

2022 KAIST Tech Fair KAIST 기술이전 설명회



인간처럼 생각하는 AI: 교육/인사/게임/광고/핀테크/행동치료용 뇌기반 인공지능 기술

이상완

신경과학-인공지능 융합연구센터장
뇌인지공학 프로그램 책임교수
바이오및뇌공학과 부교수
헬스사이언스 연구소 겸임교수

KAIST

연구목표

Exploring the nature of the brain's computation

Two-fold approach:

- “Brain→AI” aims to understand HOW the brain learns
- “AI→Brain” aims to understand WHY the brain implements such computations.

Brain→AI

Understanding HOW the brain learns

- Theory-driven task design
- Latent task policy profiling

AIxBrain

Intersection of machine learning and neuroscience

- Brain-inspired AI
- Computational psychiatry
- BCI/neural decoding

AI→Brain

Understanding WHY the brain implements such computations

- Neuroscience of reinforcement learning
- AI-human co-evolution engine

연구실적 요약

- 저널 43편, 학회 130편
- **Journal papers:**
Cell Reports (2021), *Nature Communications* (2019), *Science Robotics* (2019), *Neural Networks* (2020), *PLOS Biology* (2015), *Neuron* (2014), *PLOS Computational Biology* (2022), *PLOS Computational Biology* (2021)
- **Proceeding papers (top AI conference):**
AAAI (2021), *ICCV* oral (2021), *ICCV* (2021), *EMNLP* (2020), *ICASSP* oral (2019)
- **Workshop papers (top AI conference):**
NeurIPS (2021), *ICLR* (2021), *NeurIPS* (2019)
- **Poster presentations (top computational neuroscience conference):**
2 *COSYNE* (2022), 2 *COSYNE* (2021), 1 *COSYNE* (2020), 1 *COSYNE* (2019), 5 *COSYNE* (2018), 5 *RLDM* (2019)

수상실적 요약

- Google 교수 연구상 (2016)
- IBM 학술상 (2021)
- KAIST 송암 미래석학연구상 (2019), KAIST 국제협력상 (2022) 등

Affiliations



Sponsors



Lab information

http://aibrain.kaist.ac.kr
sangwan@kaist.ac.kr
E16-1 YangBoonSoon Bldg.
Room 401 (CNAI research center)
Room 404 (Lab)
Room 516 (Sang Wan Lee)

International collaborators

Ben Seymour (Oxford)
Joel Leibo (DeepMind)
Mattia Rigotti (IBM AI Research)
Kevin Miller (UCL/DeepMind)
Benedetto De Martino (UCL)
Doo-Sup Choi (Mayo Clinic)
John O'Doherty (Caltech)



Contents

- 01 연구개발 배경
- 02 기술 개요
- 03 기술의 특징
- 04 기술의 효과
- 05 산업분야 및 시장 규모
- 06 사업화 방안

01. 연구개발 배경

딥러닝 기반 인공지능은 인간처럼 행동할 수 있지만, 인간처럼 생각하지 않는다.

→ 인공지능과 인간(사용자) 협업 과정의 부조화 (Alignment problem)

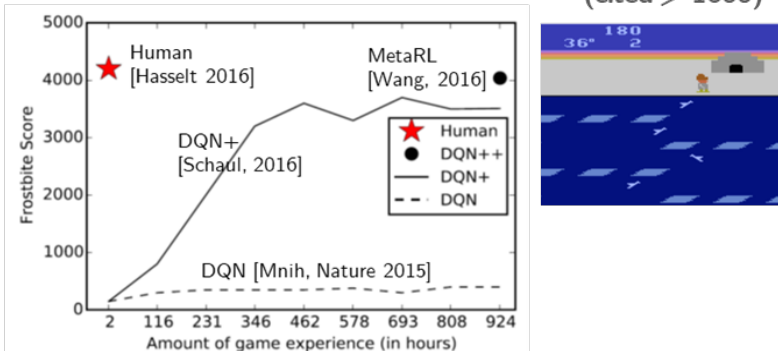
→ 관련된 7가지 이슈는 신간 ‘인공지능과 뇌는 어떻게 생각하는가’ 참조

Atari (2015), AlphaGo (2016), AlphaGo0 (2017),
POMDP환경(2019) AlphaFold (2020), ...



AS-IS

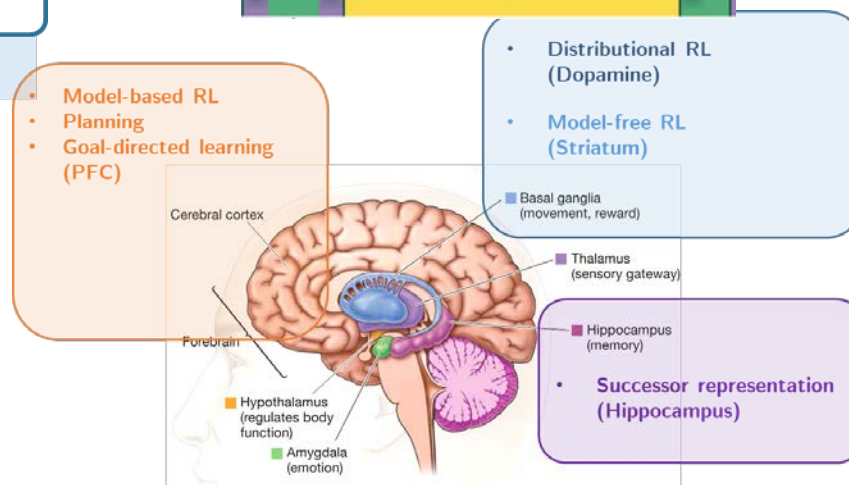
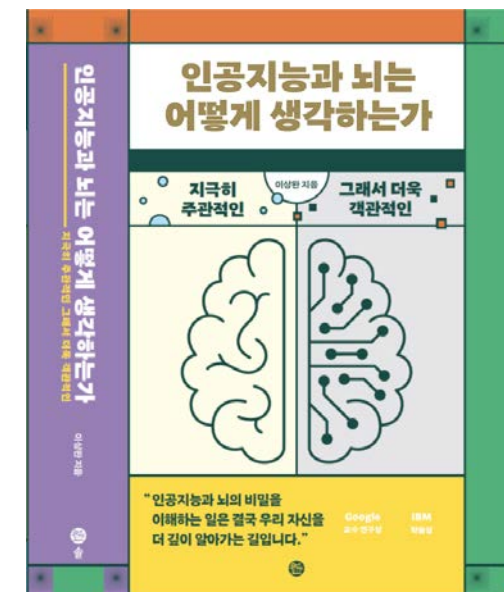
“Building machines that learn and think like people”
(cited > 1600)



TO-BE

인간만의 독특한 사고체계 예시

- 상대방을 이기지 않고 계속 플레이 (목적 기반 학습)
- 환경 물리 법칙 파악 (직관 학습)
- 작업의 근본 원리 파악 (관계 학습, 언어지도 학습)
- 한번 보고 다른 상황 예측 (이론 학습, 원샷 학습)
- 스스로의 학습능력에 대한 확신성 (메타인지 학습)
- 과소적합-과적합 없는 상황 적응력 (적응학습)



뇌의 사고체계 - AI 모델 매칭 필요

인간처럼 생각하는 AI: 전략적 의사결정/학습/추론 프로파일링 핵심 및 응용기술

최근 5년간 총 20건 = 핵심기술 13 + 응용기술 7
(*해외 특허출원 13건 포함)

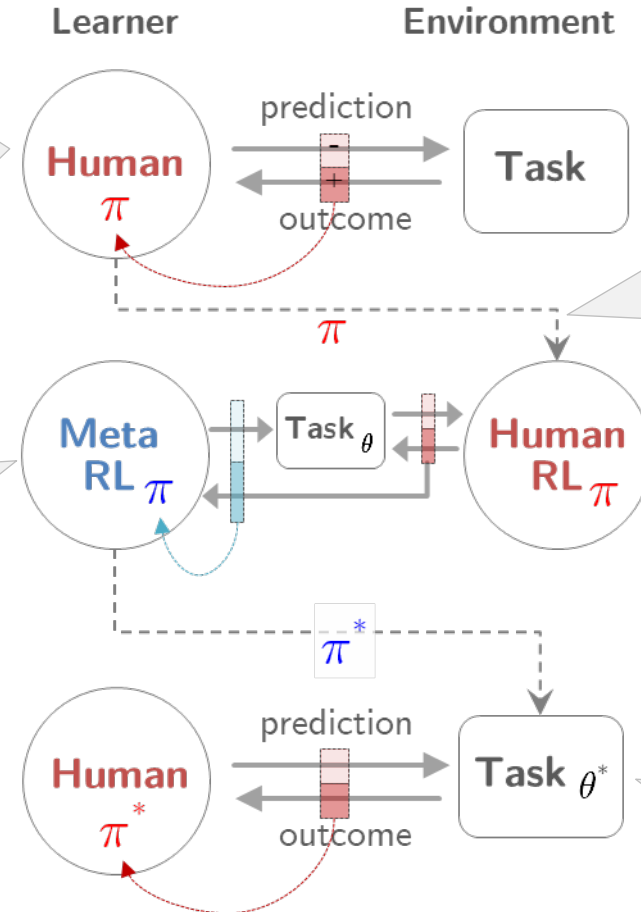
(1) 표준 데이터 취득 파이프라인

(간단한 의사결정 게임 기반)

- 인간 고위인지기능-AI 이식 파이프라인 10-2020-0028772 (KR/US/JP/PCT 출원)*
- 일반화 가능한 가상의 사용자 모델링 10-2020-0126999 (KR 출원)
- 실험 최적화 및 가설 자동생성 10-2203336-0000 (KR등록)
- 사용자 지식 베이스 자가 성장 시스템 10-2018-0145428 (KR등록)
- 다중 도메인 학습 기술 10-2020-0183859 (KR/US/PCT/CN 출원)*
- 제로 도메인 원샷학습 10-2020-0109802 (KR/US 출원)*

(3) 인간-AI 정렬 기술

- 인간학습/추론 제어 및 훈련 기술 10-2132529-0000 (KR등록, US출원)*
- 게임 사용자 전략 유도 기술 10-2111857-0000 (KR등록, US출원)*



(2) 인간의 전략적 의사결정 능력 프로파일링 기술

(교육/인사/게임/광고/핀테크/인지행동치료/BCI)

- 의사결정 프로파일링 10-2021-0035103 (KR/US/PCT/CN출원)*
- 의사결정 불확실성 추론 10-2020-0166376 (KR/US출원)*
- 행동전략 프로파일링 110-2190584-0000 (KR등록)
- 메타인지학습 프로파일링 10-2019-0056870 (KR등록, PCT/US/EP출원)*
- 신경발화 프로파일링 10-2020-0156897 (KR/US출원)*
- 정신질환 프로파일링 10-2020-0156681 (KR/US/JP/PCT출원)*
- 대중인식 프로파일링 10-2021-0158836 (KR출원)

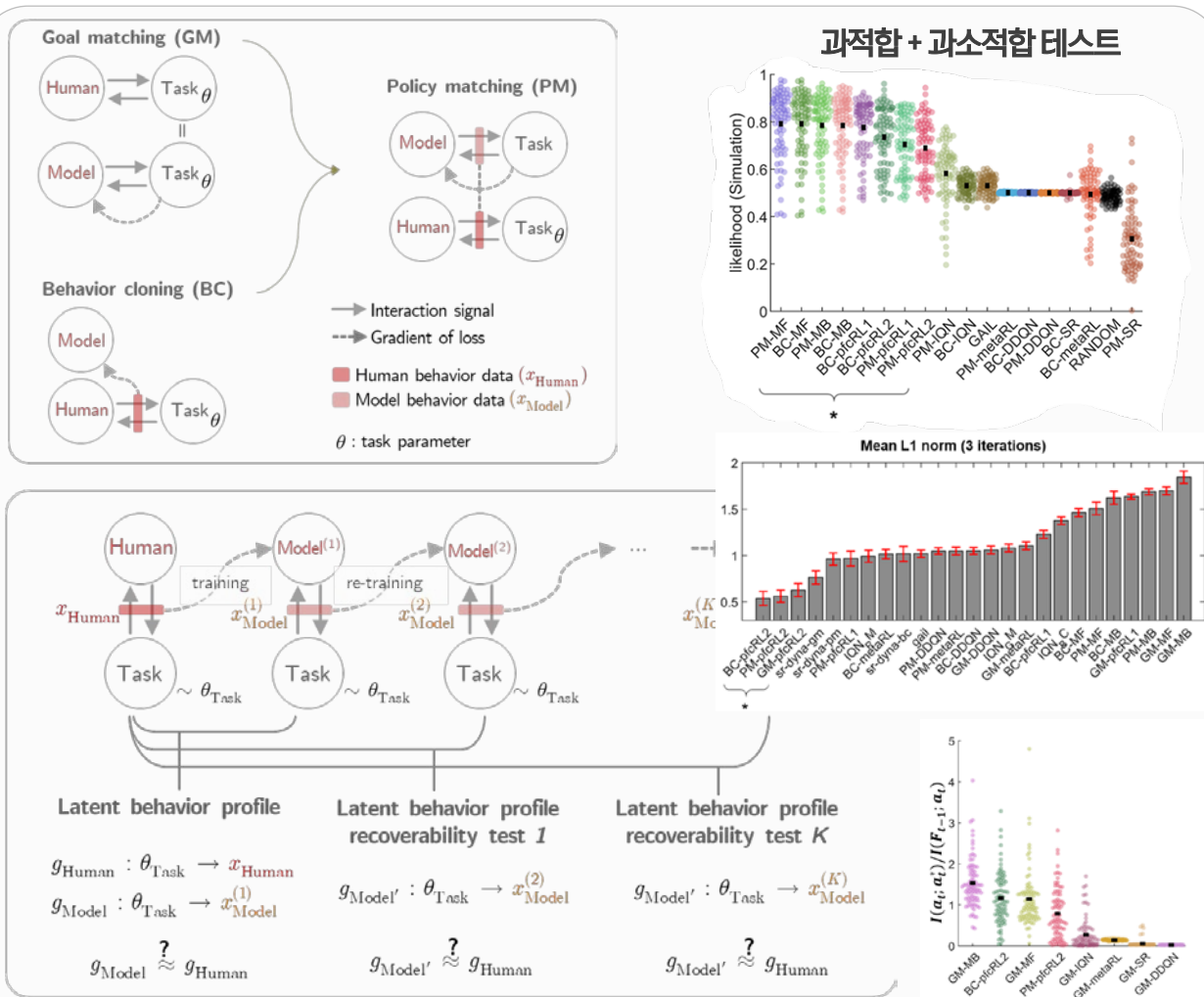
(4) 뇌기반 인공지능 응용기술

(게임/인지행동치료/BCI)

- 인간처럼 유추하는 자기지도 에이전트 10-2022-0001866 (KR/US출원)*
- 뇌기반 인공지능 BCI 기술1 10-2305124-0000 (KR등록, US출원)*
- 뇌기반 인공지능 BCI 기술2 10-2130211-0000 (KR등록, US출원)*
- 적응형 뇌파 분석 10-2111857-0000 (KR등록)
- 우울증 조기진단용 프로파일링 10-2245353-0000 (KR등록)

03. 기술의 특징

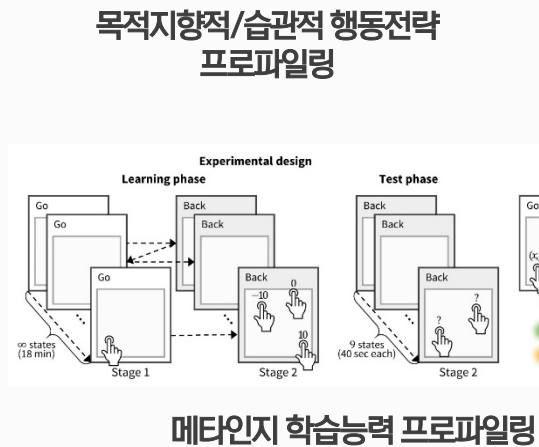
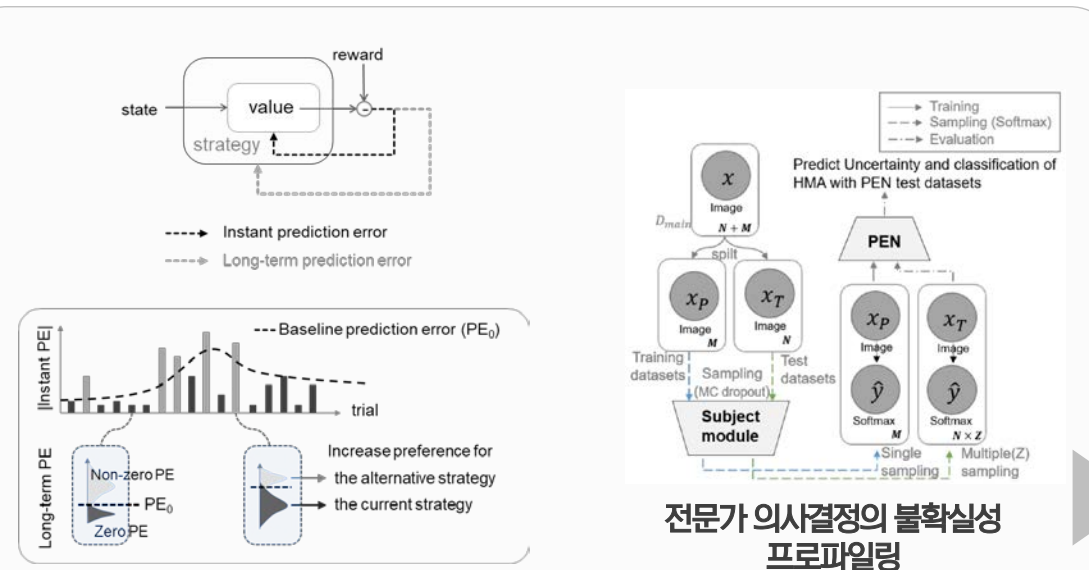
(1) 표준 데이터 취득 파이프라인



인간의 전략적 의사결정 모델링 파이프라인

(2) 인간의 전략적 의사결정 능력 프로파일링 기술

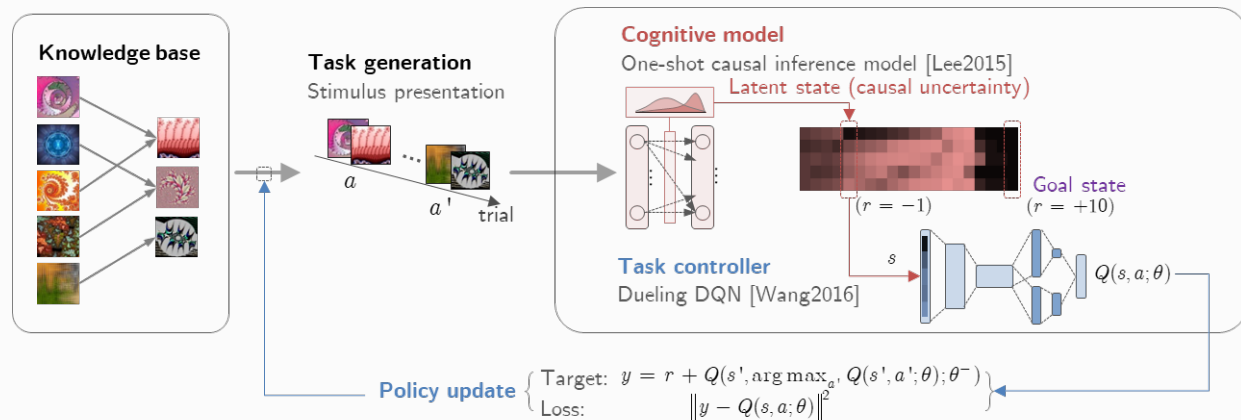
(교육/인사/게임/광고/핀테크/인지행동치료/BCI)



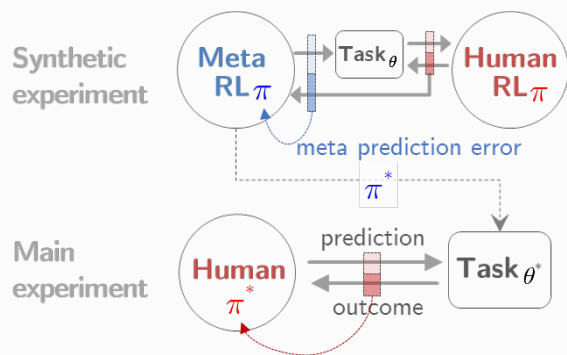
메타인지 학습능력 프로파일링

03. 기술의 특징

(3) 인간-AI 정렬 기술

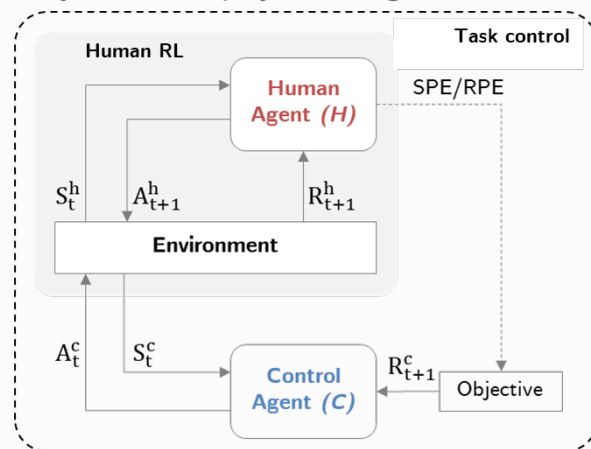


인간의 고속 추론능력 유도/훈련기술



인간의 목적기반 학습능력 유도/훈련기술

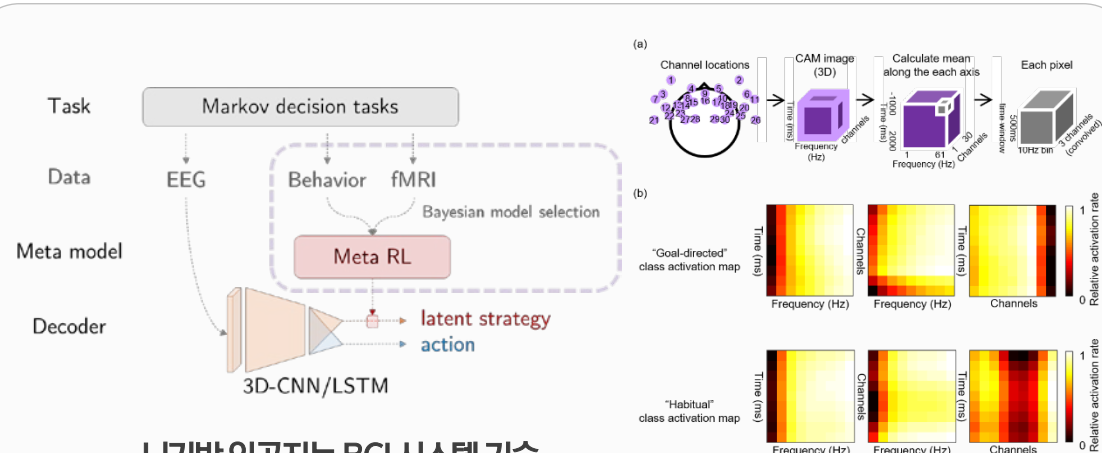
Asymmetric 2-player setting



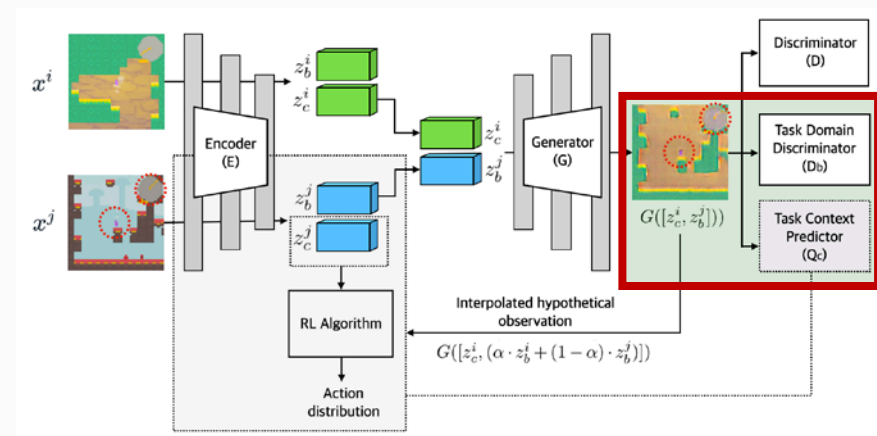
게임 사용자 전략유도 엔진

(4) 뇌기반 인공지능 응용기술

(게임/인지행동치료/BCI)



뇌기반 인공지능 BCI 시스템 기술



인간처럼 유희하는 자기지도 에이전트

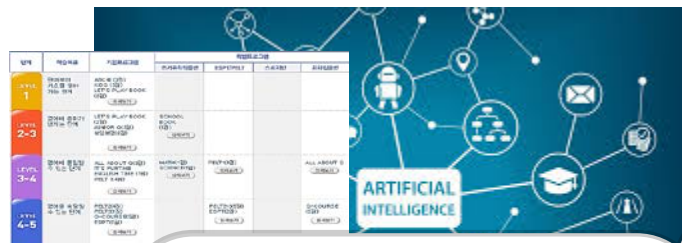
04. 기술의 효과

- ✓ **AI와 인간이 상호작용하는 모든 기술에 적용 가능**
- ✓ **복잡한 지식에 대한 학습/추론이 요구되는 다양한 분야에 적용 가능**

디지털 헬스케어



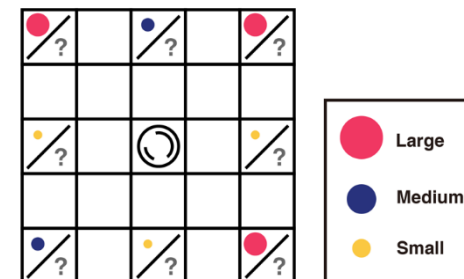
스마트 교육



뉴로모픽 반도체



인지행동치료 (중독진단/몰입 및 집중력 향상/ 실패 강건성 향상)



인간-로봇 상호작용



게임 사용자 프로파일링/컨텐츠



핀테크 (금융/보험상품개발,시장예측)



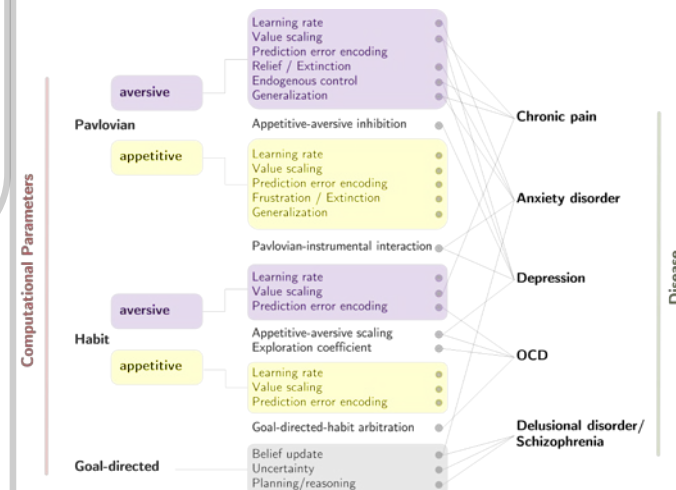
차세대 HR



스마트 광고

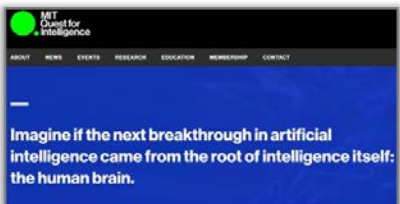


정신질환 프로파일링



05. 산업분야 및 시장 규모

세계적 동향 (딥마인드, 구글, IBM 등에서 전략적 확장)



MIT



Harvard



IBM



Google DeepMind



Oxford center for neuroscience and AI



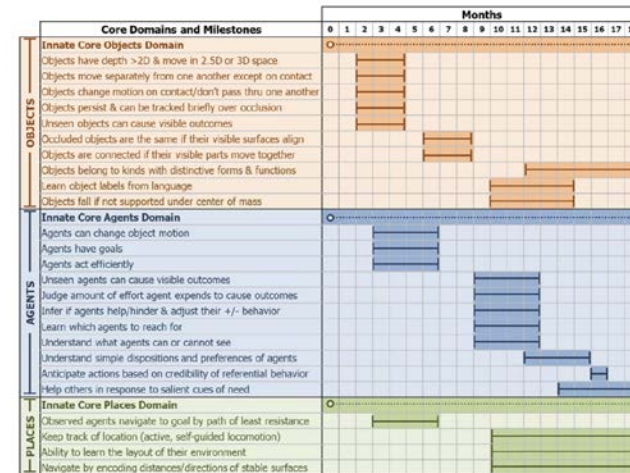
Japan brain-AI center

스마트 교육 시장

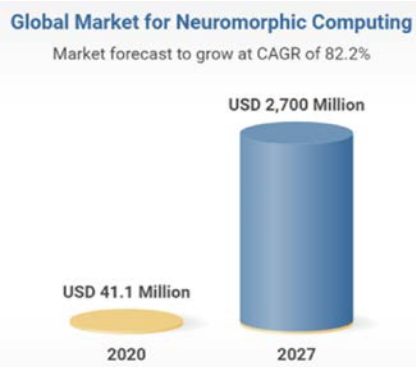


미국방성 전략적 R&D투자

DARPA의 차세대AI분야에 common sense R&D (>200억)



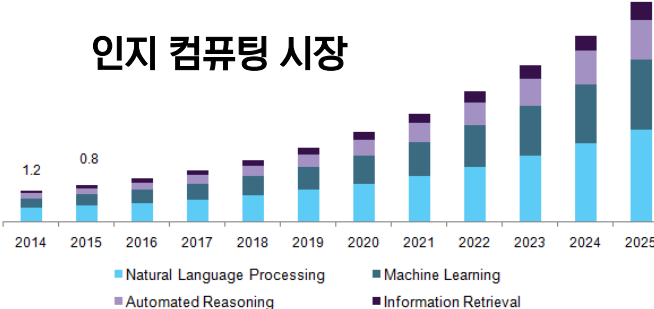
차세대 반도체 시장



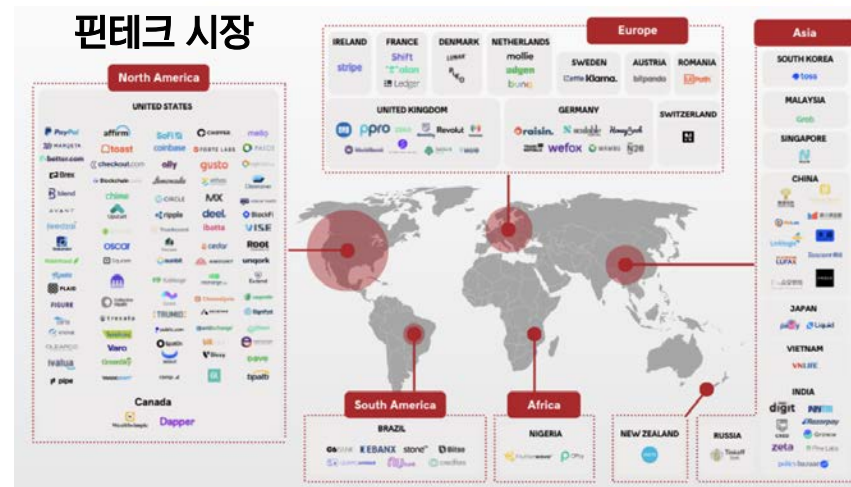
차세대 HR



인지 컴퓨팅 시장

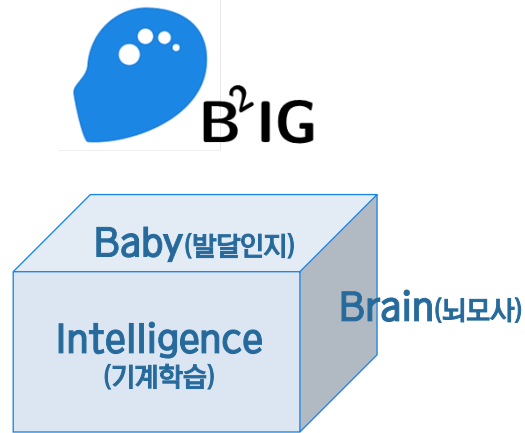


핀테크 시장

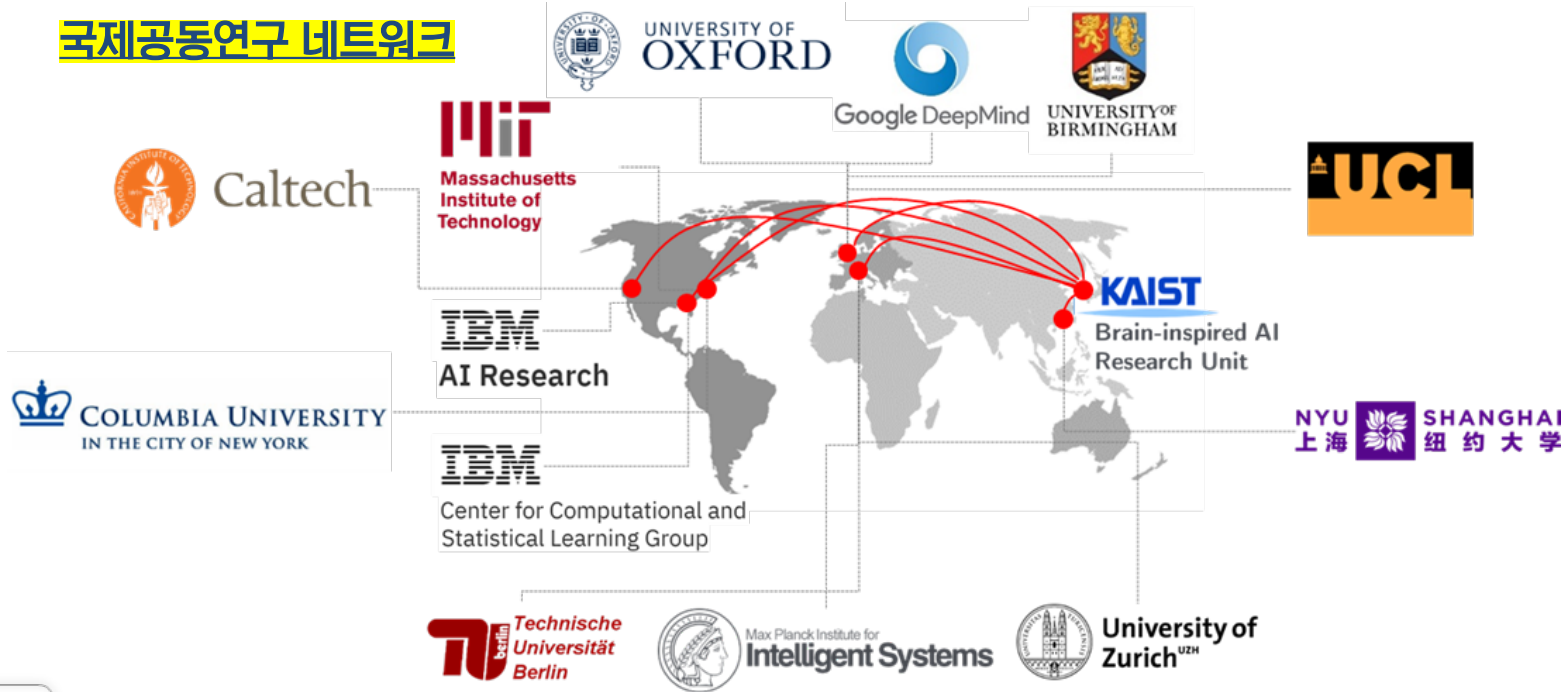


06. 사업화 방안

신경과학-인공지능 융합연구센터 (Center for Neuroscience-inspired AI)



국제공동연구 네트워크



조직도



06. 사업화 방안

신경과학-인공지능 융합연구센터 (Center for Neuroscience-inspired AI)



- 기술 리스트 <https://cna.kaist.ac.kr/32>
- 기술 사업화 문의 kaist-cnai@googlegroups.com

Center for Neuroscience-inspired Artificial Intelligence

ABOUT B&I G ON MEMBER RESEARCH NEWS

B&I G system

유틸리티 및 App

- Various environments for testing human cognitive models
 - Dynamic pong (Link)
 - Infinite bandit task (Link)
 - Unlabeled Uncertainty pong (Link)
 - Unlabeled robot hand tasks (Link)
 - Contextual POMDP meta RL task (Link)
- LAV/STARlip-Audio-Visualization-STAR (Link)
- Procedy (Link) DOI: 10.5281/zenodo.4309929
 - Byung-Ju Choi, Hummel
- VEGA: A Toolkit for Building Virtual Environments to Train and Test Human-like Agents (Link)
 - Youngho Lee, Seoul National University

Database

- Computational model of arbitration and fMRI results (Link)
 - Sang Wan Lee, KAIST
- Raw data of 'infinitely-many armed bandit task' and simulation codes for metacognitive exploration (Link)
 - Sang Wan Lee, KAIST
- Ko corpus of Korean mother-child interaction (Link)
 - Eun Suk Ko, Chosun University
- Evaluating the LENA System for Korean: Supplementary Materials (Link)
 - Eun Suk Ko, Chosun University
- Korean mothers' use of tactile cues for word learning (Link)
 - Eun Suk Ko, Chosun University
- Conversational turns (Link)
 - Eun Suk Ko, Chosun University

- Softwares
- Computational model of arbitration and fMRI results (Link) DOI: 10.5281/zenodo.4309427
 - Sang Wan Lee, KAIST
 - Raw data of 'infinitely-many armed bandit task' and simulation codes for metacognitive exploration (Link) DOI: 10.5281/zenodo.4309030
 - Sang Wan Lee, KAIST
 - Unsupervised 3D-Reconstruction (Link)
 - Songhwa Oh, Seoul National University
 - Deep-Elastic-Network (Link)
 - Songhwa Oh, Seoul National University
 - Soft Action Particle Deep Reinforcement Learning for a Continuous Action Space (Link)
 - Songhwa Oh, Seoul National University
 - Trimming the f, Regularizer: Statistical Analysis, Optimization, and Applications to Deep Learning (Link) DOI: 10.5281/zenodo.4309278
 - Eunho Yang, KAIST
 - AM2-MAML (Link) DOI: 10.5281/zenodo.4309034
 - Eunho Yang, KAIST
 - Interpreting Deep Neural Networks - Relative Attributing Propagation (Link) DOI: 10.5281/zenodo.4309506
 - Sang Wan Lee, KAIST
 - RL Task Controller Simulation (Link) DOI: 10.5281/zenodo.5751202
 - Sang Wan Lee, KAIST
 - Human RL Task Controller (Link) DOI: 10.5281/zenodo.5751206
 - Sang Wan Lee, KAIST
 - fMRI_dimensionality (Link) DOI: 10.5281/zenodo.578443
 - Sang Wan Lee, KAIST
 - Depression-study (Link) DOI: 10.5281/zenodo.578449
 - Sang Wan Lee, KAIST
 - Meta-control of social learning strategies (Link)
 - Sang Wan Lee, KAIST
 - Ensemble-diversified Actor-Critic (EDAC) (Link) DOI: 10.5281/zenodo.578595
 - Hyun Oh Song, Seoul National University
 - CKD - Distilling Content Linguistic for Language model compression (Link) DOI: 10.5281/zenodo.5751074
 - Eunho Yang, KAIST
 - WPBERT (Link) DOI: 10.5281/zenodo.5754453
 - Kyogu Lee, Seoul National University

기술설명회



2022 KAIST Tech Fair KAIST 기술이전 설명회

감사합니다.



<https://aibrain.kaist.ac.kr/>
sangwan@kaist.ac.kr