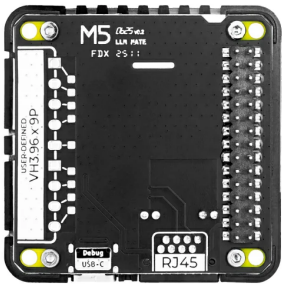
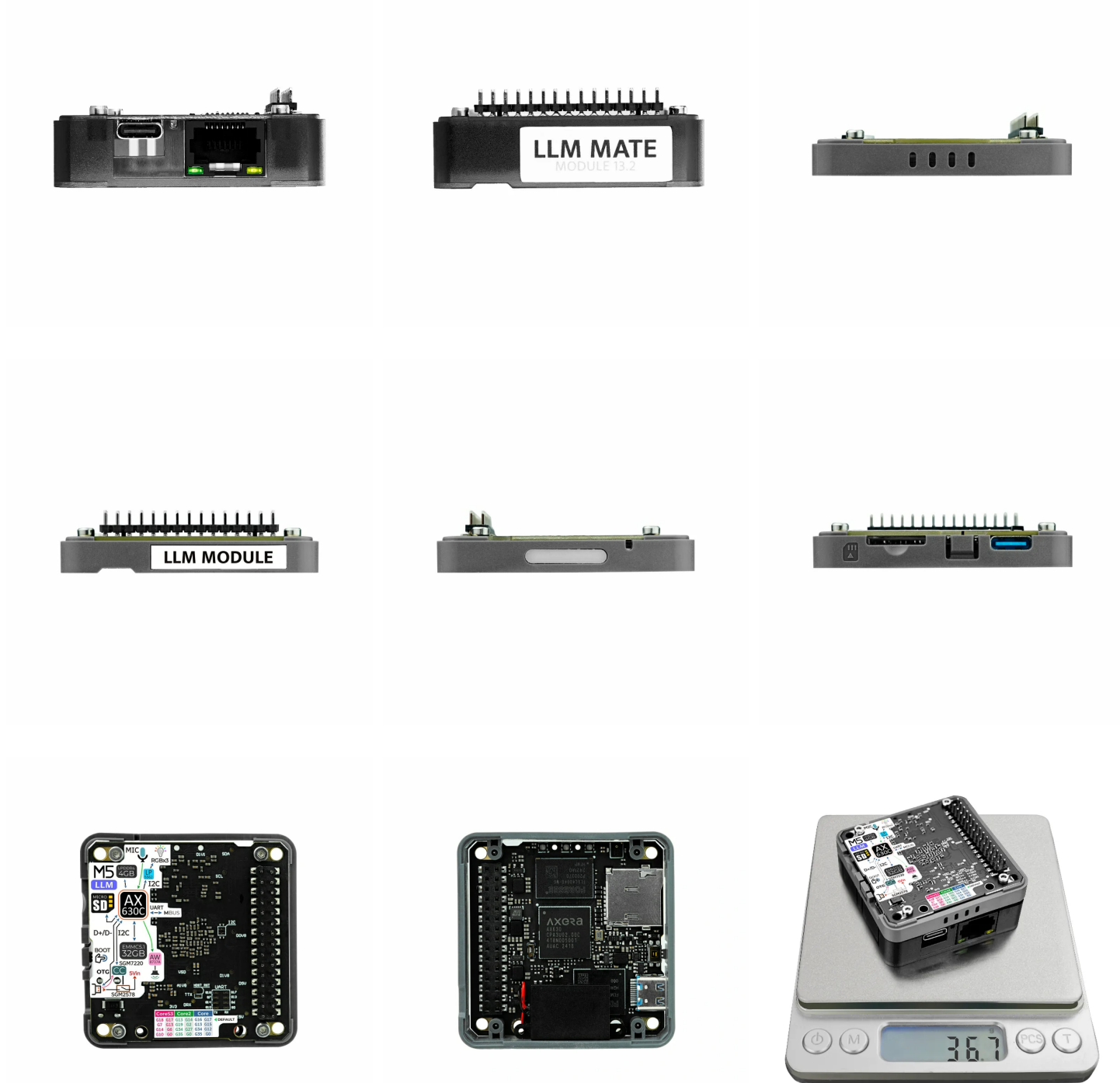


Module LLM Kit

SKU:K144





Description

Module LLM Kit is a smart modular kit focused on offline AI inference and data communication interface applications. It integrates the Module LLM and Module13.2 LLM Mate modules to meet the offline AI inference and data interaction requirements across various scenarios.

Module LLM is an integrated offline large language model (LLM) inference module designed specifically for terminal devices that require efficient and intelligent interaction. Whether for smart home applications, voice assistants, or industrial control, Module LLM delivers a smooth and natural AI experience without relying on the cloud, ensuring privacy, security, and stability.

Module13.2 LLM Mate Module provides a variety of interface functions to facilitate system integration and expansion. It achieves stacked power supply with Module LLM via the M5-Bus interface; its built-in CH340N USB conversion chip offers USB-to-serial debugging functionality, while the Type-C interface is used for USB log output. Additionally, the RJ45 interface works with the onboard network transformer to extend to a 100 Mbps Ethernet port and core serial port (supporting SBC applications); the FPC-8P interface connects directly to Module LLM, ensuring stable serial communication; furthermore, an HT3.96*9P solder pad is reserved for DIY expansion.

The Module LLM module integrates the StackFlow framework along with the Arduino/UiFlow libraries, allowing edge intelligence to be implemented with just a few lines of code. Powered by the advanced AiXin AX630C SoC processor and featuring a high-efficiency NPU delivering 3.2 TOPS with native support for Transformer models, it effortlessly handles complex AI tasks. Equipped with 4GB LPDDR4 memory (1GB for user applications and 3GB dedicated to hardware acceleration) and 32GB eMMC storage, it supports parallel multi-model loading and chained inference, ensuring smooth multitasking. With an operating power consumption of only about 1.5W, it is far more energy efficient than similar products, making it ideal for long-term operation.

Module LLM is compatible with multiple models and comes pre-installed with the Qwen2.5-0.5B large language model, featuring built-in functions including KWS (wake word), ASR (speech recognition), LLM (large language model), and TTS (text-to-speech). It also supports apt-based rapid updates of software and model packages. By installing the openai-api plugin, it becomes compatible with the OpenAI standard API, supporting chat, conversation completion, speech-to-text, and text-to-speech among various application modes. The official apt repository offers abundant large model resources—including deepseek-r1-distill-qwen-1.5b, InternVL2_5-1B-MPO, Llama-3.2-1B, Qwen2.5-0.5B, and Qwen2.5-1.5B—as well as a text-to-speech model (melotts) and speech-to-text models (whisper-tiny, whisper-base) and visual models (such as yolo11 and other SOTA models). The repository is continuously updated to support the most cutting-edge model applications, meeting the demands of complex AI tasks.

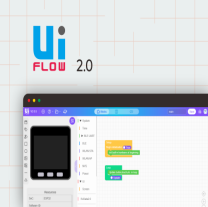
Module LLM Kit is plug-and-play, and when paired with the M5 host, it provides an instant AI interactive experience. Users can seamlessly integrate it into existing smart devices without cumbersome setup, quickly enabling intelligent features and enhancing device performance. This product is ideal for offline voice assistants, text-to-speech conversion, smart home control, interactive robots, and more.

Tutorial



Arduino IDE

This tutorial will show you how to program and control the Module LLM device using the Arduino IDE



UiFlow2

This tutorial will show you how to control the Module LLM device using the UiFlow2 graphical programming platform

Features

- Offline inference, 3.2 TOPS at INT8 precision
- Integrated KWS (wake word), ASR (speech recognition), LLM (large language model), TTS (text-to-speech)
- Parallel multi-model processing
- Onboard 32GB eMMC storage and 4GB LPDDR4 memory
- Onboard microphone and speaker
- Serial communication

- SD card firmware upgrade
- Supports ADB debugging
- RGB status LED
- Built-in Ubuntu system
- Supports OTG functionality
- Development Platform
 - UiFlow1
 - UiFlow2
 - Arduino IDE

| Includes

- 1 x Module LLM
- 1 x Module LLM Mate
- 2 x FPC-8P Wire

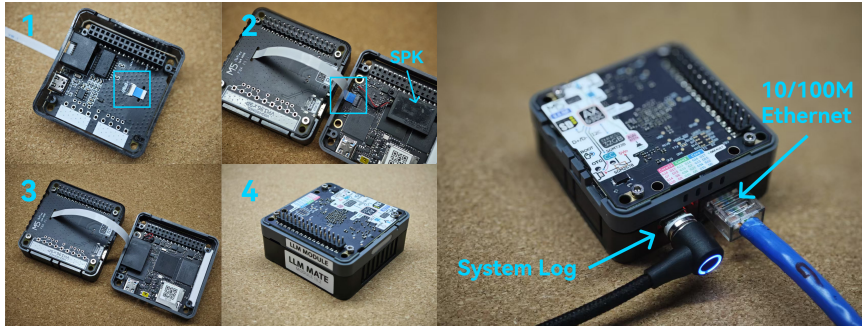
| Applications

- Offline voice assistant
- Text-to-speech conversion
- Smart home control
- Interactive robot

| Specifications

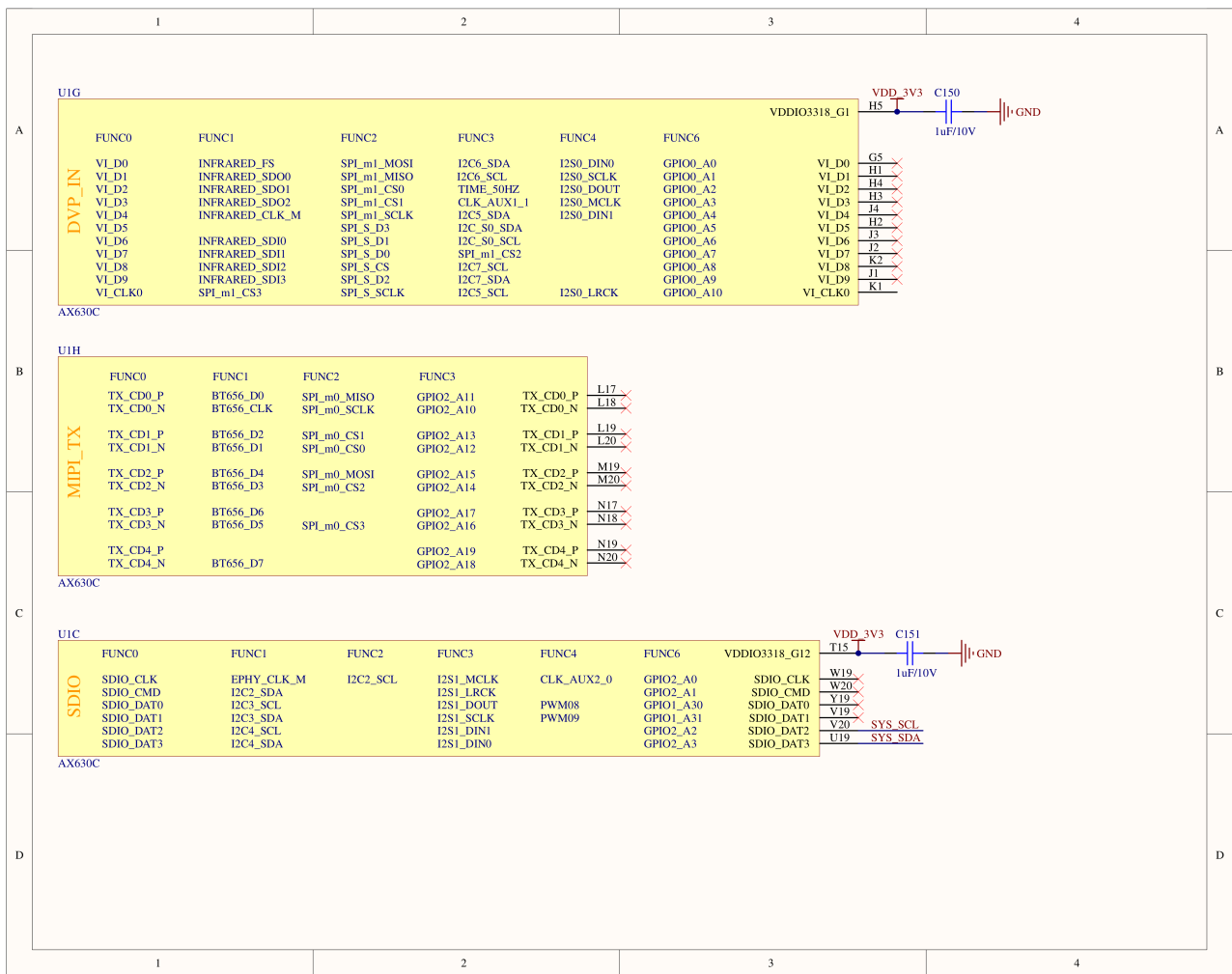
Specification	Parameter
Processor SoC	AX630C@Dual Cortex A53 1.2 GHz MAX.12.8 TOPS @INT4, 3.2 TOPS @INT8
Memory	4GB LPDDR4 (1GB system memory + 3GB dedicated to hardware acceleration)
Storage	32GB eMMC5.1
Communication	Serial communication, default baud rate 115200@8N1 (adjustable)
Microphone	MSM421A
Audio Driver	AW8737
Speaker	8Ω@1W, size: 2014 cavity speaker
Built-in Functions	KWS (wake word), ASR (speech recognition), LLM (large language model), TTS (text-to-speech)
RGB LED	3x RGB LED@2020, driven by LP5562 (status indicator)
Power Consumption	No load: 5V@0.5W, Full load: 5V@1.5W
Button	Used to enter firmware download mode
Upgrade Interface	SD card/Type-C port
Conversion Chip	CH340N
Ethernet Interface	RJ45 interface with onboard network transformer (11FB-05NL SOP-16)
Serial Interfaces	FPC-8P interface, Type-C interface, RJ45 interface
DIY Expansion	HT3.96*9P solder pad
Operating Temp.	0-40°C
Product Size	Module LLM: 54.0 x 54.0 x 13.0mm Module13.2 LLM Mate: 54 x 54 x 19.7mm
Package Size	Module LLM: 192.0 x 95.0 x 17.0mm Module13.2 LLM Mate: 192.0 x 95.0 x 21.0mm
Product Weight	Module LLM + Module13.2 LLM Mate : 36.7g
Package Weight	Module LLM: 32.0g Module13.2 LLM Mate: 34.8g

Connection Steps and Interface Overview

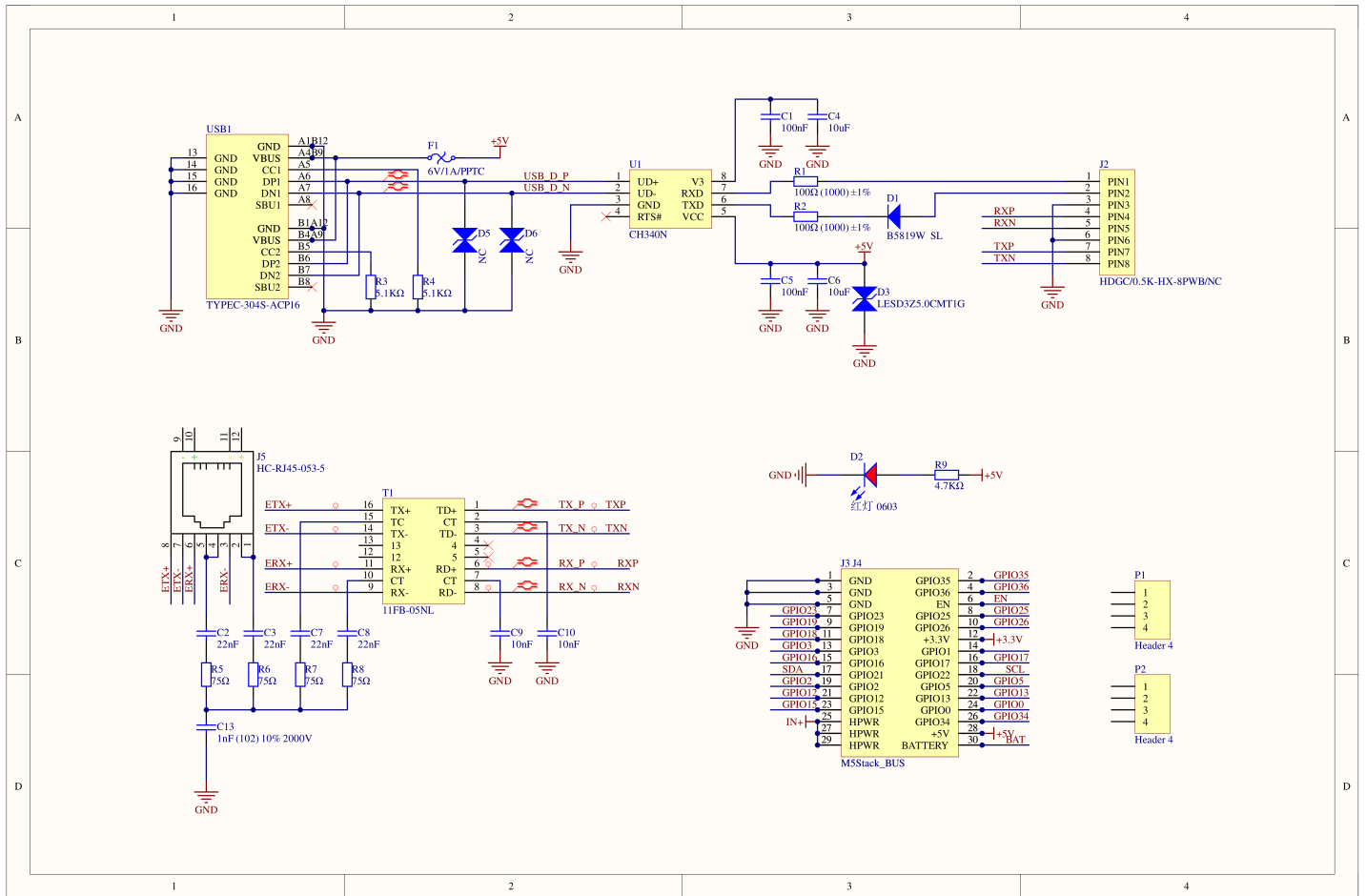


Schematics

- [Module LLM Schematic Download](#)
- [Module13.2 LLM Mate Schematic Download](#)



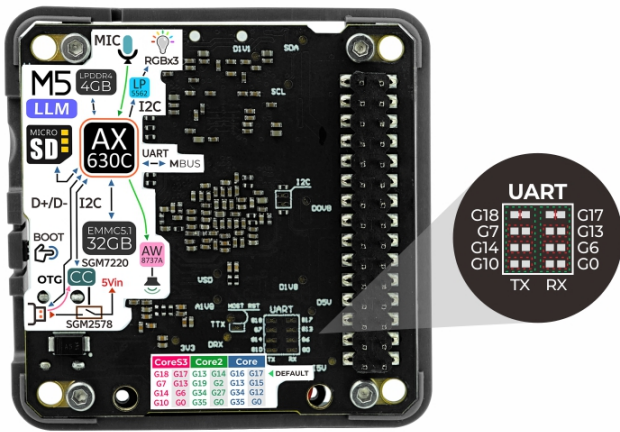
1		2		3		4	
A	MIPI_RX	FUNC0	FUNC1				
		RX_CD0_P	VI_D11	RX_CD0_P	T2		
		RX_CD0_N	VI_D10	RX_CD0_N	T1		
		RX_CD1_P	VI_D13	RX_CD1_P	R2		
		RX_CD1_N	VI_D12	RX_CD1_N	R1		
		RX_CD2_P	VI_D15	RX_CD2_P	P2		
		RX_CD2_N	VI_D14	RX_CD2_N	P1		
		RX_CD3_P	VI_D17	RX_CD3_P	N2		
		RX_CD3_N	VI_D16	RX_CD3_N	N1		
		RX_CD4_P	VI_D19	RX_CD4_P	M2		
B		RX_CD4_N	VI_D18	RX_CD4_N	M1		
		RX_CD5_P	VI_D20	RX_CD5_P	L2		
		RX_CD5_N	VI_CLK1	RX_CD5_N	L1		
C							
D							
1		2		3		4	



PinMap

Module LLM Pin Switching

Module LLM has reserved pin-switching solder pads. In cases where pin multiplexing conflicts occur, you can cut the PCB traces and jumper the connection to another set of pins. Refer to the [tutorial](#).



Tx Rx

CoreS3	Core2	Core	
G18	G17	G13	G14
G7	G13	G19	G2
G14	G6	G34	G27
G10	G0	G35	G0

M5-Bus

- Module LLM

Net-Tie

The pins marked NT in the M5-Bus table below can be switched via the above operation to adapt to different main control devices.

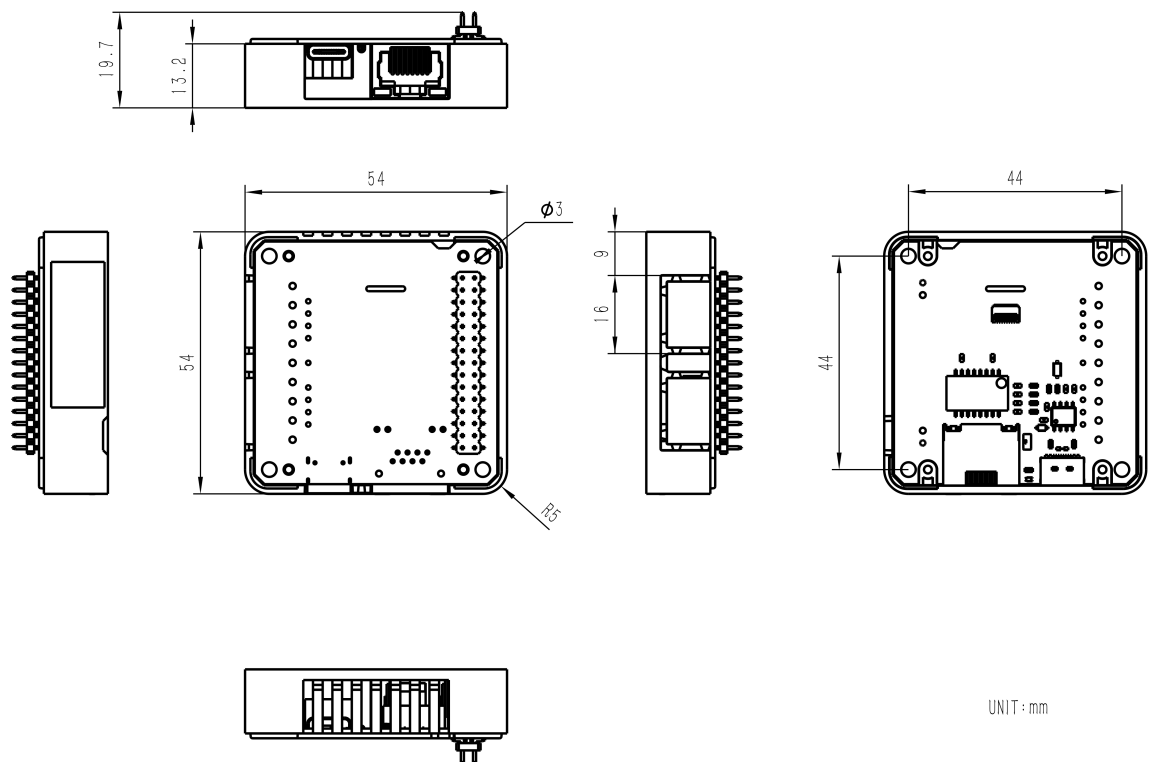
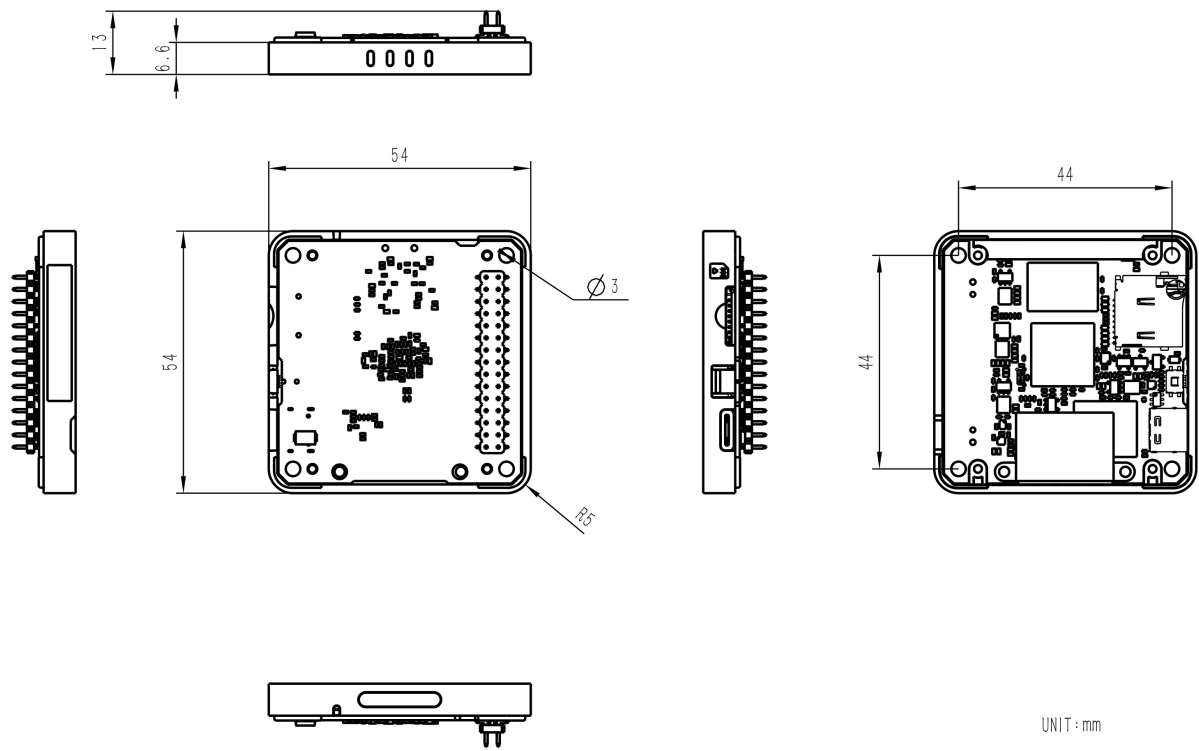
PIN	LEFT	RIGHT	PIN
GND	1	2	TRM_TXD(NT)
GND	3	4	
GND	5	6	RST
	7	8	
	9	10	
	11	12	3V3
	13	14	
TRM_TXD(NT)	15	16	TRM_RXD(NT)
SoC_SCL	17	18	SoC_SDA
	19	20	
TRM_RXD(NT)	21	22	TRM_TXD(NT)
TRM_RXD(NT)	23	24	TRM_RXD(NT)
	25	26	TRM_TXD(NT)
	27	28	5V
	29	30	

Module LLM Mate

PIN	LEFT	RIGHT	PIN
GND	1	2	
GND	3	4	
GND	5	6	
	7	8	
	9	10	
	11	12	
	13	14	
	15	16	
SCL	17	18	SDA
	19	20	
	21	22	
	23	24	
HPWR	25	26	
HPWR	27	28	5V
HPWR	29	30	BAT

|

Model Size



| Datasheets

- [AX630C](#)

| Softwares

| Quick Start

- [Module LLM ADB / SSH / UART Connect](#)
- [Module LLM Software Update](#)
- [Module LLM Image Update](#)
- Module LLM Status Indicator:
 - Red: Device initializing
 - Green: Device initialization complete
- Module LLM Application Upgrade Status Indicator:
 - Blue blinking: Application package updating
 - Red: Application package upgrade failed
 - Green: Application package upgrade successful

Module LLM Model Replacement Notes

The models supported by the Module LLM are in a special format unique to AXERA and require special processing to be used normally. Therefore, existing models on the market cannot be used directly.

| Arduino

- [Module LLM Arduino Quick Start](#)
- [Module LLM Arduino Library](#)
- [Module LLM Arduino Library API](#)

| UiFlow2

- [Module LLM UiFlow2 Quick Start](#)
- [Module LLM UiFlow2 API](#)

| OpenAI API

- [Module LLM OpenAI API tutorial](#)

| Development Framework

- [StackFlow](#)

| Development Resources

- [AX630C Databrief](#)
- [Module LLM JSON API Documentation](#)

- [Module LLM AX630C API Docs](#)
- [Module LLM AX620E MSP/Sample](#)
- [Module LLM Linux Kernel 4.19.125-head](#)
- [AXERA LLM Example](#)
- [AXERA CV Example](#)
- [Module LLM Large Model Compilation Guide](#)

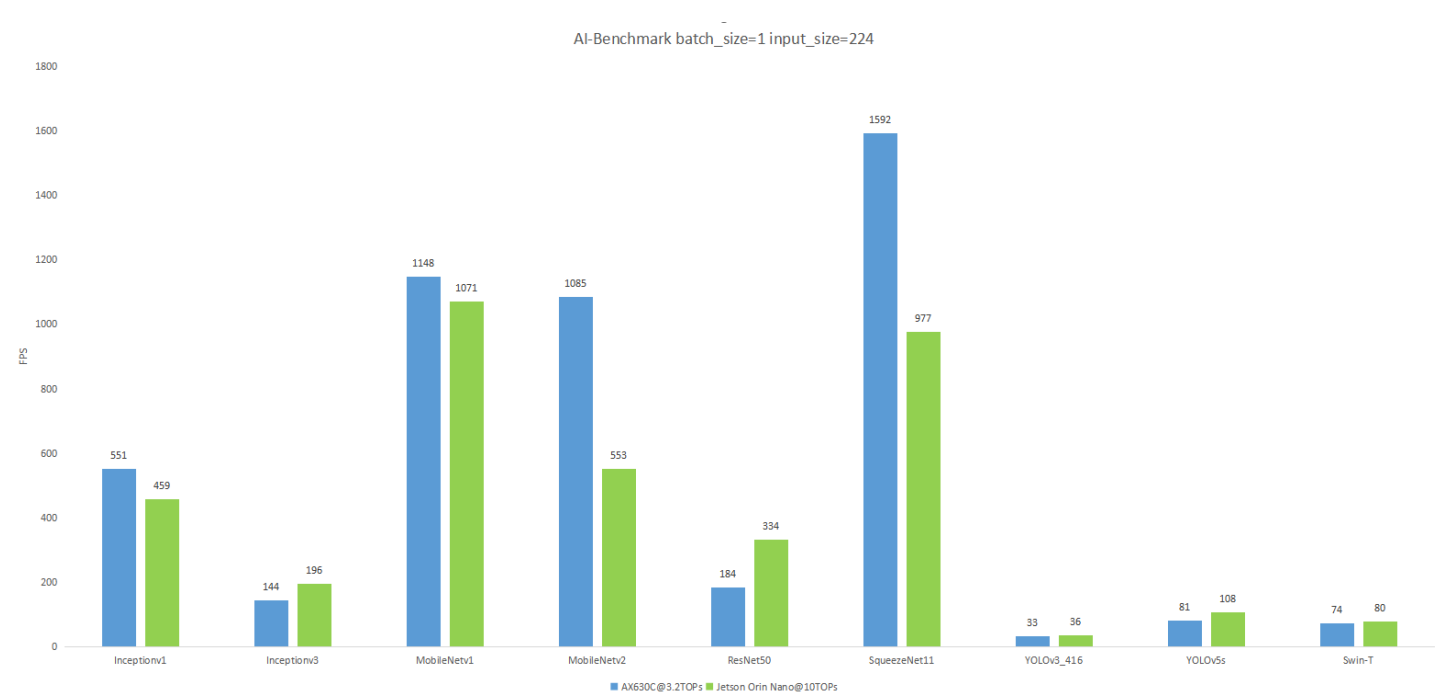
Video

- Module LLM product introduction and demonstration

[Module_LLM_Video.mp4](#)

Product Comparison

AI Benchmark Comparison



Mouser Electronics

Authorized Distributor

Click to View Pricing, Inventory, Delivery & Lifecycle Information:

[M5Stack:](#)

[K144](#)