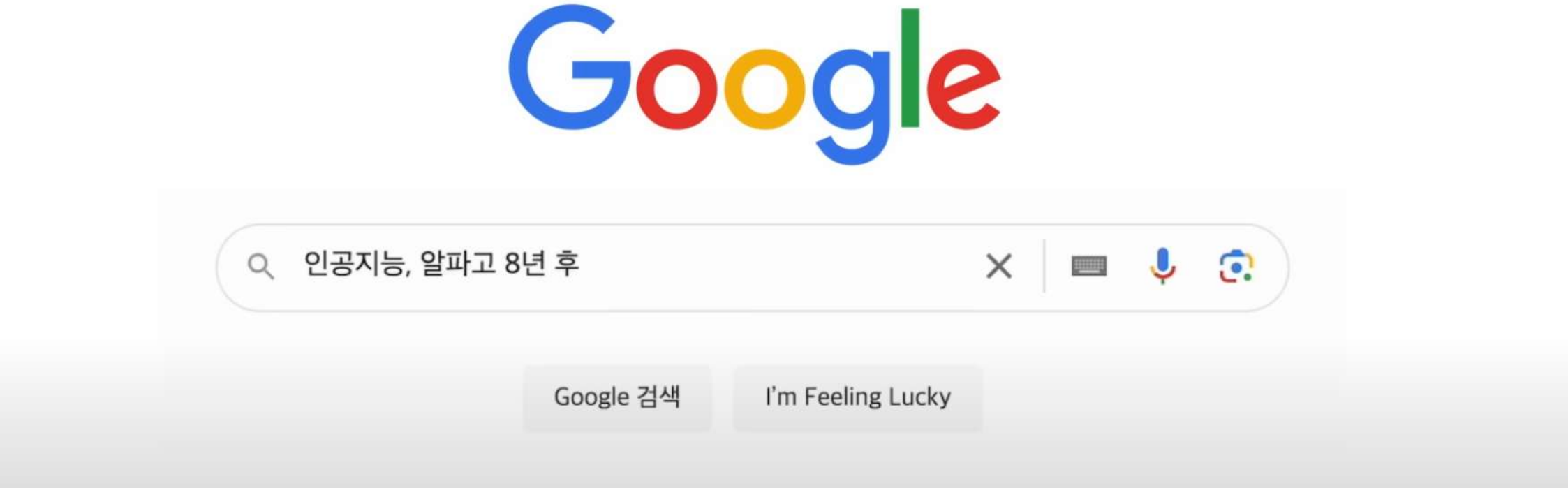
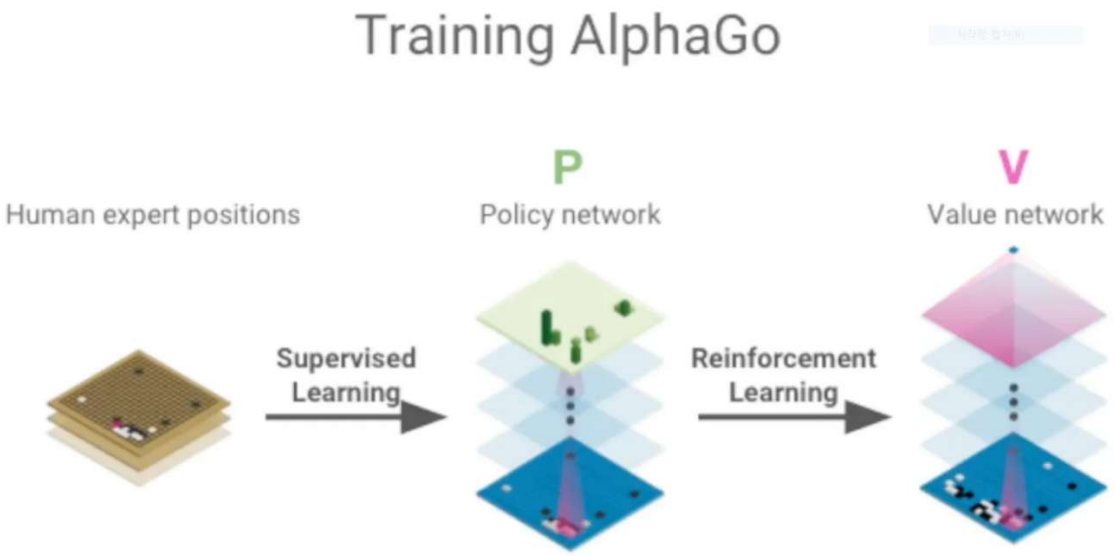
실물 환경에서의 데이터 기반 강화학습 적용 방안

("Data-Driven Reinforcement Learning for Optimal Control in Washing Machines" in IEEE CAI 2024
+
"실물 환경에서의 데이터 기반 강화학습 적용 방안" in 대한전자공학회 2024년 5월호 특집기)

강찬석
인공지능연구소 제어지능TP


Google
인공지능, 알파고 8년 후
Google 검색
I'm Feeling Lucky

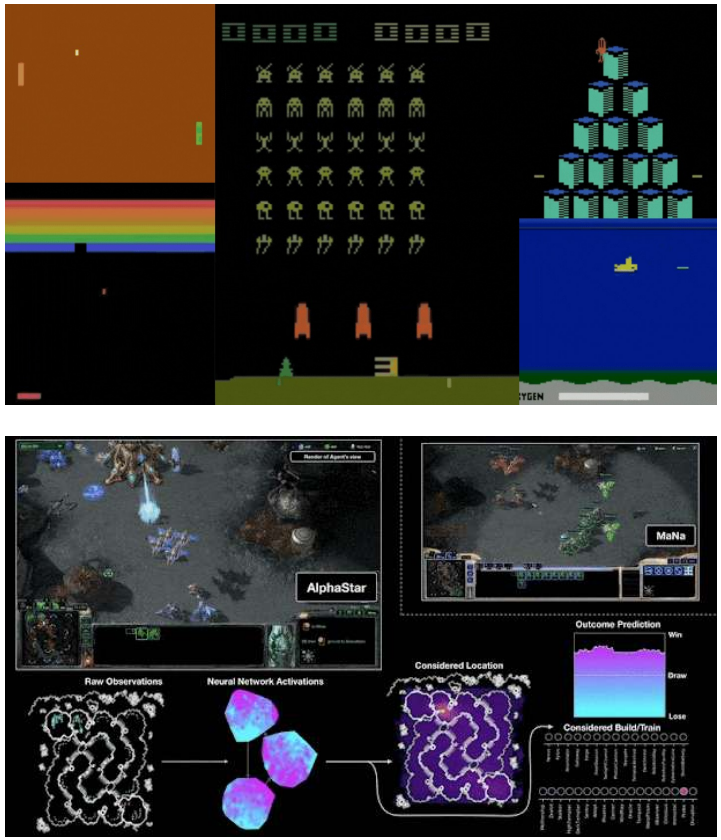
알파고에서의 강화학습



- 인간 전문가의 기보 데이터 기반의 특정 상황에서의 수 예측 (Supervised Learning)
- Self-play를 통한 다양한 상황에서의 정책 지속 개선 (Reinforcement Learning)

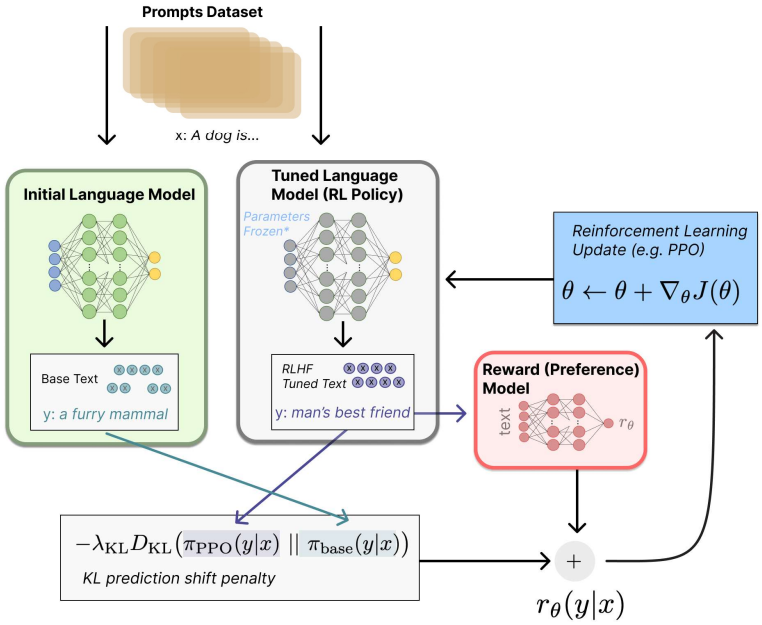
Credit: **DeepMind x UCL | Reinforcement Learning Course 2018**

실제 생활에 적용되고 있는 강화학습 사례



Game

Robot Control



RLHF

Credit: **ALE, AlphaStar, MT-Opt, HuggingFace**

목표

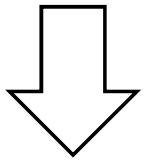
- 실환경 데이터 기반 모델 성능 향상 기술 개발

  – 데이터 기반 학습 알고리즘 개발

  – 모델 성능 개선 기간 단출

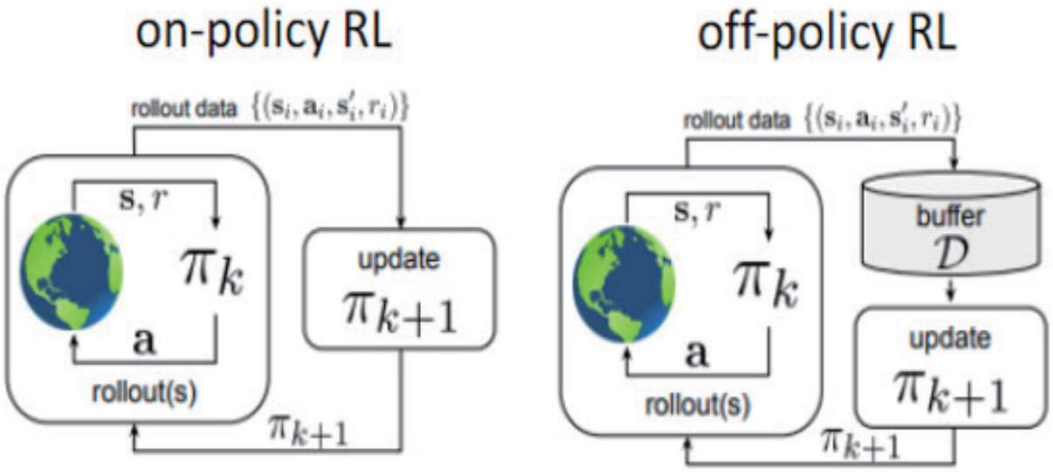
  } 데이터 기반 모델 성능 개선을 위한 강화학습 원천기술 확보

  – 데이터 기반 실환경 적응기술 개발 → 고객 환경 맞춤형 모델 업그레이드를 위한 선행 기술 탐색



Offline RL + Real World RL

오프라인 강화학습



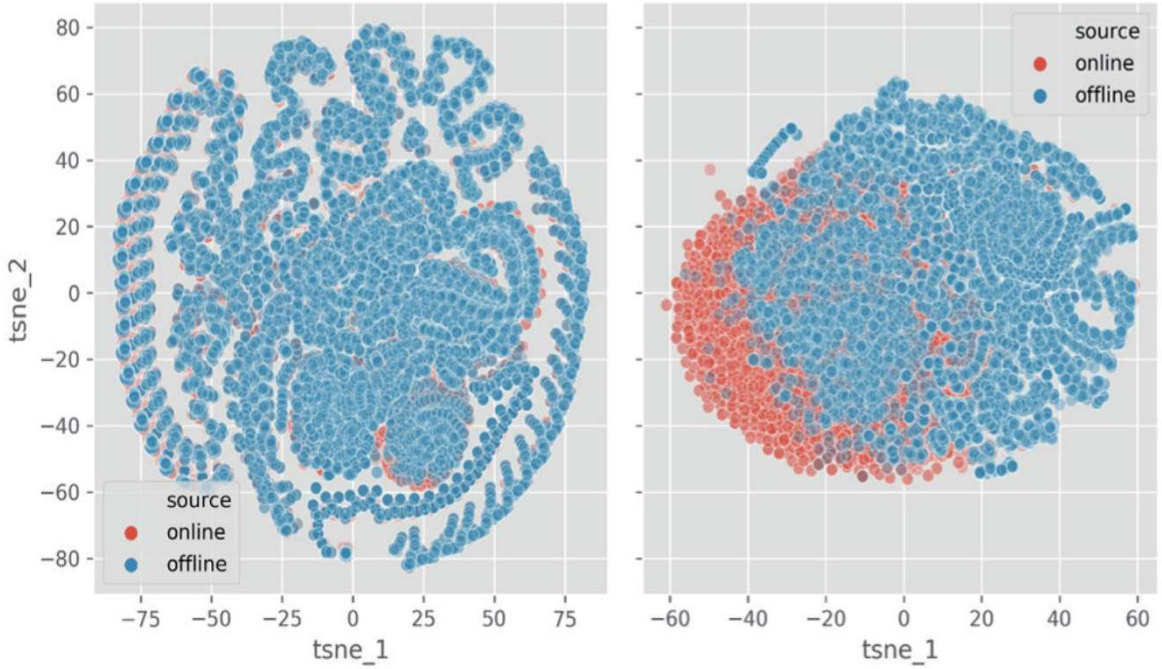
- 실험이나 전문가에 의해 수집된 다양하고 대규모의 데이터셋을 활용한 강화학습.

- 환경과의 interaction이 없는 상황에서도 모델을 학습시킬 수 있음.
  → 데이터 수집에 비용이 많이 들거나 위험한 환경에 유용

Credit: Levine, Sergey, et al. "Offline reinforcement learning: Tutorial, review, and perspectives on open problems." arXiv preprint arXiv:2005.01643 (2020).

오프라인 강화학습

- Data Distribution Shift
  - 데이터의 분포에 따른 편향성 발생



  - Counterfactual Query

실물 환경 강화학습의 어려운 점

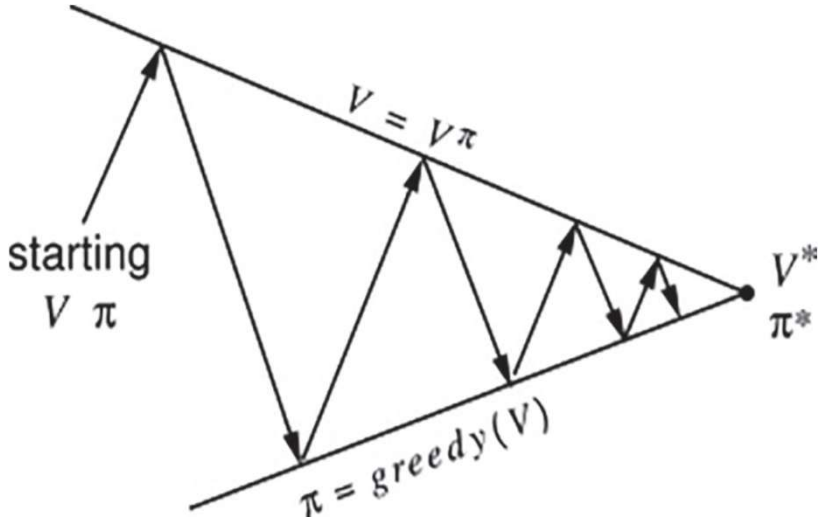
- 제한된 샘플에서의 학습
  - ALE나 MuJoCo 같은 환경을 모사할 수 있는 Simulator 부재
  - Dataset내에서의 제한된 Exploration & Low-variance Dataset
  - 실제 환경 상에서의 학습주기가 매우 김
    (ex. 추천 시스템이나 의료/의학의 경우 실제 효과가 나타나기까지 수개월 소요)

→ 샘플 효율성이 좋으면서, 실제 시스템에 적용되었을 때도 성능이 좋은 알고리즘 적용 필요

Credit: Dulac-Arnold, et al. "Challenges of real-world reinforcement learning." arXiv preprint arXiv:1904.12901 (2019).

실물 환경 강화학습의 어려운 점

- Offline/Off-policy 환경에서의 성능 평가
  - 강화학습의 근간, Policy Iteration (PI)
    : Policy Evaluation + Policy Improvement
  - 실제 환경에서는 이 과정도 실제 환경에서 수행되어야 함



⟶ Off-Policy Evaluation (OPE) 적용 필요

Behavior Policy (π_β) ≠ Target Policy (π)

Credit: Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). The MIT Press

실물 환경 강화학습의 어려운 점

- 불특정적이고 다중 목표를 띄는 보상 함수

  – Objective: To find optimal policy that maximize its total expected return

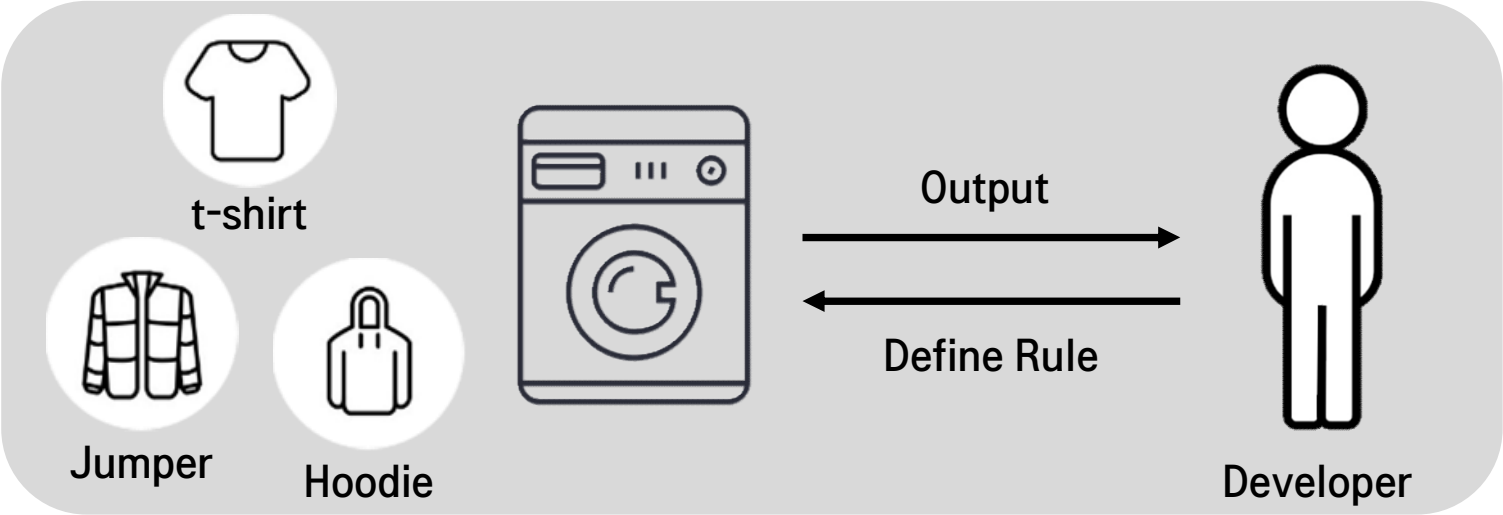
$$arg\max_{\pi} \sum_{t=0}^{T} E_{S_t \sim d^{\pi}(s), a_t \sim \pi(a|s)}[\gamma^t r(s_t, a_t)]$$

  – Unspecified : 점수를 높이는 게 목적인 게임과는 다르게 실물 환경의 보상함수는 정의가 어려움.
  (≈ MDP 설계의 어려움)

  – Multi-objective: 실물 환경에서는 보상의 목표가 여러개인 경우가 많음.

  → RL from Human Feedback (RLHF)

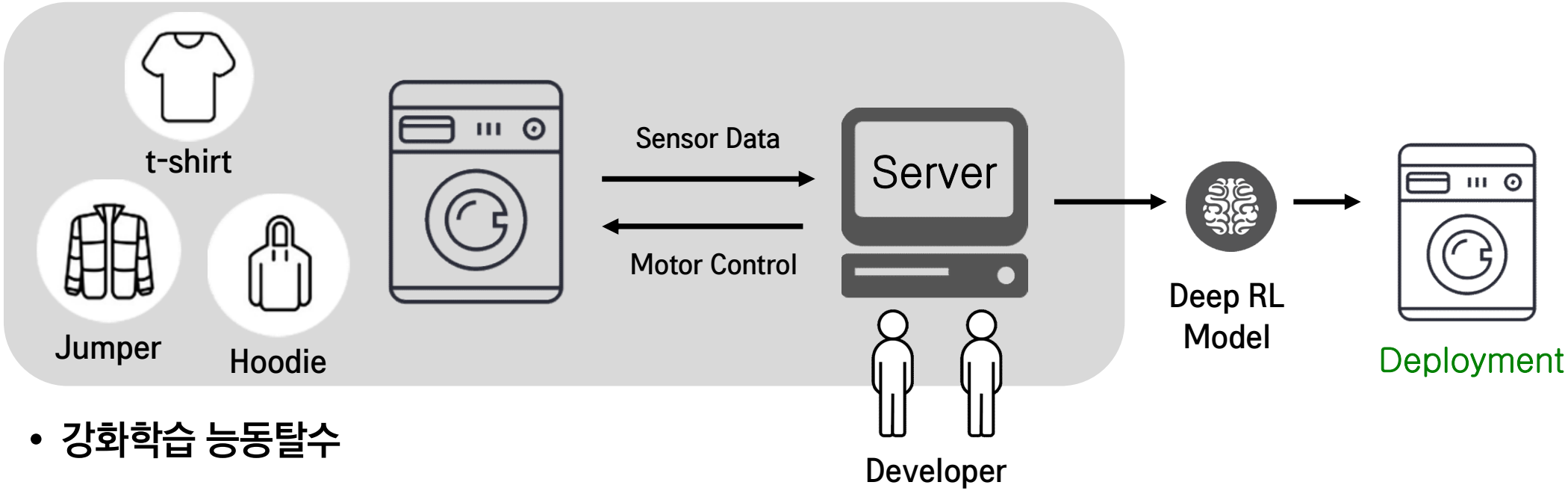
Data-Driven RL for Optimal Motor Control in Washing Machines



- 일반적인 세탁기 개발 프로세스 (Human-In-The-Loop)
  - 세탁기 개발자 ( )가 수많은 시행 착오를 반복하며, 실제 세탁기를 가지고 실험.
  - 품질 인정시험용 세탁포 ( ) 에 최적화된 제어 로직을 펌웨어 코드에 삽입.

잠바 후드티 티셔츠

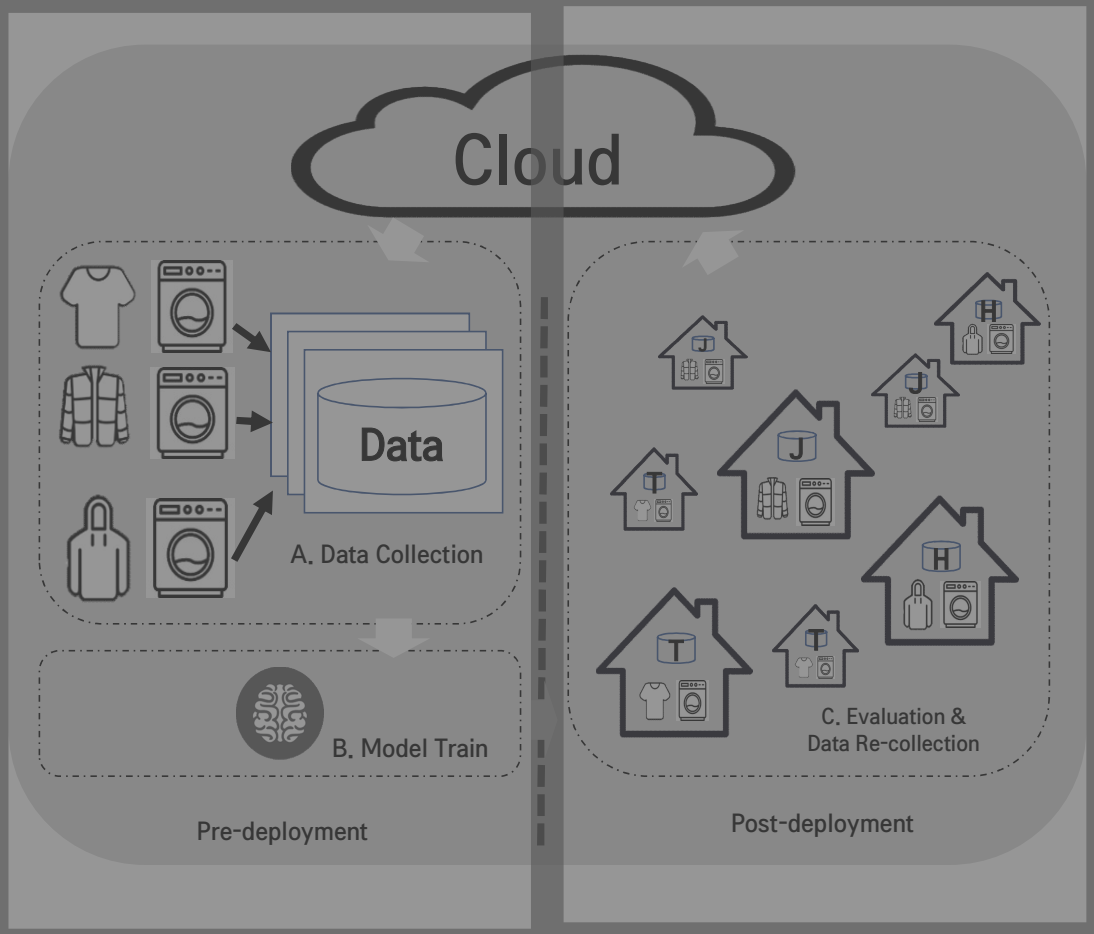
Data-Driven RL for Optimal Motor Control in Washing Machines



- 강화학습 능동탈수
  - AI 개발자 (CTO 인공(연)), 세탁기 개발자 (H&A 리빙(연))들이 강화학습 환경 설계 및 실험 진행
  - 학습된 모델을 실제 세탁기에 탑재할 수 있을 만큼 압축 (SoC센터 DQ-C)

Data-Driven RL for Optimal Motor Control in Washing Machines

- Delayed Online Update (DOU)



Data-Driven RL for Optimal Motor Control in Washing Machines

- Experiments & Results
  - 인정 시험에 사용되는 세탁 중 대표성을 띄는 6개의 세탁포 대상

Jean | T-Shirt | Jeans + Towel | Towels | JumperA Towels | JumperB Towels

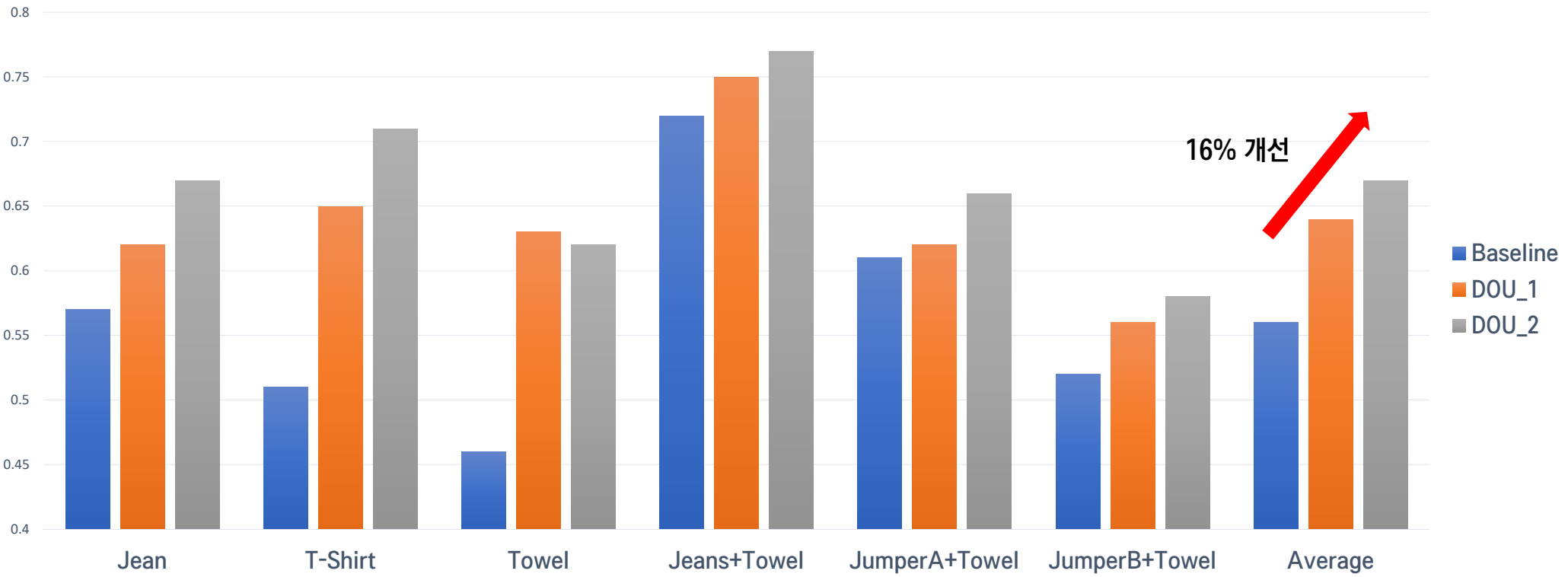
  - Offline RL의 효과 (Towel 1kg)

Motion Control in Heuristic

Motion Control with Offline RL

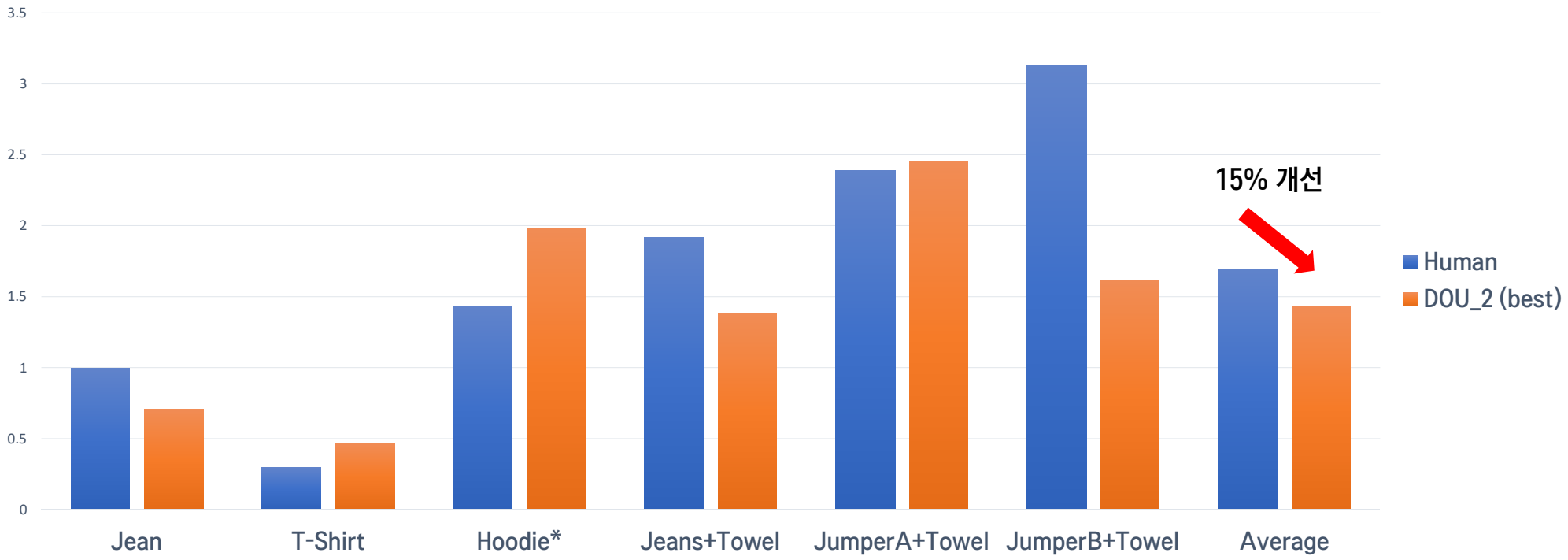
Data-Driven RL for Optimal Motor Control in Washing Machines

- Experiments & Results
  - DOU를 통한 성능 개선 효과 (탈수 성공율)



Data-Driven RL for Optimal Motor Control in Washing Machines

- Experiments & Results
  - DOU를 통한 성능 개선 효과 (양산 로직과의 내부 개발 지표 비교)



*Hoodie is included for the formal qualification control test

Data-Driven RL for Optimal Motor Control in Washing Machines

- 의미
  - LG SIGNATURE (NextCombo) 세탁건조기에 탑재되어 양산됨
  - Offline RL을 적용한 세계 최초의 생활가전
  - 실제 환경에서 수집된 데이터를 통해서 가전의 본질 성능을 개선할 수 있는 실용적인 접근 방법.

데이터 기반 강화학습 적용을 위한 다양한 시도

- 제어지능TP
  - 논문 기술의 실증(필드테스트) 예정
  - 강화학습 적용 모델 및 도메인 확대 전개
    (세탁, 탈수, 건조, 식기세척기 등)
  - 선행 강화학습 알고리즘의 실제 환경 적용 및 최적화
    - Offline-to-Online RL
    - Offline Meta RL
    - Real World RL 등...

결론

https://ai.googleblog.com/2020/04/an-optimistic-perspective-on-offline.html

약력

- LG전자 인공지능연구소 제어지능TP

- LG전자 AI Specialist (Control Intelligence)

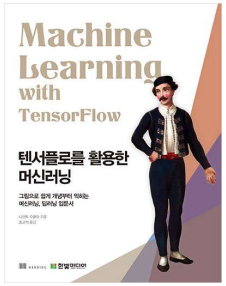
- 이력

– 18' 차세대표준(연) 5G Base Platform 개발 및 검증

– 20' 인공지능(연) 냉장고 부하대응 정온 정밀제어 PoC 개발 및 검증

– 23' 인공지능(연) 강화학습 기반 가전 능동제어 개발 및 검증

– 24' 인공지능(연) 상황인지 기반 최적 건조 모션 제어 기술 개발



- 감수/번역

- "텐서플로를 활용한 머신러닝" (한빛미디어, 2018년)
- "그로킹 심층 강화학습" (한빛미디어, 2021년)

Q&A

강찬석
(chanseok.kang@lge.com)