

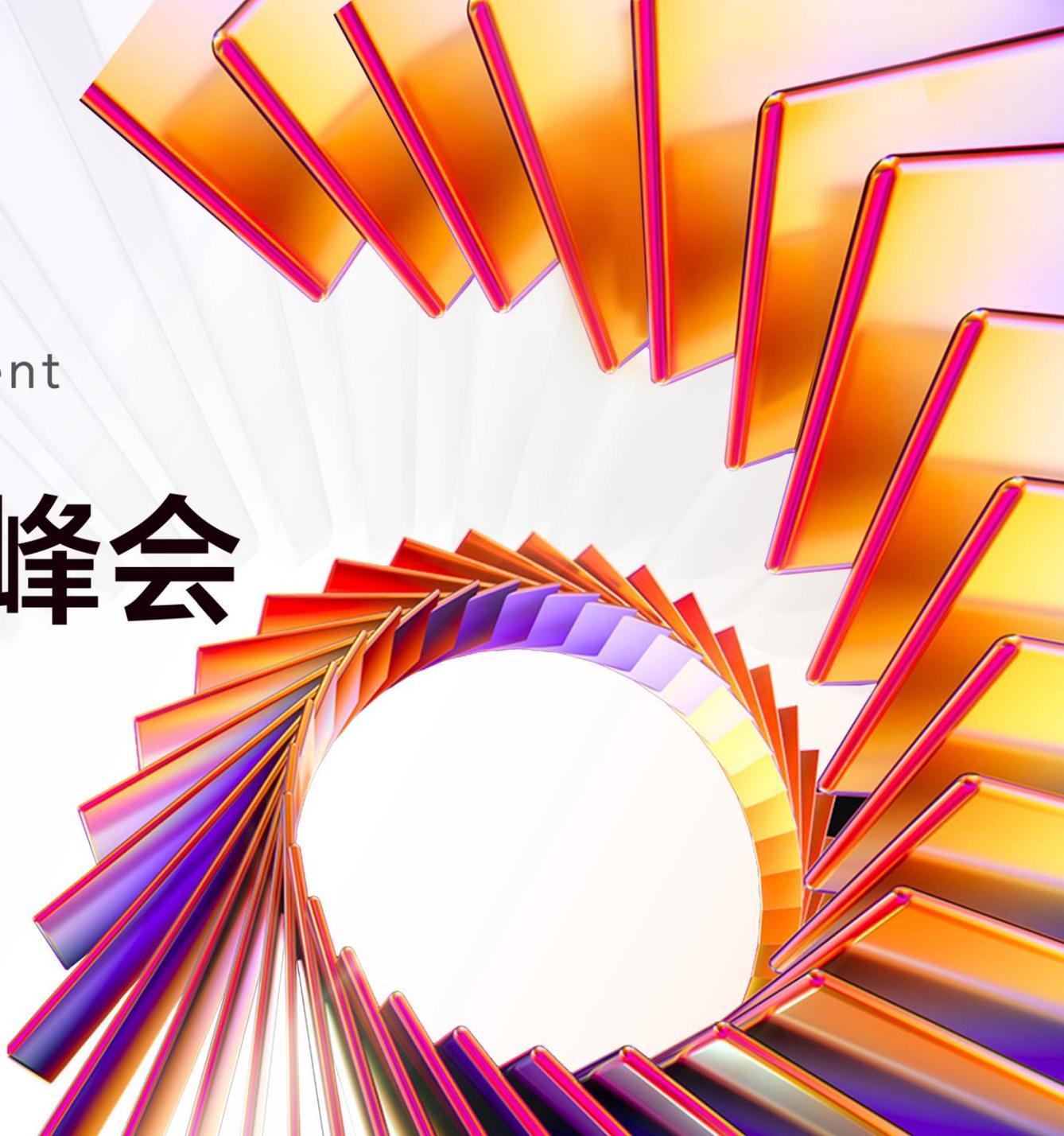


第7届 AI+ Development
Digital Summit

AI+ 研发数字峰会

拥抱AI 重塑研发

8月8-9日 | 北京站





第8届 AI+ 研发数字峰会

拥抱 AI 重塑研发 AI+ Development Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

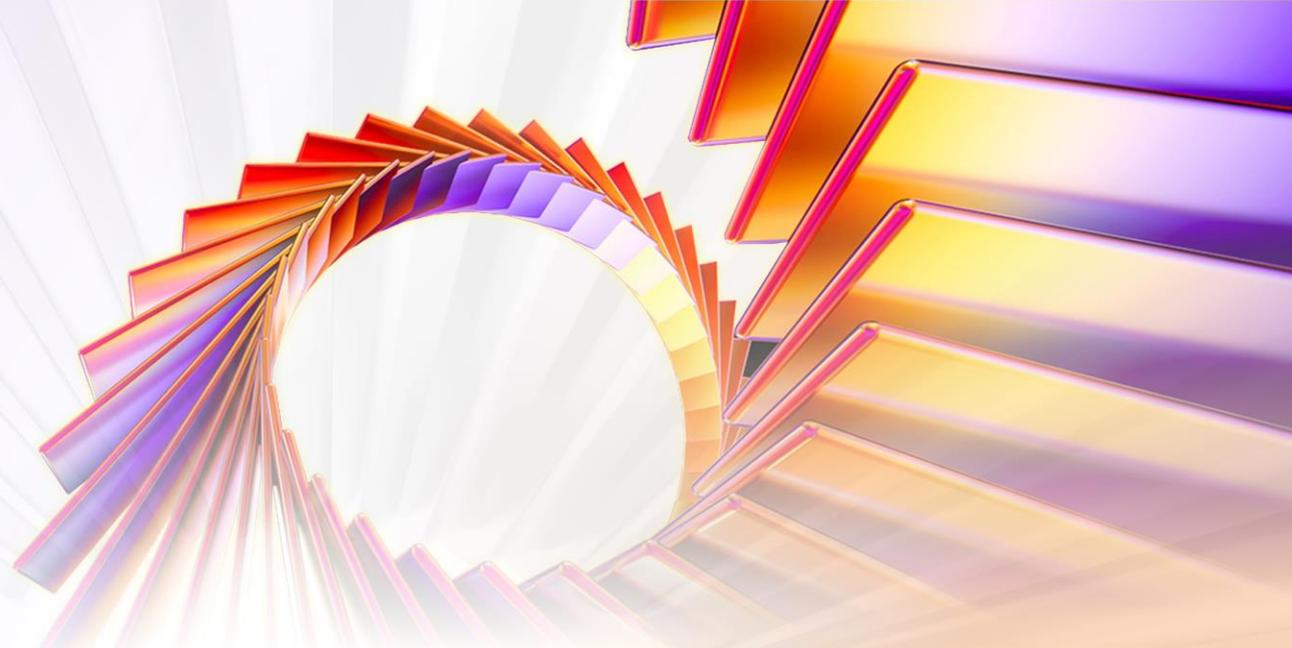
上下文工程

AiDD 7th | 8月8-9日 | 北京站
2025

第7届 AI+ Development
Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发



大模型Agent在智能硬件上的实践与应用

——大模型重塑智能硬件交互体验

王松 | 未来智能



王松

未来智能CTO

连续创业者，个人专注于软硬件技术、研发管理、前沿技术，曾经研发全国首个智能短途配送系统，利用AI提效50%以上，曾带领团队做出世界上第一台支持全场景录音的智能耳机，非常擅长将新技术转化为生产力。

目前团队产品是国内智能耳机的高端品牌Viim耳机，具备多种AI功能，与大模型无缝集成，团队正在打造下一代智能穿戴设备，希望通过AI让用户的生活和工作更加轻松和便捷。

目录

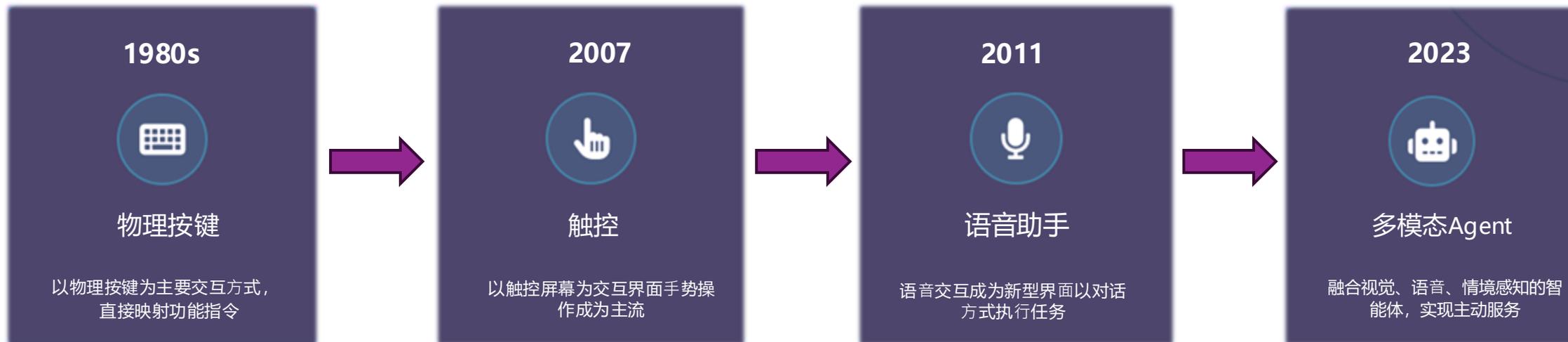
CONTENTS

- I. 交互演进史
- II. 软硬一体技术架构
- III. 三大体验重构
- IV. 落地案例

PART I

交互演进史

交互演进史



智能硬件痛点速描



唤醒词喊三遍

"小爱同学...小爱同学! ...小爱同学!!!"



跨APP复制粘贴

切换三个应用才能完成一项任务



场景断点

无法记住上下文，每次都要从头开始



沉默尴尬

"对不起，我没听清..."



唤醒词识别率低

用户平均需要重复唤醒词2.5次才能激活设备



应用割裂

跨应用任务需要用户手动串联，增加操作复杂度



上下文记忆缺失

交互中断即上下文丢失，无法延续之前的对话



交互机械僵硬

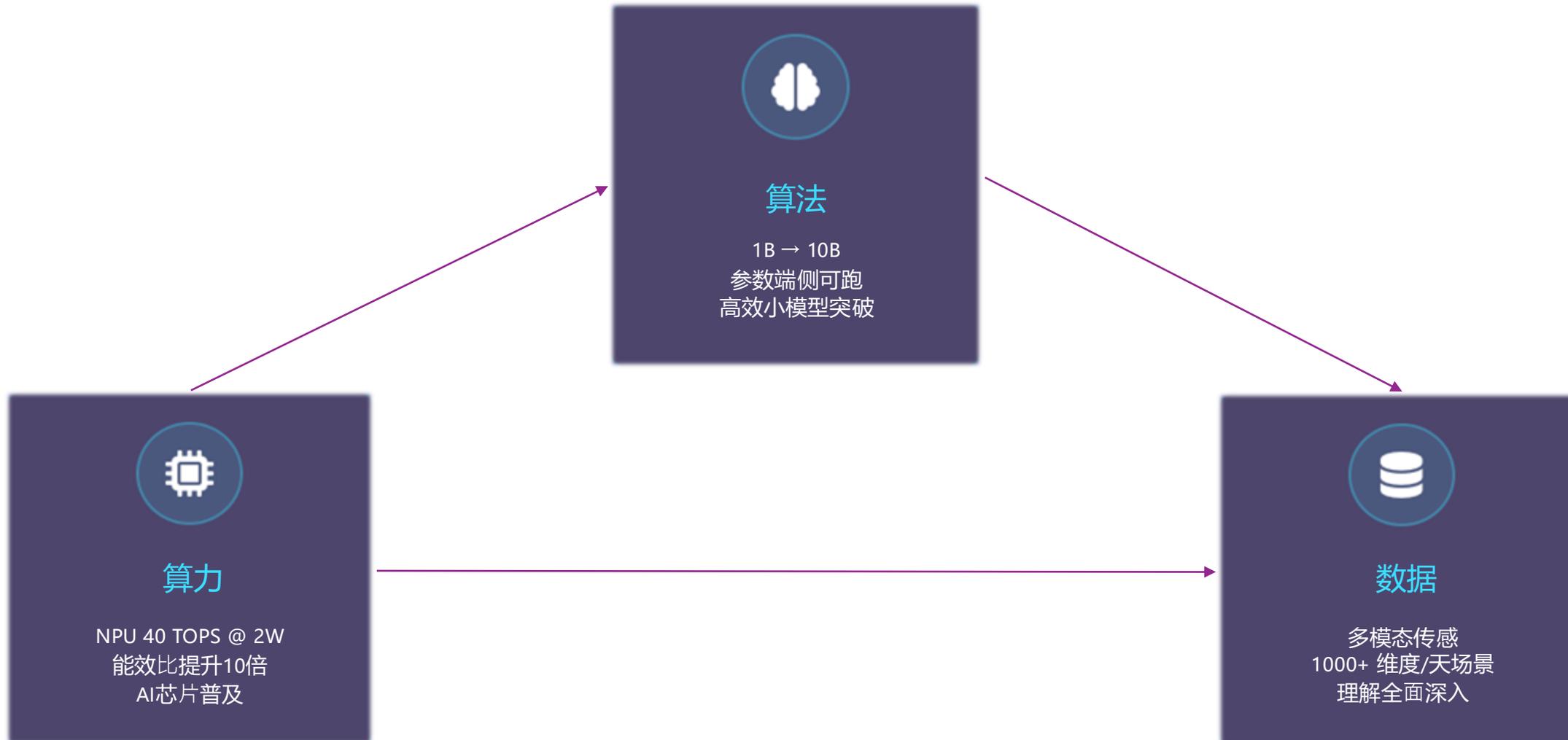
对话体验机械，缺乏自然情感和人性化反馈



PART II

软硬一体技术架构

为什么是现在?



▶▶ 端侧模型压缩黑科技



4bit 量化

将模型权重从 FP16/FP32 压缩到 4bit, 减少 4-8 倍存储

模型体积减少 **75%**



60% 稀疏化

去除冗余连接和权重, 保留结构化稀疏, 加速推理

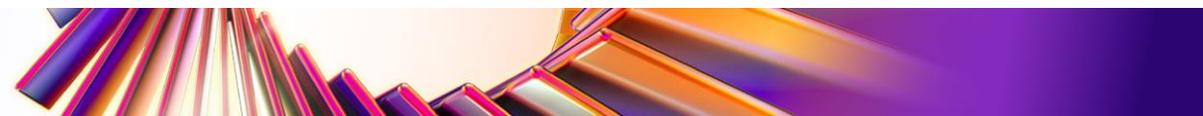
计算量减少 **60%**



KV-Cache 重计算

优化注意力机制中的键值缓存, 减少内存占用

内存占用减少 **40%**

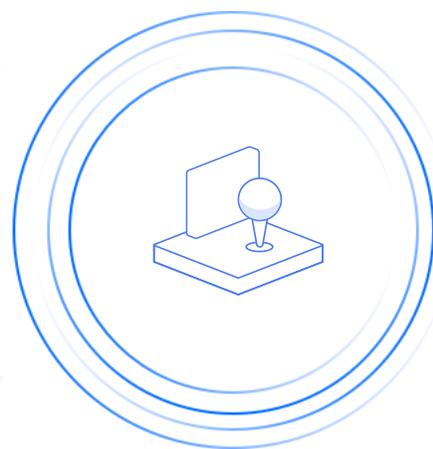


应用层 (Application Layer)

面向用户的智能应用、服务、场景适配和多模态交互界面

模型层 (Model Layer)

大语言模型、多模态感知模型、知识图谱、向量数据库



框架层 (Framework Layer)

智能Agent编排、推理框架、API调用编排、任务管理与调度

芯片层 (Chip Layer)

NPU、专用AI加速器、异构计算硬件、底层驱动优化



▶ 记忆系统突破



用户记忆云

云端存储全局长期记忆，支持跨设备访问



本地缓存

设备存储短期记忆，保护隐私不出设备



可撤销记忆

用户对历史记忆的查看、修改与删除权限

关键技术：

向量数据库

联邦学习

差分隐私



PART III

三大体验重构

▶▶ 三大体验重构：交互范式跃迁



传统语音助手

指令式交互

用户需要明确表达具体的指令和操作步骤，助手被动执行单一任务。

用户输入:

"导航去公司"

助手响应:

"好的，正在为您导航去公司"



单一任务



被动执行

VS



Agent

目标式交互

用户只需表达高层次目标，Agent理解意图并主动规划多个任务步骤。

用户输入:

"我今天想9点到公司，顺便买咖啡"

Agent响应:

"已为您规划路线，考虑到当前交通状况，建议8:25出发。已标记路线上星巴克门店，预计路上耗时30分钟，已添加8:55提醒取咖啡。"



多任务编排



主动服务



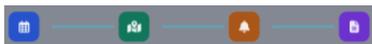
▶▶ 三大体验重构：服务闭环

结果反馈

- 已为您完成全部安排：
- 明天15:00的会议已添加到日历
 - 建议14:30出发，预计路程30分钟
 - 已设置14:00提醒
 - 演讲稿模板已创建并添加到备忘录

工具调用

```
calendar.createEvent(title="重要会议",  
startTime="2025-07-21T15:00:00", location="会议室A")  
maps.getRoute(destination="会议室A", departureTime="2025-07-21T14:30:00")  
reminder.set(time="2025-07-21T14:00:00", message="准备参加会议")
```



语音指令

"明天下午3点我有个重要会议，需要准备演讲稿，麻烦提前安排交通并提醒我"

意图解析

日程：会议(明天,15:00)
任务：准备演讲稿(优先级：高)
出行：安排交通(目的地：会议地点)
提醒：设置(时间：会前适当)

任务编排

任务1：创建日历事件
任务2：查询交通路线
任务3：设置会前提醒



▶ 三大体验重构：情感陪伴

智能设备不只是信息显示工具，更是用户的情感观察者与陪伴者，通过多模态传感器实时感知用户情绪状态，提供贴心的情感支持。

-  情绪实时感知
-  对话式情绪干预
-  自适应场景响应

情绪识别准确率
92%

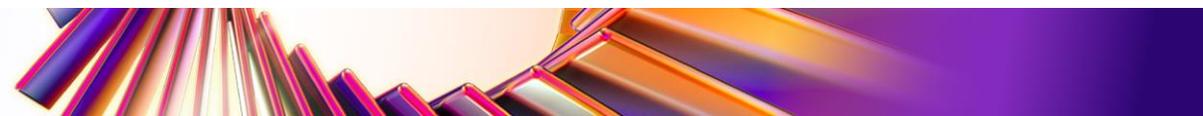
对话续聊率提升
+65%

! 情绪提醒

您今天第3次叹气，需要2分钟呼吸训练吗？

稍后再说

立即开始



PART IV

落地案例

▶ 案例一：新一代 AI Pin

最新一代 AI Pin



前置摄像头



智能可穿戴设备



激光投影单元



无屏交互



突破传统屏幕限制，通过自然语言交互、手势识别和多模态感知，实现无需屏幕的人机交互范式，体积更小，交互更自然。

全向拾音 视觉识别 空间定位

激光投影 + 手势



采用先进微型激光投影技术，将界面投射到任何表面，通过精确手势识别系统，实现直观操控，带来“随处是屏幕”的未来体验。

全空间投射 毫米级精度 自适应亮度

端侧7B模型，续航8h



内置突破性的7B参数大模型，完全端侧运行，无需云端连接，保证隐私安全。创新能效设计使其在全天候高负载场景下仍能维持8小时续航。

隐私保护 2W功耗 8小时续航 NPU加速



▶ 案例二：智能耳机 Pro

智能耳机 Pro



多语言翻译



空间音频系统

内置AI处理单元

实时翻译



采用端侧多语言翻译模型，支持40+语言的实时对话翻译，延迟低至500毫秒，语音识别精度达98%，让国际交流无障碍。

⚡ 延迟 < 500ms

🌐 40+语种

✓ 98%准确率

会议纪要



内置多说话人区分算法，自动生成会议摘要和关键点提取，支持实时查看和语音提问，纪要准确率高达96%，让会议效率提升50%以上。

👥 多人识别

☰ 自动分类

🎯 96%准确率

环境降噪自适应



基于AI算法的环境识别与自适应降噪系统，可识别多种不同场景噪声，提供个性化音频体验，降噪效果比传统ANC提升65%。

📍 场景感知

☰ 自动调节

🔊 -40dB降噪

🔋 10小时续航



▶ 开发者生态机会



插件市场

提供丰富的功能拓展模块，一键安装使用



场景脚本YAML

简洁的配置语言，快速定制个性化智能场景



端侧推理SDK

低延迟高性能的设备端AI能力，支持多种硬件平台

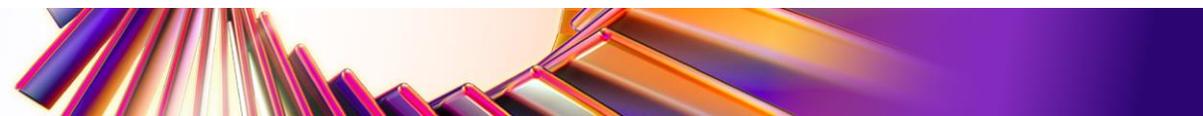


收益分成

开发者获得70%收入分成，激励创新应用开发

完整开发者生态，助力合作伙伴共赢
Agent智能时代

低门槛接入，快速实现商业化
AI能力开放，无需深度学习专家
庞大用户基础，直达目标人群
合规沙箱环境，安全无忧创新



本地优先

敏感数据优先在设备本地处理，最小化数据传输

国密算法

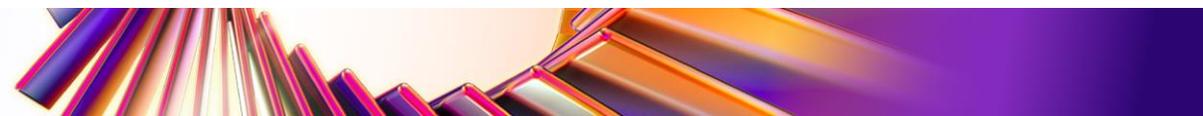
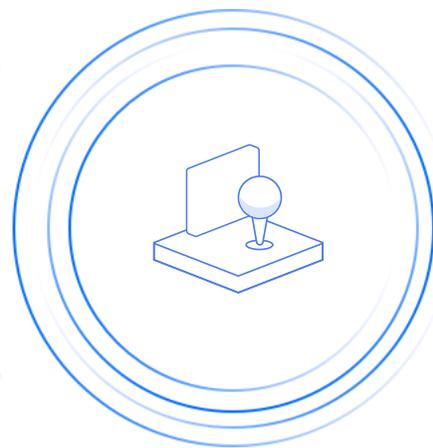
采用符合国家标准的密码算法，保障数据安全

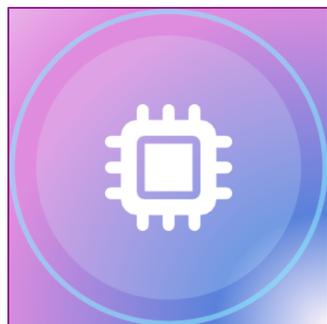
可删除记忆

用户拥有完全控制权，可随时撤销和删除历史记录

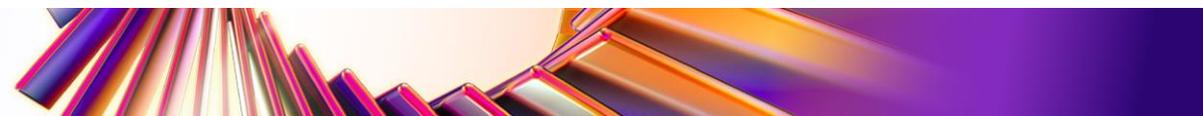
合规沙箱

所有功能在安全沙箱中运行，通过合规性审核





“ 未来的硬件，
不是装了一个大模型，
而是住进了一个 Agent。 ”





第8届 AI+ 研发数字峰会

拥抱 AI 重塑研发 AI+ Development Digital Summit

下一站预告

11/14-15 | 深圳站

12/19-20 | 上海站



查看会议详情

深圳站论坛设置

智能装备与机器人

超越“编程 Copilot”

下一代知识工程

智能网联与汽车智能化

AI 测试工具开发与应用

AI 基础设施和运维

数据智能及其行业应用

可信 AI 安全工程

大模型和 AI 应用评测

多 Agent 协同框架

从智能测试到自主测试

大模型推理优化

多模态 LLM 训练与应用

智能化 DevOps 流水线

上下文工程

AiDD

「深行 · 浅智」

Walk Deep, Think Light.

2025.11.16

AiDD首届麦理浩径徒步





科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



AiDD峰会详情





第7届 AI+ 研发数字峰会
AI+ Development Digital Summit

感谢聆听!

扫码领取会议PPT资料

