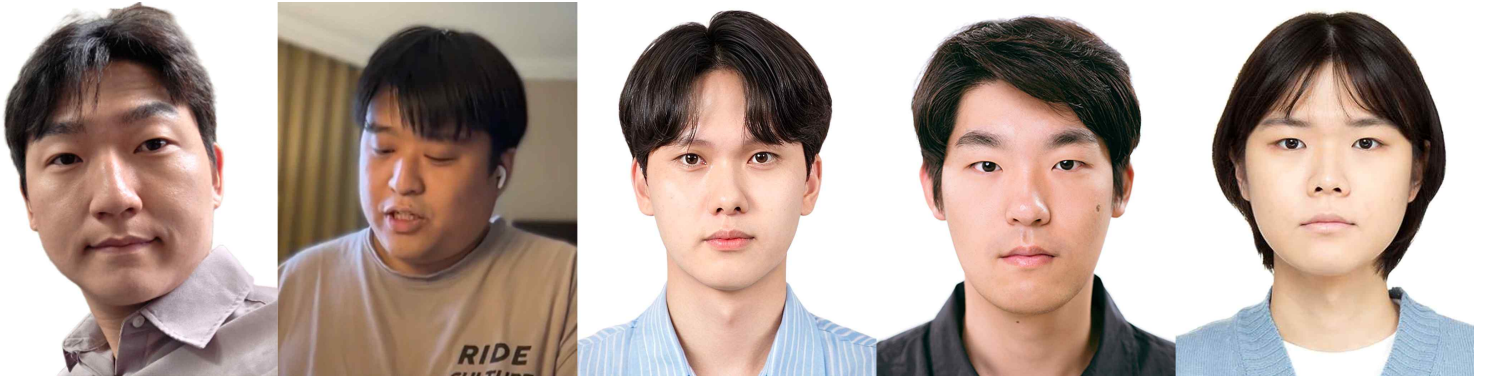


“사람의 풀이 과정을 따라 ‘의도’를 이해하는 AI”

GIST, 인간처럼 추론하는 AI 구현

기존 모델 대비 정확도 5.85%p 향상 (83.59% → 89.44%)

- AI융합학과 김선동 교수팀, 사람 풀이 과정 속 의도를 추정·정렬하는 ‘의도 학습 알고리즘’ 제안 및 다양한 풀이 경로 만들어 내는 ‘데이터 증강 기법’ 개발
- 정렬과 증강 결합해 단순 정답 도출 넘어 ‘사람 같은 추론(human-like reasoning)’ 가능성 입증... 데이터과학 최고 학회 ‘ACM SIGKDD 2025’ 구두 발표 및 국제학술지 《TMLR》 게재



▲ GIST AI융합학부 김선동 교수(KDD 및 TMLR 논문 교신저자), 김세진 박사후연구원(KDD 논문 제1저자 및 TMLR 논문 공동교신저자), 황산하 석사과정 졸업생(TMLR 논문 제1저자), 이승필 석사과정생(TMLR 논문 제2저자), 전기전자컴퓨터공학과 이호성 학사 졸업생 (KDD 논문 제2저자)

광주과학기술원(GIST, 총장 임기철)은 **AI융합학과 김선동 교수 연구팀이 사람의 문제 풀이 과정에 담긴 ‘의도’를 추정·정렬하는 학습 알고리즘과 AI 생성모델***을 결합해, **사람처럼 다양한 풀이 과정을 만들어 내는 데이터 증강*** 기법을 제안했다고 밝혔다.

연구팀은 두 접근법을 결합함으로써, **인공지능이 단순히 정답을 도출하는 수준을 넘어 ‘사람처럼 추론하는 능력’**을 구현할 수 있음을 입증했다.

* **생성모델(Generative Model)**: 기존 데이터를 단순히 분류하거나 예측하는 것이 아니라, 새로운 데이터(예: 이미지, 텍스트, 문제 풀이 과정)를 만들어내는 인공지능 모델. 대표적으로 GAN, VAE, 대규모 언어모델(LLM), GFlowNet 등이 있다.

* **데이터 증강(Solution augmentation)**: 기존 데이터에서 다양한 풀이 과정을 새롭게 생성해 학습 데이터로 활용하는 방법. 이번 연구에서는 GFlowNet을 활용해 구현했다.

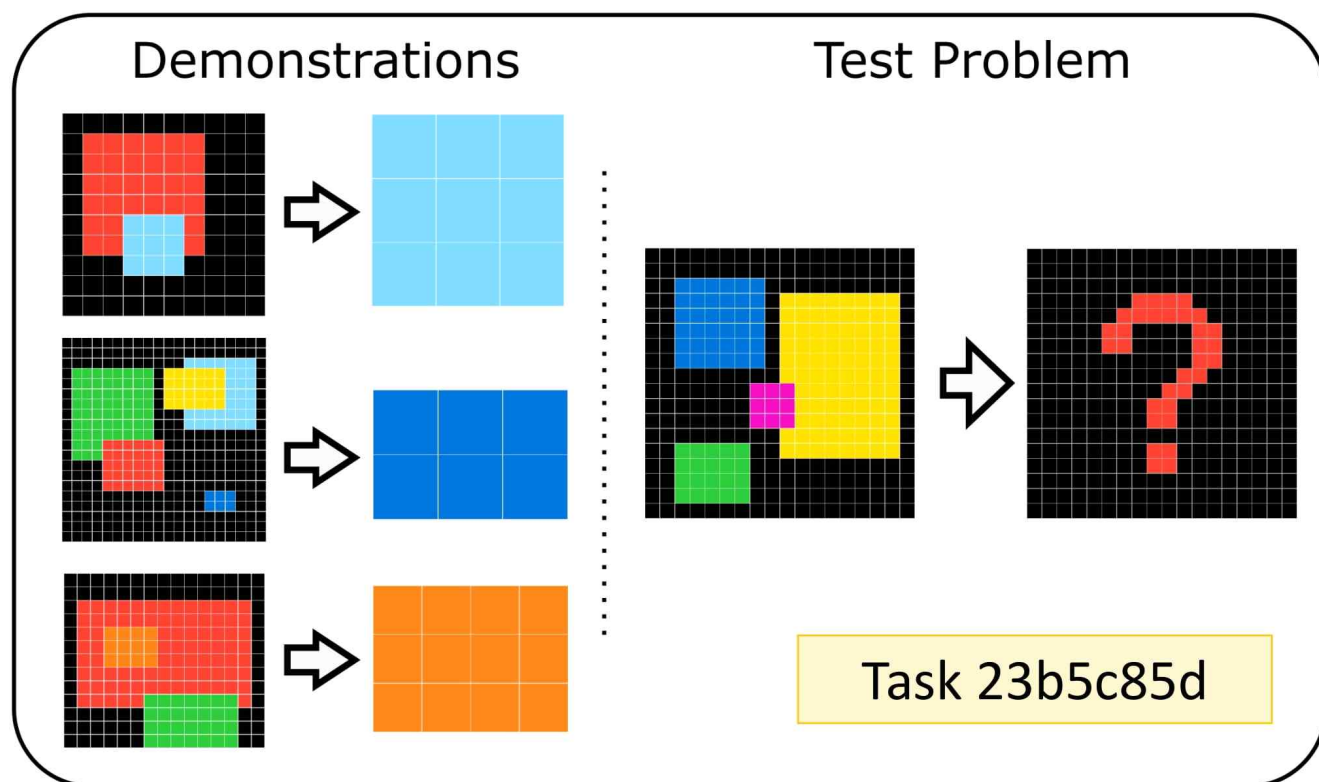
* **사람처럼 추론하는 능력(human-like reasoning)**: 단순 계산이 아닌, 인간처럼 사고 과정을 거쳐 답을 도출하는 능력.

오늘날 인공지능은 주어진 문제의 정답을 빠르게 도출하는 데 강점을 보이지만, **사람처럼 단계적 사고 과정을 거쳐 문제를 해결하는 추론 능력은 여전히 부족하다.**

인간은 문제를 풀 때 시행착오를 겪으며, 같은 목표를 여러 방식으로 시도한다. 이 과정에서 축적된 풀이 데이터에는 단순한 행동의 나열이 아니라 **의도***가 담겨 있다. 이러한 의도를 학습하는 것이 **‘사람처럼 추론하는 능력’**을 갖춘 AI의 핵심이다.

* **의도(intention)**: 문제 해결 과정에서 사람이 궁극적으로 달성하려는 목표나 전략. 같은 의도가 여러 행동 패턴으로 나타날 수 있다.

연구팀은 이러한 관점에서, 문제 풀이 과정 속 의도를 추정·정렬*하는 알고리즘을 개발하고, 생성모델 중 하나인 지플로우넷(GFlowNet*)을 활용해 다양한 풀이 궤적을 생성하는 데이터 증강 기법을 제안했다.



▲ 풀이 과정이 담긴 궤적 데이터를 수집하기 위한 문제의 예시. 왼쪽의 그림 쌍들로부터 문제의 규칙을 파악하여, 오른쪽의 그림 쌍을 완성해야 하는 ARC-AGI 벤치마크의 문제 중 하나를 시각화했다. 이 문제의 규칙은 왼쪽 그림에서 가장 면적이 작은 직사각형을 골라 오른쪽에 배치하는 것이다.

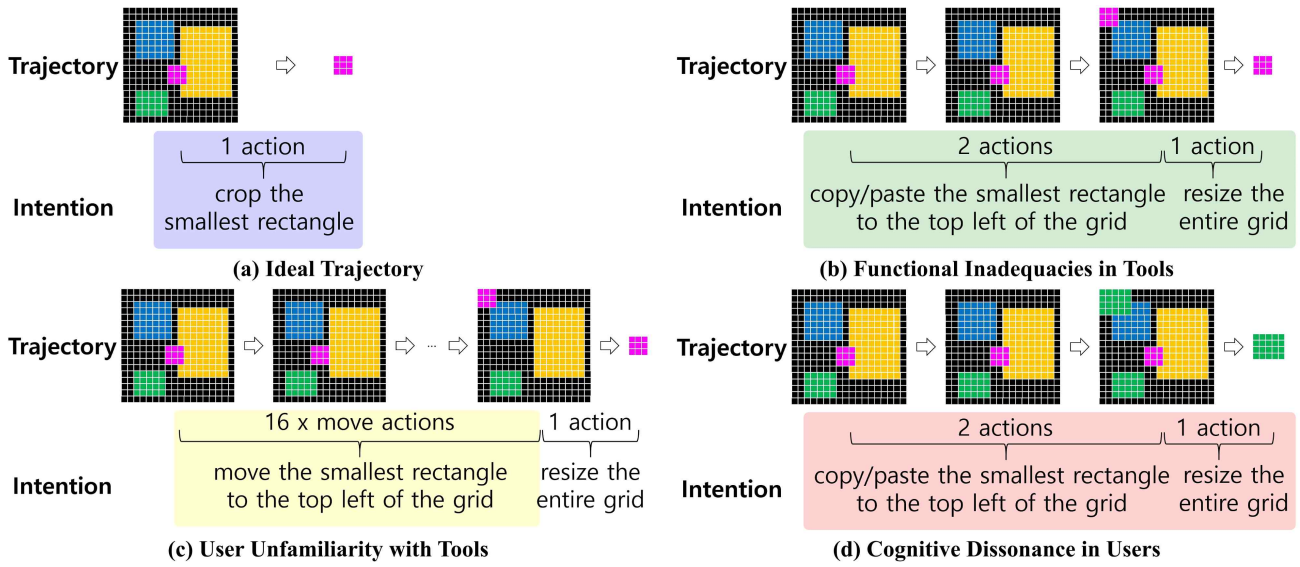
전자는 사람의 사고 과정을 반영하는 방식이고, 후자는 풀이 과정을 확장해 일반화 성능을 강화하는 방식으로, 두 접근은 상호 보완적으로 작용한다. 이를 통해 단순 정답 도출을 넘어 인간과 유사한 사고·일반화 능력을 갖춘 AI 모델 개발을 목표로 했다.

* **정렬(alignment)**: AI가 추정한 의도를 실제 문제 풀이 과정과 일치시키는 과정.

* **GFlowNet(Generative Flow Network)**: 확률적 탐색을 통해 다양한 정답 후보를 생성하는 생성 모델 기법.

연구팀은 사람의 문제 풀이 데이터를 분석해, 최적 풀이에 도달하지 못한 경우를 세 가지 유형(① 기능 부족, ② 비효율적 시도, ③ 잘못된 전략)으로 분류하고 이를 학습에 반영했다.

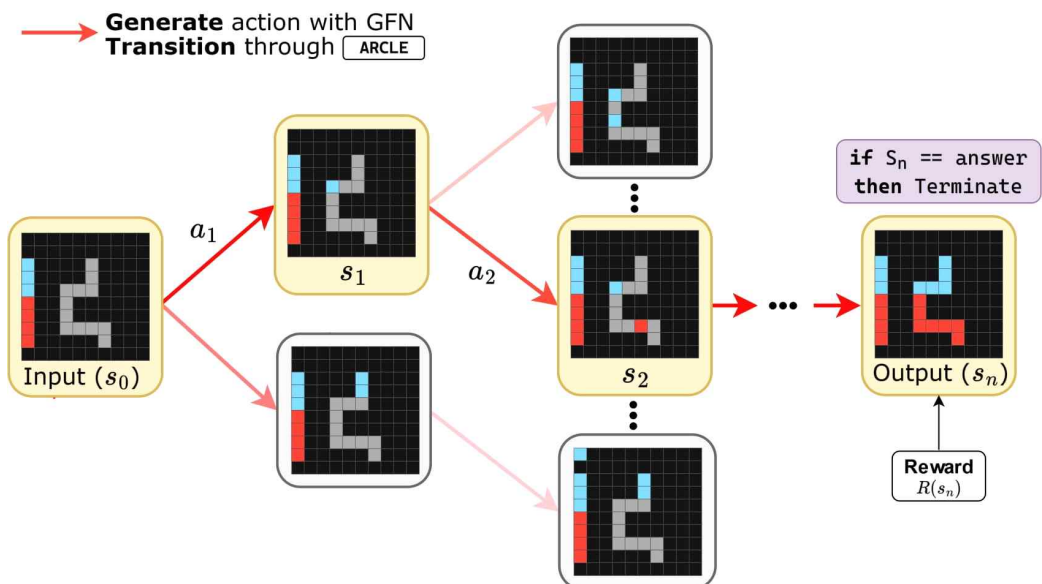
이어 문제 풀이 궤적을 여러 단계로 분할해 각 단계의 의도를 추정·정렬하는 알고리즘을 제시하고, 이를 AI 학습에 반영함으로써 사람의 사고 과정을 모방한 학습을 구현했다.



▲ **다양한 문제 풀이 궤적과 의도 사례.** 본 그림은 사람이 동일한 문제를 풀 때 나타날 수 있는 다양한 궤적과 그 이면에 존재하는 의도를 보여준다. (a) 이상적 궤적(Ideal Trajectory): 최소한의 행동으로 목표를 달성하는 가장 효율적인 풀이. (b) 도구적 한계(Functional Inadequacies in Tools): 사용 가능한 기능의 제약으로 인해 불필요한 단계가 추가된 경우. (c) 도구 미숙(User Unfamiliarity with Tools): 도구에 익숙하지 않아 불필요하게 많은 단계를 거친 풀이. (d) 인지적 혼란(Cognitive Dissonance in Users): 사용자가 다른 전략을 선택해 사용자의 의도와 실제 풀이 과정이 달라진 경우.

또한 지플로우넷 기반 데이터 증강 기법을 통해 사람처럼 다양한 풀이 과정을 생성함으로써 학습 데이터의 다양성과 일반화 성능을 크게 확장했다.

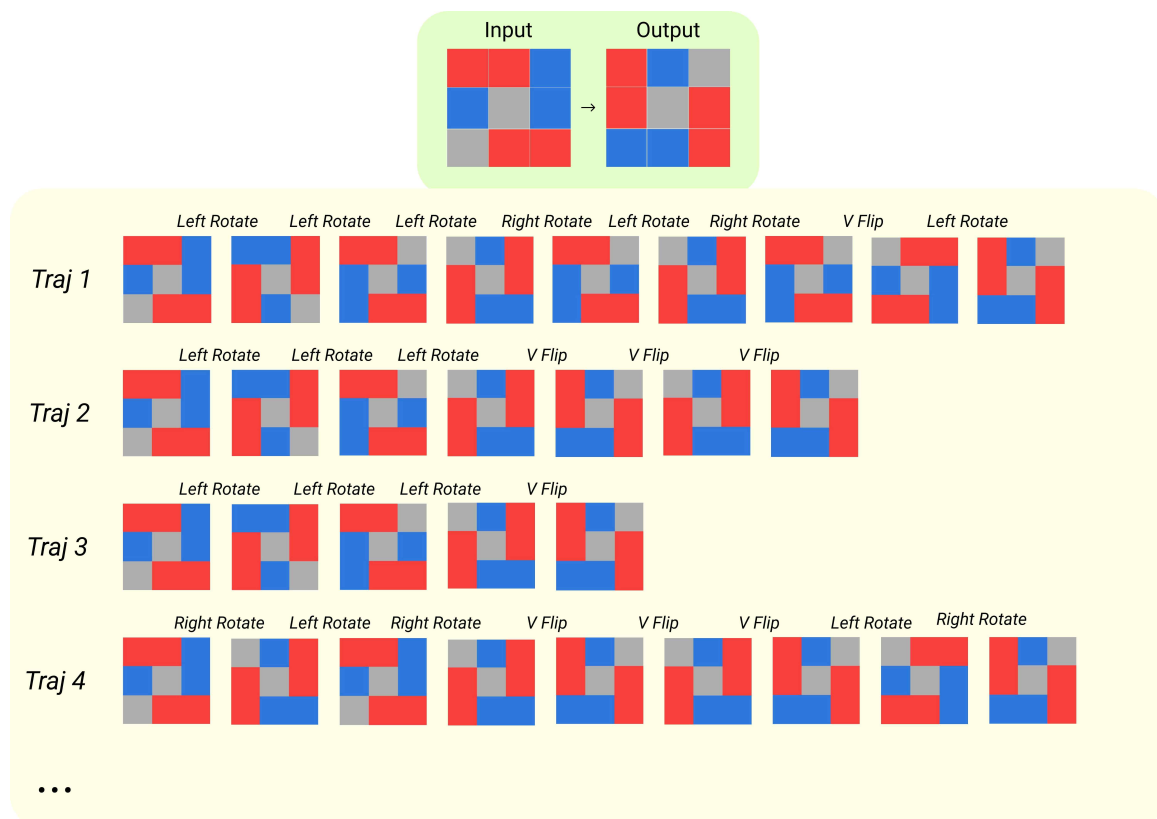
지플로우넷은 확률적 탐색으로 여러 중간 상태를 거치며 다수의 풀이 경로를 생성한다. 동일한 입력에 대해 회전·대칭 등 다양한 연산 조합으로 다수의 궤적을 만들어 학습에 활용하기 때문에, 특정 풀이 방식에 치우치지 않고 다양한 전략 분포를 학습할 수 있다.



▲ **지플로우넷 기반 문제 풀이 궤적 생성 과정.** 지플로우넷은 입력 상태에서 출발하여 일련의 행동을 통해 다양한 중간 상태를 탐색한 뒤, 최종적으로 정답 출력에 도달한다. 이러한 확률적 탐색 과정을 통해 사람처럼 여러 풀이 과정을 생성할 수 있으며, 학습 데이터의 다양성을 크게 확장해 일반화 성능 향상에 기여한다.

연구팀은 실제 사람으로부터 풀이 궤적 데이터를 수집하고, 궤적이 부족한 경우에는 지플로우넷으로 생성해 보충했다.

이렇게 확보한 다양한 풀이 데이터를 학습에 적용한 결과, 기존 모델 대비 정확도가 약 5.85%p 향상됐다(83.59% → 89.44%). 이는 AI가 사람처럼 사고하고 일반화할 수 있음을 입증한 것이다.



▲ 동일한 대각선 방향 뒤집기 문제에 대해 생성된 다양한 풀이 궤적. 하나의 입력-출력 쌍에 대해 여러 가지 변환 연산(회전, 대칭 등)을 거쳐 동일한 정답에 도달할 수 있다. 연구팀은 이러한 다양한 풀이 과정을 GFlowNet으로 생성해 학습 데이터로 활용함으로써, AI가 사람처럼 여러 전략을 고려하는 능력을 갖추도록 했다.

김선동 교수는 “인간은 늘 정석적인 풀이 과정을 따르기보다, 익숙한 방식으로 요령 있게 해답을 찾아내는 경우가 많지만 AI 모델 학습에서는 인간이 수집한 데이터를 별 고민 없이 사용하는 사례가 흔하다”며 “이번 논문에서 제시한 전·후처리 과정을 체계적으로 적용하면 데이터 내재적 한계를 극복하고, 보다 바람직한 행동을 구현하는 AI를 개발할 수 있을 것”이라고 말했다.

이어 “또한 이번 연구에서 제시한 데이터 생성 기법은 인간 풀이 데이터가 부족한 분야에서 학습 데이터를 확장하는 데 크게 기여할 것”이라고 덧붙였다.

김세진 박사는 “이번 연구는 단순히 정답을 맞히는 AI를 넘어, 사람의 문제 풀이 의도를 이해하고 다양한 풀이 과정을 스스로 학습할 수 있는 AI를 구현하기 위한 첫걸음”이라며 “향후 창의적 문제 해결형 AI, 교육 협력형 AI, 새로운 환경 적응형 AI 등으로 확장될 수 있을 것”이라고 전망했다.

GIST AI융합학부 김선동 교수가 지도하고 김세진 박사후연구원과 이호성 학사 졸업생이 수행한 '사람의 문제 풀이 과정에 내포된 의도 추정 알고리즘 개발 연구'는 과학기술정보통신부 산하 정보통신기획평가원(IITP)의 정보통신·방송 기술개발사업, 디지털혁신기술 국제공동연구 사업 및 인공지능대학원 지원사업, 한국연구재단(NRF)의 중견연구자지원사업 및 박사후국내연수지원사업, 그리고 GIST 박사후연구원 Value-up 프로그램의 지원을 받았다.

AI융합학부 황산하 석사과정생(졸업), 이승필 석사과정생, 김세진 박사후연구원이 수행한 'GFlowNet 기반 문제 풀이 궤적 증강 연구'는 과학기술정보통신부 산하 정보통신기획평가원(IITP)의 정보통신·방송 기술개발사업, 디지털혁신기술 국제공동연구 사업, 한국연구재단(NRF)의 중견연구자지원사업 및 박사후국내연수지원사업, 그리고 GIST 미래선도형특성화연구사업의 지원을 받았다.

연구 성과는 데이터과학 분야 세계 최고 수준의 학술대회인 'ACM SIGKDD* 2025'에서 8월 6일 구두 발표됐으며, 국제학술지 《트랜잭션 온 머신 러닝 리서치(Transactions on Machine Learning Research, TMLR*)》에 2025년 9월호에 게재됐다.

* **ACM SIGKDD(Conference on Knowledge Discovery & Data Mining)**: 국제 컴퓨터학회(ACM) 산하 데이터마이닝 및 지식발견 특별흥미그룹(SIGKDD)이 주관하는 세계 최고 수준의 데이터 과학 학술대회로, 인공지능·머신러닝·빅데이터 분석 분야의 최신 연구 성과와 산업 적용 사례가 발표된다.

* **Transactions on Machine Learning Research(TMLR)**: 기계학습 분야의 연구 성과를 신속하게 공유하기 위해 창간된 국제 학술지로, 신경정보처리시스템학회(NeurIPS)가 운영을 지원하고 있다. 기존 저널보다 빠른 심사와 공개 리뷰 과정을 통해 투명성과 연구자 간의 활발한 교류를 지향하며, 머신러닝 이론부터 응용까지 폭넓은 주제를 다룬다.

한편 GIST는 이번 연구 성과가 학술적 의의와 함께 산업적 응용 가능성까지 고려한 것으로, 기술이전 관련 협의는 기술사업화센터(hgmoon@gist.ac.kr)를 통해 진행할 수 있다고 밝혔다.

논문의 주요 정보

1. 논문명, 저자정보

[논문1]

- 학회명 : ACM SIGKDD 2025 (Computer Science 분야 한국연구재단 BK21+ 우수국제학술대회 기준 인정 IF 4)

- 논문명 : Addressing and Visualizing Misalignments in Human Task-Solving Trajectories
- 저자 정보 : 김세진(제1저자, GIST AI융합학과 박사후연구원)
이호성(제2저자, GIST 전기전자컴퓨터공학과 학사 졸업)
김선동(교신저자, GIST AI융합학과 조교수)

[논문2]

- 저널명 : Transactions on Machine Learning Research (2022년에 창간된 신생저널이라 Impact Factor 미산정, Scopus 등재)
- 논문명 : Solution Augmentation for ARC Problems Using GFlowNet: A Probabilistic Exploration Approach
- 저자 정보 : 황산하(제1저자, GIST AI융합학과 석사 졸업)
이승필(제2저자, GIST AI융합학과 석사과정생)
김세진(공동 교신저자, GIST AI융합학과 박사후연구원)
김선동(공동 교신저자, GIST AI융합학과 조교수)