Conversation, action, and collaboration: Communities of smart devices

ROBOTFORUM-2024
Porto, Portugal
June 17, 2024

Deborah Dahl
Conversational Technologies

Linux Foundation AI&Data Foundation, Open Voice Interoperability Initiative

dahl@conversational-technologies.com

Collaboration among Intelligent Assistants

- There are millions of chatbots and voice bots (conversational assistants) in the world
- Hosted on mobile phones, smart speakers and websites
- Hosted by many organizations – government, business, and non-profit
- Each one is independent of the others, even within an organization
- Each one has its own expertise
- This leads to:
  - implementation complexity
  - duplication of effort
  - friction for users

Open Voice Interoperability Initiative Goals

- Enable independent conversational assistants on diverse platforms to communicate
- Reduce friction for users so that they don’t have to visit multiple assistants to get the information they need
- Enable multiple legacy assistants within an enterprise to interact, even if they were developed on different platforms

If intelligent assistants can collaborate, why not other AIs?

- Robots
- IoT devices

They can collaborate, too!

Some use cases

Home robots:

- An elderly or disabled person has a personal robot that helps them with tasks such as medication reminders, companionship, and turning on lights
- They get a new robot that can vacuum their house but using it requires learning an entirely new user interface
- They get a robot pet, but have to learn how it works

Self-driving cars:

- How do you learn how these work, especially if they aren’t your own car

Public robots:

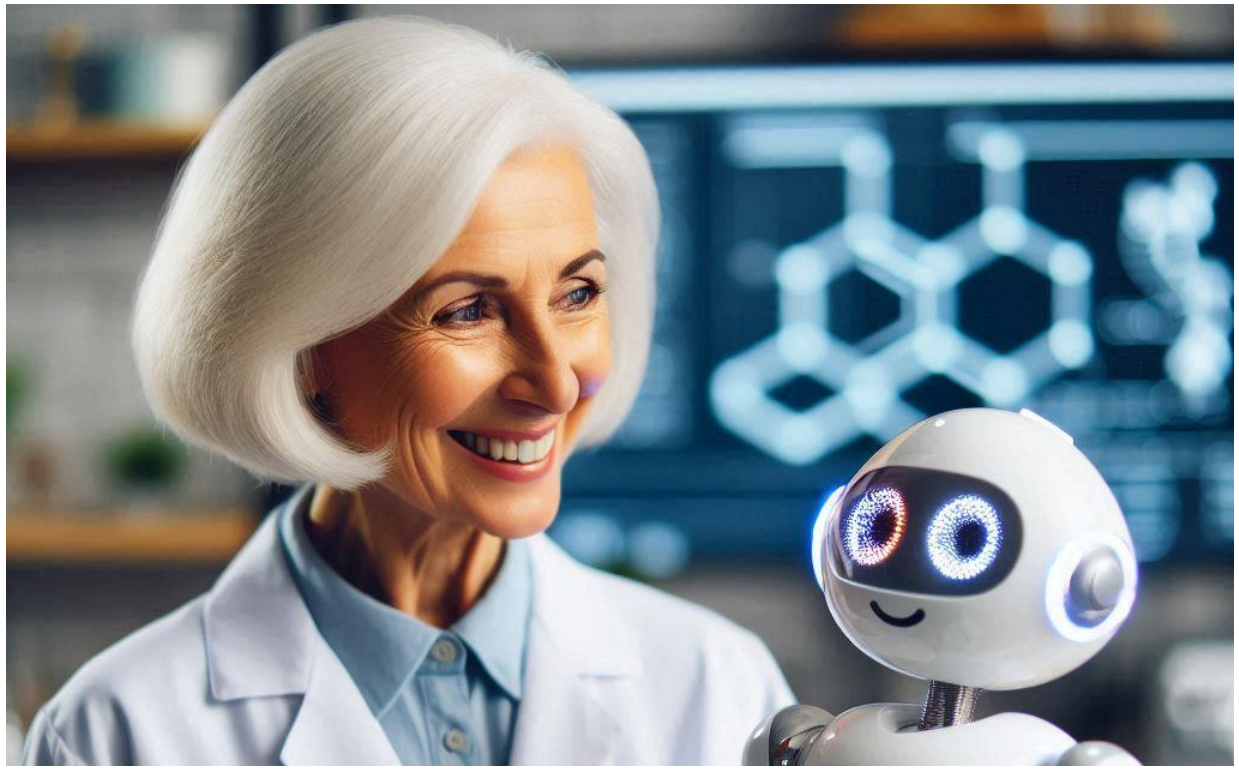
- A store has a robot that can help customers select their items, get items from inside cases or down from high shelves, but again, customers have to learn the user interface
- An airport assistant helps users find their way around the airport, for example, finding gates, services and baggage carousels

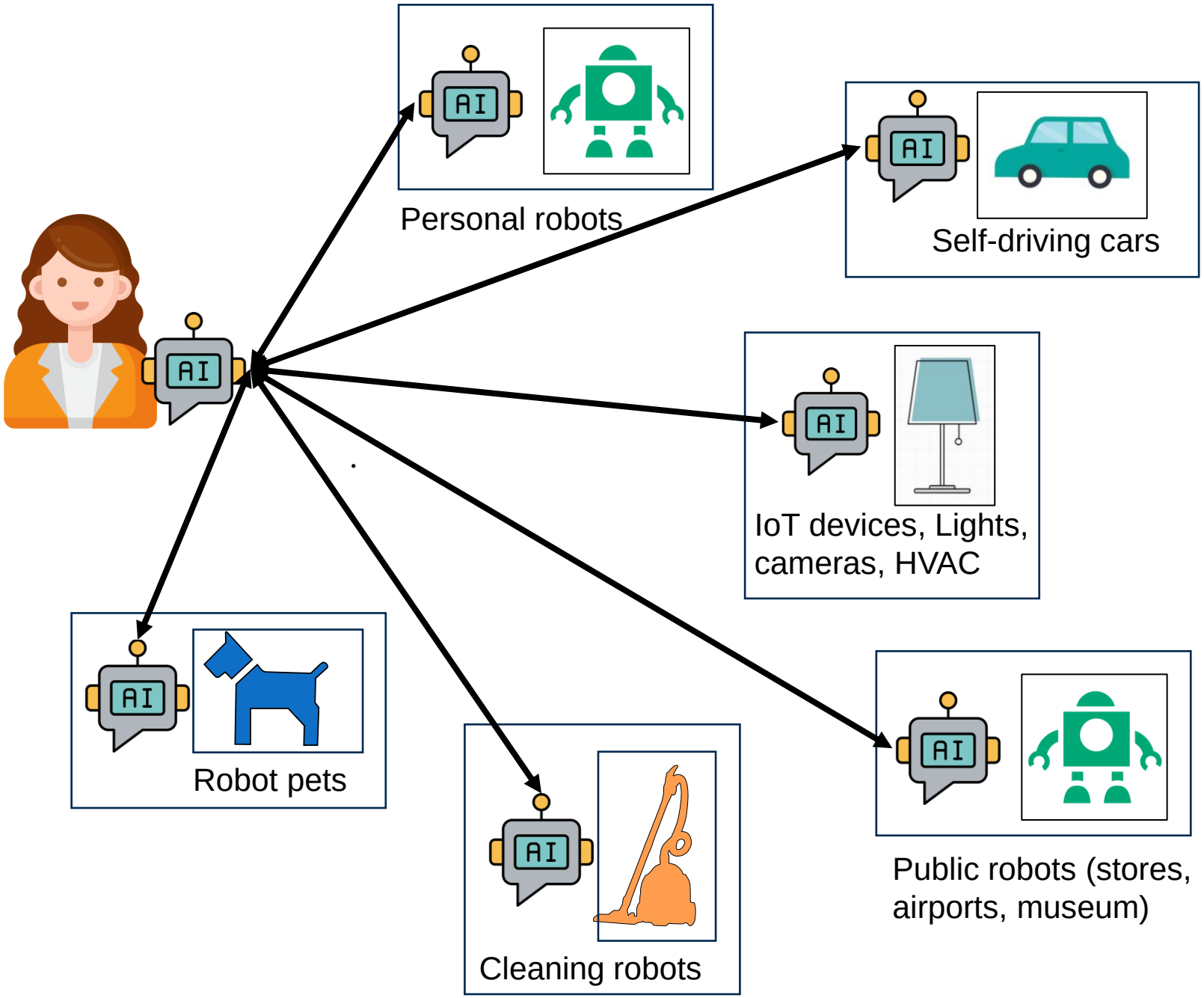
Other AI Collaborations

- We can envision conversational interaction with autonomous vehicles, which could, in turn, collaborate with the highway system or police to provide passengers with information about highway conditions.
- Conversations with smart cities could help travelers find and interact with local services of all kinds.
- Robots that assist elderly and disabled people could be instructed, using ordinary spoken conversations, to collaborate with home automation systems to adjust users' environments as needed.

The key to enabling these use cases is that their user interfaces are based on **speech and natural language**

- We already know how to speak and type
- We can ask these AIs to do things
  - without understanding much about their conventional user interfaces
  - if we can talk to them
- In the past couple of years, thanks to the availability of LLMs, it has become much easier to build speech and language interfaces



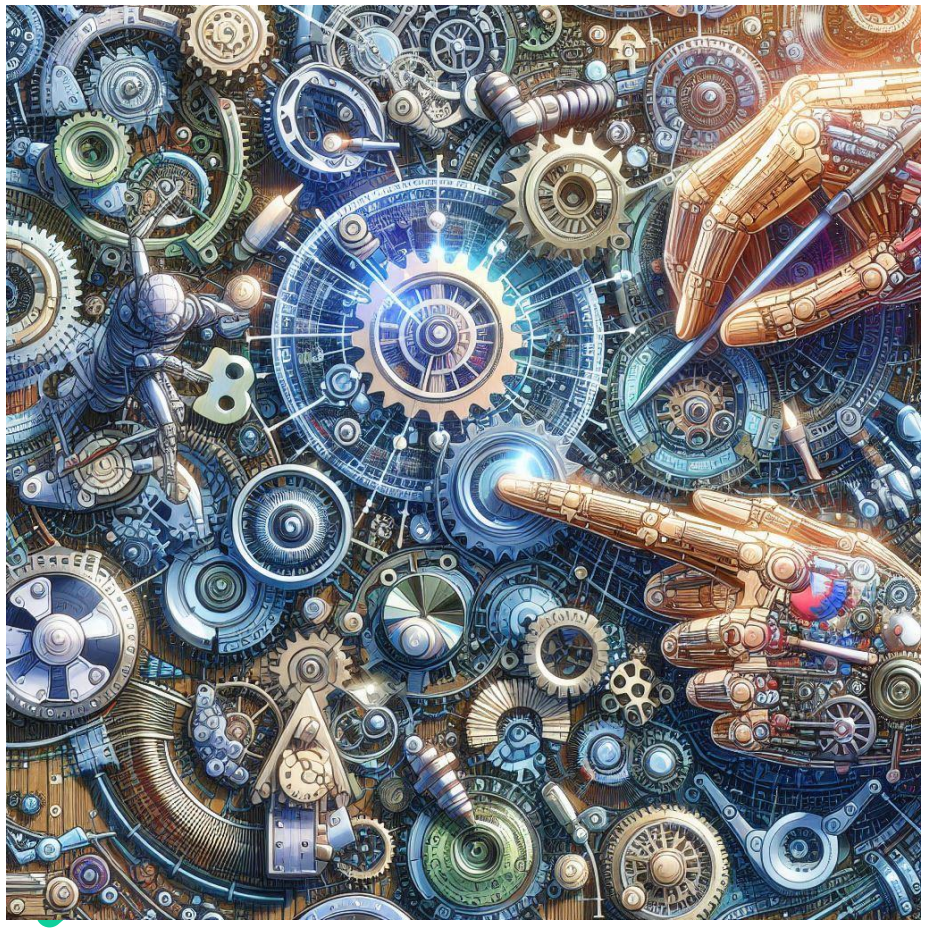

AI
Personal robots
AI
Self-driving cars
AI
IoT devices, Lights, cameras, HVAC
AI
Robot pets
AI
Cleaning robots
AI
Public robots (stores, airports, museum)

Interoperability among smart devices

How does this work?

- Connectivity
- Message formats that they all understand
- The Open Voice Interoperability Initiative provides standard message formats for
  - Assistant Description
  - Establishing and ending a connection between devices
  - Exchanging information

Open Voice Interoperability Standards



- Open source
- Dynamic, real-time configurations of AIs
- Standard messages
- Few assumptions about assistant architecture and capabilities
- Resources: implementation sandbox and specs in our GitHub repositories https://github.com/open-voice-interoperability

```
{
    "identification":
    {
        "serviceEndpoint": "https://192.168.20.20:5582",
        "organization": "My house",
        "conversationalName": "Roboty",
        "serviceName": "Everyday tasks",
        "role": "help with everyday tasks",
        "synopsis": "robot that helps with everyday tasks"
    },

    "capabilities": [
        {
            "keyphrases": [
                "lights","get mail","control TV","control music", "socializing","answer phone", "heat up
food in microwave"
            ],
            "languages": [
                "en-us"
            ],
            "descriptions": [
                "This robot can perform many everyday household tasks such as controlling
appliances. It can also have a conversation, play games, and otherwise socialize with a user"
            ],
            "supportedLayers": [
                "speech","text"
            ]
        }
    ]
}
```

AssistantManifest – how an assistant describes itself

Conversation Envelope and Dialog Events

SENT:


```
{
  "ovon": {
    "conversation": {
      "id": "convoID8403984"
    },
    "schema": {
      "version": "0.9.0",
      "url": "not_published_yet"
    },
    "sender": {
      "from": "Debbie"
    },
    "responseCode": {
      "code": 200,
      "description": "OK"
    },
    "events": [
      {
        "eventType": "utterance",
        "parameters": {
          "dialogEvent": {
            "speakerId": "Debbie",
            "span": {
              "startTime": "2024-03-21T20:35:05.234Z"
            },
            "features": {
              "text": {
                "mimeType": "text/plain",
                "tokens": [
                  {
                    "value": "which planet has more moons, Mars or Jupiter"
                  }
                ]
              }
            }
          }
        }
      }
    ]
  }
}
```

Message returned from the other assistant

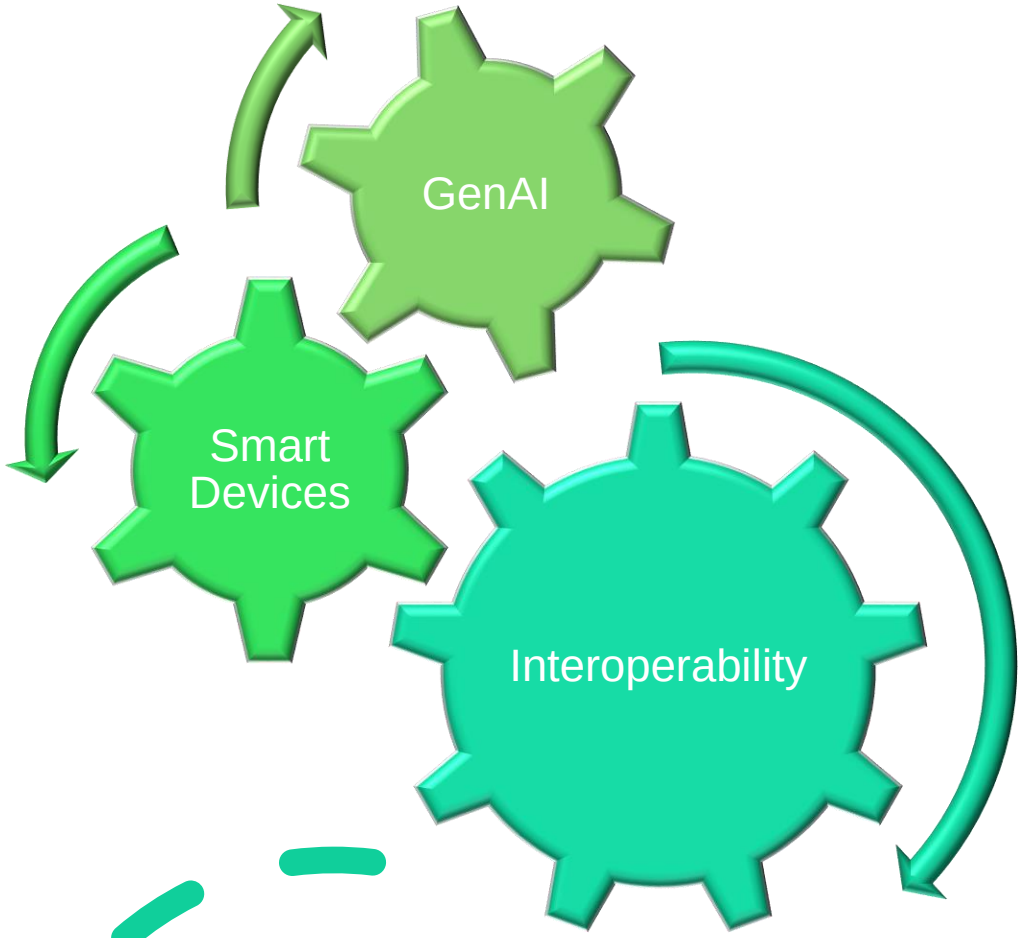
RECEIVED:


```
{
  "ovon": {
    "conversation": {
      "id": "convoID8403984"
    },
    "sender": {
      "from": "https://www.asteroute.com/ovontest"
    },
    "responseCode": 200,
    "events": [
      {
        "eventType": "utterance",
        "parameters": {
          "dialogEvent": {
            "speakerId": "wizard",
            "span": {
              "startTime": "2024-03-21 20:35:06.039699686 +0000 UTC"
            },
            "features": {
              "text": {
                "mimeType": "text/plain",
                "tokens": [
                  {
                    "value": "Jupiter"
                  }
                ]
              }
            }
          }
        }
      },
```

Privacy, safety and security?

- Underway
- We intend to use standard mechanisms like
  - passwords
  - two-factor authentication
  - biometrics
- That are in use in regular websites
- No need to reinvent the wheel

Three Converging Technologies



Generative AI

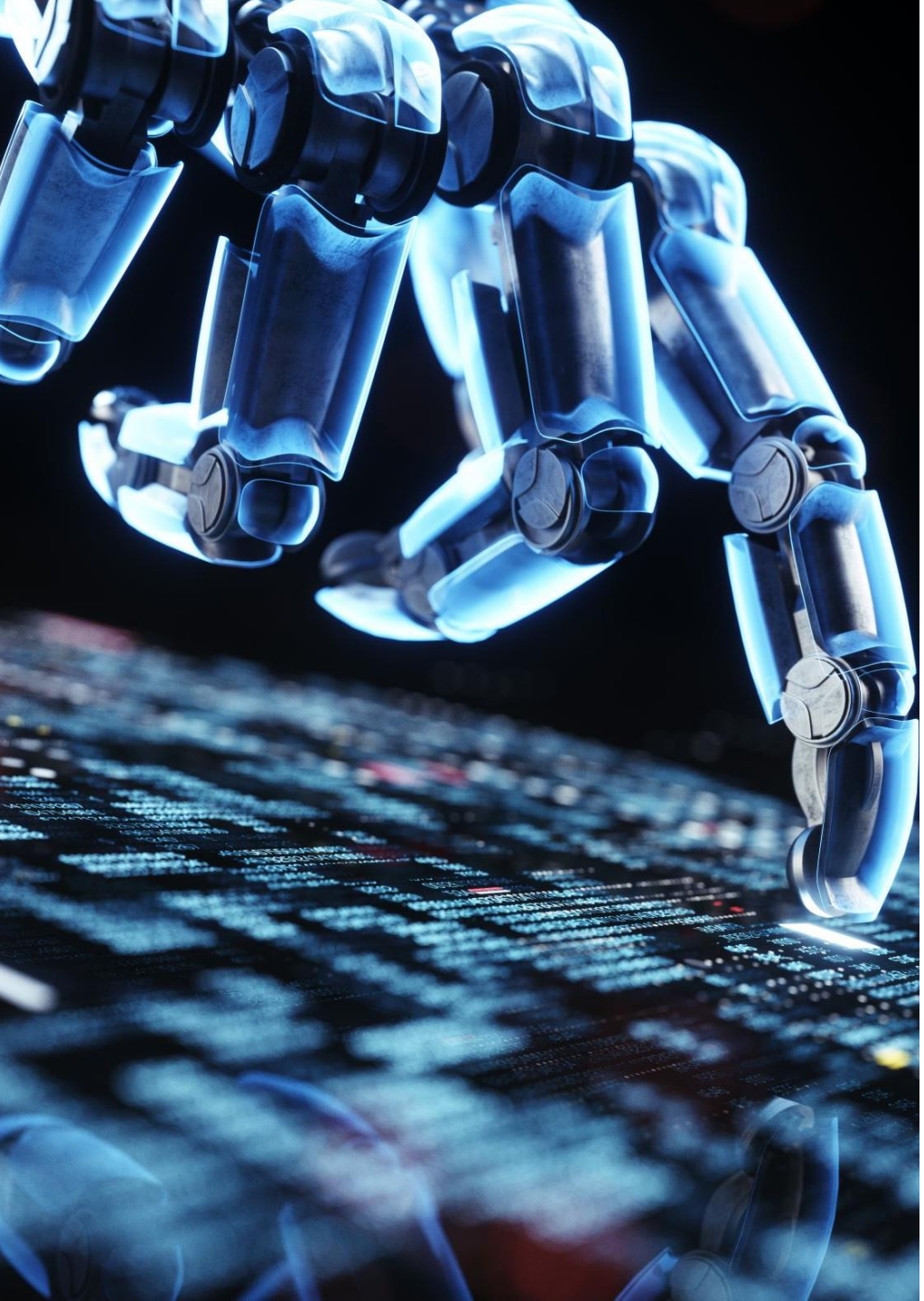
- Large Language Models are leading to a dramatic improvement in speech recognition and natural language understanding technology, making it much more cost-effective to add spoken language interfaces to systems

Smart devices

- A second accelerating technical direction is the increasing ability of devices such as cars, aircraft, smart environments, and robots to sense the state of the world and act on it.

Interoperability

- The third stream is the increasing ability of independent systems to interoperate in helping users achieve their goals.



Summary

- There are exciting opportunities to leverage LLMs, robot technologies and interoperability standards to enable many new kinds of applications

What you can do

- Try out our interoperability specs, especially try them out for robot control
- Join us in continuing to define our specs!

More information

https://github.com/open-voice-interoperability