Yongshin Kim

+82-10-2729-4727
yongshinkim@kr.kpmg.com
https://yong-shin.github.io

Research Interest

Retrieval-Augmented Generation, Embedding Model, Large Language Model, NLP

Education

Korea Advanced Institute of Science and Technology (KAIST)
M.S. in Data Science — *Aug. 2020 – Jul. 2022*
Advisor: Uichin Lee

Handong Global University (HGU)
B.A. Mathematics and Statistics & B.A. Management (Double-Major) — *Mar. 2014 – Jul. 2020*
Summa Cum Laude

Publications

[1] Lee, P., Kim, H., **Kim, Y**., Choi, W., Zitouni, M. S., Khandoker, A., ... & Jeong, Y. (2022). Beyond Pathogen Filtration: Possibility of Smart Masks as Wearable Devices for Personal and Group Health and Safety Management. JMIR mHealth and uHealth, 10(6), e38614.

[2] SeungLin Yang, **Yongshin Kim**, Doohee Chung, (2019) *"Nonlinear Relationship Between Technological Entrepreneurship and National Competitiveness: The Moderation Effect of Innovation-driven Economy"*, Journal of Technology Innovation, 27(3), 113-142. (KCI)

[3] Haejun Jung, **Yongshin Kim**, Doohee Chung, (2019) *"The Effect of Intellectual Property-Based Startups on Employment"*, Innovation Studies, 14(4), pp119-154. (KCI)

Experience

Work Experience

Machine Learning Engineer — KPMG
Aug. 2024 – Present

Machine Learning Engineer — Okestro
Sept. 2022 – Aug.2024

Academic Experience

Interactive Computing Lab — KAIST
Advisor: Uichin Lee — *Aug. 2020 – Jul. 2022*

Technological entrepreneurship Lab — HGU
Advisor: Doohee Chung — *Aug. 2018 – Jul. 2020*

Mathematics and Statistics major — HGU
Advisor: Heonjoo Kim — *Aug. 2019 – Feb. 2020*

Business Information Technology Practice (Teaching Assistant) — HGU
Advisor: I-Soo Joe — *Mar. 2018 – Jul. 2018*

Principles of Accounting (Tutoring for foreigner) — HGU
Advisor: Hyunmo Sung — *Sept. 2017 – Dec. 2017*

Other Experience

Mentor, Vision in Calling | HGU — *Mar. 2020 – Jul. 2020*
Director, Dep. of General Affairs, Student Government | HGU — *Dec. 2019 – Jul. 2020*
Director, Technological entrepreneurship Lab | HGU — *Jul. 2019 – Feb. 2020*
Executive, Residential College of International Dormitory | HGU — *Jul. 2019 – Feb. 2020*
Manager, Handong English Camp | HGU — *Mar. 2018 – Sept. 2018*
Instructor Education Certification for foreigners | HGU — *Mar. 2018 – Jun. 2018*
Accountant, International freshmen orientation | HGU — *Mar. 2017 – Aug. 2019*
English Teacher, Global Vision Christian School | GVCS) — *May. 2017 – Sept. 2017*

PROJECTS

AI-Powered Proposal Finder — KPMG
Jan. 2026 – Apr. 2026

We developed an AI-powered service designed to address one of the most time-consuming challenges for consultants: searching for relevant past project proposals. Traditionally, identifying suitable reference proposals requires extensive manual effort and deep domain knowledge. Proposal Finder transforms this process by enabling natural language search, allowing users to describe their needs intuitively. Leveraging advanced query rewriting and tagging-based search technologies, the system refines user input into optimized search queries and matches them against a structured proposal database. This enables rapid discovery of highly relevant proposals, even when exact keywords are not provided. As a result, consultants can significantly reduce time spent on research, improve proposal quality through better references, and focus more on high-value strategic work rather than manual document retrieval.

Kbank Loan & Deposit Automation System — KPMG
Dec. 2025 – Feb. 2026

In banking operations, reviewing documents for lending and deposit decisions is complex and time-consuming. Institutions must examine corporate bylaws, terms, and authorization documents to assess eligibility and collateral. We developed an AI-powered automation system that streamlines the entire review workflow for lending and deposit operations. By simply uploading key documents such as corporate bylaws, the system automatically identifies required review items, analyzes document content, and generates structured evaluation results. Leveraging advanced document understanding and LLM-based reasoning, it reduces manual effort, accelerates decision-making, and ensures consistent, scalable review quality.

AI-Based Automated Material Analysis System — KPMG
May. 2025 – Jan. 2026

We developed an AI-powered system that automates the most labor-intensive aspects of material master registration: extracting specifications from vendor documents, standardizing attribute formats, and integrating directly into ERP. By leveraging AI-based text mining and a refined Rule Book for class-specific attribute rules, the solution automatically processes vendor prints, recommends standardized values, and detects duplicates before registration. This innovation reduced manual registration time from 10–15 minutes per material to under a minute—cutting annual workload by more than 700 hours and, for large-scale projects, over 3,300 hours. The system also prevents costly duplicate or "no-code" purchases, expected to save up to ₩1.3 billion annually, while delivering a unified, high-quality material master database.

AI-Driven Smart Information Security Disclosure Automation Platform — KPMG

Mar. 2025 – May. 2025

We have developed an AI-powered platform that automates the most time-consuming aspects of corporate information security disclosure: classifying IT and security expense ledgers and calculating costs. Leveraging AI-based keyword classification, the system automatically identifies security-related entries in accounting data and extracts the information needed for headcount cost estimation, cutting preparation time by up to 90%. What once took large enterprises weeks can now be completed within a single day, while human-error risks are dramatically reduced and disclosure data accuracy is enhanced.

Automated Ad Compliance Analyzer — KPMG

Apr. 2025 – May. 2025

Before sending advertisements to advertisers, companies must undergo preliminary review based on internal advertising regulations. This preliminary advertising review is not about expertise, but rather a task of meticulously examining content according to internal guidelines. In other words, it's a simple repetitive task. We developed a service that automates this process using Large Language Models (LLMs). The system processes advertising data through image conversion, text extraction, and analysis via LLM to evaluate compliance with regulations. Through this service, company advertising managers can now know the compliance rate of advertisements without needing to conduct preliminary reviews themselves.

Trade AI: Revolutionizing Access to U.S. trade administration — KPMG

Nov. 2024 – Mar. 2025

We developed a solution that tackles the overwhelming daily document uploads on the U.S. trade administration (https://access.trade.gov/), eliminating the need for customs experts to manually search through vast amounts of data. By implementing a RAG (Retrieval-Augmented Generation) system, we've created an automated workflow that crawls, collects, and vectorizes documents in a Milvus database, allowing for efficient cosine similarity searches. Customers can now instantly discover relevant cases and make inquiries about them through our intuitive chatbot interface. Our query-rewriting technology extracts filtering elements such as year and case codes from customer queries, enabling targeted vector searches and significantly enhancing performance for customs experts.

Empowering Security Policy Analysis with AI — KPMG

Aug. 2024 – Dec. 2024

Reviewing and improving corporate security policies and guidelines is a challenging process that often leads to inconsistent quality based on the reviewer's expertise. Our innovative tool leverages Large Language Models to analyze security policies with consistent performance and accuracy. By utilizing KPMG's extensive database of over 300 security-related resources and AI-powered web searches, we comprehensively review policies to identify missing elements and ensure alignment with the latest security trends. The system delivers rapid, precise assessments while maintaining uniform quality standards regardless of the complexity of the security framework. Additionally, we provide an intuitive chatbot service that allows users to ask questions based on the analysis results, making security policy management more efficient and effective.

Create an LLM model for RAG system — Okestro

Feb. 2024 – July. 2024

Large language model(LLM) for Korean specialized in high-performance retrieval-augmented generation is rare. We created Korean LLM specialized for RAG using Orion-14B as a foundation model. A total of 800,000 data were selected from the AI hub. This data was refined and GPT-4 was used to create 8,000 high-quality RAG training data. We used data distributed parallel(DDP) to train multiple A100 GPUs in parallel and utilized the low-rank adaptation of large language models(LoRA).

Korean Embedding Model Pipeline in Closed Network RAG System — Okestro

Jan. 2024 – July. 2024

We are working on a project to secure a high-performance Korean embedding model. We researched several baseline Korean embedding models on the huggingface. We automated the construction of a dataset for embedding model fine tuning. We are working on how to generate answers that can give trust to us we were able to have a Korean embedding model with significantly better performance than the text-ada-embedding-002 provided by OpenAI.

Deploying chatbot service specific to cloud domain
Okestro
June. 2023 – July. 2024

We are developing Chatbot specific to cloud domains using Retrieval-Augmented Generation (RAG) systems. Through this, we can primarily handle product-related inquiries from in-house customers. We are working on how to generate answers that can give trust to users in the RAG system.

FAQ Classification
Okestro
Mar. 2023 – June. 2023

We present a model that uses BERT and contrastive learning to automatically classify users' inquiries. The model eliminates the need for users to have domain knowledge or for administrators to classify multiple inquiries. This can significantly improve the efficiency of the inquiry management system, reduce the workload of users and administrators and enhance the overall user experience.

Log-Level Anomaly Detection
Okestro
Mar. 2023 – June. 2023

We present FineLog, a log message-wise anomaly detection framework that enables anomaly detection for each log message in the context of a given sequence. This study shows that FineLog not only records high performance in the existing sequence unit anomaly detection, but also records high performance in log unit anomaly detection.

Baseline of SWOT Classification
Okestro
Sept. 2022 – Dec. 2022

We present baseline indicators by introducing BERT model, which is widely used in the natural language processing field, for the first time in SWOT analysis. Starting with this approach, the baseline indicators of this study are expected to be useful for business intelligence cloud platforms that can be easily accessed by all stakeholders through deep learning of SWOT analysis.

Automation of Company SWOT Analysis Using Sentence BERT
Okestro
Sept. 2022 – Dec. 2022

This study presented SWOT Sentence BERT as an AIaaS model that can intellectually automate company SWOT analysis. The SWOT Sentence BERT is a sentence embedding model that is learned through SWOT text data processed in the form of natural language inference task. In order to automate SWOT analysis, we applied K-Means clustering algorithm to make clusters with sentence embeddings and classified sentence embeddings based on their predicted clusters.

Emotion Recognition (Master Thesis)
KAIST
Oct. 2021 – July. 2022

In this study, we propose a method for predicting the emotions of the speaker in the naturalistic conversation using a speaker encoder and counterpart encoder composed of CNN-LSTM deep learning networks. We used emotion-related data called K-EmoCon collected during the debate process to empirically evaluate our model. The results showed that the counterpart's speech and the physiological signals had a positive impact on predicting the speaker's emotions.

Video & Speech synthesis
Deepbrain AI
Sept. 2021 – Oct. 2021

We produced AI Human through voice conversion using StarGAN-VC and image synthesis using FSGAN. Through the process of analyzing various voice conversion models(CycleGAN, StarGAN-VC, StarGAN-VC2), we improved the quality of voice conversion by analyzing loss terms, number of domain classes, batch size, and iteration. In addition, the average similarity between the source video and the target video was used to facilitate video synthesis.

Digital Therapeutics (DTx)
KAIST
Aug. 2020 – Sept. 2021

We develop fundamental technologies of data-driven digital therapeutic, receptivity optimization for mobile digital therapeutic development. Furthermore, we analyze the effectiveness of digital treatments by applying causal analysis.

AWARDS AND HONORS

Best Paper Award (Korean Institute of Information Technology Paper Competition)	*2022*
President Award (Top 10 World-changing Projects) | HGU	*2020*
Best Paper Award (National Technology Policy Paper Competition) | IITP & KOTIS	*2019*
Best Paper Award (Korea Society for Innovation Management & Economics Paper Competition)	*2019*
Academic Top Scholarship for Seniors | HGU	*2019*
Academic Excellence Scholarship for Juniors | HGU	*2018*
Academic Excellence Scholarship for Sophomores | HGU	*2017*
Academic Excellence Scholarship for freshmen | HGU	*2016*

CONFERENCES

[1] **Yongshin Kim**, Taehee Lee, Sanghyeon Jung, Chanjae Lee, Taewan Kwon, "*Improving Cloud FAQ Experience through Contrastive Learning-based Inquiry Classification*", Korea Computer Congress, Jeju Island, Korea (2023)

[2] **Yongshin Kim**, Sanghyeon Jung, Chanjae Lee, Jinhee Kim, "*Baseline of SWOT Classification using Bidirectional Encoder Representations from Transformers for Business Intelligence Cloud Platform*", Korean Institute of Information Technology, 2022 Fall Conference, Jeju Island, Korea (2022) – Best Paper Award

[3] **Yongshin Kim**, Panyu Zhang, Gyuwon Jung, Heepyung Kim, Uichin Lee, "*Causal Analysis of Observational Mobile Sensor Data: A Comparative Study*", Korea Computer Congress, 2021 Spring Conference, Jeju Island, Korea (2021)

[4] Haejun Jung, **Yongshin Kim**, Doohee Chung, "*The Effect of Intellectual Property-Based Startups on Employment*", Korea Technology Innovation Society, 2019 Fall Conference, Jeju Island, Korea (2019) – Best Paper Award

[5] SeungLin Yang, **Yongshin Kim**, Doohee Chung, "*Nonlinear Relationship Between Technological Entrepreneurship and National Competitiveness: The Moderation Effect of Innovation-driven Economy*", Korea Society for Innovation Management & Economics, 2019 Spring Conference, Daejeon, Korea (2019) – Best Paper Award

PATENTS

[1] Jung. S, Kwon. T, Lee. T, **Kim. Y**, Lee. C, Kim. J, "A CREATION MODULE FOR AUTOMATIC SWOT ANALYSIS TOOL USING ARTIFICIAL INTELLIGENCE AND A SWOT ANALYSIS SYSTEM COMPRISING THE SAME"
KR - Application No.10-2022-0179834

[2] Jeong. K, **Kim. Y**, Ahn. S, Lee. T, Kim. Y, Kim. M, "A CLOUD SERVER OPERATING SYSTEM IMPLEMENTING INDIVIDUAL VIRTUALIZATION OF RESOURCES AND A METHOD FOR OPERATING CLOUD SERVER"
KR - Application No.10-2022-0190619

[3] **Kim. Y**, Lee. T, Jung. S, “AN INQUIRY MANAGEMENT SYSTEM USING CLASSIFICATION METHOD BASED IN CLOUD SERVICE AND A PATFORM FOR INQUIRY-RESPONSE INTEGRATED MANAGEMENT”

KR - Application No.10-2023-0077742

COMPUTER SKILLS

Python (Advanced), R (Intermediate), SPSS (Intermediate), MS Office (Intermediate), SQL (Basic), STATA (Basic), AMOS (Basic), SAS (Basic)

Updated April 1, 2026