

# Retrospective modeling



User: What happened about 8 seconds ago?

Robot: C cut tile. ✔ 😊

User: What happened about 16 seconds ago?

Robot: C cut tile. ✔ 🤔

User: What happened about 24 seconds ago?

Robot: C marked the tile. ✘ 🤔



# Present modeling



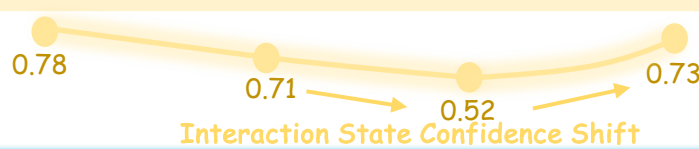
User: Describe what is happening right now?

Robot: C puts potato chips into the pan. ✔ 🤝 😊

Robot: C puts the lid on the pan. ✘ 🤝 😊

Robot: C looks at the pan. ✔ 👁 🤔

Robot: C looks at the pan. ✔ 👁 😊



# Prospective modeling



User: What will happen about 24 seconds later?

Robot: C will make pastry. ✘ 🤔

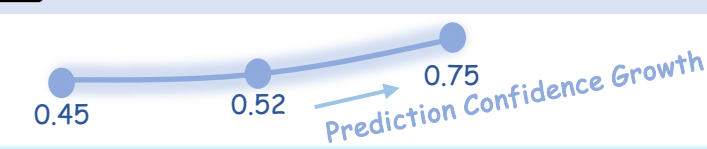
User: What will happen about 16 seconds later?

Robot: C will make dumpling. ✔ 🤔

User: What will happen about 8 seconds later?

Robot: C will make dumpling. ✔ 😊

Multi-Step Prediction



- Anchor Frame    😊 High Response Confidence
- Interaction Frame    🤔 Low Response Confidence
- Non-Interaction Frame    ✔ Correct ✘ Incorrect
- 🤝 Interaction    👁 Non-Interaction

