

* RTE 和 AI 融合生态洞察报告

→ 2024



前言

2024 年，AI 与实时互动技术的结合达到了前所未有的高度，推动了行业的发展与变革。5 月，OpenAI 发布了 GPT-4o，并展示了其对话功能，仿佛电影《HER》中的智能助手走入了现实生活。紧随其后，6 月，a16z 发布了关于语音 AI 的展望报告，详细分析了这一市场的巨大潜力，并为未来的语音 AI 发展绘制了蓝图。之后，其他行业领军企业如 Cartesia 和 Bessemer Venture Partners 也纷纷发布了各自的洞察报告，深入探讨了语音 AI 和实时互动技术的前景。

进入 2024 年 10 月，OpenAI 宣布与 Agora、Twilio 等实时互动技术公司展开合作，同时，国内各大科技公司也陆续公布了在对话 AI、多模态 AI、语音 AI 等领域的技术布局和市场战略。这一切都标志着实时互动技术与 AI 的结合进入了一个全新的发展阶段。

随着这两项技术的深度融合，我们已经看到它们在多个领域和场景中展现出巨大的应用潜力，也赋予了智能体越来越可用的能力，语音助手可以帮助用户打电话、操作终端设备；AI 能为用户提供情感陪伴；而能够纠正语音的口语陪练也让学习更加个性化和高效。这些创新的应用让智能体变得愈加智能、实用和贴近用户需求。

在这一背景下，RTE 开发者社区与 InfoQ 研究中心联合发布了本份报告，旨在深入探讨实时互动技术与 AI 结合的现状、挑战与未来趋势。我们希望通过这份报告，为行业从业者、技术开发者以及创新者们提供有价值的洞察，帮助他们更好地理解技术融合的机遇与挑战，推动技术落地与应用创新。

本报告专注于探索如何利用实时互动技术与 AI 相结合，来提升用户体验，以及社区是如何帮助 Voice Agent 生态建设和发展的。我们相信，这些洞察将为行业的持续创新和发展提供重要参考。

前言 | 实时互动 (RTE) 行业定义回顾

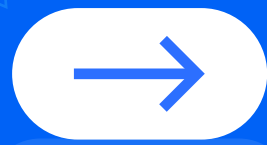
在报告的开头，我们再次回顾实时互动 (RTE) 的定义。实时互动 (RTE) 是远程互动、多维沟通、身临其境的交互活动。对互联网技术架构提出更高要求，其特性主要体现在**实时性**和**互动性**两个方面。

实时互动 (RTE) 以实时性和互动性为核心特点



来源：信通院《实时互动产业发展研究报告 2022》；声网研究院《实时万象——声网带你探寻实时互动的边界》

目录



CONTENT



01

实时互动与 AI 行业技术演进观察与分析



02

应用场景新挑战，造就 VOICE AGENT 新的产业架构



03

RTE开发者社区架起生态桥梁，加速生态沟通



04

实时互动智能生态的未来发展预测



01

实时互动与 AI 行业技术演进观察与分析

02

应用场景新挑战，造就 VOICE AGENT 新的产业架构

03

RTE开发者社区架起生态桥梁，加速生态沟通

04

1.1 AI 应用落地时代：重新定义算力、数据与算法的价值，从技术驱动到价值赋能

- 在本轮浪潮的早期，行业的焦点集中在模型的规模与通用性上，尤其是对 AGI 的追逐。然而，随着时间的推移，越来越多的从业者开始转向 AI 的实际应用落地，寻求更具现实意义的创新与突破。AI 1.0 的三驾马车——算力、数据和算法，在这一新时代中展现出新的特征。但在应用落地的关键阶段，它们必须更加强调解决实际应用中的复杂挑战，才能真正推动 AI 的商业化进程和技术赋能。

传统 1.0 的三驾马车



追求模型的大规模和全能

基础设施层面

转向

应用落地

应用落地下，三驾马车的转变与新特征

算力分布不均，云边端模式仍在探索中

当前云、边、端的算力资源分布尚未均衡，具体分配需要根据场景和行业特点动态调整。

仍需要行业和场景的专有数据积累

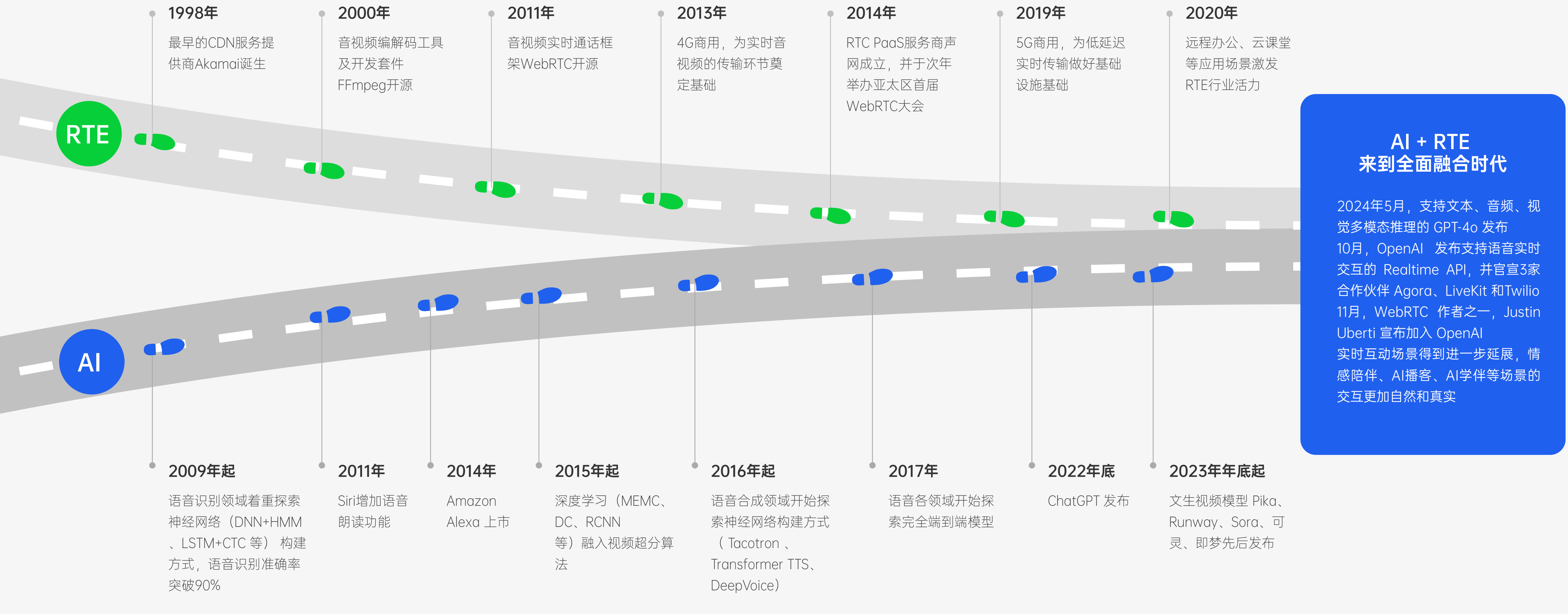
数据数量和质量的必要性已经得到普遍认可，但能捕捉行业需求独特性的专有数据，其获取和积累往往面临隐私、安全等多重挑战。

算法逐渐形成共识，但如何精准把握需求仍是挑战

尽管以 Transformer、LSTM 为代表的主流深度学习架构在性能和效率上趋于稳定和成熟，但在实际落地过程中，仍需结合具体场景和行业需求进行优化，关注落地性能。

1.2 AI 与 RTE 从独立走向交融，共同开启实时互动体验新纪元

• 2024 年，AI 和 RTE（Real-Time Engagement，实时互动）这两个长期以来被视为相对独立的领域，正式开启了融合的序幕。



AI + RTE 来到全面融合时代

2024年5月，支持文本、音频、视觉多模态推理的 GPT-4o 发布
10月，OpenAI 发布支持语音实时交互的 Realtime API，并官宣3家合作伙伴 Agora、LiveKit 和Twilio
11月，WebRTC 作者之一，Justin Uberti 宣布加入 OpenAI
实时互动场景得到进一步延展，情感陪伴、AI播客、AI学伴等场景的交互更加自然和真实

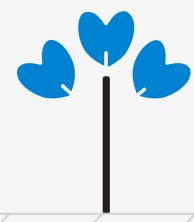
1.3 从独立到共生，历经四大阶段，行业正在进入实时互动智能时代

- 在 AI 和 RTE 逐渐融合的过程中，InfoQ 研究中心发现，AI 的发展后期，为了更好的用户体验，离不开 RTE 技术支撑，而 RTE 技术的进步又能提升 AI 应用的性能，两者相互依赖、共同推动实时互动智能生态的不断升级。

AI 与 RTE 双线并行

AI 与 RTE 独立发展,各自在算法性能、实时交互等方面不断提升，但尚未形成深度协同

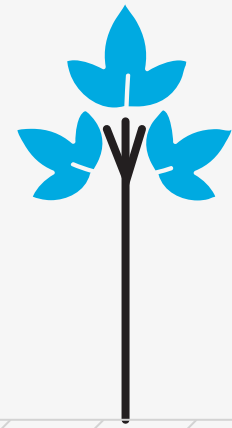
01



RTE + AI

RTE 开始主动整合 AI 能力,如利用超分辨率等 AI 算法来优化实时视频画质
AI 得益于 RTE 的支持，在终端设备上实现更好的响应速度和交互体验

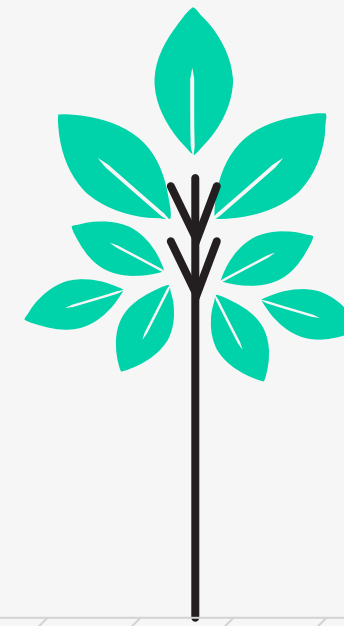
02



AI + RTE

随着 RTE 技术在计算性能和网络传输上的优化，AI算法能够更高效利用边缘计算的数据回传，加速模型迭代，同时提升多模态和语音AI的吞吐速度，为部署和推理提供更强支撑

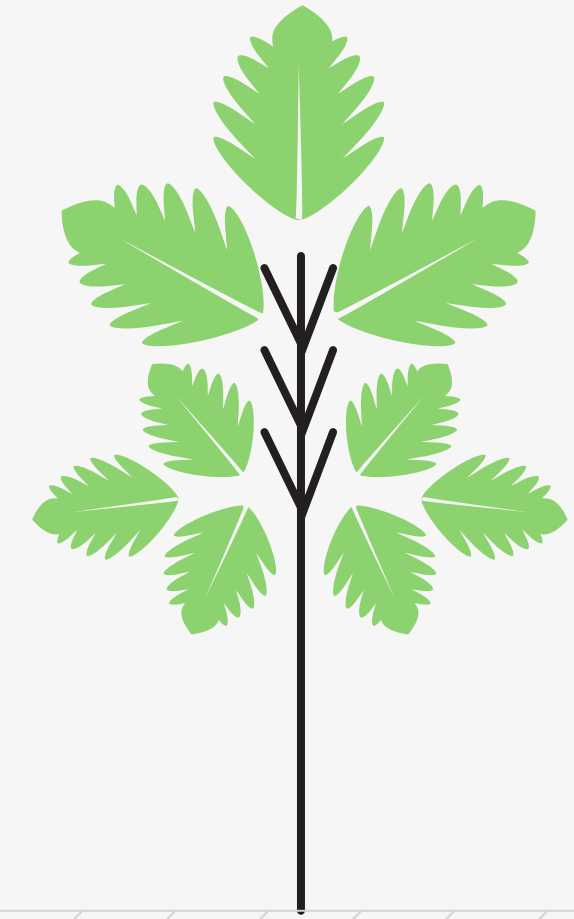
03



实时互动智能

AI 算法与 RTE 系统协同优化，实时互动与AI深度融合，彼此互为原生，形成密不可分的整体，共同构建更智能、更沉浸式的用户体验

04



交互体验新突破：实时、互动与沉浸的重塑

- 在实时互动智能中，交互体验突破了用户对实时性（毫秒级延迟）、互动性（语音、视频、表情的多模态融合）、沉浸感（AR/VR 真实程度）的感知。

01

超真实的拟人人声和情感表达

通过语音合成技术和大模型带来的理解能力，实时互动智能不仅能模拟接近人类的声音，还能根据情境表达丰富的情感，呈现高度拟人化的语音体验

02

极低延迟，极速响应

通过采集、传输等环节的优化，实现毫秒级的响应时间，确保语音交互流畅自然



03

智能打断，沉浸对话

通过 VAD 技术，支持实时智能打断功能，模拟真实对话互动，创造更自然的沉浸式对话体验

04

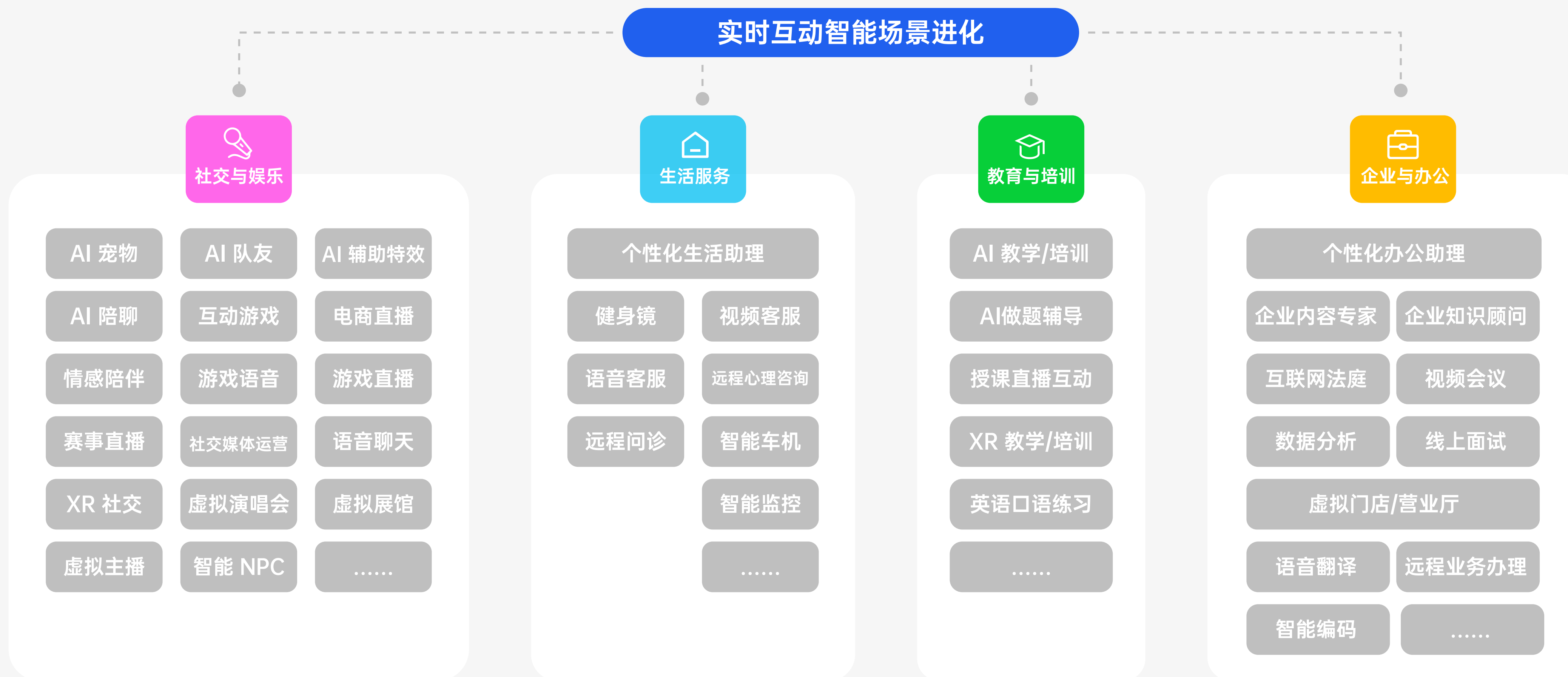
嘈杂环境，正常对话

噪声抑制、噪音过滤，语音增强，实现在嘈杂环境下的持续对话

1.5

技术进步拓展应用边界：从传统场景创新到新兴场景创造

- 通过 AI 和 RTE 技术的不断进步，许多此前存在局限性的应用场景得到了显著优化。这些进步不仅推动了传统场景的创新，也为新兴应用场景的拓展提供了更多可能，带动了社交与娱乐、生活服务、教育与培训、企业与办公等领域的变革。



1.6

高性能、高准确到好体验成为实时互动智能发展的新评价标准

- 基于实时互动智能对于场景下用户体验的重视，我们认为是时候将用户体验作为应用侧的一项重要评价标准。在这一过程中，不仅要考虑AI本身的性能，更要结合AI所应用的实际场景，评估其在特定环境中的表现。因此，实时互动智能和 AI 大模型的评价标准需要有所差别，特别是在应用场景中的互动质量和用户体验上，必须有针对性的标准来进行评估。

高性能

高性能要求：代表应用落地在实时互动场景中运行的效率要求，受到云、边、端算力协同调度和合理架构设计的影响

高准确

高准确要求：代表在特定场景中，对模型的输出精度和判断正确性要求，受场景数据的规模和算法优化程度的共同影响

新衡量标准

好体验

好体验要求：代表应用场景中用户对交互效果、实时性、准确性及沉浸感的感知和期待

实时互动智能新评价标准的三大维度

01

实时互动与 AI 行业技术演进观察与分析

02

应用场景新挑战，造就 VOICE AGENT 新的产业架构

03

RTE开发者社区架起生态桥梁，加速生态沟通

04

实时互动与 AI 行业技术演进观察与分析

2.1

Voice Agent —— 实时互动智能的破局者

- 实时互动智能生态仍处于早期阶段，生态内的关系尚未完全明晰，但各方都在快速探索业务落地的可能性。
- 在这之中，Voice Agent 作为实时互动智能中确定性较高的分支，以其自然直观的交互形式和成熟可靠的技术实现，展现出在特定场景中高效且稳定的优势。

Voice Agent 是利用语音AI和实时互动技术，借助语音等多模态的形式进行交互、解决特定场景问题的智能体，代表有情感陪伴、智能外呼、实时翻译/会议协作等。

Voice Agent



高性能

算力架构：当前实时互动智能应用，尤其是 Voice Agent，对高性能的需求主要集中在毫秒级延迟和实时响应上。这种高性能要求，离不开云、边、端三方算力的高效协同和合理调度设计。随着云端大模型的规模化以及边缘计算能力的增强，Voice Agent 在特定场景中实现更稳定的性能表现。



高准确

模型发展：相较于视频等，GPT-4o 等端到端多模态模型的出现、相关 API 的发布简化了语音交互的工作流程，并提高了实时性与语义理解的质量。这些技术进步使得 Voice Agent 在处理复杂对话时更加高效和准确。



好体验

交互方式：语音交互方式更接近人类自然沟通，具备天然的语言逻辑和情感表达优势。这使用户不需要适应复杂的操作界面或学习曲线，减少交互中的不确定性，用户体验更佳。

应用场景：语音交互通常被限定于特定场景，如语音助手、导航、实时翻译等，其使用路径和目标明确，用户的操作和系统的响应使用户体验更佳。

2.2

通用和行业玩家正在共建 Voice Agent 产业生态



2.3 Voice Agent 产业生态上下游链路较长，跨层协作难度大

- 上下游链路长、生态协作效率低的问题仍然突出。在现状中，产品侧快速变化的需求难以高效传递至基础设施和中间层，响应速度受限；同时，工具与标准缺乏统一，跨层协作复杂度高，进一步制约了整体生态的联动效能。Voice Agent 的产业架构虽在完善，但现阶段各层级的协同能力仍需提升，以应对更高效、更敏捷的市场需求。

Voice Agent 的产业架构现状

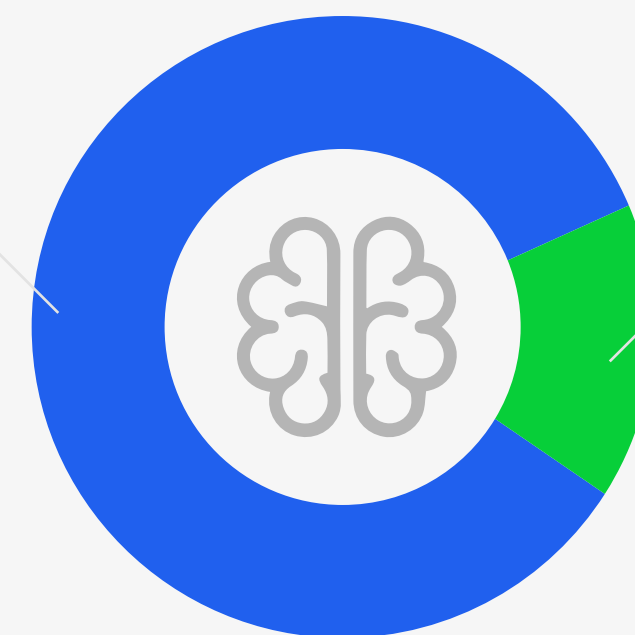
01 以基础设施为核心，产品生态协作效率受阻

Voice Agent 产品侧仍在早期探索，但产品设计、核心功能与场景适配仍在探索中，缺乏通用性或标准化方案。制约了生态协同效率的提升

80%

基础设施及中间层

包含从底层 Infra 和 API 选择到智能体平台的一系列，为了实现 Voice Agent 最终应用实现的工具及平台



20%

产品侧

包含内置语音智能体的垂类应用和硬件设备

02 上下游链条长，导致响应速度慢

应用侧快速变化的场景需求，难以通过多层链路迅速反馈到基础设施，放慢了响应速度

以实时翻译为例，可能涉及的上下游协作：



Voice Agent产品侧面临算力、工具、终端与流量的多重生态协作挑战

- 在实时互动智能领域，Voice Agent 应用的未来充满潜力，但目前仍面临算力、终端、流量和工具四大关键挑战。这些挑战不仅影响用户体验的提升，更需要整个生态系统的深度协作来应对。

端侧和云端算力的协作挑战

算力是 Voice Agent 应用落地的核心。云端算力支持高精度模型，但高成本和响应延迟限制了普及；端侧算力延迟低，但能力有限，需依赖云端补充。如何平衡云端与端侧算力分布，实现高性能与经济性的兼顾，成为整个生态的共同难题。

多终端和多场景的协作挑战

多样化的设备和场景对终端体验提出了更高要求，例如降噪处理、弱网对抗和低延迟交互等技术挑战。单凭应用层难以全面适配，亟需硬件厂商、基础设施提供方与开发者协同优化技术方案，提升多终端环境下的适配效率。

4 Voice Agent 产品侧面临的 大挑战

API/SDK与场景的协作挑战

当前 API 和 SDK 在 Voice Agent 原生场景中的设计适配性不足。交互中的打断处理、语句分割、多轮对话逻辑等需求未能得到有效支持，加剧了开发复杂性。如何围绕 Voice Agent 的场景特性，优化接口设计，提供更专用、更灵活的工具，成为提升开发效率的重要方向。

产品和流量渠道的协作挑战

在用户分散的市场中，Voice Agent 应用难以通过传统方式高效触达目标群体。如何通过数据共享、算法优化和场景协同，实现精准覆盖、减少无效流量，成为生态共建的重要课题。

2.5 Voice Agent 产品侧面临算力、工具、终端与流量的多重生态协作挑战

- Voice Agent 需要格外考虑到云端和端侧的算力分布，这主要来自于语音交互对低延迟和流畅的用户体验的追求。

高效的云端协同成为用户体验的核心

通过云端与端侧算力的智能调度和合理分配，可以根据应用场景的需求，实时动态地在云端和端侧之间平衡计算任务。这样，既能利用云端强大的算力处理复杂任务，又能在端侧实现低延迟、高效能的实时响应，从而提供流畅、丝滑的用户体验。

性能高成本高和体验低

云端算力虽然能提供足够的处理能力，能够充分展现模型的能力，却伴随着较高的成本和较慢的响应速度。这种延迟对于需要实时交互的实时互动智能场景而言，会很大程度地影响用户体验。

体验佳资源有限

端侧算力具有低延迟和本地处理的优势，能够快速响应用户需求，但其计算能力和存储空间有限，尤其是手机等终端。因此，端侧通常需要云端补充，才能处理高负载任务并确保流畅的用户体验。

云上



端侧

2.6

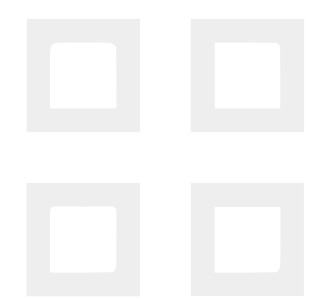
Voice Agent 产品面临不专用的 API/SDK 与场景如何协作的难题

- 除了算力分布外，Voice Agent 也需要更好用的、对语音交互场景更有针对性 API/SDK，以将高效连接底层能力与应用场景。

更好用的 API/SDK 对实时互动智能 Builder 的正向作用

专用API更适配场景需求

通用的 API 无法满足 Voice Agent 场景的特殊需求。因此，API/SDK 需要为 Voice Agent 做特定优化，能够支持复杂的语音交互、自然语言处理和多模态任务。



易用性加速开发过程

理想的 API/SDK 应提供低代码甚至无代码的支持，使非技术背景的人员也能通过简单的配置或拖拽实现复杂的功能，减少开发时间和成本。



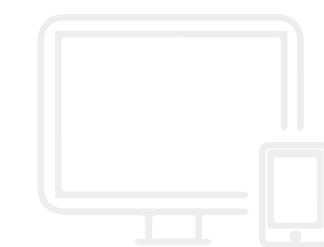
稳定性与高性能保障流畅体验

稳定性和高性能不可或缺，尤其在 Voice Agent 场景中，API/SDK 必须确保在高并发、低延迟的环境下，依然能保持流畅的性能表现，避免影响用户体验。



生态灵活性优化跨环境体验

API/SDK 的生态灵活性也至关重要。它们应支持多种适配，能够兼容不同的设备、操作系统和网络环境，并支持各种编程语言和框架，方便开发者在不同的平台和技术栈上进行灵活集成和部署。



2.7 Voice Agent产品需要考虑应用场景与不同设备终端之间的适配协作

- 在 Voice Agent 场景下，终端体验从单纯的设备兼容转向更复杂的交互适配，端侧体验的重要性日益突显。这要求开发者优化不同终端的硬件性能和网络适应性，确保在弱网和低延迟环境下仍能保持流畅、稳定的用户体验。同时，在产品交互设计时，考虑到不同终端的操作方式和用户习惯。

实时互动智能场景需要实现端侧资源和用户体验的平衡

用户体验

应用场景

- 语音、视频的快速识别与响应
- 流畅的实时互动体验
- 不同的交互方式和特点



资源消耗

端侧设备

- 有限的计算资源
- 有限的存储空间
- 网络、续航等资源



需要适配的终端逐渐增多

交互适配

- 随着应用需求变化，降噪、弱网对抗、低延迟处理等问题逐渐影响实时交互的流畅性和稳定性。弱网或设备性能限制下，语音识别、情感表达和反馈准确性成为关键。
- 不同终端的交互方式各异，开发者在适配的过程中也需要认真思考。

设备兼容

- 不同终端的硬件差异、操作系统限制、计算能力和存储空间的差距，要求开发者为每种设备设计专门的适配方案。

AI手机

AI PC

AI 耳机

AI 玩具

AI 学习机

AI 原生终端

Voice Agent产品要同精准的流量渠道协作，进行产品推广

- 通过更有性价比的流量，Voice Agent 应用能够触达更多潜在用户，积累的数据反过来促进模型优化和用户体验提升，从而建立长期优势。

更有性价比的流量对 Voice Agent 产品的助力传导路径



保持较高的产品竞争力，提高用户忠诚度，最终实现长期的竞争力增长

新兴渠道崛起，重塑用户连接

现阶段的开发者正在开发者社区之外探索高效的触达渠道：

01 自媒体平台



02 新型产品平台

- Dify、Coze 的 MarketPlace
- 各大科技企业的 Agent Store/Space

2.9

加入用户体验后，哪个领域有望诞生下一个Killer App?

新衡量标准

在重新设定的云端协同应用难度、场景准确性要求以及用户体验要求，三个衡量维度下，InfoQ 研究中心邀请了分析师和多位大模型对实时互动智能应用场景进行了打分，并征询了多位行业专家的建议，最终生成了实时互动智能应用领域分析象限图。

实时互动智能应用领域分析象限图

交融爆发区:

处在该区域内的应用场景具备一定的技术门槛和壁垒，但这些挑战和要求尚未达到难以逾越的程度

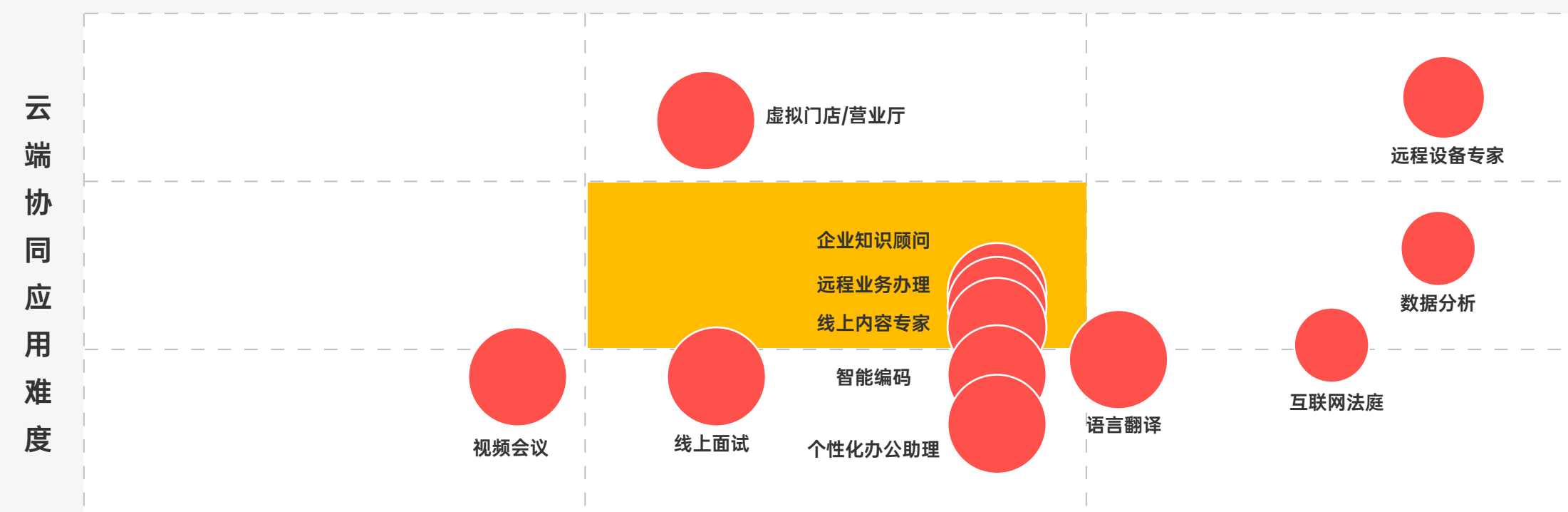
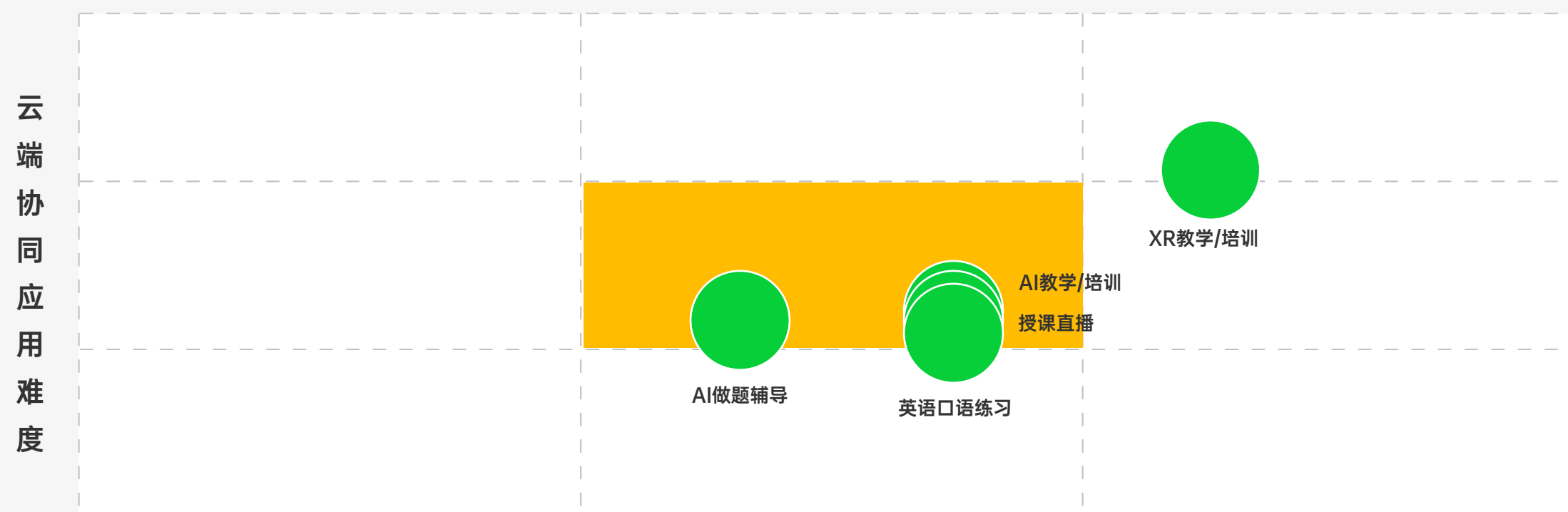
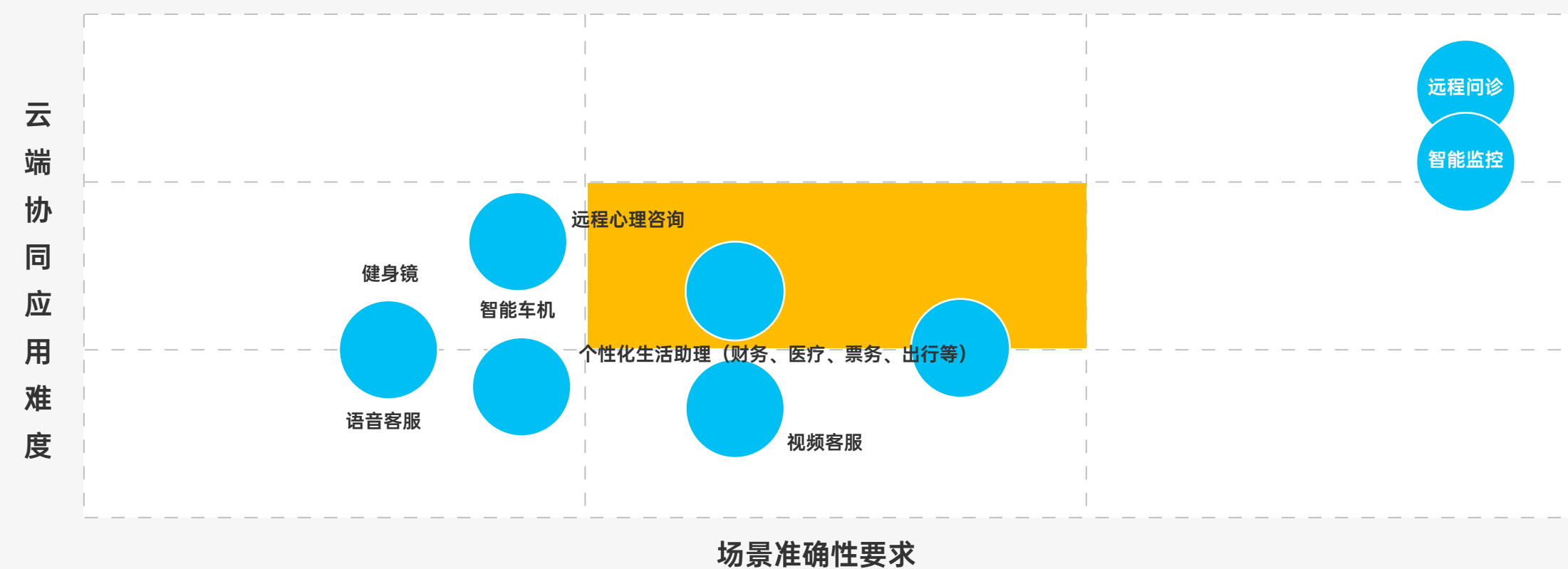
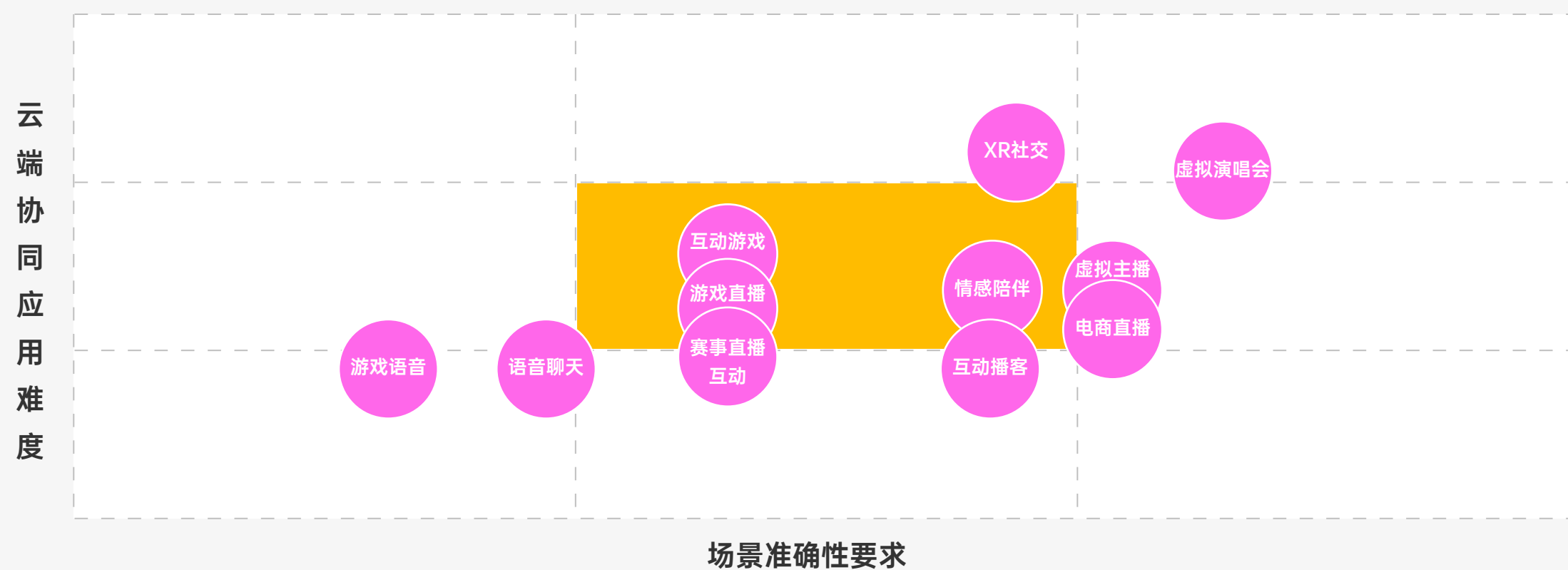
社交与娱乐

生活服务

教育与培训

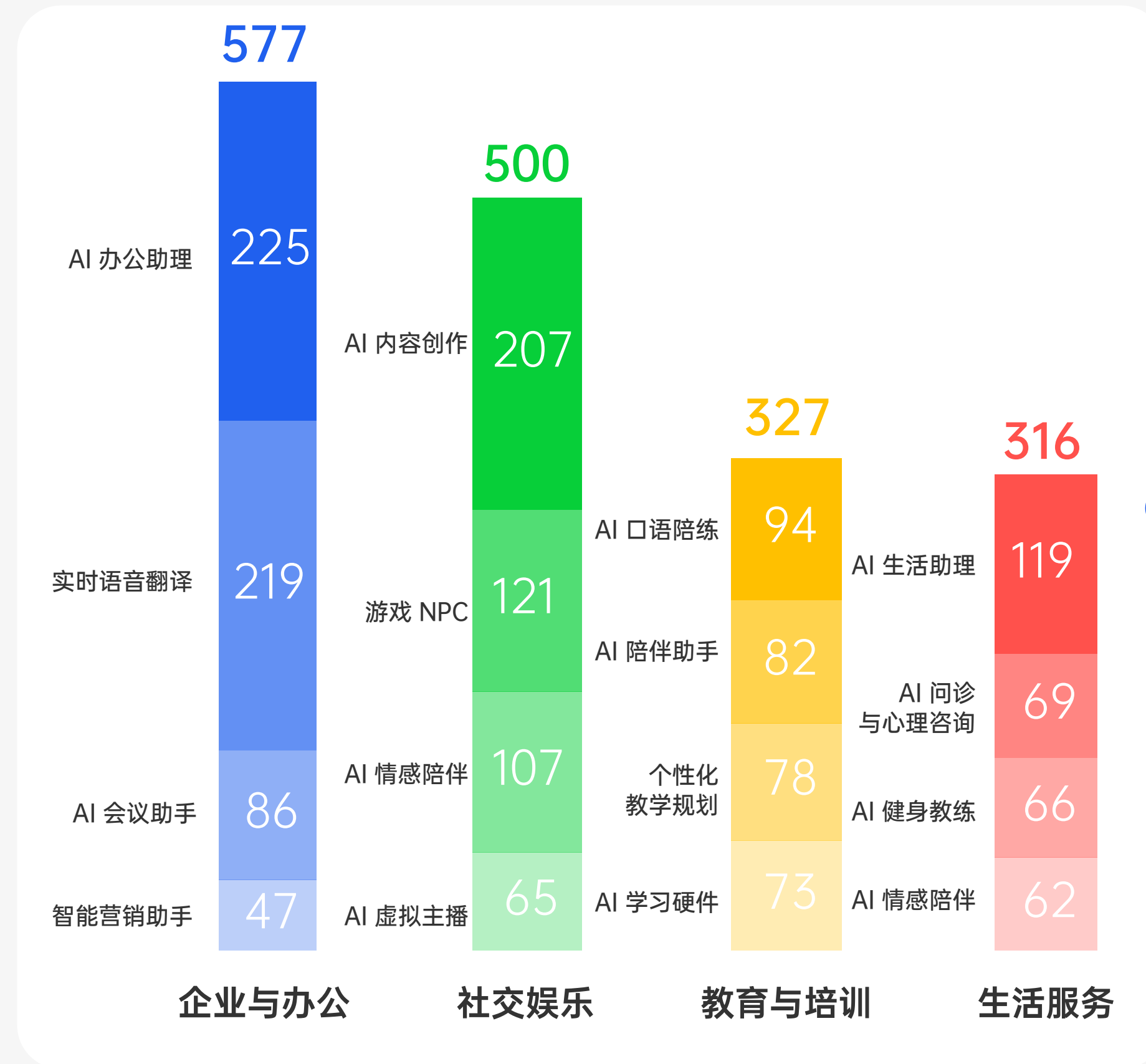
企业与办公

气泡大小代表该应用场景对用户体验的要求



2.10 开发者声音：办公助理、语音翻译和内容创作最可能诞生 Killer APP

- 2024年10月，我们也在 RTE 开发者大会现场，放置了实施互动智能场景下一个 Killer App 的投票板，超过1000位的开发者在大会现场，投票预测下一个 Killer APP。



- 01 社交与娱乐、企业与办公两大场景：最受开发者关注
- 02 办公助理、语音翻译和内容创作：Killer APP 最可能爆发应用
- 03 AI 分身社交、AI 展览、多智能体、具身智能：开发者提名领域

01

实时互动与 AI 行业技术演进观察与分析

02

应用场景新挑战，造就 VOICE AGENT 新的产业架构

03

RTE开发者社区架起生态桥梁，加速生态沟通

04

实时互动智能生态的构建与展望

3.1 算力协同新思路：基础设施、产品侧与社区的三方合力

- 在实时互动场景中，算力的云端与端侧协同是实现高性能与经济性兼顾的核心挑战。基础设施通过轻量化模型支持和透明化算力调度,为开发者提供端侧友好的工具链;产品侧聚焦功能需求拆解与资源管理优化,实现灵活的任务分布和动态模型切换;社区则通过资源整合和合作共建,推动上下游企业协同创新,共同突破云端算力瓶颈。

社区

- 提供交流场合:**通过社区力量,共建云边协同相关的技术项目、专题讨论、比赛等,推动上下游企业分享解决方案并探讨合作。
- 上下游资源整合:**发挥社区的牵头作用,整合模型压缩工具、轻量化算法等资源,加速产品前期测试。

基础设施及中间层

- 轻量化模型支持:**为垂类应用提供端侧友好的工具链如支持量化、剪枝和蒸馏的模型优化工具,帮助开发者快速部署轻量化模型。
- 云端协同框架:**提供透明化的算力调度方案,支持开发者按需调用云端资源,同时动态调整端侧推理任务。

产品侧

- 专注功能需求拆解:**明确目标场景的核心需求,合理拆解哪些功能需要高精度(依赖云端)、哪些功能可以在端侧完成。
- 优化资源管理:**在硬件允许范围内设计运行时的动态模型切换机制,例如在高负载时启用低资源消耗模式保障用户体验的稳定性。

3.2 解锁场景适配潜能：API/SDK 协同的三方实践路径

- 在 Voice Agent 应用中，API/SDK 的场景适配性是提升开发效率与交互体验的关键。基础设施通过专用接口设计和易用性优化，为典型场景提供灵活支持;产品侧聚焦场景需求深挖和快速验证，确保工具与实际应用高效对接;社区则通过插件共建与场景手册，整合资源促进协作，全面释放 API/SDK 的场景适配潜能。

社区

- 激励开发者贡献API/SDK:**通过活动激励开发者互相贡献插件，联合开源社区和云厂商、平台伙伴(如 dify、coze)合作推广 API 插件的落地应用。
- API/SDK场景手册:**按产品场景细分(如外呼、教育、情感陪伴等)，搭建场景的 API/SDK ，帮助场景开发者更快更精准的进行开发选择。

基础设施及中间层

- 打造专用API:**围绕典型 Voice Agent 场景需求，设计专用的接口与 SDK ，优化如上下文追踪、多轮对话和多模态融合的适配能力。
- 持续改进开发易用性:**通过提供灵活的参数化接口或可视化配置工具，让开发者轻松定制符合自身场景的交互逻辑。

产品侧

- 深挖场景需求:**基于具体场景的特性，定义 API/SDK 的功能诉求，例如客服需要重点优化打断处理，教育场景则需要支持长对话逻辑与上下文保持。
- 设计原型验证:**快速测试现有API的适配性，反馈改进建议，推动基础设施优化工具与应用场景的对接。

3.3 弱网、低延迟与多设备：终端适配的协作新思路

- 在多终端场景中，Voice Agent 的适配需求从设备兼容转向智能协同。基础设施通过跨终端适配框架和交互习惯优化，构建弱网环境下的低延迟体验;产品侧聚焦场景化设计和性能优化，确保终端交互逻辑与用户需求高度契合;社区则通过案例集分享与专项小组协作推动行业标准化，共同突破多设备协同的技术瓶颈。

社区

- **终端适配专项小组:**组织垂直行业工作坊，帮助终端设备明确需求，共同探讨适配语音降噪、弱网抗性和延迟优化的技术。并将相关技术提供商引入对接，
- **总结适配案例集:**发布基于 Voice Agent 的多终端适配案例集，供开发者参考学习。

基础设施及中间层

- **跨终端兼容支持:**提供统一的多终端适配框架，抽象设备差异，简化开发者的适配工作。提供支持弱网和低延迟场景的基础能力，如分布式网络节点和边缘计算优化方案。
- **交互习惯优化:**支持跨设备的用户状态和偏好同步构建支持设备间无缝切换的上下文共享机制。

产品侧

- **细化场景设计:**根据设备特点(如智能音箱的场景化对话、手机的移动性)优化终端交互逻辑，使应用更贴近具体场景的用户需求。
- **终端性能优化:**在弱网环境或资源受限的情况下，优先保障核心功能的流畅运行，避免用户体验大幅下降

3.4 重塑流量协同：生态、产品与社区的创新新实践

- 在流量分散的市场中，生态、产品与社区的协同创新成为 Voice Agent 破局的关键。基础设施通过灵活的生态合作机制，助力第三方应用和硬件厂商共同推广功能;产品侧以场景化营销策略精准触达用户;社区则通过原型测试、流量平台合作和运营支持，帮助开发者快速适应市场，推动流量高效转化。

社区

- 提供原型产品测试机会:**通过工作坊、体验活动、社区自媒体推荐等形式，实现 Demo 产品快速验证与快速迭代。
- 用户触达新方式探索:**协调新兴流量平台和智能体平台与 Voice Agent 产品/解决方案提供商，共同探索用户触达的创新方式。
- 提供产品运营帮助:**提供场景化运营指南或陪伴活动，赋能开发者快速打入分散市场

基础设施及中间层

- 生态合作机制:**提供灵活的生态合作机制，支持第三方应用和硬件厂商在平台上共同推广 Voice Agent 功能

产品侧

- 精细化用户细分:**基于用户行为和需求分析，设计个性化的产品推荐和定制化的使用场景，以提高用户粘性和转化率。
- 制定精准营销策略:**根据不同市场和用户群体的特点定制差异化的营销活动，结合数据分析优化广告投放和促销策略。

3.5 Voice Agent产品需要更多新平台新力量，实现更精准的资源对接

- 在日益独特化的应用场景中，在 Voice Agent 产品快速探索和迭代的今天，原有的开发平台、流量平台等无法实现资源的快速匹配，亟需转变。社区，通过更新资源对接的方式，帮助 Voice Agent 产品更好地应对算力、工具、终端和流量挑战，并创造更优秀的用户体验和产业结构。

社区更新资源对接以帮助 Voice Agent 产品应对算力、工具、终端和流量挑战

协助探索更合理的算力架构

- 提供交流场合:**通过社区力量，共建云边协同相关的技术项目、专题讨论、比赛等，推动上下游企业分享解决方案并探讨合作。
- 上下游资源整合:**发挥社区的牵头作用，整合模型压缩工具、轻量化算法等资源，加速产品前期测试。

联合资源开发更专用的 API/SDK

- 激励开发者贡献 API/SDK :**通过活动激励开发者互相贡献插件，联合开源社区和云厂商、平台伙伴(如 dify、coze)合作推广 API 插件的落地应用。
- API/SDK 场景手册:**按产品场景细分(如外呼、教育、情感陪伴等)，搭建场景的API/SDK，帮助场景开发者更快更精准的进行开发选择。



助力更适配的终端交互体验

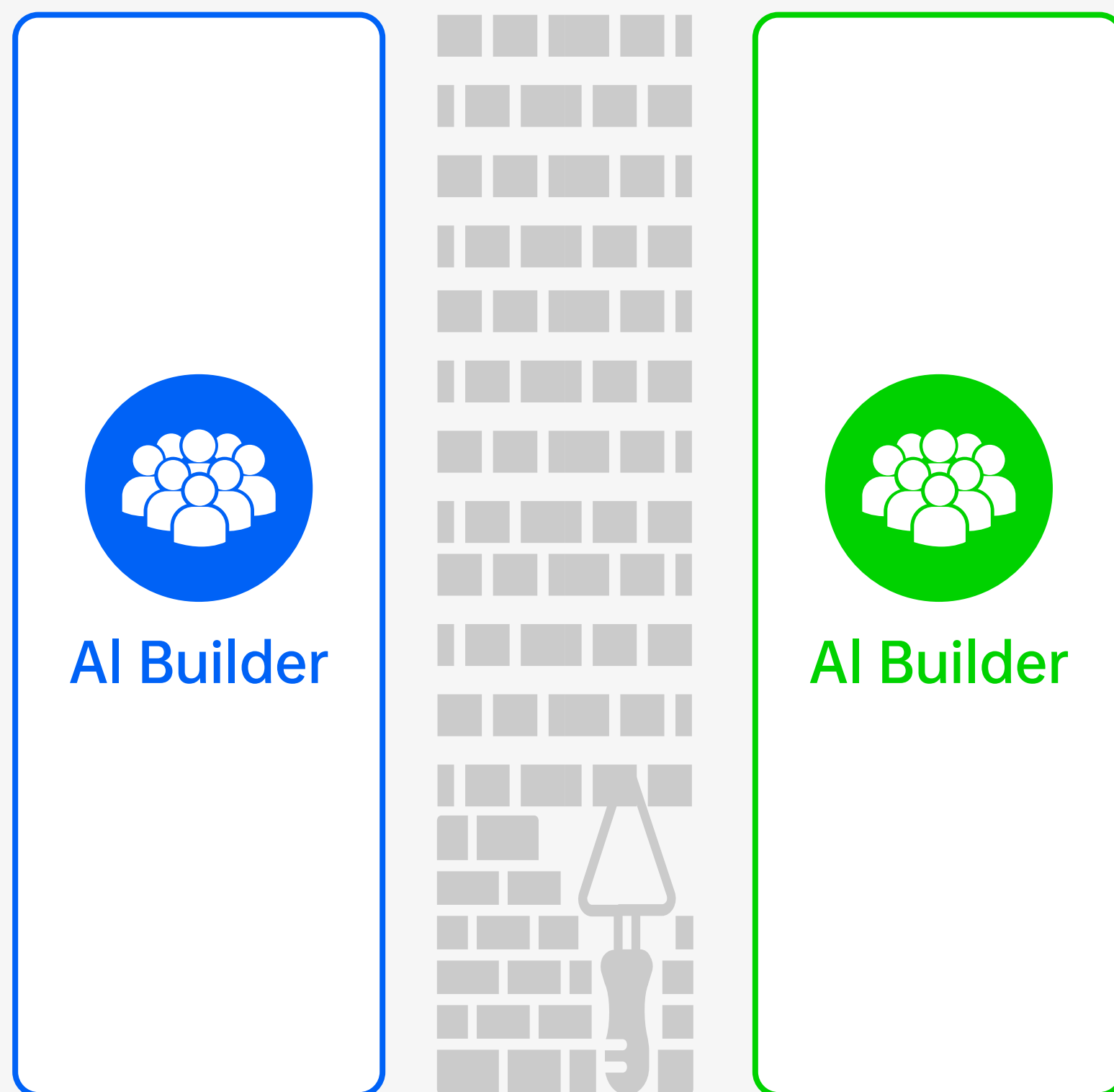
- 终端适配专项小组:**组织垂直行业工作坊，帮助终端设备明确需求，共同探讨适配语音降噪、弱网抗性和延迟优化的技术。并将相关技术提供商引入对接。
- 总结适配案例集:**发布基于 Voice Agent 的多终端适配案例集，供开发者参考学习。

提供渠道链接更精准的用户群

- 提供原型产品测试机会:**通过工作坊、体验活动、社区自媒体推荐等形式，实现 Demo 产品快速验证与快速迭代
- 用户触达新方式探索:**协调新兴流量平台和智能体平台与 Voice Agent 产品/解决方案提供商，共同探索用户触达的创新方式。
- 提供产品运营帮助:**提供场景化运营指南或陪伴活动赋能开发者快速打入分散市场。

3.6 打破短期融合障碍，社区推动实时互动智能人才交流

- 在去年社区的年度报告中，我们发现 RTE 行业本身存在着巨大的人才缺口。而根据人设部的测算，AI 开发者供求比严重失衡，预计到 2025 年人才缺口将突破 1000 万。而这两类本身在开发者中占比就较小的群体，在实时互动智能这个融合领域内，所面临的人才缺口将会更大。
- 除了面临巨大的人才缺口，RTE & AI 开发者还面临信息壁垒的挑战。这主要源于行业融合时间短、以及跨领域的技术和经验共享不足。



1. 提供交流场合

目前行业内针对 RTE 和 AI 融合的专门会议或社区较少，开发者难以获取针对性的指导和经验分享，也难以找寻到合适讨论平台

2. 整合具备易用性的工具链生态目录

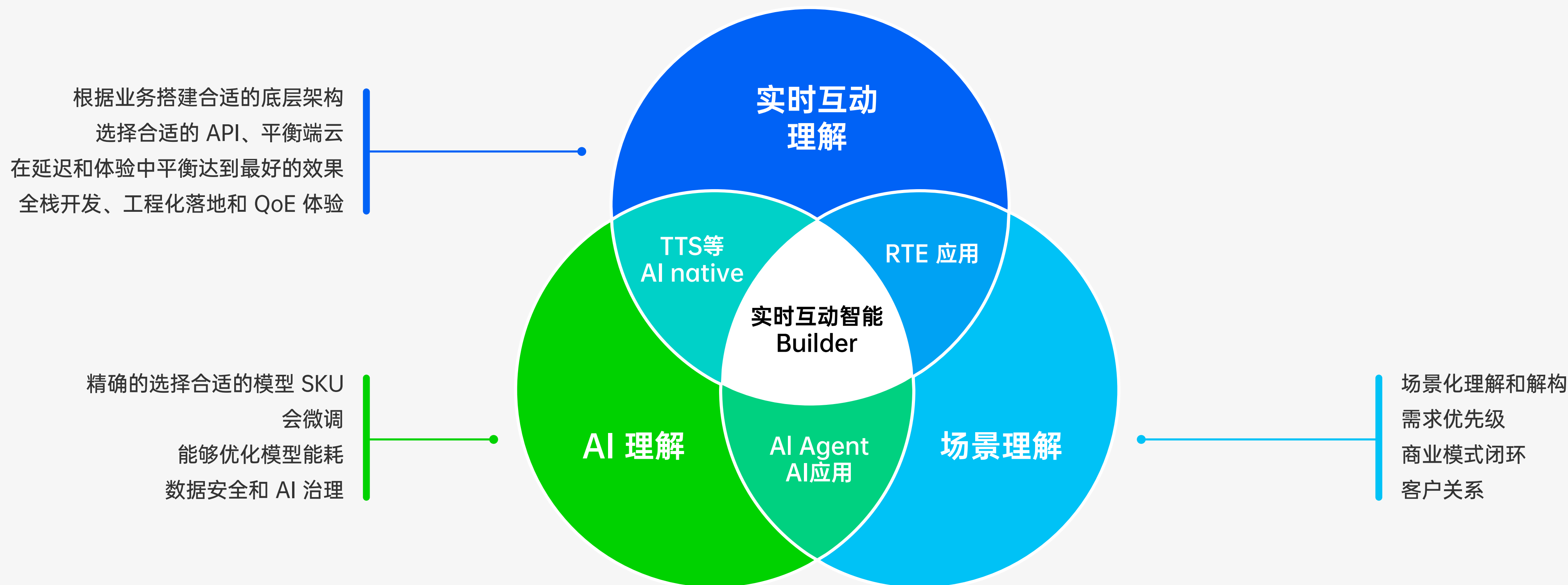
社区可以整合工具链生态，通过自媒体或使用活动推荐和完整工具链，为开发者提供使用和反馈兼备的完备工具链生态目录。

来源:《实时互动行业人才生态报告 2024 》

3.7 重塑 Builder 角色:Voice Agent 人才画像的新定义

- 除了面对算力、工具、终端和流量的挑战外，Voice Agent 的产品形态也对实时互动智能 Builder 自身的能力要求发生了转变。新时代的实时互动智能 Builder 需要既懂得选用合适的 LLM，又知道实时互动的工程化落地对于终端体验的重要性，又特别了解特定场景的实际用户痛点。

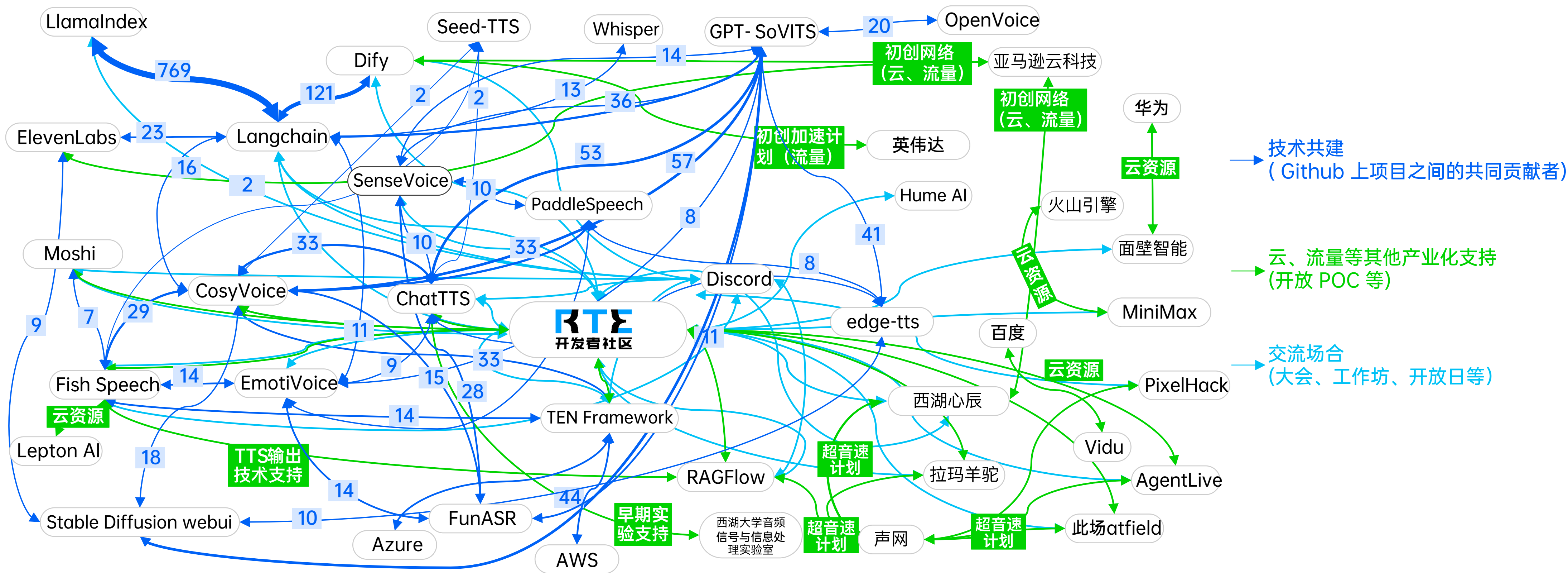
实时互动智能 Builder 的人才能力韦恩图



3.8 从社区支持到明星项目拆解：实时互动智能 Builder 需求深挖

- 除了各方合力构建新生态架构外，InfoQ 研究中心也关注实时互动智能 Builder 在社区中寻找的核心价值。为此，我们深入拆解了 Dify、CosyVoice 等一系列明星项目之间的联系，试图探究这些项目在技术共建、产业化支持和交流场合上的关键模式

实时互动智能明星项目关键模式拆解



3.9 RTE 开发者社区探索新协作模式，加速 Voice Agent 项目孵化

- 因此，除了围绕算力、工具、终端和流量的挑战提供资源整合和对接外，社区还需要加速促进人才交流与技术共建。根据这些需求RTE开发者社区也在逐渐探索生态内的新协作模型，以加速实时互动智能项目的开发和落地。



01

实时互动与 AI 行业技术演进观察与分析

02

应用场景新挑战，造就 VOICE AGENT 新的产业架构

03

RTE开发者社区架起生态桥梁，加速生态沟通

04

实时互动智能生态的未来发展预测



4.1 实时互动进化：生态协作和应用场景助推下一代多模态交互

- 实时互动智能以空间计算、空间智能为代表的空间技术，对于实时互动智能本身的互动体验、应用场景和生态协作都将带来新的机会。

1

生态协作的再构建

新玩家的深度参与

新的硬件类型、基于新硬件而产生的新交互方式

新基础设施的技术标准

构建开放、统一的空间计算技术框架

算法硬件适配

硬件和算法的互相适配，以达成场景和交互的双重赋能

2

应用场景的再拓展

更好的理解和感知物理世界

打破物理和数字空间边界

- 模拟真实环境的交互设备/流程实训
- 打破物理空间限制的创新表达
- 增强现实辅助医疗与远程手术等

3

实时互动体验的再升级

空间计算通过融合 AR、VR 等技术，借助新的设备终端，为实时互动智能提供了全新的交互维度。

代码

键盘鼠标

触控

语音

多模态

下一个交互方式

4.2 实时互动智能 Killer App 五大潜力场景

- 根据本报告中的场景选择新评价标准和开发者票选结果，选择以下五大潜实时互动智能Kiler App力场景。

	算法硬件适配	算法硬件适配
1	语音翻译/会议协作	<ul style="list-style-type: none">• 国际商务会议• 跨国团队协作• 外贸企业的实时语音订单协作
2	情感陪伴	<ul style="list-style-type: none">• 孤独人群的心理支持• 儿童早教互动
3	口语学习	<ul style="list-style-type: none">• 个人口语练习辅助• 语音陪练
4	语音客服	<ul style="list-style-type: none">• 智能营销外呼客服• 语音售后客服
5	个性化生活助理 (财务、医疗、票务、出行等)	<ul style="list-style-type: none">• 重要事项提醒和规划• 智能家居实时控制

4.3 打破多方挑战，社区支撑良性生态交流体系共建

复合型人才

- 主动进行知识分享，编写技术博客和实践指南
- 参与/牵头工具链建设，降低普通开发者的入门门槛

开发者

- 主动进行跨领域知识学习和储备
- 需要积极参与社区讨论与建设
- 需要主动参与项目，积累经验

业内企业

- 参与国家/行业相关技术标准制定
- 通过实习机会、专项培训等方式培养复合型人才
- 需要加强与社区、企业的合作



开源项目

- 完善文档和教程，提高工具和代码的可读性，并提供多样的使用案例及学习资源
- 建立开放的技术支持通道，如论坛、微信群、Slack 群组或 Discord 社区

开发者社区

- 建立专门的技术分享平台，积极推广社区成果，吸引更多开发者和企业参与
- 组织实时互动智能相关的行业峰会、研讨会黑客松和工作坊活动，促进技术交流和创新
- 整合上下游合作伙伴资源，提供更多的产业化资源和交流场合

4.4 社区助力培育兼具场景选择和场景适配的明星项目

- 除了场景选择外，能够适应场景适配也是社区参与培育的明星项目的典型特征之一。

明星项目在场景选择上的特征

高性能要求

代表应用落地在实时互动场景中运行的效率要求，受到云、边、端算力协同调度和合理架构设计的影响

场景选择

高准确要求

代表在特定场景中对模型的输出精度和判断正确性要求，受场景数据的规模和算法优化程度的共同影响

好体验要求

代表应用场景中用户对交互效果、实时性、准确性及沉浸感的感知和期待

明星项目在场景适配上的特征

高效开发适配

明星项目通常会具备良好的API设计和生态兼容性，并结合特定场景的复杂需求优化开发工具链，包括API、SDK、语音模型框架等。

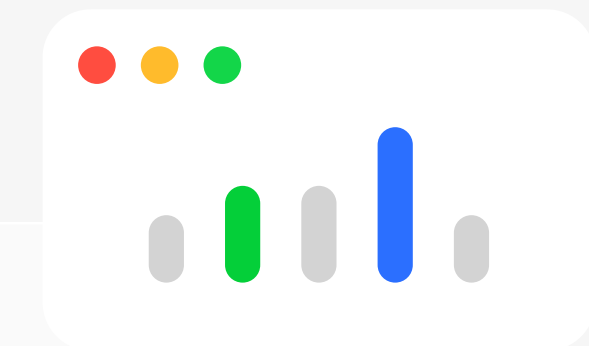
场景适配

多模态交互

明星项目将借助多模态数据(语音、文本、视觉等)的融合，提升场景理解的准确性。

包容性沟通

支持开发者与场景应用方(如企业客户普通用户)之间的反馈闭环，提升Voice Agent产品设计和服務精准度。



RTE 开发者社区由声网及多位资深实时互动社区专家联合发起，是聚焦实时互动领域的中立开发者社区。希望通过社区链接领域内的生态伙伴，激活开发者力量，萌芽更多新技术、新场景，探索实时互动领域的更多可能。

目前现任主理人团队



林旅强 (Richard Lin)

- 开源社联合创始人
- 投身开源和开发者生态领域逾 15 年
- 社区发起人: 杨慧



杜金房 (Seven Du)

- FreeSWITCH 中文社区创始人
- RTS 社区和 RTSCon 创始人
- 社区运营组: 傅丰元 孟蕾 陈韵 王尚 白宦成 魏伊培



卢恒 (Bear Lu)

- 语音 AI 资深专家
- 有声内容 AI 生成
- 大模型数据创业

45000+

RTE 领域
开发者覆盖

200+

开发者
握手次数



RTE 开发者社区通过「陪跑计划」，提供更多资源支持

- 为了更深入、更持久、更有针对性的陪伴 RTE Builder (开发者), RTE 开发者社区也发起了「陪跑计划」, 提供创业扶持、技术和家务加速、个人成长和影响力打造等资源支持。

01.技术和业务增长陪跑：超音速计划

- **创业导师 1 v 1咨询**: 针对技术型创业开发者
- 推荐加入「**超音速计划**」, 加入后可享受技术支持、投融资对接、媒体曝光、行业资源等多方位支持

02.技术和产品加速

- 来自产业生态的多项**优惠额度**: 包含云资源、AI工具免费额度、产品运营短信优惠套餐
- 行业最新开发场景 Demo 展示: RTE Openday
- 用户运营、应用市场等**产品运营建议**
- 技术共建: Github 上项目之间的共同贡献者

03.开发者个人成长

- 各类生态内的**线上线下活动**和主流技术大会的门票福利
- 社区内开发者之间的**链接交流** (RTE Meetup, RTE Hack-Day)
- 最新技术趋势洞察: 「编码人声」播客, RTE开发者日报

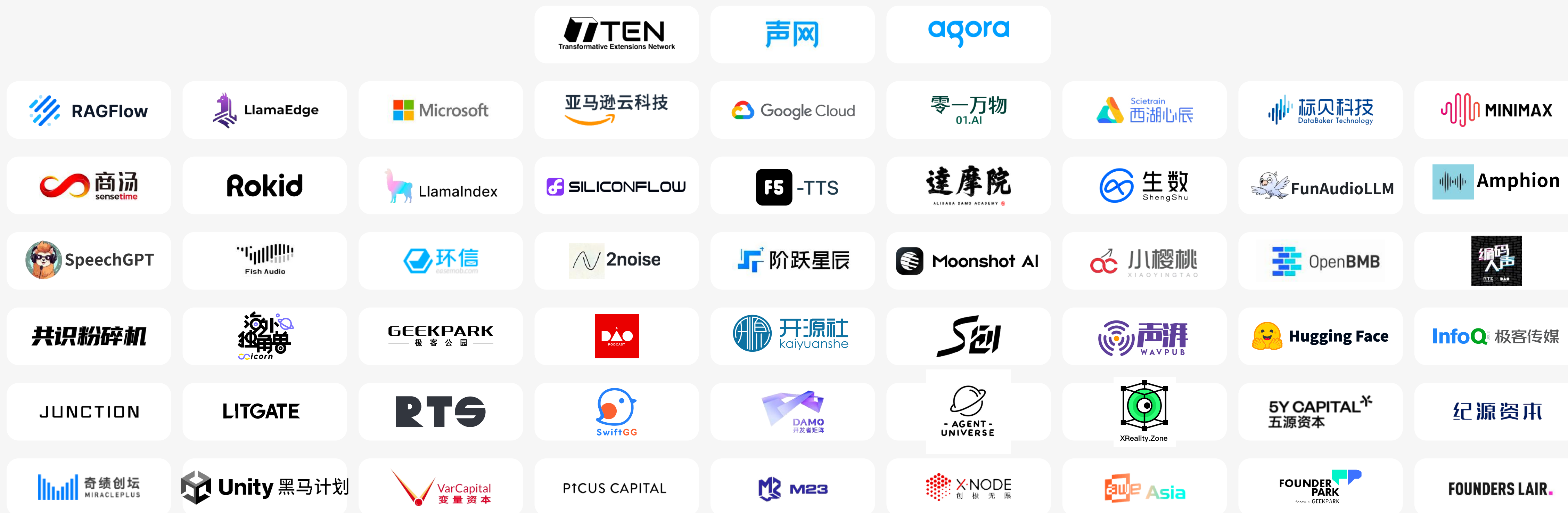
04.个人影响力打造

- 主流社区渠道曝光 (播客、采访、演讲等)
- 交流连接 (黑客松、工作坊、开放日等)



集合上下游社区伙伴

- 过去一年，这些社区伙伴通过 RTE 大会、RTE Meetup、RTE Open Day、超音速计划等活动参与到 RTE 开发者社区的共建中。(排名不分先后)





极客邦科技双数研究院

InfoQ^{ueue} 研究中心

- InfoQ 研究中心隶属于极客邦科技双数研究院，秉承客观、深度的内容原则，追求研究扎实、观点鲜明、生态互动的目标，聚焦创新技术与科技行业，围绕数字经济观察、数字人才发展进行研究。
- InfoQ 研究中心主要聚焦在前沿科技领域、数字化产业应用和数字人才三方面，旨在加速创新技术的孵化、落地与传播，服务相关产业与更广阔的市场、投资机构，C-level 人士、架构师/高阶工程师等行业观察者，为全行业架设沟通与理解的桥梁，跨越从认知到决策的信息鸿沟。

技术市场趋势洞察



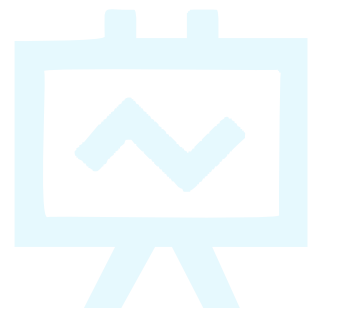
- 市场份额追踪
- 细分市场分析
- 市场规模预测
- 市场分析模型输出

技术市场趋势洞察



- 用户规模评估
- 用户认知分析
- 用户决策分析
- 用户行为分析

技术市场趋势洞察



- 技术需求洞察
- 技术实践分析
- 应用规划建议
- 发展趋势研判



内容咨询:

researchcenter@geekbang.com



商务合作:

hezuo@geekbang.com

- 极客邦科技，以“推动数字人才全面发展”为己任，致力于为技术从业者提供全面的、高质量的资讯、课程、会议、培训等服务。极客邦科技的核心是独特的专家网络和优质内容生产体系，为企业、个人提供其成功所必需的技能 and 思想。
- 极客邦科技自 2007 年开展业务至今，已建设线上全球软件开发知识与创新社区 InfoQ，发起并成立技术领导者社区 TGO 鲲鹏会，连续多年举办业界知名技术峰会（如 QCon、ArchSummit 等），自主研发数字人才在线学习产品极客时间 App，以及企业级一站式数字技术学习 SaaS 平台，在技术人群、科技驱动型企业、数字化产业当中具有广泛的影响力。
- 2022年成立双数研究院，专注于数字经济观察与数字人才发展研究，原创发布了数字人才粮仓模型，以此核心整合极客邦科技专业的优质资源，通过 KaaS 模式助力数字人才系统化学习进阶，以及企业数字人才体系搭建。
- 公司业务遍布中国大陆主要城市、港澳台地区，以及美国硅谷等。十余年间已经为全球千万技术人，数万家企业提供服务。



InfoQ

促进数字技术领域知识与创新的传播



TGO 鲲鹏会

科技领导者同侪学习社区



极客时间

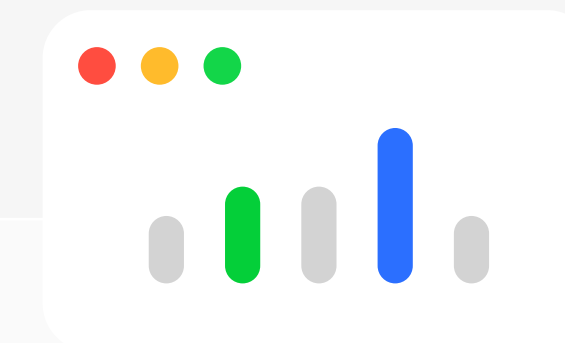
数字人才的移动知识库



极客时间 | 企业版

一站式数字技术学习 SaaS 平台

洞察技术创新趋势，推动数字化商业升级



内容咨询:

researchcenter@geekbang.com



商务合作:

hezuo@geekbang.com

●● 报告编制说明及专家团致谢

• 联合出品方

RTE 开发者社区、极客邦科技双数研究院旗下 InfoQ 研究中心

• 相关数据来源

a16z、Bessemer Venture Partners、Lightspeed、Letta、海外独角兽、Cartesia

• 专家顾问团 (按照姓名拼音顺序排序)

陈若非、冯越、halajohn、李天雨、李森、Plutoless、王铁震、武执政、俞佳

• 报告制作团队

杨慧、姜昕蔚、傅丰元、崔白洁

• 社区媒体合作伙伴

InfoQ、LitGate、语音之家、LlamaEdge、交互技术前瞻、异步社区、178A 盟