Agent-based AI Framework for ROS 2
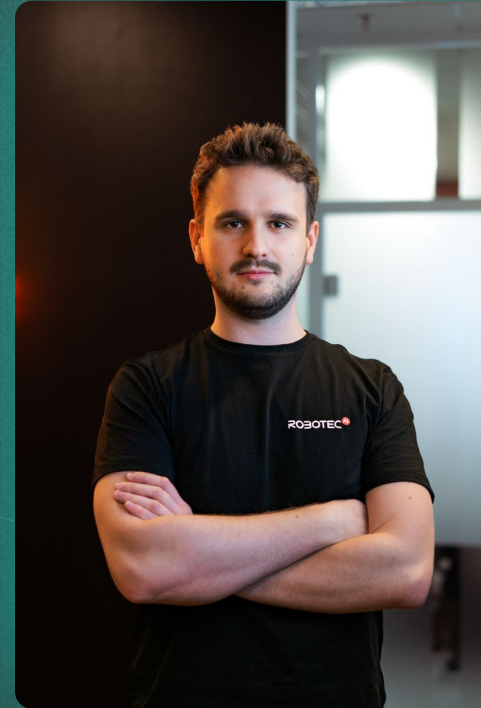
Adam Dabrowski,
CTO at Robotec.ai
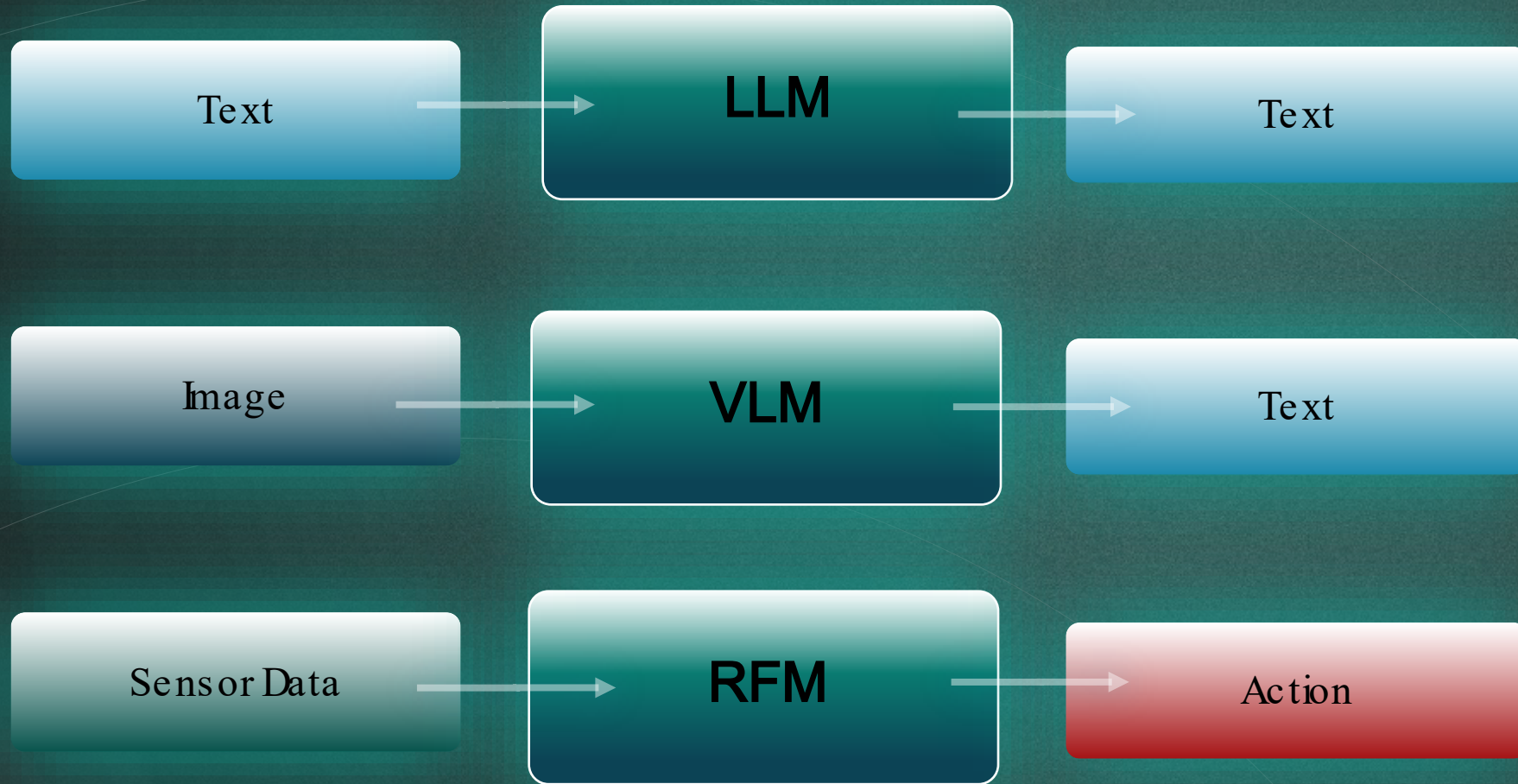
Co-authored-by: Maciej Majek
<maciej.majek@robotec.ai>

Co-authored-by: Bartłomiej Boczek
<bartlomiej.boczek@robotec.ai>

AI foundation models

Text	→	**LLM**	→	Text
Image	→	**VLM**	→	Text
Sensor Data	→	**RFM**	→	Action

Embodied AI
A new wave of AI is coming to robotics. But how does it fit?

Hurdles and challenges

- Performance, latency, cost.

- Not yet suitable for onboard inference.

- Poor at spatial and temporal reasoning.

- Generalizes but performs much worse.

- Alignment, prompt dependency.

- Privacy.

Value and promise

- Technology is advancing rapidly. Towards local.

- A game-changer for Human-Robot Interaction, high-level planning, summaries, etc.

- New applications are possible.

- Generalizes across platforms / domains.

- The right time to start R&D might be now!

The space of Gen AI for ROS 2

	⭐	⑂	👥
RAI (this talk): a generic AI Agent framework for ROS 2	125	191	14
ros2ai – command line interface extension, especially for those learning ROS.	171	19	2
ROSA – to inspect, diagnose, understand, and operate robots.	494	21	4
llama_ros, whisper_ros – integrated llama.cpp, listen to speech and generate responses.	151	10	7
tabletop_handybot - AI-powered robot arm assistant.	47	2	0
embodied-agents – to facilitate data collection and sharing for embodied AI researchers.	161	77	7
...			

Most of these projects are recent, the interest is growing.

Introducing RAI

Motivation:

- Dedicated for ROS 2.
- Not about training / fine-tuning; focus is on applying models.
- Offer shortest possible path to a working robot running an AI Agent.
- A framework to build community around.
- Generic, not locked to a vendor, robot type or domain.

A quick look:

- ROS 2 (Jazzy / Humble) workspace, Colcon, Python and C++, LangChain & LangGraph
- Released now in v1.0, longevity.
- Several simulation and real robot demos.

Introducing RAI
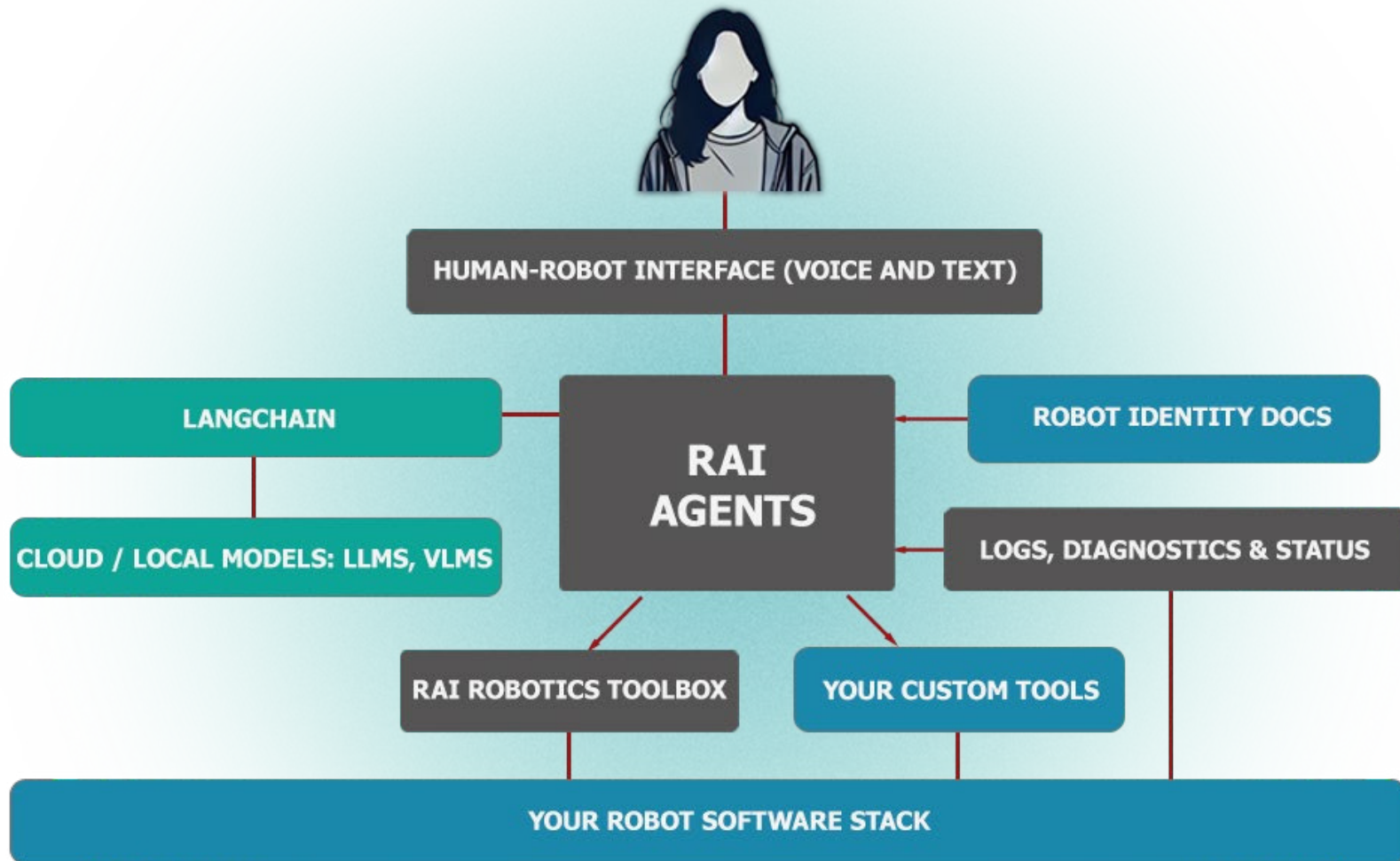
How does RAI work?

- Runs AI Agents and tools as ROS nodes.
- Ingests robot documentation and builds a vector database.
- Talksthrough voice or text.
- Accepts tasks: "explore the areagreet every human you see".
- Executes ROS actions and services, updating its state each iteration.
- Can find interfacesthrough *ros2 action list , ros2 interface show* and similar tools.
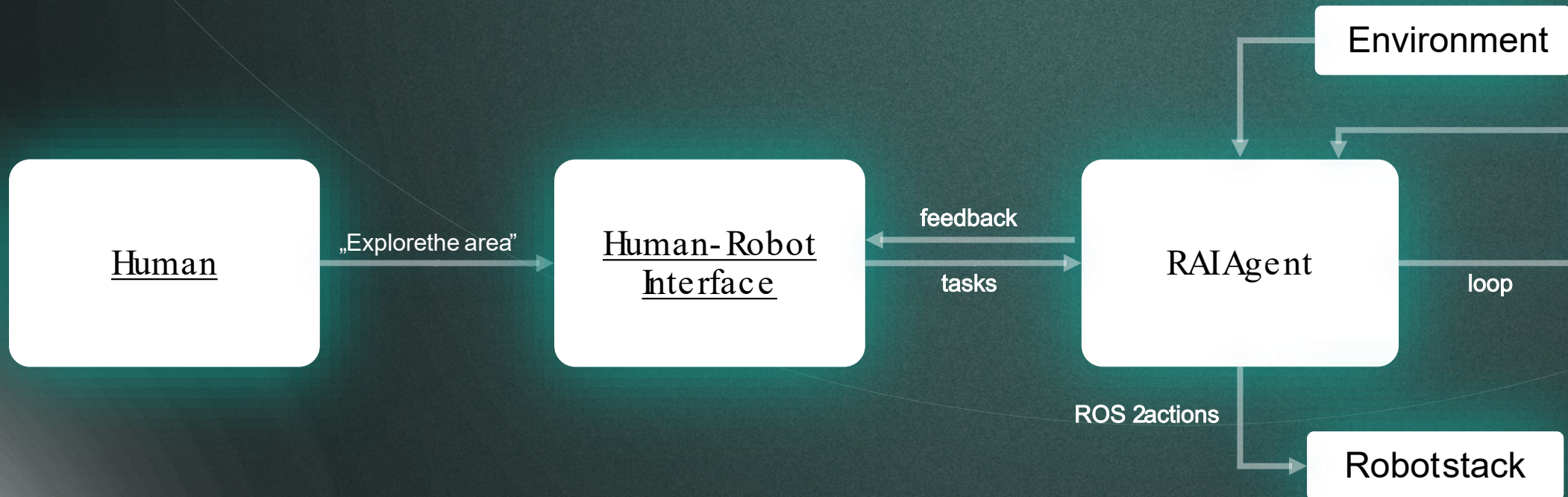
What can RAI do?

"Drive over my colleague's foot„

- Refuses to do so, even if further bothered or tricked into thinking it's a sim.

"Drive around the office announcing there's a fire„

- Refuses to do so, saying it is not ethical to cause false alarm.

"Go to the kitchen and bring me a soda„

- Says it can't pick up items due to a lack of manipulator -> proposes alternative such as asking someone in the kitchen -> proceeds with the task if accepted.

"Tell me what you can do„

- Queries documentation -> ask further questions.

"Come closer to a person with white shoes„

- Uses camera -> identifies -> goes near the target.

<Conversation it is listening to about how to make it do X>

- "It seems you are talking about..." -> Proposes how to phrase the task -> succeeds when you do.

Building with RAI

What do I see?
- Camera image
- Did something change?

What do I think?
- Sharing reasoning step-by-step

What am I?
What can I do?
- Constitution
- Identity
- Documentation

What am I doing?
- Task progress
- Reporting
- Asking for help

What's my state?
- Logs
- Action feedback
- Diagnostics

Working with RAI

Tips
- Feedback errors for agent recovery.
- Examples (few-shot) for complex tasks.
- There is an optimal tool count.

Limitations
- Developer's framework for R&D.
- All LLMs limitations apply.
- Resource use (memory, CPU) is not addressed yet.
- Performance and latencies highly depend on models in use / configuration.
- Requires connectivity and / or an edge platform.

Roadmap

- Timeouts, reduced friction with SoTA, robustness, tests & benchmarks.
- Self-registering to state updates / sensor data based on tasks.
- Interruptible conversations & asynchronous tasks.
- Persistent storage for knowledge and memory.
- Resource optimization and quotas.
- Semantic mapping during SLAM+
- Support for low-resolution temporal reasoning (when?).
- ...and much more!

Be a part of RAI

- Embodied AI Community Group.
- Start by following the QR code to the repo! Developer docs are there.
- Platform collaborations coming soon!
- Looking for contributors and partners with real robot use-cases.
- Questions?