

# 领驭科技AI语音服务解决方案

专业AI语音服务方案提供商

V2025-08

深圳领驭科技有限公司

Shenzhen LEINWIN Technology Co., Ltd

**LEINWIN**  
**领驭科技**

数字智造 领驭未来

# 目录

## CONTENTS



1

领驭科技公司介绍

2

AI语音解决方案

3

案例分享

## 领驭科技

人工智能与智能制造领域先锋企业  
专注于为企业提供前沿的数字化解决方案

**90%**  
管理者工业软件  
行业经验超15年

**70%**  
技术研发  
人员占比

“专精特新”企业

CS1/CMMI3级认证

科创型中小企业

ISO9001/27001认证

50+ 软件著作权/发明专利

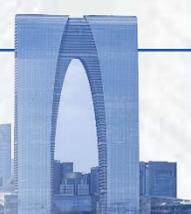
300+ AI项目案例



### 深圳市瀚鹏 人工智能科技有限公司

聚焦工业AI的深入应用与行业落地

在人工智能方向，领驭科技自主研发了瀚鹏AI应用解决方案，融合多模态大模型、知识库、智能体等核心能力，广泛应用于快销品、高科技电子、智能装备等行业



### 苏州领驭 信息技术有限公司

专注于半导体行业，为头部制造企业提供本地化支持与交付能力

在智能制造方向，公司以工业软件为核心，构建涵盖数字化工厂、系统集成、工业数据分析与AI算法的综合解决方案，已服务面板半导体、3C电子、汽车装备等多个领域



### 新加坡&香港 分公司

响应客户全球化布局

具备东南亚区域项目实施与交付能力，积极推动中国企业出海，用数字化能力助力全球制造。

# 发展历程

LEINWIN

公司使命：以人工智能与数字智造加速全球企业生产力升级

核心价值观：创新 高效 专业 分享

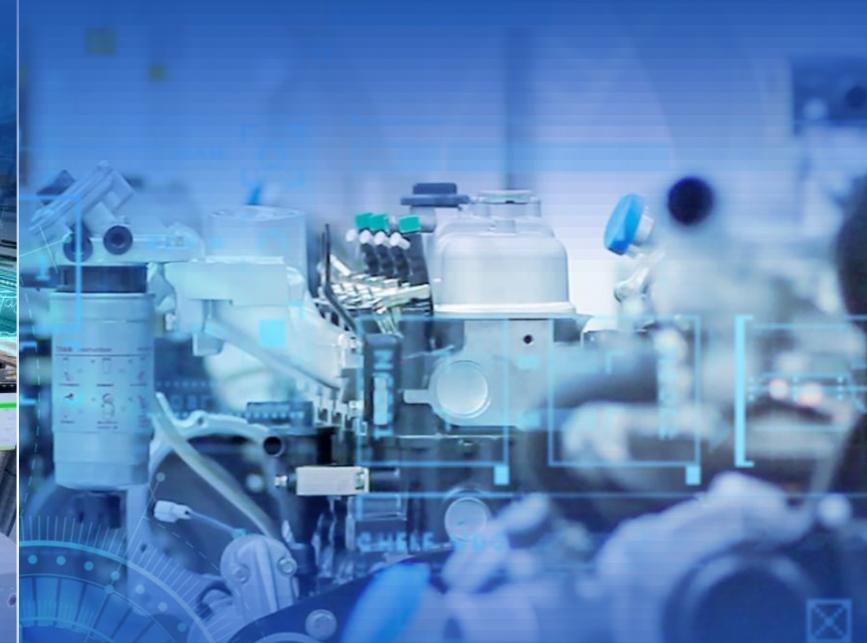
## 领跑1.0 IT信息化集成服务商

- 专注于MOM和IT信息化集成，奠定行业基础
- 西门子工业软件合作伙伴，深化数字化工厂解决方案



## 领跑2.0 2020年-2024年 数字化转型服务商

- 扎根智能制造，开启工业系统“自主研发元年”
- 深耕行业，打造行业解决方案，沉淀Know-How经验
- 连续多年获得西门子匠心服务奖
- 微软(中国)最佳出海云服务解决方案合作伙伴



## 领跑3.0 2025年 企业智能体全栈服务商

- 瀚鹏工业AI平台荣获深受企业欢迎的制造业信息化优秀产品
- 加大研发投入，华东研发子公司正式启航
- 成立海外业务中心，助力中资企业出海
- 全面布局AI，携手企业共创AI+场景

立足大湾区，辐射全球  
构建“1+2+N”全球服务体系



## 1个全球总部

深圳—全球总部

## 2个研发基地

苏州—华东研发中心

深圳鹏鼎—华南研发中心

## N个交付中心

中国香港

新加坡

泰国

马来西亚

印度尼西亚

北马其顿

# 荣誉资质

领驭已获得国家高新技术企业、专精特新企业  
知识产权管理、信息安全管理、质量管理等国家体系认证  
发明授权5项，申请中15项、软件著作权40+项  
并通过了CMMI3级认证。

# 微软相关资质

HKCSP 1T、21V NCEI、21V 2T CSP

MAICPP: Infrastructure、Data & AI、Modern Work、Digital & App  
Innovation

ASP: AI Platform on Microsoft Azure、Infra and Database Migration  
to Microsoft Azure



# 目录

## CONTENTS



1

领驭科技公司介绍

2

AI语音解决方案

3

案例分享

## 客户产品



翻译机



录音笔



普通耳机



AI眼镜



儿童玩具



蓝牙音箱



## 领驭科技解决方案

### 软件：APP赋能

|      |      |                  |
|------|------|------------------|
| 基础服务 | 实时翻译 | 出国旅游、会议翻译，海量小语种  |
|      | 同声传译 | 多向同传，平均翻译准确度超95% |
|      | AI助手 | AI快速解答疑问，提供专业建议  |
| 进阶服务 | 拍照翻译 | 手机APP或者眼镜拍照翻译    |
|      | 角色扮演 | 200+角色智能体，开箱即用   |
|      | 视频翻译 | 视频字幕和语音翻译，轻松出海   |

### 硬件：芯片集成

蓝牙方案

WIFI方案

4G方案



## 微软AI能力

SDK

API

Copilot & Grounding & Plugin

Foundation models (Speech service)

AI Infrastructure



01

## 方案一:

- 1、语音转文本 (STT)
- 2、文本翻译 (TXT translation )
- 3、文本转语音 (TTS)

02

## 方案二:

- 1、语音转文本 (STT)
- 2、大语言模型翻译 (GPT4o)
- 3、文本转语音 (TTS)

03

## 方案三:

- 1、Speech translation

Speech

04

## 方案四:

- 1、gpt4o/mini Realtime

05

## 方案五:

- 1、gpt4o/mini transcribe
- 2、大语言模型翻译 (GPT4o)
- 3、gpt-4o/mini-TTS

06

## 方案六:

- 1、gpt4o/mini transcribe
- 2、大语言模型翻译 (GPT4o)
- 3、文本转语音 (TTS)

AOAI

# 不同方案能力对比

|     | 描述   | 相对延时 | 应用场景               | 优势                   | 劣势                     |
|-----|--|------|--------------------|----------------------|------------------------|
| 方案一 | STT + TXT translation +TTS                       | 较高   | 客服录音转写<br>批量视频字幕生成 | 组合灵活, 配额高<br>支持的语种丰富 | 延时高<br>无语义优化           |
| 方案二 | STT + gpt翻译+TTS                                  | 中    | 跨语言营销文案<br>教育内容本地化 | GPT翻译效果更优            | 延时略高<br>不适合同传          |
| 方案三 | Speech translation                               | 低    | 多语言会议系统<br>旅行翻译App | 端到端语音翻译, 只需调用一个API   | 支持语种对较少                |
| 方案四 | gpt4o/mini Realtime                              | 低    | AI虚拟助手<br>实时同声传译   | 可以打断                 | 支持的声音较少Preview阶段       |
| 方案五 | gpt4o/mini transcribe + gpt 翻译 +gpt-4o/mini- TTS | 中    | 外语口语练习<br>直播角色翻译   | AI转录效果好<br>成本优势      | 配额不够生产<br>可供选择声音较少     |
| 方案六 | gpt4o/mini transcribe + gpt 翻译 +TTS              | 中    | 说书讲故事<br>游戏聊天翻译    | 声音丰富<br>AI转录效果好      | Preview阶段<br>多个API串接问题 |

# 不同方案成本对比

|                            | 方案一                                  | 方案二                 | 方案三   | 方案四   | 方案五  | 方案六  |
|----------------------------|--------------------------------------|---------------------|---|---|--|--|
| 描述                         | ASR+TextTranslation+TTS              | Speech Translation  | 4o Real Time  | 4o Real Time + TTS  | 4o mini Transcribe + 4o mini + 4o mini tts                                   | 4o mini Transcribe + 4o mini +TTS  |
| Pricing<br>(Pay as you go) | ASR: \$1/1 audio hour                | \$2.50/1 audio hour | Input:<br>\$5 /1M token (Text)<br>\$40 /1M token (Audio)  | Input:<br>\$5 /1M token (Text)<br>\$40 /1M token (Audio)  | 4o-mini-Transcribe:<br>Input: \$3/1M token (Audio);<br>Output: \$5/1M (text) | 4o-mini-Transcribe:<br>Input: \$3/1M token (Audio);<br>Output: \$5/1M (text) |
|                            | Text Translation: \$10/1M characters |                     | N/A   | Output:<br>\$20 /1M token (Text)<br>\$80/1M token (Audio) | 4o mini:<br>Input: \$0.15/1M tokens,<br>Output: \$0.60/1M Tokens             | 4o mini:<br>Input: \$0.15/1M tokens,<br>Output: \$0.60/1M Tokens             |
|                            | TTS: \$15/1M characters              |                     | Output:<br>\$20 /1M token (Text)<br>\$80/1M token (Audio) | TTS: \$15/1M characters                                   | 4o-mini-tts:<br>Input: \$0.6/1M (Text);<br>Output: \$12/1M (Audio)           | TTS:<br>\$15/1M characters   |
| \$ price per hr (PAYG)     | \$2.25                               | \$2.50              | \$ 1.50   | \$ 1.50   | \$0.27   | \$0.86   |

PS: 按1小时用量预估: 1小时5万个字符, 或1小时12500个Token

| 服务类型       | 功能         | 使用场景  |
|------------|------------|---|
| 语音转文本      | 实时语音转文本API | <ul style="list-style-type: none"><li>• 实时会议的听录、描述文字或字幕</li><li>• 说话人分离</li><li>• 发音评估</li><li>• 联系中心代理助手</li><li>• 听写</li><li>• 语音代理</li></ul> |
|            | 快速听录 API   | <ul style="list-style-type: none"><li>• 快速音频或视频听录、字幕和编辑。</li><li>• 视频翻译</li></ul>   |
|            | 批量听录API    | <ul style="list-style-type: none"><li>• 预录制音频的听录、描述文字或字幕</li><li>• 联系中心通话后分析</li><li>• 说话人分离</li></ul>  |
| 文本转语音      | TTS        | <ul style="list-style-type: none"><li>• 文本转语音可以将输入的文本转换为类似人声的合成语音</li></ul>   |
| 语音直播 (预览版) | REST API   | <ul style="list-style-type: none"><li>• 通过将语音识别、生成 AI 和文本集成到单个统一界面中，可实现语音代理的低延迟、高质量语音与语音交互。</li></ul>   |

| 服务类型    | 功能描述   |
|---------|--|
| • 语音翻译  | • 使用语音翻译可在应用程序、工具和设备中实现实时的多语言语音翻译。实施语音转语音和语音转文本翻译时可以使用此功能                          |
| • 语言识别  | • 语言标识用于在与支持的语言列表进行比较时确定音频中说话人的语言。借助语音转文本识别或语音翻译，让语言自身识别语言。                        |
| • 说话人辨识 | • 说话人辨识提供的算法可通过说话人的独特语音特征来验证和识别说话者。说话人辨识用于回答“谁在说话？”的问题。                            |
| • 发音评估  | • 发音评估可以评估语音发音，并为说话人提供有关讲话音频准确度和流利度的反馈。通过发音评估，语言学习者可以练习、获得即时反馈并改进其发音，因此能够自信地讲话和演示。 |
| • 意向识别  | • 意向识别：将语音转文本与对话语言理解结合使用，从听录的语音中获取用户意向，并根据语音命令采取行动                                 |

# Azure OpenAI部分模型清单与能力

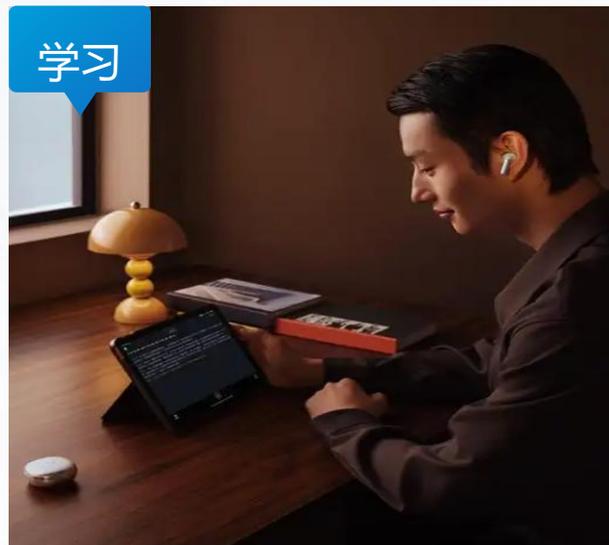
| 模型           | 主要用途             | 核心优势                           |
|--------------|------------------|--------------------------------|
| GPT-4O       | 日常任务             | 头脑风暴、总结、撰写邮件、创意内容生成            |
| GPT-4.5      | 创意类任务            | 情感理解、清晰表达、富有创意、协作式头脑风暴         |
| GPT-5        | 专业领域、日常问答        | 统一智能路由、多模态增强、持久记忆、低幻觉与高诚实度     |
| image        | 创意内容生产           | 高精度与细节理解、端到端多模态融合              |
| O3           | 复杂、多步骤任务         | 战略规划、详细分析、大规模编码、高级数学、科学推理、图像推理 |
| O1           | 深度推理、工程设计、专业科学研究 | 深度推理能力、高精度专业任务、视觉推理            |
| O4-mini      | 快速处理技术任务         | 快速解决STEM问题、编程、图像推理、支持多模态输入     |
| O4-mini-high | 深入的技术任务          | 高级编码、数学科学解释、思考更深入、准确率更高        |

# 不同使用场景方案选型



(实时翻译)  
(多语言翻译)

- ✓ 方案二
- ✓ 方案五



(录音转写) (会议助手)  
(实时翻译) (文件翻译)

- ✓ 方案一
- ✓ 方案五



(实时翻译) (多人同时翻译)  
(同声传译) (多语言翻译)

- ✓ 方案三
- ✓ 方案四

推荐方案

# 文本翻译解决方案

## Text translation 方案特点:

- **支持多语言**能够精准翻译超过100种语言的文本内容。
- **可定制翻译**可创建专属模型，以处理特定领域的专业术语。
- **适合生产环境**采用为微软产品每天提供数十亿次翻译的同一技术。
- **内置安全保障**你的数据始终归你所有，翻译过程中不会保存你的文本输入。

|   |   |
|---|---|
| From  | To  |
| 中文 (简体)   | 英语  |
| Microsoft 文本翻译是一项基于云的机器翻译服务，支持多种语言，其支持的语言覆盖全球国内生产总值 (GDP) 95% 以上的区域。使用 Translator 可构建应用程序、网站、工具或任何需要多语言支持的解决方案。 | Microsoft Text Translation is a cloud-based machine translation service that supports multiple languages and covers more than 95% of the world's gross domestic product (GDP). Use Translator to build applications, websites, tools, or any solution that requires multilingual support. |

## GPT翻译 方案特点:

- **多语言高精度**，支持上百种语言，尤其擅长低资源语言和专业术语；
- **多模态能力**，可处理文本、图像、语音输入，实现拍照翻译和实时对话；
- **自然流畅**，生成符合目标语言习惯的表达，避免机械翻译的生硬感；
- **上下文理解**，能捕捉文化隐喻、指代关系，确保逻辑连贯；
- **交互优化**，支持动态调整风格（如正式/口语化）。  
相比传统工具，GPT-4o更智能灵活，适合商务、学术、旅行等场景，但长文本处理需分段。

Azure AI Foundry | gpt-4o-mini-2024-07-18 AI-generated content may be incorrect

### Original:

GPT通过生成式预训练，能够生成自然流畅的文本，减少直译现象，尤其在非标准用法和惯用语句处理上表现更佳。例如，在处理专业术语或复杂语境时，GPT生成的翻译更符合目标语言习惯

### Translated:

Through generative pre-training, GPT is capable of producing natural and fluent text, reducing instances of literal translation, especially excelling in handling non-standard usage and idiomatic expressions. For example, when dealing with technical terms or complex contexts, the translations generated by GPT are more in line with the conventions of the target language.

## voice live语音对话解决方案

Voice Live API 整合了多项微软语音技术，形成统一的接口。

### 功能包括：

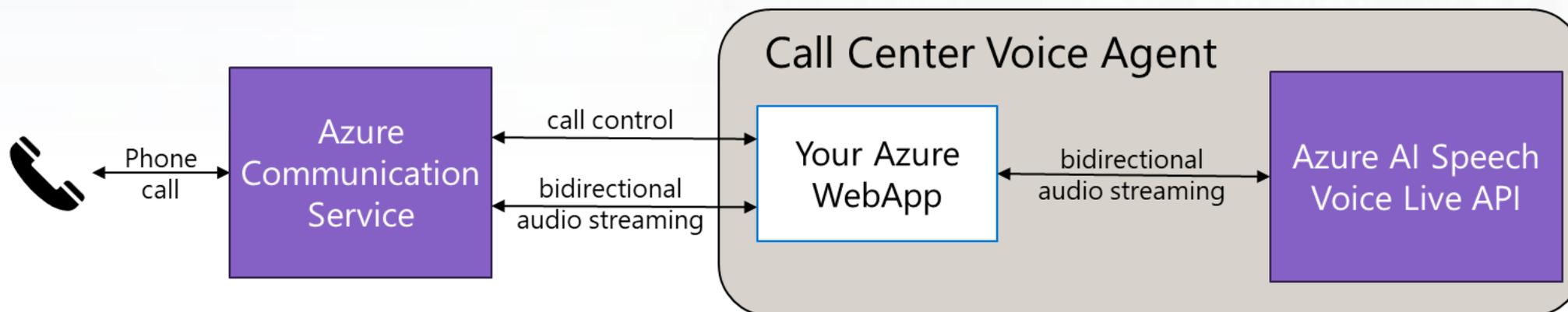
- **语音识别 (STT) 与合成 (TTS)** 支持超过 150 种语言的输入和输出，提供 600 多种逼真语音选项，包括 30 多种专为对话优化的超自然神经语音
- **生成式 AI 集成** 允许开发者选择基础模型（如 GPT-4o Realtime、Phi 等），并通过微调提升准确性和品牌一致性
- **实时音频处理**：内置降噪、回声消除、打断检测等技术，确保流畅交互
- **虚拟数字人支持**：可配置虚拟形象，增强视觉参与度

## 方案优势

- 低延迟
- 模块整合
- 灵活定制

## 典型应用场景

- 客服中心
- 车载助手
- 虚拟教师
- 政务助手
- 人力资源助手

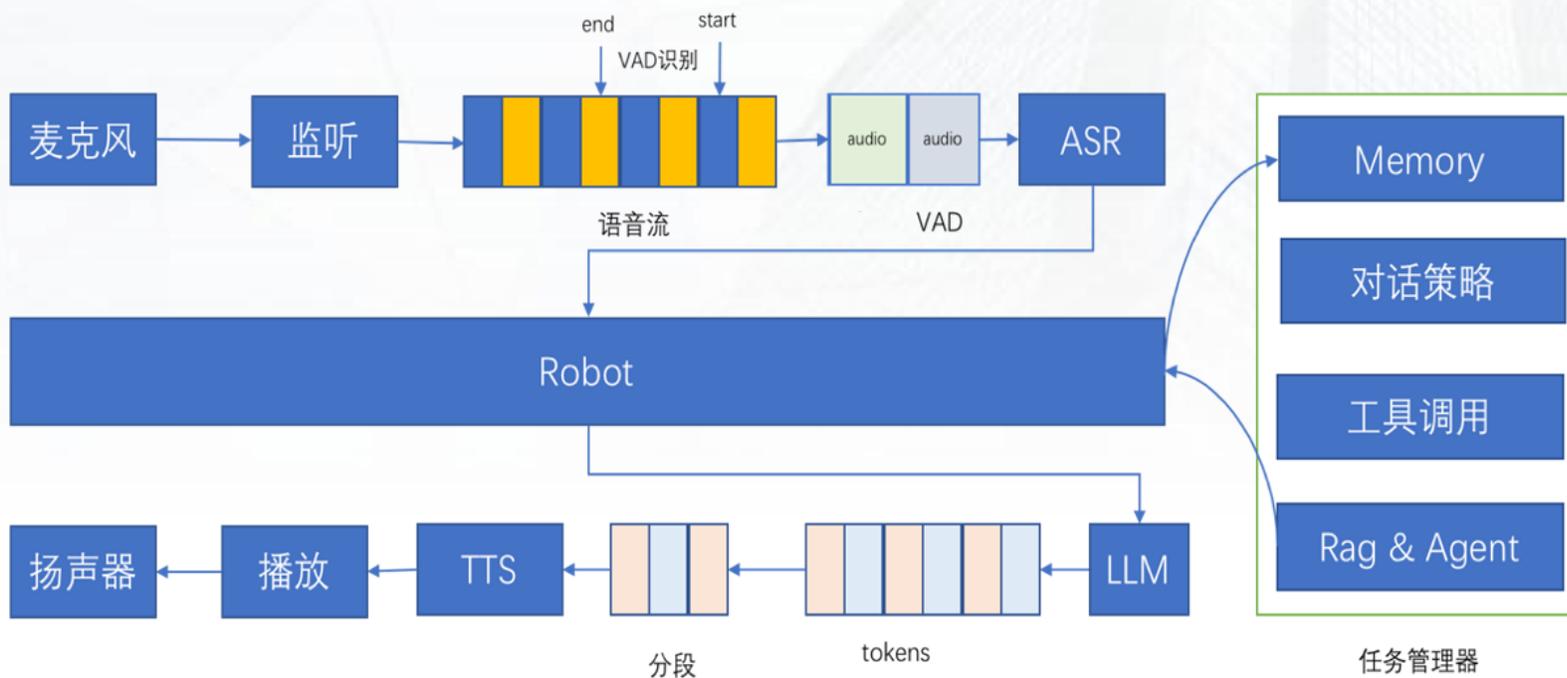


## 方案优势

- 高质量语音对话**：整合了优秀的ASR、LLM和TTS技术，确保语音对话的流畅性和准确性。
- 轻量化设计**：无需高性能硬件即可运行，适用于资源受限的环境。

## 功能特性

- 语音活动检测**：采用 Server-VAD 过滤无效音频，提高识别效率。
- 智能对话生成**：基于 Azure OpenAI 强大语言理解能力，生成自然回复。
- 支持打断**：灵活配置打断策略，识别关键词和语音打断，确保即时反馈，提升交互流畅度。
- 支持记忆功能**：记忆用户偏好与历史对话，提供个性化互动。
- 支持工具调用**：集成外部工具，用户可语音请求信息或执行操作，增强实用性。
- 支持任务管理**：高效跟踪任务进度、设置提醒和动态更新，帮助用户把握重要事项。



# 联网聊天解决方案

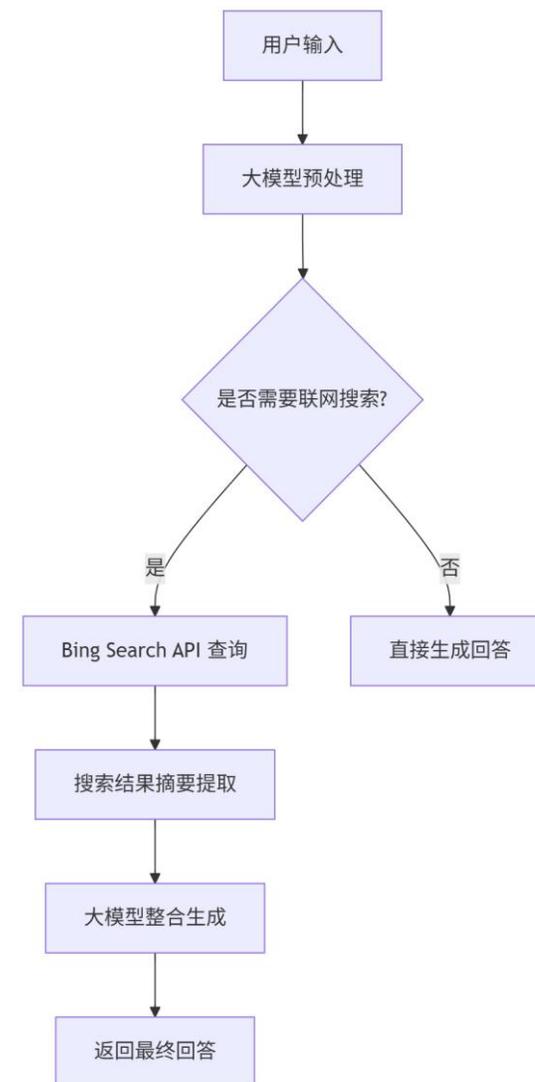
## 适用场景：

- 实时信息查询（新闻、股价、赛事比分）
- 长尾知识问答（小众技术问题）
- 对抗大模型知识陈旧问题

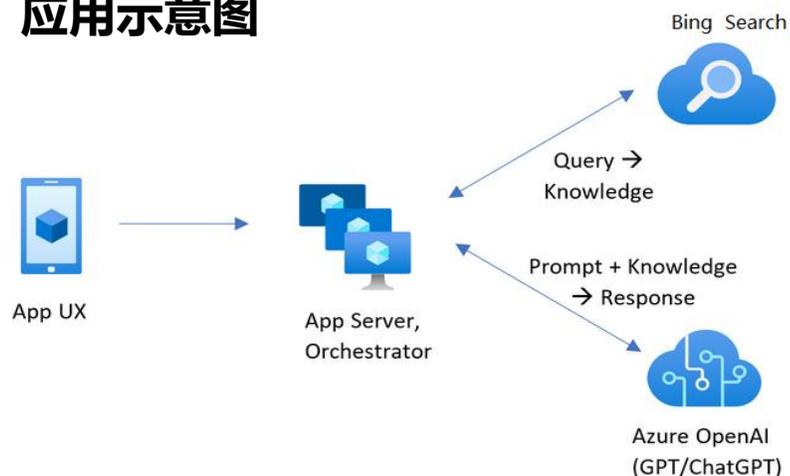
## 部署建议：

- 1.先用Azure OpenAI + Grounding with Bing Search快速验证效果
- 2.逐步引入本地模型降低成本
- 3.通过A/B测试优化搜索触发策略

## 联网聊天流程图



## 应用示意图



## 应用场景

- ◆ **电商直播**  
AI虚拟主播支持多语言实时翻译和互动，提升跨境销售。
- ◆ **博物馆展览**  
数字人复现历史人物，增强观众沉浸体验。
- ◆ **企业营销**  
数字人助力品牌宣传、产品推广及培训，支持多场景应用。
- ◆ **客户服务**  
智能客服数字人提高咨询处理效率。
- ◆ **教育培训**  
虚拟教师提供个性化教学服务。
- ◆ **法律咨询**  
数字人参与辩论和法律普及，提升专业效率。
- ◆ **新闻播报**  
数字人替代真人播报新闻及天气等内容。

## 方案特点

- ✓ 深度神经网络合成视觉
- ✓ 跨语言声音
- ✓ 语音语调合成



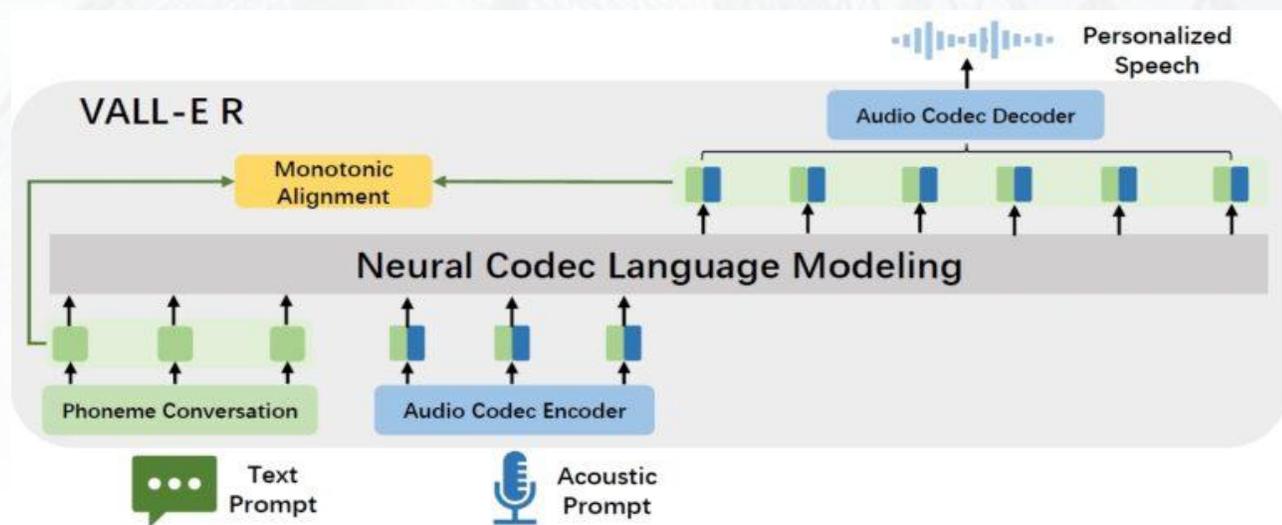
# 声音复刻解决方案

**方案背景：**利用AI技术（如Azure AI Speech）实现声音快速克隆，支持多种语言和情感调节，广泛应用于无障碍交流、内容制作及企业服务领域，同时重视伦理规范和防止滥用的保障措施。

- ◆ **Azure AI Speech (个人语音功能)** 提供 “DragonV2.1Neural” 模型，几秒音频即可生成高保真语音克隆，支持100多种语言，提升自然度和发音准确性，适用于聊天机器人和多语言配音，并通过水印和授权防止滥用。
- ◆ **NaturalSpeech 3** 是与高校合作开发的零样本语音克隆系统，3秒音频实现人声复制，支持情感和语调调节，基于FACodec模型和大量训练数据，实现精准音色控制。
- ◆ **Custom Voice** 支持企业用1小时录音生成专属语音，应用于培训和客服，集成Azure云并通过API/SDK部署。

## VALL-E

只需3秒音频样本即可复制声音的音色、情感和环境声学特征，采用神经编解码器语言模型，擅长零样本语音合成和编辑。



**注：**所有方案均要求用户授权、添加水印，并遵守《深度合成管理规定》，明确标识AI生成内容

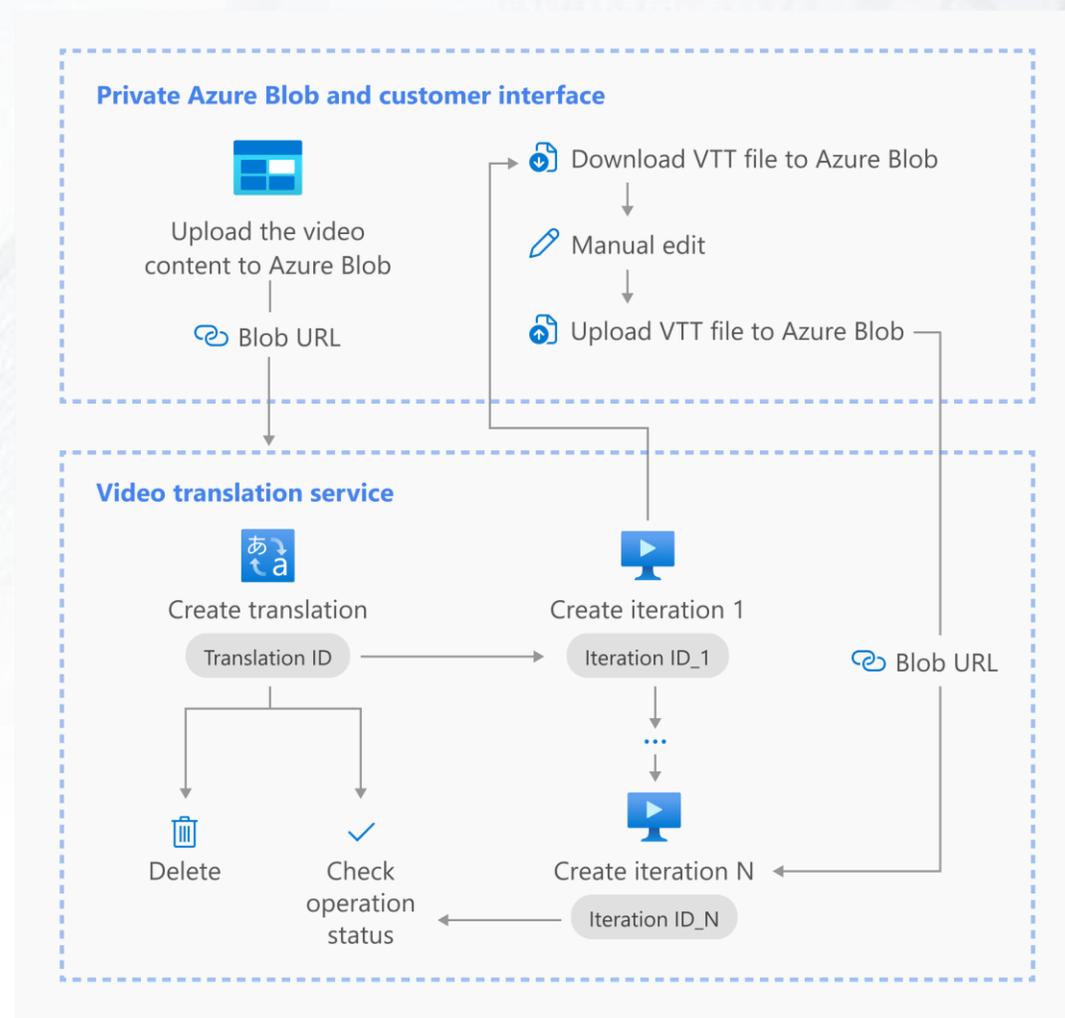
## 传统视频翻译挑战

- **费用高昂**: 依赖人工语音演员和繁琐后期制作
- **时间消耗大**: 录制编辑过程耗时长, 影响发布速度
- **质量波动**: 语音演员难以精准还原原声, 影响观众体验

## AI视频翻译优势:

- ✓ 降低昂贵人工语音和后期制作需求
- ✓ 大幅减少制作时间
- ✓ 精确复制原始声音
- ✓ 快速高效生产大量内容

## 视频翻译方案



# 目录

## CONTENTS



1

领驭科技公司介绍

2

AI语音解决方案

3

案例分享

# 案例1:翻译耳机

## 客户简介

XX科技, 传统耳机制造商,  
ODM代工厂

## 客户诉求

耳机销量大, 用户强需求要有翻译功能, AI聊天  
等功能。客户对微软语音不了解, 缺乏开发能力

## 客户收益

从 **普通耳机** 到 **AI 翻译耳机** 大转型, 支持小语种  
翻译, 同声传译等潮流功能, 翻译准确率超95%

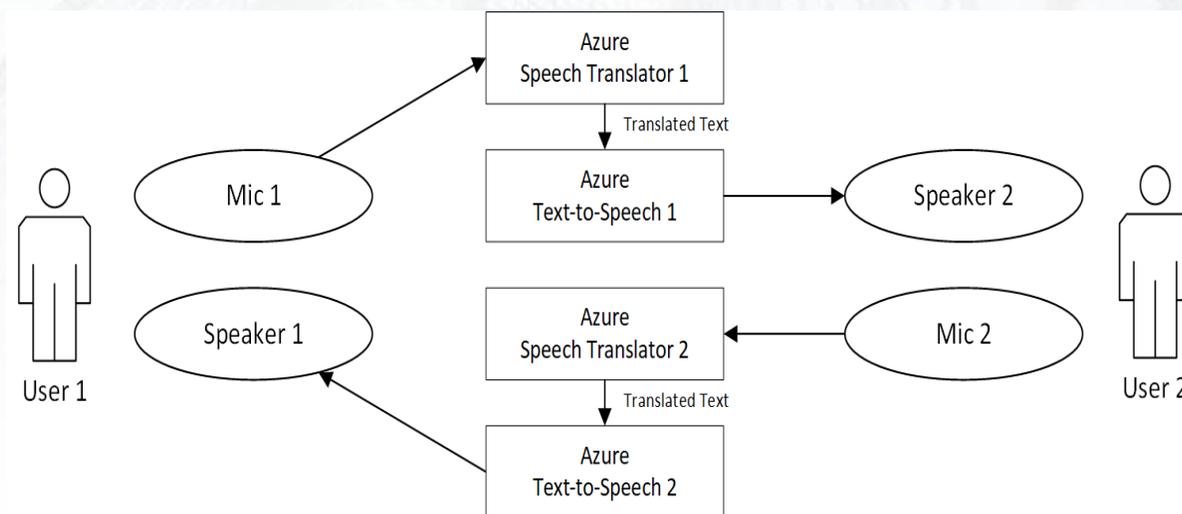
## 领驭提供的服务

提供基于微软语音服务的端到端解决方案, 从  
方案设计, 选型到落地实施, 再到售后支持,  
领驭团队提供全程服务。

帮助客户实现了实时语音翻译, 同声传译, 会  
议助手, AI智能体等功能, 支持多语种高精度  
同传, 适配耳机硬件SDK。

针对**语音延时**, **翻译质量**, **成本优化**三个方面  
为客户提供深度全链路技术支持

## (耳机) 同声传译解决方案



## 客户简介

XX科技, AI眼镜制造商  
海外销售

## 客户诉求

出海销量大, 产品需求翻译、AI助手等功能。客户对微软语音不了解, 缺乏开发能力

## 客户收益

解锁高端AI眼镜的关键卖点: 跨语言对话, 识图翻译。打破沟通壁垒, 以极致的用户体验, 驱动销量与溢价

## 领驭提供的服务

提供基于微软语音服务的端到端解决方案, 帮助客户集成Azure AI翻译, 快速为眼镜注入强大的“沟通超能力”, 从而增加销量、提升品牌价值、并控制成本。实现了实时语音翻译, 同声传译, 拍照翻译, 视频字幕, 导航等功能。

选用的AI模型:

- GPT-4o-realtime
- GPT-4o
- Speech STT

## (AI眼镜) 语音翻译解决方案



AI眼镜结构示意图

本报告中所研究的「AI眼镜」主要是指面向消费者的, 接入大语言模型或多模态模型, 具备语音交互、视觉识别、信息提示等典型功能的智能眼镜。

### 语音交互

- 通过语音与智能助手进行沟通问答, 查询和获取信息, 同时还可通过语音发送指令, 使眼镜执行相关操作

### 视觉识别

- 使用眼镜搭载的摄像头拍摄图像或视频, 利用多模态大模型进行物体、场景、文字等识别并给予反馈

### 信息提示

- 利用光学显示系统, 将虚拟信息通过眼镜叠加到现实场景中, 提供AR导航、演讲提词等可视化信息提示

# 案例3:短剧出海

## 客户简介

XX科技, 短剧出海

## 客户诉求

apps上线海外短剧出海, 买1000部短剧, 需要翻译成其他语言, 北美巴西阿根廷日韩, 中东。

## 客户收益

帮助客户加速变现, 利用 Azure CDN 和全球基础设施, 确保视频秒开、播放流畅, 降低用户流失, 直接提升付费率和广告展示量

## 领驭提供的服务

**成本降低:** 为客户设计弹性部署架构, 应对短剧流量波动大的特点, 自动伸缩, 避免资源闲置, 显著节省IT成本。

**AI驱动:** 借助Azure OpenAI视频翻译和AI语音功能, 帮助客户生成优质的出海视频作品。

**安全与合规:** 利用微软全球合规认证帮助客户快速进入新市场, 避免合规风险, 保护收入来源。

Azure弹性部署: 每50用户资源配置

|     | 配置               | 单位    | 单价        | 备注                 |
|-----|------------------|-------|-----------|--------------------|
| 计算  | D2asv6           | 小时    | US\$0.054 | 2 vCPUs, 8GB RAM   |
| 存储  | Azure Blob       | 1TB/天 | US\$0.67  | 1TB热存储LRS冗余        |
| CDN | Azure Front door | GB    | US\$0.17  | 全球负载均衡, 192 个边缘接入点 |



# 谢谢!

# Thank you!



400-8889-356

[www.leinwin.com](http://www.leinwin.com)



**深圳领驭科技有限公司**

SHENZHEN LEINWIN TECHNOLOGY CO.,LTD

广东省深圳市南山区粤海街道国人通信大厦 9F

9F, Guoren communication building, Yuehai street,  
Nanshan District, Shenzhen City, Guangdong Province