# Multi-Modality —
# For Conversational AI

## SiMa Value Proposition

- Enables natural, real-time conversation with your equipment to simplify set-up, use, and repair.

- Accommodates multiple interaction modalities — text, voice, and images to enhance ease of use and accelerate decision making.

- Provides actionable insight exactly when it's needed.

- Supports Retrieval Augmented Generation (RAG) to expand model's domain expertise without retraining.

- Brings the power of generative AI anywhere—even off-grid—in under 10W.

**Demo Controls**

Chat · Talk · Snap · Select Image · Settings · Refresh

**Text  Voice  Image**

*"How do I clean the variator drive belt in a New Holland combine harvester?"*

**Expand model context — without retraining!**

OPERATOR'S MANUAL
TC5040
TC5050
TC5060
TC5070
TC5080
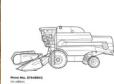
*To clean the variator drive belt on a New Holland combine harvester, follow these steps: 1. Move the variator to the middle position. Stop the engine. 2. Loosen the screw and adjust with the nut. 3. Correct belt tension: obtain an upper belt deflection of 14 mm halfway up the front section when applying a force of 30 N.*
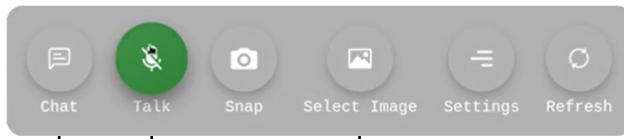
**Natural interaction with on-device AI**

## Modalix brings Large Multi-Modal Models (LMMs) to life for Physical AI

Across factories, vehicles, and field equipment, systems are evolving from command-based control to natural interaction. SiMa.ai Modalix enables that shift — running large multi-modal models (LMMs), LLMs, and VLMs entirely on-device for real-time intelligence without cloud latency, cost, or privacy concerns.

- Transforms equipment into an intelligent assistant that understands intent, not just inputs.
- Enables natural human-machine interaction through text, voice, and vision.
- Runs fully on-device — no cloud or host required — cutting cost, latency, and privacy risk.
- Accelerates decision-making and innovation, driving measurable ROI from on-device intelligence.

# Conversational AI, Multi-Modality Demo
## Powered by SiMa.ai Modalix Running LLaVA & TinyLLaVA at Under 10W

## Technical Details

Multi-modal AI — LMMs, LLMs, and VLMs work together to understand text, voice, and images.

Runs CLIP, Whisper, Piper TTS, and LLaVA or TinyLlaVA (among other models) fully on-device — no cloud or host needed.

Private by design since no need to access cloud.

Real-Time Response — delivers up to 13 TPS and 0.5 s TTFT for human-like interaction.

Under 10 W for efficient & thermally stable operation.

**MLSoC™ Modalix**

**Application Processing Unit** (ARM A65) | **ML Accelerator (MLA)**

**Piper** (TTS)

**TinyLLaVA** — CLIP, multi-modal understanding with RAG context

**Whisper-Small** (STT)

**RAG System** — Knowledge base, vector embeddings, retrieval engine
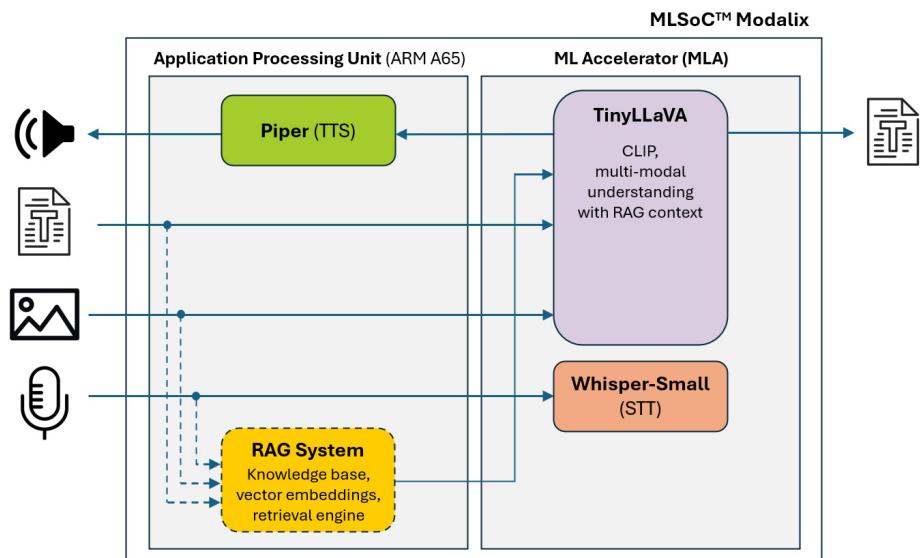
Convert Hugging Face models into optimized, binaries with **SiMa's LLiMa framework**, integrated into the **Palette SDK**.

## Hardware

### Our Silicon
**Modalix**

### Our Boards
**SoM**  **SoM Dev**  **PCIe HHHL**

## Software

Palette Edgematic™   Palette™ SDK   LLiMa by SiMa

🤗 **Hugging Face**   huggingface.co/simaai

**Model Zoo**   sima.ai/model-browser/

## About SiMa

SiMa.ai is a leader in Physical AI, delivering a purpose-built, software-centric platform that brings best-in-class performance, power efficiency, and ease of use to Physical AI applications. Focused on scaling Physical AI across robotics, automotive, industrial automation, aerospace & defense, smart vision, and healthcare, SiMa.ai is led by seasoned technologists and backed by top-tier investors. Headquartered in San Jose, California.
Learn more at **www.sima.ai**.