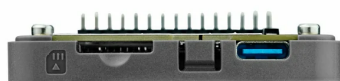
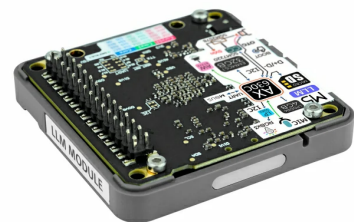
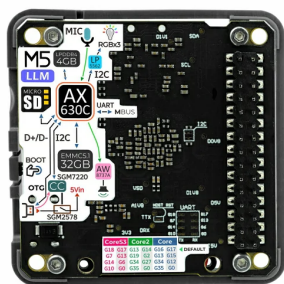


Module LLM

SKU:M140





描述

Module LLM 是一款集成化的离线大语言模型 (LLM) 推理模块，专为需要高效、智能交互的终端设备设计。无论是在智能家居、语音助手，还是在工业控制中，Module LLM 都能为您带来流畅、自然的 AI 体验，无需依赖云端，确保隐私安全与稳定性。

该模块集成 **StackFlow** 框架，配合 **Arduino/UiFlow** 库，几行代码就可轻松实现端侧智能。搭载爱芯 **AX630C** SoC 处理器，集成 3.2TOPs 高性能 NPUs，原生支持 Transformer 模型，轻松应对复杂 AI 任务。配备 **4GB LPDDR4** 内存（其中 1GB 供用户使用，3GB 专用于硬件加速）和 **32GB eMMC** 存储，支持多模型并行加载与串联推理，确保多任务处理流畅无阻。运行功耗仅约 1.5W，远低于同类产品，节能高效，适合长时间运行。

Module LLM 集成麦克风、扬声器、TF 存储卡、**USB OTG** 及 RGB 状态灯，满足多样化应用需求，轻松实现语音交互与数据传输。在灵活扩展方面，板载 SD 卡槽支持固件冷 / 热升级，**UART** 通信接口简化连接与调试，确保模块功能持续优化与扩展。USB 口支持主从自动切换，既可以做调试口，也可以外接更多 USB 设备如摄像头。搭配 LLM 调试套件，用于扩展百兆以太网口及内核串口，作为 SBC 使用。

此模块多模型兼容，出厂预装 **Qwen2.5-0.5B** 大语言模型，内置 **KWS**（唤醒词）、**ASR**（语音识别）、**LLM**（大语言模型）及 **TTS**（文本转语音）功能，且支持分开调用或 **pipeline** 自动流转，方便开发。后续将支持 Qwen2.5-1.5B、Llama3.2-1B 及 InternVL2-1B 等多种端侧 LLM/VLM 模型，支持热更新模型，紧跟社区潮流，适应不同复杂度的 AI 任务。在视觉识别能力上，支持 CLIP、YoloWorld 等 Open world 模型，未来将持续更新 DepthAnything、SegmentAnything 等模型，赋能智能识别与分析。

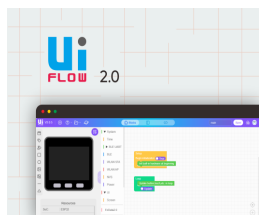
Module LLM 即插即用，搭配 **M5 主机**，实现即插即用的 AI 交互体验。用户无需繁琐设置，即可将其集成到现有智能设备中，快速启用智能化功能，提升设备智能水平。该产品适用于离线语音助手、文本语音转换、智能家居控制、互动机器人等领域。

教程 & 快速上手



Arduino IDE

本教程将向你介绍，如何通过 Arduino IDE 编程控制 Module LLM 设备，需搭配 M5Core 系列主机使用



UiFlow2

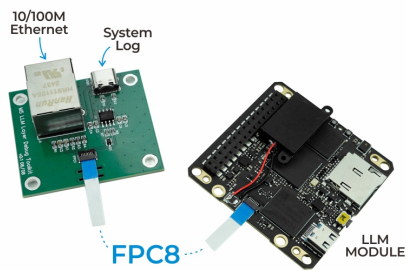
本教程将向你介绍，如何通过 UiFlow2 图形化编程平台控制 Module LLM 设备

产品特性

- 离线推理, 3.2T@INT8 精度算力
- 集成 KWS (唤醒词), ASR (语音识别), LLM (大语言模型), TTS (文本生成语音)
- 多模型并行
- 板载 32GB eMMC 存储和 4GB LPDDR4 内存
- 板载麦克风及扬声器
- 串口通信
- SD 卡固件升级
- 支持 ADB 调试
- RGB 提示灯
- 内置 ubuntu 系统
- 支持 OTG 功能
- 支持 Arduino, UiFlow

包装内容

- 1 x Module LLM



随产品附赠调试板 (仅首发限量赠送)

应用场景

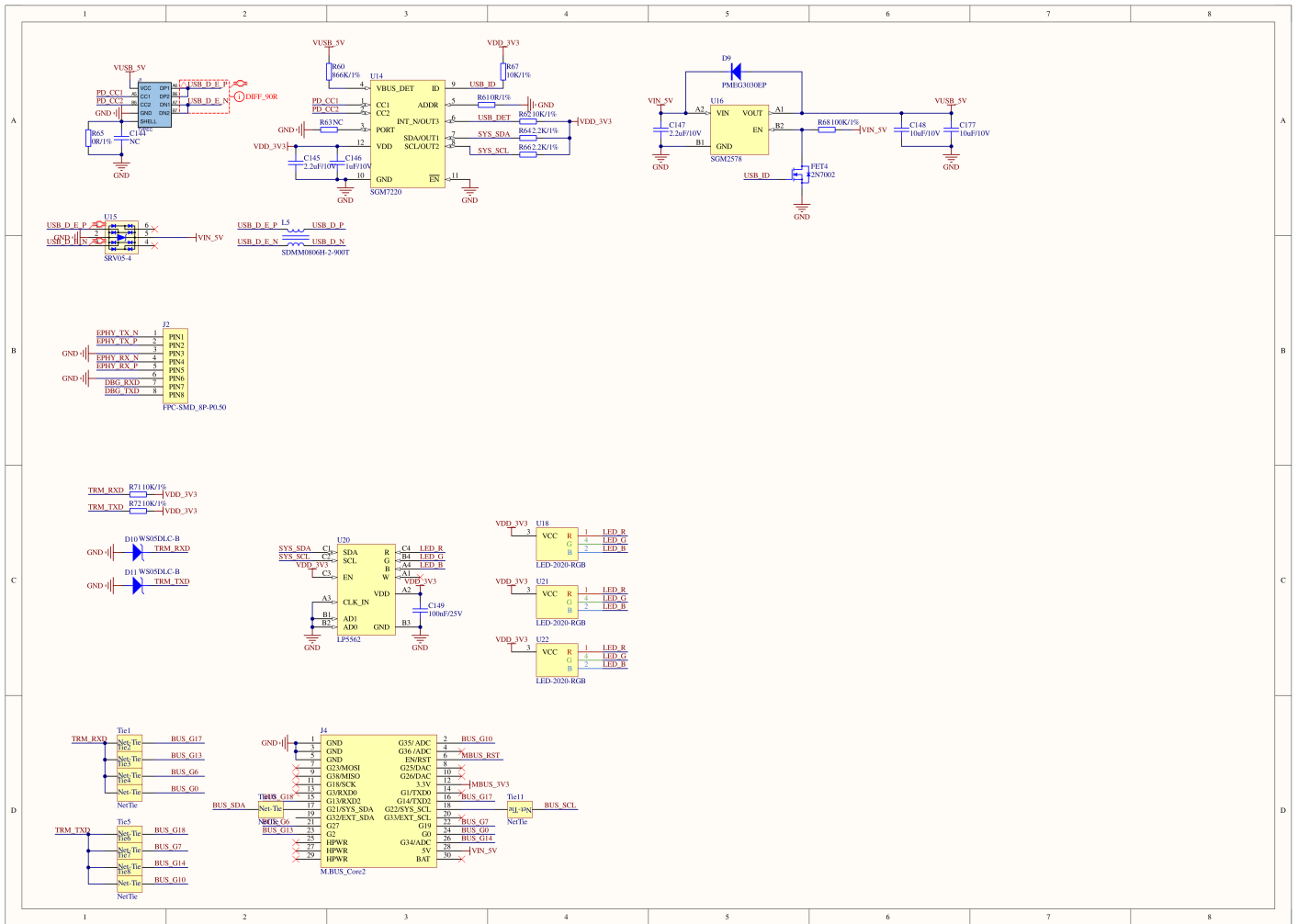
- 离线语音助手
- 文本语音转换
- 智能家居控制
- 互动机器人

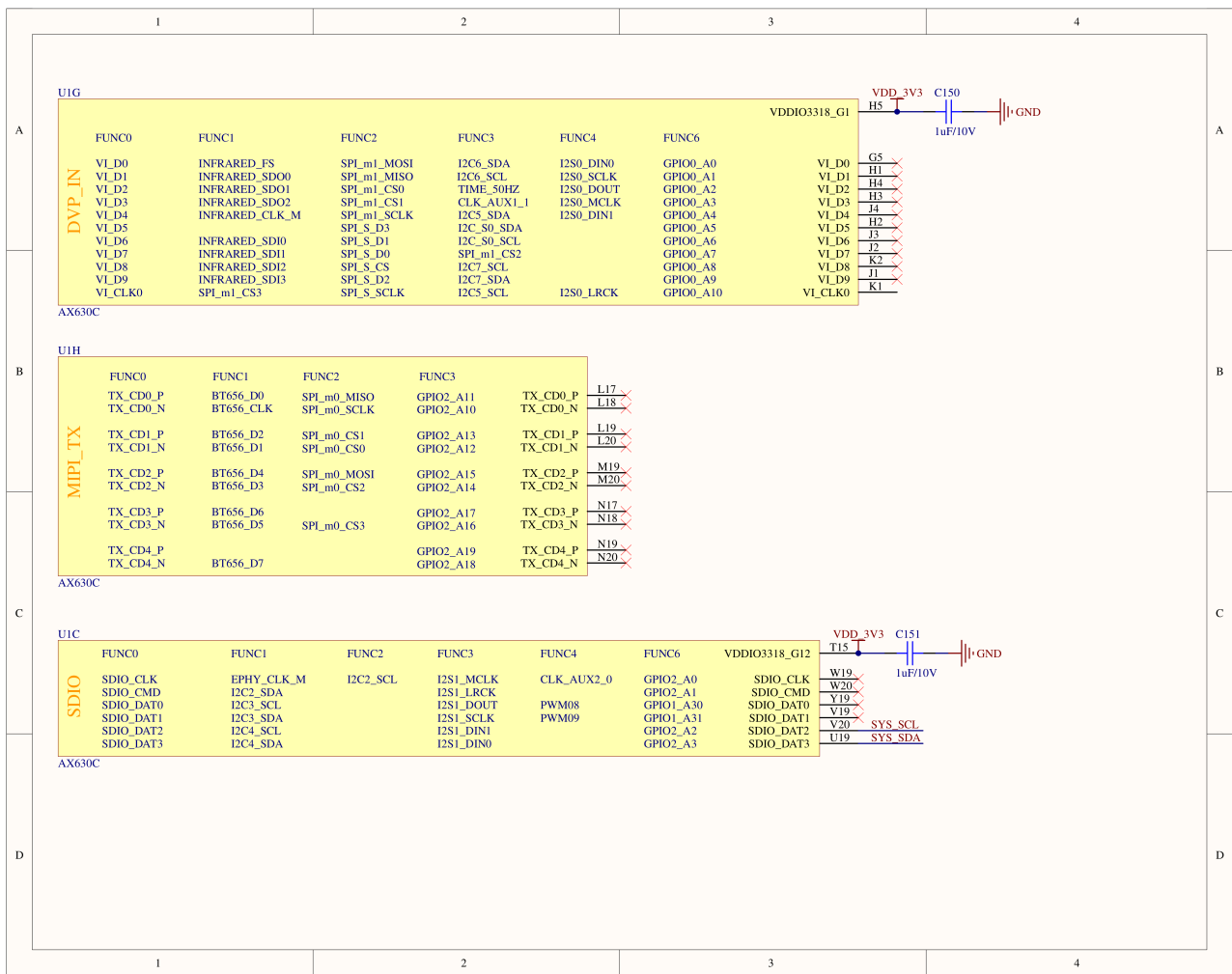
规格参数

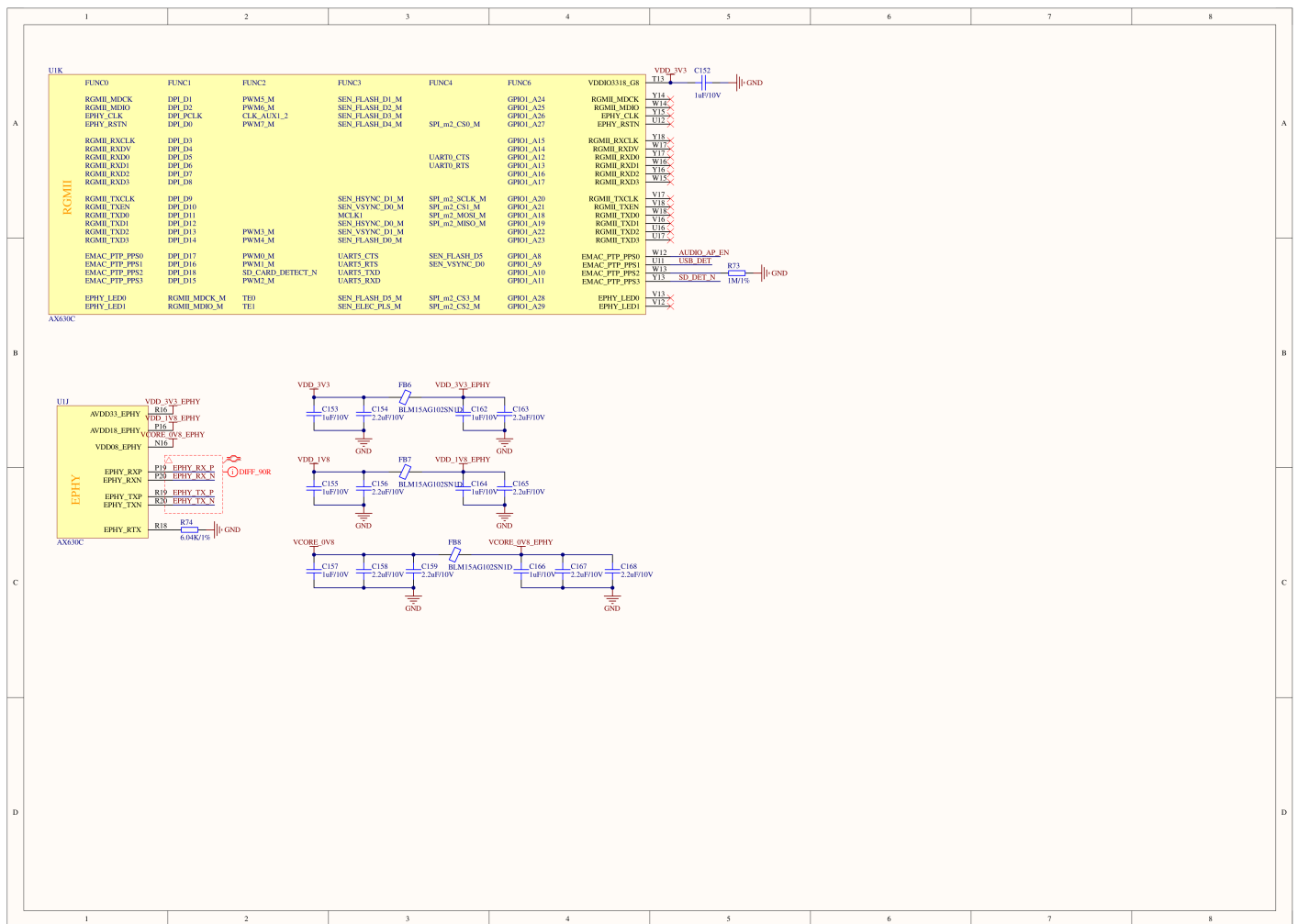
规格	参数
处理器 SoC	AX630C@Dual Cortex A53 1.2 GHz MAX.12.8 TOPS @INT4,and 3.2 TOPS @INT8
内存	4GB LPDDR4 (1GB 系统内存 + 3GB 硬件加速专用内存)
存储	32GB eMMC5.1
通信接口	串口通信 波特率默认 115200@8N1 (可调)
麦克风	MSM421A
音频驱动	AW8737
扬声器	8Ω@1W, 尺寸：2014 腔体喇叭
内置功能单元	KWS (唤醒词)，ASR (语音识别)，LLM (大语言模型)，TTS (文本生成语音)
RGB 灯	3x RGB LED@2020 由 LP5562 驱动 (状态指示)
功耗	空载：5V@0.5W，满载：5V@1.5W
按键	用于升级固件进入下载模式
升级接口	SD 卡 / Type C 口
工作温度	0~40°C
产品尺寸	54.0 x 54.0 x 13.0mm
产品重量	17.1g
包装尺寸	132.0 x 95.0 x 16.0mm
毛重	30.7g

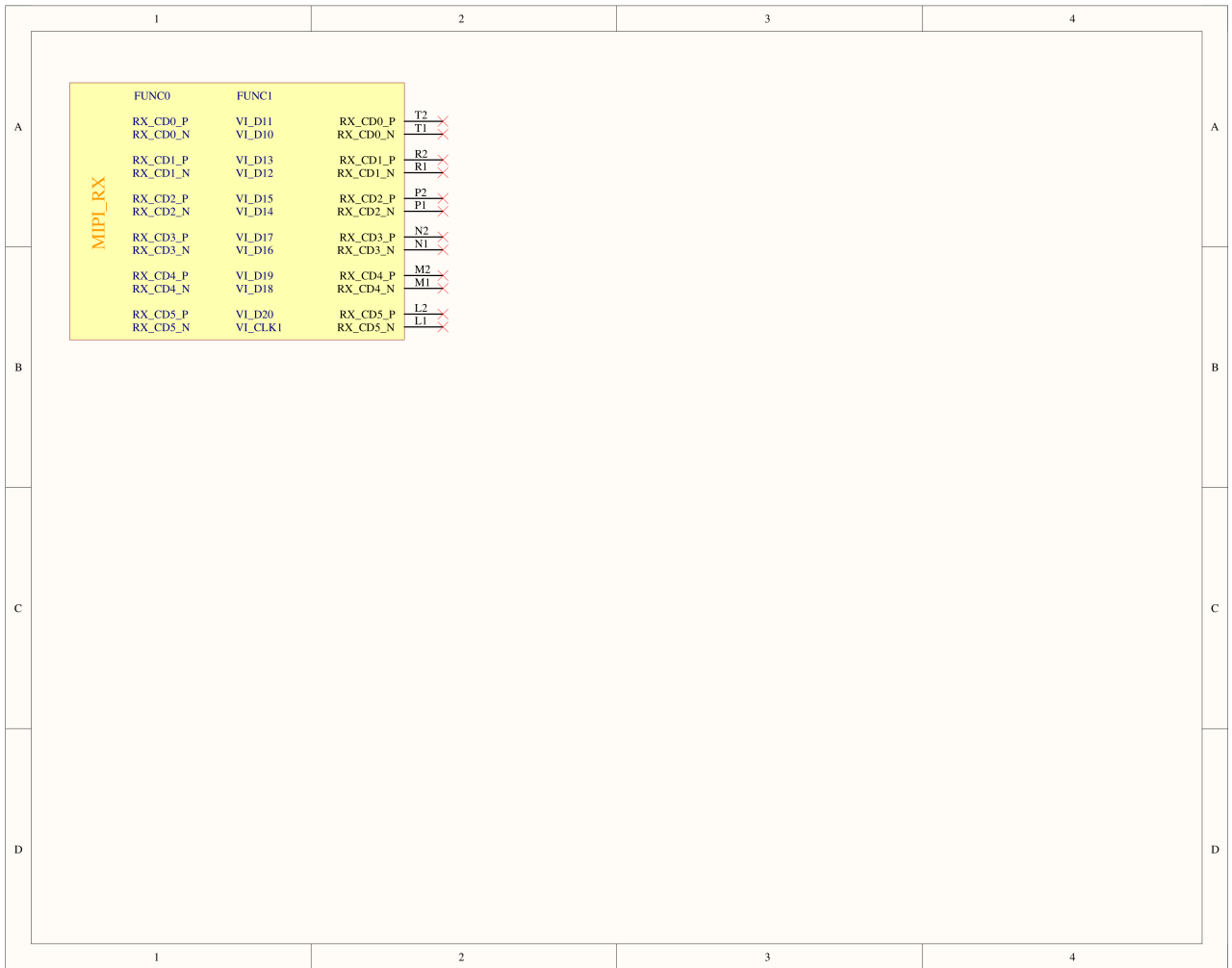
原理图

- [Module LLM 原理图下载](#)





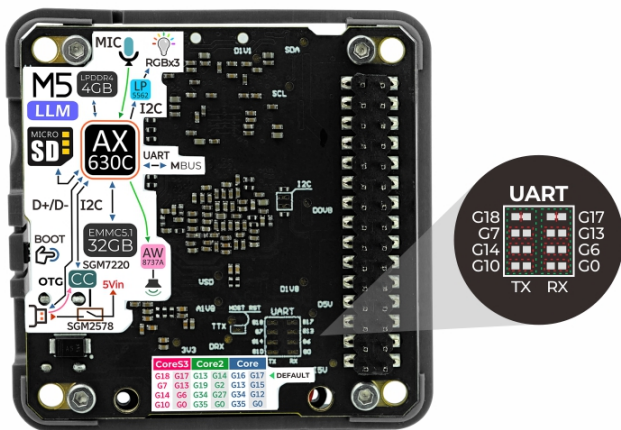




管脚映射

Module LLM 引脚切换

Module LLM 预留了引脚切换焊盘，一些出现引脚复用冲突的情况下，可以通过切割 PCB 线路然后跳线连接至其他组引脚。



Tx Rx

CoreS3	Core2	Core	
G18 G17	G13 G14	G16 G17	← DEFAULT
G7 G13	G19 G2	G13 G15	
G14 G6	G34 G27	G34 G12	
G10 G0	G35 G0	G35 G0	

以**CoreS3**为例子，第一列 (左绿色框) 是串口通讯的 TX 引脚，用户根据需要四选一，(从上到下分别代表引脚 G18 G7 G14 G10)，默认是 IO18 引脚，如需要切换其他引脚，切断焊盘连接线 (红线处)(建议使用刀片切割)，然后连接下面三个引脚之间的一个即可；第二列 (右绿色框) 是 RX 引脚切换，如上所述，也是四选一操作。

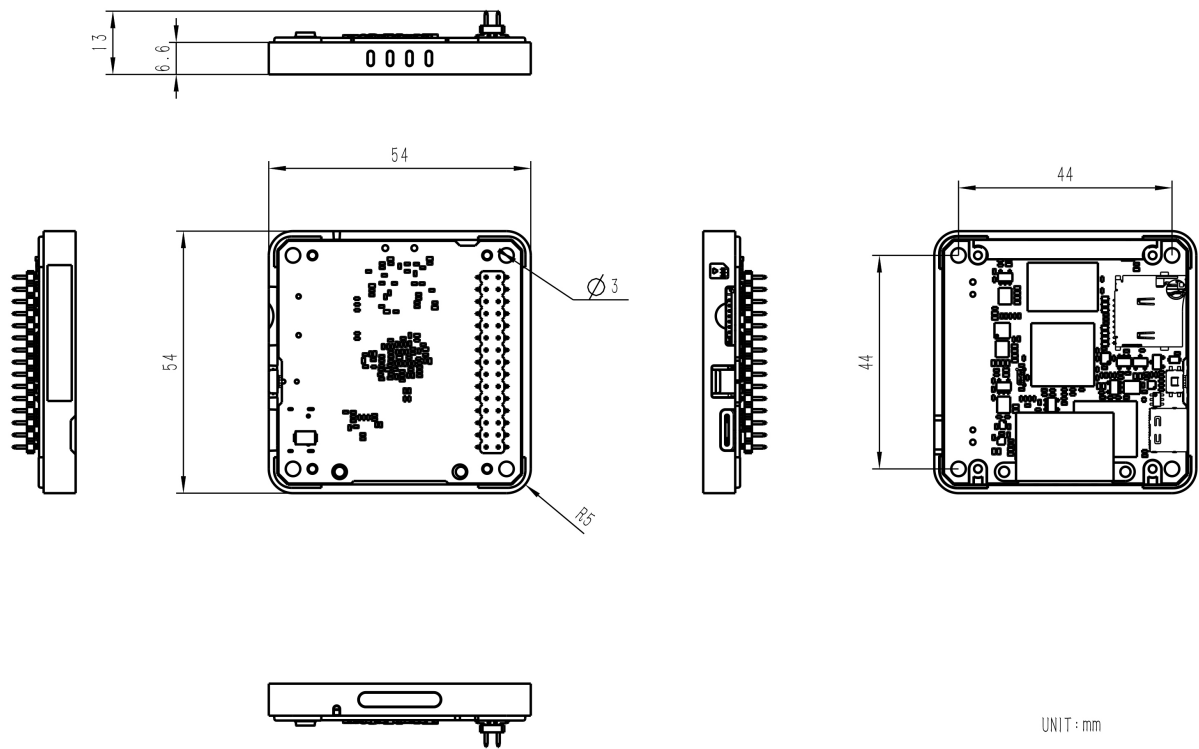
M5-Bus

Net-Tie

下方 M5-Bus 中标记 **NT** 的引脚，可通过上述操作进行切换，用于适配不同的主控设备。

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_SDA	17	18	SoC_SCL
	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	

尺寸图



数据手册

- [AX630C](#)

软件开发

快速上手

- [Module LLM ADB / UART / SSH 连接调试](#)
- [Module LLM 软件包应用更新](#)
- [Module LLM 镜像底包更新](#)
- Module LLM 工作状态指示灯:
 - 红色: 设备正在初始化
 - 绿色: 设备初始化完成
- Module LLM 应用升级状态指示灯:
 - 蓝灯闪烁: 应用包更新中
 - 红色: 应用包升级失败
 - 绿色: 应用包升级成功

LLM Module 支持的模型为爱芯特有格式，需经过特殊处理才可正常使用。因此并不能直接使用市面现有模型。

Arduino

- [Module LLM Arduino 快速上手](#)
- [Module LLM Arduino 驱动库](#)
- [Module LLM Arduino API](#)

UiFlow1

- [Module LLM UiFlow1 文档](#)

UiFlow2

- [Module LLM UiFlow2 快速上手](#)
- [Module LLM UiFlow2 文档](#)
- [Module LLM UiFlow2 API](#)

OpenAI API

- [Module LLM Kit OpenAI API 使用指南](#)

开发框架

- [StackFlow](#)

开发资料

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

相关视频

- [Module LLM 产品介绍以及例子展示](#)

[Module_LLM_Video.mp4](#)

产品对比

AI Benchmark 对比

AI-Benchmark batch_size=1 input_size=224

