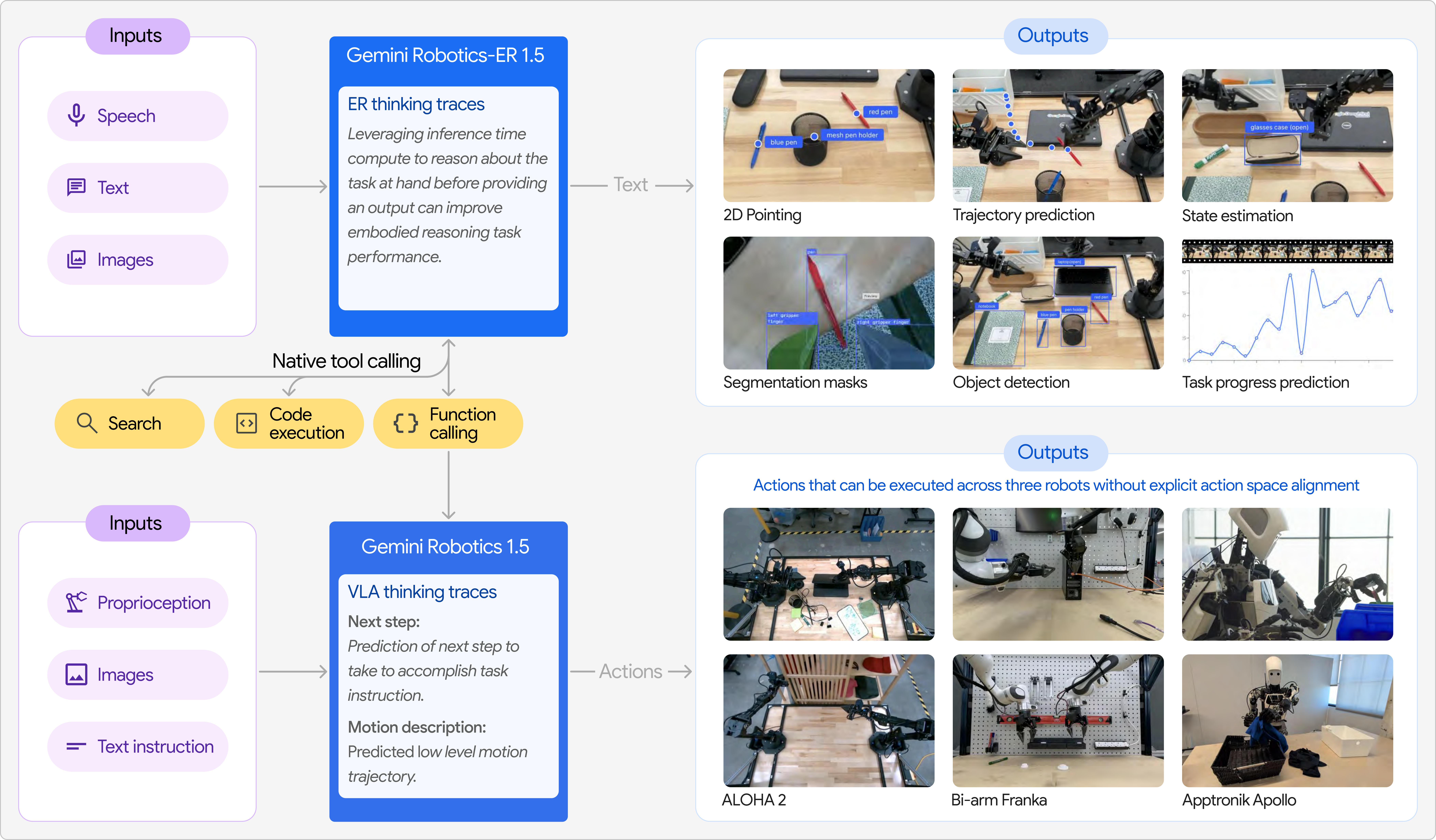

Inputs

Speech

Text

Images

Gemini Robotics-ER 1.5

ER thinking traces

Leveraging inference time compute to reason about the task at hand before providing an output can improve embodied reasoning task performance.

Text

Outputs

2D Pointing

Trajectory prediction

State estimation

Segmentation masks

Object detection

Task progress prediction

Native tool calling

Search

Code execution

Function calling

Inputs

Proprioception

Images

Text instruction

Gemini Robotics 1.5

VLA thinking traces

Next step:

Prediction of next step to take to accomplish task instruction.

Motion description:

Predicted low level motion trajectory.

Actions

Outputs

Actions that can be executed across three robots without explicit action space alignment

ALOHA 2

Bi-arm Franka

Apptronik Apollo