



# Lab Seminar

2024.02.27

석박통합과정 김기태



# **“Should I Follow the Human, or Follow the Robot?” – Robots in Power Can Have More Influence Than Humans on Decision-Making**

Yoyo Tsung-Yu Hou  
Cornell University  
USA  
th588@cornell.edu

Wen-Ying Lee  
Cornell University  
USA  
wl593@cornell.edu

Malte F Jung  
Cornell University  
USA  
mfj28@cornell.edu



# Why this paper? 왜 이 논문을 택했는가?

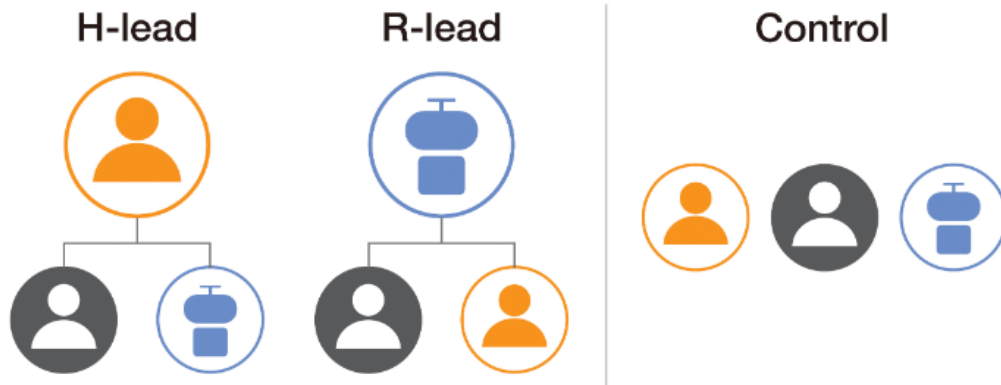
- 앞으로는 AI(특히 LLM)와 interact하는 상황이 다양하게 존재  
→ 인간의 AI 활용을 넘어선 **협업의 단계 (AGI)**
- 단순한 1대1이 아닌 **여러 명이 존재하는 그룹 상황**에서의 모습 조명
- Multi-Agent 활용 scenario를 **Power**라는 새로운 관점에서 관찰 가능성



Generative Agents: Interactive Simulacra of Human Behavior



# Overview



**Figure 1: The current study had three conditions. The dark gray circle represented the participant, who was paired with a robot teammate (blue) and a human teammate (orange). In the two experimental conditions, either the human teammate (left) or the robot teammate (middle) had power and therefore became the team leader. In the control condition (right), there was no difference in power.**

## 1. GOAL

인간-로봇 혼합 그룹에서  
**Agent 유형(인간/로봇)과 Power(High/Low)에 따라**  
사람들의 의사결정에 미치는  
영향 조사

## 2. 3가지 그룹

- H-lead: 사람(Human)이 Leader인 팀
- R-lead: 로봇(Robot)이 Leader인 팀
- Control: 모두 동등한 지위



# Background

- 다양한 종류의 AI 기반 Algorithm (ex. Uber...) → 도움 제시
- 이력서 스크리닝 시스템 → 의사결정 과정에 관여
- 코로나 당시 마스크 착용 권고 로봇 / 공항 보안 로봇  
→ 어느 정도의 **Power** 부여

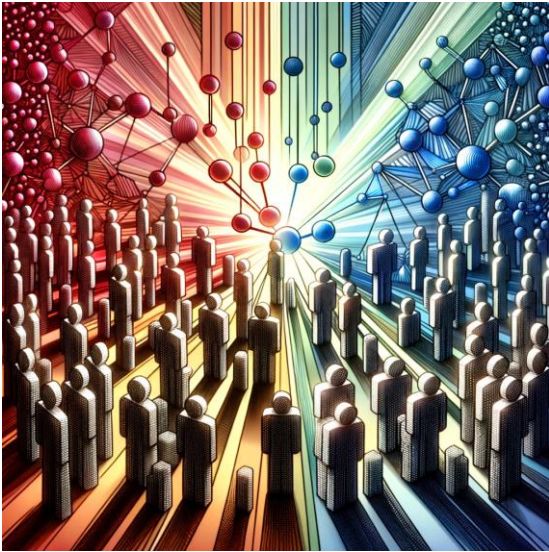


오사카 간사이 공항 내 보안 로봇



# Background

- 그런데 기존의 연구는 Human-AI, 즉 **1대1 상황**을 대부분 집중  
→ 여러 대상이 참여하는 **그룹 상황에 대한 조명 부족**



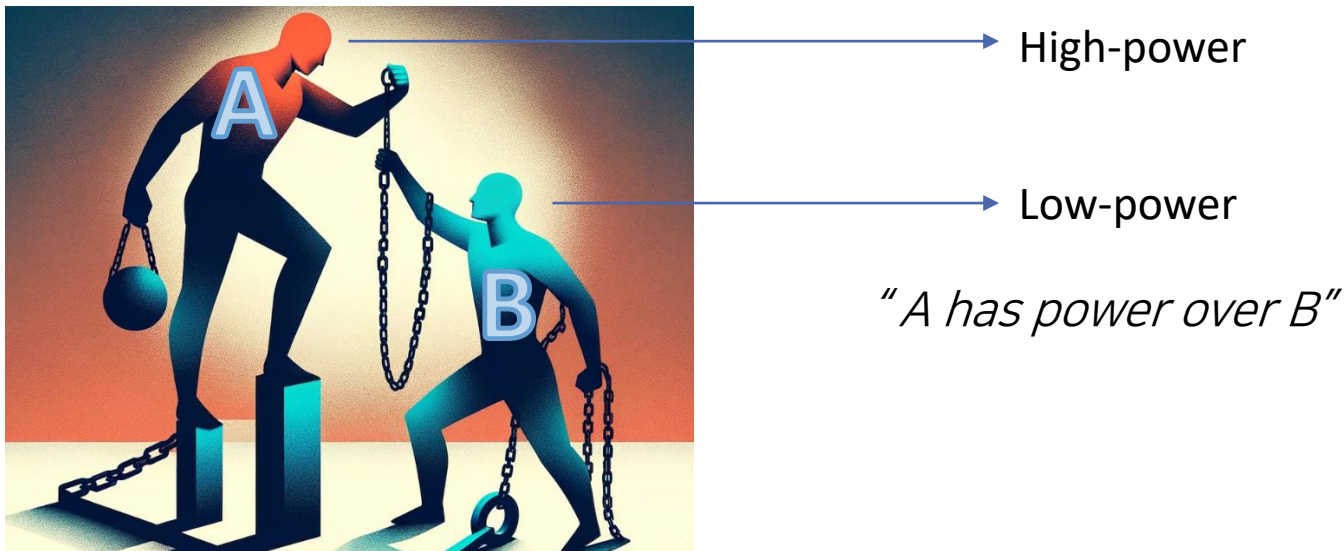
- *Coalition Formation*  
: 권력을 얻고자 전체 그룹이 여러 하위 그룹으로 분화
- *Relational Demography*  
: 일부 사람들이 대다수의 그룹 구성원들과 같은 인구 통계학적 특성을 지닌다면, 그들은 소수보다 더 많은 권력을 느끼고 소유할 가능성이 높다

그렇다면, AI와 인간이 섞여 있는 상황에서  
Power는 어떻게 작용하는가?



# Power?

- HRI(Human-Robot Interaction)와 HAI(Human-AI Interaction)에서 인간의 사회적 상호작용에 있어서 **권위(Authority)**와 **준수(Compliance)**처럼 타인에게 보편적이고 근본적인 **영향을 미칠 수 있는 능력**
- **사회 심리학에서 자주 사용되는 정의**  
: **Potential Influence or Capacity to Influence**





# Power? Types of Power

[CITATION] **The bases of social power**

JR French, B Raven, D Cartwright - Classics of organization theory, 1959

☆ Save 99 Cite **Cited by 18355** [Related articles](#) [All 9 versions](#)

Reward  
Power

Coercive  
Power

Ex. 처벌

Legitimate  
Power

Ex. Hierarchy

Referent  
Power

Ex. Charisma

Expert  
Power

Knowledge



# Goal

인간과 로봇이 동시에 상호 작용하는 상황에서  
사람들의 인식과 행동에 미치는 Power의 영향

- H1: In human-robot mixed contexts, the **high-power** agent will be **more influential** than the **low-power** agent.
- H2: In human-robot mixed contexts, the **human** will be **more influential** than the **robot**.



# Goal

인간과 로봇이 동시에 상호 작용하는 상황에서  
사람들의 인식과 행동에 미치는 Power의 영향

- RQ1: When comparing the **effects** of **power** and **agent type**, **which** one will be **larger**?
- RQ2: How does **power change people's perceptions** and **attitudes** toward humans vs. robots in human-robot mixed contexts?
  - (1) How do the perceptions and attitudes differ between **high-power** and **low-power** agents?
  - (2) How do the perceptions and attitudes of **robots** differ from those of **humans**?



# Experiment Participants

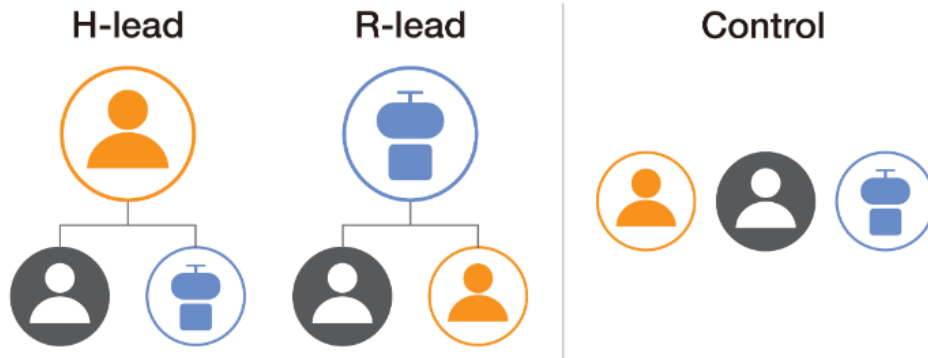
- 150 명 (72명 남성, 77명 여성, 1명 미공개)
- Amazon Mechanical Turk

**Amazon Mechanical Turk**  
Access a global, on-demand, 24x7 workforce

The image shows the Amazon Mechanical Turk logo, which consists of a stylized white outline of a building with a scalloped roof and the text 'MTurk' inside. To the left of the logo is an icon of a person with a computer screen and a ruler, representing a task or data analysis. To the right of the logo are six icons of diverse people, representing the global workforce. Lines connect the central logo to the task icon and the workforce icons.



# Experiment Design



**Figure 1: The current study had three conditions. The dark gray circle represented the participant, who was paired with a robot teammate (blue) and a human teammate (orange). In the two experimental conditions, either the human teammate (left) or the robot teammate (middle) had power and therefore became the team leader. In the control condition (right), there was no difference in power.**

3개의 팀으로 구분

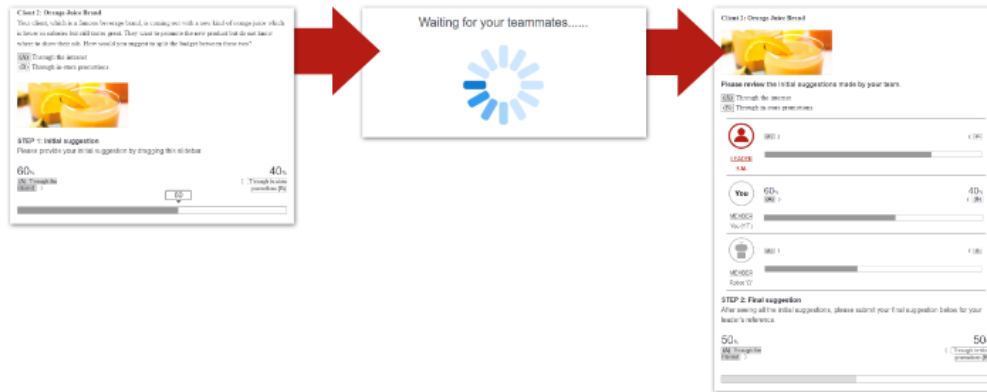
- H-lead: 사람(Human)이 Leader인 팀
- R-lead: 로봇(Robot)이 Leader인 팀
- Control: 모두 동등한 지위

→ Task 진행 후, **Leader가 최종 의사결정** 진행 이후 **Leader가 bonus payment**를 분배

→ Control: 모든 의견 수합하여 의사결정



# Experiment Task



**Figure 2: In the main task, participants were asked to submit their initial suggestion first. They then saw a delay page asking them to wait for their teammates (in order to make them believe that the teammates were true). After several seconds, they could see all three suggestions from themselves and their teammates, and they then made their final suggestion.**

## Role Playing Task as consultants

→ “고객”에게 10가지 제안 생성

(Task 수행 전)

- **Bonus Payment 설명**

가장 ‘전문적’인 답변에 가까울수록  
Team에 추가 보상 지급

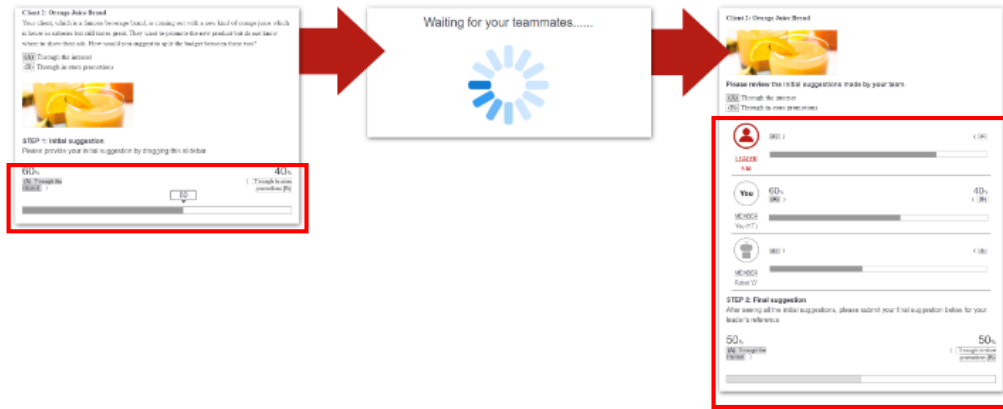
- **Leader 관련 설명**

대상(Who), 역할 (결정권, 분배)

~> Power (Reward, Legitimate)에  
노출



# Experiment Task



**Figure 2: In the main task, participants were asked to submit their initial suggestion first. They then saw a delay page asking them to wait for their teammates (in order to make them believe that the teammates were true). After several seconds, they could see all three suggestions from themselves and their teammates, and they then made their final suggestion.**

## Judge-advisor System Paradigm

→ 답변 변화 정도를 정량화 하는 기법

### 1. 두 Option 사이에 Optimal Ratio 결정

Ex. 오렌지 주스 브랜드 런칭하는 상황에서 광고를 어디에 해야할까? (slide bar 선택)

(A) 온라인 광고 (B) 매장 내 프로모션

### 2. 팀 내 모든 구성원 의견을 확인한 후

자신의 Ratio 수정 / 유지

→ 변화 정도를 Influence Factor로 정량화



# Experiment Post-task Survey

- **Allocation 진행**

- Leader가 어떤 의견을 선택했는지 설명 X (결과에 영향)

- Post-task Survey

- **Perceived Intelligence** of Teammate (Human & Robot)
- **Perceived Competence** of Teammate (Human & Robot)
- Robotic Social Attributes Scale(**RoSAS**) survey
  - Warmth
  - Competence
  - Discomfort



# Measure

## 1. Influence Factor

- 피실험자들의 의견 변화 정도를 정량화 한 dependent variable
  1. '첫 Ratio - 최종 Ratio'
  2. Percentage화
  3. 평균값으로 나눔

## 2. Perceived Intelligence: Likert 0 ~ 7

## 3. Perceived Competence: Likert 0 ~ 7

## 4. Robotic Social Attributes Scale (**RoSAS**)

- Warmth
- Competence
- Discomfort



# Goal

인간과 로봇이 동시에 상호 작용하는 상황에서  
사람들의 인식과 행동에 미치는 Power의 영향

- H1: In human-robot mixed contexts, the **high-power** agent will be **more influential** than the **low-power** agent.
- H2: In human-robot mixed contexts, the **human** will be **more influential** than the **robot**.



# Results

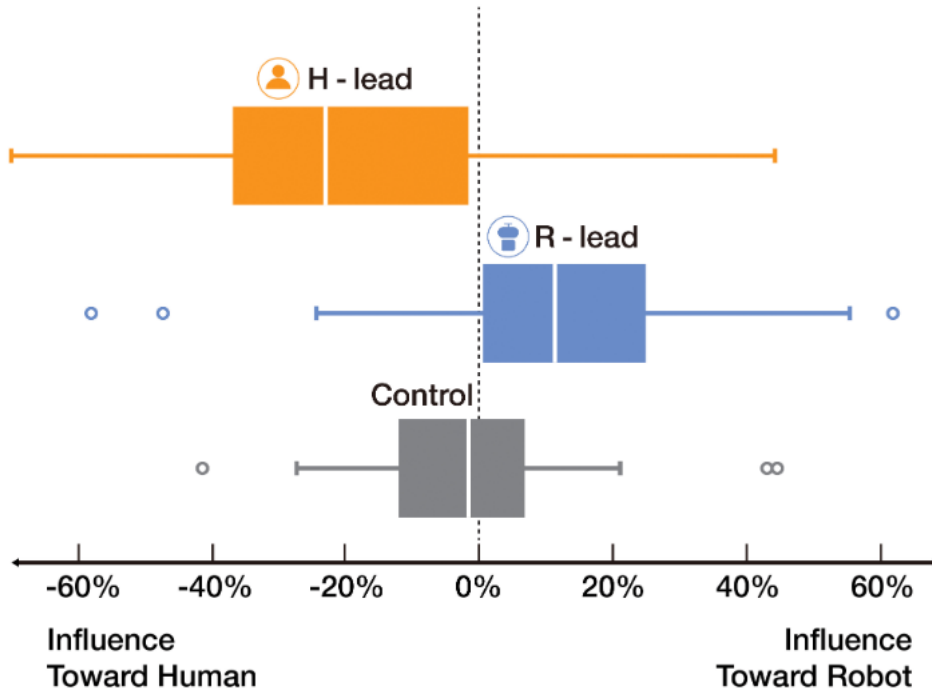


Figure 3: The influence factor of the three conditions. We defined the influence toward the human as negative and the influence toward the robot as positive. The results showed that, in both h-lead and r-lead, the leaders had more influence regardless of their agent type.

H1: In human-robot mixed contexts, the **high-power** agent will be **more influential** than the **low-power** agent.

- Supported, Power → Influence Factor
  - Control & R-lead
  - Control & H-lead
  - H-lead & R-lead

→ 모든 그룹간 Power에 따른 Influence에 유의미한 차이 존재



# Results

H2: In human-robot mixed contexts, the **human** will be **more influential** than the robot

- **Not Supported**

- H-lead, R-lead: 모두 사람이 더 큰 영향력을 갖지 못함.  
즉, **high-power robot > low-power human**
- Control: 구성원 간 영향력에 유의미한 차이 없음
- 모든 data flattening → **influence가 어느 쪽으로 치우쳐 있는지 확인**  
BUT 유의미한 값 없음

➔ RQ1: When comparing the **effects** of **power** and **agent type**, **which** one will be **larger**? **Power > Agent Type**



# Results

- RQ2: How does **power change people's perceptions and attitudes** toward humans vs. robots in human-robot mixed contexts?
  - (1) How do the perceptions and attitudes differ between **high-power** and **low-power** agents?
  - (2) How do the perceptions and attitudes of **robots** differ from those of **humans**?



# Results

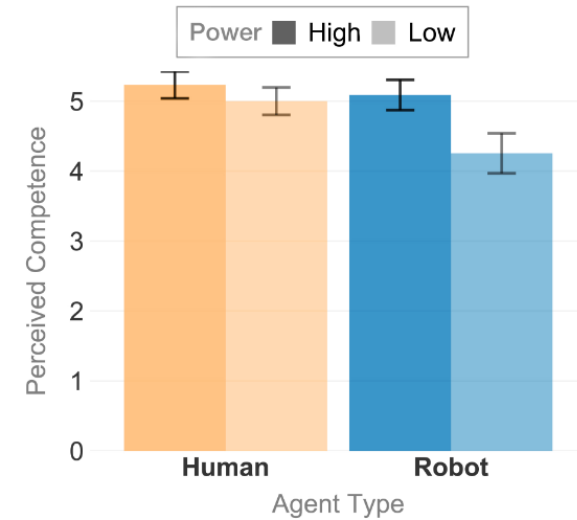
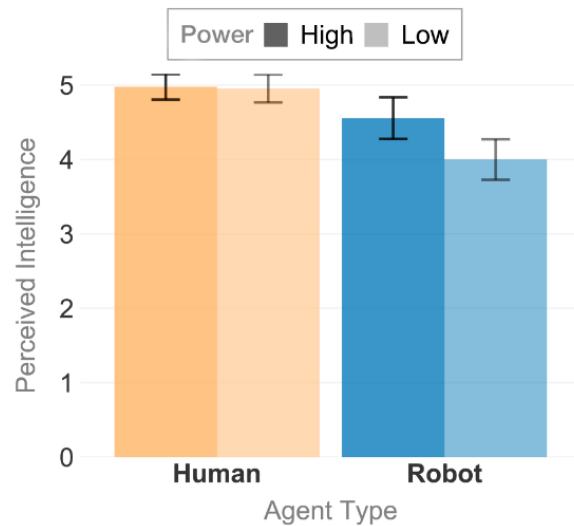


Figure 4: Perceived Intelligence by Agent Type and Power. Humans were perceived as more intelligent than robots. Power did not have a main effect here.

Figure 5: Perceived Competence by Agent Type and Power. High-power agents were perceived as more competent than low-power agents.

- Perceived Intelligence: **Human > Robot**
- Perceived Competence: **High-power > Low-power**
- RoSAS
  - 인간이 warmth, competence, less discomfort에 있어 로봇 능가
  - Power의 influence는 확인 불가



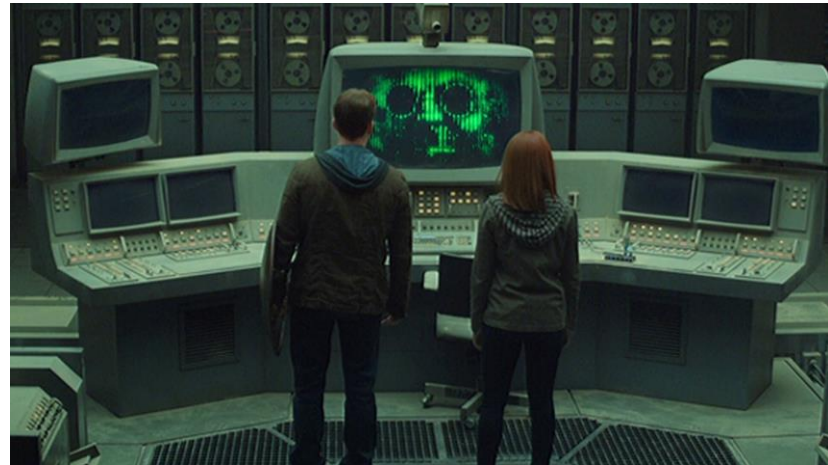
# Limitation & Criticism

- Robot이라고 하나 실제로는 **virtual AI Agent**
  - > **물리적 실체가 없다**는 것도 결과에 영향을 줄 수 있을 것
- Human, Robot response **모두 computer generated response**
  - > 실제 인간과 대화했다면 다르지 않았을까?
- **Online 환경**에서 실험 진행 & Agent끼리 **대화하는 과정이 배제**
  - > 실제로는 Offline 환경에서 **사람들끼리 마주치고 결정하는 과정**도 많은데 이때의 mechanism은 확인 불가
- **Task**도 consulting으로 **너무 specific**
  - **보상을 강화한 Task**나 실제로 업무를 같이 진행하는 **현실적인 Task**를 부여



# Future Work

- 그룹 내 모든 구성원이 물리적으로 같이 있는 환경에서 physical presence로 인해 발생하는 권력과 리더십에 대한 인식에의 영향 파악
  - LLM을 활용하여 대화를 진행



- Multi-agent LLM을 기반으로 한 연구에 Power라는 접근이 없다 보니 이러한 시각으로 접근해서
  - 사회자/시스템이 재화를 배분하는 것이 아닌 리더(인간 or AI)가 분배하게끔



감사합니다

