

Language and Robotics: Toward Building Robots Coexisting with Human Society Using Language Interface

Introduction

Yutaka Nakamura, Shuhei Kurita, Koichiro Yoshino

Guardian Robot Project, RIKEN

Center for Advanced Intelligence Project, RIKEN



知識獲得・対話研究チーム
Knowledge Acquisition &
Dialogue Research Team



ガーディアンロボット
プロジェクト
Guardian Robot Project



革新知能統合研究センター
Center for Advanced
Intelligence Project



Discussion on Dory

- ◆ We will use “Dory” for discussion
- ◆ You can put any comments at any timing of the tutorial
- ◆ You can vote on other people’s questions
- ◆ We will have a discussion at the last of the tutorial based on your comments
- ◆ <https://dory.app/events/ORCeryF2DQT5E45K8oS9/ijcnlp-aacl-tutorial-langrobo/>
- ◆ Any tutorial materials are on:
<https://github.com/riken-grp/langrobo-tutorial>

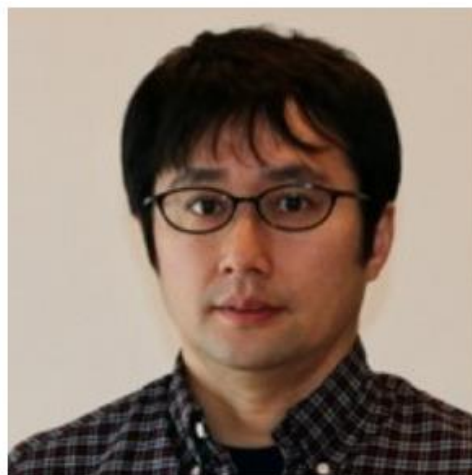


Tutorial speakers



Shuhe Kurita

Language Information
Access Technology
Team,
RIKEN Center for
Advanced Intelligent
Project (AIP)



Yutaka Nakamura

Behavior Learning
Research Team,
RIKEN R-IH Guardian
Robot Project (GRP)



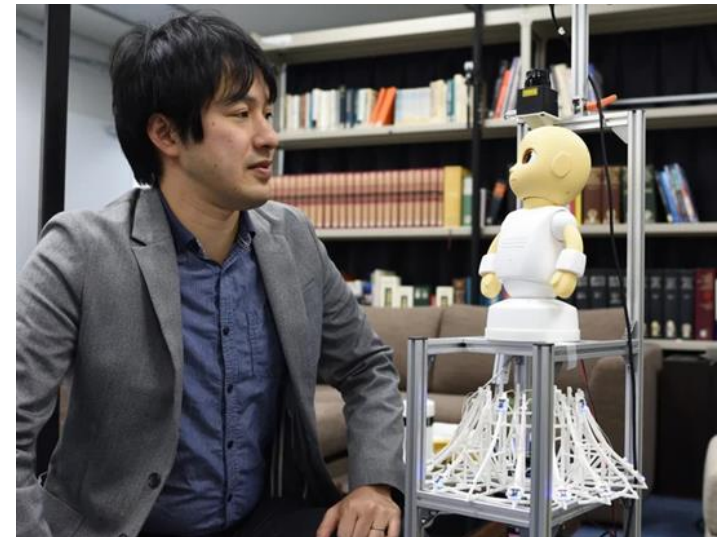
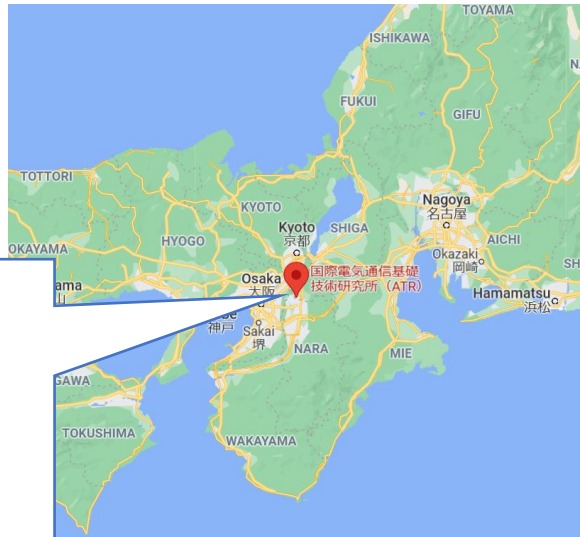
Koichiro Yoshino

Knowledge Acquisition
& Dialogue Research
Team,
RIKEN R-IH Guardian
Robot Project (GRP)

Self introduction

◆ Koichiro Yoshino

- Team Leader (PI) of knowledge acquisition and dialogue research team, Guardian robot project, RIKEN, Kyoto, Japan
- Dialogue system technology challenge committee, action editor of ACL rolling review, IEEE-SLTC member, SIGdial board



Robots in our living space



Astro, Amazon



HSR, Toyota



Pepper, Softbank



Stretch, hello-robot

We expect them to help us,
using conversational interface
for enriching our life

Language in intelligence

- ◆ It is related to “**human mind**” and “**human awareness**”
 - What is mind? What is awareness?
 - How do we define “**intelligence**”
- ◆ Language is an important tool for “**communication**” and “**knowledge-building**”
 - Abilities to use language are related to the work of intelligence
 - *nani gigantum umeris insidentes*
 - We can receive, accumulate, and rebuild the knowledge using language communication

Language and robotics

◆ Current robots do not have abilities to:

- accumulate and use one's experience
- use the experience and knowledge of others

◆ How humans do that?

- by using language!



I dropped a glass cup and broke it.

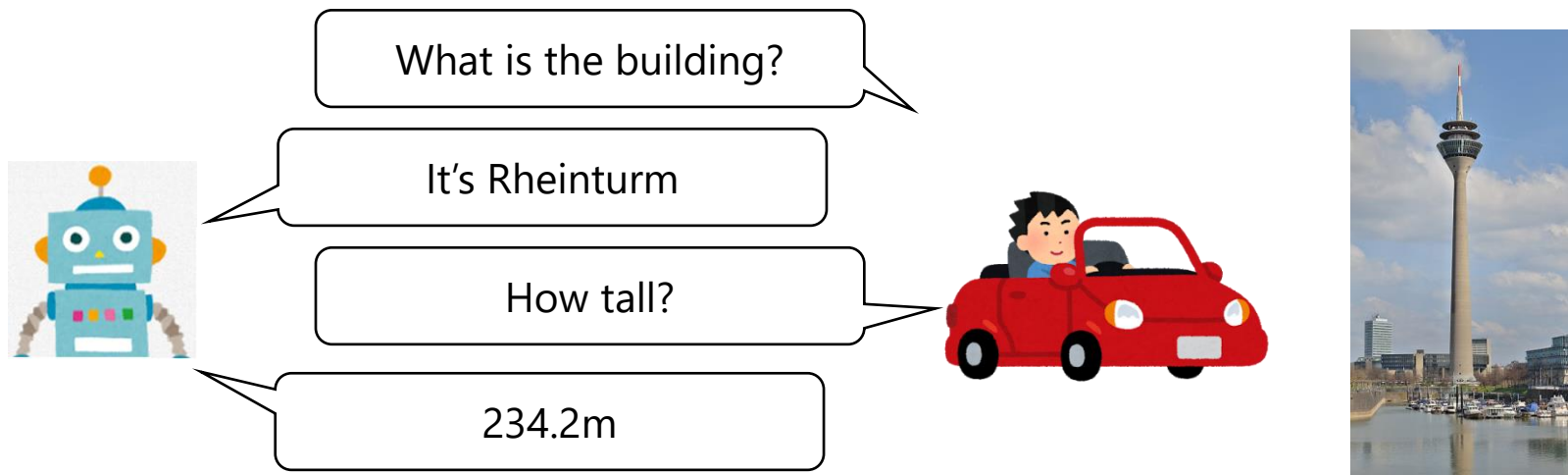
Oh, I would be careful when holding a glass cup.



*Nur ein Idiot glaubt, aus den eigenen Erfahrungen zu lernen.
Ich ziehe es vor, aus den Erfahrungen anderer zu lernen, um von vorneherein eigene Fehler zu vermeiden.*

Using natural language in real world

- ◆ Systems should consider the dialogue contexts
- ◆ Dialogue contexts contains not only dialogue history but also shared information
 - Dialogue context: shared information in real-world
 - Map, visibility of the user (dialogue in car, [Misu14])
 - Visual question answering, visual dialogue [Alamri 19]



Using real-world knowledge

◆ Using both visual information and dialogue history (DSTC7 visual dialogue track)

- Extension of visual question answering (VQA)
- The system changes the answer according to the dialogue context

Example Visual dialogue

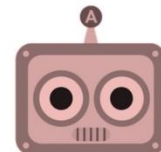


C: A dog with goggles is in a motorcycle side car.
 Q: Is motorcycle moving or still?
 A: It's parked
 Q: What kind of dog is it?
 A: Looks like beautiful pit bull mix
 Q: What color is it?

Image

Dialog history

Question



Visual Dialog model

Talking about bull-mix
 -> answer the color of bull-mix

Answer

A: Light tan with white patch that runs up to bottom of his chin

Understanding a situation



Another one is missing

User has a glass
There is another glass on the table

May I bring another glass?

× understand the language

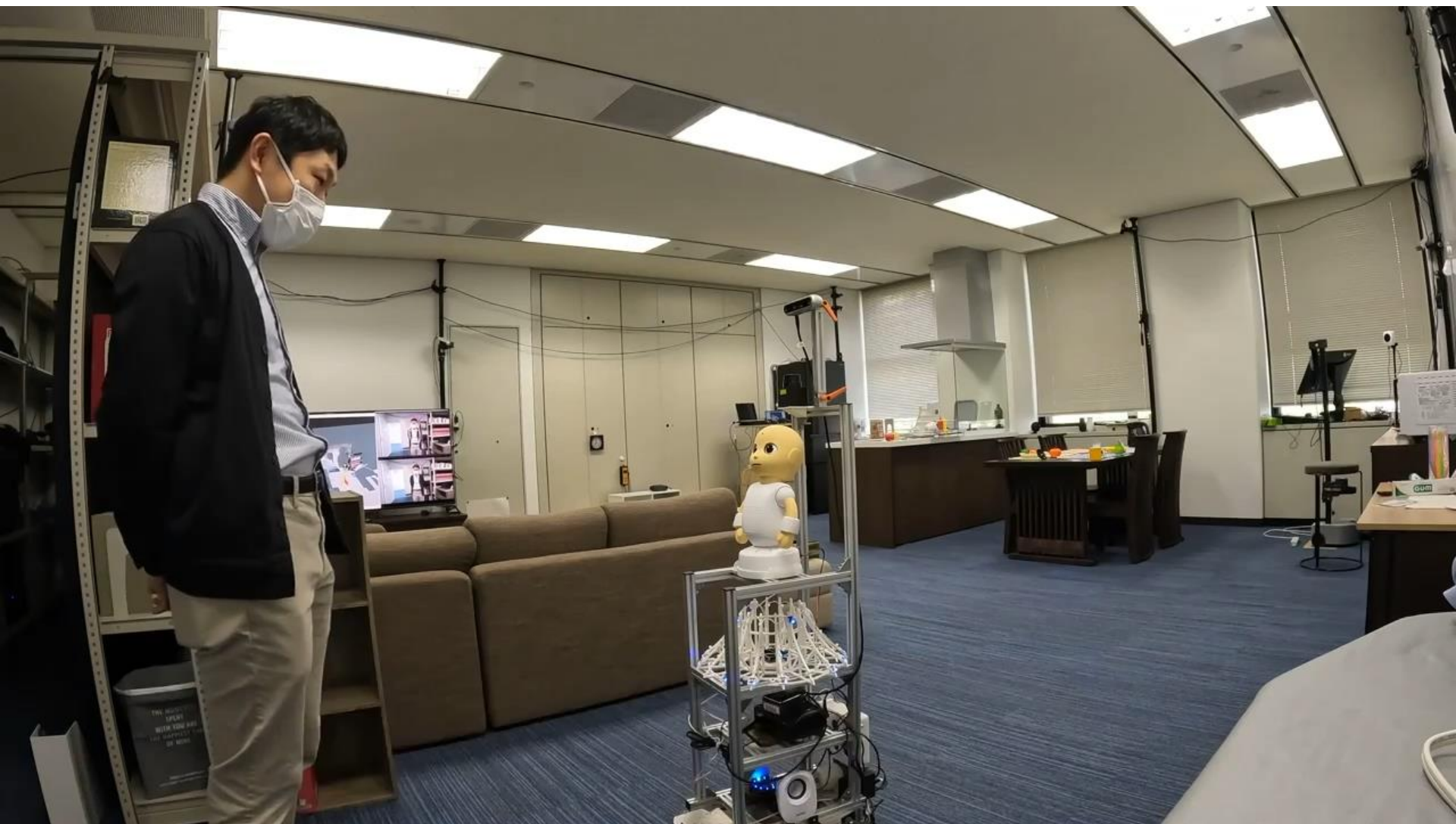
○ understand the situation and the intent

Situation recognition

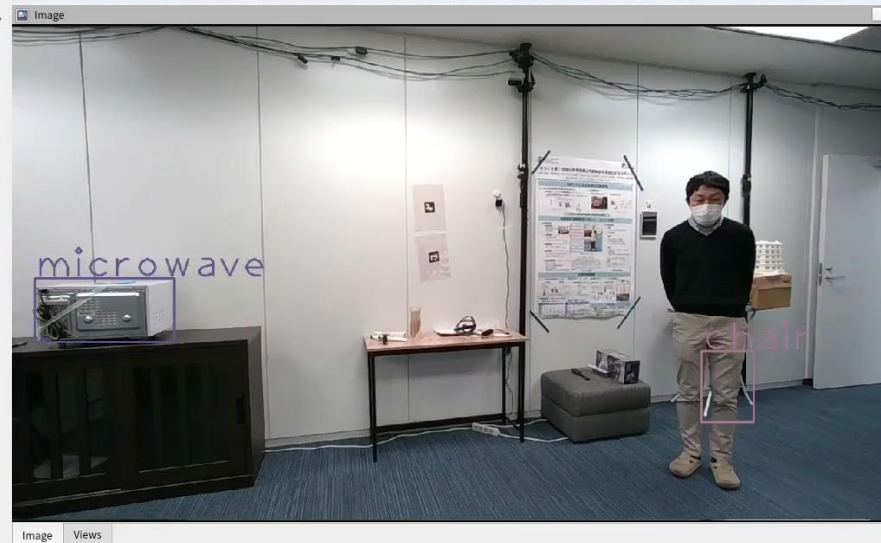
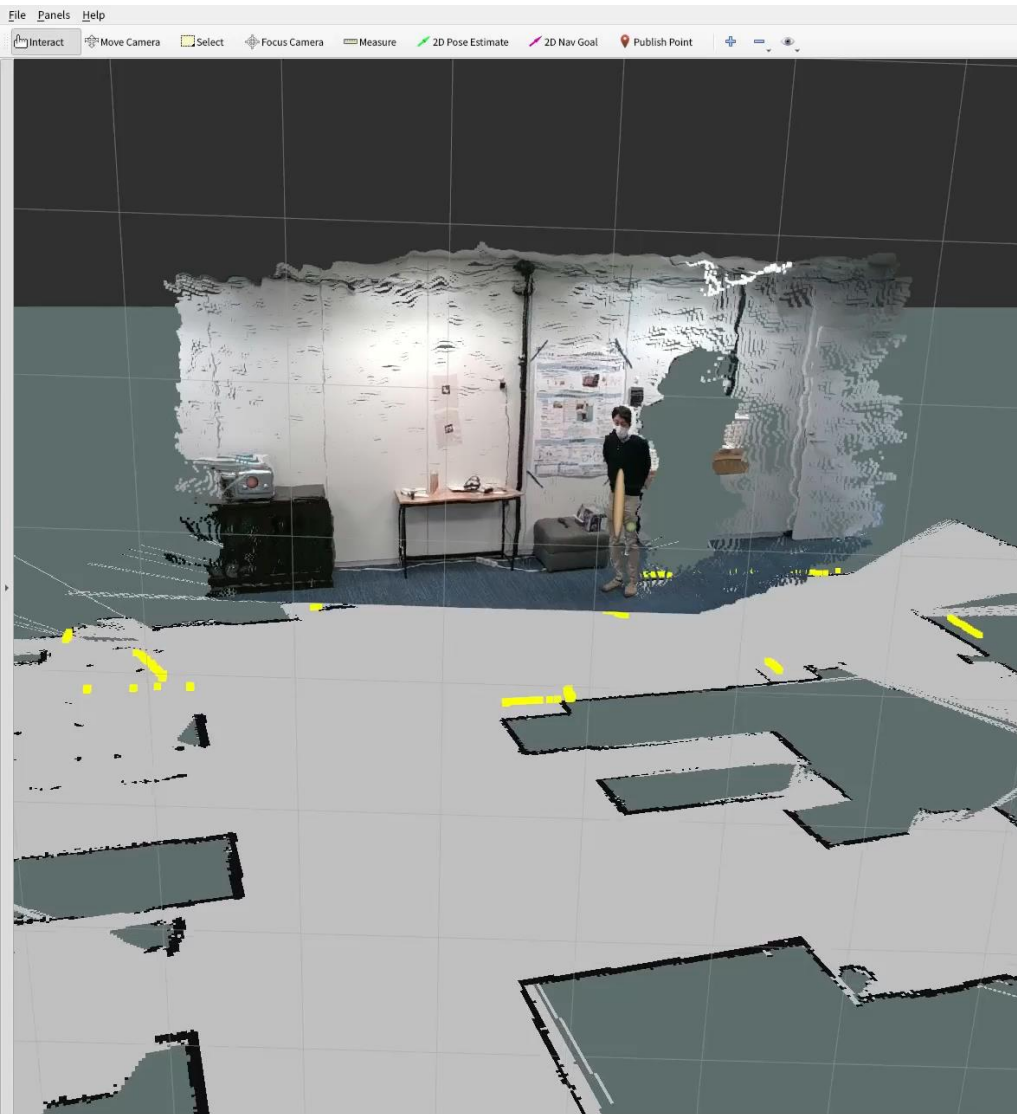


To realize appropriate robot action selection, situation understanding corresponding to robot's task is important

Understanding the surrounding situation



The first-person view



Reasonable action selection



Thanks for the meal

Say "Thanks for the meal"

↓ Is after

Finish eating lunch

↓ Is before

Return cutlery

Candidate Actions
Bring dishes
Throw out the trash
...
Returning catsup

May I return the catsup on the table?

How do we align the robot action with the situation?
Inference of the robot required? (for explainability?)

Barriers to entry

◆ Robotics research fields had barriers to entry

- Using physical robots requires costs (money, place)
- Robots are complicated systems (many things before working)

◆ The goals of this tutorial are,

● Removing such barriers related to physical robots

- Try to run a simulator of manipulation robot

● Increasing the number of people who are interested in using language to control robots

- Empowered by large language models!

Outline

- ◆ Introduction (Yoshino)
- ◆ Robots and actuations (Nakamura)
- ◆ Understanding from vision (Kurita)
- ◆ Future directions (Yoshino)

- ◆ Break

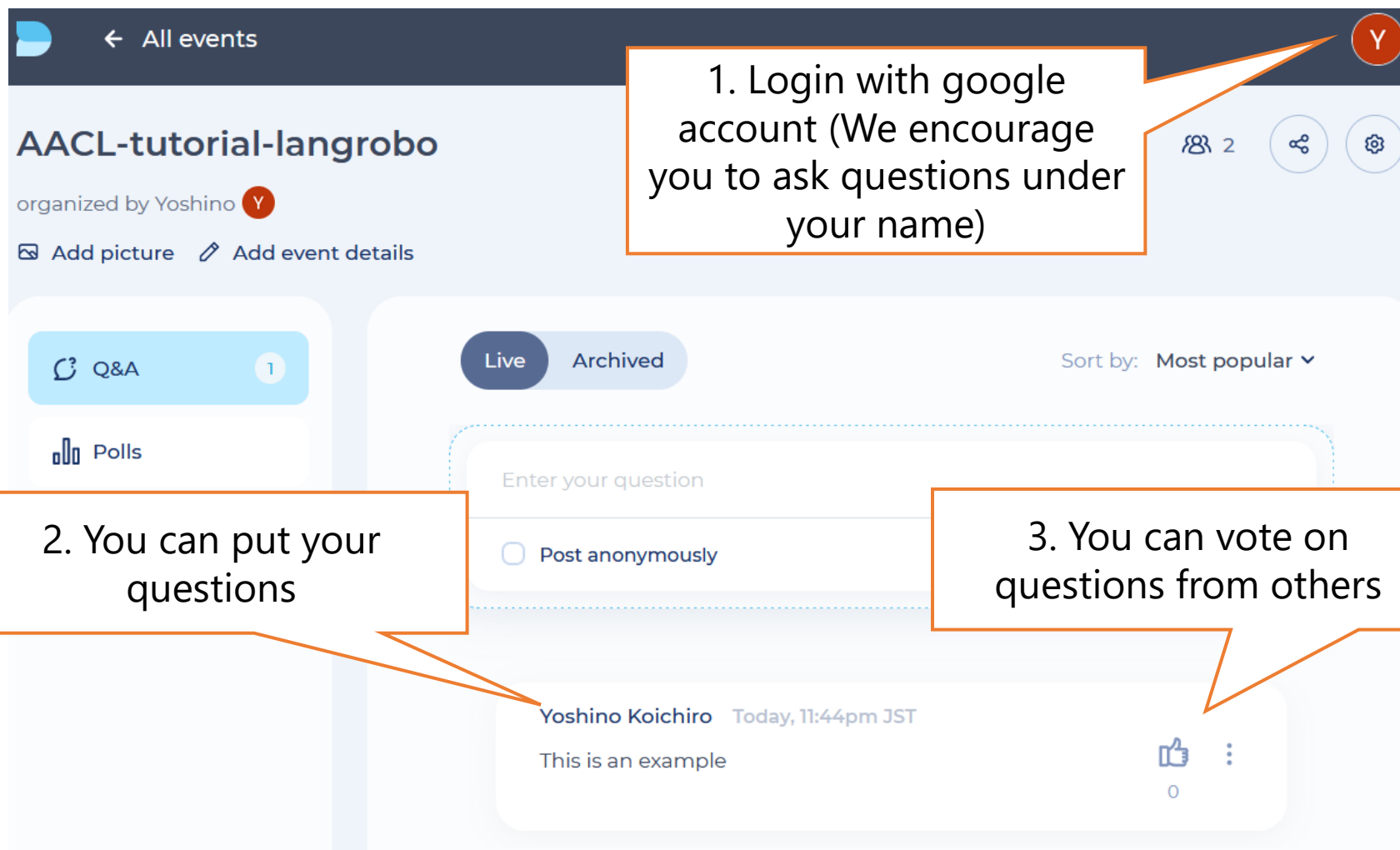
- ◆ Try to run HSR (manipulation robot) on ROS2
- ◆ Discussion using Dory

Discussion on Dory

- ◆ We will use “Dory” for discussion
- ◆ You can put any comments at any timing of the tutorial
- ◆ You can vote on other people’s questions
- ◆ We will have a discussion at the last of the tutorial based on your comments
- ◆ <https://dory.app/events/ORCeryF2DQT5E45K8oS9/ijcnlp-aacl-tutorial-langrobo/>
- ◆ Any tutorial materials are on:
<https://github.com/riken-grp/langrobo-tutorial>



Dory interface



← All events

AACL-tutorial-langrobo
organized by Yoshino

Add picture Add event details

Q&A 1

Polls

Live Archived

Sort by: Most popular

Enter your question

☐ Post anonymously

Yoshino Koichiro Today, 11:44pm JST

This is an example

0

1. Login with google account (We encourage you to ask questions under your name)

2. You can put your questions

3. You can vote on questions from others