb RAM / 60Gb HDD

Structure of Project Smart Sandbox

PROJECTS

(2023)

Monitoring System

: for In-House Projects

Tech Stack

- ▶ Monitoring : Prometheus, Grafana, Spring Boot Actuator
- ▶ Infrastructure: NHN Cloud, Docker, Kubernetes

My Role (Team of 1)

- System Design & Implementation (*Primary Role*)
 - Designed and developed a real-time monitoring system for company-wide applications and cloud infrastructure.
 - Integrated Spring Boot Actuator, Prometheus, and Grafana to enable comprehensive system observability.
 - Deployed custom monitoring dashboards to track CPU, memory, request latency, database performance, and API response times.
- Prometheus-Based Metric Collection & Grafana Visualization
 - Designed interactive Grafana dashboards to visualize system health, error rates, and traffic patterns.
 - Enabled custom query-based alerts for downtime, high CPU/memory usage, and API failures.
 - Integrated multiple data sources (Prometheus, NHN Cloud logs) for a centralized monitoring view.

Contribution ★★★★★

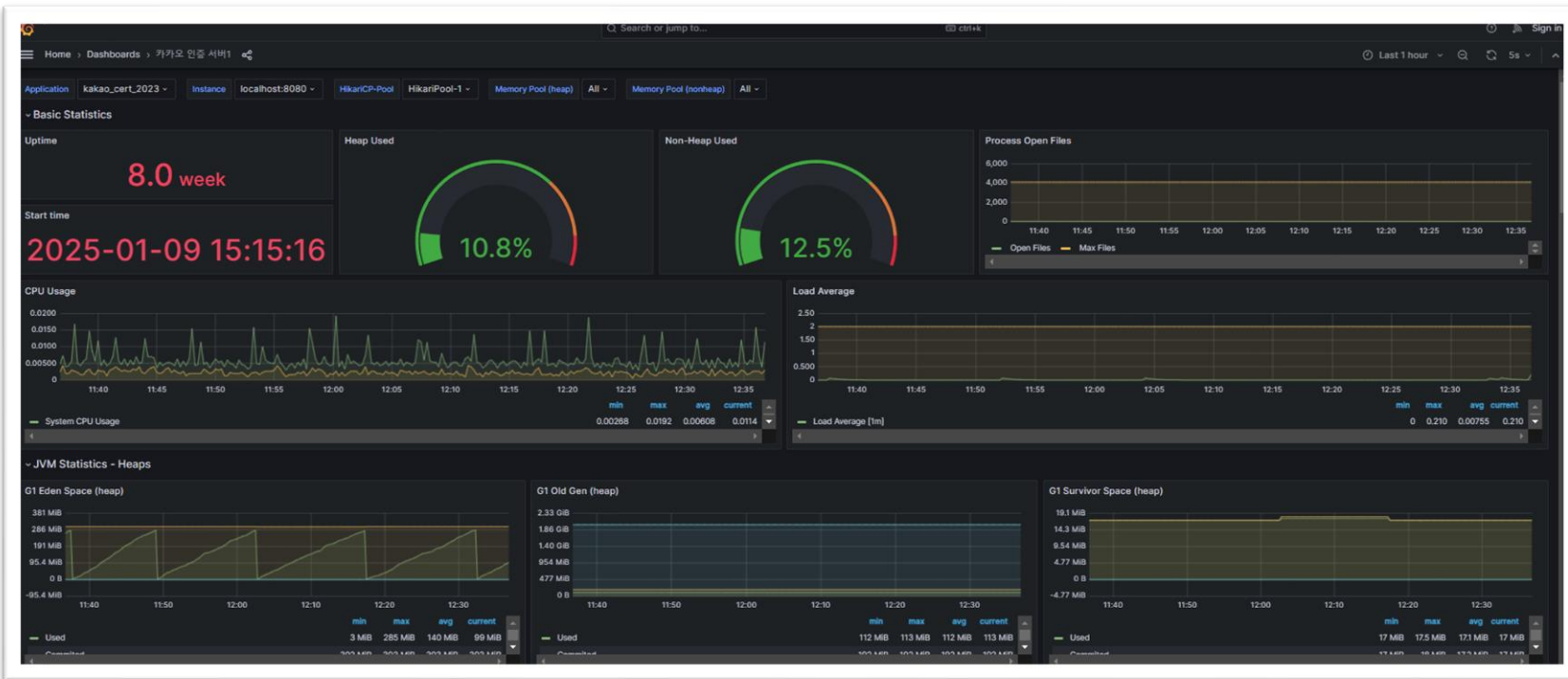
<https://mobileentropy.notion.site/Springboot-Prometheus-Grafana-cccd24d628ab4e7cb6c8b0977b67947a>

Springboot + Prometheus + Grafana

이 글은 플랫폼 서비스에 적용한 (중인) 모니터링 시스템 설치 방법을 설명한다.



Monitoring System Installation Guide



Monitoring Dashboard

PROJECTS

(2023)

ElementaryGPT

: LLM for Elementary School Teachers

Tech Stack

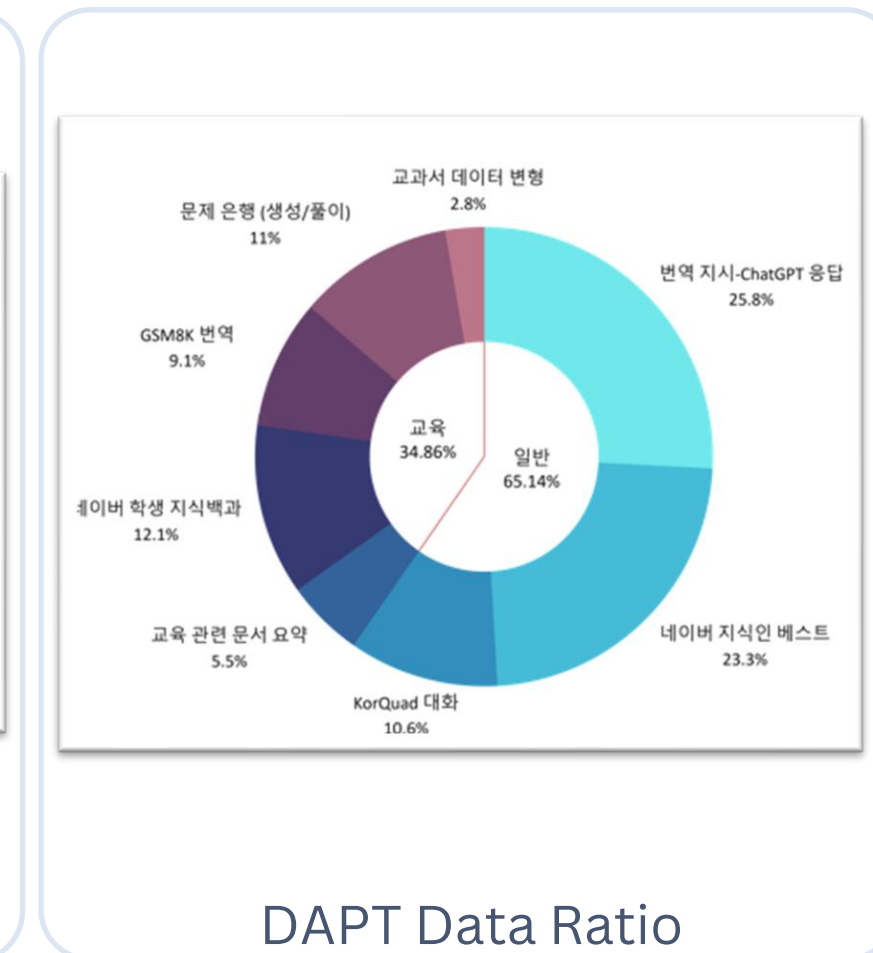
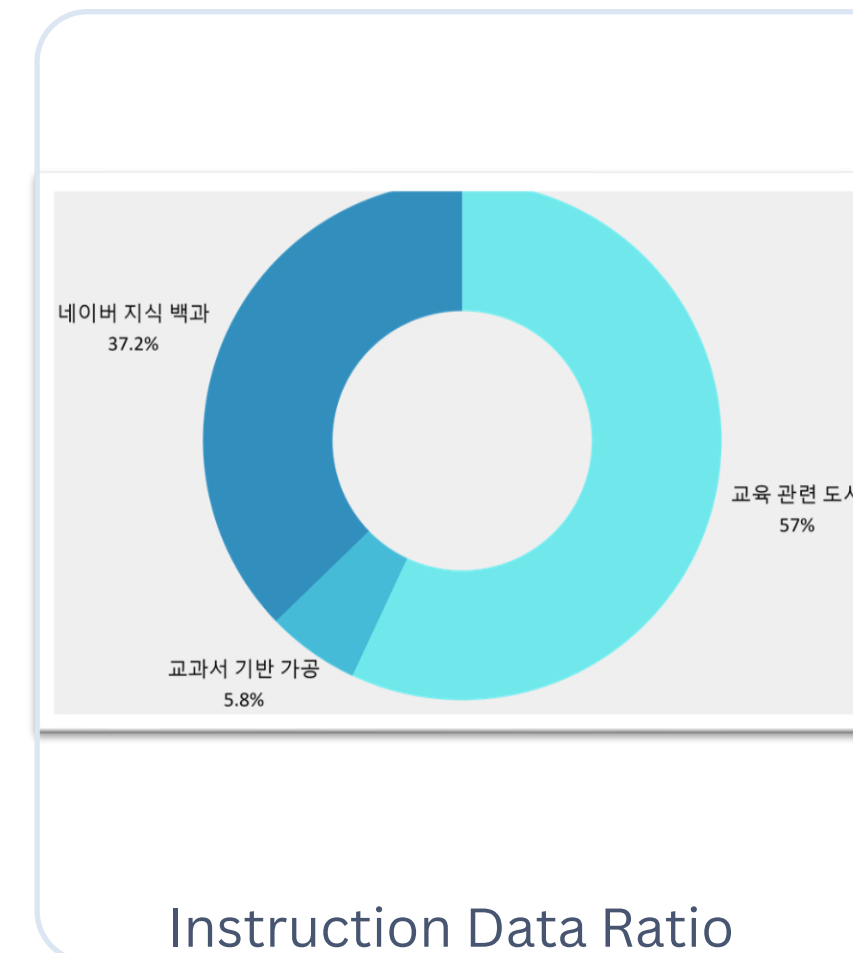
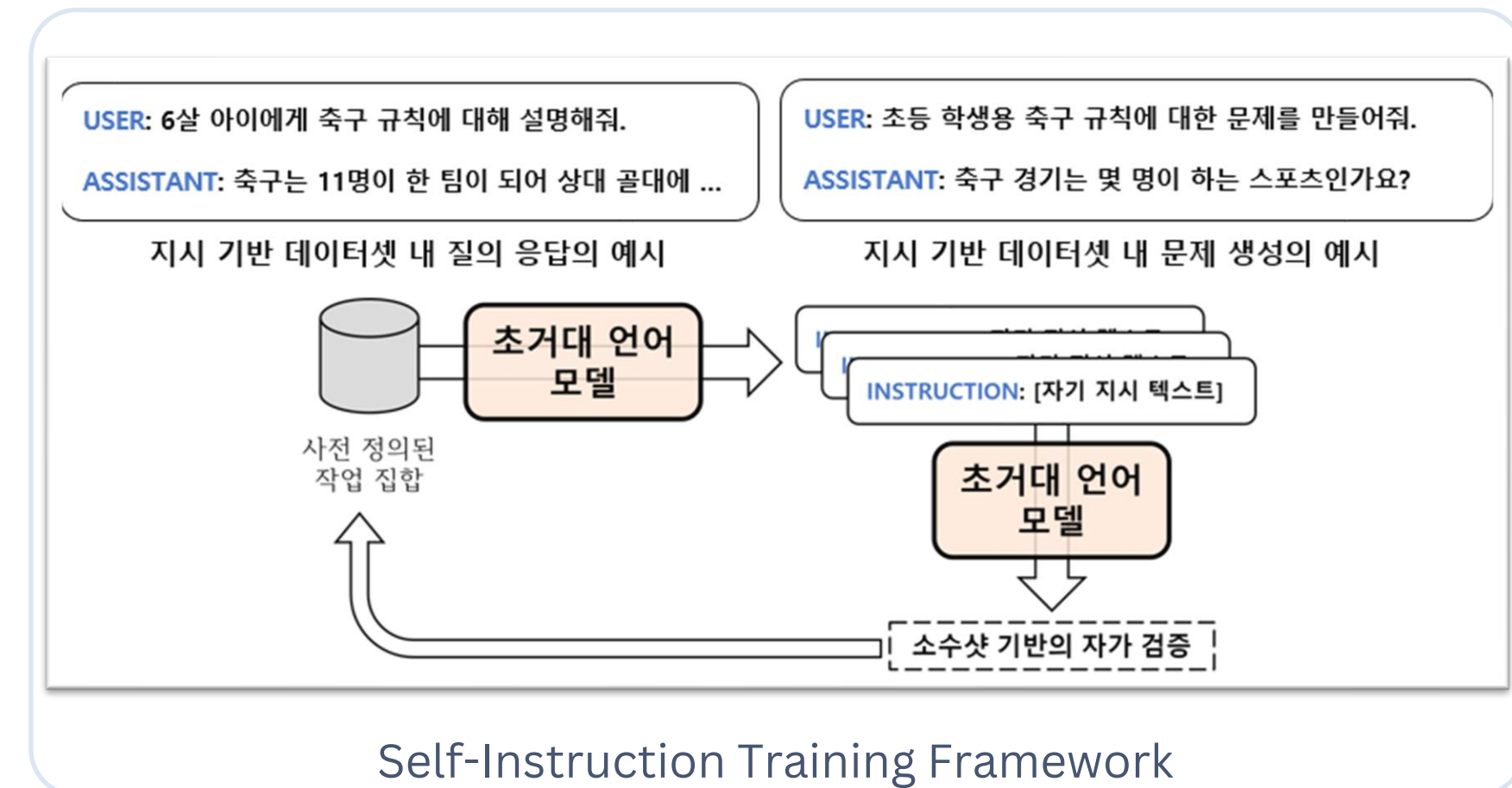
- ▶ LLM Models: Polyglot-Ko[1], Mistral[2], Solar[3]
- ▶ Pretraining: Domain Adaptive Pre-training (DAPT)[4], NEFTune[5]
- ▶ Data Processing: Python, Hugging Face Datasets, Pandas

My Role (Team of 3)

- AI Researcher & LLM Engineer (*Primary Role*)
 - Developed an LLM-based assistant for elementary school teachers, fine-tuned for curriculum-based Q&A.
 - Applied Domain Adaptive Pre-training (DAPT) to inject unseen textbook knowledge into the model.
 - Implemented instruction-following fine-tuning using datasets from KorQuAD, Naver Q&A, GSM8K translation, and problem banks.
 - Used NEFTune to improve robustness against text distortions and incomplete queries.
 - Trained and evaluated multiple LLMs (Polyglot-Ko 12.8B, Mistral 7B, Solar 10.7B) for accuracy and instructional clarity.
 - Addressed challenges in math reasoning and English textbook alignment using AlphaGeometry insights and tailored dataset curation.

Contribution ★★★★★

- [1]Ko, Hyunwoong, et al. "A technical report for polyglot-ko: Open-source large-scale korean language models." *arXiv preprint arXiv:2306.02254* (2023).
[2]Jiang, Albert Q., et al. "Mixtral of experts." *arXiv preprint arXiv:2401.04088* (2024).
[3]Kim, Dahyun, et al. "Solar 10.7 b: Scaling large language models with simple yet effective depth up-scaling." *arXiv preprint arXiv:2312.15166* (2023).
[4]Gururangan, Suchin, et al. "Don't stop pretraining: Adapt language models to domains and tasks." *arXiv preprint arXiv:2004.10964* (2020).
[5]Jain, Neel, et al. "Neftune: Noisy embeddings improve instruction finetuning." *arXiv preprint arXiv:2310.05914* (2023).



[1] Dubey, Abhimanyu, et al. "The llama 3 herd of models." arXiv preprint arXiv:2407.21783 (2024).
[2] Abdin, Marah, et al. "Phi-3 technical report: A highly capable language model locally on your phone." arXiv preprint arXiv:2404.14219 (2024).
[3] Team, Gemma, et al. "Gemma: Open models based on gemini research and technology." arXiv preprint arXiv:2403.08295 (2024).
[4] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in neural information processing systems 33 (2020): 9459-9474.
[5] Ampazis, Nicholas. "Improving rag quality for large language models with topic-enhanced reranking, Springer Nature Switzerland, 2024.

PROJECTS (2024)

CBRNGPT

: LLM for Chemical, Biological, Radioactive, and Nuclear Situations

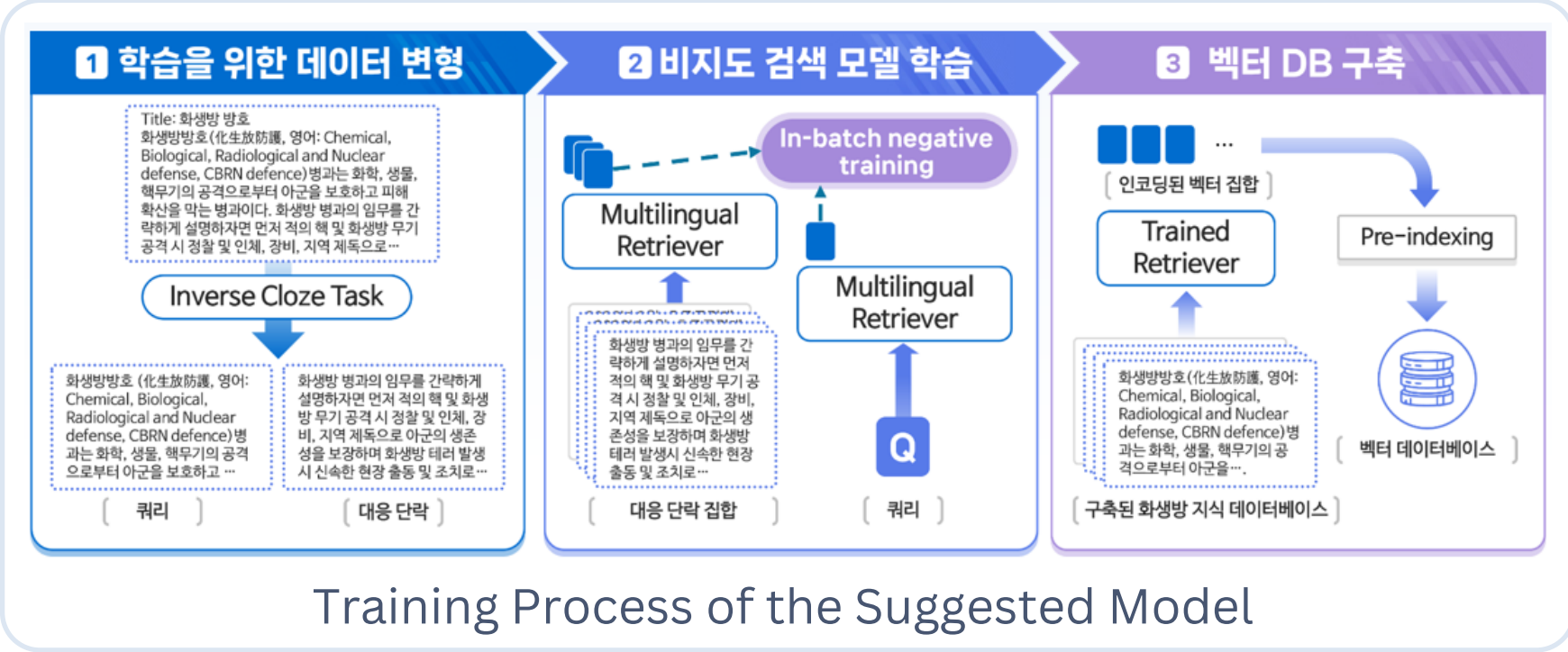
Tech Stack

- ▶ LLM Models: LLaMA-3[1], Phi-3[2], Gemma[3]
- ▶ RAG[4]: Dense & Sparse Vector Search, Cross-Encoder Reranking[5]
- ▶ Data Processing: Military Doctrines, Response Manuals

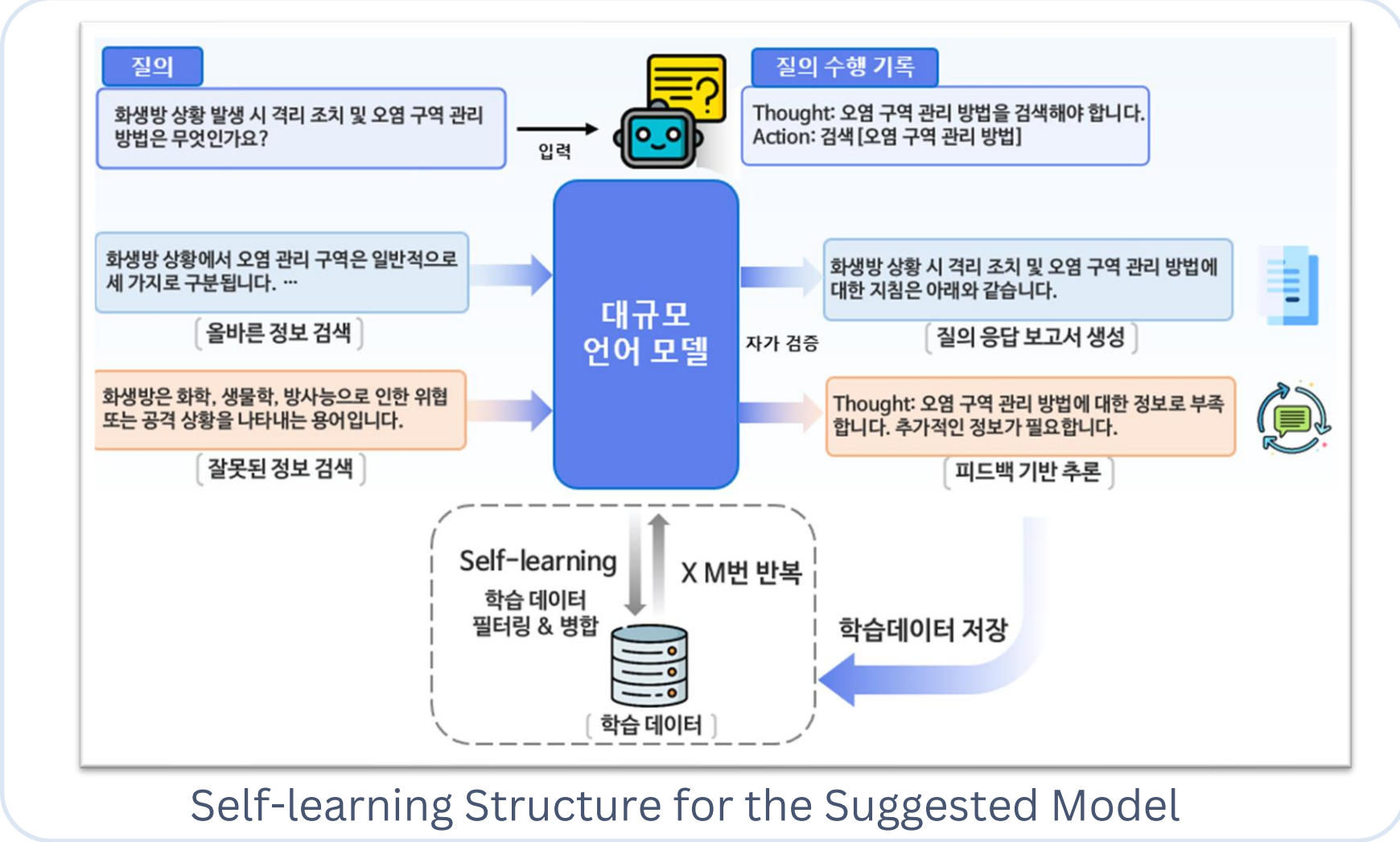
My Role (Team of 3)

- AI Researcher & LLM Engineer (Primary Role)
 - Developed an AI-based decision support system for Chemical, Biological, Radiological, and Nuclear (CBRN) operations.
 - Built a domain-specific knowledge database using NBC-RAMS data, military doctrines, and response manuals.
 - Implemented Retrieval-Augmented Generation (RAG) to enhance LLM accuracy in CBRN scenarios.
 - Designed a CBRN-specific search model with pre-indexing for vector DBs to improve retrieval speed and accuracy.
 - Developed an AI agent for real-time situational analysis, integrating NBC-RAMS API for contamination spread prediction.
 - Enhanced LLM robustness with self-learning mechanisms, improving decision-making capabilities through historical scenario-based training.

Contribution ★★★★★



Training Process of the Suggested Model



Self-learning Structure for the Suggested Model

PROJECTS (2024)

COMA

: Career Optimization and Management Assistant

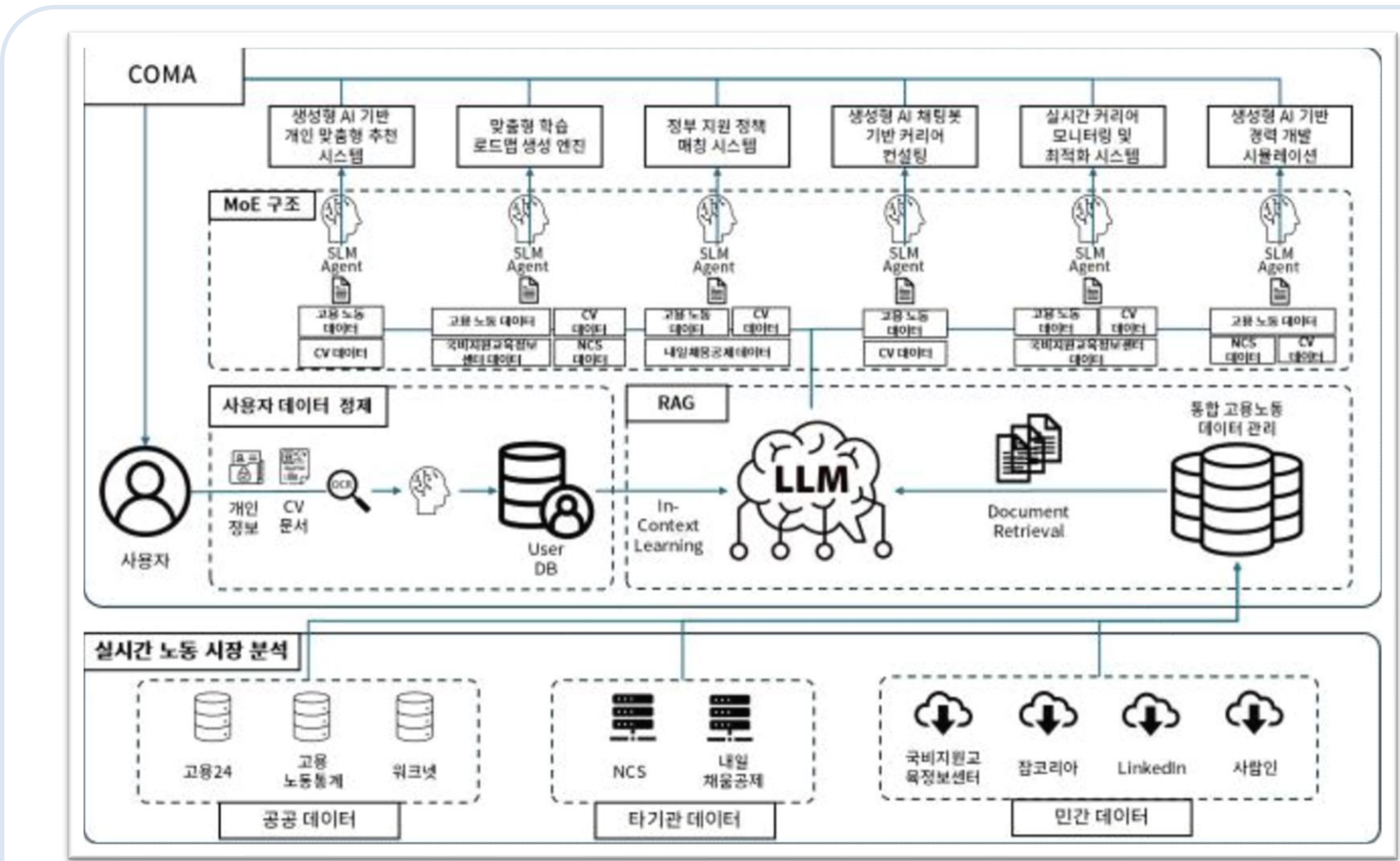
Tech Stack

- ▶ Frontend: React.js
- ▶ Backend: FastAPI
- ▶ Data Processing: Government labor data APIs (EIS, NCS, ELDS)

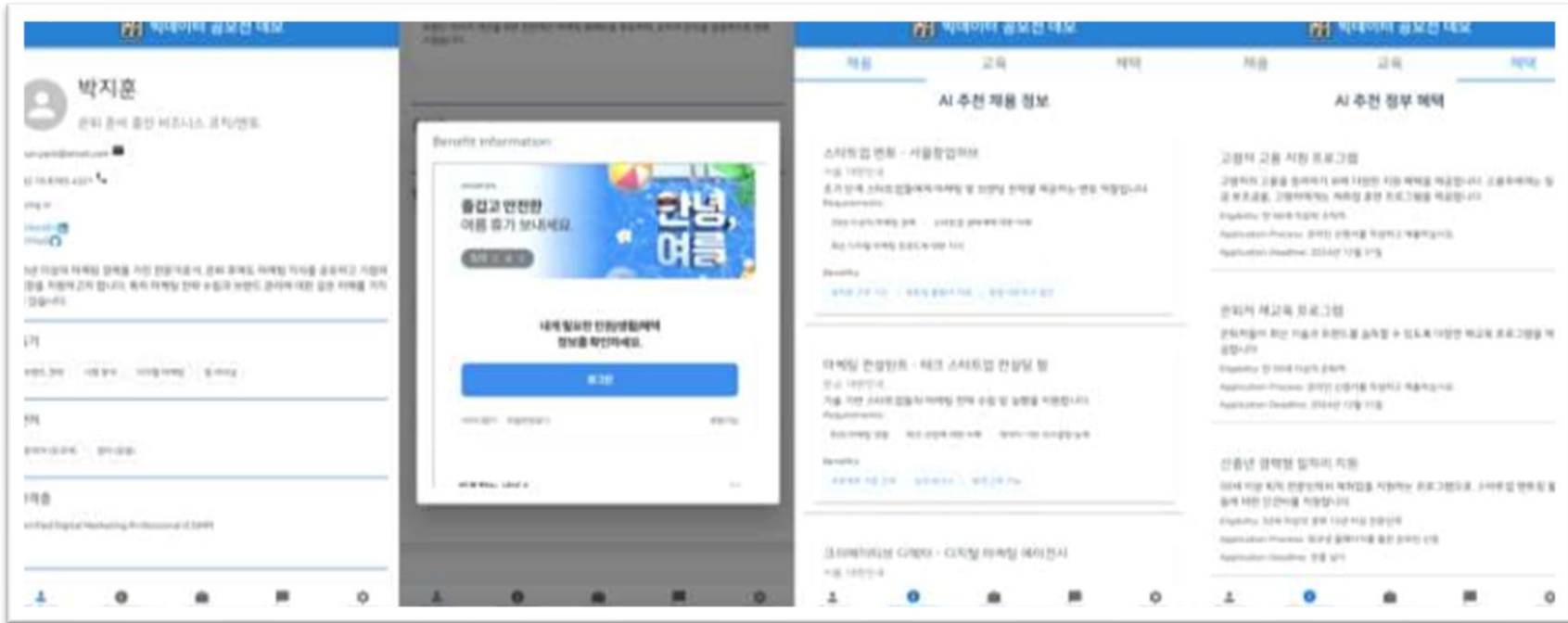
My Role (Team of 5)

- Full-Stack Developer & AI Engineer (*Primary Role*)
 - Developed an AI-powered career management platform integrating LLM-based job market analysis.
 - Built a web service with a React.js frontend and a FastAPI/Spring Boot backend, handling real-time labor data processing.
 - Designed an AI-driven recommendation system for career paths, education, and job matching using government labor data (EIS, NCS, ELDS).
 - Implemented RAG-based career consulting and a custom AI chatbot for job transitions and policy matching.
 - Developed real-time labor market monitoring and career simulation models to predict optimal job transitions and salary trends.
- Awarded the Grand Prize at the 3rd Ministry of Employment and Labor Public Data Utilization Contest.

Contribution ★★★★★



Overall Structure of COMA



Sample Pages of COMA

PROJECTS

(2024)

GenAI Academy

: Generating Educational Tools Using Generative AI and Digital Twins

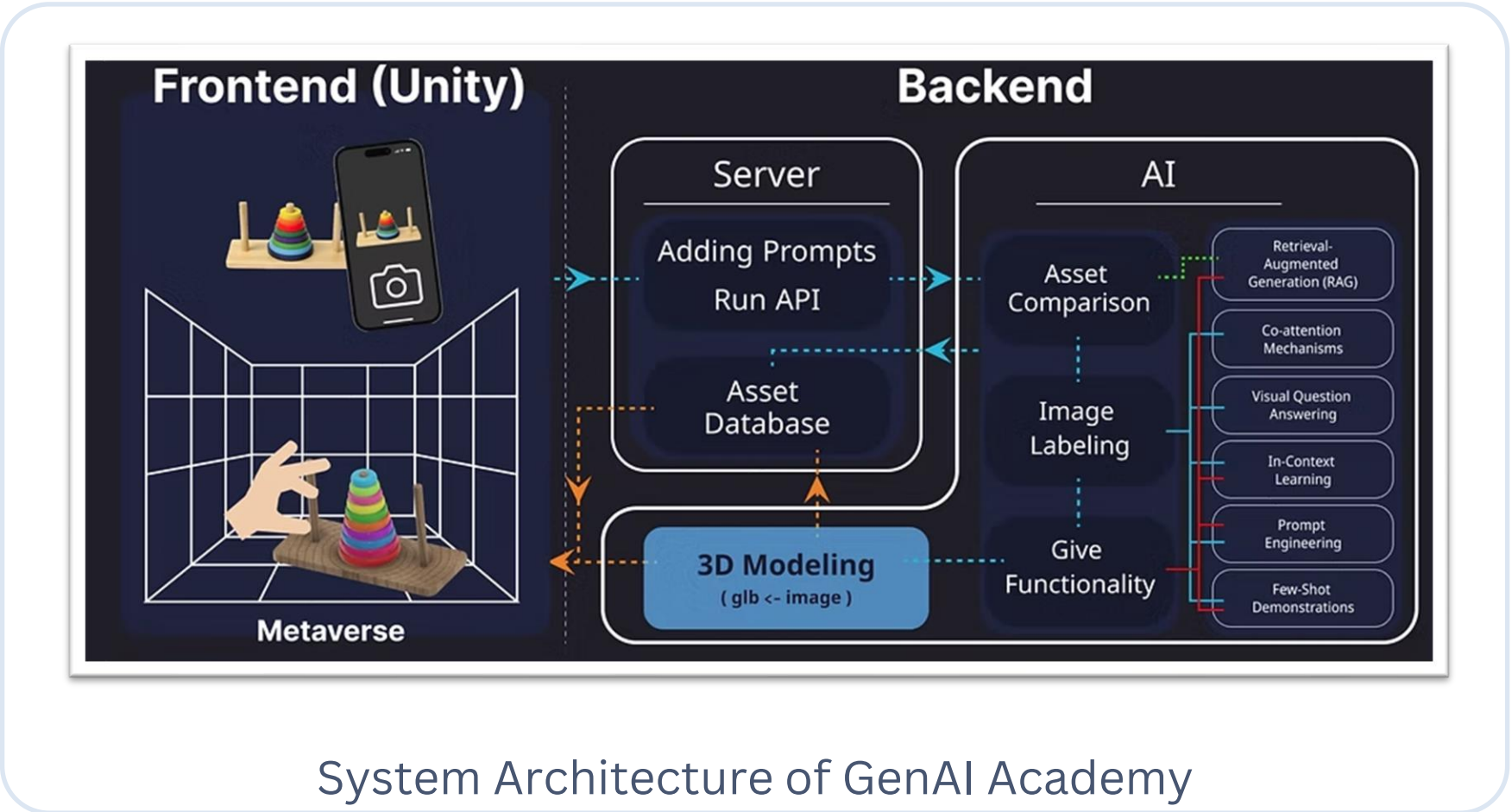
Tech Stack

- ▶ XR Development: Unity
- ▶ AI Models: GPT-based Generative Models, MeshyAI
- ▶ App Development: React Native, FastAPI

My Role (Team of 5)

- XR Developer & AI Integration Specialist (*Primary Role*)
 - Developed GenAI Academy, an XR-based educational platform integrating Generative AI to enhance interactive learning experiences.
 - Designed immersive Extended Reality (XR) environments for educational content delivery, enabling interactive and engaging learning modules.
 - Implemented Generative AI models to create adaptive learning materials, personalized quizzes, and real-time feedback mechanisms for students.
 - Collaborated with a multidisciplinary team from Korea University, including experts in AI, XR development, 3D design, and educational technology, to ensure the platform's effectiveness and user engagement.
- Awarded the Encouragement Prize at the 2024 Metaverse Developer Competition.

Contribution ★★★★★



Thank You!

심정우

Phone

010-9205-0971

Email

jshim0978@gmail.com