参考文章：[BERT改进版对比：RoBERTa、DistilBERT、ALBERT、ELECTRA 以及 DeBERTa\_roberta的升级版-CSDN博客](https://blog.csdn.net/qq_36803941/article/details/144158695)

本地中文nlp：

**DistilBERT** 或 **ALBERT**

[ChatLM-mini-Chinese - 小参数中文对话模型，支持低显存预训练，优化SFT和DPO性能 - 懂AI (dongaigc.com)](https://www.dongaigc.com/p/charent/ChatLM-mini-Chinese)

[**~~DeepSeek-R1 蒸馏~~**~~模型~~](https://blog.csdn.net/llm_way/article/details/145365802)**~~要求显存8GB，模型大小4.9GB~~**

**云端推理：deepseek**

**图像：EdgeYOLO**