RK3588 NPU运行大模型项目：

[ttps://github.com/Pelochus/ezrknn-llm](https://github.com/Pelochus/ezrknn-llm)

NPU运行视频：

<https://www.youtube.com/watch?v=KB9qRwj1pnU&t=4s>

Radxa rock 5手册：

<https://docs.radxa.com/rock5/rock5b/getting-started/install-os/boot_from_sd_card?Platform=Windows>

Llama2 huggingface Transformer格式模型

<https://huggingface.co/meta-llama/Llama-2-7b-chat-hf>

Tinyllama

<https://huggingface.co/TinyLlama/TinyLlama-1.1B-Chat-v1.0/tree/main>

Llama2 转换后rkllm格式模型

<https://huggingface.co/Pelochus/llama2-chat-7b-hf-rk3588>

gcc交叉编译工具

<https://developer.arm.com/downloads/-/gnu-a/10-2-2020-11>

linux 动态链接库文件 so解析

[Linux的so文件到底是干嘛的？浅析Linux的动态链接库](https://www.zhihu.com/tardis/zm/art/235551437?source_id=1005)

window上交叉编译所需工具

[GCC - 基于win10平台搭建Cmake + MinGW + gcc-arm-none 开源开发环境\_gcc-arm-none-eabi-10.3-2021.10-win32.exe-CSDN博客](https://blog.csdn.net/mido94/article/details/128144684)

Openai-api

<https://platform.openai.com/docs/api-reference/chat/create>

llm数据集

<https://github.com/leobeeson/llm_benchmarks>

llama.cpp

<https://github.com/ggerganov/llama.cpp>

whisper

<https://github.com/ggerganov/whisper.cpp>

whisperhuggingface

<https://huggingface.co/models?sort=downloads&search=openai%2Fwhisper->

监控rknputop

<https://github.com/ramonbroox/rknputop>

piper

<https://github.com/rhasspy/piper/tree/master>

piper huggingface

<https://huggingface.co/rhasspy/piper-voices/tree/v1.0.0/en/en_GB/southern_english_female/low>

github gradio项目

<https://github.com/JakeFurtaw/Chat-RAG?tab=readme-ov-file>

gradio语音项目

<https://huggingface.co/learn/audio-course/zh-CN/chapter0/introduction>

llama.cpp支持的模型对话模板：

<https://github.com/ggerganov/llama.cpp/wiki/Templates-supported-by-llama_chat_apply_template>

llama2 huggingface

<https://huggingface.co/TheBloke/Llama-2-7B-Chat-GGUF>