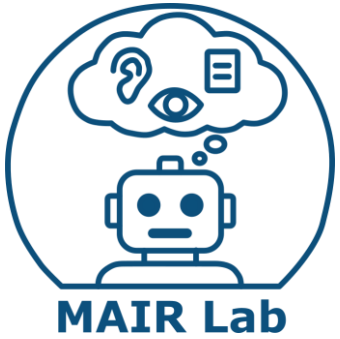
Introduction

안인규 (Inkyu An)

Speech And Audio Recognition
(오디오 음성인식)

https://mairlab-km.github.io/



수업계획

- 수업계획

Week 1: Introduction
Week 2: Digital Signal Processing
Week 3: Speech Recognition 1
Week 4: Speech Recognition 2
Week 5: Self-Supervised Models for Audio
Week 6: Source Separation 1
Week 7: Source Separation 2
Week 8: Midterm exam
Week 9: Audio-Visual Deep Learning
Week 10: Robot Audition
Week 11: Multi-channel Audio Processing
Week 12: Sound Source Localization
Week 13: Multi-channel Speech Separation & ASR
Week 14: Diffusion-based TTS
Week 15: Final Project

수업계획

- 진행
  - PPT 강의교재 (https://mairlab-km.github.io/courses/speech-audio-recognition-2025fall)
  - 이론수업, 실습수업 (Google Colab) 병행
- 평가
  - 중간고사 (40%)
  - 프로젝트 (40%) – 중간고사 이후 공지
  - 과제 (10%)
  - 출석 (5%)
  - 보너스 점수 (5%) – 수업참여(질문, 대답)+0.5, 학우를 도와주면+0.5

수업계획

- Prerequisites
  - Python
- 참고자료
  - https://github.com/yandexdataschool/speech_course?tab=readme-ov-file
  - https://github.com/markovka17/dla
- 성취기반평가
  - 정확한 기준은 중간고사 이후에 결정 및 공지
  - 상대평가가 아닌 절대평가

My Past Research into Three Areas

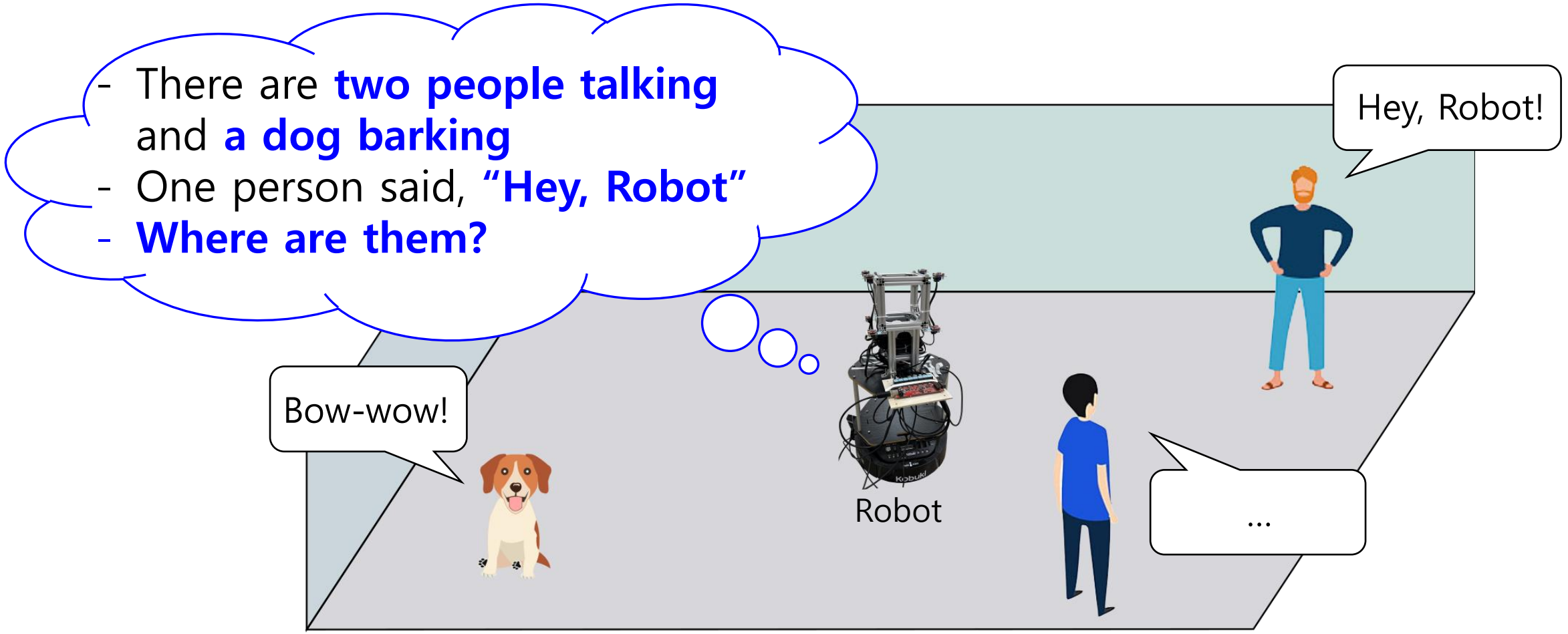
- Research on **robot audition technologies** (overcoming real-world challenges) [ICRA 18, 19, 20; T-RO 22; RA-L 24]
  - Sound source localization
  - Speech separation for overlapping speech signals and noise
- **Multimodal sensing integration** [NeurIPS 23; T-RO 19; CAVW 23]
  - Robot mapping
  - Image retrieval
  - Acoustic material estimation
- **Real-time robotic applications**
  - Research on path planning for drones

My Past Research into Three Areas

- Research on **robot audition technologies** (overcoming real-world challenges) [ICRA 18, 19, 20; T-RO 22; RA-L 24]
  - Sound source localization
  - Speech separation for overlapping speech signals and noise
- **Multimodal sensing integration** [NeurIPS 23; T-RO 19; CAVW 23]
  - Robot mapping
  - Image retrieval
  - Acoustic material estimation
- **Real-time robotic applications**
  - Research on path planning for drones

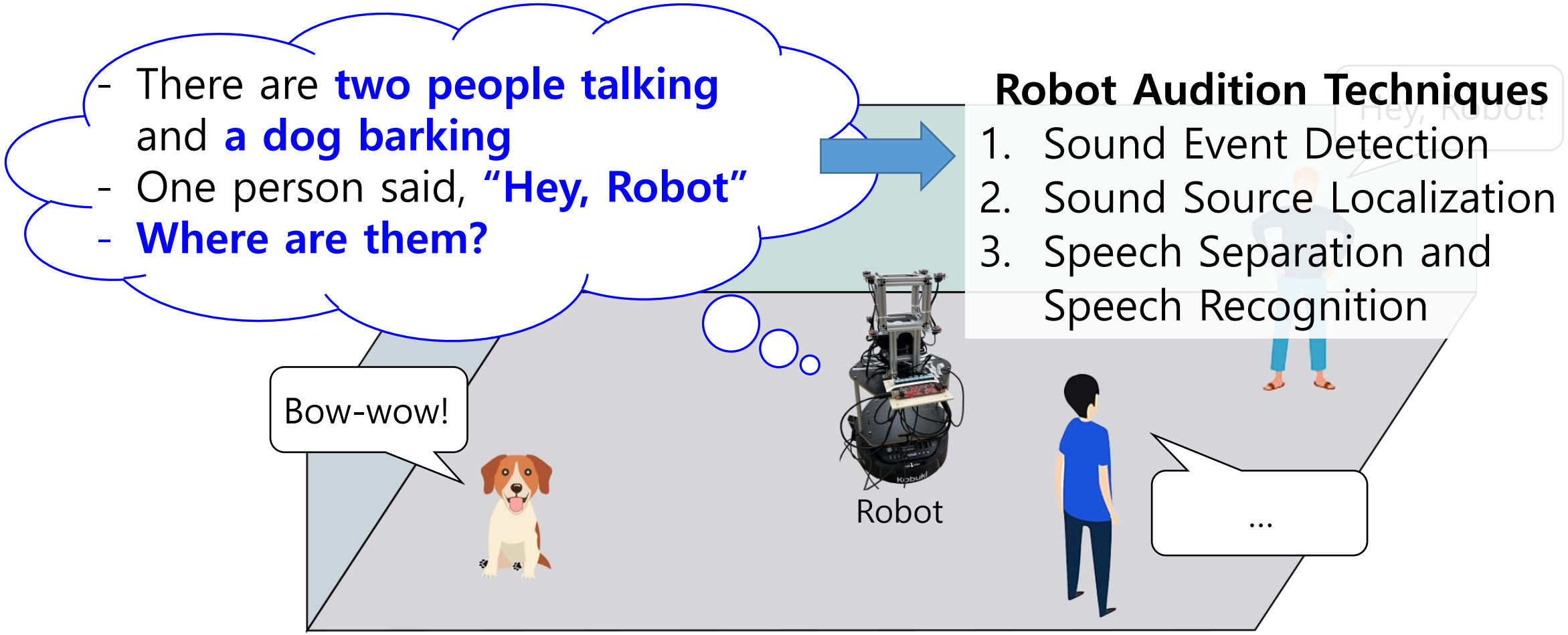
Robot Audition?

- **The ability of a robot to perceive and process auditory information from its environment**

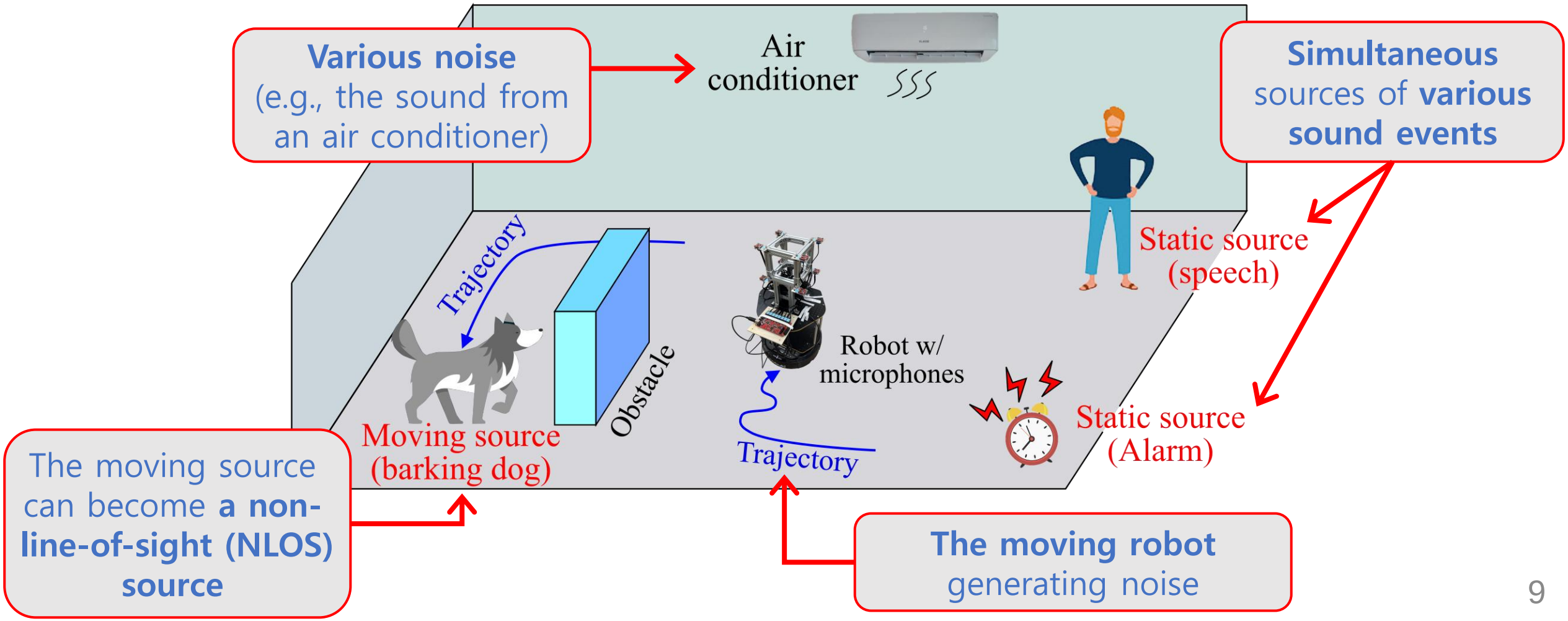


Robot Audition?

- **The ability of a robot to perceive and process auditory information from its environment**

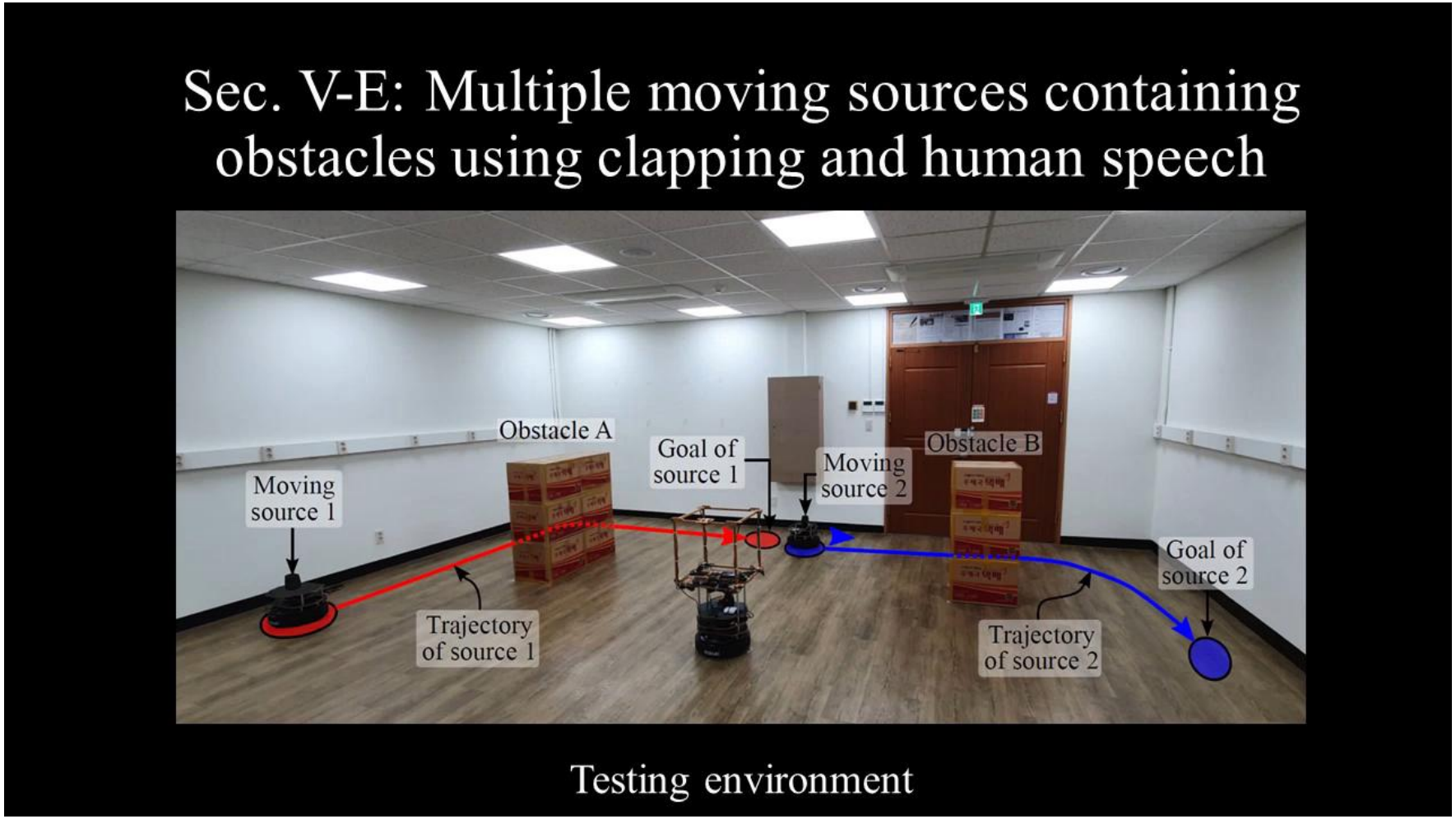


Challenge of Robot Audition in Real-world environments



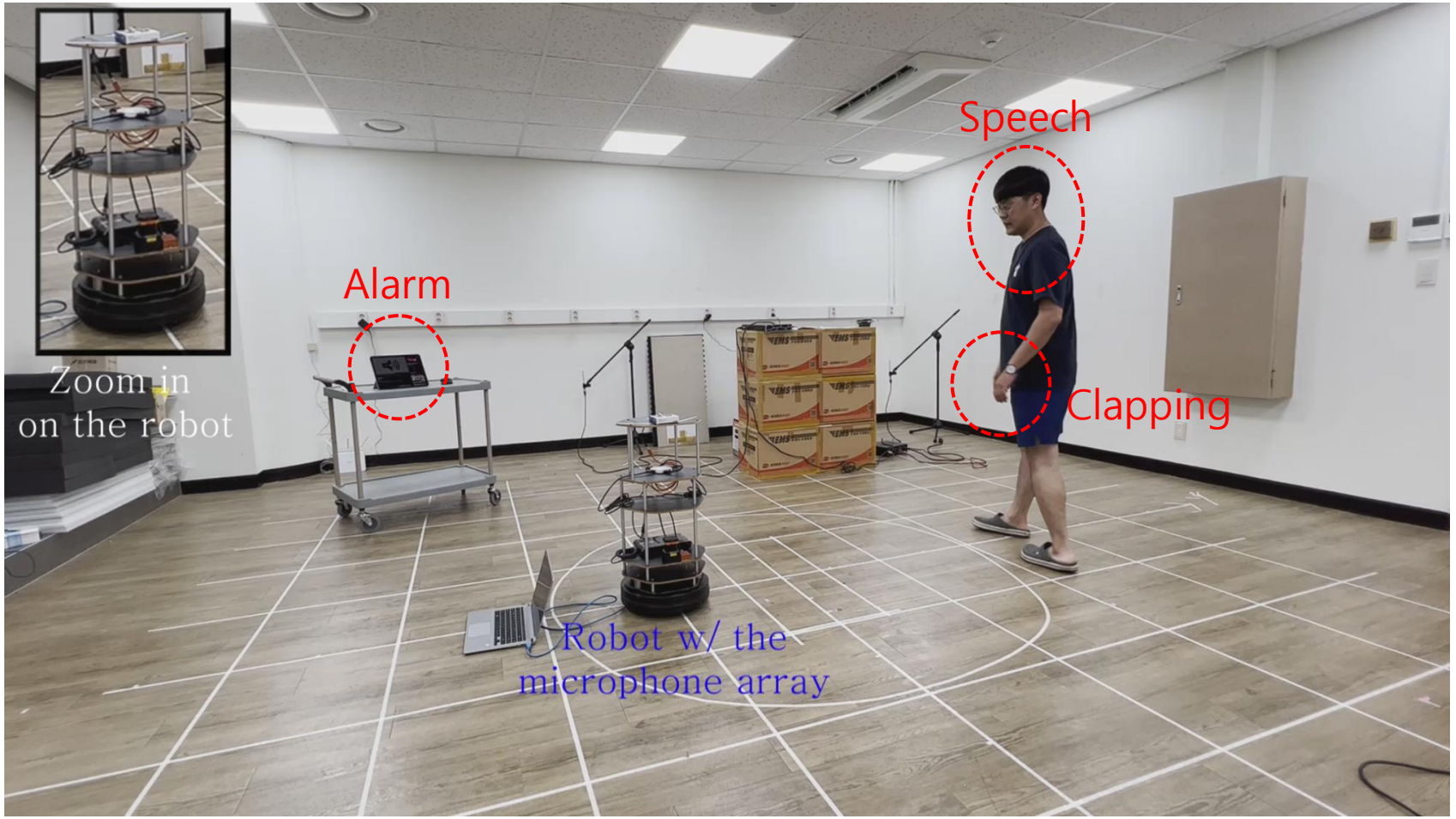
Challenges: NLOS sources

- **The demonstration video of the acoustic ray tracing-based SSL**



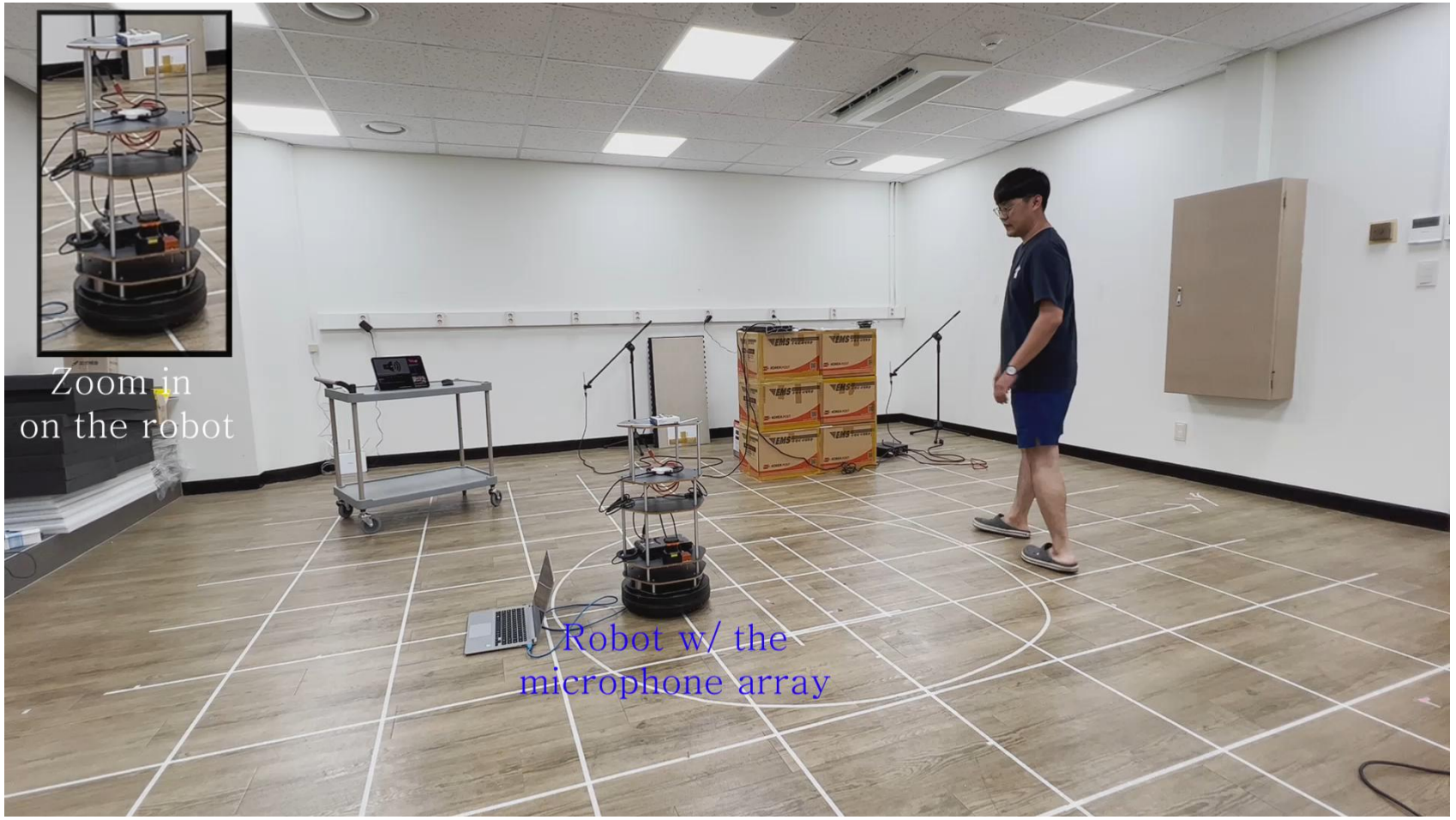
Challenges: Simultaneous sources of different sound events

- **The demonstration video of the robust-TDoA model with microphone pair training**



Challenges: Simultaneous sources of different sound events

- **The demonstration video of the robust-TDoA model with microphone pair training**



Navigation

Sec. V-G: Navigating to the NLOS source (clapping sound)

Testing environment

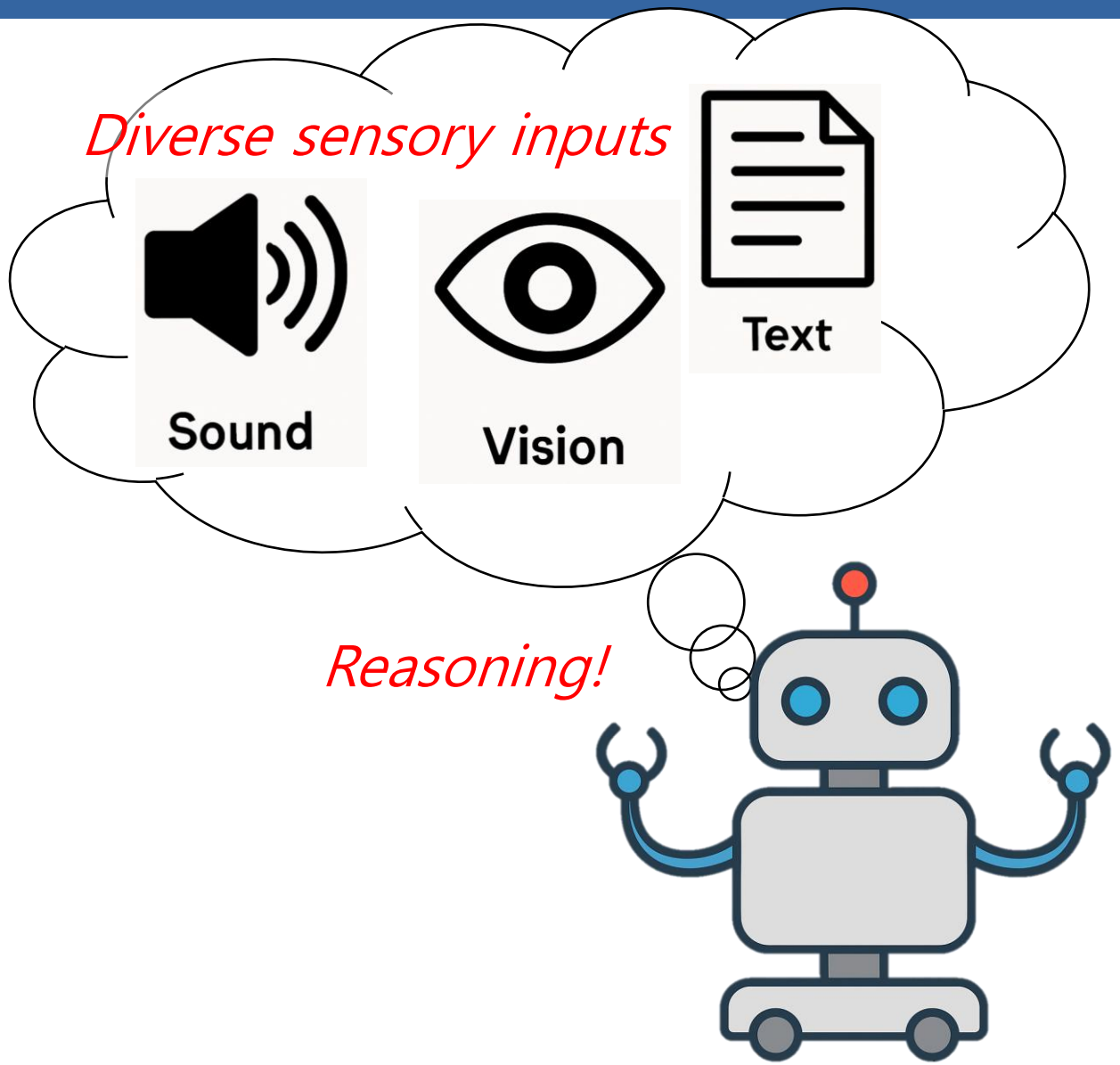
Research Topics Prior to Joining Kookmin Univ.

- Research on robot audition technologies (overcoming real-world challenges) [ICRA 18, 19, 20; T-RO 22; RA-L 24]
  - Sound source localization
  - Speech separation for overlapping speech signals and noise
- Multimodal sensing integration [NeurIPS 23; T-RO 19; CAVW 23]
  - Robot mapping
  - Image retrieval
  - Acoustic material estimation
- Real-time robotic applications
  - Research on path planning for drones

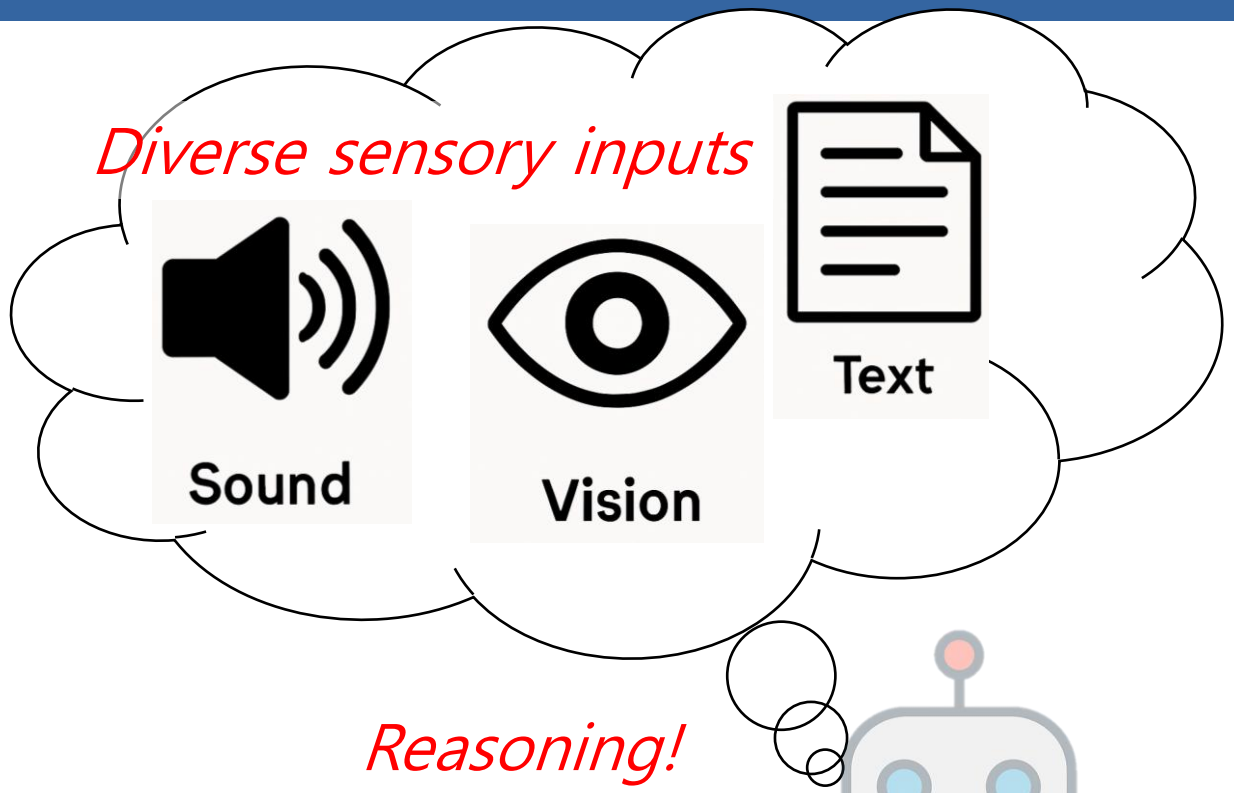
Research on Robot Audition

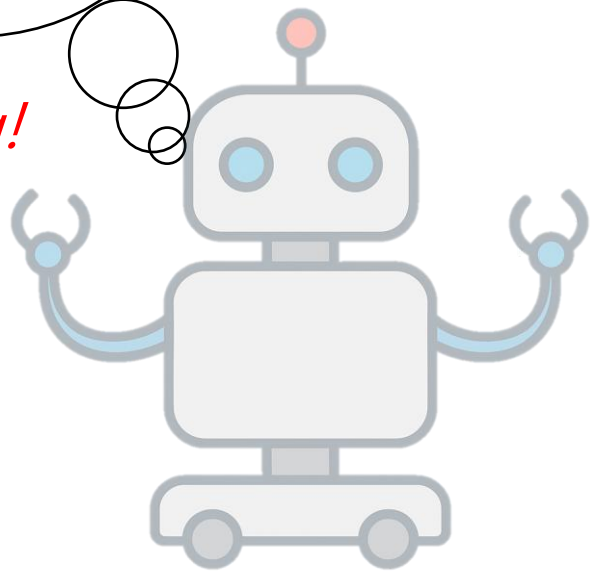
<u>Expanding it into Multimodal Embodied AI for Robots</u>

Research Topics

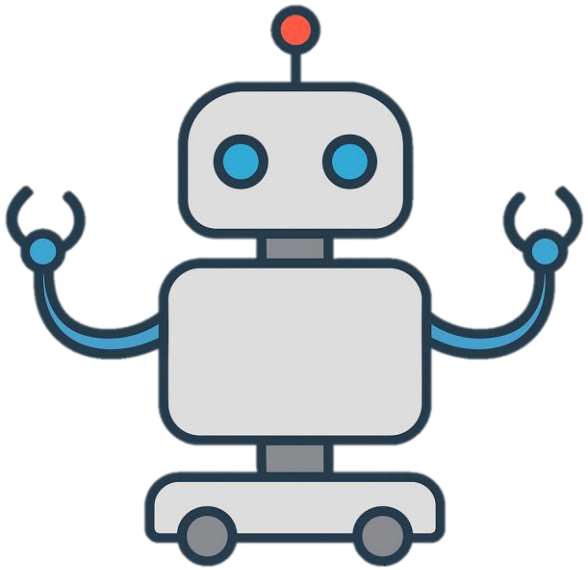


Research Topics





Take an action!



Research Topics

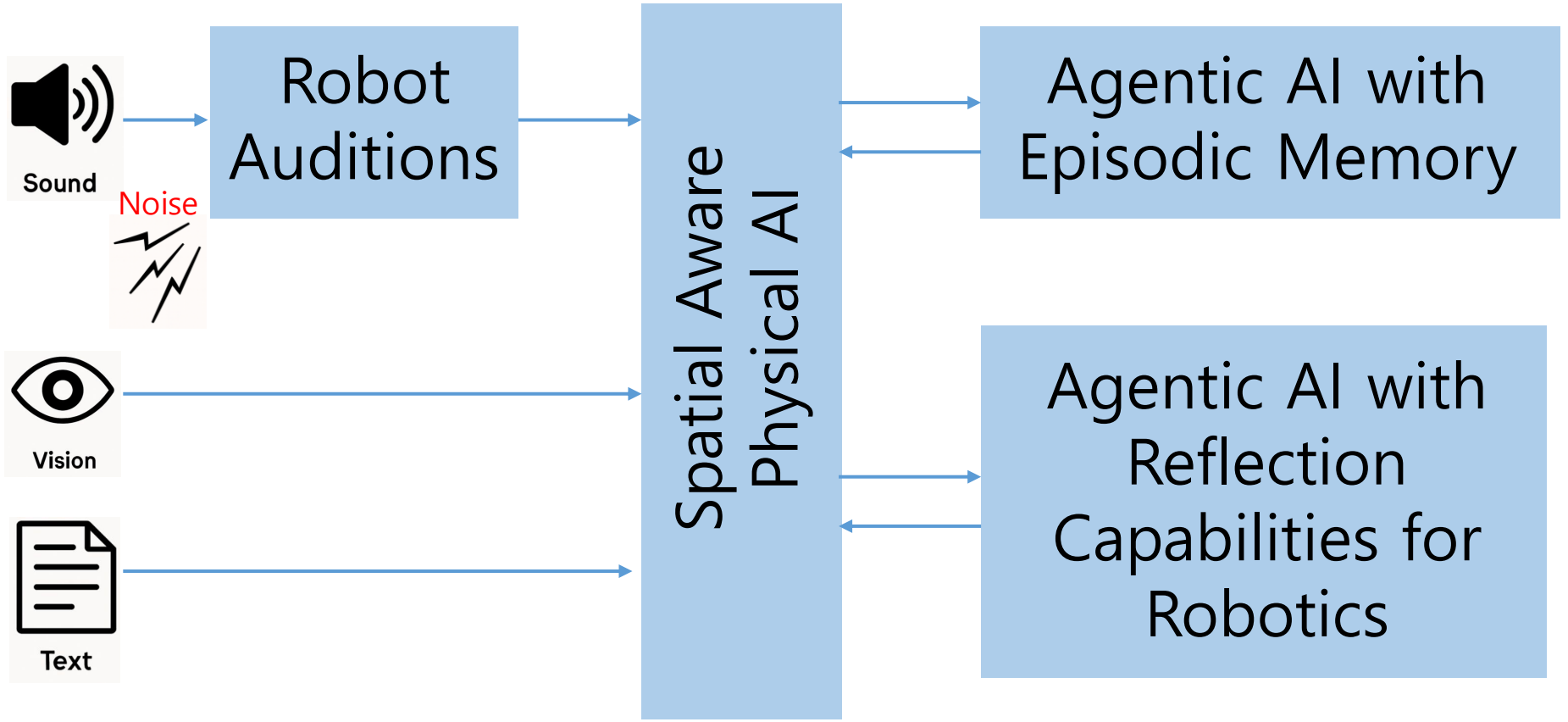
Multimodal Embodied AI

Multimodal Agentic AI

Robot Auditions in Real Environments

Research Topics

Ongoing Core Technologies

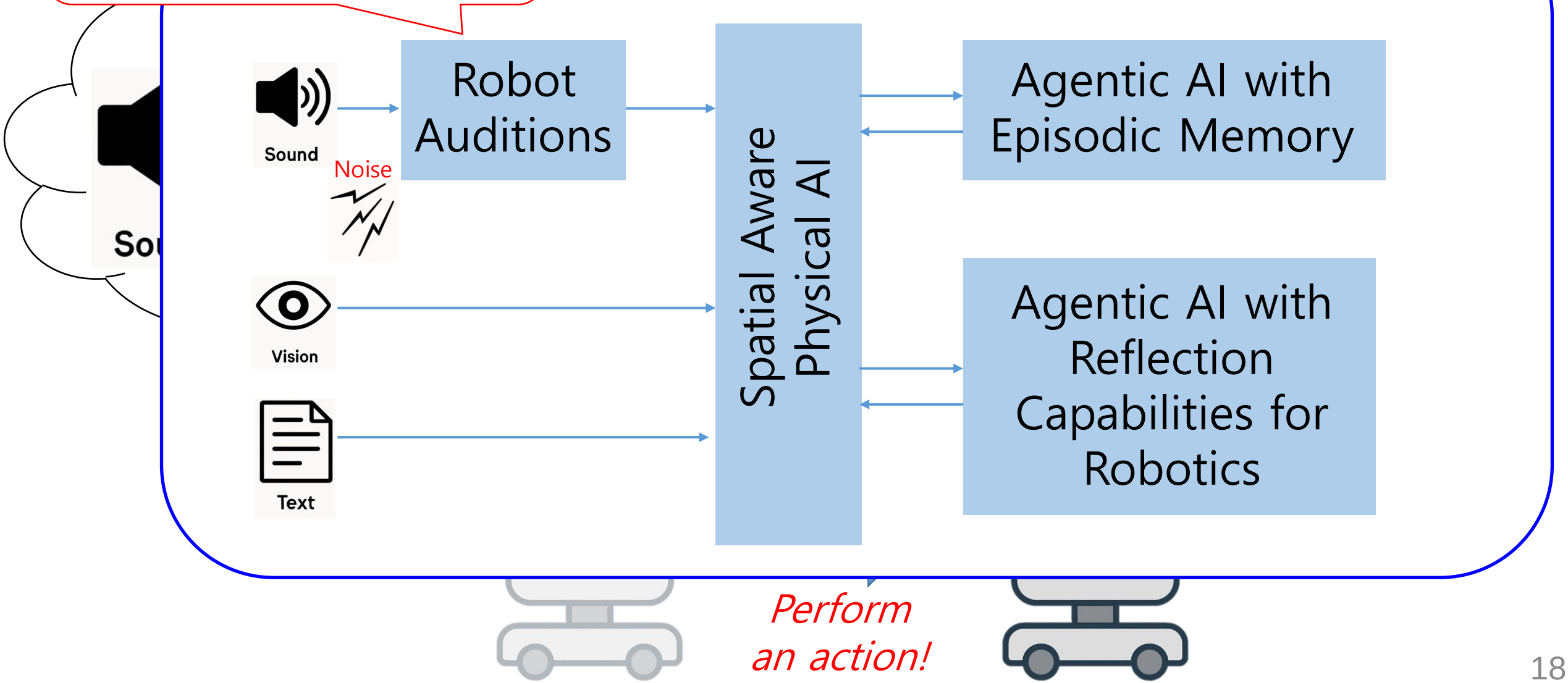


Perform an action!

Research Topics

Operate reliably in real-world environments

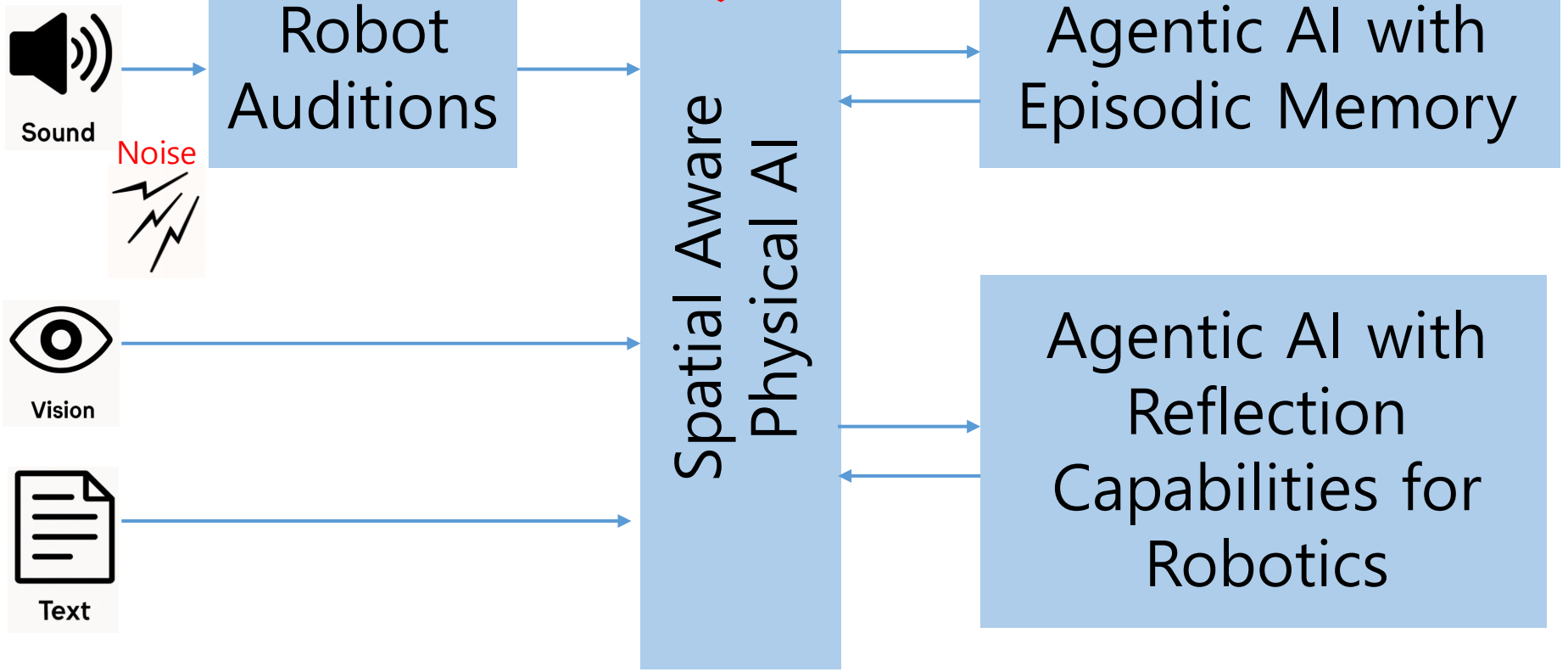
Ongoing Core Technologies



Res

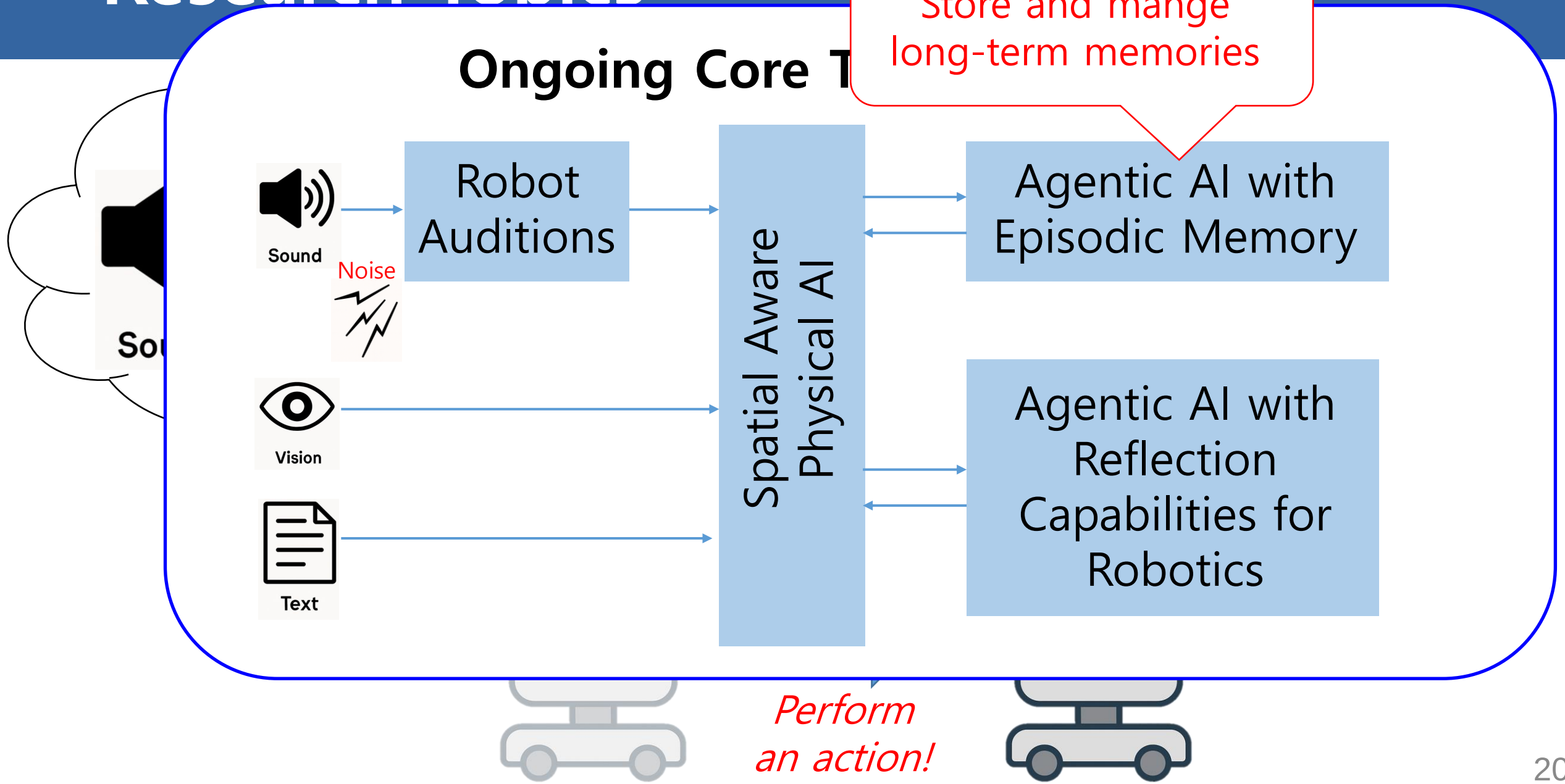
logies

Perceive spatial information from multimodal data and perform reasoning



Perform an action!

Research Topics



Research Topics

Ongoing Core Technologies

Robot Auditions

Spatial Aware Physical AI

Agentic AI with Episodic Memory

Agentic AI with Reflection Capabilities for Robotics

Allow robots to improve themselves over time

Path planning (Drone)

Don’t be afraid of failure!