

Vol. 1 · 2026

Proceedings of the Conference on PyGeek

2026 International Conference on PyGeek (PyGeek 2026)

PyGeek

Proceedings of the Conference on PyGeek

**2026 International Conference on PyGeek
(PyGeek 2026)**

Volume 1 · 2026

Published by Marketian

Publication and Copyright Information

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

Official publication title	Proceedings of the Conference on PyGeek
Conference	2026 International Conference on PyGeek (PyGeek 2026)
Volume / year	Vol. 1 · 2026
Publication date	2026-09-01
Publisher	Marketian
Official postal address	3F, A-02, Changdae Building, 224, Baekjegobun-ro, Songpa-gu, Seoul, 05583, Rep. of KOREA / 서울특별시 송파구 백제고분로 224 창대빌딩 3층 에이-02호 (삼전동)
Proceedings website	https://pygeek.acin.kr/proceedings/2026
Online ISSN	3140-5312
Stable URL	https://pygeek.acin.kr/proceedings/2026
Conference contact	pygeek.admin@gmail.com

Copyright and publication terms

Copyright in each paper remains with its authors. Marketian publishes this proceedings volume and the submitted camera-ready versions under the authors' recorded publication consent. No proceedings-wide Creative Commons license is asserted unless explicitly stated for an individual work.

Message from the General Chair

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

Welcome to PyGeek 2026.

It is my great pleasure to welcome researchers, industry professionals, developers, students, authors, and speakers to the 2026 International Conference on PyGeek. Held from August 19 to 21, 2026, at Hongik University in Seoul, Republic of Korea, this conference provides a forum for sharing research, practical experience, and new ideas across the rapidly evolving Python and artificial intelligence ecosystem.

This year's conference explores the theme of smart manufacturing and artificial intelligence, with particular attention to the development of a skilled workforce and emerging technology trends. The program also highlights Large Language Models, Multimodal Computing, and Agent-based AI. These areas are reshaping how intelligent systems are designed, evaluated, and deployed, while creating new opportunities for collaboration among universities, industry, and public institutions.

PyGeek 2026 brings together perspectives from academic research and real-world practice. Across three days, the program includes keynote speeches, technical sessions, research paper presentations, seminars, workshops, and opportunities for professional networking. Together, these activities encourage participants to examine both foundational advances and practical applications, including intelligent manufacturing, digital transformation, efficient AI systems, scientific computing, developer tools, and responsible deployment.

The papers included in these proceedings represent the authors' contributions to the PyGeek 2026 research program. They reflect a range of approaches to current challenges in artificial intelligence, computing, engineering, and interdisciplinary applications. We hope that this volume will serve not only as a record of the conference, but also as a useful resource for continued research, education, technical development, and collaboration.

I extend my sincere appreciation to all authors and presenters for sharing their work and experience. I also thank the keynote speakers, Technical Program Committee, organizing committee, session contributors, and everyone who contributed to the review and preparation process. Their time, expertise, and careful efforts have made both the conference and these proceedings possible.

We are also grateful to the organizations jointly hosting and organizing PyGeek 2026: DeepShark Research Lab, the Industry-Academic Convergence Talent Center at Hongik University, and Henan University of Engineering. I would also like to recognize our sponsors and partners for their valued support. Their contributions help create an environment in which researchers, practitioners, and students can exchange ideas and build meaningful connections. Sponsorship does not imply influence over the paper review, acceptance, or publication process.

The publication of this first volume of the Proceedings of the International Conference on PyGeek marks an important step in preserving and sharing the conference's scholarly and technical contributions. I hope these proceedings inspire new questions, productive partnerships, and practical innovations beyond the conference itself.

On behalf of the organizing committee, I warmly welcome you to PyGeek 2026 and wish you an engaging and rewarding conference.

Giseop Noh General Chair Department of Software and Communications Engineering, Hongik University, Republic of Korea

Giseop Noh
General Chair
Hongik University

2026 PYGEEK CONFERENCE

Sponsors and Partners

We gratefully acknowledge the organizations whose support contributed to the successful delivery of this conference.

PyGeek Conference gratefully acknowledges the organizations whose support contributed to the successful delivery of this conference. Sponsorship does not imply influence over the paper review, acceptance, or publication process.

PLATINUM SPONSOR



Ytron

"We Sense. We Connect. We Complete Safety." Ytron protects industrial workplaces through high-precision AI sensing and highly reliable communication solutions, while creating new business value through intelligent, data-driven connectivity.

와이트론

‘센싱하고 연결하여, 안전을 완성합니다.’ 와이트론은 고정밀 AI 센싱과 고신뢰 통신 솔루션을 통해 산업 현장의 안전을 지키고, 데이터 기반의 지능형 연결로 비즈니스의 새로운 가치를 창출합니다.

<https://www.ytron.co.kr/>

PLATINUM SPONSOR



Korea Industrial Development Association (KOIDA)

The **Korea Industrial Development Association (KOIDA)** promotes innovation and competitiveness across manufacturing and service industries. By fostering collaboration among industry, academia, research institutes, and government, KOIDA supports industrial growth and Korea's economic development.

(사)한국산업진흥협회

사단법인 한국산업진흥협회는 제조업의 성장동력 확충 및 경쟁력 강화와 서비스업의 차별화된 수익창출의 장마련을 위한 활동을 선도하고, 지식기반산업, 글로벌산업 등 신성장 산업과 유·무형의 서비스 산업의 동반성장 및 융·복합을 도모하며, 유관 산학연관 간 정보교환 및 상호협조를 유도함으로써 국민과 기업의 경쟁력 제고와 국가경제발전에 이바지 하고자 2016년에 설립된 기관입니다.

<http://www.koida.or.kr/>

PLATINUM SPONSOR



Sejong Regional Institute of Industrial Advancement

The Sejong Regional Industry Promotion Agency (RIIA) drives regional industrial growth by connecting national policies with Sejong's strategic industries. RIIA supports strategic planning, R&D programs, businesses, and workforce development to foster innovation and sustainable growth in the region.

(재)세종지역산업진흥원

(재)세종지역산업진흥원은 정부 산업정책을 세종의 성장기회로 연결하는 지역산업 육성 전문기관입니다. 5극 3특 성장엔진 프로젝트, 지역혁신클러스터육성사업 등 지역산업 전략기획부터 정부·부처 R&D 사업의 평가·관리, 기업지원, 인력양성까지 지역산업 육성의 전 과정을 지원하며, 세종의 전략산업 육성과 지역기업의 혁신성장을 함께 만들어가고 있습니다.

<https://sj.riia.or.kr/>

PLATINUM SPONSOR

신일특허사무소 (SHINIL PATENT & LAW FIRM)

Shinil Patent & Law Firm

Shinil Patent & Law Firm, located in Gojan-dong, Ansan, provides comprehensive intellectual property services for patents, trademarks, and designs. From application to registration, we are committed to delivering tailored solutions that fully meet our clients' needs.

신일특허법인

신일특허사무소는 안산시 고잔동에 위치하고 있으며, 특허, 상표 및 디자인에 대하여 출원부터 등록까지 고객의 니즈를 100% 충족시켜 드립니다.

<http://www.shinillaw.co.kr>

PLATINUM SPONSOR



Learning Clover

Learning Clover operates Learncler, a SaaS platform that automates education and event administration. It integrates registration, attendance, assignments, evaluation, certification, and performance management, helping organizations reduce workload and improve data accuracy.

(주)러닝클로버

러닝클로버는 교육·행사 운영의 복잡한 행정 업무를 자동화하는 SaaS 플랫폼 '런클'을 운영합니다. 참가자 모집부터 출결·과제·심사·수료·성과관리까지 전 과정을 하나의 시스템으로 통합합니다. 공공기관·대학·기업의 업무 부담은 줄이고, 운영 데이터의 정확성과 성과 활용도는 높입니다.

<https://learncler.com/>

PLATINUM SPONSOR



newdic

NEWDIC Co., Ltd. (NEW Daily Intensive Care) develops mobility solutions for people with walking difficulties, including walking aids and smart insoles for gait analysis. Through innovative R&D, NEWDIC aims to provide people of all ages with greater mobility and the joy of walking.

주식회사 뉴딕

(주)뉴딕은 NEW Daily Intensive Care의 약어로, 당사는 보행약자를 지원하기 위한 보행보조기 및 보행분석을 위한 스마트 인솔 연구개발과제를 수행하고 있습니다. '26년 중기부 창업성장기술개발 디딤돌 과제 및 사회적기업인증원의 예비사회적기업 육성지원사업에 선정, 남녀노소 누구에게나 이동의 자유와 보행의 즐거움을 제공하고 있습니다

<https://www.newdic.com/>

Conference Overview

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

Main theme	Smart Manufacturing and AI: Training Skilled Workforce and Technology Trends
Format	on_site
Official language	English
Venue	Conference Hall, 2nd Floor, WoW Building, Hongik University
Address	Conference Hall, 2nd Floor, WoW Building, Hongik University 94 Wausan-ro, Mapo-gu, Seoul 04066, Republic of Korea
Hosts	DeepShark Research Lab; Industry-Academic Convergence Talent Center (ICTC), Hongik University; Henan University of Engineering, China

PyGeek 2026 is an international conference that brings together researchers, educators, developers, students, and industry professionals working across Python’s rapidly expanding ecosystem. The conference provides a forum for sharing original research, practical experience, emerging technologies, and interdisciplinary applications in artificial intelligence, computing, software, science, engineering, and industry.

Under the theme **“Smart Manufacturing and AI: Training Skilled Workforce and Technology Trends,”** PyGeek 2026 explores how artificial intelligence and advanced computing can contribute to industrial innovation, digital transformation, and the development of a future-ready workforce. Particular attention is given to Large Language Models, multimodal computing, agent-based AI, efficient AI systems, and the deployment of intelligent technologies in real-world environments.

The three-day program combines peer-reviewed research paper presentations with keynote speeches, technical sessions, industry seminars, workshops, and professional networking. By connecting academic research with industry practice, PyGeek 2026 encourages the exchange of ideas, international collaboration, and the responsible development of computing technologies with practical and societal value.

PyGeek 2026 is held from August 19 to 21, 2026, at Hongik University in Seoul, Republic of Korea.

Scope

The scope of PyGeek 2026 covers foundational research, system design, software tools, evaluation methods, education, and real-world applications related to artificial intelligence and modern computing.

The conference welcomes contributions concerning foundation models, Large Language Models, multimodal models, agent-based and agentic AI, human-computer interaction, generative AI, efficient AI infrastructure, scientific computing, data science, developer tools, and responsible AI. It also encourages interdisciplinary work applying computing and AI to smart manufacturing, healthcare, finance, robotics, communications, education, and other industrial or societal domains.

PyGeek seeks to bridge theory and practice by providing opportunities for researchers and practitioners to present validated results, discuss implementation experience, identify emerging challenges, and establish collaborative relationships across academia and industry.

Topics

- Foundation Models, Large Language Models, and Multimodal Models
- Agent-based AI, Agentic AI, and Human-Computer Interaction
- Generative AI and Creative Applications
- Efficient AI Systems and Infrastructure
- AI for Smart Manufacturing and Industrial Digital Transformation
- AI Applications in Healthcare, Finance, Robotics, and Communications
- Evaluation, Ethics, Safety, and Responsible AI
- Python for Scientific Computing and Data Science
- Developer Tools and Frameworks for AI
- AI Education and Skilled Workforce Development
- Edge Intelligence, IoT, and Autonomous Networks
- Interdisciplinary Computing Applications

Conference Organization

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

General Chair

Giseop Noh	Dept. of SW & Communications Engineering, Hongik University	Republic of Korea
-------------------	---	-------------------

Organizing Committee

Junjun Zhang	School of Software, Henan University of Engineering	China
Sungwoong Hong	Dept. of AI & SW, Cheongju University	Republic of Korea

Technical Program Committee Chair

Chaeyun Seo	Dept. of SW & Communications Engineering, Hongik University	Republic of Korea
--------------------	---	-------------------

Conference Program

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

Day 1 — 2026-08-19

Foundations & Systems · Asia/Seoul

Time	Session / Topic	Presenter	Location
13:30– 15:00	Workshop Workshop on LLM fundamentals and advanced Python for AI.	Pre-registered workshop panel only.	—
15:00– 15:30	Coffee Break Have a fun! And Socials	N/A	—
15:30– 17:00	Industry & Academic Joint Seminar Industry & Academic Joint Seminar for Collaborative Team Building and Create Things to Go	Only pre-registered industry, society VIP members.	—
17:00– 18:00	Administrative Meeting Steering Committee Meeting for Final Review for Conference States	Steering Committee, and TPC members	—
18:00– 20:00	Dinner Networking and Social Event near Conference Venue (TBA)	Those wishing to attend	—

Day 2 — 2026-08-20

Multimodality & Agents · Asia/Seoul

Time	Session / Topic	Presenter	Location
09:50– 10:00	Opening Remarks Welcome & Hello Message, Daily Orientation	Giseop Noh, General Chair	—
10:00– 10:20	Keynote 1 From Crude Oil to Intelligence	Medical Dr. Sangwon Ko	—
10:20– 10:40	Keynote 2 AI Agentification over Wireless Networking: The Future of Autonomous Networks Driven by 6G and Decentralized Edge Intelligence	Prof. Byung-Seo Kim	—
10:40– 11:00	Technical Session #1 BOARS: Development of an AI-Based Real-Time Wild Boar Integrated Control System	Prof. Hyunjun Park (Korea Maritime & Ocean University)	—
11:00– 11:20	Technical Session #1 Changes in Engineers' Job Competencies in the AI Era	Prof. Heungseob Kim (Changwon National University)	—
11:20– 11:40	Technical Session #1 Formalizing Energy-Safe Inference Scheduling for Physical AI under Non-stationary Energy Harvesting	Prof. Young-myung Kang (Sungkyul University)	—
11:40– 12:00	Oral Presentation #2 A Co-evolutionary Strategy of AI Technology for Expanding Regional Industrial Value Chains A Co-evolutionary Strategy of AI Technology for Expanding Regional Industrial Value Chains — Eujung Kim	PhD. Candidate, Eujung Kim (Industry-academic Convergence Talent Center, Hongik University)	—
12:00– 13:30	Lunch Break Eat first. Solve AI later. 😊 And lunch is on... you!	Mission: Find food. Objective: Return before 1:30 PM	—
13:30– 13:50	Technical Session #2 Research on Mission-Driven Intelligent AI Agents: From Defense Mobility to Sports Science	Prof. Youngsoo Kim (Jeonju University)	—
13:50– 14:10	Technical Session #2 Development and Application of AI Solutions in the Semiconductor Industry	Mr. Seungwon Seo (GSFSolution)	—
14:10– 14:30	Oral Presentation #1 A Factor Decomposition Method for Separating Store Substitution and Embedding Placement Effects in Vector Database Migration A Factor Decomposition Method for Separating Store Substitution and Embedding Placement Effects in Vector Database Migration — Junha Yoon	Prof. Sangsoon Lim et. al. (Chung-Ang University)	—

Time	Session / Topic	Presenter	Location
14:30– 14:50	Oral Presentation #1 Multi-scale Temporal-aware Text-Enhanced Network for Multimodal Sentiment Analysis Multi-scale Temporal-aware Text-Enhanced Network for Multimodal Sentiment Analysis — Junjun Zhang	Prof. Junjun Zhang (Henan University of Engineering)	—
14:50– 15:20	Coffee Break Have a fun & enjoy networking!	-	—
15:20– 15:40	Oral Presentation Distinguishing Professional and Amateur Piano Performances using 2D Audio Features and Convolutional Neural Networks Distinguishing Professional and Amateur Piano Performances using 2D Audio Features and Convolutional Neural Networks — Changheun Oh	Prof. Changheun Oh (Hongik University)	—
15:40– 16:00	Technical Session #1 Are AI Agents Tools or Agents? A Shift in Human-Machine Relationships	Prof. Hayoung Oh (Sungkyunkwan University)	—
16:00– 16:20	Oral Presentation #2 Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting — Junggyu Park	Mr. Junggyu Park (Hongik University)	—
16:20– 16:40	Oral Presentation #2 Semantic Cohesion Analysis of the Chunk of Code with an LLM Semantic Cohesion Analysis of the Chunk of Code with an LLM — Eunjin Jeong	Ms. Eunjin Jeong (Hongik University)	—
16:40– 17:00	Oral Presentation A Self-RAG-Based Triple Factuality Verification Pipeline for Local LLM KARMA A Self-RAG-Based Triple Factuality Verification Pipeline for Local LLM KARMA — Seung-gyeom Kim	Mr. Seung-gyeom Kim (Hongik University)	—
17:00– 17:20	Oral Presentation Automatic Generation of Emotion Retrospective Content Based on Generative AI Automatic Generation of Emotion Retrospective Content Based on Generative AI — Jun Sun Chang	Ms. Jun Sun Chang (Hongik University)	—
17:20– 17:30	Daily Closing Remarks Briefly introduce tomorrow's schedule and provide an overview of the evening social networking banquet.	Giseop Noh, General Chair	—
18:00– 21:00	Conference banquet No slides, no equations—just conversations, food, and new connections.	If you're free, hungry, or just curious—come join us! ★	—

Day 3 — 2026-08-21

Applications & Future Directions · Asia/Seoul

Time	Session / Topic	Presenter	Location
10:30–12:00	Academy workshop The Business of AI: From Research to Revenue.	Academic Leaders & Students	—
12:30–13:30	Lunch Break -	-	—
13:30–14:00	Closing Remarks Summary and Future of PyGeek.	Prof. Giseop Noh, General Chair	—
15:00–16:00	Steering Committee Meeting Quick Review Board	TBD	—

Table of Contents

Message from the General Chair	ii
Sponsors and Partners	iv
Conference Overview	viii
Conference Organization	x
Conference Program	xi
Day 1 — 2026-08-19	xi
Day 2 — 2026-08-20	xii
Day 3 — 2026-08-21	xiv
Keynote Speeches	1
From Crude Oil to Intelligence	
Sangwon Ko	2
AI Agentification over Wireless Networking: The Future of Autonomous Networks Driven by 6G and Decentralized Edge Intelligence	
Byung-Seo Kim	22
Technical Sessions	23
BOARS: Development of an AI-Based Real-Time Wild Boar Integrated Control System	
Hyunjun Park	24
Research on Mission-Driven Intelligent AI Agents: From Defense Mobility to Sports Science	
Youngsoo Kim	30
Changes in Engineers' Job Competencies in the AI Era	
Heungseob Kim	42
Formalizing Energy-Safe Inference Scheduling for Physical AI under Non-stationary Energy Harvesting	
Young-myyoung Kang	53
Development and Application of AI Solutions in the Semiconductor Industry	
Seungwon Seo	54
Accepted Papers	67
A Co-evolutionary Strategy of AI Technology for Expanding Regional Industrial Value Chains	
KIM EU JUNG	68–72

A Factor Decomposition Method for Separating Store Substitution and Embedding Placement Effects in Vector Database Migration	
Junha Yoon, Ohhyeon Gwon, Junhaeng Lee, Hyeji Roh, Soowan Cho, Suhee Kim, Sangsoon Lim	73–77
Multi-scale Temporal-aware Text-Enhanced Network for Multimodal Sentiment Analysis	
Junjun Zhang, Giseop Noh	78–83
Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting	
junggyu, Park, Giseop Noh	84–89
Semantic Cohesion Analysis of the Chunk of Code with an LLM	
Eunjin Jeong, Janghwan Kim, Chaeyun Seo, R.Young Chul Kim	90–93
Distinguishing Professional and Amateur Piano Performances using 2D Audio Features and Convolutional Neural Networks	
Changheun Oh	94–99
A Self-RAG-Based Triple Factuality Verification Pipeline for Local LLM KARMA	
Seung-gyeom Kim, Giseop Noh	100–106
Automatic Generation of Emotion Retrospective Content Based on Generative AI	
JunSun Chang, Janghwan Kim, R. Young Chul Kim	107–109
Award Recipients	110

Keynote Speeches

From Crude Oil to Intelligence

Sangwon Ko

Marketian · Republic of Korea

Experience

- CEO, Marketian (Current)
- Adjunct Professor, Graduate School of Medicine, CHA University (Current)
- Former Chief Executive of Management Division, Kwangdong Hospital
- Former Vice President, K-DOC
- Former GL&S Korea Manager, Pfizer Korea
- Former Senior Consultant, Gallup Korea Consulting
- Former Team Leader, Healthcare Information Systems Development Team, Hallym University Medical Center



Academic Records

- M.Ph.D. in Medicine, CHA University
- M.M.S. in Marketing, University of Colorado Denver
- M.B.A. in Accounting, Dongguk University

Abstract

This keynote examines the fundamental challenges and opportunities facing computer science education and software professionals amid the rapid paradigm shift from the analog era through digital transformation to the emerging Intelligence Economy. Drawing parallels with the rise and decline of past industries and academic disciplines, the talk explores how the widespread adoption of Generative AI has significantly reduced the scarcity value of programming syntax itself. As the world transitions from the 20th-century paradigm of oil refining to the 21st-century era of “intelligence refining” powered by computing resources, developers must evolve beyond the role of coders to become architects of systems and definers of problems. The lecture highlights how value creation is increasingly driven not by writing code, but by identifying meaningful problems, designing intelligent systems, and orchestrating AI capabilities. Through practical industry cases and real-world examples, the speaker will present five actionable principles that professors of artificial intelligence, computer science students, and industry professionals can immediately adopt to secure competitive advantages in the age of AI-driven transformation.

KEYNOTE

원유에서 **지능(Intelligence)**으로

파이썬학회 기초강연

인공지능학과 교수 · 연구진 · 개발자 · 산업계를 위한 이야기

ChosunBiz

이코노미조선 | English · 日本語

재테크 증권 부동산 IT 금융 산업 중소기업·벤처 바이오 유통 정책 정치 사회 국제 사이언스조선 문화 오피니언

산업 > 업계 소식

인바이츠 생태계, K-디지털헬스케어 40개사 얼라이언스 출범식 성료

정민기 기자

입력 2026.03.18. 09:25



국내 디지털헬스케어 기업 40개사가 참여하는 'K-디지털헬스케어 얼라이언스'가 공식 출범했다.

디지털헬스케어 분야 기업들은 지난 17일 서울 여의도 글래드호텔에서 공동 협약(MoU)

Lenovo

레노버 ThinkPad X1 Carbon Gen 13

3,484,050원 **33%**

레노버 ThinkPad X1 Carbon Gen 13

3,784,838원 **36%**

■ 4 strengths of Marketian Healthcare: IT 개발능력

① 디지털 치료분야 개발사례



→ 인공지능 기술 기반으로 시각장애인을 위한 스마트 침대, 뇌병변 장애인을 위한 상지재활운동기기 그리고 운동기능검사를 통한 경도인지장애 판별 도구 등을 개발하였습니다.

*당사 자체 리서치 운영 프로그램



② 기타개발분야 사례



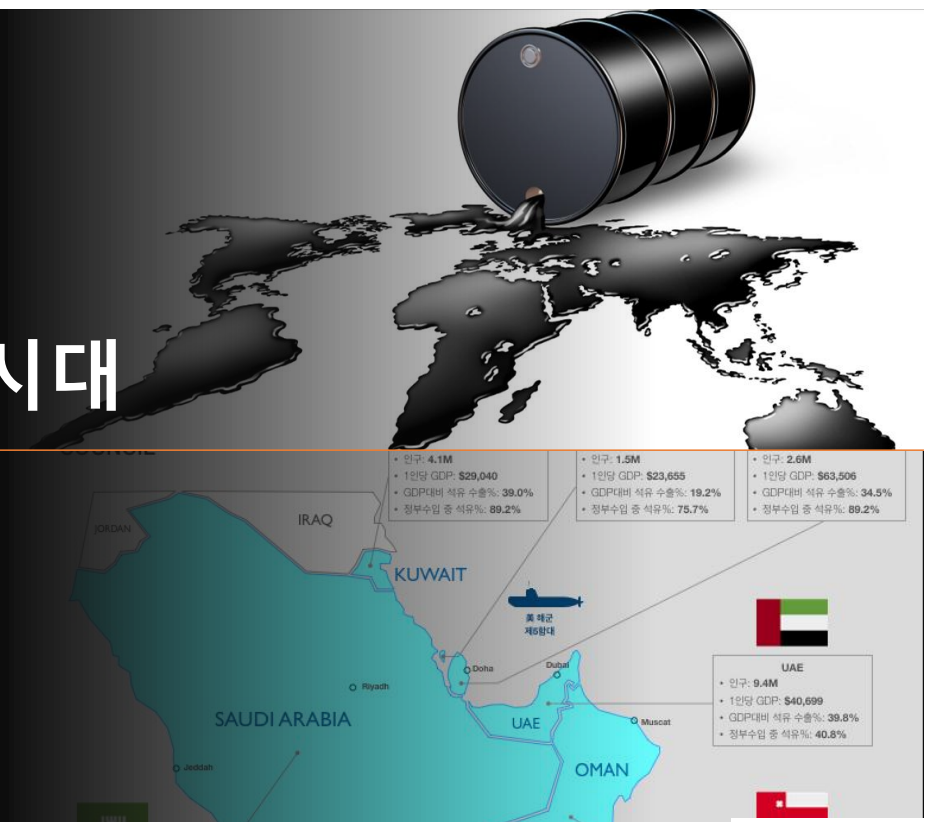
→ 제약회사 업무를 보조하기 위한 ERP시스템, 수요예측 시스템, 사내 지식관리 시스템, 외부 행사 진행에 도움을 주는 전자방명록 등을 개발하였습니다.

세상이 어떻게 변해가고 있고,
그 안에서 우리는 어떤 인사이트를 필요로 하는지에 대해
이야기 해 볼까 합니다.

지능 경제(Intelligence Economy)의 서막

원유경제의 시대

20세기를 간단히 일컬어





현대 경제는 원유를 통해 발전해 왔다.

인류 현대사의 발전과 부의 공식(20세기까지)

원유 → 정제 → 휘발유/등유 → 제품/서비스

그렇다면 21세기는?

원유의 자리를
컴퓨팅 파워가 대체하게 될 것이다.

부의 공식

인류 현대사의 발전과 부의 공식(20세기까지)

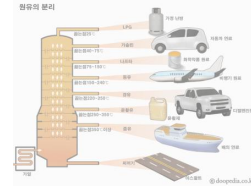
원유 → 정제 → 휘발류/등유

21세기의 부의 공식(?)

컴퓨팅파워 → AI팩토리에서 정제 → 지능

*데이터 입력하고 GPU 돌려서 학습시키는 과정

원유자체는 아무 쓸모가 없음



컴퓨팅파워? 그냥 전기먹는 하드웨어

→ 모델+알고리즘으로 ‘정제’ 해야 지능으로 발전

시사점

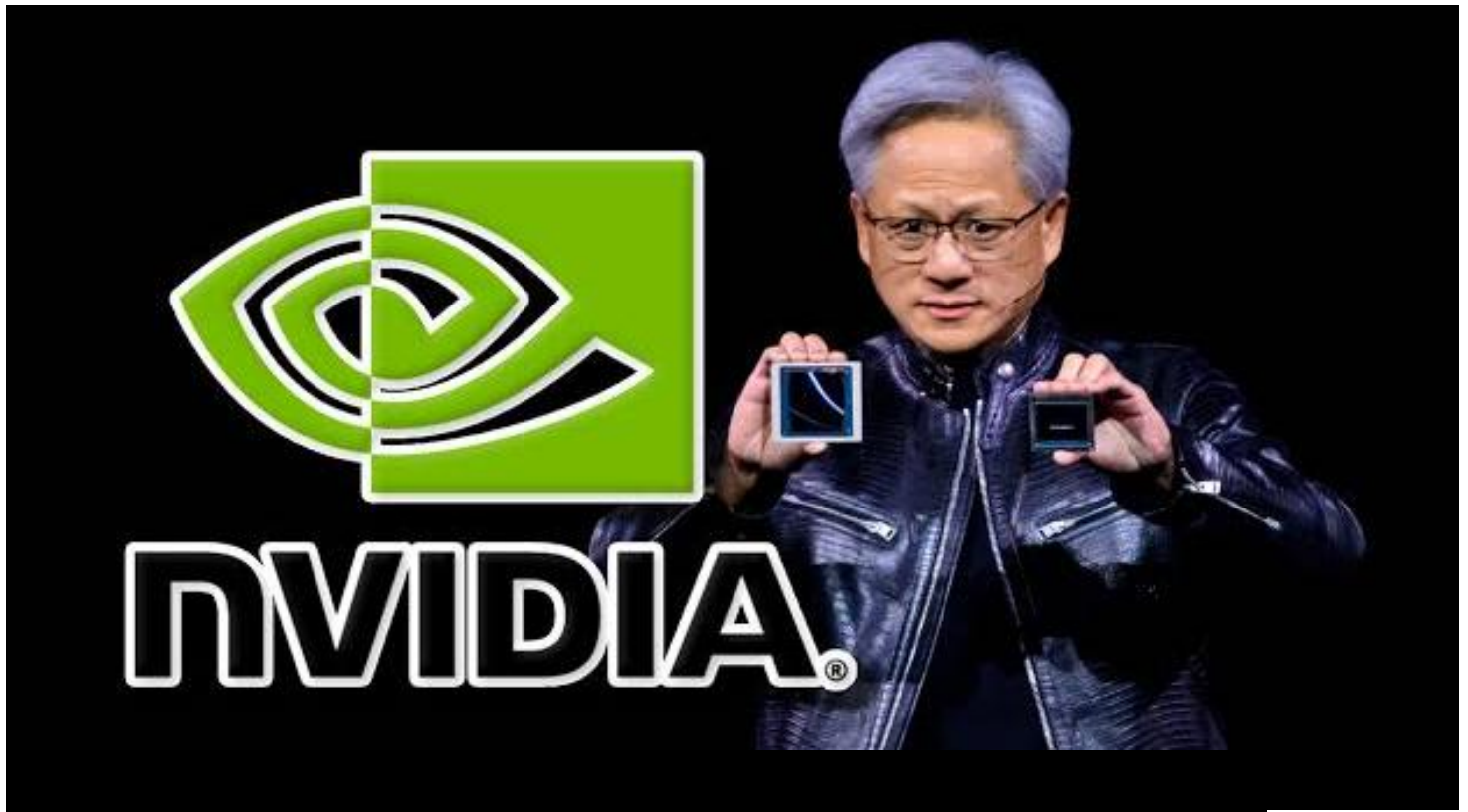
1. 과거의 방식은 통하지 않음

2. 현재의 방식을 만능으로 착각하면 안됨

컴퓨터 사이언스가 만능키가 될 수 없고, 우리의 연구와 분야가 성공을 보장하지 않는다

2026년 상반기

세계경제 시가총액 1위 기업은?



1990년대

세계경제 시가총액 1위 기업은?

석유에너지 개발기업

ExxonMobil

석유와 천연가스 유전의 개발 및 생산을 기반으로 세계 에너지 산업의 큰 형님 대접을 받는 거대 기업이며, 세계 정유 업계의 슈퍼 메이저로 유명한 대기업으로서 그 명성이 널리 알려져 있다. 텍사스 주 휴스턴에 본사가 자리 잡고 있다. 세계 각지에 대규모의 유전을 보유하고 있어 글로벌 경제 전반에 막대한 영향력을 행사하고 있다.

20세기 · 원유의 시대

부(富)는 땅 속에 있었다

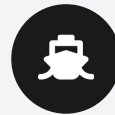
대한민국도 원유 경제의 핵심 산업으로
지금의 부의 근간을 이루었음



ExxonMobil

1990년대 세계 시총 1위

원유를 땅에서 캐내 정제하는 것만으로
세계 최고 부(富)를 만들었다.
자원을 가진 자, 정제하는 자가 이겼다.



포항제철 · 현대중공업

한국 고도성장의 상징

쇠물을 녹이고 배를 만들던 굴뚝 산업이
대한민국의 부를 일으켰다.
땅을 파고, 굴뚝 세우고, 배 만든 사람이
세상을 지배했다.

패러다임 : 물리적 자원을 가공하는 자가 부를 독점한다

20세기 · 원유의 시대 · 한국편

국제시장

덕수처럼 — 몸을 팔고, 땅을 파고, 직접 팔았다

01



독일 탄광

파독 광부

지하 1,000m
석탄을 캐다.
몸이 곧 자본이었다.

02



베트남 전쟁

파월 기술자

목숨 걸고
달러를 벌었다.
외화가 곧 국력이었다.

03



울산 조선소

중화학 공업

쇠물을 녹이고
배를 만들었다.
굴뚝이 곧 성장이었다.

04



부산 국제시장

장사·무역

새벽부터 물건을
직접 팔았다.
발품이 곧 생계였다.

패러다임 : 부(富)는 몸과 땅에서 나왔다 — 자원을 가진 자, 직접 일한 자가 이겼다

하지만

지능경제 시대는 조금 다릅니다

21세기 · AI 팩토리 (모바일 플랫폼) 시대

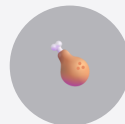
플랫폼을 가진 소수가 생태계 전체를 장악했다



카카오 · Apple

플랫폼 제국의 지배자

카카오는 수수료·광고로 생태계를 독점했고,
Apple은 앱스토어 30%를 가져갔다.
플랫폼을 가진 소수가 나머지 전체를 지배했다.



치킨집 사장님 · 일반 개발자

플랫폼에 종속된 하청 구조

열심히 튀겨도 수수료·광고비로
이익은 플랫폼이 가져갔다.
개발자도 마찬가지로 — 밤새 짠 앱의
수익 30%는 애플 몫이었다.

패러다임 : 플랫폼(네트워크)을 장악한 소수가 다수의 가치를 흡수하는 생태계 구성

아무리 열심히 치킨을 튀겨도...

왠지 재주는 곰이 넘고
돈은 왕서방이 챙기는 느낌은...



2018년: 시간당 7,530원 (전년 대비 16.4% 인상)

뭔가

우리에게 **장미빛** 미래가???

개발자 전성시대의 종말

인공지능(AI) 동작평가도구 10분만에 만들기



한국체육대학교 윤지운 교수
jwyoona@knsu.ac.kr



상황에 따라

비전공자와 전공자의 GAP

크게 차이나 나지 않을 수도 있다

개발자 전성시대의 종말

5~6년 전

- 파이썬 알고리즘·밤샘 코딩이 곧 경쟁력
- 코더(Coder)의 희소성이 곧 몸값
- 삼성전자 임원: 코더 좀 빌려주시게



지금 · AI 혁명

- LLM·생성형 AI가 코딩을 스스로 마스터
- 코딩의 장벽 붕괴 — 누구나 중급 이상
- 코더의 희소성은 급락, 가치의 무게중심이 이동

“

"이제 최고의 프로그래밍 언어는 파이썬이 아니라, 인간의 자연어다."

— 젠슨 황(NVIDIA CEO)의 메시지를 빌리며

80년대 '영문과 훈풍' 이 불던 시절이 있었다.

영어만 잘하면 어디든 가던 시절이 있었다. 그러나 영어 능통자가 흔해지자, '영어' 자체가 아니라 '영어로 무엇을 하느냐'가 경쟁력이 되었다. 종급 파이썬 코딩을 AI가 다 해주는 지금, 우리에게 던져지는 질문은 다음과 같다 —

파이썬은 종말인가, 새로운 진화인가?

Survival Game

본론 2

현장에서 느낀 개발자가 살아남는 5가지 행동방침

01	 문제 정의	Syntax보다 Semantics를 지배하라
02	 도메인 융합	나만의 '영토(Domain)'를 확보하라
03	 AI 협업	AI를 10명의 주니어처럼 거느려라
04	 아키텍처 설계	코드보다 '지능의 흐름'을 짜라
05	 검증과 책임	최종서명은 당신의 몫

01.

Syntax보다 Semantics를 지배하라

Syntax 구문, 문법 the arrangement of words and phrases to create well-formed sentences in a language

Semantics 의미, 의미론 the study of the meanings of words and phrases in language

01



문제 정의 — Syntax보다 Semantics를 지배하라

AI는 명령한 대로 코딩할 뿐, '왜 이 코드를 짜야 하는가'는 모른다.

무엇이 진짜 문제인지 정의하는 기획자가 살아남는다.

현장에서...

쇼핑몰 앱 과제를 받은 학생은 결제 API 연동에 밤을 새운다. 하지만 뛰어난 개발자는 묻는다 — '왜 결제 단계에서 고객의 40%가 이탈하는가?' 코드 타이핑은 AI 덕분에 10분의 1로 줄었다. 남은 90%의 시간은 '진짜 풀어야 할 문제'를 정의하는 데 써야 한다.

단순한 문법을 아는자? 의미와 맥락을 이해하는자가 필요한 세상

* 수없이 문을 닫은 코딩학원들/ 그리고 수없이 지원하는 개발자들

02. 나만의 '영토(Domain)'를 확보하라

02



도메인 융합 — 나만의 '영토(Domain)'를 확보하라

코딩만 잘하는 개발자는 너무 많다. AI 기술을 '어느 필드'에 꽃을지가 중요.
자신만의 전문 도메인을 가진 융합형 인재가 살아남는다.

현장에서...

중급 파이썬 실력의 컴공 학생. 몇 년전과 달리 취업이 그렇게 호락호락하지 않다.
그러나 의료 데이터의 특성을 이해하거나 제약 마케팅 메커니즘을 안다면 이야기가 달라진다.
순수 코딩은 AI가 메워주지만, '환자의 이상 행동 패턴을 잡으려면 어떤 데이터 센싱이 필요한가'
같은 도메인 지식은 AI가 대신 공부해주지 않는다.

03



AI 협업 — AI를 10명의 주니어 개발자로 거느려라

일자리를 뺏길까 두려워 말라. AI를 부하 직원으로 두고 지휘하는 시니어 PM으로 일하라

현장에서...

일 잘하는 2~3년 차 개발자는 혼자 일하지 않는다. Copilot, Claude 같은 도구를 양옆에 끼고 '이 부분 보안 점검해줘', '리팩토링으로 효율 20% 높여줘'라 지시한다. 혼자 일주일 걸릴 분량을 반나절에 끝낸다. 이제 경쟁력은 'AI라는 10명의 인턴을 얼마나 능숙히 지휘하느냐'다.

04



아키텍처 설계 — 코드보다 '지능의 흐름'을 따라

세부 문법 오류는 AI가 다 잡는다. 개발자는 전체 시스템의 뼈대와 데이터가 흐르는 파이프라인을 설계해야 한다.

현장 예시 · 건축 비유

과거 개발자는 벽돌을 쌓고 시멘트를 바르는 '조적공'까지 겸했다. 오타 하나에 건물이 무너졌으니까. 이제 그 노동은 AI가 완벽히 해준다. 인간은 '몇 층으로 지을 것인가', '사용자를 어떤 동선으로 움직이게 할 것인가'를 그리는 종합 건축가가 되어야 한다. import 문법 암기보다, 빅데이터와 거대 모델을 어떻게 연결할지 설계하는 훈련을.

05



검증과 책임 — AI가 만들어도, 최종 서명은 인간이 한다

'AI가 짤으니 맞겠지'라는 안일함은 사고로 이어진다. AI는 그럴듯하지만 틀린 코드를 자신 있게 내놓는다. 개발자의 가치는 이제 타이핑이 아니라, 그 결과를 검증하고 책임지는 마지막 게이트에 있다.

현장 예시

AI가 짜준 결제 모듈은 정상 케이스에서 완벽하게 동작했다. 하지만 통화 단위 반올림 오차를 검증하지 않고 배포한 팀은 대량 정산 사고를 냈다. 반면 같은 기능을 받은 다른 개발자는 AI 코드를 그대로 믿지 않고 엡지 케이스 테스트를 직접 설계해 배포 전 오류를 잡아냈다. 코드를 짜는 손은 AI였지만, 사고를 막은 것은 인간의 검증이었다.

맺음말

파이썬 문법 자체가 만능인 세상은 아닐지 몰라도,
파이썬은 **지능을 연결하는 최후의 언어**로 진화한다.

우리는 더 이상 코드를 '타이핑'하는 사람이 아니라, 지능을 설계하고 지휘하는 아키텍트가 되어야 한다.

문제를 정의하라 · 도메인을 확보하라 · AI를 지휘하라 · 흐름을 설계하라 · 검증과 책임을 다하라

AI Agentification over Wireless Networking: The Future of Autonomous Networks Driven by 6G and Decentralized Edge Intelligence

Byung-Seo Kim

Hongik University · Republic of Korea

Experience

- Former Computer Integrated Manufacturing Engineer, Advanced Technology R&D, Motorola Korea
- Former Senior Software Engineer, Networks and Enterprises, Motorola Inc., Schaumburg, Illinois
- Professor at Hongik University since 2007; formerly Department Chair, Vice President of Graduate School, and Director of Planning & Management Office
- Adjunct Professor, Ton Duc Thang University, Vietnam (since 2023)
- IEEE Senior Member and journal associate editor



Academic Records

- B.S. in Electrical Engineering, Inha University
- M.S. and Ph.D. in Electrical and Computer Engineering, University of Florida

Abstract

The explosive growth of 6G networks and the Internet of Things (IoT) has ushered in the era of 'Edge Intelligence,' shifting data processing from centralized clouds to the network edge. However, conventional edge AI systems still rely on static models and lack the autonomy required to dynamically adapt to rapidly changing wireless environments. This keynote presentation introduces a revolutionary paradigm to overcome these limitations: 'Agentic AI' and its implementation process, Agentification. Agentification represents a profound shift that transforms passive edge devices and network nodes into autonomous agents capable of perceiving multi-modal environments (Perception), understanding complex contexts (Reasoning), and executing self-directed workflows to achieve specific goals. In this talk, we delve into the four core pillars for building practical and scalable Agentic AI systems within resource-constrained 6G edge environments: Compactness, Efficiency, Knowledge & Reasoning, and Migration. Furthermore, we showcase how agentified edge intelligence maximizes network efficiency and operational autonomy through some real-world case studies. Ultimately, this keynote envisions a future where edge computing transcends merely running pre-trained AI models to becoming a unified, self-evolving system—Edge General Intelligence—offering deep insights into the key enabling technologies, structural designs, and future research directions for next-generation autonomous networks.

Technical Sessions

BOARS: Development of an AI-Based Real-Time Wild Boar Integrated Control System

Hyunjun Park

Korea Maritime & Ocean University · Republic of Korea



BOARS

인공지능 기반 실시간 멧돼지 통합관제 시스템 개발

 국립한국해양대학교
해양과학기술융합대학 인공지능공학부

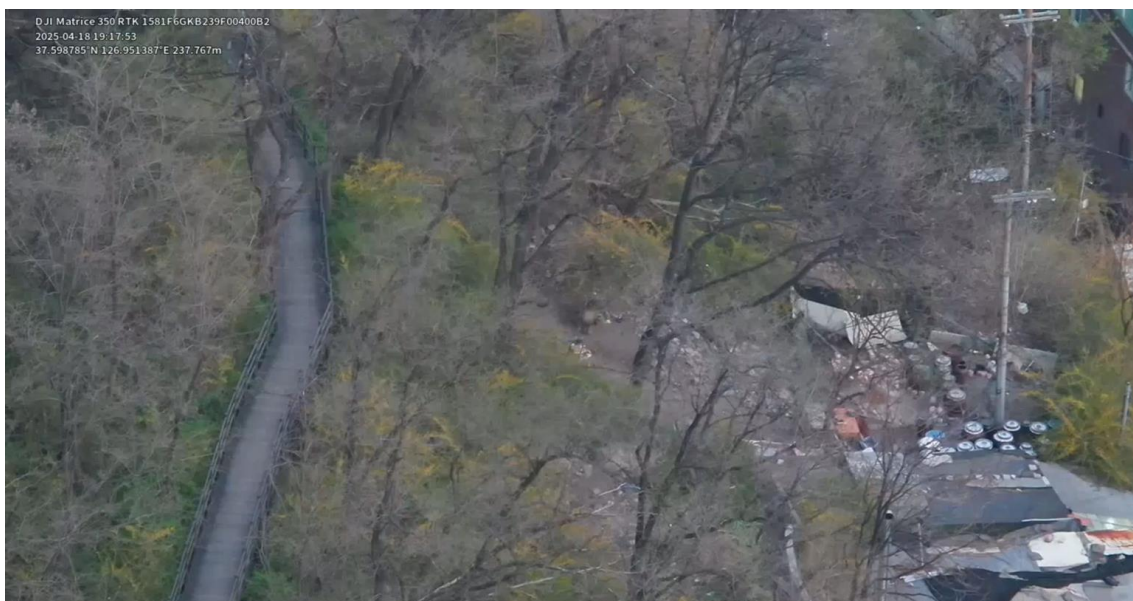
 박현준 교수

 hyunjun@kmou.ac.kr



에이~ 멧돼지가 어딴어..?

2



2025-04-18
상명대학교 주변
포방터2나길



도심 생활권으로 확장되는 멧돼지 출현

3



■ 2026.07.20 - 부산, 나홀간 멧돼지 출몰 신고 8건

부산에서 여름철인데도 멧돼지가 잇따라 출몰. 2026년 1~6월 출몰 75건
상반기 포획 211마리로, 2024년 상반기 114마리와 2025년 상반기 116마리보다 크게 늘어나, "멧돼지가 특정 계절에만 나타나는 문제가 아님"

■ 2026.05.01 - 서울 마포구 아파트 + 이화여대 연속 출몰

마포구 아파트, 인근 이화여대에서도 멧돼지가 나홀 연속 캠퍼스를 활보해 학생들에게 야간 출입 자제 경보
서울의 최근 멧돼지 목격 신고가 400건 이상. "서울 한복판에서도 멧돼지가 나타남"

■ 2026.01.20 - 대전 도심에서 닳새 연속 출몰, 9마리 포획

대전 서구에서 5일 연속 멧돼지 출몰 신고. CCTV에는 야산에서 도심으로 내려왔다 다시 돌아가는 모습도 포착



멧돼지 관리는 왜 필요한가

4

1. 생활권 안전 위험

- 도심·공원·주거지 인근 출현 증가 → 시민 안전, 교통사고, 야간 보행 위험

2. 농작물 및 시설 피해

- 농경지 훼손, 먹이 활동으로 인한 반복 피해 → 지역 주민과 지자체의 대응 부담 증가

3. ASF 등 감염병 확산 위험

- 야생 멧돼지는 아프리카 돼지 열병 확산 관리의 핵심 대상 → 위치·이동경로·개체정보 관리 필요

4. 데이터 기반 대응 필요

- 출현 이후 대응을 넘어, 출현 패턴을 추적·분석해야 함 → "어디에 나타났는가"에서 "어디에 나타날 가능성이 높은가"로

효과적인 멧돼지 관리를 위해서 개별 출현에 대한 사후 대응을 넘어, 분산된 현장·위치·영상 데이터를 통합하고 출현 패턴을 분석하여 선제적 대응으로 연결할 수 있는 체계가 필요함

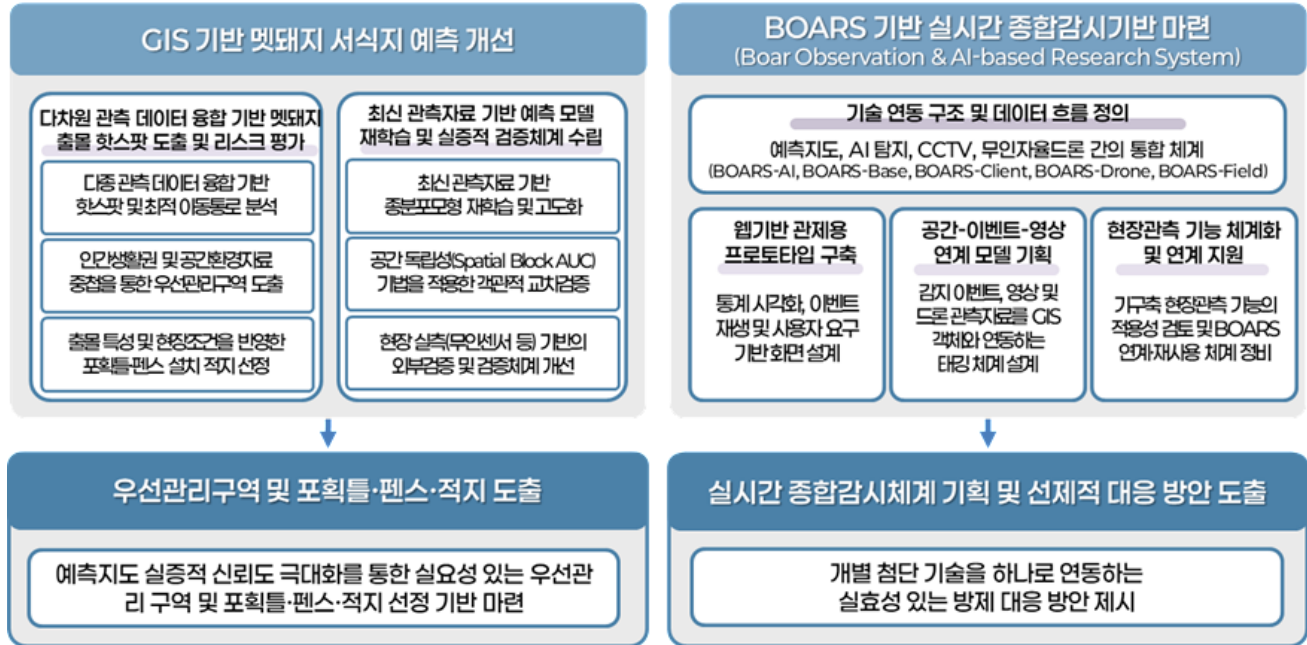


도심 출몰 멧돼지 서식지 예측 및 종합감시 기술 개발 연구



국립생물자원관
National Institute of Biological Resources

5

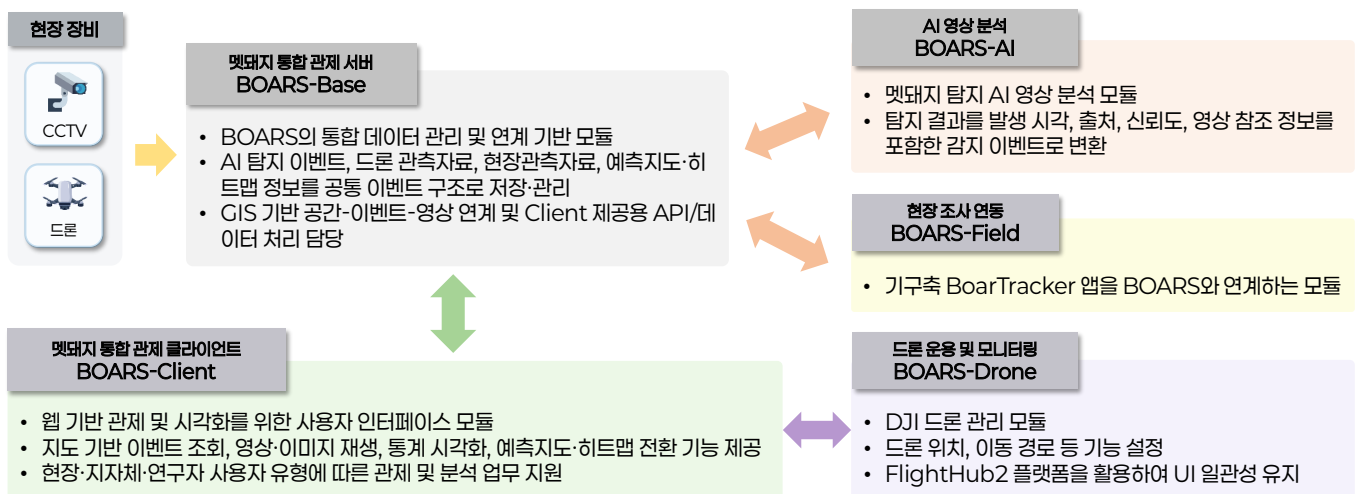


BOARS

Boar Observation & AI-based Research System

6

멧돼지 관측 및 AI 기반 연구 시스템 BOARS





BOARS 요구사항 정의 및 구현 현황

7

- AI 탐지, 통합 관제, 지도 시각화, 드론 연계, 현장조사 연계 기능을 중심으로 총 61개의 요구사항을 정의
- 기능 요구사항: 42건, 데이터 요구사항: 7건, 인터페이스 요구사항: 9건, 비기능 요구사항: 3건



완료 69% · 부분 26% · 외부담당 3건 · 미구현 0건 (실제 코드 전수 검증 기준, 2026-07)

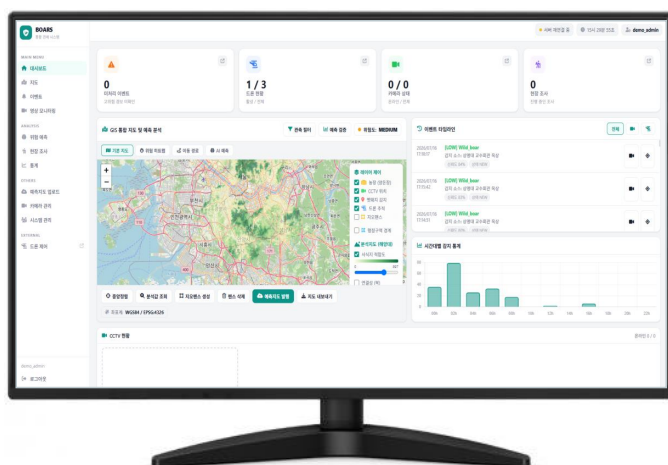
세부 모듈	완료	부분	미구현
BOARS-AI	7	1	0
BOARS-Base	23	9	0
BOARS-Client	11	3	0
BOARS-Drone	1	1	1
BOARS-Field	0	2	2
총계	42	16	3



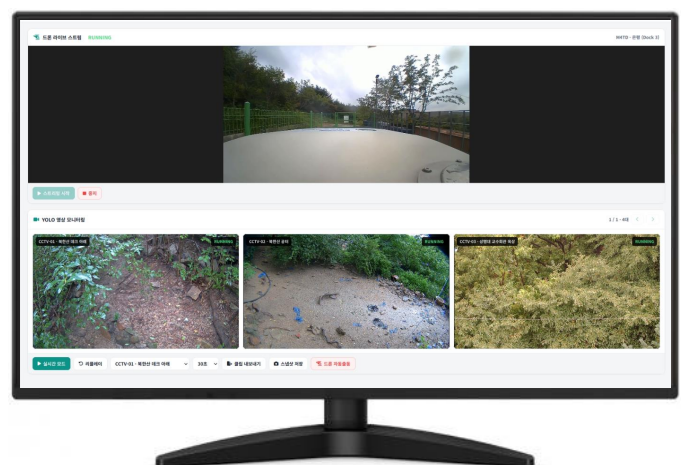
BOARS 중간 구현 결과 예시

8

메인 화면



CCTV 및 드론 영상 모니터링





향후 발전 방향

9

1. 전국 단위 멧돼지 감시 체계 구축

- 최근 멧돼지 출현이 증가하고 있어 멧돼지 감시 체계에 대한 수요가 높아짐
- 관리자/생물자원관/지자체 등 권한에 따른 맞춤형 기능을 제공하고 있으며, 지자체에서 실시간으로 멧돼지 출현 현황 등을 직접 조회할 수 있음
- 멧돼지 검출시 지자체 담당자에게 카카오톡 알람 전송
- 향후 더 많은 지자체와 협력하여 다양한 위치의 CCTV 영상을 획득하여 전국 단위 멧돼지 감시 체계 구축 가능

2. 멧돼지 외의 다양한 야생 동물 관리 시스템으로 확장

- 국립생물자원관에서 연구, 관리하고 있는 다양한 동물 대상으로 확장 가능
- 국립생물자원관에서 보유한 학습 데이터 외 추가 데이터 수집 필요

3. CCTV에서 인식된 멧돼지 위치로 드론이 출동하고, 멧돼지를 추적하여 선제적 대응 가능

- 현재 드론 제조사, API 문제 등으로 지자체에 설치하는데 한계가 있으나, 향후 개선 가능할 것으로 예상됨



수고하셨습니다.



국립한국해양대학교
해양과학기술융합대학 인공지능공학부



박현준 교수



hyunjun@kmou.ac.kr

Research on Mission-Driven Intelligent AI Agents: From Defense Mobility to Sports Science

Youngsoo Kim

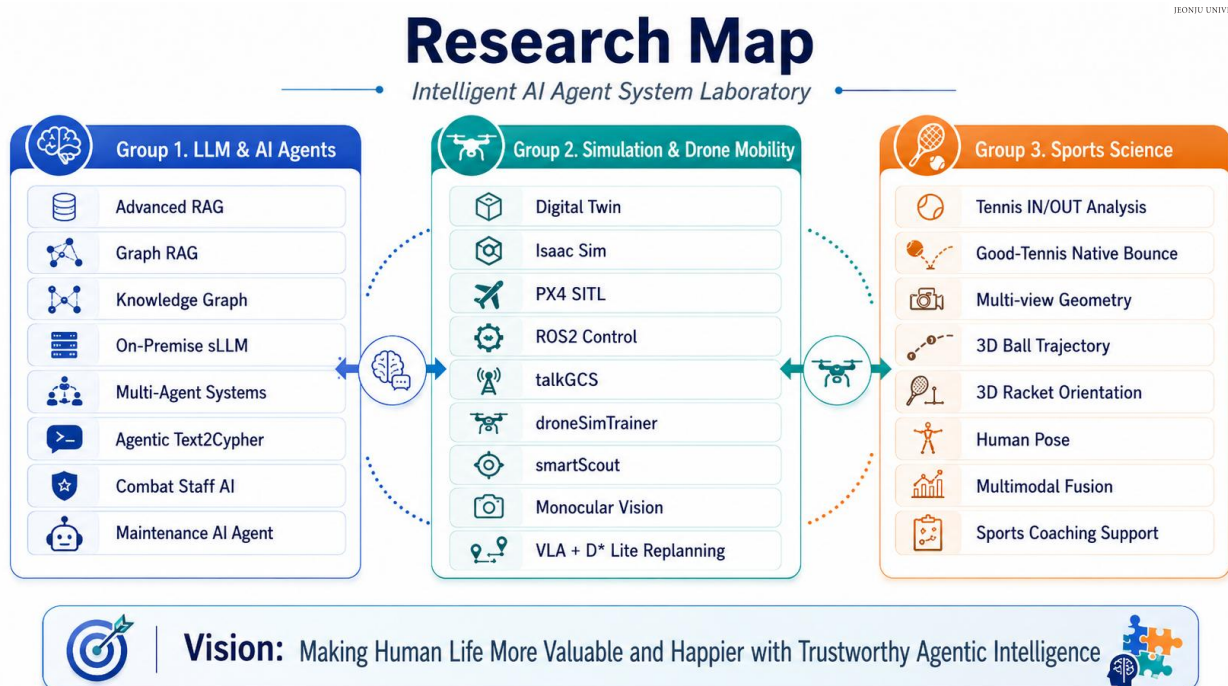
Jeonju University · Republic of Korea

Research on Mission-Driven Intelligent AI Agents

[From Defense Mobility to Sports Science]

Prof. Kim, Youngsoo
(pineland@jj.ac.kr)

Supported by Hangil Co., Ltd.



RISE project



Multi-PurPose Defense AI Drone Development Platform

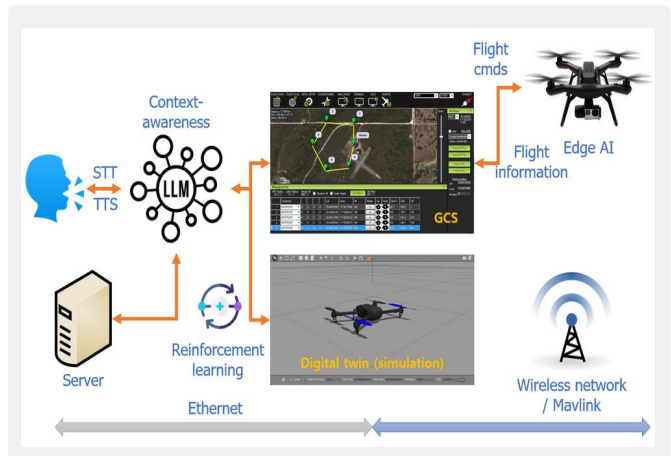
This study aims to develop and demonstrate an all-in-one intelligent mobility (drone) software package that automates, adds intelligence to and streamlines land, sea and air mission execution, using current AI technologies such as speech recognition, large language models (LLM), simulation and reinforcement learning, and edge AI.

Research Background

- Rapid growth of K-Defense and the widening importance of the defense industry
- Jeonbuk State selected as the national defense innovation cluster region
- Jeonbuk State selected as the 'Physical AI Proving-Ground Development Project' final target site



System Architecture



RISE project



Core Modules

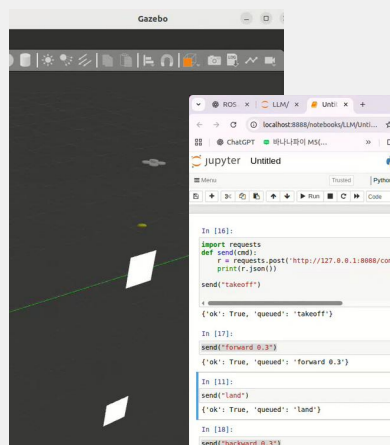
talkGCS

- Fine-tuning and integration of Small Language Models (SLMs) for flight control command generation



droneSimTrainer

- Development and Advancement of Gazebo/IssacSim-based Drone Control Simulation Based on Flight Control Commands



smartScout

- Development and Optimization of YOLO/VLA-based Object Detection and Identification Algorithms
- Currently conducting research on the development of a context-aware TinyLLM



RISE project



AI-based Autonomous Flight Path Planning on Drones

This study implements and validates a path-planning system that flies to the mission area and returns autonomously using one monocular camera and no LiDAR, even when the link to the ground control station is lost. A three-stage pipeline — A* global planning, ROI risk detection with VLA avoidance-point prediction, and D* Lite incremental replanning — was fully flight-verified on a survey-based campus digital twin (0.62 m registration residual) and a real flight stack (PX4 SITL).

Photorealistic Environment Modeling

Reference Orthophoto

(National Geographic Information Institute)

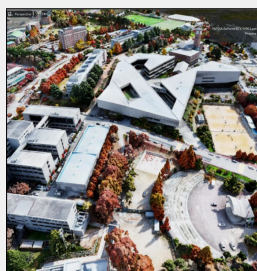


Precision Calibration

Alignment & Point Cloud
- Median residual <0.62m



3D Mesh Reconstruction
& Texturing



Completed 3D Mesh Campus



RISE project



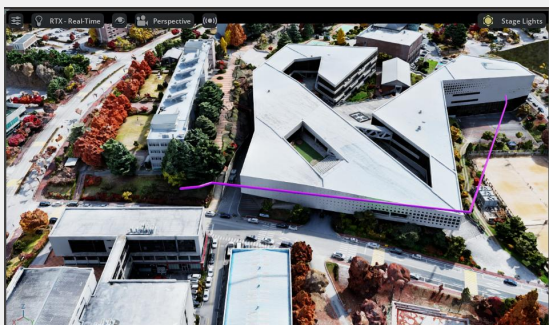
AI-based Autonomous Flight Path Planning on Drones

AI based Path Planning

A* Global Path Planning

Automatic waypoint generation when operator designates a target point

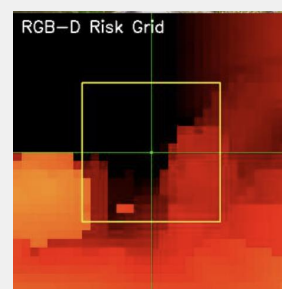
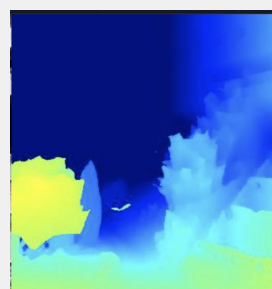
- Maintains a 1.5m safety radius over the grid
- Automatic targeting of approach points via 8-directional search



Monocular Depth-based ROI Risk Detection

Estimated depth of 0–100m via monocular camera

- Intensive monitoring of the absolute defense zone covering the center 25% (16×16 cells) of the 32×32 danger grid
- Risk level $R \geq 0.70$; immediate evasion upon entering within 15m



RISE project



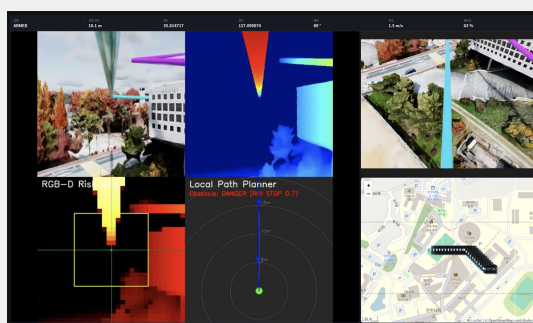
AI-based Autonomous Flight Path Planning on Drones

AI based Path Planning

D* Lite Incremental Replanning

Route replanning when deviating by more than 20m

- 83% computational reduction by reusing previous search data
- Re-path generation speed: <14~78ms



VLA based Human Avoidance Learning

SmolVLM-500M + LoRA (Lightweight Edge Inference Optimization)

- Predicts the optimal single avoidance point (V1) in 3D space by learning on 90 human avoidance demonstrations

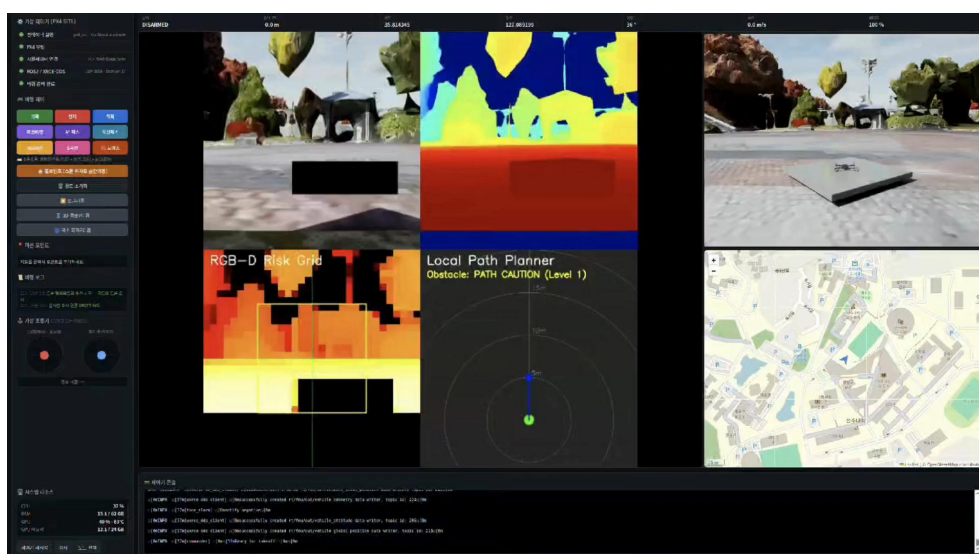


RISE project



AI-based Autonomous Flight Path Planning on Drones

Demonstration



Supported by Hangil Co., Ltd.



Optimization of Advanced RAG System for Aircraft Mechanics

This study implements and validates a RAG-based AI agent capable of real-time question answering on the aircraft maintenance floor, and proposes an optimization approach. A configuration that is open-source, commercially deployable and Korean-capable, with optimized accuracy and response speed, was derived experimentally. As a result, answer accuracy on user queries improved by 20% and response time was cut to roughly 5 seconds, demonstrating the feasibility of an AI agent for aircraft maintenance support.

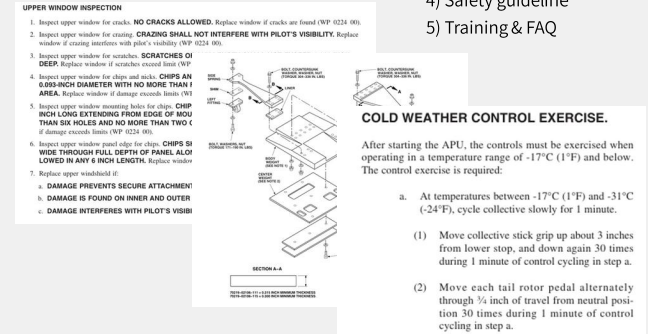
Research Background

- Aircraft accidents are critical events that cause severe human and material loss
- Accidents arise mainly from pilot or technician inexperience (human error); aircraft maintenance demands a high level of expertise
- A maintenance AI agent can stand in for a veteran technician and help new technicians build proficiency quickly



Aircraft Maintenance Data

- 1) Maintenance manuals
- 2) Parts lists & diagrams
- 3) Diagnostic data
- 4) Safety guideline
- 5) Training & FAQ



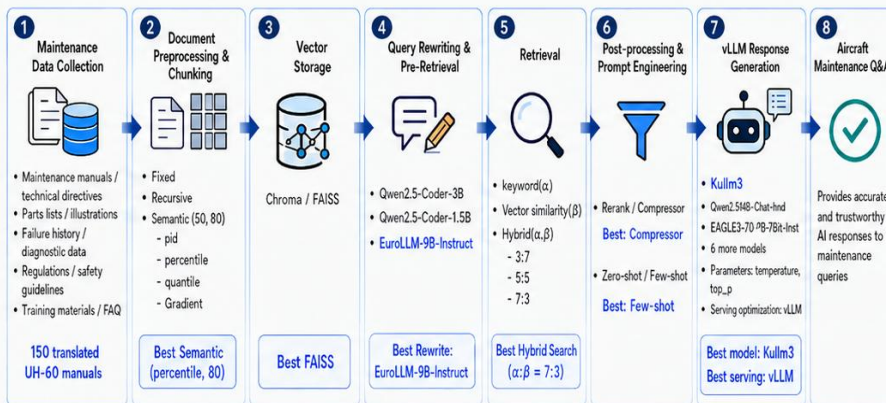
Supported by Hangil Co., Ltd.



RAG System Development Process and Key Experimental Results for Aircraft Mechanics

AI Maintenance Support Agent Based on the UH-60 Utility Helicopter Maintenance Manuals

1. System Development Process



2. Key Experimental Results



Effect of Applying vLLM (KULLM3)

Category	Average Response Time (sec)	Average Quality (Avg. Qual.)
Naive RAG	166.72	4.05
RAG + vLLM	11.32	11.05 3.46

Average Number of Correct Answers by Question Type

Question Type	Naive RAG	Improved RAG	Improvement
Elements	1.8	4.2	+2.4
Management Method	2.6	3.0	+0.4
Procedure	2.2	3.4	+1.2
Terminology	3.8	3.8	0.0

Average Generation Time by Question Type

Question Type	Naive RAG	Improved RAG	Improvement
Elements	11m	5.5s	-9m 52s
Management Method	15m	5.72s	-9m 15s
Procedure	46m	5.76s	-40m 24s
Terminology	4m	3.75s	-16s

RISE project



Tactical Strategy Support AI Agent for Small Unit Commanders

Research on AI agents for Tactical Course of Action (COA) support focuses on empowering small-unit commanders with rapid, data-driven decision-making tools in complex combat environments. Facing extreme time pressure and information overload, small-unit leaders require real-time assistance to evaluate options and maintain a tactical advantage over opposing forces.

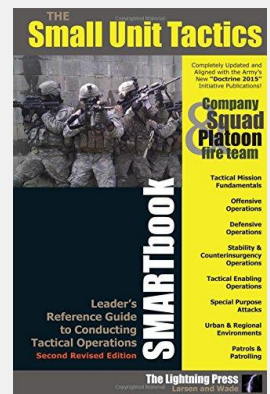
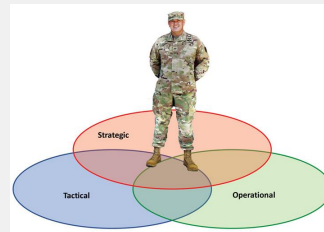
Background

- Data Deluge vs. Cognitive Limits
- Compressed Decision Timelines
- Friction & Fog of War



Key Research Pillars

- Multi-Domain Data Synthesis
- Predictive Risk Assessment
- Automated COA Generation & Evaluation
- Human-in-the-Loop Collaboration

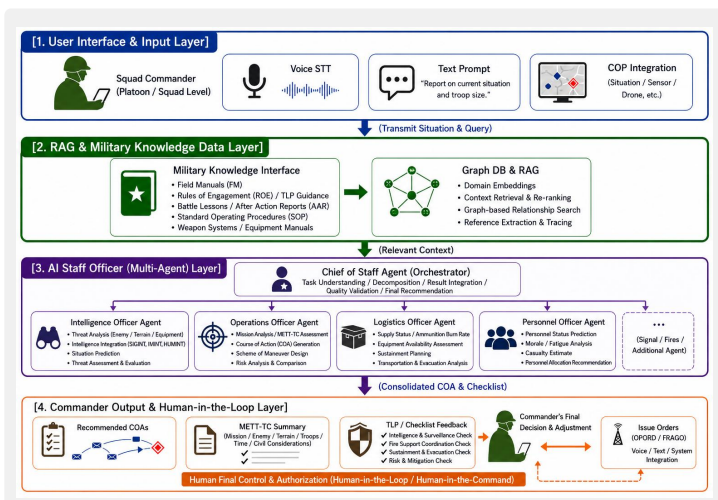


RISE project

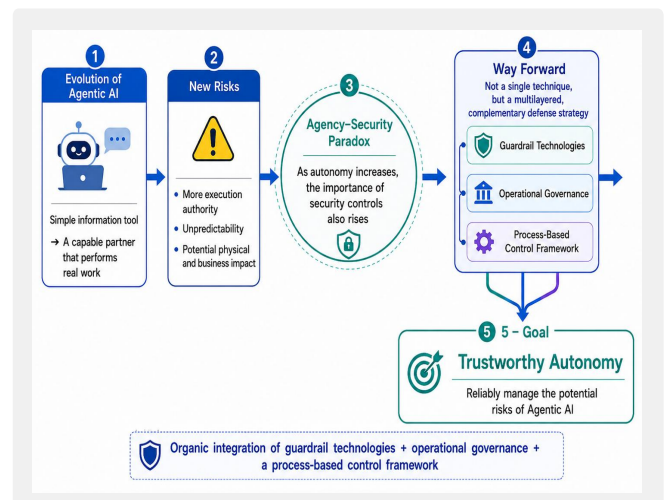


Tactical Strategy Support AI Agent for Small Unit Commanders

System Architecture



Reliable & Securable Agentic AI



RISE project



Building Medical Ontology and Knowledge Graph

This study aims to construct a medical ontology and a Knowledge Graph using LLM to enhance the quality of medical question-and-answer services. By utilizing LLM to identify medical entities and their inter-entity relationships to build the ontology and graph databases, the study seeks to improve search accuracy and response quality for complex medical queries.

Background

- Naive RAG's plain similarity search falls short; complex medical queries need relation-based retrieval and reasoning
- Relations among diseases, symptoms, drugs and treatments are defined in an ontology, making the semantic structure explicit
- LLM-based entity and relation extraction builds a knowledge graph, structuring unstructured medical information
- Combining graph relations and multi-hop traversal with Graph RAG raises retrieval accuracy and reasoning on compound queries

Extract Entity & Relation Types

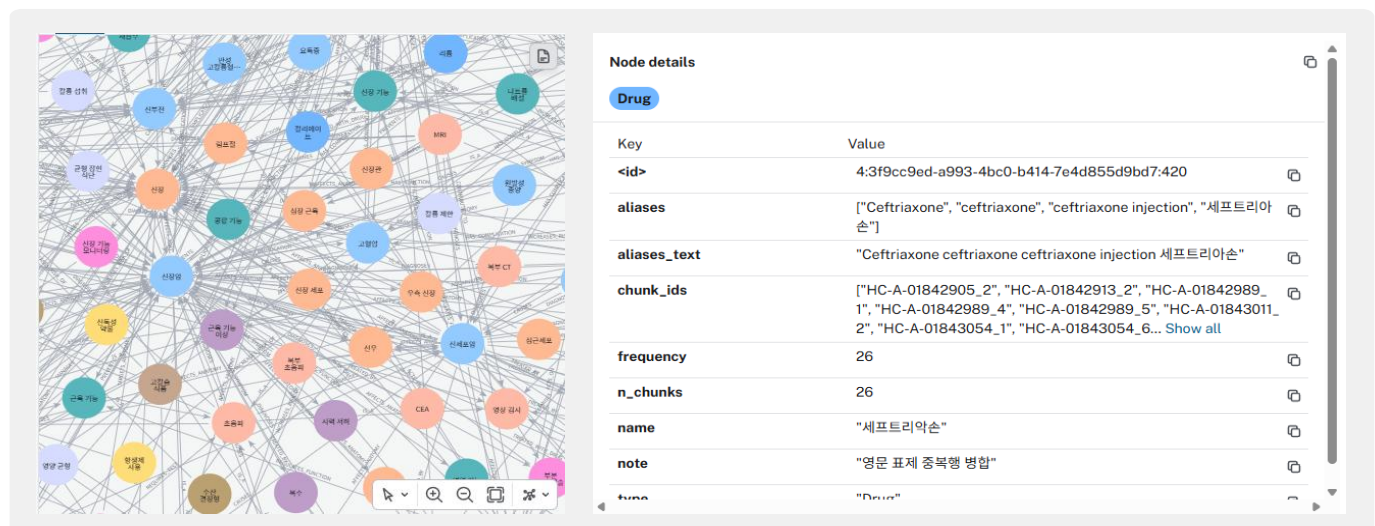
- 임상 질환 및 증상 (Clinical Conditions & Pathology) - Disease (질환 / 질병) - Symptom (환자 자각 증상) - Finding (의학적 객관적 소견) - Pathogen (병원체 - 세균, 바이러스 등) - Etiology (병인 / 발생 원인) - Prognosis (예후 / 경과 예측) - 치료 및 중재 (Treatments & Interventions) - Drug (검증된 정식 의약품) - UnverifiedDrug (미검증 / 대체 의약품) - Treatment (치료법 / 시술 / 수술) - TreatmentOutcome (치료 결과 / 반응) - Equipment (의료 기기 및 장비) - 진단 및 생체 측정 (Diagnostics & Measurements) - DiagnosticTest (진단 검사 - 혈액검사, MRI 등) - Measurement (측정 수치 / 생체 지표)	- 증상, 경과 및 질환 관계 (Symptoms, Disease Progression & Complications) - has_symptom (증상을 동반함) - causes_symptom (증상을 직접 유발함) - manifests_symptom (증상으로 발현됨 / 표출됨) - causes (질환 / 병인을 직접 유발함) - has_complication (합병증을 동반함) - comorbid_with (동반 질환 / 동반 관계임) - induces_disease (다른 질환을 유도 / 유발함) - triggers (발병 / 발작을 촉발함) - leads_to (질환 / 상태로 이행함) - contributes_to (질환 / 경과에 요인으로 기여함) - co_occurs_with (다른 증상 / 상태와 동시 발생함) - 치료, 약물 및 중재 (Treatments, Interventions & Drugs) - treated_by (치료법 / 시술 / 수술로 치료됨) - treated_with_drug (약물 치료로 경과를 관리함) - uses_drug (치료 과정에서 특정 약물을 사용함) - relieves_symptom (증상을 완화 / 경감시킴) - achieves_outcome (특정 치료 결과 / 임상 목적을 달성함)	- 신장 기능 검사 - DIAGNOSES -> 신장 질환 - 신장 기능 검사 - MEASURES -> 혈류 농도 - 요속 검사 - IS_A -> 요역동학 검사 - 요속 검사 - DIAGNOSES -> 과민성 방광 - 요속 검사 - DIAGNOSES -> 요로 감염 - 근 배양 검사 - DIAGNOSES -> 임질 - 심전도 검사 - DIAGNOSES -> 고혈압성 - 심전도 검사 - HAS_FINDING -> 심전도 이상 - 심전도 검사 - MEASURES -> 심률 농도 - X-ray 검사 - EVALUATES_SYMPTOM -> 흉구리 불투명
---	---	---

RISE project



Building Medical Ontology and Knowledge Graph

Constructed knowledge graph



RISE project

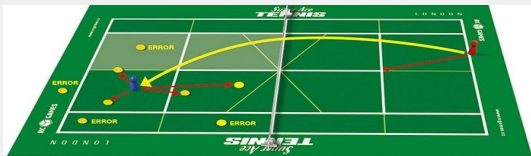


Vision-based Tennis Player Motion Analysis

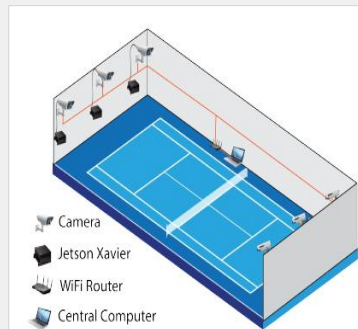
This study reconstructs tennis ball and racket motion in a single 3D coordinate system from synchronized multi-camera video, then combines it with human pose. Detection supplies 2D ball centres and racket regions; timestamps are synchronized and balls matched across views; triangulation with reprojection checks yields the 3D trajectory, while racket keypoints and masks are fused conservatively into a 3D racket direction.

Background

- A tennis ball is small, fast and often motion-blurred, so display must follow detections rather than free prediction
- A single camera cannot recover metric depth reliably, so multi-view geometry is required
- Racket orientation links player movement to ball contact and the resulting trajectory
- Ball tracking and racket pose have to be solved together, in one 3D coordinate system

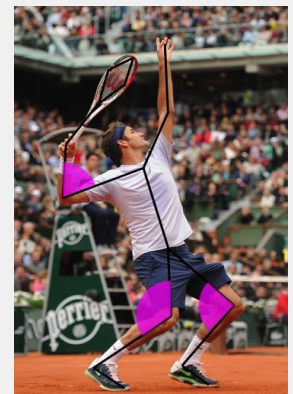


Target System (e.g.)



Multi-camera setup for locating a tennis ball.

Source: Xiao, Zaidi, and Gombolay, ICRA 2024, Fig. 4.



RISE project

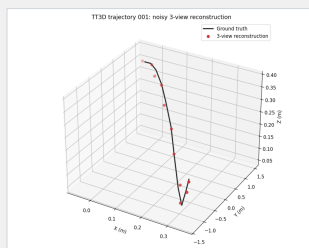


Vision-based Tennis Player Motion Analysis

Core Technologies

3D Ball Recovery

- Predicting the trajectory of a tennis ball based on geometric principles
- Predicting the trajectory of the ball using training data
- Research methods to improve accuracy by combining the two methods above



Racket Movement Analysis

- Additional Annotation and Learning to analyze Racket Face Movement
- Analyze Racket Movement in conjunction with Key Points of the Body and Hand
- Research on more accurate exercise postures, including rackets



Multi-Camera Fusion

- Time/Spatial Synchronization & Calibration among Images from Multi-cameras
- Reconstructing Precise 3D Trajectories and Racket Direction in 3D space through Triangulation using Multi-view Geometry



RISE project



Vision-based Tennis Player Motion Analysis

Demonstration



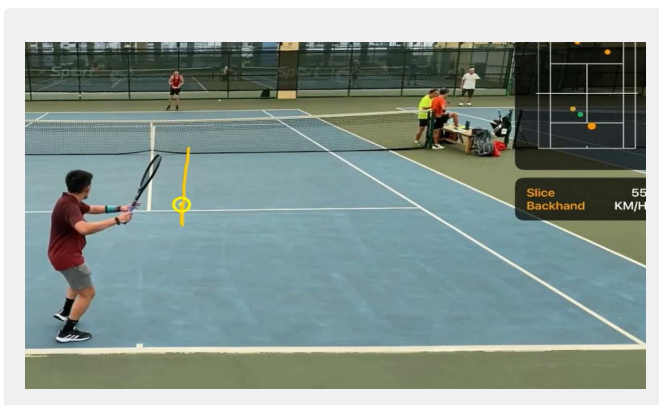
RISE project



Tennis IN/OUT Landing Judgment

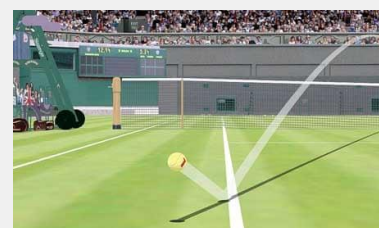
This project reuses Good-Tennis native outputs to build a conservative decision and visualization pipeline: native bounce, court mapping, IN/OUT call, then a synchronized left video and minimap. Native landing events and a homography to court coordinates drive every decision, and both views read the same authoritative event table so they can never disagree.

Background



Core Technologies

- High-Speed Multi-Camera Synchronization
- Court Area Calibration for Images from Multiple Cameras
- Ball Detection & 3D Trajectory Restoration
- Ball Bounce Point Detection and Ball Deformation (Compression) Physical Modeling



RISE project



Video Captioning

The Video Captioning (VC) combines temporal event localization with natural language generation for untrimmed videos. Unlike standard video captioning that yields a single summary, VC outputs a set of temporal segments with textual descriptions. This research investigates its technical evolution to model complex multi-event video narratives effectively.

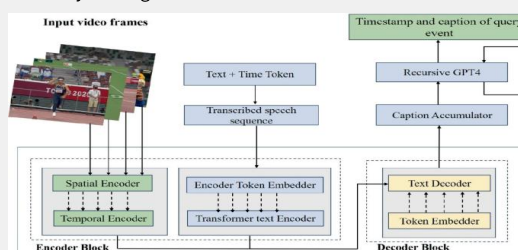
Background

- Explosive increase in video contents (e.g. YouTube)
- A lot of time spent watching and understanding video contents
- Needed for situational assessment through streaming video, such as in surveillance and reconnaissance



Core Technologies

- Event Proposal & Localization in video
- Inter-Event Context Modeling
- Multimodal Data Fusion – Images, Audio, etc.
- Redundancy & Length Control

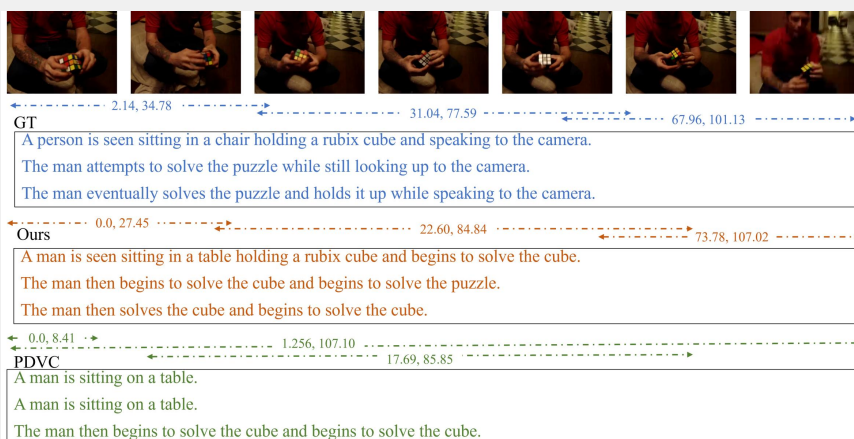


RISE project



Video Captioning

Experimental Results



- PDVC repeats 'sitting on a table' and only mentions the core object, 'cube' in the last sentence.
- In contrast, our method identifies the rubix cube in the first sentence.
 - As time progresses, the process of 'begins to solve' → 'solves the cube' is revealed, making the temporal logic more complete. In the surfing example, while PDVC exhibits grammatical errors like 'beach beach' and the illusion of static description, this method captures dynamic elements such as large waves and riding along the water.

Thank you for your attention

Q & A

Changes in Engineers' Job Competencies in the AI Era

Heungseob Kim

Changwon National University · Republic of Korea



International Conference on 2026 PyGeek

AI 시대 엔지니어 직무능력 변화

국립창원대학교 산업시스템공학과

Smart System Laboratory,

Department of Industrial and Systems Engineering,

Changwon National University

2026년 08월 20일

발표자: 김 흥 섭

heungseob79@changwon.ac.kr



스마트시스템연구실 (Smart Systems Lab.)



발 표 순 서

- I 생성형 AI 시대의 엔지니어 Job Crafting
- II 실 증: AI 도입 후 조직 변화
- III 결 언: AI 시대 엔지니어의 핵심 역량

국립창원대학교
CHUNGNAM NATIONAL UNIVERSITY

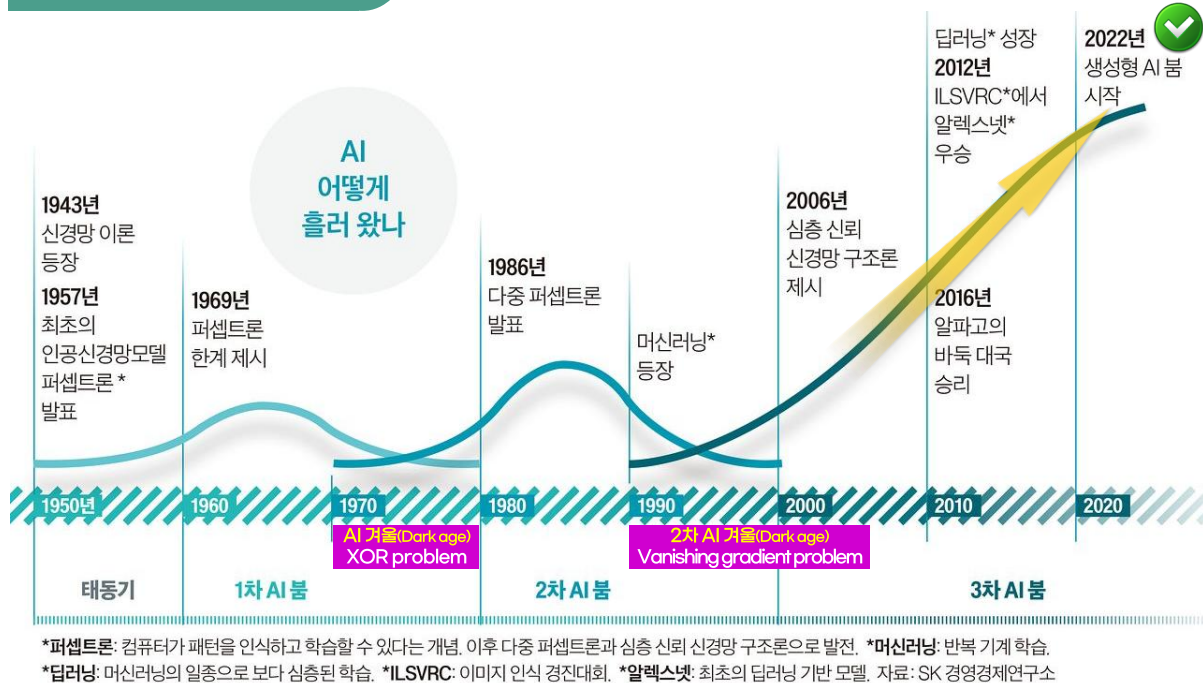
스마트시스템연구실 (Smart Systems Lab.)

PYGEEK

I · 생성형 AI 시대의 엔지니어 Job Crafting

I II III

인공지능(AI)의 발전 동향



국립창원대학교
CHUNGNAM NATIONAL UNIVERSITY

스마트시스템연구실 (Smart Systems Lab.)

PYGEEK

I · 생성형 AI 시대의 엔지니어 Job Crafting

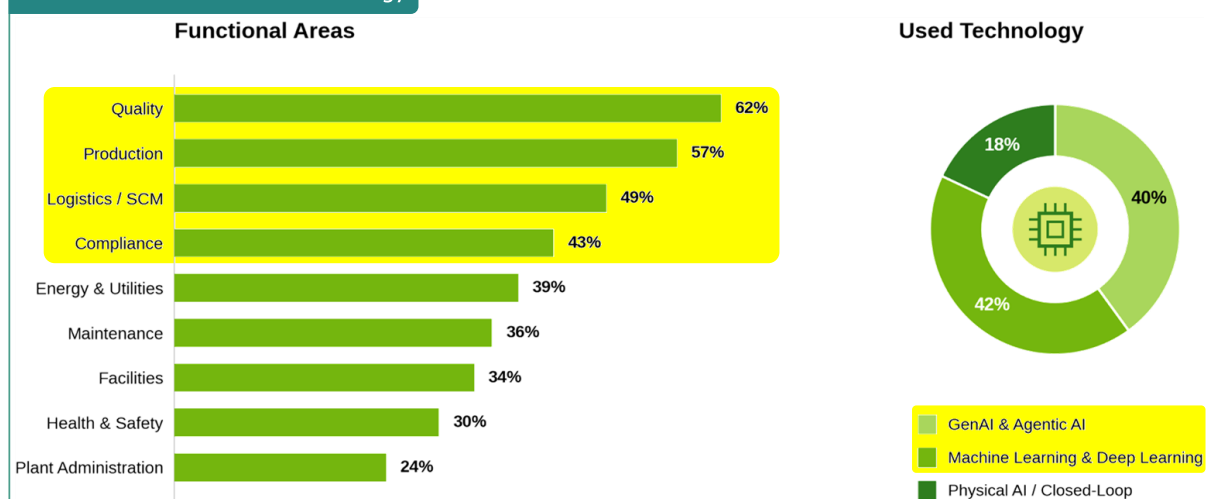
I II III

제조업부문 AI 활용 현황

AI in Manufacturing (Deloitte, June, 2026)

- ☑ 주요 AI 활용 분야 : 품질(62%) > 생산(57%) > 물류(49%) > 내부통제 및 준법감시(43%)
- AI 기술 : 머신러닝/딥러닝(42%) > 생성형 및 에이전틱 AI (40%) > 피지컬 AI/Closed-Loop(18%)

Functional area & Used AI technology

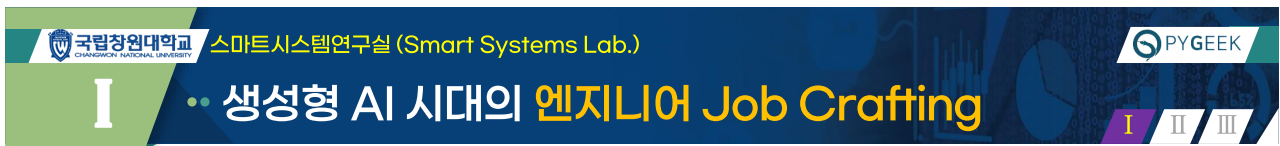
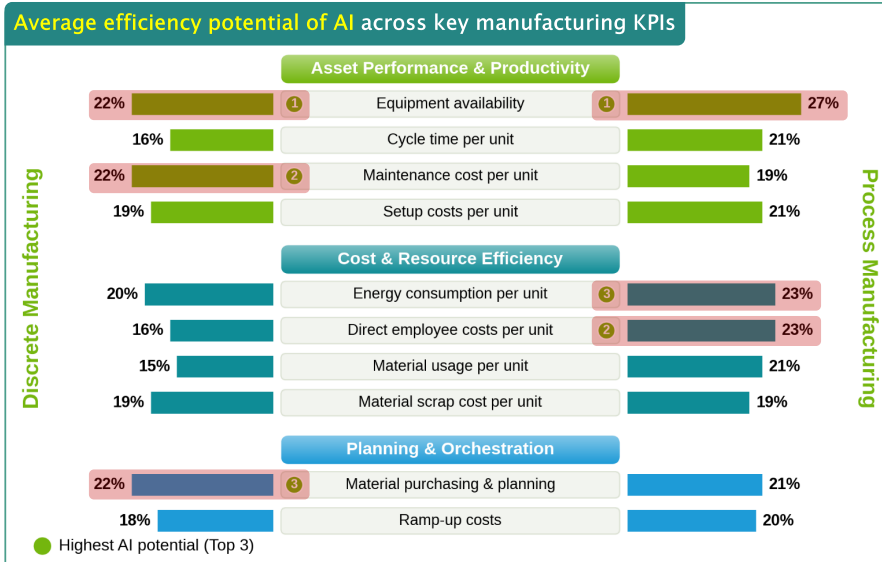




제조업부문 AI 활용 현황

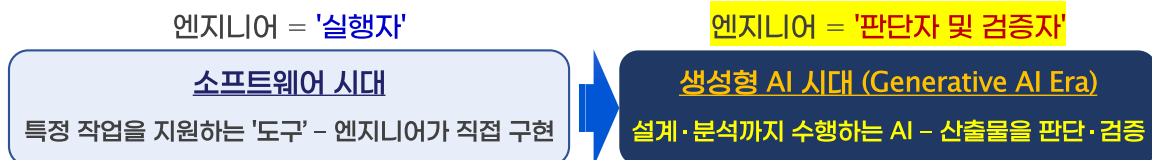
AI in Manufacturing (Deloitte, June, 2026)

- 이산 제조(Discrete Manufacturing) : [자산 및 생산성] 및 [자재 소요 계획]에서 높은 잠재력
- 프로세스 제조(Process Manufacturing) : [비용/자원 효율성] 및 [자산 및 생산성]에서 높은 잠재력

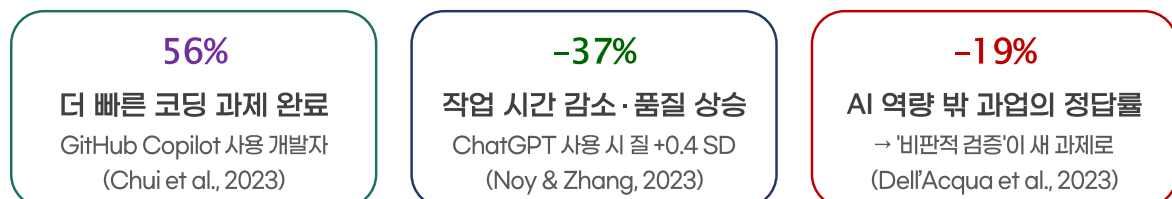


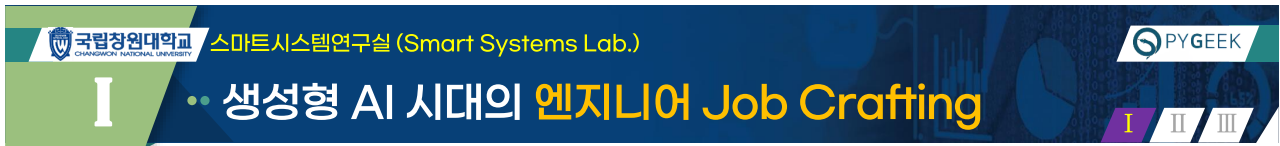
생성형 AI 시대 엔지니어의 변화

- 생성형 AI(Generative AI)의 확산 → 엔지니어 업무 수행 방식의 근본적 변화
 - 코드 생성, 문서 작성, 설계 초안 도출, 정보 탐색 등 반복적·절차적 작업이 자동화
 - 지식 노동의 설계 방식, 요구 역량, 직무의 경계 자체가 재편되는 중



- 생산성 효과와 한계 : 실증 연구 결과





Job Crafting : 일하는 방식을 보는 렌즈

☑ 잡크래프팅(Job Crafting)이란? (Wrzesniewski & Dutton, 2001)

· 개인이 자신의 욕구·역량·열정에 맞춰 직무의 경계와 의미를 스스로 재설계하는 과정

과업 크래프팅 (Task)

업무의 수나 범위, 수행 방식을
물리적으로 변경

관계 크래프팅 (Relational)

상호작용 대상의 범위·질을 조정,
대인 관계적 맥락을 변화

인지 크래프팅 (Cognitive)

업무의 심리적 의미·가치를 재정의,
직무 정체성을 변경

생성형 AI 환경에서의 확장

직무 요구를 줄이고 직무 자원을 늘리는
자원 보존의 관점으로 확장

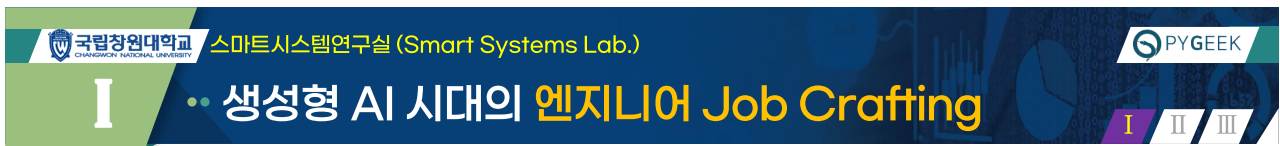
AI Crafting (Li et al., 2024)

AI를 단순한 도구가 아니라 업무 방식을
재구성하는 '행위적 파트너'로 인식

시그널 크래프팅 (Mayer et al., 2025)

AI 산출물에 자신의 전문적 통찰·검증 과정을
의도적으로 노출 → 개인의 독자적 기여 증명

[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28



연구결과 ① : 과업 크래프팅

☑ 과업 크래프팅 : 실행의 위임과 직무의 고차화

· 물리적 실무를 인공지능(AI)에게 위임하고, 인지적 자원을 고차원적 설계·검증에 재배치
→ 엔지니어의 역할이 「직접 구현」에서 『시스템 오케스트레이션』으로 변화

[과거] '백지' 상태
처음부터 직접 작성



AI가 초안 생성
1차 프로토타입



비판적 검토·수정
'초안 리팩토링'으로 시작

[설문자 : S/W담당, 30세(경력 4년 / AI 활용 1년)] “예전에는 처음부터 끝까지 직접 코드를 작성했는데, 지금은 AI에게 먼저 초안을 생성해 보라고 하고 그걸 기반으로 수정합니다. 업무의 시작점 자체가 완전히 바뀌었습니다.”

01 업무 시작점의 이동

'백지' 탈피 — AI 초안을 검토·수정하며 업무 시작

02 정보 탐색 경로 단일화

파편화된 웹 검색 → 통합된 AI 대화창

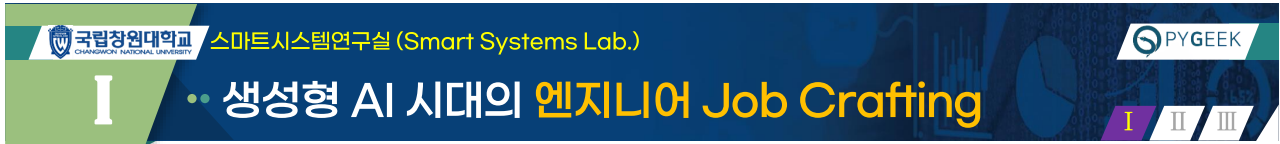
03 저부가가치 업무 위임

반복적·정형적 과업은 AI, 자신은 판단 활동에 집중

04 업무의 양적·질적 확장

여력을 구조 설계·병렬 연구에 재배분 → '질적 심화'

[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28



연구결과 ② : 관계 크래프팅 (사람)

☑️ 사람과의 관계 크래프팅 : 소통의 효율화와 협업 방식의 재조정

▶ 기술적 문제 해결의 일차적 대상을 인간 동료에서 인공지능(AI)으로 전환

→ 대인 의존도를 낮추고 소통의 질을 선택적으로 조정



[설문자 : 백엔드 엔지니어, 30세(경력 5년 / AI 활용 1.5년)] "요즘은 먼저 AI에게 물어보고 어느 정도 해결한 다음에 사람에게서는 최종 확인만 합니다."

인적 상호작용의 효율화

저수준 질의 축소 : 리뷰·방향 논의 중심 '고부가가치 소통'으로

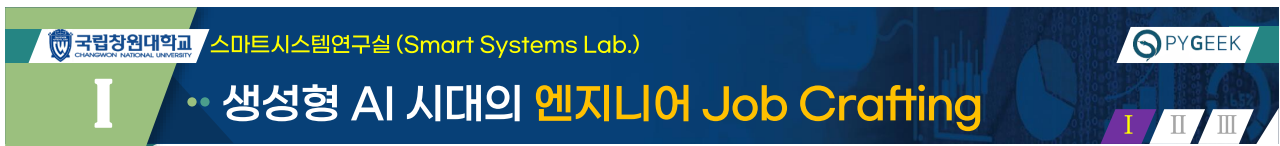
개인 완결형 협업 구조

타 직군·동료 의존도 감소, 업무의 주도권 확보

기술기반 의사소통 참여 확대

정보 격차 완화 : 비엔지니어도 기술 논의에 참여

[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28



연구결과 ② : 관계 크래프팅 (AI)

☑️ AI와의 관계 크래프팅 : 기술적 파트너십과 지능형 협업

▶ 인공지능(AI)을 「지적 동료」, 「가상 팀원」으로 정의하며 능동적 파트너십 구축

→ 생성형 AI 환경에서 새롭게 확장된 관계 크래프팅

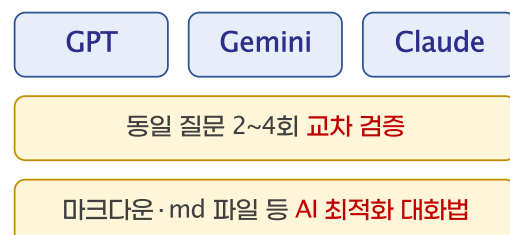
[설문자 : S/W개발, 34세(경력 11년 / AI 활용 2.5년)] "AI에게 설계자, 구현자, 리뷰어의 역할을 각각 부여하여 가상의 팀을 조직하고 나는 리더로서 이들을 지휘합니다. 혼자 일하지만, 팀으로 일하는 기분을 느끼며 체계적으로 결과물을 만들어냅니다."

[설문자 : 양자컴퓨터·양자센서 개발 및 시뮬레이션, 26세(경력 1.6년 / AI 활용 3.2년)] "AI의 답변을 결코 맹신하지 않습니다. 여러 엔진에 반복하여 교차 검증하며 AI와 밀고 당기는 관계를 유지합니다."

① 기술적 파트너십 : AI 의인화·역할 부여



② 상호작용 고도화 : 교차 검증



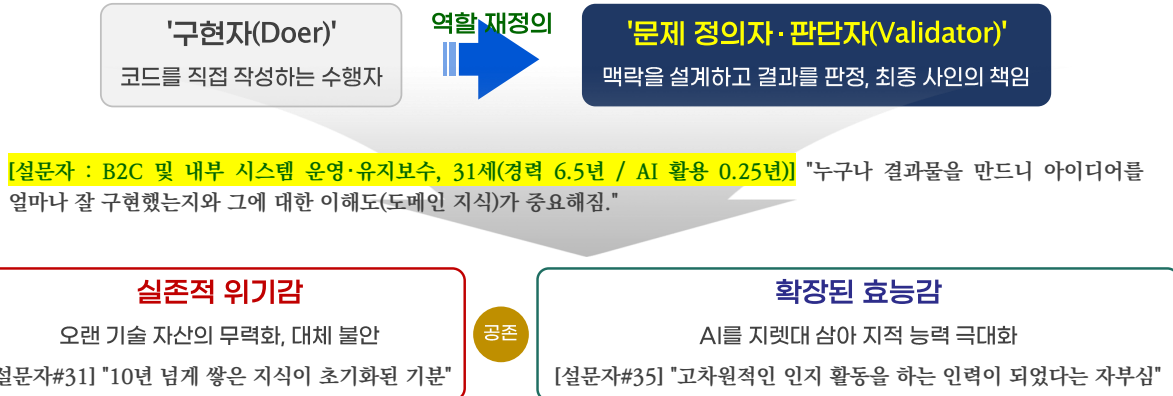
[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28

국립창원대학교 CHANGWON NATIONAL UNIVERSITY 스마트시스템연구실 (Smart Systems Lab.) PYGEEK I · 생성형 AI 시대의 엔지니어 Job Crafting I II III

연구결과 ③ : 인지 크래프팅

① 인지 크래프팅 : 엔지니어 정체성의 재정의를

- 기술적 수행 능력의 하락이 아닌, 「무엇이 나를 전문가로 만드는가」에 대한 근본적 정체성 이행 과정
- 「전문성 가치 재정립」 AI가 대체할 수 없는 도메인 지식과 비판적 사고를 핵심 가치로 상정

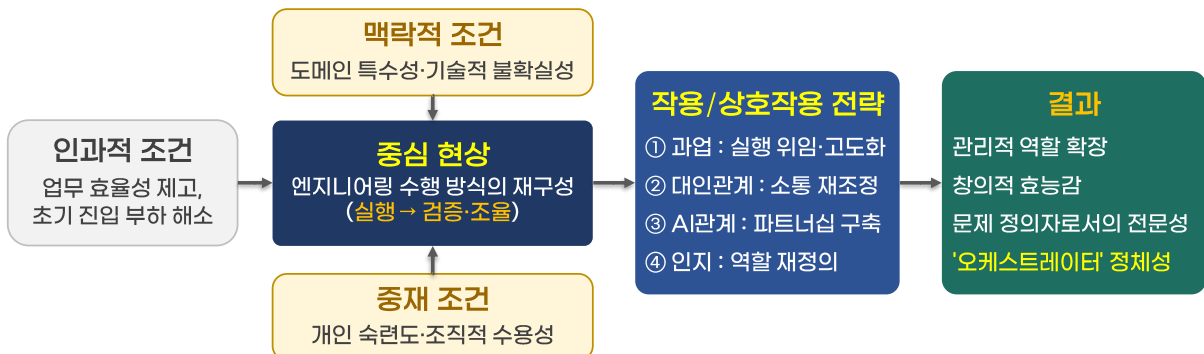


[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28

국립창원대학교 CHANGWON NATIONAL UNIVERSITY 스마트시스템연구실 (Smart Systems Lab.) PYGEEK I · 생성형 AI 시대의 엔지니어 Job Crafting I II III

패러다임 모델과 오케스트레이션 이론

① 패러다임 모델 : 「실행(Doing)」에서 「검증·조율(Validating)」로



「오케스트레이션 기반 '엔지니어 전문성 성장 이론' : 4단계 질적 도약」



[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environments, *Journal of Engineering Education Research*, 29(3), pp.3-28

국립창원대학교
CHANGWON NATIONAL UNIVERSITY

스마트시스템연구실 (Smart Systems Lab.)

PYGEEK

II

실 증 : AI 도입 후 조직 변화

I II III

인공지능(AI) 도입 후 조직 변화

아시아경제 : 2025.11월 / 대상 : 국내 주요 100대 기업

☑️ [가장 먼저 변화된 업무영역]

▸ 문서 초안 작성(39.2%) > 반복 업무 자동화(35.1%) > 고객 문의 대응(14.9%) > 의사결정 지원(10.8%)

☑️ [기존 업무를 AI로 대체 후 인력 변화]

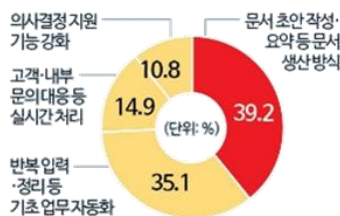
▸ 감독 역할 중심 전환(37.8%) > 재교육 후 새로운 역할 부여(27.0%) > 다른 부서/직무로 이동(17.6%)

☑️ [인력 재배치 시 최우선 고려요소]

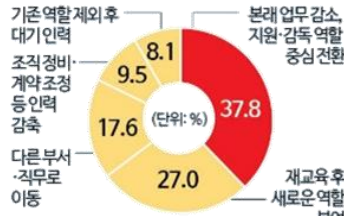
▸ 직무 구조 개편(29.7%) > 새로운 직무 역량 보유(25.7%) > 인력 부족/업무 공백(20.3%)

설문조사 결과

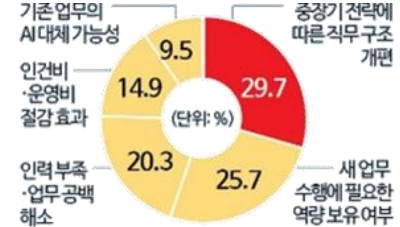
【가장 먼저 변화된 업무영역】



【기존 업무 AI 대체 후 인력 변화】



【인력 재배치 시 최우선 고려요소】



국립창원대학교
CHANGWON NATIONAL UNIVERSITY

스마트시스템연구실 (Smart Systems Lab.)

PYGEEK

II

실 증 : AI 도입 후 조직 변화

I II III

인공지능(AI) 도입 후 조직 변화

아시아경제 : 2025.11월 / 대상 : 국내 주요 100대 기업

☑️ [개인 평가에서 강화된 역량 기준]

▸ AI 결과물 검토/판단(32.4%) > AI 활용 가치/성과 창출(24.3%) > 신기술 역량 습득(21.6%)

☑️ [AI 업무 리스크 판단 기준]

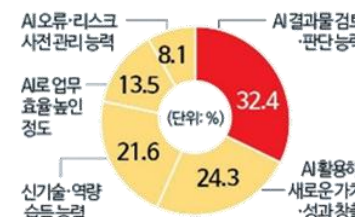
▸ 업무 최종 승인자(33.8%) > 경영진·조직 차원의 책임(25.7%) > AI 시스템 운영/관리 부서(21.6%)

☑️ [AI 사용 시 가장 중요한 윤리 기준]

▸ AI 결과 오류·왜곡 방지(35.1%) > 책임 주체 기준 명확화(28.4%) > 사용과정 기록·추적(20.3%)

설문조사 결과

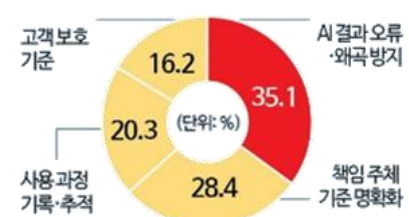
【개인 평가에서 강화된 기준】

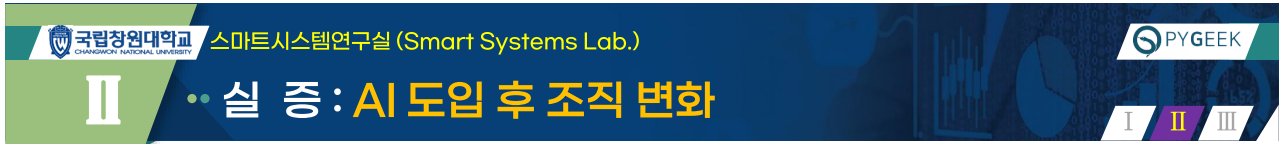


【AI 업무 리스크 책임 판단 기준】



【AI 사용 시 가장 중요한 윤리 기준】





인공지능(AI) 도입 후 조직 변화

아시아경제 : 2025.11월 / 대상 : 국내 주요 100대 기업

☑ [AI 시대 가장 필요한 인재 유형]

▶ 판단·해석 중심 인재(35.1%) > 데이터 해석·의미 도출 인재(24.3%) > AI 이해·활용 인재(21.6%)

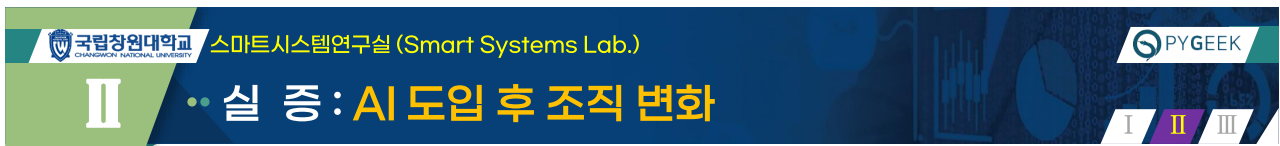
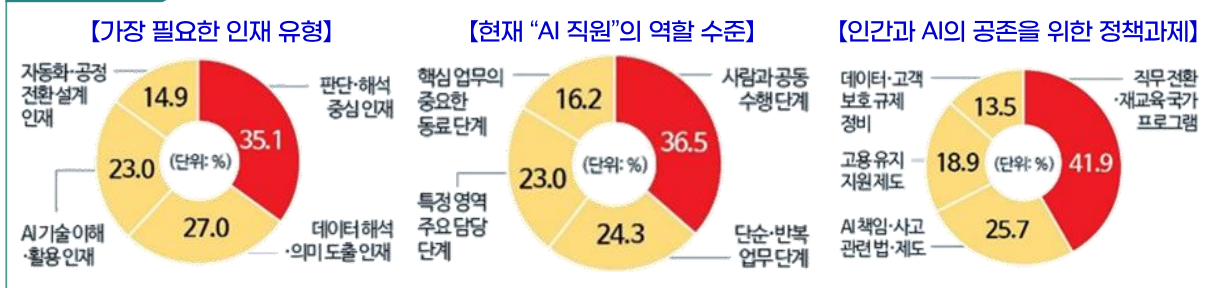
☑ [현재 AI 직원의 역할 수준]

▶ 사람과 공동 수행 단계(36.5%) > 단순·반복 업무 단계(24.3%) > 특정 영역 주요 담당 단계(23.0%)

☑ [인간과 AI가 공존하기 위한 정책과제]

▶ 직무전환·재교육 프로그램(35.1%) > AI 책임·사고 관련 법·제도(28.4%) > 고용유지 지원(18.9%)

설문조사 결과



인공지능(AI) 도입 후 조직 변화

아시아경제 : 2025.11월 / 대상 : 국내 주요 100대 기업

☑ [인문·철학 기반 사고력 중요성 변화] : (높아질 것 : 총 70.2%)

▶ 크게 높아질 것(37.8%) > 다소 높아질 것(32.4%) > 현재와 비슷(23.0%) > 낮아질 것(6.8%)

설문조사 결과



서울경제 (2026. 6. 27)

“문송합니다” 시대 저무나...미국에선 철학과가 컴공보다 취업 잘 된다

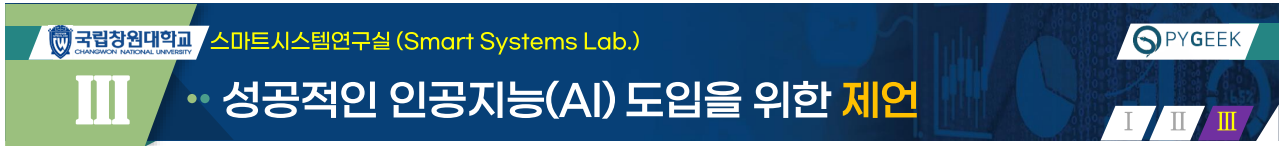
박민주 기자 · 2026. 6. 27 · 10:16

Q 49

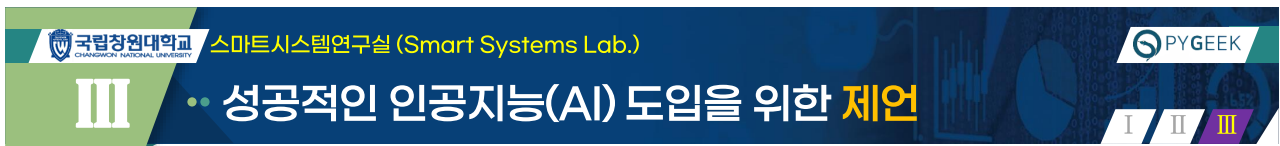
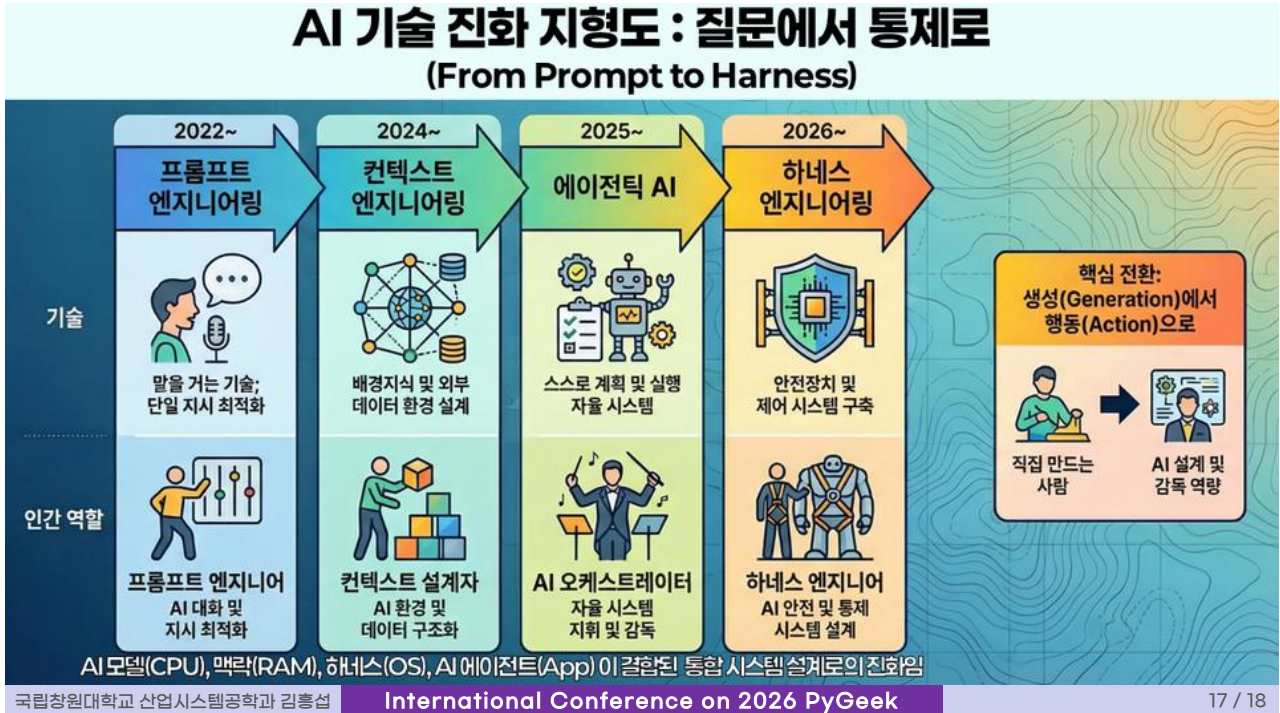
철학 특유 논리적 겸손, AI 오류 줄여 철학과 졸업생·교수, 졸업이 스카우트 안전 중요시하는 클로드, 헌법까지 제정 연쇄를 개선하는 강력한 수단”이라고 평가 AI 규제 강화 속 철학자 위상 높아질 전망

AI 시대, 서점에 쏟아진 인문학의 관심 (한국출판문화산업진흥원)





AI 시대 엔지니어의 핵심 역량



AI 시대 엔지니어의 핵심 역량

☑ (미래) 엔지니어에게 요구되는 핵심 역량

문제 정의 역량 (Problem Definition)

복잡한 현상을 구조화하고 무엇이 본질적인 문제인지 판별

비판적 검증 역량 (Critical Validation)

AI 한각 리스크를 제하고 산출물의 정확성을 도메인 지식으로 판단

AI 오케스트레이션 (AI Orchestration)

AI를 가상의 팀원·파트너로 인식하고 업무를 전략적으로 분담·지휘

도메인 전문성 (Domain Expertise)

기술이 보편화된 시대, 인간 고유의 가치를 보장하는 깊이 있는 현장 지식

☑ 포워드 디플로이드 엔지니어(FDE : Forward Deployed Engineer)

성공적인 FDE가 되기 위한 핵심 역량

T자형을 넘어선 엔드투엔드(End-to-End) 기술력



[Ref.] Chang, K., Jeon, D. and Park, S. (2026) From 'Doing' to 'Orchestrating': A Qualitative Study on Engineers' Job Crafting in Generative AI Environment
Journal of Engineering Education Research, 29(3), pp.3-28



Formalizing Energy-Safe Inference Scheduling for Physical AI under Non-stationary Energy Harvesting

Young-myoung Kang

Sungkyul University · Republic of Korea

Development and Application of AI Solutions in the Semiconductor Industry

Seungwon Seo

GSFSolution · Republic of Korea

GSF SOLUTION

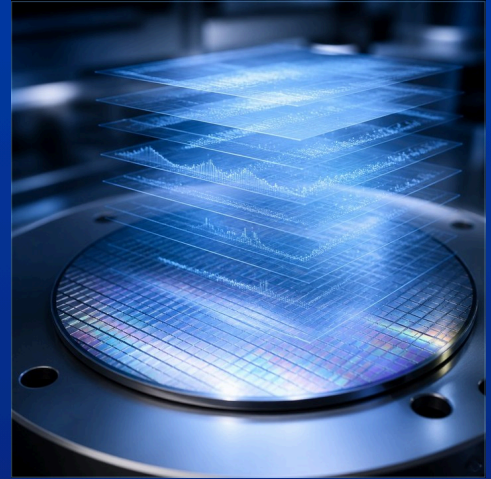
Development and Application of AI Solutions in the Semiconductor Industry

Experience-Grounded Reasoning for Equipment and Process Intelligence

Seungwon Seo · GSF Solution

www.gsfsolution.com

대외 공개본 · PUBLIC RELEASE



GSF SOLUTION

AGENDA

발표 순서

문제에서 시작해 신호 · 경험 · 검증을 거쳐 사업으로 — 6개 파트 · 22개 섹션

PART I 01 - 03

문제 정의

- 01 회사 소개
- 02 문제 정의 · 통합 구조
- 03 산업 구조적 제약

PART II 04 - 05

데이터 기반

- 04 수집 · 통신 기반과 SVID 연동
- 05 P&ID 지능화

PART III 06 - 11

신호화 · 공정 적용

- 06 거동 신호화
- 07 Wobbling 특징량 · Vision 실증
- 08 착수점 이상 탐지
- 09 논문 검증 · 산학협력
- 10 신호 - 공정 해석
- 11 신호-물리 맵 · 예측 파이프라인

PART IV 12 - 16

경험 재사용 · 추론

- 12 사건 정렬 · 경험 재사용
- 13 공정 인식 유사도 · 조치 등급
- 14 가설 검증 · 통합 시나리오
- 15 지식 기반 SENTIX
- 16 하이브리드 추론 스택

PART V 17 - 19

검증 · 최적화 과제

- 17 PEALD 증착 공정 검증
- 18 기술 성숙도 · KPI 검증
- 19 최적화 과제 · 온프레미스 추론

PART VI 20 - 22

사업 · 결론

- 20 시장성 · 구조적 진입장벽
- 21 익명화 · IP 경계
- 22 결론

검증된 항목과 확장 중인 항목을 각 파트에서 구분해 말씀드립니다.

01 회사 소개 About GSF Solution

공정과 장비를 같은 축에서 해석해 이상을 사전에 차단한다

Customized Process & Equipment Solutions for Semiconductor FAB

회사 개요

회사명	GSF솔루션
대표	김승환
설립	2022년 5월 · 사내벤처로 출범
사업 분야	반도체 공정 · 장비 기술 맞춤형 솔루션
거점	본사 충북 청주 · 지사 / 기업부설연구소 경기 성남

공정 레시피와 장비 거동을 동일한 축에서 해석해 온 20년의 현장 경험 — 사후 분석이 아니라 사전 개입을 전제로 솔루션을 설계합니다.

단계별 성장 · Growth Path

1	원천기술 개발	전문가 인사이트 · Hybrid 아키텍처 · 수집 기반	완료
2	실증 · 지식 축적	ETCH · PEALD · Wet Clean 실증 · PHOTO TRACK 확대	진행 중
3	초기 성장	개선품 양산 · 무인화 FAB 진입	진행 중
4	중기 성장	공정 확대 · 플랫폼 서비스	계획
5	고도 성장	해외 진출 · 세정 공정 확대	계획

현재까지 축적된 자산

10건+

특허 출원 · 등록
연구 개발, 인력

3+α공정

식각 · 증착 · 세정

1,000건+

구조화 지식 문서
공정 레시피, 장비

본 발표의 사례는 자사 실증 · 구축 환경에서 확보한 것입니다



02 문제 정의 · 통합 구조 The Problem → Context Fabric

데이터의 양은 해석력을 보장하지 않는다 — 결손된 계층은 문맥이다

장비는 소프트웨어가 아니라 하드웨어 · 공정의 언어로 말합니다 — Equipment speaks in physics, electricity and chemistry.

분리된 다섯 개의 데이터 흐름

- 영상 Vision — 프레임 단위, 고빈도
- 센서 진동 · 변위 · 정전기, 개별 수집
- FDC RF · Gas · Pressure · Temperature
- 알람 이벤트 시퀀스, 코드 체계 상이
- 정비 문서 비정형 텍스트, 사후 기록

다섯 개의 흐름이 서로 다른 시간축과 서로 다른 ID 체계를 가진다.
수집은 이미 되어 있으나 같은 사건으로 묶이지 않는다.

판단 지점

Vision	Sensor	FDC	Alarm	Docs
--------	--------	-----	-------	------

공유되는 문맥 없음

엔지니어의 판단이 멈춘다

다섯 개의 화면과 공유되지 않는 문맥 — 엔지니어가 수동으로 대조합니다.
수집의 문제가 아니라 문맥 결손의 문제입니다.

FDC 단독으로는 원인을 설명할 수 없다

GSF 통합 구조 · CONTEXT FABRIC — 구조에서 시작해 증거 기반 조치로 끝난다

WHERE

장비 구조
Equipment
Topology

WHEN

공정 문맥
Recipe · Step
Chamber State

WHAT

거동 · 신호 특징
Motion & Signal
Features

BEFORE

Expert Case
Memory
Validated Cases

NEXT

증거 기반 조치
Evidence-Based
Action

하나의
수평 흐름

해석 축 — P&ID 한 장에 의지하지 않습니다

배관 위상은 해석의 시작점입니다. 물리 · 전기 · 화학 거동을 함께 읽어야 같은 신호가 Step마다 다른 의미를 갖는 이유를 설명할 수 있습니다.

반복성 — 같은 지점에서 다시 일어납니다

지식 체계는 방대하지만 이상은 무작위가 아닙니다. 같은 지점(Topology) · 같은 Step에서 반복되므로 탐색 범위는 유한한 집합으로 좁혀집니다.

Not another dashboard — a reasoning layer for semiconductor operations.

방대한 지식은 반복되는 지점에서 재사용 가능한 판단 카드로 축적됩니다 — Experience is captured as reusable decision cards.

03 산업 구조적 제약 Structural Constraints

AI 도입을 제약하는 것은 알고리즘이 아니라 산업 구조다

Four structural constraints of semiconductor manufacturing — 우회할 수 없으므로 전제로 설계해야 하는 조건들

01 데이터 폐쇄성 · NO DATA EGRESS

팹도, 장비사도 데이터를 내주지 않습니다

- 팹 차원 — 인터넷 연결이 허용되지 않아 클라우드 AI를 쓸 수 없고, 추론은 장비 PC 연산 안에서 끝나야 합니다 · IP 보안 우려 73%
- 장비사 차원 — 장비 데이터가 공급사 플랫폼 · 포맷에 귀속됩니다. 태그 사전 비공개, 구조화 export 미제공, EDA Freeze Version 불일치

현상 — ① 필요한 신호를 필요한 주기로 확보하기 어렵고 ② 이상의 원인 규명이 공급사 리드타임에 종속되며 ③ 학습에 쓸 수 있는 데이터 범위가 공급사 허용 범위로 축소됩니다

PDF Solutions 2025 · Accenture 2024 · Semiconductor Engineering 2022 · Cimetrix

02 데이터 사일로 · DATA SILOS

통신 표준은 50년째, 의미의 통일은 아직입니다

SEMI 표준화 1970 설립 · 1973 프로그램 출범 · SECS-I 78 · SECS-II 82 · GEM 92 · HSMS 94 · 현재 1,000건+

- 50년간 표준화를 이어왔지만 채택은 자발적이고 인증 체계가 없습니다
- 같은 표준을 지켜도 OEM마다 변수 · 메타데이터 정의가 다르고, EDA Freeze Version이 어긋나면 통신 자체가 불가능합니다

현상 — MES · FDC · SPC · YMS가 각각 저장하고 구형 · 후공정 장비는 표준 밖에 있어, 같은 사건으로 묶이지 않고 챔버 기록은 테스트 시점엔 이미 삭제됩니다

SEMI 1973 · 2019 · Cimetrix · DR Yield · Semiconductor Engineering 2022

03 도제식 전수 · TACIT KNOWLEDGE

판단 기준이 사람에게만 남습니다

- 현장 지식은 감각적 · 문맥 의존적 · 무의식적 — 매뉴얼과 교육 영상으로 포획되지 않습니다
- 2030년까지 반도체 신규 숙련 인력 100만 명 이상 필요
- 미국 반도체 신규 기술직의 80% (약 6.7만 명) 미충원 위험

Dirac 2024 · Deloitte 2022 · SIA 2023

04 희소 이벤트 · RARE FAILURES

고장 사례는 지도학습에 부족하고, 장비 상태는 계속 변합니다

- 센서 1,000만 건 중 라벨된 고장은 37건 — 오퍼팅과 오토퍼로 신뢰를 잃은 실제 사례
- 챔버 상태 drift로 예측 모델 성능이 빠르게 저하됩니다
- AI 가치를 전사로 확장하지 못하는 기업 74%

PDF Solutions 2025 · IEEE T-SM 2019 · BCG 2024

반도체는 SEMI 표준에 힘입어 여기까지 왔습니다 — 장비 구매 1건당 표준 20~25건 인용 · 300mm 팹은 GEM300 사실상 의무 · 표준화는 지금도 진행 중입니다

GSF는 이 제약을 우회하지 않고 설계 전제로 채택했습니다 — 팹 내부에서 데이터를 직접 생성하고, 동일한 축으로 정렬하고, 전문가 판단을 카드로 축적합니다

04 추론의 전제 — 수집 · 통신 기반과 SVID 연동 Acquisition Layer · Data Pipeline

데이터를 직접 생성하고 동일한 축으로 정렬한다

데이터를 직접 만드는 계층

기존 장비 센서
온도 · 압력 · 펄스 · RF · 가스

추가 센서 Install
비전 · 변위 · 정전기 · 진동

설치 환경 분석 → 최적 배치 설계 → 보정

Raw Data + Solution Data

다채널 통신 · 자체 구축

40ch 최대 확장

0.1초 수집 주기

7종 산업 프로토콜

SECS/GEM · PROFINET · EtherCAT · EtherNet/IP · CAN · Modbus · DNP3

NON-SECS 장비도 변환 모듈을 통해 동일 축으로 편입

SVID 정의 5종

SVID 01 와류 · 파동 Data 대조 Tool의 일치도 값

SVID 02 파형의 개수 Data
설정 Edge Level을 초과한 파형 수 Count

SVID 03 X-bar 위치 Tool 대상 물체의 X-axis Position

SVID 04 Y-bar 위치 Tool 대상 물체의 Y-axis Position

SVID 05 편차값 Tool 기준 흔들리는 범위의 편차

적용 솔루션 범위

CHAMBER

Wobbling Nozzle 유량 Nozzle 처짐

Spin Plate 진동 Wafer Notch Suck Back

EFEM · TRANSFER

Blade 처짐 Blade 진동 Wafer 정전기

Chemical Tank 수위

동일한 수집 · 정렬 · 추론 구조를 공정별 문제에 반복 적용합니다.

채널 구조 · 로드맵

1 Set
Sensor 단위

15 SVID
대조 · 엡지 · 폭 X/Y

8 Set = 120
확장 예비 포함

Set당 4개 실측 채널과 11개 Spare 채널로 구성되어 Tool 추가 시 세 설계 없이 확장됩니다.

후속 로드맵

1 Real Data 취합 및 분석 제품 구현

2 FDC LAN 선 포설

3 Nozzle 상태 모니터링 Protocol 개발

4 FDC 연결 및 Data 관리

IP · 구축 이력

특허 출원

Wobbling 감지 (미국 · 대만 · 일본)

용액 토출 감시

장비 부품 모니터링

블레이드 처짐 감시

ESD 관리

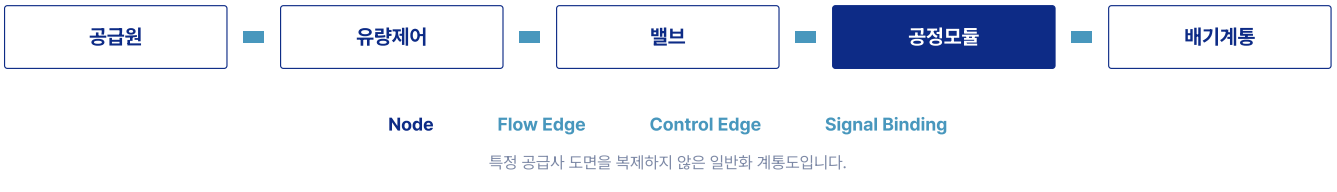
반도체 팹 대상 FDC 구축 및 추가 센서 연동 수행 이력

수집 계층을 직접 보유해 필요한 신호를 필요한 주기로 확보하고, Vision Tool 출력도 FDC SVID 채널에서 동일한 축으로 관리합니다

05 P&ID 지능화 Equipment Map

P&ID를 기계가 해석하는 설비 위상 지도로 변환한다

일반화된 배관 계통도



그래프화 4단계

- ✓ 심볼 · 태그 인식 Symbol / Tag Detection
- ✓ 배관 위상 복원 Line Topology Reconstruction
- ✓ 신호 결합 Signal Binding
- ✓ 전문가 검증 Human Validation

왜 필요한가

A deviation matters only when we know where it occurs and what it can affect.

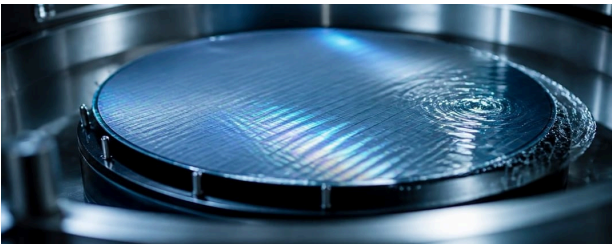
이상이 발생한 지점과 영향 경로를 알면, 연결되지 않은 부분은 원인 후보에서 제외됩니다.



06 거동 신호화 Wobbling · Wet Clean Process

관측되지 않던 거동을 정량 측정 가능한 신호로 변환한다

현상 — 세정 공정 회전 중 미세 흔들림



Wet Clean 공정 — 회전축 편심과 DIW 국소 와류 (설명용 이미지)

신호화 — 중심 궤적 Signature



특정 위상에서 궤적이 반복 확대되고 UCL에 근접한다

실측 — 폭 Tool 거리 편차 (정규화 지수)

구분	측정 기준	정규화 범위	상대 편차
정상 Chamber-A	Cup – Wafer 테두리	Index 1.00 (기준)	약 1.8%
와블링 Chamber-B	Cup – Wafer 윤곽	Index 약 1.20	약 4.0%
와블링 Chamber-B	Cup – 기준 마커 윤곽	Index 약 1.20	약 3.5%

특징량 · Motion Signature

대조(일치도) · Edge(파형 수) · 폭(거리) 3개 Tool 병행
검출 구간: Wafer 최고 위치 2nd Step (DIW), Spin & Dry
속도 · 조명 · 보정 상태로 정규화

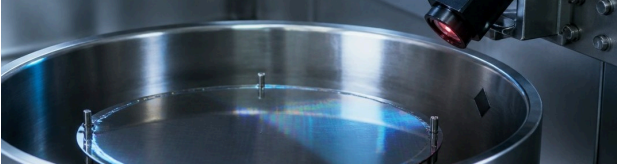
정규화 지수로 표기 · 독립 평가 전



07

Wobbling 특징량과 Vision 실증 Feature Definitions · PoC

센서 설치 구성



IR 센서 · 전용 브라켓 · Chuck Pin · Cup 내벽 기준 마커 — 설명용 이미지

- ✓ 적외선 Type 선정 — 백색광은 Flash가 공정에 영향을 줄 가능성을 배제할 수 없음
- ✓ 전용 브라켓 제작 — 챔버 내 고정 수단이 없어 각도 · 위치 재현성 확보
- ✓ 정상/이상 동시 설치 — Chamber-A(정상) · Chamber-B(와블링)에 각 1 Set, Dummy Wafer 비교 Test

3개 Tool 검출 방식

일치도



대조 Tool

DIW 와류로 발생하는 파동의 일치도 값. 정상은 낮은 범위, 와블링은 80~120 범위로 기준점 분리 가능.

Edge Level



Edge Tool

정상은 잔잔한 파형, 와블링은 Peak가 높은 위치에 형성. 설정 Level을 초과한 파형 수를 Count해 판정.

Cup



폭 Tool

Cup 기준 거리와 편차를 측정. Chuck Pin이 오차를 유발할 수 있어 기준 마커 윤곽 검출을 병행.

대조 Tool을 base로 Edge Count Tool을 결합해 정상 · 와블링을 분리합니다.

특징량 6종

dx / dy

프레임별 중심 이동량

radius

중심 - 외곽 반경 r(t)

P-P 진폭

구간 내 최대 - 최소

RMS

대표 경고 후보 지표

ellipticity

타원률 — 편심 지표

frequency band

주파수 대역 성분

Calibration & 정규화

- 01 특징량 추출 전에 렌즈와 장착 오프셋을 제거
- 02 RPM 설정에 따른 속도 정규화
- 03 조명 변화에 따른 밝기 · 대비 정규화
- 04 동일 임계값을 조건 변화와 무관하게 유지

경고 후보 판단 로직

연속성 단일 프레임 이상은 채택하지 않음

신뢰도 추적 품질이 낮으면 보류

운전상태 가감속 구간은 별도 처리

세 조건을 결합해 경고 후보를 결정

영상 단독 이상은 즉시 장비 고장으로 단정하지 않습니다.

6종 특징량과 3개 Tool로 거동을 좌표 분포로 변환하고, 세 조건을 결합해 경고 후보를 확정합니다



08

착수점 이상 탐지 Wet Clean Landing Point · Vision

고속 회전면의 용액 착수 위치를 통계적 정상 영역으로 판정한다

대상 공정 · 파이프라인

Wet Clean 공정(Clean & CMP 연결 영역) — 고속 회전하는 웨이퍼에 세정 용액을 토출하는 구간으로, 용액 착수점의 위치 편차가 미세 결함으로 이어지는 고위험 구간입니다.

3단 구성

① 고속 영상 획득

고휘도 조명 · 반사 · 회전 대응

② 용액이 닿는 지점 찾기

닿는 지점을 좌표로 변환

③ 판정 · 시각화 · 기록

범위를 벗어난 장면을 바로 표시

경량 구조로 설계 — 스마트팩토리 · 디지털트윈 감시 체계에 통합 가능한 형태입니다.

이상을 판정하는 기준

① 정상만 보고 기준을 만듭니다

속행자가 정상이라고 판단한 토출 장면만 학습합니다. 불량 사례를 일부러 만들 필요가 없어, 이상이 드물게 나는 공정에도 적용됩니다.

시중의 범용 영상 인식 도구는 이 착수점을 찾기 못해, 자사 공정 장면으로 다시 학습시켰습니다.

1,650

지점 직접 표기

8,396

학습한 장면 수

≈900

정상 지점 표본

② 정상 산포가 관리 범위가 됩니다

정상 토출은 용액이 닿는 지점이 한곳에 모입니다 — 그 묶음의 중심과 폭이 공정 허용 범위입니다

③ 범위를 벗어나면 이상으로 봅니다

SPC 관리 한계를 영상에 적용한 것과 같은 개념입니다. 정상 토출 30초에서 닿는 지점 약 900개로 중심과 폭을 정하고, 벗어난 장면은 시각과 위치가 함께 기록됩니다.

Mock-up 완료 · 양산 라인 적용 중

현재 상태와 확장 계획

Mock-up 검증을 마치고 양산 환경 적용 단계입니다 — 추가 확장을 진행할 예정입니다.

남은 기술 과제

① 한 지점을 둘로 잡는 경우 — 학습을 과하게 시키면 착수점을 2개로 잘못 잡습니다. 착수점 자체가 눈으로도 희미한 데서 오는 현상입니다.

② 판정까지 걸리는 시간 — 단독 환경은 짧지만, 실제 라인에서는 서버를 거치는 단계가 늘어 지연이 커집니다.

③ 정밀하게 볼수록 느려집니다 — 정밀도와 속도 중 어디에 맞추는 지는 공정 기준으로 정합니다.

적용 대수가 아직 적어 성능은 수치로 증명하지 않습니다. Wobbling과 동일한 **거동 감시 체계**로 통합할 예정입니다.

본 기술은 동료심사 학술지 게재와 정부 지원 R&D 과제 수행을 통해 검증되었습니다 · 서지 정보와 과제 정보는 대외 공개본에서 생략

거동을 좌표 분포로 변환하는 방식은 Wobbling과 동일하며, 적용 단계만 다릅니다

09 논문 검증 · 산학협력 Peer-Reviewed Validation

산학협력 과제로 수행하고 결과를 학술지에 게재했다

앞 페이지의 착수점 탐지 기술을 산학협력 과제로 수행했습니다 — 주황색 표기는 모두 학술지에 게재된 내용입니다

게재 상태 · PUBLICATION

딥러닝 기반 고속 회전면
용액 착수 이상 탐지 방법

게재 국내 동료심사 학술지 게재 완료
Peer-Reviewed Journal · 서지 정보 비공개

시점 2025년 상반기 게재

수행 대학 연구진 · 교신저자
GSF솔루션 기술부

심사 동료심사 접수 → 수정 → 게재확정
절차 완료

연구비 정부 지원 R&D 과제
과제번호 비공개

논문이 밝힌 것 · FINDINGS

① 기성 도구로는 탐지되지 않습니다
사전학습 상태의 영상 인식 도구 3종을 비교한 결과, 착수점을 **전혀 다른 사물로 오탐**했습니다. 학습 데이터에 이런 대상이 없었기 때문입니
다.

② 착수점은 논문으로도 잘 안 보입니다
위에서 토출하면 용액이 닿는 즉시 퍼져 웨이퍼 표면이 거울처럼 작동합
니다. 그래서 1,650장을 사람이 직접 표기했습니다.

③ 데이터를 늘리면 오히려 틀립니다
8,396장으로 늘려 학습하니 착수점을 2개로 잡았습니다. 최종 모델은
증강 없이 1,650장으로 확정했습니다.

④ 기준은 운영자가 정합니다
정상만으로 만든 통계로 비정상성을 찾는 것은 논리적 모순이라, **허용 폭**
을 운영자가 지정하는 방식을 택했습니다.

시뮬레이션 결과 · TEST

Accuracy	0.99	Precision	0.93
Recall	0.89	F1 Score	0.91

허용 폭 10 조건 · 비정상 장면 50개를 무작위로 섞어 넣고 10회 반
복한 평균값입니다.

논문이 스스로 밝힌 한계

지표는 허용 폭을 어떻게 정하느냐에 따라 달라지므로 절대적인 값
이 아니며, Mock-up은 실제 팜을 완전히 재현하지 못합니다.

함께 구현한 것 — 웹 기반 감시 화면에 허용 범위 기준선과 현재 좌
표를 실시간으로 표시하고, 벗어난 순간을 로그로 남기는 서비스까
지 만들었습니다.

과제 수행 형태

대학 연구진과 산학협력 과제로 수행했고, 고객 정보를 제외한 방법과 실험 결과를 논문으로 정리했습니다.

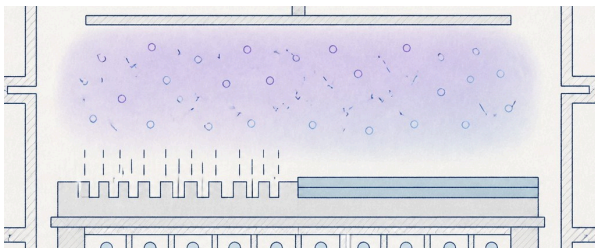
이 수치의 범위

위 지표는 논문의 Mock-up 시뮬레이션 결과입니다. 양산 라인 성능은 적용 대수가 쌓인 뒤에 별도로 보고합니다.

방법과 실험은 논문으로 정리하고, 공정 데이터는 팜 안에 남깁니다

10 신호 · 공정 해석 Plasma · Etch · Deposition

플라즈마 반응을 신호로 해석하면 식각과 증착은 동일한 문법을 갖는다



설명용 도해 — 좌: 이온 충격에 의한 이방성 식각 / 우: 박막 증착

식각 · ETCH [검증 완료]

RF 출력으로 **종횡비(Asspect Ratio)**를 맞추는 것이 식각의 핵
심입니다.
Step별 특징량 → Etch Rate · Uniformity 예측 입력
ER 97.2% · Uniformity 92.1%
DOE 15 Run 독립 평가

증착 · PEALD (MT-NIT) [검증 완료]

Cycle별 특징량 — DCS · NH₃ 유량 · RF Power · Purge
Mode
→ 두께 · 슬롯간 균일도(U/F) 예측 입력
약 85Å σ ± 0.4Å · U/F 약 1.4%
80 LOT 마라톤 · 합격률 100%

공통 문법 · SHARED KEY

Equipment — Chamber — Recipe — Step — Timestamp
물리 상관 **4중** 내재화 — ER ∝ Source Power,
Uniformity ∝ Bias Power
같은 RF · 압력도 Step · Chamber 이력이 다르면 의미가 달라
집니다.

실증 수집 규모 — Plasma Etch Class-A

70여 Active Field 4천여 Step 900여 Recipe Pair 1,200+ 수집 태그

GSF interprets a signal by its physical role, timing and equipment history.

11 ETCH 신호-물리 맵과 예측 파이프라인 Signal-Physics Map · Prediction Pipeline

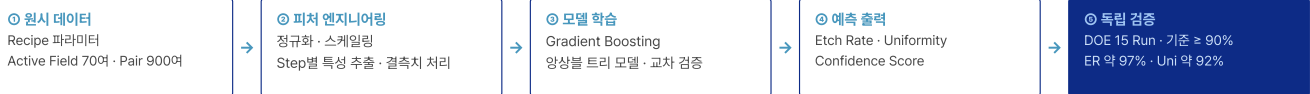
신호에 물리 역할을 부여하면 예측은 검증 가능한 수치가 된다

신호 · 물리 역할 대응

신호 그룹	물리적 역할	Step 내 의미	파생 특징량
RF Source / Bias	이온 에너지 · 종횡비(Aspect Ratio) 제어	점화 · 안정화 구간 구분	Setpoint-Actual, Settling
Gas / MFC	Reaction supply	Step별 조성 전환 시점	전환 지연, 유량 편차
Pressure / APC	Pressure balance	배기 · 공급 균형 상태	Drift, 밸브 위치 변화
ESC / Temperature	Thermal state · 적 상태	누적 이력에 따른 기준 이동	Chamber History 보정
Event Log	상태 전이 기록	Step 경계 확정 근거	Event Sequence

같은 값이라도 Recipe Step이 다른면 정상 · 이상 판정이 달라집니다. 화학식과 상세 Recipe 조건은 노출하지 않습니다 · OES / VI Probe는 본 구축 범위의 성과로 주장하지 않습니다.

회귀 모델 기반 가상 공정 예측



Etch Rate 예측 성능

약 97%

정확도 = 100% - MAPE
(MAPE 약 2.8%)

MAE	약 1.7 Å/min	오차 절대값 평균
RMSE	약 2.2 Å/min	큰 오차 민감 지표
R ²	약 0.89	결정계수
15 Run 중 14 Run APE ≤ 10% · 해당 14 Run 평균 APE 약 1.9%		

Uniformity 예측 성능

약 92%

정확도 = 100% - MAPE
(MAPE 약 7.9%)

이상치 제외	약 97%	특이 조건 1 Run 제외 시
MAE	약 0.08 %p	절대오차 평균
RMSE	약 0.17 %p	이상치 영향 포함
특이 조건 1 Run(APE 약 71%) 포함한 값을 대표 수치로 보고		

신호에 물리 역할과 Step 의미를 부여해 특징량을 구성하고, 성능은 독립 평가셋으로만 보고합니다



12 사건 정렬 · 경험 재사용 StepSync → Expert Case Memory

정렬된 사건은 검색 가능한 경험 자산으로 축적된다

01 정렬 · StepSync



02 저장 · Event Envelope

- 01 기준 구간 Baseline
- 02 전조 구간 Precursor
- 03 사건 구간 Event
- 04 회복 구간 Recovery
- 05 검증된 원인 · 조치 · 결과

Envelope 한 건 = Case 한 건.
검색 키는 Context(장비 · Recipe · Step)와 Symptom(특징량 백터)입니다.

03 재사용 · 4R Cycle

① Retrieve

유사 사례 검색

② Reuse

가설 · 조치 재사용

③ Retain

검증본만 축적

④ Revise

현장 검증 · 수정

③ Retain → ① Retrieve로 재순환

CASE 레코드 구성



상태 관리 · LIFECYCLE

Draft → Reviewed → Validated → Retired

엔지니어가 Validated로 승인한 Case만 다음 추론에 인용됩니다.

The unit of intelligence is not a data point. It is a verified, retrievable event.



13 공정 인식 유사도와 안전 조치 등급 Similarity · Safe Action Ranking

유사도는 경험을 인출하고, 조치 등급은 안전 제약이 결정한다

유사도 구성 · 5요소

Similarity = Process Context + Topology + Trace Shape
+ Alarm Sequence + Maintenance History

Process Context	Recipe Family · Step · Chamber 상태가 같은 조건인가
Topology	P&ID 상에서 영향 경로가 연결되어 있는가
Trace Shape	파형의 형상 · 기울기 · 상승 패턴이 유사한가
Alarm Sequence	알람 발생 순서와 조합이 일치하는가
Maintenance History	직전 정비 · 교체 이력이 유사한가

가중치 정책 · 보정 방식

가중치 정책
고정 가중치는 공개하지 않습니다 — project-specific parameter입니다.

보정 방식
장비군 · 공정군별 검증 데이터와 전문가 판정 결과로 보정합니다. 유사도는 경험을 불러오는 역할이며, 적용 여부는 물리와 안전이 결정합니다.

두 Confidence의 분리

Detection Confidence

이상이 존재한다는 판단의 확신도

Action Confidence

권고 조치가 지금 적용 가능하다는 확신도

화면의 Confidence는 모델 정확도가 아닙니다

점수 구성 요소

- + Evidence Coverage
- + Context Match
- + Topology Consistency
- Counter-evidence
- Uncertainty Penalty

조치 등급화

추가 데이터 요청 증거가 부족해 판단을 보류하는 단계

확인 권고 기본 출력 원인 후보와 확인 순서를 제시하는 단계

자동 조치 제한적 허용 Validated Case와 Safety 제약을 모두 만족할 때만

Validated 상태이고 유효기간 내인 Case만 Action Candidate로 사용합니다

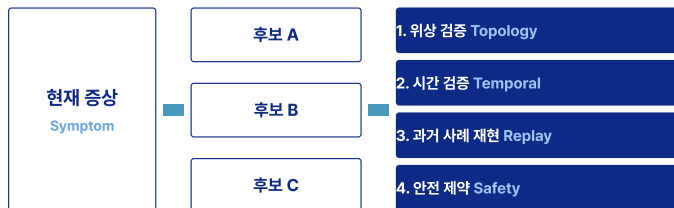


14 가설 검증과 통합 시나리오 Evidence Reasoner · Scenario

단일 추정을 내놓지 않고 경쟁 가설을 순차 검증한다

CANDIDATES

VERIFICATION GATES



우선순위 조치

- 01 근거가 남은 원인 후보
- 02 후보를 지지하는 증거
- 03 권고 확인 순서
- 04 금지 조치 목록

AI의 출력은 원인 확정이 아니라 근거 있는 확인 순서입니다.
연결되지 않은 부품, 비활성 Step, 안전 충돌 조치는 게이트에서 제외됩니다.

통합 시나리오 · 하나의 사건이 추론 루프 전체를 통과하는 8단계



No recipe change or automatic restart without verified evidence.

기술 설명용 합성 시나리오입니다 — 실제 고객 사례가 아닙니다.

Similarity retrieves experience. Physics and safety decide whether it applies.

15 지식 기반 · 전문가 인사이트 SENTIX Knowledge Base

데이터 반출이 불가능한 환경에서는 지식을 내부에 선축적이다



설명용 이미지 — 패세망 내부에 구축된 SENTIX 지식 서버

구조화 축적 규모

228문서 · 2.96 GB

장비 지식 — 매뉴얼 · 정비 · 절차

388문서 · 667 MB

공학 지식 — 식각 · 플라즈마 · 진공 · FDC

5공정군 스키마

Oxidation · Diffusion · CVD · Etch · Vacuum

문서를 그대로 쌓지 않습니다. **장비 지식**과 **공학 지식**을 분리 색인하고, 현장 용어 사전과 시각 용어 사전으로 같은 부품을 같은 이름으로 정규화합니다.

축적된 지식의 형태

이중 지식베이스

Manual — 운전 · 점검 · 안전 절차
Theory — 현상의 물리 메커니즘

물리 계산 도구 7종

Arrhenius · Deal-Grove · 평균자유행로
체류시간 · 쉬스 전위 · CVD 성장속도 · 확산길이
물리 질문은 추정하지 않고 계산합니다.

전문가 인사이트의 역할

- ① 무엇을 볼지 정한다 — 피쳐 선정과 가중치
- ② 어떤 순서로 확인할지 정한다 — Check Sequence
- ③ 무엇을 하면 안 되는지 정한다 — Prohibited Action
- ④ 언제 사람에게 넘길지 정한다 — Escalation Rule

AI는 후보와 근거를 제시하고, 판단의 경계는 전문가가 정합니다.

왜 빠르게 답하는가

질문을 **의도별로** 분기합니다 — 절차 · 정의 · 시각 · 물리
키워드 검색과 의미 검색을 **결합**하고,
속연도에 따라 매뉴얼과 이론의 비중을 바꿉니다.
변경된 문서만 재색인하는 **중분 구조**입니다.

패세망 안에서 동작 — 데이터를 밖으로 내보내지 않습니다.

지식 검색 · 물리 도구 · 의도 라우팅은 구현 완료 · 현장 Case 연결은 확장 중

SENTIX는 Expert Case Memory의 자료 참고입니다



16 하이브리드 추론 스택 Neuro-Symbolic

하이브리드 추론은 탐지와 판단을 분리한다



Pattern recognition detects a deviation. Domain reasoning decides what it means.

가설 점수 — 개념식

$HypothesisScore(h) = PatternMatch$
+ ProcessContextMatch + TopologyConsistency
+ TemporalConsistency + VerifiedCaseEvidence
- ContradictionPenalty - UncertaintyPenalty
가중치는 장비군 · 공정군별 project-specific parameter이며,
피쳐 선정과 가중치는 전문가 지식을 기반으로 설정됩니다.

Release Gate

- ✓ 필수 증거 Coverage 충족
- ✓ P&ID 경로와 모순 없음
- ✓ 해당 Step에서 활성
- ✓ 반대 증거 없음
- ✓ Safety 제약과 충돌 없음
- ✓ Case Validated · 유효기간 내



17 PEALD 증착 공정 검증 MT-NIT · Golden Recipe · 조건값 비공개

실험 시리즈 · 4단계

단계	Run	핵심 결과
① DOE1	21	변수 감도 · Arrhenius 거동
② DOE2 (ToF)	19	Golden Recipe 확정 ToF 영향 없음 ($-0.4\text{\AA}$) 약 $85\text{\AA}$ $\sigma \approx \pm 0.4\text{\AA}$
③ 80 LOT 마라톤	80	U/F 약 1.4% · 합격 100% 약 $84\text{\AA}$ · U/F 약
누적 123 Run — 양산 유효성 검증		

사이클 수 변경 시 두께 차이 $< 0.2\text{\AA}$ — 설계된 ALD 공정 창 안에서 Self-limiting 거동 유지 확인

변수 감도 · 튜닝 우선순위

1순위 · FINE 전구체 유량

선형 $R^2=0.98$ · 유량에 선형 비례 — 두께 미세 조정

2순위 · COARSE 온도

Arrhenius · 온도에 지수 비례 ($r=0.95$) — 고온 구간 과민성 주의

3순위 · U/F RF Power

저출력 U/F 불량 → 일정 출력 이상 Plateau ($r=-0.92$)

4순위 · 슬롯간 Purge 모드 최적화

U/F 약 0.8% — 기본 대비 70% 감소 · CVD-like 기상 반응 억제

확정 GOLDEN RECIPE · 조건값 비공개

온도 · 전구체 유량 · 반응가스 유량
RF 출력 · Purge 모드 조합 확정 → 목표 U/F $\leq$ 약 0.8%

고유량 조건에서 U/F 약 0.7%로 최우수 — 표준 조건은 별도 기준으로 유지

이상 감지 기준 · Control Limit

일계 기준	우선 점검 항목
두께 하한 이탈	RF 인터락 · 플라즈마 상태 NO PLASMA 시 두께 급감
두께 상한 이탈	전구체 라인 압력 · 사이클 설정값
U/F $> 2.0\%$	RF 출력 하한 미만 / 상한 초과 Purge Mode 3 확인
F2 Clean 직후 두께 상승 구간	챔버 시즈닝 — 정상 패턴 2 LOT 더미 런 후 양산 진입

이상 2건의 원인 계층 분리

LOT-A 통신 끊김 → 두께 급감

LOT-B 전구체 투입 이상 → 과증착

두 건 모두 장비 전기 · 제어계 이상으로 분류 — 공정 레시피 자체의 이상이 아님을 확인했습니다.

이상 2건의 원인은 레시피가 아니라 장비 제어계였습니다 — 원인 계층의 분리가 판정의 전제입니다

18 기술 성숙도 · KPI 검증 Validation Result

성능은 독립 평가 기준으로만 보고하고, 검증 공정과 확장 공정을 구분한다

식각 · ETCH — 계획 대비 실적 (독립 평가)

KPI 항목	계획 기준	실적	판정
AI 예측 정확도 (Etch Rate)	$\geq 90\%$	약 97%	충족
AI 예측 정확도 (Uniformity)	$\geq 90\%$	약 92%	충족
이벤트 감지율	$\geq 95\%$	약 98%	충족
장비 가동률	$\geq 95\%$	약 95%	충족
데이터 무결성	$\geq 95\%$	약 99%	충족
실시간 데이터 수집 태그	≥ 60 개	76개	충족

6개 항목 전 항목 충족 · 실증 규모 70여 Active Field · 4천여 Step · 900여 Recipe Pair
데이터 무결성은 캘리브레이션 주기 단축 · 불안정 구간 필터링으로 약 99% 확보

검증 4단계 · VALIDATION PROTOCOL

- 독립 평가셋 — 학습 구간 배제 · 현장 측정값을 정답으로 사용
- 지표 산출 — MAE · RMSE · MAPE · R^2 · 정확도 = $100\% - \text{MAPE}$
- 운영 지표 — False Alarm / Lot · Top-k Diagnosis · MTTR
- 재현 가능성 — 계산 스크립트와 원본 데이터 제출 · Leakage 차단

증착 · PEALD (MT-NIT) — 양산 검증

두께 재현성 · 유효 LOT 기준

약 $85\text{\AA}$ 편차 $1\text{\AA}$ 미만

평균 균일도 U/F

약 1%대 목표 2% 이내

U/F 2% 이내 합격률

100% 전 LOT 통과

Self-limiting 마진

LOOP 변경 · 계측기 장착에도

두께 변동 $1\text{\AA}$ 미만

Golden Recipe 고정 조건 · 120여 Run 연속 진행으로 양산 재현성 확인

Confidence ≠ 독립 평가 정확도

화면 표시값 · Confidence Score

모델 내부 신뢰도 — 학습 패턴과의 일치 정도, 실측값 대비 정확도를 보장하지 않습니다.

보고 기준 · 독립 평가 정확도

DOE 15 Run 현장 측정값 대비 MAPE로 산출 — 약 97% / 약 92%의 근거입니다.

두 값을 구분해 말하는 것이 성능 주장의 전제입니다.

확장 중 · IN PROGRESS

세정(Wet Clean) Wobbling 검출

Vision · SVID 연동 구현 완료 — 독립 평가셋 확보 중

포토(PHOTO) TRACK 장비 확대

동일 신호 체계 이식 진행 중 — 도포 · 이송 거동 대상

챔버 · Recipe 전반의 일반화

장비군을 넘는 적용 범위 판정 기준 정립 중

확장 단계도 같은 기준 적용

식각 · 증착과 동일한 독립 평가 절차를 그대로 사용

식각 · 증착에서 검증한 동일 신호 체계를 Wet Clean과 포토 TRACK으로 확장하고 있습니다



19 최적화 과제 — 스마트함과 속도 On-Prem Optimization

대형 LLM의 판단 품질을 폐쇄망 장비 PC 안에서 재현한다

The constraint is not model size — it is judgment quality per unit of on-premise compute.

관건 · THE CHALLENGE

엔지니어가 기대하는 수준은 대형 LLM의 스마트함과 대화형 응답 속도입니다.
그러나 팍은 인터넷이 차단되고 추론은 장비 PC의 제한된 연산에서 끝나야 합니다 — 모델을 키우는 방향으로 풀 수 없는 문제입니다.

연산 한도

장비 PC · 폐쇄망
클라우드 사용 불가

실시간성

공정은 초 단위
사후 분석은 무의미

근거 필수

확률적 답변만으로는
조치 승인 불가

01 구조 최적화

큰 모델 하나가 아니라 역할을 분리합니다

탐지는 신경망이, 판단은 문맥 그래프와 근거 검증이 맡습니다. 단계를 쪼개면 각 단계는 작은 모델로 충분해집니다.
거대 모델의 추론을 흥내내지 않고, 판단 경로 자체를 설계합니다.

구현 완료 — 하이브리드 스택

02 지식 최적화

파라미터 대신 검색으로 지식을 담니다

질문을 의도별로 라우팅하고(절차 · 정의 · 시각 · 물리), 키워드 검색과 의미 검색을 결합합니다.
변경된 문서만 증분 재색인하므로 지식이 늘어도 응답 경로는 늘지 않습니다.

구현 완료 — SENTIX

03 모델 최적화

공정별로 필요한 최소 모델을 고릅니다

예측은 Gradient Boosting 앙상블, 영상은 RT-DETR 경량 구조로 구성했습니다.
정확도와 복잡성의 트레이드오프는 공정 기준에 맞춰 선택합니다 — 항상 큰 모델이 정답은 아닙니다.

공정별 진행 중

04 실행 최적화

응답 시간을 설계 변수로 다룹니다

수집은 0.1초 주기 · 최대 40ch로 이미 실시간입니다.
남은 구간은 End-to-End Latency — 서버-클라이언트 다단 경유를 줄이고 판단 경로를 장비 측으로 당기는 작업입니다.

과제 — 수치 확보 전

앞으로의 개선 과제 · NEXT

자가 진단 엔진 · Self-Diagnostic

이상이 난 뒤 찾는 대신, 장비가 주기적으로 자기 상태를 점검합니다. 정상 baseline 대비 열화 추이를 스스로 판정해 사전 경고 지점을 확보합니다.

응답 지연 예산

공정별 허용 지연을 정의하고
구간별 예산으로 분해합니다

모델 크기 상한

장비 PC 연산 한도 안에서
운용 가능한 상한을 확정합니다

재학습 주기

дре프트에 대응하는
자동 재학습 주기를 설계합니다

다음 목표는 이상을 빠르게 찾는 것이 아니라, 이상이 생기기 전에 장비가 스스로 알리게 하는 것입니다 — 사후 대응에서 사전 감지로

20 시장성과 구조적 진입장벽 Market & Moat

시장 규모 · 성장률

AI 예지보전 · 글로벌

49조 원 · CAGR 32.6%

2035년 전망 · TAM
Bizwit Research & Consulting

아시아 · 국내

15조 / 7.6조원

SAM 35.1% · SOM 34.0%
2035년 전망 · 같은 출처

제조 AI 중 반도체 성장률

34.7% CAGR

제조 산업군 중 최고
2024-2035 · 같은 출처

300mm Fab 장비 투자

\$133B 2026 · +18%

2027 \$151B → 2029 \$172B
SEMI, 2026-04

전방 시장이 커지는 것과 우리 매출은 다른 문제입니다. 우리가 파는 것은 **장비 대수 × 공정별 비중**으로 환산되는 팍 내 설치 단위입니다.

문제의 비용 — 왜 지금인가

비계획 정비 · 직접 손실	\$1.5~1.8M / 시간	Siemens · Aberdeen
숨은 비용 포함 총비용	\$5~9M / 시간	같은 출처
예지보전 도입 효과 (벤치마크 평균)	정지 최대 50% ↓	SNS Insider
AI 적용 Scrap 감소 사례	5% → 2%	학술 서베이

시간당 손실은 펌 노드 · 시설 가치 · 환산 기준에 따라 크게 흔들립니다. **직접 손실과 총비용을 구분해** 인용해야 하는 수치입니다.

구조적 진입장벽

- Ⓞ 물리적 접근 — 팍 클린룸 진입 자체가 통제됩니다.
- Ⓞ 데이터 반출 금지 — 공정 · 장비 데이터는 외부로 나가지 않습니다. 외부에서 학습한 모델로는 시작할 수 없습니다.
- Ⓞ 도메인 해석 — 접근이 가능해도 전문가 없이 신호의 의미를 판정할 수 없습니다.

GSF는 팍 안에서 데이터를 직접 만들고, 폐쇄망 안에서 지식을 축적합니다 — 세 장벽의 안쪽에서 출발합니다.

진입장벽은 우리를 막는 벽이 아니라, 이미 진입한 쪽을 보호하는 벽입니다

21

익명화와 IP 경계 IP Boundary

대체 명칭 규칙

공급사명	Equipment Vendor-A
고객사명	Customer Fab-A / Pilot Site-A
장비명	Plasma Etch Class-A / PEALD Class-B Wet Clean Class-C / Photo Track Class-D
Chamber	Chamber-A / B / C
Recipe	Recipe Family-R1 / R2
Gas · 전구체명	Reactive Gas-A / B · Precursor-P1
Tag · Alarm	GSF Canonical Tag / Alarm Family-A
실제 수치 · Lot	Normalized Index · LOT-### / WAFER-##

공급사 매뉴얼의 문장 · 표 · 도면은 캡처하거나 재사용하지 않습니다.

범용 기반

Sensor · FDC · SPC · FFT
XGBoost · Retrieval · Knowledge Graph

공급사 매뉴얼은 도메인 근거로만 사용합니다.

GSF 고유 통합

Domain Ontology · Canonical Signal Dictionary
StepSync & Event Envelope
Wobbling / Etch / Deposition / Wet Clean / Track Feature Set
Expert Case Memory · Process-Aware Similarity
Evidence Reasoner · Safety Gate
Expert Feedback and Validation Workflow

고객 · 장비 · Recipe · Tag 정보는 공개하지 않습니다.

GSF SOLUTION

22 · 결론

범용 기술 위에 축적한 도메인 지식이 GSF의 IP다

알고리즘 하나가 아니라, 공정의 시간 · 물리 · 경험을 연결하는 페루프 운영지식입니다.

FOUNDATION · 범용 기반

Sensors
FDC · SPC · FFT
Machine Learning
Retrieval
Knowledge Graph

누구나 접근 가능한 공통 기술



GSF DOMAIN IP · 우리가 만든 계층

Equipment Topology
Process Signal Dictionary
장비 구조와 신호 사전

Expert Case Memory
검증된 사례만 축적하는 경험 저장소

StepSync &
Event Envelope
사건 단위 정렬과 저장

Evidence Reasoner
& Safety Gate
근거 검증과 안전 제막

Motion · Etch · Deposition
Wet Clean Feature Sets
공정별 특징량 체계

Expert Feedback Loop
전문가 판단이 다시 자산이 되는 순환

축적된 자산 · PROOF

10건+ 특허 출원 · 등록

Wobbling · 토출 · 처짐 · ESD
미국 · 대만 · 일본

3+α 공정 적용

식각 · 증착 · 세정 검증 + 포토 TRACK 확장 중

1,000건+ 구조화 지식

장비 지식 · 공학 지식 분리 색인

폐쇄망

안에서 축적
웹 안에서 데이터를 직접 만들고 학습

GSF makes semiconductor AI remember context, evidence and outcomes.

GSF는 데이터를 보는 AI를 넘어, 공정의 시간 · 물리 · 경험을 기억하는 반도체 도메인 AI를 만듭니다.

감사합니다 — 질문은 각 파트의 검증 자료로 바로 되돌아가 답변드리겠습니다.

Accepted Papers

A Co-evolutionary Strategy of AI Technology for Expanding Regional Industrial Value Chains

Eu jung KIM¹

PhD. Candidate, Industry-academic Convergence Talent Center, Hongik University, South Korea

Corresponding author: Eu jung KIM(eujeje@hongik.ac.kr)

Abstract

This study introduces a co-evolution framework designed to substantively expand the value chains of regional industries through organic integration with artificial intelligence (AI) technologies. While AI continues to advance rapidly, regional small and medium-sized enterprises (SMEs) consistently face significant barriers to tech adoption, primarily driven by critical shortages in both infrastructure and specialized personnel. Moving beyond the limitations of unidirectional technology transfers, this research constructs a bidirectional growth model by leveraging hub cities with established manufacturing infrastructures as optimal testbeds. Specifically, the paper examines mechanisms through which regional strategic sectors—namely, materials, components, equipment (MCE) and mobility—can deploy data analytics and machine learning models to maximize operational efficiency across the entire spectrum of the value chain, from initial planning to production and distribution. Concurrently, it outlines a virtuous technical cycle wherein empirical data generated directly from industrial sites serves to continuously refine and advance the AI algorithms themselves. To ensure successful technological internalization, we propose an industry-academia matching platform as a core governance structure, dedicated to cultivating practical AI talent and integrating them directly into real-world corporate projects. Ultimately, the proposed co-evolution strategy not only generates new economic value for regional industries but also establishes a highly scalable and adaptable standard innovation model applicable to other municipalities and industrial clusters.

Keywords: Regional Industry, Artificial Intelligence, Value Chain Expansion, Co-evolutionary Strategy, Machine Learning

1. Introduction

Artificial intelligence (AI) has recently transcended mere technological progress to establish itself as a core driving force that fundamentally transforms paradigms across entire industries. Furthermore, the widespread adoption of Python-based machine learning and open-source AI algorithms is significantly enhancing the efficiency of corporate manufacturing processes, thereby driving the creation of new added value[1]. However, the benefits of such technological advancements and innovative outcomes tend to be disproportionately concentrated within metropolitan areas and large enterprises. Consequently, this trend exacerbates the divide in AI adoption and digital transformation for regional small and medium-sized enterprises (SMEs) operating with relatively inadequate infrastructure. In

particular, the materials, components, and equipment (MCE) as well as the future mobility sectors—which serve as the central pillars of national strategic industries—are facing profound external pressures driven by the restructuring of global supply chains and rapid technological advancement. Consequently, due to insufficient capabilities in utilizing field data and a critical shortage of specialized personnel, these regionally based enterprises are currently struggling to achieve innovation across their entire value chains, from initial product planning through to production and distribution[2]. Therefore, to secure the self-sustaining viability of regional industries and foster sustainable growth, there is an urgent need for a novel approach that organically integrates advanced AI technologies in alignment with the specific characteristics of the regional industrial ecosystem.

To date, local government-led technology support initiatives and vendor-driven AI adoption models have predominantly leaned toward unidirectional "technology transplantation." The approach of indiscriminately deploying externally developed, general-purpose AI solutions or merely matching hardware to regional enterprises has failed to account for the unique manufacturing processes and operational characteristics of the field, ultimately resulting in substantial underutilization and the obsolescence of the introduced technologies[3]. Furthermore, in the absence of internal personnel capable of sustainably operating and refining these technologies, the digital transformation (DX) of supply chains has frequently devolved into one-off pilot initiatives. This has created a fundamental disconnect between two essential pillars: the axis of innovation, where technology transforms the industry, and the axis of development, where empirical field data and human capital mature the technology itself. Consequently, this fragmentation has manifested in a clear limitation—an inability to achieve a substantive expansion of the value chain relative to the resources invested.

To overcome the limitations inherent in this unidirectional technology transplantation, this study proposes a "co-evolution" strategy wherein AI technologies and the regional industrial ecosystem interact bidirectionally to foster mutual growth. By designating a hub city equipped with a manufacturing-based empirical infrastructure as a testbed, this research analyzes how Python-based data analytics and machine learning models can optimize efficiency across the entire value chain. Furthermore, to ensure successful technological internalization, we present a governance model centered on an industry-academia matching platform. This platform is designed to cultivate practical, AI-proficient talent and closely integrate them with the real-world projects of regional enterprises. Ultimately, the co-evolution strategy derived from this study aims not only to generate new added value for regional industries but also to provide a standardized framework that can be flexibly adapted across various municipalities and industrial clusters.

2. Theoretical Background

2.1 Regional Industrial Value Chain and Artificial Intelligence Transformation

The value chain refers to a series of value-adding activities through which a company plans, designs, produces, markets, and distributes its products or

services[4]. Amid the rapid restructuring of global supply chains, the DX and integration of AI technologies across the entire value chain are essential for the survival of regional SMEs and the MCE industries. Previous studies have demonstrated that integrating machine learning and data analytics into traditional manufacturing-based clusters can substantially enhance productivity through process optimization and predictive maintenance[5]. However, the artificial intelligence transformation (AX) of regional industries is inherently limited if it remains confined to merely transplanting high-performance equipment or general-purpose solutions. A genuine expansion of the value chain is only fully realized when the unique domain knowledge of the industrial field is organically combined with flexible open-source AI algorithms, thereby enabling the autonomous derivation of enterprise-specific predictive models[6]. Therefore, moving beyond piecemeal technology adoption, a structural approach that organically integrates data across all stages of the value chain is required.

2.2 Techno-Social Co-evolution Theory

The concept of co-evolution has been extensively employed in the fields of macro-organizational theory and the sociology of technology to elucidate the dynamic interactions in which technology and the environment—or technology and organizations—continuously evolve together as mutually reciprocal causes and effects[7]. According to research in technological innovation, for a novel technological system to be successfully embedded within a society or a regional industrial cluster, the evolution of the technology itself must be accompanied by the simultaneous co-evolution of the surrounding institutions, infrastructure, and the capabilities of the adopting actors[8]. From the perspective of the regional industrial ecosystem, co-evolution with AI technology is predicated on a bidirectional virtuous cycle. The high-quality data generated from the specialized empirical infrastructure and manufacturing sites of hub cities serves as a vital resource for optimizing AI algorithms and advancing models (Region → Tech). In turn, these sophisticated, field-tested AI models drive innovation within the value chains of regional enterprises and generate high added value (Tech → Region) [9]. Consequently, this study proposes this co-evolutionary mechanism as the core framework for regional industrial innovation.

2.3 Cultivation of Practical AI Talent and Industry-Academia Collaboration

Ultimately, the critical bridge that sustains the co-evolution of AI technologies and regional industries is

human capital. The primary reason many local governments fail in technology adoption lies in the absence of internal specialized personnel capable of sustainably operating, modifying, and advancing general-purpose solutions[10]. Furthermore, recent studies demonstrate that practical, application-oriented AI education—which focuses on solving real-world projects in the industrial field—is significantly more effective in achieving regional technological internalization than traditional, theory-centric higher education[11-12]. In particular, field-tailored AI training programs—encompassing everything from Python-based data manipulation to no-code automation tools—are gaining significant traction as a viable solution to rapidly elevate the AX capabilities of SME incumbents and young talent [13]. To prevent the outflow of these trained professionals from the region and ensure their skills directly contribute to expanding the value chains of local enterprises, it is imperative to establish a self-sustaining, regional university-led industry-academia matching governance. This governance structure must be capable of mediating, in real time, the identification of problems within the supply chain and the corresponding supply of specialized talent..

3. Proposed Framework

3.1 Bidirectional Co-evolution Framework

The co-evolution framework proposed in this study is designed as a bidirectional cyclical structure wherein the regional industrial value chain (planning, production, and distribution) and the AI technology ecosystem continuously interlock and interact. To break away from the limitations of unidirectional solution adoption, enterprises continuously feed empirical data—generated during the production processes and product planning stages—into data analytics pipelines. By training on this data, AI models provide actionable feedback to enterprises in the form of process optimization and demand forecasting, thereby expanding the value chain. For this technological cycle to be sustained autonomously, a practical, project-based industry-academia matching platform—which connects field-ready talent with enterprises—must be integrated as an essential governance mechanism.

3.2 Value Chain Optimization and Application in Manufacturing

The technological axis of this framework lies in directly embedding machine learning models into the practical

operations of manufacturing-oriented enterprises, specifically within key regional strategic industries such as MCE and future mobility. For instance, at the R&D planning stage—the frontline of the value chain—chemical and materials companies developing novel metal oxide-based materials can leverage Python libraries like Scikit-learn or TensorFlow to analyze experimental data. By constructing predictive models that incorporate variables such as precursor mixing ratios and heat treatment temperatures, these enterprises can drastically reduce the costs associated with repetitive experiments and maximize efficiency during the new product planning phase. Furthermore, in the mobility sector, large-scale sensor and traffic data collected during the empirical testing of autonomous driving can be preprocessed using Pandas to refine driving anomaly detection models. This constitutes an empirical case wherein the infrastructure of regional enterprises serves directly as a testbed for AI algorithms.

3.3 Cultivation of Practical Talent and Matching Platform

To proactively secure the human resources that will serve as the core driving force to sustainably lead this framework, this study proposes a two-stage industry-academia collaboration governance led by regional universities. First, it is necessary to support incumbents and regional talent by introducing new curricula focused on the utilization of no-code platforms and AI agents, departing from conventional theory-centric education. This will empower them with the capabilities to perform task automation and data analytics even without prior software development knowledge. Second, regional universities must function as project-based matching platforms, enabling local talent to directly solve the immediate challenges faced by enterprises. This approach not only alleviates the labor shortages of these companies but also connects regional youth with high-quality jobs they may have otherwise overlooked, ultimately strengthening the foundational AI capacity of the entire regional industry.

When this process establishes an iterative virtuous cycle, this governance model emerges as the most viable solution to simultaneously address the dual challenges of preventing youth talent outflow and successfully establishing AX in SMEs. Furthermore, its significance extends beyond a localized success story; it possesses robust scalability as a "Region-Customized Technology-Talent Co-evolution Standard Model" that can be flexibly adapted to the specific on-site conditions of other municipalities and industrial clusters currently facing population decline and industrial stagnation.

4. Conclusion

This study proposed a "Technology-Talent Co-evolution Framework" designed to overcome the concentration of AI technological benefits in metropolitan areas and large corporations, and to substantially expand the value chains of regional SMEs struggling with relatively fragile infrastructure. Recognizing the limitations of unidirectional technology transplantation traditionally led by local governments or supplier enterprises, we engineered a bidirectional virtuous cycle structure wherein empirical data from regional industrial sites and machine learning models continuously interact to foster mutual growth.

In particular, focusing on the region's key strategic industries—such as MCE and future mobility—this study presented a practical approach to minimize R&D costs and maximize process efficiency by applying data collected during enterprises' planning and production stages to AI predictive models and anomaly detection algorithms. Furthermore, the research elucidated that a novel industry-academia collaboration governance, with regional universities at its core, is essential to ensure that such technological innovations do not remain isolated, one-off projects but rather achieve self-sustainability within the region. A matching platform that cultivates field-ready talent through practical, application-oriented education based on no-code tools and AI agents—and subsequently connects them with the real-world data projects of enterprises—serves as the core driving force to simultaneously achieve technological internalization and human resource acquisition.

In conclusion, the co-evolutionary governance proposed in this study emerges as the most viable solution to prevent the outflow of trained youth talent and drive the successful AX of local SMEs. Its scalability and policy implications extend far beyond the infrastructure of any single hub city; it serves as a customized "Technology-Talent Co-evolution Standard Model" that can be flexibly adapted to the on-site conditions of various municipalities and industrial clusters currently grappling with the dual hardships of population decline and industrial stagnation. For future research, we plan to apply this framework to the actual operational challenges of regional enterprises, analyze the resulting quantitative performance metrics, and, based on these findings, verify the feasibility of expanding the framework's application to other industrial sectors.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

Acknowledgement

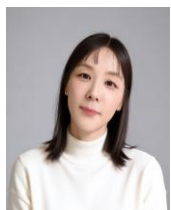
The authors would like to express their sincere gratitude to the members of the DeepShark Laboratory for their valuable support and assistance throughout this research. This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency (NIPA) (Grant No. PJTS0801260410 070001000300200).

References

- [1] Kwon Young-hwan, "Open Source AI Activation Method to Secure Global AI Competitiveness", SPRI Research Report, RE-202, 2026. <https://spri.kr/posts/view/23990?code=research>
- [2] Korea Chamber of Commerce and Industry, "Current Status of Corporate AI Transition and Improvement Measures", K-Growth Series (7), 2025.
- [3] M. S. Park., "The Real Reason 80% of Industrial AI Adoption Fails", ETNews, 2025-12-29. <https://www.etnews.com/20251229000073>
- [4] M. E. Porter, "Competitive Advantage: Creating and Sustaining Superior Performance", The Journal of Business Strategy, 1985.
- [5] Vlad Krotov, "Beyond AI automation: Redesigning organizations for human-AI synergy", Science Direct, 2025. <https://doi.org/10.1016/j.bushor.2025.06.007>
- [6] C. S. Lee, "User-Modified Artificial Intelligence Causal Prediction Model for Process Improvement Using Design Document Interface", J. Korean Soc. Manuf. Technol. Eng. 34:1 (2025) 64~71., 2025 <https://doi.org/10.7735/ksmte.2025.34.1.64>
- [7] Han Se-eok, "Co-evolution of Technology and Institutions in the Era of the Fourth Industrial Revolution", Proceedings of the Seoul Association of Public Administration Academic Conference, 2018.
- [8] Frank W. Geels, "Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study", Research Policy, Volume 31, Issues 8–9, 2002. [https://doi.org/10.1016/S0048-7333\(02\)00062-8](https://doi.org/10.1016/S0048-7333(02)00062-8).
- [9] K. J. Arrow, "The economic implications of learning by doing, Review of Economic Studies", vol. 29, no. 3, pp. 155-173, 1962. <https://doi.org/10.2307/2295957>

- [10] H. S. Nam, M. S. Ahn, “Research of Artificial Intelligence Usage in Public Sector”, SPRI Research Report, RE-156, 2023.
https://spri.kr/posts/view/23653?code=data_all&study_type=research
- [11] J. H. Han, Y. J. Oh, “Exploring the Institutionalization of Industry-Linked AI Education Policy: Policy Frame Redesign through the Case of Gwangju Metropolitan City”, Asia-pacific Journal of Convergent Research Interchange Vol.12, No.1, January 31 (2026), pp.1-11, 2026
<http://dx.doi.org/10.47116/apjcri.2026.01.01>
- [12] Jin Hoe-seung, Yoon Bo-seong, “Analysis of Korea’s AI Education System and Implications for Developing Practical AI Talent”, SPRI Issue Report, IS-223, 2026
https://spri.kr/file_html/is223/index.html
- [13] Kim Ju-mi, Kim Mi-jeong, “A Study on the Developing Strategies to Strengthen DX/AX Capabilities of Regional Small And Medium Manufacturing Enterprises”, KOSI Policy Research Report, 25-07, 2025
<https://db.kosi.re.kr/kosbiDB/front/subjectResearchDetail?dataSequence=B260410K02&issueID=f80253ce81d14f489d8a78d18787de41#>

Author Information



Eu-jung Kim
2023: MS, Smart City & Urban
Regeneration Converge, Hongik
University
2024~now: Ph.D. Candidate, Smart
City & Convergence, Hongik
University

Email: eujeje@hongik.ac.kr

Research Interest: Environmental Chemistry,
Adsorbents, Nanomaterials, Water Treatment, Toxicity
Assessment

A Factor Decomposition Method for Separating Store Substitution and Embedding Placement Effects in Vector Database Migration

Junha Yoon¹, Ohhyeon Gwon², Junhaeng Lee³, Hyeji Roh³, Soowan Cho⁴, Suhee Kim⁴, and Sangsoon Lim⁵

¹BS Course, Department of Media Software, Sungkyul University, South Korea

²BS Course, Department of Computer Engineering, Sungkyul University, South Korea

³MS Course, College of Art & Technology, Chung-Ang University, South Korea

⁴Deep Insight Lab, South Korea

⁵Professor, College of Art & Technology, Chung-Ang University, South Korea

Corresponding author: Sangsoon Lim (slim@cau.ac.kr)

Abstract

When the vector layer of a retrieval-augmented generation (RAG) system is migrated to a different database, the effect is usually assessed by an end-to-end comparison before and after the migration. That comparison, however, blends two independent but co-occurring changes into a single measurement, namely substituting the store engine and relocating where the embedding is computed, so it cannot tell which factor caused the change. This paper introduces a bridge condition, in which the target store is loaded with the embeddings the source system has already produced, so that the entangled comparison splits into two comparisons that each differ in one factor only. It requires no re-embedding, so it is inexpensive and leaves the running service untouched. Applied to a Korean product-review RAG service in production (1,347 chunks, 43 queries), the store substitution showed no detectable change in retrieval quality (top-10 Jaccard 0.971, relevance difference -0.042 , $p = 0.523$), whereas relocating the embedding into the database produced a significant, medium-sized effect ($p = 0.003$, Cohen's $d_z = 0.473$). Virtually all of the observed change therefore comes from the placement factor. The method does not judge any particular product and applies to migrations to converged databases in general.

Keywords: Retrieval-Augmented Generation, Vector Database, In-Database Embedding, Migration, Factor Decomposition

1. Introduction

A retrieval-augmented generation (RAG) system consists of an embedding model that turns queries and documents into vectors, a store that indexes and searches those vectors, and a language model that generates an answer grounded in the retrieved results [1], [2]. Among these, the component changed most often after deployment is the vector layer. A representative path leads from a general-purpose relational database with a vector extension to a converged database that has a vector type and an inference runtime inside it, that is, one in which the database itself performs the embedding computation.

The effect of such a migration is usually assessed by an end-to-end comparison that weighs the whole system before and after the change at once. In this comparison, however, two factors change at the same time. The first is the store, that is, the engine that holds the vectors and the indexing method it uses. The second is the placement, that is, whether the embedding is computed by an external service or by an inference runtime inside the database. When a migration is motivated by governance, moving the embedding inside as well is the preferred configuration, and a database that hosts a model imposes a model size limit. This is why the two factors change together in practice.

As a result, the end-to-end difference alone does not reveal which factor caused the change. When quality drops, it cannot be told apart whether the store or the

model is at fault, and when nothing changes, it cannot be known whether the two factors cancelled each other out. In either case the conclusion cannot be carried over as is to an environment with a different instance class or model size limit.

This paper inserts an intermediate condition, the bridge condition, that loads the target store with the embeddings the source system has already produced, and thereby splits one comparison in which the two factors are entangled into two comparisons that each differ in one factor at a time. We apply the proposed method to a Korean review-analysis service migrated from a PostgreSQL vector extension to a converged database that supports in-database embedding, and measure it on 1,347 review chunks from the production corpus and 43 fixed queries. This paper does not judge which product is superior, and the latency figures reflect the measurement environment rather than the products.

2. Proposed Method

2.1 System Architecture

The target system retrieves Korean product reviews and generates answers that cite the retrieved evidence. Both architectures operate in four stages: query input, query embedding, vector search, and answer generation. Retrieval is a hybrid scheme that combines vector similarity with metadata filters and source-based prioritization, and this rule and its parameters were kept identical across the three conditions. Because both are

expressed as a single SQL statement, the store and the embedding placement can be swapped while the application logic is left as it is.

Fig. 1 shows the Oracle-based architecture. The application sends a single SQL statement carrying the query string (step 1), and the database converts the query into a 384-dimensional vector with its built-in Open Neural Network Exchange (ONNX) inference runtime (step 2) and then, within the same SQL statement, performs an exact search over the VECTOR column together with metadata filtering (step 3). The language model cites the returned top- k chunks to generate an answer (step 4).

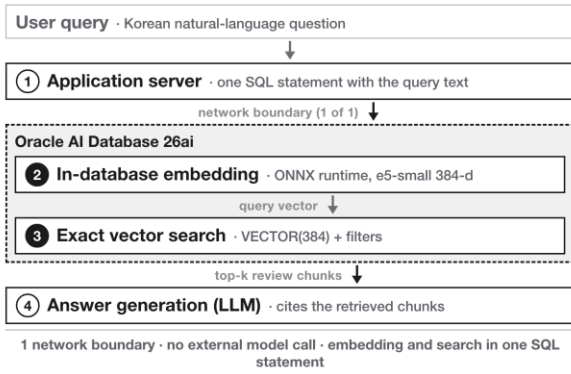


Fig. 1. Oracle-based system architecture

Fig. 2 shows the PostgreSQL-based architecture. The application calls an external embedding service to convert the query into a 1024-dimensional vector (step 2) and then requests the search with that vector as an argument (step 3). The store explores candidates with a hierarchical navigable small world (HNSW) approximate index and applies the same filters, and the remaining steps are as in Fig. 1.

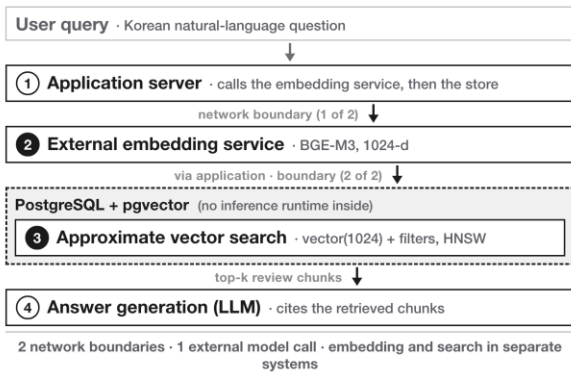


Fig. 2. PostgreSQL-based system architecture

The two architectures differ in two things at once, the store and the embedding placement. One is which engine holds the vectors and with what indexing method, and the other is whether the embedding is computed outside or inside the database. An end-to-end comparison alone cannot tell which of the two produced the difference. The next two subsections summarize the characteristics of

each architecture from these two viewpoints, and Section 3 presents a way to separate them.

2.2 Characteristics of the Oracle-Based Architecture

First, it provides a native VECTOR type and a built-in ONNX inference runtime, so embedding and search are combined into a single SQL statement [3]. Second, it performs an exact search that compares every vector, without an approximate index that narrows candidates in advance, so no correct result is missed because of the index. Third, the network boundary is reduced to one, and because the query and the retrieved results stay inside the database, controls such as encryption and auditing apply to the entire retrieval path. Fourth, a built-in model must satisfy the model size limit the database allows, and in this environment the largest multilingual model within that limit was multilingual-e5-small (384 dimensions) [4].

2.3 Characteristics of the PostgreSQL-Based Architecture

First, it provides vector search through an extension module (pgvector) and contains no inference runtime [5]. The embedding model can be chosen separately from the store, so there is no size constraint, and this architecture uses the 1024-dimensional BGE-M3 [6]. Second, it uses an HNSW approximate index that narrows candidates as it searches, so search time grows gently as the data grows, at the cost of missing some correct results [7]. Third, because the embedding is computed externally, the query text leaves the store boundary and one more network round trip is required.

3. Performance Evaluation

3.1 Experimental Setup

In this experiment we denote a retrieval configuration by the triple (S, E, P) , where S is the store, E is the embedding model, and P is where the embedding is computed. Comparing the two architectures of Section 2 as they are amounts to comparing $(S_0, E_0, \text{external})$ with $(S_1, E_1, \text{in-database})$; all three coordinates differ, so the contribution of each factor cannot be separated. We therefore add $(S_1, E_0, \text{external})$, a bridge condition that uses the target store but is loaded with the vectors the source system produced, and the three conditions are summarized in Table 1. The bridge condition needs no recomputation of the vectors. Because the columns in production are left as they are and only one more vector column is created and filled, its preparation cost is a single bulk load and the running service is not touched.

Table 1. Experimental conditions

	A (baseline)	B (bridge)	C (migrated)
Store	PostgreSQL + pgvector	Oracle 26ai	Oracle 26ai

Embedding model	BGE-M3 (1024)	BGE-M3 (1024)	e5-small (384)
Placement	external	external	in-database
Vector search	HNSW approx.	exact	exact

A→B changes only the store factor, and B→C only the placement factor. Oracle 26ai denotes Oracle AI Database 26ai [3], and e5-small denotes multilingual-e5-small [4].

In condition C, however, moving the placement also changes the model, because the model size limit of Section 2.2 makes it impossible to host the source model inside the database. Choosing in-database placement means accepting that limit as well, so this coupling was left in place on purpose. Separating the two would require one more condition, (S_l , E_l , external), which runs the smaller model outside the database, and we leave this to future work.

The corpus consists of 1,347 review chunks from a single product domain of the production service, loaded identically into both stores after verifying in advance that the chunk identifiers correspond one to one. The query set was fixed at 43 Korean queries, 13 drawn from production logs and 30 synthetic queries written so as to cover evenly the evaluation aspects and purchase-journey stages the service handles.

Three metrics are measured. The first is the agreement of the retrieved results: how much the top-10 results the two conditions return for the same query overlap (Jaccard), how similar their rankings are (Spearman rank correlation), and, taking the exact search of the bridge condition as ground truth, how many of those the baseline recovers (Recall@10). The second is the relevance of the retrieved evidence: a large language model scores the top-5 chunks on a 0–2 scale [8], and 215 judgments per condition are averaged per query and then compared in pairs across the same queries. The third is the search latency, the actual time taken to process a single query.

The target database ran on the lowest-tier free instance (Always Free, 2 GB RAM) and the external embedding service on a local laptop (Apple M5 Pro). Confidence intervals (CIs) were obtained by bootstrap resampling of the measurements 10,000 times.

3.2 Experimental Results

The measurements by factor are summarized in Table 2.

Table 2. Results by factor

Metric	A → B (store)	B → C (placement)
Top-10 Jaccard [95% CI]	0.971 [0.946, 0.992]	0.175 [0.133, 0.221]
Spearman rank correlation	0.958	—
Recall@10 (vs. exact search)	0.984	—
Relevance score (0–2)	1.228 → 1.270	1.270 → 1.014
Paired difference [95% CI]	−0.042 [−0.172, 0.084]	0.256 [0.098, 0.423]
Significance / effect size d_z	$p = 0.523$ / −0.098	$p = 0.003$ / 0.473
Mean search latency (ms)	150.2 → 110.0	110.0 → 947.0
External calls / boundaries	1→1 / 2→2	1→0 / 2→1

Relevance is the per-query mean over the 43 queries (215 judgments per condition), and p values are from a paired t -test.

Looking first at ranking agreement, the two stores loaded with identical vectors returned exactly the same top-10 set for 37 of the 43 queries, and even the minimum was 0.667. The results the baseline missed because of its approximate index amount to about 1.6%.

For the store factor, the paired difference in relevance scores across the same queries was not statistically significant and the effect size was negligible ($t(42) = -0.643$). A Wilcoxon test on the 28 queries that remain after excluding those with identical scores leads to the same conclusion ($p = 0.908$).

Since “no difference” is itself our claim, we did not rely only on a non-significant p value but tested equivalence separately [9]. Two one-sided tests examining whether the difference falls within ± 0.15 points gave $p = 0.052$: the lower bound of the confidence interval, -0.172 , falls outside this range, so the significance level was not met. We therefore report not that equivalence has been established but that no effect was detected and that its magnitude is bounded to a small range. Since the direction in which the interval extends is the one where the target store scores higher, there is no basis for concluding that substituting the store degraded quality.

By contrast, between B and C, which share the same store and the same search method, the results differ greatly. The top-10 Jaccard falls to 0.175, so mostly different evidence is retrieved for the same query, and the relevance score decreases from 1.270 to 1.014. This difference is significant with a medium effect size ($t(42) = 3.104$), and a Wilcoxon test on the 38 queries that remain after excluding those with identical scores leads to the same conclusion ($p = 0.006$). Because no store effect was detected in the change from A to B, virtually all of the end-to-end quality change is due to the placement factor and the model size limit that comes with it. Had only an end-to-end comparison been performed, the roughly 20% drop in relevance would have been blamed on the store, and reverting the store, the remedy that follows from that diagnosis, would not bring the quality back.

For search latency, the ratio of C to B was 8.27 times at the median (Wilcoxon $p < 0.001$). The increase in C is not a property of in-database embedding itself but a consequence of running inference on a free-tier instance, so it cannot be generalized to higher-tier hardware. Since A and B produce the same vectors with the same external service, the query embedding time was measured once and applied to both conditions, and the latency difference between them therefore reflects only the store search time.

Finally, because 30 of the 43 queries are synthetic, we checked whether they behave differently from the 13 real queries. Relevance scores were systematically lower on the synthetic subset (A: 1.147 vs. 1.415, $p = 0.045$; C:

0.880 vs. 1.323, $p = 0.007$; Mann–Whitney), so the absolute level of relevance must not be read as service quality. In contrast, ranking agreement, the basis on which the cause is attributed, showed no significant difference between the two subsets ($p = 0.441$, $p = 0.570$). Absolute scores depend on where the queries come from, but the conclusion about the cause is the same on both.

4. Conclusion

Adding a single bridge condition makes it possible to split an end-to-end comparison into two single-factor comparisons. Applied to a Korean review RAG service in production, no effect of the store factor was detected, and the entire observed change originated from the placement factor. Changing the placement is a trade in governance rather than in performance. In exchange for keeping the query and the review text inside the database boundary, one has to accept the model size limit and the inference performance of the database tier. Therefore, if governance of the stored data is the only goal, replacing the store alone is enough, and if the query path must also stay inside the boundary, one should first confirm that the largest model that can be hosted is sufficient for that domain. The bridge condition makes that confirmation possible before the migration is carried out.

The limitations are as follows. The study targets a single Korean corpus of 1,347 chunks, human assessors were not involved in the relevance scoring, and the measurements were taken on an entry-level instance. In addition, as stated in Section 3.1, placement and model size are coupled by construction, and the result for the store factor does not establish equivalence.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

References

- [1] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459-9474, 2020.
- [2] Y. Gao et al., “Retrieval-augmented generation for large language models: A survey,” *arXiv preprint arXiv:2312.10997*, 2023.
- [3] Oracle, “Oracle AI Vector Search User's Guide,” Oracle AI Database 26ai documentation, Available: <https://docs.oracle.com/en/database/oracle/oracle-database/26/vecse/>, 2026, [Accessed: Aug. 8, 2026].
- [4] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, and F. Wei, “Multilingual E5 text embeddings: A technical report,” *arXiv preprint arXiv:2402.05672*, 2024.
- [5] pgvector, “pgvector: Open-source vector similarity search for Postgres,” GitHub repository, Available: <https://github.com/pgvector/pgvector>, 2026, [Accessed: Aug. 8, 2026].
- [6] J. Chen, S. Xiao, P. Zhang, K. Luo, D. Lian, and Z. Liu, “M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation,” *Findings of the Association for Computational Linguistics: ACL*

2024, Bangkok, Thailand, pp. 2318-2335, <https://doi.org/10.18653/v1/2024.findings-acl.137>, Aug., 2024.

- [7] Y. A. Malkov and D. A. Yashunin, “Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 824-836, <https://doi.org/10.1109/TPAMI.2018.2889473>, 2020.
- [8] L. Zheng et al., “Judging LLM-as-a-judge with MT-Bench and Chatbot Arena,” *Advances in Neural Information Processing Systems*, vol. 36, Datasets and Benchmarks Track, 2023.
- [9] D. Lakens, “Equivalence tests: A practical primer for t tests, correlations, and meta-analyses,” *Social Psychological and Personality Science*, vol. 8, no. 4, pp. 355-362, <https://doi.org/10.1177/1948550617697177>, 2017.

Author Information



Junha Yoon

2026 ~ now: B.E. Candidate, Media Software, Sungkyul University
Email: wnskg314@gmail.com
Research Interest: Human-Computer Interaction, Applied AI, RAG



Ohhyeon Gwon

2026 ~ now: B.E. Candidate, Computer Engineering, Sungkyul University
Email: ruud5521@gmail.com
Research Interest: Large Language

Models (LLM), Data-driven Analysis, Applied AI



Junhaeng Lee

2026: B.S., Computer Engineering, Sungkyul University
2026 ~ now: M.S., Dept. of Applied Art and Technology, Chung-Ang University

Email: hkhk9495@cau.ac.kr

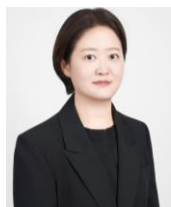
Research Interest: Natural Language Processing, Faithfulness in Natural Language Generation, Retrieval-Augmented Generation, Large Language Model-based Agent Systems



Hyeji Roh
2021: BE, Computer Engineering,
Sungkyul University
2021 ~ now: Integrated MS-PhD, Dept.
of Applied Art and Technology,
Chung-Ang University

Email: shgpw1509@cau.ac.kr

Research Interest: AIoT, AI, Physical AI, Deep
Learning-based Remote Sensing Analysis



Soowan Cho
2004: BS, Mathematics, Sogang
University
2007: MA, Consumer Psychology,
Ewha Womans University
2008 ~ 2019: Senior Researcher,

Global Research Company (HANKOOK, IPSOS)
2019 ~ 2026: Director of PR & Marketing, Education
Industry

2026 ~ now: COO, Deep Insight Lab

Email: soowanc@naver.com



Suhee Kim
2004: BA, Business Administration,
Sookmyung Women's University
2009: MA, International Business with
Marketing, University of Exeter, UK
2006 ~ 2019: Senior Researcher,

Global Research Company (KANTAR, IPSOS)
2019 ~ 2021: Manager of Consumer Research, Global
Manufacturer

2025 ~ now: CEO, Deep Insight Lab

Email: shkim8161@gmail.com



Sangsoon Lim (IEEE Senior Member)
2013: Ph.D, Electrical Engineering and
Computer Science, Seoul National
University
2013 ~ 2017: Senior Engineer,
Samsung Electronics

2017 ~ 2025: Assistant Professor, Dept. of Computer
Engineering, Sungkyul University

2023 ~ 2024: Visiting Professor, Dept. of Electrical and
Computer Engineering, the University of British
Columbia

2025 ~ now: Associate Professor, College of Art &
Tech., Chung-Ang University

Email: slim@cau.ac.kr

Research Interest: AIoT, AI, ML, Image Analysis,
Intelligent Computing

Multi-scale Temporal-aware Text-Enhanced Network for Multimodal Sentiment Analysis

Junjun Zhang¹, Giseop Noh²

1. School of Software, Henan University of Engineering, Zhengzhou 451150, China

2. Department of Software and Communications Engineering, Hongik University, 2639 Sejong-ro, Jochiwon-eup, Sejong-si, 30016, Republic of Korea

Corresponding author: Giseop Noh (e-mail: kafa46@hongik.ac.kr)

Abstract

Multimodal Sentiment Analysis (MSA) has achieved remarkable progress by leveraging textual, visual, and acoustic cues. However, adequately modeling the dynamic, fine-grained temporal dependencies both within and across modalities in long video sequences remains a formidable challenge. Existing Transformer-based fusion networks, while adept at capturing cross-modal interactions, often fall short in comprehensively extracting and integrating sentiment evolutions across different temporal scales. Building upon the prevailing paradigm of text-enhanced Transformer fusion networks, this paper addresses their limitation in modeling fine-grained temporal dependencies in long video sequences by proposing a novel Multi-scale Temporal-aware Speech-Enhanced Transformer Fusion Network (MSTETFN). Specifically, we introduce: 1) a Multi-scale Temporal Sequence Learning Module (MTS-TLM) to capture sentiment variations at different temporal granularities; 2) a Hierarchical Temporal Attention Mechanism (HTAM) to dynamically weight and aggregate these multi-scale temporal features; and 3) a Temporally-Aligned Text-Enhanced Transformer (TA-TET) to integrate these temporally rich representations into cross-modal interactions. Extensive experiments conducted on the benchmark datasets CMU-MOSI and CMU-MOSEI demonstrate that MSTETFN significantly outperforms current state-of-the-art methods, particularly for long video segments and complex sentiment dynamics.

Keywords: Multimodal Sentiment Analysis; Temporal Modeling; Multi-scale Feature Learning; Hierarchical Attention

1. Introduction

The proliferation of multimedia content on social platforms has propelled Multimodal Sentiment Analysis (MSA) into a research hotspot. Unlike unimodal approaches, MSA integrates complementary cues from text, audio, and visual modalities to achieve a more comprehensive understanding of human affective states, a process that mimics the human cognitive mechanism of expressing emotions through language, facial expressions, and vocal intonation. Early MSA research primarily focused on fusion strategies, evolving from simple concatenation to decision-level integration [1], with subsequent advancements in attention mechanisms and tensor fusion networks [2–3] further enhancing performance. The advent of the Transformer architecture [4], particularly its self-attention and cross-attention mechanisms, has enabled Transformer-based MSA models to effectively model complex intra-modal and inter-modal relationships, yielding significant progress.

Among various approaches, the text-enhanced Transformer fusion network has garnered considerable attention by recognizing the central role of text in emotional expression. Text typically conveys the most explicit sentiment information; such methods augment non-verbal modalities through text-oriented cross-modal mappings, while incorporating unimodal label prediction to preserve modality-specific characteristics. However, despite their notable fusion performance, these methods remain deficient in modeling fine-grained temporal dependencies and multi-scale sentiment evolution within long video sequences.

Human affect is inherently dynamic, evolving at varying speeds and intensities over time. Emotions may shift abruptly within seconds—such as a sudden burst of laughter—or gradually unfold over longer durations, as seen in sustained irony. Accurately capturing these different temporal granularities is crucial for sentiment analysis in real-world scenarios. Existing Transformer

models typically process sequences at a fixed temporal resolution or rely on generic temporal encoders, making it difficult to simultaneously capture instantaneous reactions and overarching affective trajectories. This limitation is particularly pronounced when facing subtle or delayed sentiment transitions, often resulting in an incomplete understanding of affective dynamics.

To address this critical issue, we propose a novel Multi-scale Temporal-aware Text-Enhanced Transformer Fusion Network (MSTETFN). This network directly embeds sophisticated multi-scale temporal awareness into the feature learning and fusion processes. Our main contributions are as follows:

- (1) Multi-scale Temporal Sequence Learning Module (MTS-TLM). We introduce parallel temporal convolutional networks with varying kernel sizes and dilation rates. This module independently processes features from each modality to extract sentiment-relevant patterns at different temporal granularities.
- (2) Hierarchical Temporal Attention Mechanism (HTAM). To intelligently integrate the diverse information derived from MTS-TLM, we design a hierarchical attention mechanism. HTAM dynamically assigns weights to features across different temporal scales for each modality, enabling the model to adaptively focus on the most relevant temporal resolutions according to the specific affective characteristics of the input.
- (3) Temporally-Aligned Text-Enhanced Transformer (TA-TET). The refined multi-scale temporal-aware features are subsequently fed into a modified text-enhanced Transformer. This ensures that cross-modal interactions and text-guided augmentations are performed on representations that inherently preserve and exploit fine-grained temporal context, thereby facilitating more robust and temporally coherent multimodal fusion.
- (4) We conduct comprehensive experiments and ablation studies on two benchmark datasets, demonstrating the state-of-the-art performance of MSTETFN and validating the effectiveness of its novel components.

2. Related Work

2.1 Multimodal Sentiment Analysis (MSA)

MSA aims to overcome the limitations of unimodal analysis by integrating information from multiple modalities. Early fusion concatenates features from different modalities before feeding them into a single model. Although simple, it often fails to capture complex cross-modal relationships and is sensitive to misalignment. Late fusion processes each modality independently and fuses their predictions at the decision level [1]. This approach preserves modality-specific

information but may miss deep and subtle cross-modal interactions. Intermediate fusion combines modalities within the neural network architecture. Methods such as TFN [5] and LMF [6] leverage tensor operations to model outer-product interactions. Attention-mechanism-based fusion models [2] utilize cross-modal attention to facilitate information exchange across modalities. Some advanced methods [7–8] further refine fusion by treating text as the core modality and augmenting non-verbal representations during cross-modal mapping..

2.2 Temporal Modeling in MSA

The temporality of emotional expression is crucial in MSA, as many tasks involve analyzing video or audio segments where sentiments evolve dynamically. Early recurrent neural networks were widely used to capture sequential dependencies within each modality, and CHFusion [9] further employed a hierarchical RNN structure to model fine-grained local correlations. However, traditional RNNs struggle with very long sequences and implicitly capture temporal dynamics without explicitly considering multi-scale patterns. TCN[10] emerged as an alternative, offering parallelism and long-range dependency capture through dilated convolutions, yet they typically operate with a single fixed receptive field. Transformers, equipped with positional encoding and self-attention, inherently process sequence data and capture global temporal context, but they treat all time steps uniformly, which may be suboptimal for discerning sentiment variations across different temporal scales.

Our proposed MSTETFN uniquely combines the strengths of the text-enhanced fusion paradigm with an explicit, multi-scale, and hierarchically attentive temporal modeling strategy. Unlike prior methods that focus on single-scale dependencies or global context, MSTETFN ensures that fine-grained affective dynamics across various temporal resolutions are captured, dynamically weighted, and effectively integrated into cross-modal fusion. This direct remedy to the multi-scale temporal limitation constitutes a novel contribution to MSA.

3. Methodology

In this section, we elaborate the architecture of our proposed Multi-scale Temporal-aware Text-Enhanced Transformer Fusion Network (MSTETFN). We first formulate the problem and present the overall architecture and then describe each component in detail.

3.1 Problem Formulation and Overall Architecture

The task of Multimodal Sentiment Analysis (MSA) is to predict the sentiment score $y \in [-3, 3]$ for a given video segment, which contains three modalities: text I_t , visual I_v , and acoustic I_a . Each modality $m \in \{t, v, a\}$ is represented as a sequence $I_m \in \mathbb{R}^{T_m \times d_m}$. As shown in Figure 1, MSTETFN consists of four main stages: (1) feature extraction and initial context encoding; (2) Multi-scale Temporal Sequence Learning Module (MTS-TLM); (3) Hierarchical Temporal Attention Mechanism (HTAM); (4) Temporally-Aligned Text-Enhanced Transformer (TA-TET) and Unimodal Label Generation Module (ULGM) for final sentiment prediction.

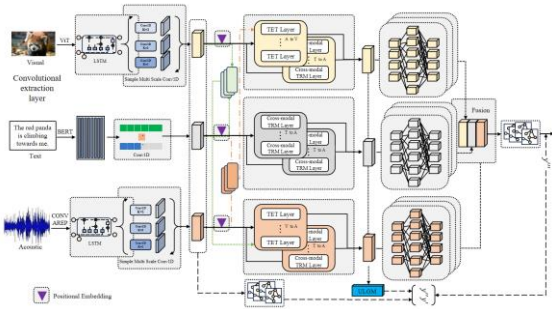


Fig. 1. Architecture diagram of the MSTETFN model

3.2 Feature Extraction and Initial Context Encoding

Drawing upon the successful paradigm of existing text-enhanced fusion networks, we leverage robust pre-trained models for unimodal feature extraction to guarantee high-fidelity initial representations. Specifically, we adopt BERT for textual encoding, ViT for visual encoding, and the COVAREP toolkit for acoustic encoding, yielding the initial feature sets F_t , F_v and F_a for the three respective modalities.

3.3 Multi-scale Temporal Sequence Learning Module (MTS-TLM)

MTS-TLM aims to explicitly capture sentiment-relevant patterns at different temporal granularities. For each modality m , we adopt a parallel architecture of Temporal Convolutional Networks (TCN), specifically one-dimensional convolutional layers with varying kernel sizes and optional dilation rates. Each branch learns features at a distinct temporal scale.

Let h_m^0 be the input sequence for modality m . For each modality, we define k parallel convolutional branches. Each branch $k \in \{1, 2, \dots, K\}$ applies a one-dimensional convolution with a specific kernel size s_k and dilation rate d_k :

$$F_m^{scale,k} = \text{Conv1D}(h_m^0; \text{Kerel_size} = s_k, \text{dilation_rate} = d_k) \quad (1)$$

where $F_m^{scale,k} \in \mathbb{R}^{T_m \times d^t}$ denotes the features extracted at scale k . Small kernel sizes with a dilation rate of 1 capture fine-grained, immediate temporal dynamics, while large kernel sizes or dilated convolutions capture broader contextual temporal patterns. The outputs of the parallel branches are then concatenated:

$$F_m^{multi-scale} = \text{ConcatD}(F_m^{scale,1}, F_m^{scale,1}, \dots, F_m^{scale,K}) \quad (2)$$

3.4 Hierarchical Temporal Attention Mechanism (HTAM)

The MTS-TLM provides a rich set of multi-scale temporal features. The Hierarchical Temporal Attention Mechanism (HTAM) is designed to dynamically weight these multi-scale features. For each modality m , given its multi-scale feature representation $F_m^{multi-scale} \in \mathbb{R}^{T_m \times (K \cdot d')}$, we first split it back into its K scale-specific components. Then, HTAM computes attention weights over these K scales:

$$a_{m,t,k} = \frac{\exp(W_{HTAM} \cdot F_{m,t}^{scale,k})}{\sum_{j=1}^K \exp(W_{HTAM} \cdot F_{m,t}^{scale,k})} \quad (3)$$

where $a_{m,t,k}$ is the attention weight for modality t at time step k and scale t . The final temporal-aware feature representation $H_m \in \mathbb{R}^{T_m \times d'}$ for modality m is obtained by a weighted sum over the scale-specific features.

$$H_{m,t} = \sum_{k=1}^K a_{m,t,k} \cdot F_{m,t}^{scale,k} \quad (4)$$

3.5 Temporally-Aligned Text-Enhanced Transformer (TA-TET)

The core fusion mechanism of MSTETFN draws upon the idea of the text-enhanced Transformer. Let

H_t, H_v, H_a be the temporal-aware representations from HTAM. First, to augment the non-verbal modalities with textual information, H_v and H_a query H_t .

$$c_v = \text{Mul_Att}(Q = H_v W_{Q_v}, K = H_t W_{K_{t1}}, V = H_t W_{V_{t1}}) \quad (5)$$

$$c_a = \text{Mul_Att}(Q = H_a W_{Q_a}, K = H_t W_{K_{t2}}, V = H_t W_{V_{t2}}) \quad (6)$$

Here, c_v and c_a are intermediate representations of the visual and acoustic modalities.

Next, we model pairwise cross-modal mappings. For the i -th layer:

$$h_{a \rightarrow v}^i = \text{Mul_Att}(Q = H_v W_{Q_v}', K = c_a W_{K_a}', V = c_a W_{V_a}') + \text{LN}(H_v) \quad (7)$$

$$h_{t \rightarrow v}^i = \text{Mul_Att}(Q = H_v W_{Q_v}'', K = H_t W_{K_t}'', V = H_t W_{V_t}'') + \text{LN}(H_v) \quad (8)$$

After L layers, the final text-enhanced representation of the visual modality becomes:

$$H_v^{fused} = \text{Concat}(h_{a \rightarrow v}^L, h_{t \rightarrow v}^L) \quad (9)$$

Similarly, we obtain H_v^{fused} by concatenating $h_{a \rightarrow v}^L$ and $h_{t \rightarrow v}^L$. For the text modality itself, we apply self-attention to H_t to gather its intrinsic temporal context:

$$H_t^{fused} = s_Att(H_t) + \text{LN}(H_t) \quad (10)$$

3.6 Unimodal Label Generation Module (ULGM)

For each modality $m \in \{t, v, a\}$, its final temporal-aware representation H_m is passed through a modality-specific fully connected layer to predict a unimodal sentiment score.

$$\hat{y}_s^m = FC_m(H_m^{pooled}) \quad (11)$$

where H_m^{pooled} is typically the output at the final time step or an average pooling of H_m .

3.7 Sentiment Prediction and Loss Function

The final fused representations H_t^{fused} , H_v^{fused} , and H_a^{fused} are concatenated and passed through a fully connected layer to generate the multimodal sentiment prediction:

$$\hat{y}_{mul} = FC_{final}(\text{Cpncat}(H_t^{fused}, H_v^{fused}, H_a^{fused})) \quad (12)$$

The total loss function is a combination of the multimodal regression loss and the unimodal prediction loss.

$$\mathcal{L}_{total} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_{mul}^i - y^i| + \lambda \frac{1}{N} \sum_{m \in \{t, a, v\}} w_i^m \cdot |y_s^{m, (i)}| \quad (13)$$

where λ is a hyperparameter balancing the multimodal and unimodal tasks.

4. Experiments

4.1 Datasets

We evaluate MSTETFN on two widely used benchmark datasets for multimodal sentiment analysis. The CMU-MOSI dataset contains short video clips from YouTube reviews, each manually annotated with sentiment intensity labels ranging from -3 to +3. CMU-MOSEI is the largest MSA dataset, containing 23,353 annotated video clips from 2,928 videos.

4.2 Evaluation Metrics

We adopt standard metrics for regression and classification tasks. For regression, we employ Mean Absolute Error (MAE) and Pearson Correlation Coefficient (Corr). For classification, we binarize sentiment scores (negative vs. non-negative / negative vs. positive). We report binary classification accuracy (Acc-2), i.e., the percentage of correctly classified samples, for both "negative vs. non-negative" and "negative vs. positive" classifications.

4.3 Baseline

To comprehensively evaluate the performance of MSTETFN, we compare it with several state-of-the-art MSA models:

- TNF [3]: Tensor Fusion Network, an early and influential intermediate fusion method.
- LMF [6]: Low-rank Multimodal Fusion, an efficient variant of tensor fusion.
- MulT [2]: Multimodal Transformer, which employs cross-modal Transformers for modality conversion.
- MISA[7]: Modality-Invariant and Modality-Specific Representations, which learns disentangled features.
- TETFN [11]: Text-Enhanced Transformer Fusion Network.

4.4 Implementation Details

The experiments employ the Adam optimizer, with a fine-tuning learning rate of 5×10^{-5} for BERT and 1×10^{-3} for other parameters, and a batch size of 64. The LSTM hidden size is 128. The Conv1D kernel size for all modalities is 3, and the visual feature length is 50 frames. For the Multi-scale Temporal Sequence Learning Module (MTS-TLM), we employ $K = 3$ parallel one-dimensional convolutional branches for each modality. Scale 1 uses a kernel size of 3 with a dilation rate of 1; Scale 2 uses a kernel size of 5 with a dilation rate of 1; Scale 3 uses a kernel size of 3 with a dilation rate of 2.

HTAM module uses 5 heads for the text-oriented multi-head attention in the Transformer layers. The cross-modal Transformer employs $L = 3$ layers. The loss weight λ is empirically set to 0.7 for MOSI and 0.5 for MOSEI. All models are trained for 50–100 epochs with early stopping based on validation MAE. We report average results over 5 runs with different random seeds.

4.5 Quantitative Results

Table 1. presents the performance of MSTETFN and various baseline models on the CMU-MOSI dataset. All results are averaged over multiple runs

Model	MAE↓	Corr↑	Acc-2	F1-score
LMF(B)	0.917	0.695	-/82.5	-/82.4
MuT(B)	0.861	0.711	81.5/84.1	80.6/83.9
MISA(B)	0.783	0.761	81.8/83.4	81.7/83.6
TETFN*	0.717	0.800	84.05/86.10	83.83/86.07
MSTETFN (Ours)	0.693	0.812	85.31/87.19	85.15/87.19

Table 2. Performance results on the CMU-MOSEI dataset

Model	Corr↑	Acc-2	F1-score
TNF(B)	0.700	-/82.5	-/82.1
LMF(B)	0.677	-/82.0	-/82.1
MuT(B)	0.703	-/82.5	-/82.3
MISA(B)	0.756	83.6/85.5	83.3/85.3
TETFN*	0.748	84.25/85.18	84.18/85.27
MSTETFN (Ours)	0.771	85.46/86.32	85.39/86.28

In terms of regression, MSTETFN achieves an MAE of 0.693 and a Pearson correlation of 0.812 on CMU-MOSI, and 0.522 and 0.771 on CMU-MOSEI, outperforming all baselines including TETFN. The higher correlation indicates more accurate sentiment intensity capture. For classification, MSTETFN attains Acc-2 of 85.31/87.19% on CMU-MOSI and 85.46/86.32% on CMU-MOSEI, demonstrating enhanced polarity identification even in complex cases. The consistent improvements, particularly on the larger and more diverse CMU-MOSEI dataset, strongly suggest that explicitly modeling and dynamically weighing multi-scale temporal dependencies significantly enhances the network's understanding of affective dynamics.

4. Conclusion

In this paper, we have proposed a Multi-scale Temporal-aware Text-Enhanced Transformer Fusion Network (MSTETFN), aiming to overcome the limitation of existing multimodal sentiment analysis models in modeling fine-grained temporal dependencies in long video sequences. Building upon the solid foundation established by the text-enhanced Transformer fusion network paradigm, MSTETFN integrates three key innovations: a Multi-scale Temporal Sequence Learning Module (MTS-TLM) for extracting sentiment-relevant patterns at different temporal granularities, a Hierarchical Temporal Attention Mechanism (HTAM) for dynamically weighting and aggregating these multi-scale features, and a Temporally-Aligned Text-Enhanced Transformer (TA-TET) for robust cross-modal interactions.

Our comprehensive experimental evaluations on two publicly available benchmark datasets demonstrate that MSTETFN achieves state-of-the-art performance, validating the effectiveness of our proposed multi-scale temporal awareness and hierarchical attention mechanism..

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

Acknowledgments

This work was supported by the Henan Provincial Key Scientific Research Projects Program and the for Higher Education Institutions under Grant 25A520053 and National Innovation and Entrepreneurship Training Program for College Students 2026 (Grant No. S202611517033).

References

- [1] Poria, S., Cambria, E., Howard, N., Huang, G. B., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, pp. 98-125.
- [2] Tsai, Y.-H.H., Bai, S., Liang, P.P., Kolter, J.Z., Morency, L.-P., Salakhutdinov, R. Multimodal Transformer for Unaligned Multimodal Language Sequences. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 28 July–2 August 2019, pp. 6558–6569..
- [3] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Tensor fusion network for multimodal sentiment analysis. In *Empirical Methods in Natural Language Processing, EMNLP. Copenhagen, Denmark, Sep. 9-11, 2017*, pp. 1103–1114, doi:10.18653/v1/D17-1115.
- [4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, pp. 5998-6008, 2017.
- [5] Zadeh, A., Liang, P. P., Poria, S., Vij, P., Cambria, E., & Morency, L. P. Multi-attention recurrent network for human communication comprehension. In *Proceedings of the AAAI conference on artificial intelligence*, Feb. 2–7, 2018, 32(1), pp. 5622–5630.
- [6] Liu, Z., Shen, Y., Lakshminarasimhan, V. B., Liang, P. P., Zadeh, A., & Morency, L. P. Efficient low-rank multimodal fusion with modality-specific factors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 2018, vol.1, pp. 2247–2256.
- [7] Wang, D., Guo, X., Tian, Y., Liu, J., He, L., & Luo, X. TETFN: A text enhanced transformer fusion network for multimodal sentiment analysis. *Pattern Recognition*, vol. 136, pp. 109259, 2023.
- [8] Hou, J., Omar, N., Tiun, S., Saad, S., & He, Q. . TCHFN: Multimodal sentiment analysis based on text-centric hierarchical fusion network. *Knowledge-Based Systems*, vol. 300, pp. 112220 , 2024.
- [9] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the*

2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) ,pp. 1724–1734 , 2014.

[10] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. Temporal Convolutional Networks for Action Segmentation and Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , pp. 1003-1012 , 2017.

[11] Hazarika, D.; Zimmermann, R.; Poria, S. Misa: Modality-invariant and-specific representations for multimodal sentiment analysis. In *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1122–1131. doi: 10.1145/3394171.3413678.

Author Information



Junjun Zhang

2008: BE, Computer Science and Technology, Nanyang Institute of Technology, China

2017: ME, Computer Technology, Zhengzhou University, China

2023: PhD, Computer Information

Engineering, Cheongju University, South Korea

2009–2019: Professor, School of Software, Zhengzhou University of Industrial Technology, China

2024–Present: Professor, School of Software, Henan University of Engineering, China

Email: zhangjunjun@zzuit.edu.cn

Research Interest: AI, NLP, Computing, Emotion Analysis



Giseop Noh

1998: BE, Industrial Engineering, Korea Air Force Academy, Korea

2009: MS, Computer Science, University of Colorado Denver, USA

2014: PhD, Computer Science, Seoul National University, Korea

1994–2005: Officer, Republic of Korea Air Force, responsible for operation and maintenance of advanced weapon computer systems

2017–2018: Faculty, Department of Computer Science, Korea Air Force Academy

2018–2024: Professor, Department of Artificial Intelligence Software, Cheongju University, Korea

2024–Present: Professor, Department of Software and Communications Engineering, Hongik University

Email: kafa46@hongik.ac.kr

Research Interest: AI, NLP, Computing

Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting

Jung-kyu Park¹, Giseop Noh²

¹BE Course, Department of Software and Communications Engineering, Hongik University, South Korea

²Professor, Department of Software and Communications Engineering, Hongik University, South Korea

Corresponding author: Giseop Noh (kafa46@hongik.ac.kr)

Abstract

In this study, we conducted experiments using U-Net segmentation and LaMa-based inpainting to remove marine snow from high-resolution underwater videos. Underwater videos often suffer from visual degradation caused by light scattering, absorption, and suspended particles. In particular, marine snow appears as small bright dots or blurred spots, reducing the accuracy of subsequent tasks such as object detection and structural analysis. To preserve the detailed information of the original 4K videos, the frames were divided into 512×512 tiles, and a U-Net-based segmentation model was trained to detect marine snow regions as binary masks. The generated masks were then used as input to the LaMa inpainting model to restore the particle regions with surrounding background information. The experimental results showed that tile-based U-Net training is applicable for detecting small marine snow regions in high-resolution underwater videos. However, background noise with visual characteristics similar to marine snow, such as barnacles on underwater structures and bright reflection regions, was also detected. In addition, the LaMa-based inpainting method showed limitations in naturally restoring pixel-level small particle regions. When undetected residual particles remained around the masked region, the restored area was sometimes regenerated in a form similar to marine snow. Therefore, this study confirms the applicability and limitations of U-Net-based marine snow detection and LaMa-based inpainting removal in high-resolution underwater videos. The results suggest that further research is needed on temporal-information-based detection methods and restoration models specialized for marine snow removal.

Keywords: Marine Snow, U-Net, segmentation, inpainting, Underwater Video

1. Introduction

Recently, the investigation and analysis of underwater environments have become increasingly important in various fields, including marine structure inspection, underwater ecosystem observation, marine resource exploration, and the operation of underwater robots and drones. In particular, advances in underwater cameras and equipment such as ROVs (Remotely Operated Vehicles) and AUVs (Autonomous Underwater Vehicles) have made it possible to acquire high-resolution underwater images. Such image data are used as essential resources for the precise analysis of marine environments.

However, unlike images captured in terrestrial environments, underwater images suffer from various types of visual degradation due to the characteristics of

water as a medium. Light scattering and absorption reduce image contrast and sharpness, while fine particles or suspended matter drifting underwater appear in images as bright spots, stains, blurring, or snow-like patterns. This phenomenon is generally referred to as marine snow, and it is a major factor that degrades the visibility of underwater images and reduces the accuracy of subsequent processing tasks such as object detection, structure analysis, and image restoration [1].

Marine snow does not have consistent size, brightness, or shape within images, and it appears in various forms depending on camera motion, lighting conditions, and underwater turbidity [2]. In particular, in videos captured by a moving camera such as an underwater drone, background structures and suspended particles may move together or overlap visually. Therefore, it is difficult to

stably remove only suspended particles using simple filtering or threshold-based methods. In addition, excessive image restoration may damage the texture of background structures, making it important to balance suspended particle removal and preservation of the original image.

Existing studies on underwater image enhancement have mainly focused on methods such as color correction, contrast enhancement, and denoising. Although these methods can be effective in improving overall image quality, they have limitations in precisely detecting and removing marine snow located at specific positions in an image. Recently, deep learning-based removal studies have been conducted; however, when suspended particles in underwater images have irregular shapes and are difficult to distinguish from the background, it remains difficult to obtain stable removal results [2].

In this study, a pipeline-based approach consisting of segmentation and inpainting is experimentally applied to remove marine snow present in underwater images. First, a segmentation model is used to extract marine snow regions as masks. Then, based on the extracted masks, an inpainting method is applied to naturally restore the regions where suspended particles were present using the surrounding background. Through this process, this study aims to construct an underwater image restoration method that explicitly detects and removes suspended particle regions, rather than simply enhancing the image.

The purpose of this study is to specifically identify the major problems that occur when attempting to remove marine snow from real underwater images. In the experimental process, variations in the size and brightness of suspended particles, similarity to background structures, temporal instability caused by camera motion, and background damage caused by excessive restoration were analyzed. Based on this analysis, this study summarizes the limitations of existing image processing and deep learning-based restoration methods in real underwater environments and suggests directions for improvement toward the development of more stable marine snow removal techniques.

2. Related Work and Limitations

Marine snow appearing in underwater images takes the form of small dots, bright particles, or blurred stains, degrading both the visual quality of images and the performance of subsequent analysis tasks [2]. To mitigate this problem, filtering-based methods and segmentation-based detection methods can be used in existing studies and image processing procedures. Filtering-based methods aim to reduce noise across the entire image,

whereas segmentation-based methods aim to detect suspended particle regions to be removed at the pixel or object level. However, in real underwater images, both approaches have limitations because the size and shape of marine snow are irregular, and visually similar elements often coexist with background structures.

The median filter is a representative nonlinear filter widely used for noise removal in OpenCV-based image processing. However, since the median filter is applied uniformly to the entire image, it cannot selectively remove only marine snow regions. When the kernel size of the filter is set to a large value, small suspended particles or dot-like noise can be partially reduced, but the boundaries and fine textures of background structures may also be damaged [3]. In particular, the contours of underwater structures, surface textures, and fine patterns in the seafloor background can serve as important information for subsequent analysis. Therefore, excessive filtering may cause a decrease in image sharpness and blurring of boundaries. Conversely, when the kernel size is set to a small value, the detailed information of the original image is relatively well preserved, but marine snow that appears small or blurred may not be sufficiently removed. In addition, in underwater videos captured by a moving camera, the positional relationship between suspended particles and the background changes from frame to frame. Thus, it is difficult to obtain temporally stable removal results using simple spatial filtering alone. Therefore, although the median filter can be used to reduce noise in underwater images, it has limitations as a method for precisely detecting marine snow or selectively removing it without damaging the background.

Segmentation-based methods play an important role in image restoration tasks that require explicit specification of the target regions to be removed, because they can separate specific objects or regions in an image at the pixel level. In particular, deep learning-based segmentation models can learn features such as color, brightness, boundaries, and shape from input images and predict object regions in the form of masks. The generated masks can then be used as inputs for restoration techniques such as inpainting.

However, existing segmentation-based methods often focus on problems in which object shapes and boundaries are relatively clear or the target regions are larger than a certain size. In contrast, the marine snow addressed in this study appears as very small dots, blurred particles, or bright stains within images, and occupies a very low pixel ratio in the entire frame. Such small suspended particles do not have clear boundaries with the background, and their size and brightness vary from frame to frame.

Therefore, it is difficult to detect them stably using general segmentation methods alone.

Furthermore, underwater images contain background elements with visual characteristics similar to those of marine snow. Barnacles on the surfaces of underwater structures, lighting reflections, and bright textures in the seafloor background may have colors and shapes similar to suspended particles, causing false detections. Conversely, blurred or very small marine snow may be classified as background and remain undetected. Therefore, for marine snow detection, it is necessary to construct separate training data and design a segmentation model that considers the characteristics of small pixel-level suspended particles, rather than directly applying a general object segmentation model. Accordingly, in this study, high-resolution underwater images were used to train a U-Net-based segmentation model to detect marine snow regions as binary masks [4]. The generated masks were then used as inputs to the LaMa inpainting model, and experiments were conducted to restore the detected suspended particle regions using surrounding background information [5].

3. Research Method

3.1 U-Net Based Segmentation

3.1.1 Dataset Construction

In this study, a U-Net-based binary segmentation model was trained to detect marine snow regions in underwater images. To this end, videos captured by an underwater drone were divided into individual frames, and the marine snow regions present in each frame were manually annotated at the pixel level to construct the dataset. The input images consisted of actual underwater video frames, while the ground truth masks were generated as binary images in which marine snow regions were represented in white and background regions in black.

In the ground truth masks, the white regions indicate marine snow regions that the model should detect, whereas the black regions indicate background regions that should be preserved. By using these binary masks, the U-Net model was trained as a segmentation problem that separates marine snow regions from the input image at the pixel level.

When high-resolution underwater frames are used for training at their original size, GPU memory usage can increase rapidly, making training difficult. On the other hand, when the entire image is simply resized to reduce the input size, it was experimentally observed that marine

snow pixels, which occupy a very small proportion of the total pixels, can be lost or weakened during the resizing process. Therefore, in this study, to preserve the detailed information of the original 4K images as much as possible while reducing the GPU memory burden, the entire images were not directly resized but were instead divided into 512×512 tiles to construct the training data.

Finally, the U-Net training dataset consisted of 21,711 image-mask pairs, and the validation dataset consisted of 5,258 image-mask pairs.

3.1.2 U-Net Training

The U-Net model was trained in a supervised learning manner using pairs of original images and their corresponding binary masks. The input images were loaded in RGB format and normalized to values between 0 and 1. For the masks, regions with pixel values greater than 127 were converted into marine snow regions, while the remaining regions were converted into background regions, thereby producing binary masks. During training, both the input images and masks were used at a size of 512×512 .

The input image consists of a 512×512 RGB image and is converted into 64 feature channels through the first convolution block. Subsequently, in the encoder path, the spatial resolution of the feature maps is gradually reduced using max pooling, while the number of channels is increased to 64, 128, 256, 512, and 1024 to extract high-level features from the image.

Each encoder stage used a DoubleConv block composed of two 3×3 convolutions, batch normalization, and ReLU activation functions. Through this structure, the model was trained to learn local features of marine snow, such as brightness, shape, and boundaries. In the bottleneck stage, 1024 feature channels were used at the smallest spatial resolution to represent the abstract features of the input image.

In the decoder path, the resolution of the feature maps was gradually restored using transposed convolution. At this stage, feature maps generated at the corresponding resolution levels in the encoder were connected through skip connections and combined with the decoder feature maps. Skip connections help compensate for positional information and detailed boundary information that may be lost during the downsampling process. This is important for detecting marine snow regions that appear as small dots or blurred stains at the pixel level [4].

In the final output layer, a 1×1 convolution was applied to generate a one-channel output mask. The output value represents the probability that each pixel

corresponds to marine snow, and a threshold was subsequently applied to convert it into a binary mask that separates marine snow regions from background regions.

Since marine snow occupies a very small pixel ratio in the entire image, using only general binary cross entropy may cause the model to be biased toward the background region during training [6]. To alleviate this problem, this study used a combination of Weighted Binary Cross Entropy Loss, which assigns a weight to the positive class, and Dice Loss [7]. Weighted BCE assigns higher importance to marine snow pixels, thereby compensating for the learning of small suspended particle regions, while Dice Loss directly reflects the area overlap between the predicted mask and the ground truth mask [8].

The final loss function was constructed by combining Weighted BCE Loss and Dice Loss at an equal ratio.

$$L_{U-Net} = L_{WBCE} + L_{Dice}$$

Table 1 shows the main hyperparameters used for training the U-Net model in this study.

Table 1. Training Hyperparameter Settings

Parameter	Value
Batch size	4
Epoch number	100
Learning rate	0.0001
Early Stopping Patience	15

3.2 LaMa-Based Inpainting

In this study, the LaMa inpainting model was applied based on the marine snow masks generated through U-Net segmentation. Inpainting is a technique that restores regions to be removed from an image or video using surrounding background information. In this study, the marine snow regions marked in white in the binary masks predicted by U-Net were set as the target regions for removal, and these regions were restored using the LaMa model.

LaMa is a deep learning-based image restoration model proposed for large-mask inpainting and is used to restore large missing regions and complex structures. In this experiment, each frame and its corresponding U-Net-predicted mask were used as inputs, and the suspended particle regions corresponding to the mask regions were filled in a form similar to the surrounding background.

The processing was performed on a frame-by-frame basis. The original frame was input into the U-Net model to generate a marine snow mask, and the generated mask and the original frame were then input into the LaMa model. Subsequently, the frames restored by LaMa were reassembled into a video according to the frame order and FPS of the original video.

However, the LaMa inpainting used in this study is a frame-level image restoration method and therefore does not directly consider temporal consistency between consecutive frames. As a result, in videos captured by an underwater drone, where camera motion and suspended particle motion occur simultaneously, the restoration results may differ slightly from frame to frame. Considering this characteristic, this study visually compared the LaMa-based inpainting results and analyzed the background restoration quality after marine snow removal as well as the occurrence of residual suspended particles.

4. Experimental Results

In this study, marine snow regions were detected using the trained U-Net model. The experimental results showed that small suspended particle regions distributed in the image could be detected by dividing the original 4K video into tiles for training, without directly downscaling the video. Figure 1 shows the original underwater video frame, and Figure 2 shows the marine snow segmentation result generated by inputting the corresponding frame into the U-Net model. Some small suspended particles that were difficult to identify visually in the entire frame were also detected, which is considered to be the result of training while preserving the detailed information of the high-resolution image.

However, the detection results of U-Net showed that regions other than suspended particles were also detected. In particular, barnacles on the surfaces of underwater structures and bright reflection regions tended to be falsely detected when they exhibited features similar to those of suspended particles. This can be regarded as a limitation caused by the fact that a single-frame-based segmentation model cannot consider the motion information of suspended particles or temporal differences from the background.

Figure 3 shows the marine snow removal result obtained by applying an existing inpainting method using the U-Net-based segmentation result as a mask. For suspended particles with relatively clear sizes and distinguishable boundaries from the surrounding background, the masked regions were restored in a form similar to the surrounding background. However, for

suspended particles appearing at the pixel level, the mask regions were very small, and the restoration effect was not clearly observed. In some regions, traces similar to the removed suspended particles were generated again.

In addition, when undetected residual suspended particles remained in the surrounding area, the inpainting model recognized them as background information and generated patterns similar to the residual suspended particles again during the restoration process. Therefore, it was confirmed that existing inpainting methods can be applied to general object removal, but have limitations in completely removing fine marine snow that is irregularly distributed in underwater images.

Overall, this experiment confirmed that the U-Net trained using a 4K tile-based learning approach can be applied to the detection of small suspended particles. However, false detections of background noise similar to suspended particles and the limitations of existing inpainting methods in restoring fine suspended particle regions were observed together. These results indicate that, for marine snow removal in underwater images, additional approaches that consider temporal continuity or the motion characteristics of suspended particles are required, beyond simple segmentation and restoration methods.

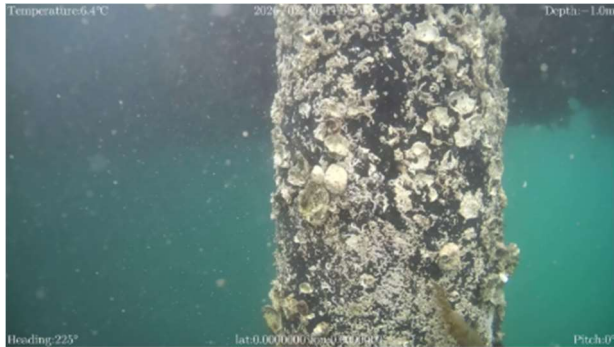


Fig. 1. Natural Image



Fig. 2. U-Net Output Mask



Fig. 3. LaMa inpainting Output Image

5. Conclusion

In this study, to examine the feasibility of removing marine snow from 4K underwater videos, the original images were divided into 512×512 tiles, and a U-Net-based segmentation model was trained. Through this approach, it was confirmed that small suspended particle regions can be detected at the pixel level without directly downscaling high-resolution images. In particular, the tile-based learning method showed the advantage of preserving the detailed information of the original images while allowing the model to focus relatively more on marine snow regions, which occupy a very small proportion of the entire image.

However, the experimental results showed that although U-Net-based segmentation could detect small marine snow regions relatively sensitively, it also detected regions with visual characteristics similar to marine snow, such as barnacles on the surfaces of underwater structures, bright reflection regions, and fine background textures. This indicates that, because marine snow appears in images as small dots, blurred particles, or bright stains, it is difficult to stably distinguish suspended particles from background noise based only on the spatial features of a single frame.

In addition, when existing inpainting methods were applied based on the detected masks, limitations were observed in naturally restoring small suspended particle regions at the pixel level. In particular, when undetected residual suspended particles existed around the mask, the restoration process referred to the surrounding information and filled the corresponding regions again in a form similar to suspended particles. This implies that, for marine snow removal, it is important not only to restore the detected regions but also to generate precise masks that consider fine suspended particles that remain undetected.

Limitations were also observed in video-level application. The same marine snow was not consistently detected across consecutive frames, and cases occurred in

which it was detected in some frames but not in others. As a result, a flickering phenomenon was observed in the removal results, where certain suspended particles disappeared and then reappeared. This phenomenon can be considered a problem caused by the frame-level segmentation model not sufficiently reflecting temporal continuity.

Therefore, future research should consider detection methods that go beyond conventional approaches relying on the spatial features of a single frame and jointly reflect motion information and temporal consistency between consecutive frames. In addition, further research is needed to analyze the effect of the weight ratio between Weighted BCE Loss and Dice Loss during U-Net training on detection sensitivity and false detections, and to derive an optimal loss function combination that can more stably distinguish background regions from marine snow regions.

Acknowledgement

This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency(NIPA) (Grant No. PJTS0801260410070001000300200).

References

- [1] Fernando Galetto, Guang Deng “A deep learning approach for marine snow synthesis and removal.”, 2023
- [2] Jeremy Paul Coffelt, Nicolas Nowald, Peter Kampmann, “Marine Snow Simulation and Elimination in Video”, 2023
- [3] Reina Kaneko, Yuya Sato, Takumi Ueda et al., “Marine Snow Removal Benchmarking Dataset”, 2024
- [4] Olaf Ronneberger, Philipp Fischer, Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, 2015
- [5] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin et al., “Resolution-robust Large Mask Inpainting with Fourier Convolutions”, 2022
- [6] Saeid Asgari Taghanaki, Yefeng Zheng, S. Kevin Zhou et al., “Combo loss: Handling input and output imbalance in multi-organ segmentation”, 2019
- [7] Jun Ma, Jianan Chen, Matthew Ng et al., “Loss odyssey in medical image segmentation”, 2021
- [8] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb et al., “Unified Focal loss: Generalising Dice and cross

entropy-based losses to handle class imbalanced medical image segmentation”, 2022

Author Information



Jung-kyu Park
2022~ now: BE, Dept. of Software & Communications Engineering, Hongik University
Email: pjg367@naver.com
Research Interests: Deep Learning, Computer Vision, Image Processing



Giseop Noh
1998: BE, Engineering, Korea Air Force Academy
2009: MS, Computer Science, University of Colorado, USA
2014: PhD, Computer Engineering, Seoul National University
2016.12~2018.02: Assistant Professor, Dept. of Computer & Information Science, Korea Air Force Academy
2018.03~2025.02: Assistant Professor, Dept. of Artificial Intelligence Software, Cheongju University
2025.03~now: Assistant Professor, Dept. of Software & Communications Engineering, Hongik University
Email: kafa46@hongik.ac.kr
Research Interests: NLP, Neural Network Theory, Reinforcement Learning

Semantic Cohesion Analysis of the Chunk of Code with an LLM

Eunjin Jeong¹, Janghwan Kim², Chaeyun Seo³, R. Young Chul Kim⁴

¹Undergraduate Student, Dept. of Software and Communications Engineering, Hongik University, South Korea

²Ph.D Candidate, Dept. of Software and Communications Engineering, Hongik University, South Korea

³Ph.D, Dept. of Software and Communications Engineering, Hongik University, South Korea

⁴Professor, Dept. of Software and Communications Engineering, Hongik University, South Korea

Corresponding author: R. Young Chul Kim (bob@hongik.ac.kr)

Acknowledgement: This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency (NIPA) (Grant No. PJTS0801260410070001000300200).

Abstract

Recent evaluations of software quality have raised concerns that structural metrics fail to capture design intent adequately. Metrics such as LCOM measure only the physical connectivity between code elements, missing the semantic notion of cohesion that reflects a developer's actual intent. To address this, we propose a Large Language Model (LLM)-based semantic cohesion evaluation model built on the 'LLM-as-a-Judge' paradigm, reformulating classical cohesion theory — grounded in Stevens' taxonomy and the Single Responsibility Principle — into five semantic rubric criteria evaluated through an automated process. This builds on prior findings that top-tier LLMs, such as Claude Sonnet and GPT-4o-mini, correlate strongly with human expert reviewers when interpreting identifiers, comments, and logic flow, letting them catch design-intent distortions that static analysis misses. We detail a five-stage pipeline — from preprocessing and AST-based context extraction to rationale-backed scoring — that converts subjective judgments into an objective, data-driven metric, laying groundwork for a semantics-centered code quality framework. Future work will validate the rubric across multiple LLMs and extend it into an automated refactoring recommendation tool.

Keywords: Large Language Model, Semantic Cohesion, LLM-as-a-Judge, Source Code Analysis

1. Introduction

As software systems grow more complex, managing cohesion has become increasingly important for preserving module independence and maintainability [1]. Cohesion measures the degree to which a module's components cooperate toward a single purpose, and high-cohesion design is central to improving readability and minimizing side effects [2]. However, existing structural metrics such as Lack of Cohesion in Methods (LCOM) rely solely on syntactic connectivity and therefore fail to

capture the semantic notion of cohesion, leaving design intent and code context unreflected [3, 4].

To address this limitation, we propose a semantic cohesion measurement method based on the 'LLM-as-a-Judge' paradigm. Our goal is to compute a deeper level of cohesion — one that is difficult to capture through static analysis — by leveraging semantic information such as identifiers and comments. To this end, we re-establish traditional cohesion theory and build a five-criterion semantic rubric evaluation system optimized for LLMs' qualitative reasoning capabilities [5]. This enables quantification of subjective code interpretation against a

consistent standard, enabling semantic quality evaluation while providing an in-depth rationale for each judgment.

Section 2 reviews related work, Section 3 details the proposed methodology, and Section 4 concludes the paper with directions for future work.

2. Related Work

2.1 Cohesion Models in the Object-Oriented Paradigm

Traditional cohesion is a measure of the association among components within a module; in the object-oriented paradigm, it has been concretized primarily in terms of relationships between classes and methods. Prior work has proposed schemes that identify cohesion levels based on the syntactic structure of code, such as method call frequency, parameter sharing, and return-value usage [6], with sequential and communicational cohesion — where data flow is clear — widely used as key indicators of module reusability. As noted above, however, such purely structural measures cannot infer the logical consistency or design intent latent in the code, motivating the LLM-based semantic analysis model proposed in this paper.

2.2 LLM-as-a-Judge Based Rubric Evaluation Model

Recently in software engineering, the ‘LLM-as-a-Judge’ paradigm — which uses LLMs as the evaluating agent for software artifact quality — has drawn significant attention, building on the foundational LLM-as-a-Judge framework originally proposed for evaluating open-ended natural-language responses [7]. A subset of high-performing LLMs — such as Claude Sonnet and GPT-4o-mini — has been empirically shown to achieve high correlation with human expert reviewers when interpreting naming conventions, comment context, and logic flow [8], suggesting that this interpretive capability can be leveraged to analyze semantic cohesion that reflects design intent.

However, simple prompting approaches suffer from hallucination and inconsistent, subjective interpretation. To address this, our study adopts a rubric-based evaluation process (Fig. 1), since prior work shows that a concrete rubric improves both the validity and reproducibility of LLM-based evaluation [5].

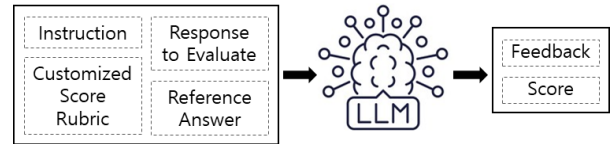


Fig. 1. LLM-as-a-Judge evaluation process

This process jointly analyzes the Instruction, the Response to Evaluate, the Reference Answer, and a Customized Score Rubric, generating both a score and accompanying feedback that grounds the evaluation in a clear, traceable rationale.

3. Proposed Methodology

3.1 Design of an LLM-Based Semantic Cohesion Evaluation Rubric

We accept the criticism raised in prior work that the traditional seven-level cohesion hierarchy, being merely an ordinal scale, is unsuited to automated quantitative measurement [2], and that applying it to object-oriented systems requires subjective semantic analysis, making automation difficult [1]. To address this, we restructure the complex hierarchy into five core semantic indicators that an LLM can evaluate in fine detail, and we design a 100-pt rating scale for each indicator. Throughout this rubric, the function is used as the primary unit of evaluation, while module refers to the class- or file-level container that groups related functions.

Table 1. LLM-Based Semantic Cohesion Evaluation Rubric

Evaluation Indicator	Evaluation Criteria
Responsibility Focus (RF)	Does the function exist to fulfill only a single, clearly defined domain purpose?
Functional Consistency (FC)	Do the multiple functions grouped within the module share a common domain context?
Data Cohesion (DC)	Do the functions within the module communicate closely around the same core data or state?
Flow Cohesion (FIC)	Does the execution flow between functions reflect an organic exchange of data, or merely an arbitrarily imposed order?
Role Clarity (RC)	Do identifiers (function/variable names) and comments clearly and truthfully describe the actual internal logic?

Table 1 presents the five evaluation indicators and their corresponding criteria. They are grounded in the classical cohesion taxonomy proposed by Stevens et al. [9] and the Single Responsibility Principle (SRP) from the SOLID principles [10]. Specifically, Responsibility Focus (RF)

and Functional Consistency (FC) correspond to functional cohesion and the SRP; Data Cohesion (DC) corresponds to communicational cohesion; and Flow Cohesion (FIC) corresponds to sequential cohesion. Role Clarity (RC) has no counterpart in that taxonomy — it was newly introduced so the LLM's natural-language reasoning can judge whether identifiers and comments truthfully reflect the code's actual logic, exposing 'distortions of design intent' that static analysis tools cannot detect.

Table 2. Deduction Severity Criteria for LLM-Based Scoring

Severity Level	Deduction Range	Description
Trivial	−1 to −5 pt	A minor flaw that barely deviates from the principle
Minor	−6 to −15 pt	Deviates from the principle but does not lead to a functional problem
Moderate	−16 to −30 pt	A clear violation with a substantive impact on maintainability
Major	−31 to −45 pt	Multiple principles violated at once, or impact spread across the module
Critical	−46 to −60 pt	A fundamental violation that effectively collapses cohesion

Table 2 summarizes the deduction severity criteria used to score each indicator. If multiple issues are found within a single indicator, the LLM deducts for each and accumulates the total, with a floor of 0. A single violation cannot exceed −60 pt; a more severe case must instead be expressed as multiple individually listed violation reasons whose deductions accumulate, thereby keeping the deduction rationale auditable and helping constrain cross-model variance.

3.2 Proposed Semantic Cohesion Evaluation Process

By applying the single-standard evaluation rubric defined in Section 3.1 to the system, we propose an automated ‘semantic cohesion evaluation process’ that spans from source-code input to the derivation of a final quality score. As shown in Fig. 3, the overall proposed architecture comprises five core processing stages.

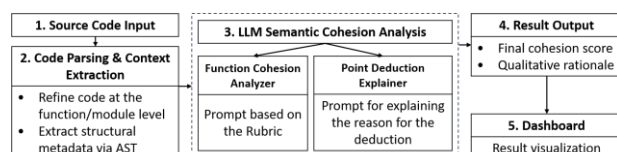


Fig. 3. Rubric-based semantic cohesion evaluation process

Step 1. Source Code Input: The source code to be analyzed is entered into the system.

Step 2. Preprocessing and Context Extraction: The code is refined at the function and module levels, with structural metadata extracted from the AST to capture both syntactic and semantic context.

Step 3. LLM Semantic Cohesion Analysis: An LLM injected with the rubric prompt compares the code's design intent against each of the five indicators. Within this single LLM call, two functional roles operate together: the Function Cohesion Analyzer assigns a numeric score to each rubric indicator, while the Point Deduction Explainer generates a natural-language rationale for every point deducted, referencing the severity levels defined in Table 2.

Step 4. Result Output: Structured data — including the cohesion score, matched indicators, and qualitative rationale — is produced for each function.

Step 5. Dashboard: The results are rendered on a web-based dashboard for intuitive interpretation by the developer.

4. Conclusion and Future Work

We proposed an LLM-based semantic cohesion evaluation model to overcome the limitations of existing static analysis, converting subjective, qualitative judgments into an objective, rubric-driven metric through an automated evaluation process. As future work, we plan to validate the rubric empirically across multiple LLMs and extend the system into an automated refactoring recommendation tool that directly improves code based on the evaluation results.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

Acknowledgement

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency (NIPA) (Grant No. PJTS0801260410070001000300200).

References

- [1] L. C. Briand, J. W. Daly, and J. Wüst, "A unified framework for cohesion measurement in object-oriented systems," *Empirical Software Engineering*, vol. 3, no. 1, pp. 65-117, 1998.
- [2] J. M. Bieman and L. M. Ott, "Measuring functional cohesion," *IEEE Transactions on Software Engineering*, vol. 20, no. 8, pp. 644-657, Aug. 1994.
- [3] A. Marcus and D. Poshyvanyk, "The conceptual cohesion of classes," in *Proc. IEEE Int. Conf. Software Maintenance (ICSM'05)*, Budapest, Hungary, 2005, pp. 133-142.
- [4] L. H. Etzkorn and H. S. Delugach, "Towards a Semantic Metrics Suite for Object-Oriented Design," *Journal of Object-Oriented Programming*, vol. 12, no. 10, pp. 35-44, 2000.
- [5] S. Bi et al., "Prometheus: Inducing Fine-Grained Evaluation Capability in Language Models," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024.
- [6] E. Y. Byun, B. K. Park, W. S. Jang, and Y. C. Kim, "Best practices for cohesion level identification for code reuse in the object-oriented paradigm," in *Proc. KIPS (Korea Information Processing Society) Fall Conf.*, vol. 23, no. 2, pp. 455-458, 2016.
- [7] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," in *Proc. 37th Int. Conf. Neural Information Processing Systems (NeurIPS)*, 2023.
- [8] M. Kawalerowicz, M. Pietranik, and K. Stępnik, "LLMs as Code Review Agents: A Rapid Review and Experimental Evaluation with Human Expert Judges," in *Computational Collective Intelligence (ICCCI 2025)*, *Lecture Notes in Computer Science*, vol. 16138, Springer, Cham, 2026.
- [9] W. P. Stevens, G. J. Myers, and L. L. Constantine, "Structured design," *IBM Systems Journal*, vol. 13, no. 2, pp. 115-139, 1974.
- [10] R. C. Martin, *Agile Software Development: Principles, Patterns, and Practices*. Upper Saddle River, NJ: Prentice Hall, 2003.

Author Information



Eunjin Jeong
now: Undergraduate Student,
Department of Software and
Communications Engineering,
Hongik University, South Korea
Email: je031020@naver.com

Research Interest: Natural Language Processing,
Machine Learning, Prompt Engineering



Janghwan Kim
now: Ph.D Candidate, Department
of Software and Communications
Engineering, Hongik University, South
Korea
Email: lentoconstante@hongik.ac.kr

Research Interest: AI software validation, reinforcement
learning-based software validation, software
visualization, and requirements engineering.



Chaeyun Seo
2014: Ph.D, Department of Computer
Engineering, Hongik University
Email: chaeyun@hongik.ac.kr
Research Interest: Metaverse,
Metaverse Contents, Design Thinking,

CBD, Software Engineering, BPM, AI, Business Process
Modeling.



R. Young Chul Kim
2000: Ph.D. Degree,
Software Engineering from the
Department of Computer Science,
Illinois Institute of Technology, USA
Email: bob@hongik.ac.kr

Research Interest: Test Maturity Model, Model-Based
Testing, Metamodeling, Software Process Model,
Software Visualization, AI software validation,
reinforcement learning-based software validation.

Distinguishing Professional and Amateur Piano Performances using 2D Audio Features and Convolutional Neural Networks

Authors and affiliations:

Changheun Oh*, Department of Software and Communication Engineering, Hongik University, Sejong 30016, Republic of Korea; choh@hongik.ac.kr

Junyeong Song, Department of Software and Communication Engineering, Hongik University, Sejong 30016, Republic of Korea; june227321@gmail.com

Abstract

Evaluating the skill level of a pianist traditionally requires expert human listeners [1]. In this paper, we propose a machine learning approach to automatically classify piano performances into two categories: professional and amateur. We approach this as a binary audio classification problem. Utilizing the Maestro v3.0.0 dataset for professional recordings [2] and the newly introduced PianoVAM v1.0 dataset for amateur recordings, we extract various 2D audio representations—specifically, Linear Magnitude Spectrograms, Mel-spectrograms, and Mel-Frequency Cepstral Coefficients (MFCCs) [3], [4]. These representations are used to train a Convolutional Neural Network (CNN) [5], [6] consisting of three convolutional blocks followed by adaptive average pooling and fully connected layers.

To understand the optimal input characteristics for skill classification, we conducted extensive grid-search experiments over varying audio clip durations (1.0, 2.0, and 5.0 seconds) and Short-Time Fourier Transform (STFT) frequency analysis parameters (N_FFT and Hop Length). Our results demonstrate that shorter duration clips (1.0 second) utilizing standard Spectrogram representations (N_FFT=2048, Hop=1024) achieve the highest test accuracy of 99.01%. This finding provides strong empirical evidence that a model primarily relies on localized acoustic features, such as articulation precision, attack transients, and timbre consistency, rather than long-term musical phrasing or structural context, to distinguish between professional and amateur skill levels.

Keywords: Audio Classification, Convolutional Neural Networks (CNN), Music Information Retrieval (MIR)

1. Introduction

The assessment of musical proficiency is a subjective and highly complex task, often requiring years of training for a human expert to perform reliably [7]. In piano performance, the distinction between a seasoned professional and a novice amateur lies not only in the correct execution of notes but also in nuanced temporal dynamics, articulation, pedaling, and consistency of timbre [8]. Automating this classification process has profound implications for various fields, including

computer-assisted music education, automated audition screening, and music information retrieval (MIR) [1], [9].

In this study, we cast skill-level assessment as a binary audio classification problem. Rather than relying on symbolic data (such as MIDI) which strips away much of the acoustic nuance, we process raw audio waveforms directly. We evaluate the effectiveness of different 2D audio feature representations and temporal window durations using a relatively shallow, naive Convolutional Neural Network (CNN) [5]. By systematically analyzing

different audio cropping lengths and 2D feature representations, we aim to uncover which temporal and spectral characteristics are most discriminative of a pianist's skill.

In this study, we systematically investigate three primary objectives. First, we aim to determine the optimal audio feature representation—such as Spectrogram, Mel-spectrogram, or MFCC—for capturing the nuances of piano skill level. Second, we evaluate how the temporal duration of the analyzed audio clip affects classification performance, specifically examining whether the model requires long-term musical context or if brief localized snapshots are sufficient. Finally, we seek to identify the optimal time-frequency resolution trade-off by analyzing various STFT window sizes and hop lengths.

2. Methodology

2.1 Datasets

We utilized two distinct datasets to represent the two skill levels, mapping them to a binary classification task where Professional is denoted as Class 1 and Amateur as Class 0.

- Professional (Class 1): The Maestro (MIDI and Audio Edited for Synchronous Tracks and Organization) v3.0.0 dataset [2]. This dataset comprises highly skilled, virtuosic classical piano performances recorded on Yamaha Disklaviers. It is widely recognized as a benchmark for professional-level piano audio.

- Amateur (Class 0): The PianoVAM v1.0 dataset. This dataset contains performances by beginner to intermediate piano learners. The metadata allows us to properly split the data (using train, extended train, valid, and test splits) to avoid data leakage.

2.2 Audio Preprocessing and Feature Extraction

All audio files were standardized to a sample rate of 22,050 Hz. Rather than processing entire performances—which can span several minutes—audio files were split into short, fixed-duration segments to provide consistent inputs to the neural network.

2.2.1 Temporal Durations

We evaluated three distinct temporal durations to determine the necessity of long-term context: - 1.0 second - 2.0 seconds - 5.0 seconds

During training, we implemented a dynamic loading mechanism that extracts random continuous segments (of the specified duration) from the full audio files. If a file was shorter than the target duration, it was zero-padded to ensure uniform tensor dimensions.

2.2.2 2D Feature Representations

For each audio segment, we extracted three types of 2D features: 1. Spectrogram: A linear magnitude spectrogram, providing a direct representation of frequency content over time. Values were converted to a decibel (dB) scale relative to the maximum amplitude. 2. Mel-Spectrogram: A perceptually scaled frequency representation aligned with human hearing, utilizing 128 Mel bands ($N_{\text{MELS}} = 128$) [6]. 3. Mel-Frequency Cepstral Coefficients (MFCC): A highly compact representation of the spectral envelope, utilizing the first 13 coefficients ($N_{\text{MFCC}} = 13$) [3].

2.2.3 Time-Frequency Resolution (STFT Parameters)

To investigate the optimal time-frequency resolution, we performed a grid search over various Short-Time Fourier Transform (STFT) window sizes (N_{FFT}) and hop lengths (Hop). We evaluated the following combinations of (N_{FFT} , Hop): $\{(1024, 256), (1024, 512), (2048, 512), (2048, 1024), (4096, 1024)\}$.

2.3 Model Architecture

We intentionally designed a compact, naive CNN architecture to process the 2D audio features as single-channel “images”. The simplicity of the model ensures that high performance is attributed to the discriminative power of the features themselves rather than complex architectural priors.

2.3.1 Conv Blocks

The network, implemented in PyTorch, consists of three sequential convolutional blocks:

Conv Block 1: 2D Convolution (1 channel in, 16 channels out, 3×3 kernel, padding=1) $\rightarrow$ 2D Batch Normalization $\rightarrow$ ReLU Activation $\rightarrow$ 2D Max Pooling (2×2).

Conv Block 2: 2D Convolution (16 channels in, 32 channels out, 3×3 kernel, padding=1) $\rightarrow$ 2D Batch

Normalization → ReLU Activation → 2D Max Pooling (2×2).

Conv Block 3: 2D Convolution (32 channels in, 64 channels out, 3×3 kernel, padding=1) → 2D Batch Normalization → ReLU Activation → 2D Max Pooling (2×2).

Following the convolutional feature extraction, an AdaptiveAvgPool2d(1, 1) layer is employed. This crucial layer aggregates the feature maps down to a fixed size of 64×1×1, making the model robust to any minor variations in the temporal dimension caused by different hop lengths or duration settings.

The aggregated features are flattened and passed through a fully connected network: - Linear Layer (64 input, 32 output) → ReLU Activation - Dropout Layer (probability = 0.5) to mitigate overfitting - Linear Layer (32 input, 1 output)

2.3.2 Loss Function

The network outputs a raw logit for binary classification. We optimize the network using Focal Loss ($\alpha=0.25$, $\gamma=2.0$) [10]. Focal Loss dynamically scales the cross-entropy loss based on classification confidence, compelling the model to focus on hard-to-classify examples.

3. Experiments and Results

The model was trained and evaluated iteratively across all parameter combinations of Duration, Feature Type, N_FFT, and Hop Length. For each configuration, we recorded the Test Accuracy.

3.1 Overall Performance

Our empirical results reveal exceptionally high accuracy for the binary classification task. The single highest performing configuration utilized 1.0-second crops, standard linear Spectrograms, N_FFT=2048, and Hop=1024. This optimal model achieved a 99.01% Test Accuracy.

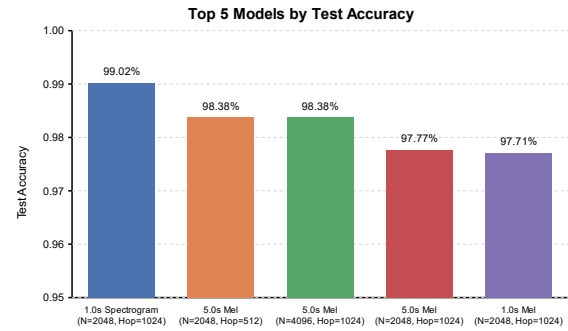


Figure 1. Comparison of the top 5 models across all parameter grid combinations.

Other highly competitive configurations included: - 5.0s, Mel-spectrogram, N_FFT=2048, Hop=512 (Accuracy: 98.37%) - 5.0s, Mel-spectrogram, N_FFT=4096, Hop=1024 (Accuracy: 98.37%) - 1.0s, Mel-spectrogram, N_FFT=2048, Hop=1024 (Accuracy: 97.71%) - 1.0s, Spectrogram, N_FFT=4096, Hop=1024 (Accuracy: 97.64%)

3.2 The Impact of Clip Duration

A primary research question was determining the necessity of long-term musical context. Interestingly, our results consistently show that models trained on 1.0-second clips outperformed those trained on longer 2.0-second and 5.0-second clips on average, and achieved the absolute highest peak accuracy.

For example, while the best 5.0-second model achieved an excellent 98.37% accuracy, it still fell short of the 1.0-second peak of 99.01%. Furthermore, several 2.0-second and 5.0-second models exhibited severe degradation under certain STFT parameters (e.g., 2.0s Spectrogram with 1024/512 completely failed, achieving only ~9% accuracy, indicating catastrophic training instability or extreme overfitting to a majority class mapping).

This counterintuitive finding suggests that the CNN does not require extended musical phrasing, structural understanding, or macro-tempo analysis. Instead, the network effectively categorizes the audio based on highly localized, instantaneous acoustic events. While factors such as micro-timing variations and note attack precision play a role, we must also critically consider confounding variables.

3.3 Critical Limitations: The “Clever Hans” Effect and Acoustic Confounders

From a rigorous scientific perspective, achieving 99% accuracy from merely 1.0 second of audio is not a triumph of model architecture; rather, it is a glaring red flag. Music, inherently a temporal art form, requires the analysis of phrasing, dynamics over time, and structural coherence to assess true proficiency. The fact that the CNN succeeds almost perfectly on a 1.0-second microscopic window strongly implies that the network is entirely bypassing musical semantics and instead relying on static timbral differences. Essentially, the model is performing acoustic scene classification rather than musical skill evaluation.

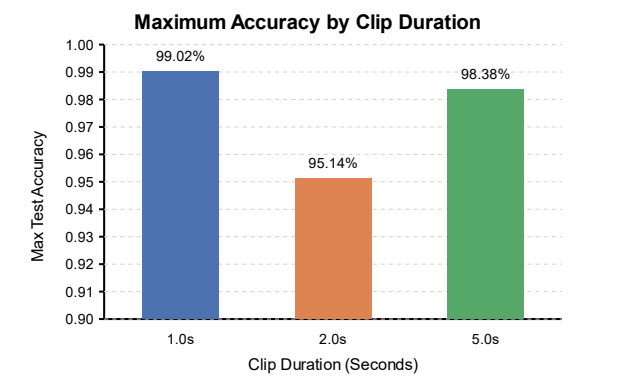


Figure 2. Maximum test accuracy achieved across different audio clip durations. The shortest evaluated duration (1.0 second) yielded the highest peak performance.

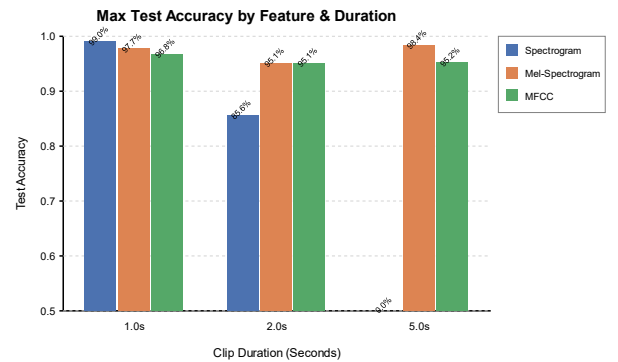


Figure 3. Maximum test accuracy achieved by each feature representation across the three evaluated temporal durations. Spectrograms consistently yielded the highest peak performance.

This phenomenon is a textbook example of the “Clever Hans” effect in machine learning—where a model arrives at the correct answer for entirely the wrong reasons [4]. The Maestro dataset consists of virtuosic performances recorded on pristine Yamaha Disklavier grand pianos in highly controlled, professional studio environments [2]. Conversely, the PianoVAM dataset features amateur learners recorded on a chaotic variety of consumer-grade upright pianos, cheap digital keyboards, and smartphone microphones in untreated, reverberant rooms.

Consequently, the distinct acoustic characteristics, baseline noise floors, and instrument resonance profiles serve as a massive, unmitigated confounding variable. The model’s “exceptional” performance at the 1.0-second scale is an illusion; it has merely learned the trivial task of classifying “High-End Studio Audio” versus “Noisy Home Recording Audio.” Treating this as a successful assessment of musical proficiency is fundamentally flawed. Future research in this domain must either strictly normalize acoustic environments across datasets or explicitly decouple timbral artifacts from performance timing, lest models continue to exploit these superficial acoustic shortcuts.

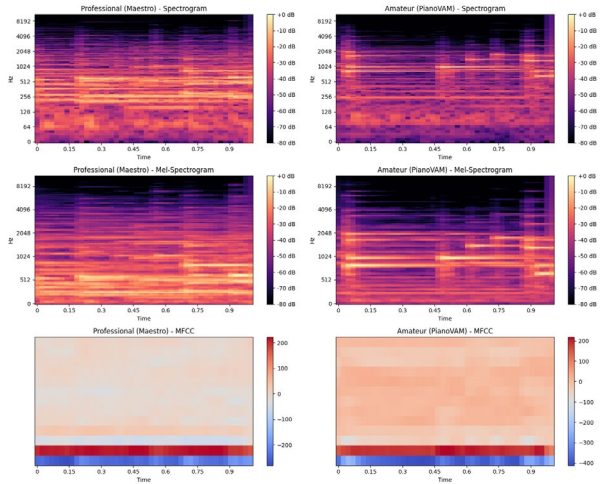


Figure 4. Visual comparison of Spectrogram, Mel-spectrogram, and MFCC features extracted from both professional and amateur performances. Notice the loss of harmonic detail in the MFCC representations.

3.4 Feature Representation Comparison

We evaluated three different 2D representations. While Mel-spectrograms provided very stable and competitive results across various durations (securing two of the top

five spots), the absolute best result was obtained using standard linear Spectrograms.

MFCCs, conversely, generally yielded the lowest performance across the grid search. For instance, the best 1.0-second MFCC model only achieved 96.78% accuracy, and many MFCC configurations hovered in the 60-80% accuracy range. MFCCs heavily compress spectral information to capture the broad spectral envelope, which is traditionally useful for speech recognition. However, in the context of piano music, this extreme compression discards the fine-grained harmonic details, overtone structures, and sympathetic resonance artifacts that are crucial for identifying the nuanced physical interactions a professional pianist has with the instrument compared to an amateur.

3.5 Impact of STFT Parameters

The choice of N_FFT and Hop Length proved to be highly impactful, though highly dependent on the chosen feature and duration. The optimal model utilized $N_FFT=2048$ and $Hop=1024$. The relatively large hop size (1024 at 22.05 kHz equals $\sim 46ms$) suggests that extremely high temporal resolution is less critical than having a stable, well-defined frequency representation (provided by the 2048 FFT window) for this specific classification task.

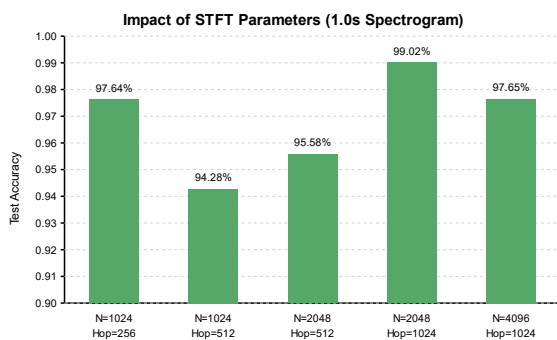


Figure 5. Test accuracy across different STFT window sizes (N_FFT) and hop lengths for the 1.0-second Spectrogram models.

4. Conclusion

In this paper, we initially set out to distinguish between professional and amateur piano performances using a shallow Convolutional Neural Network applied directly to raw 2D audio representations.

Through an extensive grid search of temporal and spectral parameters, we discovered that 1.0-second audio snippets utilizing standard linear Spectrograms ($N_FFT=2048$, $Hop=1024$) yielded an astronomical peak test accuracy of 99.01%.

However, rather than claiming this as a definitive success in automated music evaluation, we critically conclude that this high accuracy is an artifact of severe dataset bias. The model's reliance on microscopic 1.0-second timbral snapshots exposes that it is circumventing actual musical analysis—such as phrasing, rhythm, and structural context—in favor of exploiting acoustic confounders. By inadvertently classifying the pristine studio environments of the Maestro dataset against the untreated home environments of the PianoVAM dataset, the model acts as an acoustic environment detector rather than a music critic.

Our findings serve as a stark cautionary tale for the Music Information Retrieval (MIR) community: when applying deep learning to subjectively complex tasks like skill evaluation across heterogeneous datasets, raw accuracy metrics can be highly deceptive [4], [9]. Moving forward, the field must prioritize rigorous acoustic normalization and the creation of acoustically homogeneous datasets. Without these strict controls, machine learning models will inevitably continue to take the path of least resistance, capitalizing on superficial acoustic shortcuts rather than learning the profound intricacies of musical performance.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

References

- [1] A. Lerch, An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics. IEEE Press, 2012.
- [2] C. Hawthorne et al., "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset," in International Conference on Learning Representations (ICLR), 2019.
- [3] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling," in International Symposium on Music Information Retrieval (ISMIR), 2000.
- [4] B. L. Sturm, "A simple method to determine if a music information retrieval system is a 'horse'," IEEE Transactions on Multimedia, vol. 16, no. 6, pp. 1636-1644, 2014.

- [5] S. Hershey et al., “CNN architectures for large-scale audio classification,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
- [6] K. Choi, G. Fazekas, and M. Sandler, “Automatic tagging using deep convolutional neural networks,” in International Society for Music Information Retrieval (ISMIR) Conference, 2016.
- [7] E. Benetos et al., “Automatic music transcription: challenges and future directions,” *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 407-434, 2013.
- [8] S. Dixon, “Automatic extraction of tempo and beat from expressive performances,” *Journal of New Music Research*, vol. 30, no. 1, pp. 39-58, 2001.
- [9] C. J. Lapuschkin et al., “Unmasking Clever Hans predictors and assessing what machines really learn,” *Nature Communications*, vol. 10, no. 1, p. 1096, 2019.

A Self-RAG-Based Triple Factuality Verification Pipeline for Local LLM KARMA

Seung-gyeom KIM¹, Giseop Noh²

¹BE Course, Department of Software and Communications Engineering, Hongik University, South Korea

²Professor, Department of Software and Communications Engineering, Hongik University, South Korea

Corresponding author: Giseop Noh (kafa46@hongik.ac.kr)

Abstract

Large language models (LLMs) can extract entities and relations from unstructured documents to construct or expand knowledge graphs. However, generated triples may be insufficiently grounded in the source text or contain incorrect relations derived from a model's internal knowledge. KARMA is a knowledge graph enrichment framework in which multiple LLM agents perform document analysis, entity and relation extraction, schema alignment, and knowledge integration. To reduce dependence on commercial services, this study implements KARMA with a local LLM and combines it with Self-RAG retrieval and self-evaluation to assess external evidence. The pipeline consists of document preprocessing and semantic chunking, chunk-level evidence evaluation, candidate triple extraction and source-based quality checking, triple-level factuality verification, and knowledge graph acceptance. In a preliminary experiment, an NLP paper was divided into 14 chunks, from which 85 initial triples were extracted; 28 passed the quality check. External documents confirmed the head entity for six triples, but no complete claim including the relation and tail entity was verified. Consequently, no triple was automatically accepted. The results reveal limitations in the temporal coverage of the 2018 Wikipedia corpus and in filtering non-body PDF content, motivating more recent retrieval corpora, improved preprocessing, and review procedures for potentially novel facts.

Keywords: Knowledge Graph, Local LLM, KARMA, Self-RAG, Factuality Verification

1. Introduction

Advances in LLM have increased interest in knowledge graphs as a means of extracting structured knowledge from unstructured data and supporting reasoning over relationships between entities [1]. Conventional knowledge graph construction often relies on domain experts who manually analyze documents or on rule-based natural language processing systems. Although these approaches can provide relatively high reliability, they require substantial time and labor as the volume of documents increases [2].

KARMA addresses this limitation through nine commercial-LLM-based agents that collaboratively and iteratively analyze unstructured documents, extract and validate triples in the form of head entity, relation, and tail

entity, and integrate them into a knowledge graph [3]. The agents divide responsibilities across document normalization, entity and relation extraction, schema alignment, knowledge validation, and conflict resolution.

To reduce API costs in large-scale literature processing, this study replaces the commercial LLMs in KARMA with a local LLM. Response preprocessing and a domain-adaptive schema induction mechanism are also used to mitigate unstable triple output, duplicate entities, and fragmented relation expressions. However, these improvements primarily target the formal and structural quality of extracted results and do not independently verify whether generated triples are factually supported by the source text and external knowledge.

Therefore, this study applies the Self-RAG framework to strengthen evidence-based verification of triples extracted by local LLM KARMA. Input papers or documents are divided into semantically coherent chunks based on paragraph, sentence, and section boundaries to minimize context loss. The complete pipeline is designed to determine whether external retrieval is necessary before retrieving relevant documents and assess their relevance to the input. Candidate triples extracted from the chunks are checked against both the source text and external documents. Finally, the preceding evaluations are combined with knowledge graph acceptance criteria to determine whether each triple should be integrated.

The objectives of this study are threefold. First, it analyzes the structural characteristics of KARMA and Self-RAG. Second, it proposes a local LLM KARMA pipeline that incorporates Self-RAG-based retrieval and self-evaluation, while designing adaptive retrieval as a component for later end-to-end integration. Third, it analyzes limitations of the retriever, preprocessing procedure, and evaluation model identified in the preliminary experiment and presents directions for further improvement.

2. Background

2.1 KARMA

KARMA is a multi-agent framework that automatically enriches an existing knowledge graph by extracting new knowledge from unstructured documents [3]. Knowledge graphs generally represent facts as triples consisting of a head entity, a relation, and a tail entity [1]. For example, the statement "Self-RAG uses adaptive retrieval" can be represented as follows

(Self-RAG, uses, adaptive retrieval)

A key characteristic of KARMA is that no single LLM performs the entire process from document analysis to final knowledge integration. Multiple agents divide the work, and each downstream agent reviews the output produced in the preceding stage. KARMA includes entity discovery, relation extraction, schema alignment, and conflict resolution. These stages help ensure that newly extracted entities and relations conform to a domain-specific knowledge graph schema and reduce the integration of information that logically conflicts with existing knowledge or has low confidence [3].

The original KARMA employs high-performing commercial GPT models so that individual agents can

follow complex instructions and review their outputs [3]. In a local environment, however, model size, GPU memory, and inference speed are constrained. Furthermore, when an error propagates to downstream agents, an incorrect triple may remain in the final output despite multiple review stages.

Accordingly, local LLM KARMA requires not only the preservation of the original agent structure but also a verification process that directly compares generated triples with the source text or external documents. This study applies Self-RAG retrieval-necessity assessment and evidence-based self-evaluation to that verification process.

2.2 Conventional RAG and Self-RAG

Retrieval-Augmented Generation (RAG) enables a language model to retrieve external documents related to an input and use them during generation rather than relying only on knowledge acquired during training [4]. This approach can reduce factual errors by grounding a response in external information when the model's internal knowledge is insufficient or inaccurate. However, conventional RAG commonly retrieves a fixed number of documents without adequately determining whether retrieval is necessary or whether the retrieved documents are relevant to the input [5].

Table 1. Four types of reflection tokens in Self-RAG

Type	Definitions	Output
Retrieve	Is external retrieval necessary?	{yes, no, continue}
ISREL	Is the retrieved document relevant to the input?	{relevant, irrelevant}
ISSUP	Is the generated output supported by the retrieved document?	{fully supported, partially supported, no support}
ISUSE	Is the final output useful to the user?	{5, 4, 3, 2, 1}

Self-RAG addresses these limitations by allowing a language model to retrieve external documents only when needed, generate an output using the retrieved evidence, and evaluate the quality of that output [5]. In addition to

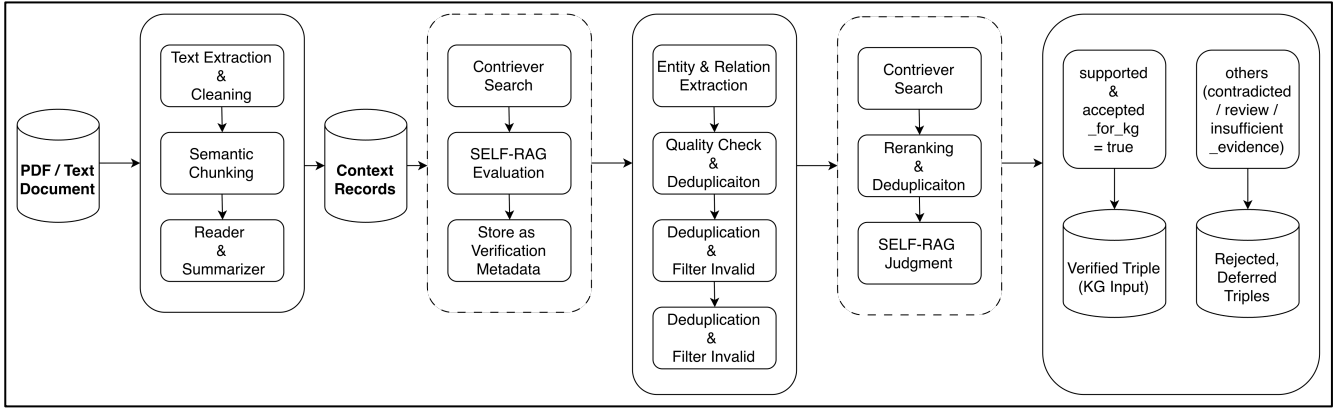


Fig. 1. Flowchart of the Self-RAG-Based Local LLM KARMA Pipeline

ordinary output tokens, the model generates reflection tokens that represent its decisions. These tokens assess whether external retrieval is necessary, whether a retrieved document is relevant to the input, how strongly the generated output is supported by the document, and whether the final output is useful. Self-RAG therefore adaptively determines retrieval necessity instead of retrieving the same number of documents for every input, and it progressively evaluates retrieved evidence and generated content before selecting the final output. The reflection token types are summarized in Table 1 [5].

3. Methods

3.1 Self-RAG-Based Local LLM KARMA

The proposed pipeline comprises document preprocessing and semantic segmentation, chunk-level external evidence evaluation, KARMA-based candidate triple extraction, triple-level factuality verification, and generation of verified knowledge graph input. Both the language model and the retriever run locally without relying on commercial external APIs.

First, the main text is extracted from a PDF or text document and divided into semantic chunks according to paragraph, sentence, and section boundaries. Sections with low relevance to the main content, such as references and acknowledgments, are excluded through section classification. KARMA's Reader then selects passages relevant to knowledge extraction from each chunk, and the Summarizer condenses them to produce a TARGET for triple extraction. The source passage corresponding to each TARGET is retained separately to verify the extraction result. The preceding and following chunks are also attached as auxiliary context to reduce context discontinuity. However, entities and relations are extracted only from the TARGET to prevent surrounding

context from being incorrectly treated as new knowledge. External evidence retrieval uses a corpus built from a 2018 Wikipedia snapshot split into passages of approximately 100 words. In this study, a retrieved document refers to one such passage rather than an entire Wikipedia article. Contriever encodes each TARGET and passage as numerical vectors and computes semantic similarity between them [6]. Passage vectors are stored in a FAISS index. FAISS is a vector similarity search library that efficiently finds items near a query vector in a large vector collection; the index stores search vectors rather than the Wikipedia source text itself [7]. Each preprocessed chunk undergoes Self-RAG-based external evidence evaluation using Contriever and the Wikipedia FAISS index. Contriever returns the top five passages for each TARGET, and Self-RAG evaluates their relevance to and support for the TARGET. A passage is accepted as external evidence only when both the relevance and support scores are at least 0.5 and Self-RAG classifies it as fully supporting the TARGET. The system also checks whether at least 80% of the TARGET's key content words are preserved and whether major entity names and numerical values remain intact. The current implementation retrieves evidence for every chunk without separately predicting retrieval necessity. A chunk is not removed even when no external evidence is accepted. Evaluation results and retrieved passages are stored as verification metadata and passed to the triple extraction stage.

During triple extraction, KARMA's Entity Extraction Agent, Relationship Extraction Agent, and Schema Alignment Agent are executed. These agents extract entities and relations from the TARGET and generate candidate triples after normalizing entity aliases and relation expressions. Each generated triple is checked against the retained source passage for the presence of its head and tail entities, the correspondence of its relation expression, and structural completeness. Results are removed when their components cannot be found in the

Table 2. Results of the pipeline

Metric	Measurement	Result
External Evidence Acceptance Rate	Accepted evidence chunks / Total chunks	0.000
Triple Quality Pass Rate	Valid triples / Initially extracted triples	0.329
Head Evidence Coverage	Triples with head evidence / Valid triples	0.214
Complete Claim Evidence Rate	Triples with complete claim evidence / Valid triples	0.000
Automatic KG Acceptance Rate	Automatically accepted triples / Valid triples	0.000

source, when the relation is excessively long, or when the relation is generated as a full sentence. Ambiguous results are assigned for further review. The original KARMA Conflict Resolution Agent and Evaluator Agent are not directly used in the current integrated pipeline; instead, the subsequent rule-based quality check and Self-RAG verification determine final acceptance.

Triples that pass the quality check are converted into natural-language verification queries. For each query, Contriever retrieves the top ten passages from the same Wikipedia FAISS index. The results are reranked using not only initial embedding similarity but also the presence of head aliases, relation expressions, and the tail entity. Passages that do not correspond to the head entity and near-duplicate passages are removed. Up to three passages are then selected as final evidence from a maximum of five candidates.

During triple verification, a rule-based check first determines whether the head entity, relation, and tail entity can be identified in the retrieved passages. If no passage corresponding to the head entity is found, the triple is labeled `insufficient_evidence`. When verifiable evidence is available, Self-RAG relevance and support assessments are additionally applied. The integrated decision labels a triple as `supported` when external evidence sufficiently supports it, `contradicted` when an explicitly conflicting fact is identified, `review` when partial evidence or conflicting judgments require further examination, and `insufficient_evidence` when sufficient evidence cannot be found.

Only triples labeled `supported` and satisfying `accepted_for_kg=true` are stored as verified knowledge graph input. Triples labeled `contradicted` are rejected, whereas `review` and `insufficient_evidence` cases are retained with their decisions and retrieved evidence for subsequent examination. By separating chunk-level evidence evaluation before extraction from triple-level factuality verification after extraction, the proposed

pipeline preserves source information while automatically accepting only triples with sufficient evidence.

4. Experimental Results and Analysis

This study conducted an experiment on a natural language processing paper to examine the initial behavior and performance of the integrated Self-RAG-based local LLM KARMA pipeline. The experiment was conducted before integration of Reader and Summarizer preprocessing, and the semantically segmented source chunks were therefore used directly as TARGETs. The purpose was to first verify the operation of the core sequence comprising chunk-level evidence evaluation, KARMA-based triple extraction, triple quality checking, and external factuality verification. In addition, the experiment used unconditional retrieval, meaning that external evidence was retrieved for all 14 chunks without applying the retrieval-necessity classifier. Rule-based chunking based on paragraph, sentence, and section boundaries divided the input into 14 chunks, from which KARMA extracted 85 initial triples. The overall results are shown in Table 2.

The chunk-level external evidence acceptance rate is the proportion of chunks for which retrieved passages sufficiently support the TARGET and are accepted as verification evidence. The external corpus consisted of Wikipedia passages of approximately 100 words. Self-RAG evaluated the top five passages returned by Contriever for each TARGET. A passage was accepted only when both relevance and support scores were at least 0.5, Self-RAG labeled it as fully supporting the TARGET, at least 80% of the TARGET's key content words were preserved, and major entity names and numerical values were retained. A total of 70 evidence candidates were retrieved for the 14 chunks, but none was accepted. Self-RAG classified seven chunks as insufficiently relevant to the retrieved passages. The remaining seven achieved

relevance scores of at least 0.5 but failed the support threshold and were judged not to sufficiently support the TARGET or to potentially conflict with it. Consequently, the chunk-level external evidence acceptance rate was 0.000. This result does not indicate that retrieval returned no passages; rather, none of the retrieved passages met the evidence acceptance criteria.

The triple quality pass rate is the proportion of initial triples whose head, relation, and tail are structurally complete and correspond to the source text. Of 85 initial triples, 28 passed the quality check, yielding 0.329. Among the remaining triples, 34 were classified as invalid and 23 as requiring further review. The main causes were a head entity not found in the source text, an insufficient correspondence between the relation and the source expression, and ambiguity requiring additional interpretation. Because some triples contained more than one issue, the sum of issue counts does not equal the number of excluded triples.

The head entity evidence coverage is the proportion of quality-approved triples for which the head entity was found in an external passage. Both the original entity expression and its abbreviations and aliases were considered. The head entity was confirmed for 6 of the 28 valid triples, producing a coverage of 0.214. No directly corresponding external evidence was found for the remaining 22 triples.

The complete claim evidence rate is the proportion of valid triples for which the head, relation, and tail were all confirmed in external passages. Of the six triples with confirmed head entities, four lacked an expression supporting the relation, and the remaining two did not provide evidence for the entire claim including the tail entity. Thus, none of the 28 valid triples had complete claim evidence, resulting in a rate of 0.000. This finding shows that identifying a head entity in an external document alone is insufficient to establish the factuality of the complete triple.

The knowledge graph automatic acceptance rate is the proportion of valid triples that were sufficiently supported by external evidence and accepted as final knowledge graph input. Of the 28 valid triples, 22 were labeled `insufficient_evidence` by the rule-based step because their head entities were not identified in the retrieved passages. The remaining six also lacked sufficient Self-RAG evidence for the complete fact, including the relation and tail entity. Consequently, all 28 triples were classified as `insufficient_evidence`, and none was automatically accepted, yielding an automatic acceptance rate of 0.000. Importantly, `insufficient_evidence` does not mean that a triple is false. It indicates that the current external corpus

did not provide sufficient evidence to determine its factuality. Therefore, unverified triples were not immediately discarded; their decisions and retrieved evidence were retained for later review.

The preliminary experiment also revealed that non-body content, including author names and paper titles in the references as well as table titles and captions, was recognized as entities and used as the head or tail of generated triples. This likely occurred because source chunks were passed directly to triple extraction without prior selection of relevant passages. To mitigate this issue, the subsequent pipeline added a preprocessing stage in which the KARMA Reader selects passages relevant to knowledge extraction and the Summarizer organizes them into TARGETs. Because the complete experiment has not yet been repeated with the Reader and Summarizer enabled, further experiments are needed to quantify improvements in triple quality and external evidence acceptance.

5. Conclusion and Future Work

5.1 Conclusion

This study proposed an integrated pipeline combining Self-RAG retrieval and self-evaluation to verify the factuality of triples generated by local LLM KARMA. The pipeline comprises semantic segmentation of input documents, chunk-level external evidence evaluation, KARMA-based candidate triple extraction, source-based quality checking, triple-level factuality verification, and generation of verified knowledge graph input. The complete design includes adaptive retrieval based on retrieval-necessity assessment. However, because this module had not yet been integrated into the end-to-end pipeline, the preliminary experiment used unconditional retrieval and searched external evidence for every evaluation target.

In the preliminary experiment, the source document was divided into 14 semantic chunks and used directly as TARGETs without Reader and Summarizer preprocessing. KARMA extracted 85 initial triples, of which 28 passed the structural completeness and source-grounding quality criteria. The chunk-level evaluation returned 70 external passages, but none met the evidence acceptance criteria. Only six of the 28 valid triples had a head entity identified in external evidence, and none had evidence for the complete fact including its relation and tail. Accordingly, all 28 triples were labeled `insufficient_evidence`, and none was automatically accepted into the knowledge graph. These results

demonstrate that the current conservative acceptance policy prevents automatic integration of triples lacking sufficient external evidence.

The experiment also exposed limitations in the temporal coverage of the retrieval corpus and in document preprocessing. The corpus was built from a 2018 Wikipedia snapshot, whereas the input paper was published in 2025; therefore, entities and findings introduced after 2018 may not be represented in the corpus. In addition, non-body information such as reference titles and author names, table titles, and captions was recognized as entities and used in generated triples. Because no controlled comparison was conducted to measure the effects of corpus age and preprocessing quality on evidence acceptance, these factors should be interpreted as constraints of the current experimental setting rather than direct causes.

The proposed pipeline operated as a verification mechanism that did not automatically integrate triples lacking sufficient evidence. However, the current experiment cannot determine whether the low acceptance rate primarily resulted from the temporal limitations of the corpus, retrieval performance, document preprocessing errors, or stringent acceptance thresholds. Future work should therefore isolate and evaluate the effect of each factor.

5.2 Future Work

Future experiments should expand the temporal coverage of the retrieval corpus by using up-to-date Wikipedia documents. To objectively evaluate the effect of corpus construction time, comparative experiments should use the 2018 Wikipedia snapshot, a recent Wikipedia snapshot, and recent scholarly literature, and analyze changes in the chunk-level external evidence acceptance rate and complete claim evidence rate.

Preprocessing must also be improved to reliably select valid body text from unstructured documents. The current implementation identifies some non-body sections using section headings and string patterns; however, if PDF conversion destroys layout and paragraph boundaries, references, table titles, and captions are difficult to distinguish from body text. A layout-aware preprocessing method should therefore use positional, font, and page-layout information to separate titles, author information, headers and footers, references, tables, and captions. The full pipeline should also be rerun with the Reader's passage selection and the Summarizer's summarization functions added after the preliminary experiment, and the

reduction in non-body entities and invalid triples should be evaluated.

Because the integrated experiment retrieved evidence for every evaluation target, the separately tested retrieval-necessity classifier should be connected to the chunk and triple verification stages. Adaptive retrieval should then be compared with unconditional retrieval in terms of reductions in unnecessary retrieval, omission of required evidence, external evidence acceptance, and final verification performance.

Finally, triples labeled `insufficient_evidence` may represent not only false information but also novel facts absent from the existing corpus. Such triples should not be immediately discarded. Instead, a separate review process should determine their eligibility for knowledge graph integration by consulting recent scholarly literature, domain-specific databases, or domain experts.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

Acknowledgement

This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency(NIPA) (Grant No. PJTS0801260410070001000300200).

References

- [1] B. Zhang and H. Soh, "Extract, Define, Canonicalize: An LLM-based framework for knowledge graph construction," in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, FL, USA, pp. 9820–9836, 2024, doi: 10.18653/v1/2024.emnlp-main.548.
- [2] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 2, pp. 494–514, 2022, doi: 10.1109/TNNLS.2021.3070843.
- [3] Y. Lu, W. Wu, X. Zhao, R. Peng, and J. Wang, "KARMA: Leveraging multi-agent LLMs for knowledge graph construction," in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, FL, USA, pp. 9820–9836, 2024, doi: 10.18653/v1/2024.emnlp-main.548.

- automated knowledge graph enrichment,” in Advances in Neural Information Processing Systems 38, 2025.
- [4] P. Lewis et al., “Retrieval-Augmented Generation for knowledge-intensive NLP tasks,” Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.
- [5] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, “Self-RAG: Learning to retrieve, generate, and critique through self-reflection,” in The Twelfth International Conference on Learning Representations, 2024.
- [6] G. Izacard, M. Caron, L. Hosseini, S. Riedel, P. Bojanowski, A. Joulin, and E. Grave, “Unsupervised dense information retrieval with contrastive learning,” Transactions on Machine Learning Research, 2022.
- [7] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” IEEE Transactions on Big Data, vol. 7, no. 3, pp. 535–547, 2021, doi: 10.1109/TBDATA.2019.2921572.

Author Information



Seung-gyeom Kim
2022~ now: BE, Dept. of Software &
Communications Engineering, Hongik
University
Email: ksg072896@naver.com
Research Interest: LLM, NLP, RAG



Giseop Noh
1998: BE, Engineering, Korea Air
Force Academy
2009: MS, Computer Science,
University of Colorado, USA
2014: PhD, Computer Engineering,
Seoul National University
2016.12~2018.02: Assistant Professor, Dept. of
Computer & Information Science, Korea Air Force
Academy
2018.03~2025.02: Assistant Professor, Dept. of
Artificial Intelligence Software, Cheongju University
2025.03~now: Assistant Professor, Dept. of Software &
Communications Engineering, Hongik University
Email: kafa46@hongik.ac.kr
Research Interest: NLP, Neural Network Theory,
Reinforcement Learning

Automatic Generation of Emotion Retrospective Content Based on Generative AI

Junsun Chang¹, Janghwan Kim², Chaeyun Seo³, R. Young Chul Kim⁴

¹Graduate Student, Department of Software and Communications Engineering, Hongik University, South Korea

²Ph.D Candidate, Department of Software and Communications Engineering, Hongik University, South Korea

³Ph.D, Department of Software and Communications Engineering, Hongik University, South Korea

⁴Professor, Department of Software and Communications Engineering, Hongik University, USA

Corresponding author: R. Young Chul Kim (bob@hongik.ac.kr)

Acknowledgement: This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency (NIPA) (Grant No. PJTS0801260410070001000300200).

Abstract

With advances in generative AI, the technology for automatically generating images and videos is rapidly developing. However, existing technology for creating videos from static images requires users to write video prompts directly, which requires specialized knowledge and has limitations in producing retrospective content that reflects personal memories and emotions. To solve this, we propose a generative AI-based framework that automatically creates emotional retrospective videos using two photos (static images) taken at the start and end. Our method analyzes the input photo to extract scene, object, and mood information using a large language model (LLM), then automatically generates video prompts that include actions, camera movements, environment, lighting, and emotional elements. The generated prompts are fed into a generative model that converts images into videos, creating emotional retrospective videos. Additionally, the entire generation process is automated to reduce the user's burden of writing prompts and simplify content creation. This study presents a method for the automatic creation of photo-based emotional retrospective content and suggests potential applications of digital retrospective services that utilize generative AI.

Keywords: Machine Learning, Dual Learning, Natural Language Processing

1. Introduction

Image-to-Video technology, enabled by recent advances in generative AI, is increasingly applied to digital reminiscence services alongside content creation. However, existing services remain limited to static content such as photos, failing to convey individuals' movements or the atmosphere of the time, and typically require users to write prompts—a barrier for senior users unfamiliar with generative AI [1].

To address these limitations, we propose an AI-based interface that automatically generates emotional reminiscence videos from a single photograph. The system analyzes the input photo, automatically generates video prompts using a large language model (LLM), and produces short reminiscence videos depicting human movement and environmental atmosphere via an Image-to-Video model. By eliminating the need for manual prompt writing, the proposed system transforms static photos into vivid videos, enhancing user immersion and demonstrating the potential of generative AI for emotional reminiscence content.

Chapter 2 discusses DreamVideo as a related study, and Chapter 3 discusses the proposed Automatic Generation of Emotion Retrospective content. Finally, Chapter 4 discusses the conclusion and future research.

2. Related Works

DreamVideo [2] is an Image-to-Video (I2V) model that generates videos from a single input image conditioned on a text prompt. Since conventional I2V models rely solely on the semantic information of the input image, visual attributes such as facial features, clothing, and background often change inconsistently across frames. To address this, DreamVideo introduces an Image Retention Branch, which extracts CNN-based features from the input image and passes them to the Video Diffusion model to preserve fine-grained visual details, and a Double-condition Classifier-Free Guidance (CFG) scheme, which jointly conditions generation on both the input image and the text prompt—enabling diverse actions (e.g., walking, running, dancing) from the same image. Experimental results show that DreamVideo achieves better preservation of input image features and more natural video generation than existing I2V models.

However, DreamVideo primarily focuses on preserving visual consistency and does not account for semantic information such as a character's personality or role. It has limitations in generating long videos or in expressing complex storylines. Thus, in domains requiring character consistency and narrative continuity—such as educational storytelling—additional methods are needed to complement this approach.

3. Automatic Generation of Emotion Retrospective Content

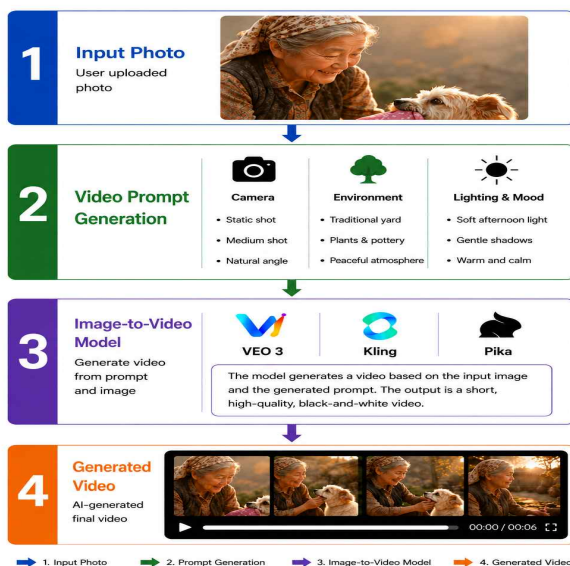


Fig. 1. Flowchart of data processing and analysis

As shown in Figure 1, the proposed system consists of four stages. In Step 1 (Input Photo), the user uploads a single photograph, requiring no additional text input and making the system accessible even to users unfamiliar with generative AI.

In Step 2 (Video Prompt Generation), the LLM analyzes the image to extract Action, Camera, Environment, Lighting, and Mood/Style elements, converting them into video-generation prompts. The Camera element defines shot composition (e.g., static shot, medium shot), Environment defines the background context (e.g., traditional yard, peaceful atmosphere), and Lighting/Mood sets the overall tone (e.g., soft afternoon light, warm and calm). A Negative Prompt is also generated to minimize distortions and improve video quality.

In Step 3 (Image-to-Video Model), the generated prompts and the original image are jointly fed into the Image-to-Video model, which preserves the visual characteristics of the person and background while capturing natural movements such as walking or interaction.

In Step 4 (Generated Video), a short reminiscence video capturing both human movement and the atmosphere of the time is delivered to the user, conveying a sense of vividness that static photographs alone cannot provide.

Overall, this pipeline removes the burden of prompt writing while automatically generating emotional reminiscence videos from a single photograph.

4. Conclusion

We propose an AI-based interface that automatically generates emotional reminiscence videos from a single photograph. The system analyzes the input image, uses an LLM to generate video and negative prompts automatically, and applies an Image-to-Video model to produce short videos depicting human movement and the atmosphere of the time. By eliminating the need for manual prompt writing, the proposed system lowers the barrier for users unfamiliar with generative AI while enhancing immersion in digital reminiscence content.

However, the system does not account for semantic information such as a character's personality or narrative context and remains limited in generating long videos or complex storylines. Future work will focus on incorporating character-consistent generation techniques, extending the system to multi-photo, story-based video

generation, and validating its effectiveness through usability studies with senior users.

Conflict of Interest

The authors declare that there are no potential conflicts of interest related to this paper.

Acknowledgement

This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency (NIPA) (Grant No. PJTS0801260410070001000300200).

References

- [1] Z. Zhang, L. Xu, X. Wang, X. Zhang, & M. Fan, (2025). "Understanding and co-designing photo-based reminiscence with older adults.", in *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1-30.
- [2] C. Wang, J. Gu, P. Hu, Y. Guo, X. Dong, H. Xu, & X. Liang (2025, April). "Dreamvideo: High-fidelity image-to-video generation with image retention and text guidance." In *Proceedings of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

Author Information



JunSun Chang
now: Graduate Student
Hongik University, South Korea
Email: cjsun0331@hanmail.net
Research Interest: Natural Language
Processing, Machine Learning, Prompt

Engineering



Janghwan Kim
now: Ph.D Candidate, Department
of Software and Communications
Engineering, Hongik University, South
Korea
Email: lentoconstante@hongik.ac.kr

Research Interest: AI software validation, reinforcement learning-based software validation, software visualization, and requirements engineering.



Chaeyun Seo
2014: Ph.D, Department of Computer
Engineering, Hongik University
Email: chaeyun@hongik.ac.kr
Research Interest: Metaverse,
Metaverse Contents, Design Thinking,
CBD, Software Engineering, BPM, AI, Business Process
Modeling.



R. Young Chul Kim
2000: Ph.D. Degree,
Software Engineering from the
Department of Computer Science,
Illinois Institute of Technology, USA
Email: bob@hongik.ac.kr

Research Interest: Test Maturity Model, Model-Based
Testing, Metamodeling, Software Process Model,
Software Visualization, AI software validation,
reinforcement learning-based software validation.

Award Recipients

2026 International Conference on PyGeek (PyGeek 2026) · Vol. 1, 2026

Best Paper Award

A Factor Decomposition Method for Separating Store Substitution and Embedding Placement Effects in Vector Database Migration

Junha Yoon, Ohhyeon Gwon, Junhaeng Lee, Hyeji Roh, Soowan Cho, Suhee Kim, Sangsoon Lim

Sungkyul University; Chung-Ang University; Deep Insight Lab

Accepted Papers: 73–77

Outstanding Paper Award

Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting

junggyu, Park, Giseop Noh

Deep Shark Lab, Hongik University

Accepted Papers: 84–89

A Co-evolutionary Strategy of AI Technology for Expanding Regional Industrial Value Chains

KIM EU JUNG

Hongik University

Accepted Papers: 68–72

Best Presentation Award

Multi-scale Temporal-aware Text-Enhanced Network for Multimodal Sentiment Analysis

Junjun Zhang, Giseop Noh

Henan University of Engineering; Hongik University

Official presenter(s): Junjun Zhang

Accepted Papers: 78–83

Proceedings of the Conference on PyGeek

Vol. 1 · 2026

2026 International Conference on PyGeek (PyGeek 2026)

ISSN (Online): 3140-5312

<https://pygeek.acin.kr/proceedings/2026>

