

证券代码： 002230

证券简称：科大讯飞

科大讯飞股份有限公司投资者关系活动记录表

编号：20260821

投资者关系活 动类别	☐特定对象调研 ☐分析师会议 ☐媒体采访 ☒业绩说明会 ☐新闻发布会 ☐路演活动 ☐现场参观 ☐其他
参与单位名称 及人员姓名	本次业绩说明会面向全体投资者以及关心科大讯飞和人工智能产业发展的朋友，本次业绩说明会采用现场、价值在线会议及网络直播相结合的方式举办。 通过价值在线会议接入的投资者共50名。此外，据不完全统计，通过视频直播观看的投资者，全景网15.5万人次、同花顺及万得2万人次、科大讯飞微信视频号1.8万人次。 为便于投资者阅读本次说明会的内容，参加价值在线会议的人员名单请详见文末。
时间	2026.08.21
地点及方式	科大讯飞合肥总部、全景路演、价值在线会议，以及同花顺、万得等媒体平台网络视频直播
上市公司接待 人员姓名	董事长刘庆峰先生、总裁吴晓如先生、副总裁/董事会秘书江涛先生、财务总监蔡尚女士
投资者关系活动主要内容介绍	
一、业绩解读：公司总裁吴晓如先生做 2026 年半年度业绩解读（相关内容包含会前收集的投资者普遍关心的问题）。PPT 另详见附件。 二、线上投资者实时交流：投资者通过价值在线会议提问，与公司董事长、总裁、副总裁/董秘、财务总监交流。	
问答交流纪要	

问：国产算力在智能体的强化学习训练等方面，与 H200 的采样推理效率差距有五倍，星火作为唯一坚持在国产算力上训练的主流大模型，在这种硬件的约束下，通过算法工程的补偿，或垂直场景专家思维链等技术，能做到什么水平？怎样定义 B 端客户共用的模型的能力边界，以及星火 X2 (300B) 与 GLM5.1 能力相当之后，与海外最新的如 GPT5.6，Cloud 5.0 等模型相比，差距如何收敛？

答：

一、目前国产算力差距客观存在，但对讯飞核心产业化场景影响有限

一方面，在模型训练上，由于目前国产算力存在显存带宽方面的限制，主要算力集群在采样推理效率上与 H200 相比确实存在约五倍的差距。但这一差距并不对应所有模型能力，而主要体现在超长上下文场景——即 1M 以上、或至少 256K 以上的超长上下文训练任务，如复杂智能体和代码领域的训练效率。

另一方面，在实际应用上，科大讯飞主导的业务场景，教育、医疗、汽车，央企国企行业应用等，90%甚至 95%以上的场景根本用不到这么长的超长上下文，数十万 token 的上下文长度已经足够。例如，依托软硬件一体化优势，教育业务中的自动批阅机已经形成了产品代差优势；星火-X2 大模型在 2026 年高考中取得 708 分，业界并列第一。因此，算力层面的差距，主要影响极少数的头部应用，但本质上影响的并不是人工智能产业的全部场景，对公司核心产业化场景冲击极小。

二、国产算力正快速追赶，性能差距即将收窄

目前华为 950DT 对标 H200，根据不同任务，综合性能已达其 65%-85%，扭转了此前五倍的性能差距。公司已与华为开展多轮深度研讨，将完成现有算力设备的必要迭代置换，同时成立联合攻关组前置开展技术攻关研究。因此，国产算力短板并非长期存在，预计今年底至明年，国产算力性能将显著改善。

三、算力掣肘的行业差异正在重塑，国产算力生态将迎来红利期

在国内整个生态环境中，近两年讯飞面临的一个比较大的压力，来自其他模型厂商普遍使用性能领先两代的英伟达芯片，其生态更完善，各类算法的训练与

移植效率更高。在此情况下，公司凭借讯飞研究院和生态合作伙伴的持续努力，依然保持了行业模型的领先地位，星火基座模型的基本能力大多处于行业第一梯队，部分甚至领先。随着美国对顶尖芯片出口管制越来越严格，未来两年内国内同行同样很难再采购到英伟达高端芯片，全行业都会被倒逼转向国产算力，国产算力的生态将得到极大完善。随着国产算力本身的进步、算力生态体系的完善，以及公司自身在模型尺寸选择上更加精准、聚焦解放生产力和释放想象力这类真实产业刚需，国产算力的硬件差距对公司的影响会越来越小。未来两三年内——尤其从今年底到明年开始——公司会呈现出一种苦尽甘来的态势：由于国产算力此前受到各种限制而导致的一定程度不公平竞争现状，将逐步得到改善。

四、讯飞差异化布局，深耕产业刚需，以极致性价比定义模型应用

对于模型能力对比 GPT5.6、Claude5.0 的差距如何收敛的问题，其背后反映的是中美在大模型路线选择上的差异——一种是持续追求参数规模的摸高，从几 T 到未来可能的 10T 甚至更高，在更长程任务或更复杂的超长上下文任务上确实能带来一定收益，但这些收益本身存在边际递减效应；并且这类能力所对应的场景，在工业应用、个人应用及民生服务等领域中的实际占比也非常低。因此，公司未来的主要发力方向仍然是解决社会刚需，采用云端协同——云侧 300B 至 1T 模型，搭配端侧 1B 至几 B 模型，以极致性价比服务刚需。今年"1024"节点，公司模型能力不仅在代码和智能体领域将进入第一梯队，同时 token 服务兼具竞争力与成本优势。总体来说，公司不会在参数规模上与头部海外模型“摸高”竞赛，而是算清产业层面的经济账，打赢落地战，形成更可持续的利润与现金流，待国产算力不断进化，逐渐拉平差距后，届时再考虑参与那些追求极致参数规模的长程任务。

问：目前很多家人形机器人厂商与科大讯飞深度合作，请问公司在机器人业务方向的进展和未来战略规划？

答：

讯飞机器人业务的定位是：在大模型公司中最懂具身智能，在机器人公司中

最懂大模型。

这两句话最终要落脚到——真正懂工业场景、懂产业实践，能够把账算清楚。聆动通用是讯飞将大模型能力与多模态能力应用于工业和商用领域、在集团内部孵化的控股公司。

一、锚定真实产业刚需，落地可盈利的商业化场景，数据飞轮闭环

基于自身多模态识别与交互技术优势，公司优先切入物流分拣赛道。目前已获得国内头部物流企业 PoC 订单。在十余家主流机器人厂商同台测评中独家胜出，解决真实刚需能力获得行业权威落地认可。业务刚需不仅是分拣、面单识别 99% 的高准确率，更重要的是具备强泛化能力和抓取准确率，实现工业化稳定落地。机器人部署到工业或分拣场景后，能否在一年到两年内收回整体投资，这笔账必须算得清清楚楚。该场景替代了高强度、高负荷的人工分拣岗位，面对人口结构变化，这类重复性重体力岗位机器替代趋势明确，商业逻辑清晰扎实。同时，业务落地后，真实业务流程与数据飞轮形成更紧密的耦合，持续迭代优化聆动机器人核心能力，形成正向增长闭环。

二、具备强泛化能力，筑牢规模化扩张底座

我们的 VLM、VLA 底座能力，能够让原本已经在使用的工业机械臂、机械手等自动化设备，接入后立即具备智能化能力。这种泛化能力并非行业内某一家公司天然具备，而我们已经和多家原有的物流自动化设备厂商完成了适配，充分验证了这一点。这种泛化能力，不仅支撑公司实现物流场景全国规模化覆盖、构筑行业主导优势，更为后续拓展商超等多元场景奠定基础。

三、领跑多模态交互技术，赋能全行业机器人生态

科大讯飞与宇树、优必选等国内 500 余家智能机器人公司建立了良好合作，提供多模态、远场多人高噪环境下的交互能力，这是机器人未来真正走进生活和工业场景的关键。以多模态交互为核心的讯飞机器人超脑平台，持续为全行业机器人厂商赋能，构建开放共赢的产业生态。

问：当前硬件涨价对于产品线是否有影响，如有影响，供应链何时可以改善与缓解？

答：

供应链保障和硬件成本问题，已经上升到公司战略层面。

一、以供应链安全锚定发展底线，讯飞在供应保障上具有相对优势

半年度报告中提到，公司上半年现金流原本表现良好，影响现金流的一个重要原因是提前增加了备货。如果当时没有备货，下半年公司将因为涨价多支出数亿元。

就行业层面而言，存储等元器件涨价是行业性趋势，从目前判断，未来一到两年内供需偏紧的状态可能仍将持续。对公司来说，首要任务是确保供应链的稳定与安全，尤其是在讯飞被列入美国实体清单的背景下，更需要在全栈自主可控的国产芯片及配套硬件元器件供应上做好保障。这方面，公司与多家厂商保持着良好的战略合作关系；同时，由于公司承担了重要的国家战略任务，在供应保障上也相对更有优势。此外，以存储芯片为例，长鑫存储同样是合肥本地的龙头企业，双方合作关系紧密。总体而言，现阶段的首要任务是保障供应链安全，而不是寄望于单方面压低采购价格，价格走势需跟随行业整体趋势。

二、从被动承压到主动传导：用产品代差提升产品竞争力

在保障供应链安全的前提下，公司今年上半年毛利率出现下降，在一定程度上也影响了部分产品线的推广力度。上半年智能硬件收入下滑超过 15%，原因既包括整体消费电子市场的消费降级趋势，也包括成本上涨后，公司主动放缓了部分中低端产品的推广节奏。但同行之所以陷入“涨价丢份额、不涨价亏利润”的两难，根源在于产品缺乏差异化、困在同质化红海里。讯飞的破局路径很清晰：持续强化技术刚需价值与代差优势，在此基础上合理提价，将上游成本压力有序传导至市场，而非由制造端独自吞下。产品力够硬，定价权才够稳。

以学习机业务为例：上半年学习机销售出现下滑，与此前保持 50%以上增长的态势相比，上半年还下滑了 1%，一定程度上低于市场预期；同期行业整体也出现下滑，公司部分低端产品线因供货能力不足，也主动没有加大铺货力度。但进入暑期后，公司围绕“用讯飞学习机，无需额外报班刷题”这一核心卖点展开推广，通过学习机的过程性的批改、个性化辅学能力等，向市场证明产品本身的刚

需价值，多个省市的高考状元也是学习机用户，口碑逐步扩散。7月单月SO销量实现明显反弹同比增长30%，在行业整体仍然下滑的情况下实现逆势增长，公司也相应上调了部分产品价格。

三、不与趋势为敌，用硬实力穿越周期

总的来看，公司无法改变行业整体的供需关系和成本上涨趋势，但可以通过提升具体产品的核心竞争力和性价比，来对冲消费降级和元器件涨价带来的压力，从而实现相对良性的产业发展态势。周期不可控，产品可控；趋势不可逆，竞争力可累积。这是讯飞在行业波动中保持良性发展态势的底层逻辑。

问：当前大模型市场尤其是 Coding 这一块，模型加速在收敛，市场更像是一个通缩和内卷的状态，如何看待这样的现象？新成立的词元星火公司的商业闭环方面，有哪些考虑？

答：

未来一段时间，工具类的代码能力逐步收敛，最终各家的代码能力差距不会很大，只是部分企业会通过使用更大尺寸的模型、承担更长程的任务来体现差距，这在一定程度上更多是为了回应资本市场的关注，而非完全出于客户的真实需求。

一、从 token 计费转向任务价值付费，讯飞靠底座能力+行业刚需+数据飞轮，已找到新商业模式。目前行业内关注点正在从强调 token 的用量，转向讨论任务完成后的付费方式。讯飞一直在思考，大模型时代的商业变现要如何才能打破当下这种“不断开辟出一片蓝海市场，却很快又坠入红海竞争”的内卷循环。我们认为，答案还是要落在底座模型能力和行业刚需的结合上，依靠数据飞轮在行业中形成领先一步的优势，并找到与之匹配的商业模式。这样讲可能显得比较概念化，结合几个具体例子来说明：

以教育业务的讯飞批阅机为例，软件上，我们依托模型实现了比友商精准的作业批改，不仅覆盖理科，文科主观题同样可用；并且具备能进行过程性错因分析，并据此生成个性化辅学推荐的深层能力。硬件上，我们自研适配，支持多种

规格纸张、褶皱、顺序错乱等复杂条件的自动快速扫描，友商需二三十分钟完成扫描，讯飞五分钟完成，这意味着，老师下课十分钟就可以完成全班作业和试卷的批改，小时级学情闭环，能够直接指导教学。基于软硬一体优势，我们的模式不是直接按 token 计费，而是按学校、按设备、按年收取服务费，硬件和后台服务共同构成年度服务包。目前一些优质民办学校单校采购超 20 台，在学校场景中已经成为刚需。上半年该细分市场招投标约 5000 台，讯飞份额超 90%，续费率同样达 90%。这不是卖一次硬件，而是一种全新的、闭环的商业模式。

再以医疗业务的影像云为例，运用讯飞 AI 进行影像质控和影像压缩，实现三秒调阅，效果等同于在院内阅片，单次费用约 5 元，由国家医保局承担相关费用。诊后患者管理方面，患者出院时，系统会根据其病情提供用药时间、复查安排的指引，一旦出现紧急情况，也能协助调整用药和康复方案，患者可以随时通过电话联系到主任医生——约 90% 的常规问题可由 AI 自动处理，剩余部分由高年资护士介入，必要时安排主治医师到场。这种服务模式相当于患者在出院时购买了未来三个月至半年的院外管理服务，费用从几百元到上千元不等，患者满意度从此前的 80% 多提升到 95%，部分场景达到 98%；三个月内的意外入院率从 2% 降至 0 点几个百分点，出院后并发症发生率也从原来的 3% 降至接近 0。

上述这些教育、医疗 AI 服务的本质，就是把过去纯粹按流量计费的 token 服务，转化为真正按完成任务、达到既定效果来持续收费的服务，这样才不会陷入目前商业逻辑不断被削弱的内卷局面。

在行业场景中，比如招投标类的大模型应用，讯飞采用 FDE 的方式，深入企业的招投标业务场景中联合共创，形成持续的、年年付费的合作模式。token 经济下一阶段的竞争重点，一方面在于 token 本身，当各家代码能力趋同时，就要比拼谁能在国产算力上针对不同类型的代码实现最高的推理性价比：相对简单的工业代码可能用 3B、4B 级别的模型就足够，复杂场景可能需要 300B 甚至 1T 级别的模型，通过闭源与开源模型的组合，找到效率最优的搭配方式。基于此，我们希望在软硬一体、面向智能硬件的代码开发领域做到业界第一。

所以，token 经济要从同质化的内卷竞争，转向在特定用户群和特定场景下，

基于自身已有算力和技术优势、形成极致性价比或相对比较优势的差异化竞争，即便我们身处行业第一梯队，名次或许是第三或第五，但这并不重要——别人打别人的，我们打自己的，在特定用户群体和特定场景中形成比别人更强的比较优势，并进一步深入行业，形成数据飞轮优势，让竞争对手难以复制。以医疗为例，诊后患者管理积累的数据是行业独有的：患者的用药方案、诊疗方案、回家康复情况、生命周期延续情况，以及健康状况与用药、生活习惯、气候等因素的关联，这类信息此前全球范围内都不存在，而我们通过全科整合的患者管理能力和语音能力构建起这一闭环，所积累的数据是其他厂商都不具备的。正是通过构建这样的比较优势，大模型行业才能逐步从单纯依靠算力堆叠、模型规模竞赛、资本红海投入的模式中走出来，走向一条更从容、更可持续的路径，我们会坚定地沿着这个方向推进。

二、以适中尺寸行业模型+数据反哺，讯飞走出一条与超大模型不同的性价比路径。目前 1T 以上、甚至两三万亿参数规模的模型，在解决非常复杂的问题上确实仍然有效，但在解决大量相对简单的问题时，性价比未必是最优的。随着多家模型同步发展，未来一些超大规模通用模型的训练，有可能会成为沉没成本——如果没有真正形成应用闭环，模型逐渐失去使用场景，前期投入的训练成本也就随之沉没。对讯飞而言，我们希望尽量避免出现这种情况。此外，我们观察到在一些重要行业场景中，比如今天提到的教育、医疗，以及在央国企招投标等场景中的应用，通过适当尺寸的行业模型，结合与本体知识的关联，构建出的智能体已经能够解决相当多的重要问题，并且在这种方法论下的性价比更高。公司通过这类行业场景模型，更快建立起应用的根据地。在教育、医疗这类重要场景的应用中，我们能够获得高质量的脱敏数据，这些数据对下一步反哺模型非常重要。与此同时，我们也在尺寸适中的模型上持续进行技术攻关。一方面持续推进技术攻关，另一方面在行业中积累高质量数据，两条线同步推进——就像登山选择南坡还是北坡，路径不同，但最终都是往上走。

三、出海业务正形成讯飞另一个重要的商业化出口。8 月 20 日，巴西政府宣布投入 23 亿雷亚尔加强人工智能建设。华为、科大讯飞作为仅有的两家中国

企业此次参与巴西国家级 AI 基础设施建设，这是中国人工智能出海具有标志性的一个案例。科大讯飞全栈自主可控的技术底座，叠加多语种核心优势，正构成科大讯飞大模型出海的独特竞争力。当前，人工智能出海布局已不限于巴西，在拉美、东盟、非洲、非盟及中东等多个地区均有项目稳步推进——无论是面向政府的顶层大模型合作，还是教育、医疗等民生领域的落地应用，都具备较大的商业价值。此前世界人工智能大会期间，已有 29 个国家与中国签署人工智能合作联盟协议。而会后首个引发全球关注的落地成果，便是此次巴西的人工智能建设项目。在众多中国大模型厂商中，巴西方面经过大量调研与多家产品对比后最终选择讯飞——核心原因在于，他们认为讯飞的模型能够真正赋能巴西社会、解决发展中的实际刚需，让普通民众切实感知到社会福利的改善，这本身也将成为巴西政府的一项重要政绩。类似的合作模式与进展，目前也正在多个国家同步推进。

问：公司大模型 API 调用及 MaaS 平台收入增速较快，能否介绍一下 MaaS 业务净利率情况？未来公司在这块业务上有怎样的规划？

答：

目前讯飞开放平台的 API 及 MaaS 业务扣除算力等成本后，净利率在百分之十几。整体增长态势良好，这得益于 API 经济本身是讯飞历史上持续发展较好的一项业务——今天大家看到的大语言模型能力，很大程度上延续自讯飞在语音应用领域长期积累的技术基础，多语言应用等场景持续存在刚性需求。目前受限于国产算力在超长上下文处理上的瓶颈，公司正在集中攻关，8 月 30 日将有阶段性模型版本发布，1024 开发者节将迎来更明显的能力跃升。届时智能体、代码能力，以及 token 服务的性价比都将进入国内第一梯队。目前受限于国产算力在超长上下文处理上的技术瓶颈，公司正在集中攻关，8 月 30 日前后即可看到阶段性成果。

在前述基础上，公司将进一步加大对 MaaS 经济的推广力度（上半年这块业务的推广节奏相对有所收敛），未来提升空间主要体现以下方面：

第一，是模型尺寸与行业刚需的精准匹配。不同尺寸的模型适合不同领域的

需求，这既需要行业经验的积累，也需要模型自身具备自动判断能力，从而为用户提供性价比更高的解决方案。如果所有场景都统一使用最大尺寸的模型，无异于杀鸡用牛刀，毛利率自然难以提升。

第二，是推理效率的极致优化。以语音领域为例，通过技术创新，推理效率已提升百分之八十几，相应地将成本降至原来的百分之十几，这套优化方法论也可以复用到其他模型能力上。目前国内大模型海外算力获取渠道已大幅受限，绝大部分场景都必须依托国产算力完成推理部署。在这方面，讯飞正与华为联合推进。950DT 主要用于模型训练，同时在 910B、910C 等平台上开展推理性能优化；此外也与海光等展开深度合作，双方共同组建了联合实验室，持续加大该方面的投入。

第三，在开发者生态方面，目前公司总开发者数量已超过 1150 万，其中大模型相关开发者约 323 万，既有大型企业客户，也有个人创业者，已经覆盖涵盖教育、媒体、办公、社交、文化创意等广泛的领域。在面向智能硬件和机器人的开发者群体中，讯飞开放平台的市场占有率均已超过 90%。通过讯飞提供的软硬一体化开发平台，开发者能够快速完成产品开发，其中机器人领域，讯飞机器人超脑平台已服务了 90% 的机器人厂商。这些都是公司下一步在 coding 能力上可以形成差异化优势、进一步帮助开发者提升效率的重点方向。

综合来看，模型能力本身的迭代、面向应用场景的模型组合策略，以及在国产算力上的推理效率适配优化，三者共同构成了公司在 API 和 Mass 业务未来的核心竞争优势。

问：目前公司大模型相关项目的中标数量和金额持续领先行业，随着大模型在企业端的渗透逐渐进入深水区，公司层面是否感知到，在边际影响上，我们在央国企客户中的入口价值有所提升？以及企业级应用交付的时间周期和毛利率相关情况。

答：

第一，模型定价持续收窄，价值转向深度绑定与数据治理。今年以来，我们观察到两个比较明显的趋势。第一个趋势是，底座模型本身的定价空间正在持续

收窄。由于开源模型的普及，即便部分场景采用开源与闭源结合、通过模型路由整合形成解决方案，模型本身的价格也很难维持在过去数千万甚至上亿元的水平，市场上模型能力本身已经相对容易以较低成本获取，这一点行业内已形成共识。但正如大家都在关注“谁是中国的 Palantir”这个问题一样，前面的问题也提到 FDE 模式——真正的价值在于谁能够依托模型能力，深入帮助央国企和行业解决真实刚需，与客户共同完成更深度的数据治理，并通过 AI Collection 这类能力，把企业内部原本互不连通的系统联动起来、被模型理解，使原本割裂的系统能够作为一个整体运行，这类应用未来会越来越多。上半年，公司在大模型相关的企业级招标采购项目上虽然仍保持行业第一，但总量较去年有所下滑，主要原因是去年存在个别体量较大的国家级项目，剔除该因素后，整体口径下降幅度为个位数百分比。这在一定程度上反映了客观情况：今年上半年市场变化速度较快，行业信息更新频繁，不少客户在决策和投入上处于观望阶段。但目前趋势正在逐渐清晰——客户对于如何建立自身的智能体平台、如何沉淀形成内部数据闭环并持续优化的场景，付费意愿明确。

第二，讯飞已形成卡位优势，并有可复制验证案例。过去几年，讯飞在这一领域的市场份额持续保持行业第一，已经在与央国企合作的底座平台建设中占据了重要的生态位置。在此基础上，凭借底座平台自主可控、叠加开源模型路由能力，我们具备天然优势，并已经开始在客户内部推进数据治理工作，形成了良好的合作基础，客户复购率维持在较高水平。进一步来看，我们发现越来越多的智能体和深度应用场景具备可复制、可推广性。以招投标场景为例，国家八部委已联合发文，要求到明年底前，不仅是央国企，各级地方政府也需要运用人工智能能力来减少招标过程中的不规范行为，营造风清气正的社会采购环境。在这一方向上，国家发改委目前给予专项支持的案例，正是科大讯飞在合肥落地的项目，国资委官网也对该项目进行了推荐，国家监察委员会与中央纪委也对我们在国家能源集团落地的这一模式给予了肯定。这也意味着，这一在国能项目中验证成熟的产品，具备向全行业推广复制的基础——通常一个客户场景可能是投入大于产出的，做到三到五家客户基本可以做到收支平衡，推广到二十家左右就能实现盈

利。随着与更多央企及行业客户的合作深入，大家也逐渐从对模型能力本身的焦虑和追捧中回归理性，重新聚焦到真正的价值创造上——即哪些应用真正有助于企业的产业竞争力和经营效益提升，客户在决策时也更加沉稳。这也使得公司此前的战略布局，与当前模型尺寸选择、应用场景逻辑之间的匹配度在逐步提升。

第三，公司交付效率与毛利率同步改善。交付层面也体现出同样的趋势。过去部分项目的交付周期可能长达半年甚至更久，目前公司已经越来越少采用开口合同，转而在签约阶段就把合同范围和质量界定得更清晰，交付效率随之提升，交付完成后的利润率也呈现改善趋势。今年我们在这类项目上的“交付费毛比”——也就是每交付1元毛利，大致需要投入多少费用——今年出现了明显下降。下降的原因主要有两个方面。第一，公司内部沉淀了一批工具，可以在不同项目间复用，比如 AI Collection：过去每对接一次企业内部的 CRM 等业务系统，都需要投入定制化开发工作量；现在这部分能力已经沉淀为标准化工具，只需要做配置即可完成对接，成本因此明显下降。第二，部分行业知识本身具有较高的复用价值，以能源行业招标为例，不同物资品类背后有其自身规律，我们通过知识图谱的方式将这些规律梳理出来——第一次为某个企业梳理某一品类的知识图谱需要投入较多时间，但完成之后，再向同行业其他企业推广复制时，边际成本会大幅下降。综合来看，随着内部 AI 工具的持续积累，以及行业本体知识图谱的不断完善，整体毛利率呈现持续提升的趋势。不过也需要说明，今年硬件、存储等相关成本的上涨，在一定程度上抵消了交付费用下降带来的部分直接利润率改善。

问：近年来公司在国产算力领域持续加大投资与合作力度，与华为昇腾、海光信息、寒武纪等均有较为深入前瞻的布局。请问公司 2026 年和 2027 年的资本开支规划大致是怎样的，其中用于采购和使用 AI 算力的部分，具体安排如何？另外，由于目前涉及的芯片架构和种类较多，在这种多元异构的情况下，公司如何保障未来训练和推理效率？

答：

首先需要说明的是，公司目前的算力来源有几种，有自有资金投入的算力，

还有承担国家相关任务所获得的算力等，因此严格意义上，公司真正的资本开支所对应的算力获取总量，与公司实际可使用的算力总量，并不完全一致。

一、算力布局更重结构效率：今年上半年公司研发投入约 30 亿元，同比增长 6 亿元，研发投入占比由去年的 20% 提升至 26%，其中部分增长来自算力投入增加。整体来看，下半年研发投入预计约为上半年的两倍略多（这与公司下半年收入通常高于上半年、相应投入力度也略高的季节性特征一致）。其中约一半用于大语言模型及语音、多模态等方向研发，涵盖自建算力折旧摊销与租赁支出。基于当前相关贷款利率已降至 1% 左右，自建与租赁成本已趋近，公司下一阶段将更倾向以租赁为主，联合政府及厂商共建基础设施并承租使用，而非大规模一次性采购。同时，国家层面举国体制优势亦为公司提供算力支持。从中巴合作等案例可见，中美在 AI 领域的国际影响力竞争日益激烈，公司承担重要战略任务，国家亦将在现有及未来算力资源上持续给予支持。

公司在今年五一战略务虚会后，集团上下已形成高度共识，将星火大模型相关业务独立运作，采用端到端的极致效率和完全扁平化的管理模式——目前该公司 300 余人，仅设 CEO 和 CTO 两位正式任命的管理者。这一平台一方面用于加快端到端的市场竞争与研发效率，另一方面也将引入外部融资，用于持续推动大模型相关战略的发展。但即便引入了外部融资，公司也不会再在算力投入上不计成本地去“摸高”。

二、多元异构布局是讯飞的重要优势之一。公司在与大量产业客户合作中，面对客户算力设备来源多元的客观现实——华为、海光、寒武纪乃至英伟达并存——如何在这一异构环境中实现高效适配与统一调度，已成为公司的核心竞争壁垒之一。未来两年模型能力供给将面临全社会需求的爆发式增长，国内外极有可能继续出现明显的算力吃紧的局面。届时，能否获取充足算力，将是决定能否对外服务、进而谈性价比的前提。在算力供给方面，讯飞既具备承担国家战略任务的优势，也有与华为、海光的深度合作，以及公司作为寒武纪早期股东在供应上的优势。更重要的是，公司不仅仅停留在现有芯片的推理优化层面，而是在合作伙伴下一代芯片设计阶段即深度参与联合研发——这一机制在与华为、海

光的合作中均已建立。综合来看，这些因素共同构成了公司未来在多元异构算力平台上、面向 token 经济时代的核心优势。

问：国务院此前发布了人工智能+行动，整个社会氛围都在强调 AI 应用，从公司自身感受来看，目前 B 端业务的景气度如何？

答：人工智能在 B 端业务的景气度，在公司教育、医疗、产业应用等业务领域，均进入到关键阶段。

教育领域，教学精准化与学习个性化已成为行业重点建设目标，讯飞覆盖指标加速跃升。无论从素质教育还是学业发展，学校、家庭和老师最关注的仍然是教学场景的数字化问题，尤其是作业环节的数字化，此前该方面的深度应用主要集中在少数头部学校。但从近期几次重要会议来看，包括在中国召开的世界教育数字大会，以及国内多个重要教育论坛，教学精准化和学习个性化已经被明确列为未来最重要的建设目标。在国家政策引导下，虽然部分区域财政资金供给出现一定紧张，但很多地方仍然优先挤出经费，学校也把相关项目作为当前最重要的建设方向。从需求端可以看到，公司相关业务的合同额增长明显，包括完全通过校内资金采购的英语听说考试、批阅机等相关产品，增速都较快。用两个指标来观察：一是作业覆盖人数，一季度末覆盖 200 多万人，三个月后增长到 381 万，增幅 56%，今年底目标是 973 万，相当于每个季度需要保持 55% 以上的增长。二是批阅机业务，这是软硬件一体化形成的个性化作业与学情分析产品，公司计划两年内完成对 3 万所头部学校的覆盖。之所以能设定这一目标，是因为 8 月初启动专项债反映出各学校存在迫切需求——大家普遍认为，人工智能时代赋能因材施教和学生综合素质发展已经到了关键时点，教育部在教育强国纲要中明确了相关行动方向，并通过世界数字教育大会向全球推荐相关方案，讯飞正协助越南、泰国、印尼以及巴西等国家推进人工智能+教育的国家级方案。

医疗领域，基层辅助诊断与影像云业务全面提速。国家卫健委在“十五五”医疗发展规划中明确提出，将基层辅助诊断能力的全覆盖作为下一阶段目标——此前基层数字化能力相对薄弱，这项工作过去并未被明确提出。这说明医疗

AI 应用已经从早期个别区域的自发探索，转变为全行业层面形成的共识。近期相关部门在合肥召开现场会，各省实地考察了讯飞在医院和社区的落地情况后，更是将原定 2030 年实现全覆盖的目标提前到 2028 年。目前公司已覆盖 800 个县区，剩余约 2000 个县区，讯飞目标是获得其中 80% 的份额。影像云方面，安徽是最早落地的省份，今年上半年广西也已落地，按照目前的招标进度，预计到今年年底覆盖省份将超过 10 个，讯飞在这些省份的份额预计将超过 60%。目前基础的影像阅片收费为每次 5 元，未来还将延伸到泛影像云和 AI 复诊等场景，覆盖用户数将从此前每年约 3000 万，到今年年底突破 1 亿，两年内有望超过 3 亿。

在产业领域，模型带动 AIcar 需求旺盛，开发者生态快速增长。以汽车行业为例，大模型的出现带动了智能座舱对多模态、端到端大语言模型理解能力的需求——车辆变得更懂用户、交互更生动、也更安全，能更好满足个性化刚需。今年上半年，尽管国内汽车整体销量有所下滑，讯飞汽车相关收入仍实现 20% 的增长，定点数增长超过 30%，新增定点车型的单车平均销售额相比过去提升了数倍，部分车型达到两倍以上，部分达到五倍，平均约 3 倍多。开发者生态方面，去年 7 月到今年 7 月，公司开发者团队数量增长 111%，其中约一半是企业开发者、一半是个人开发者。

上述数据组合起来，基本反映了当前整体行业的景气程度。

景气度转化为收入利润，讯飞商业模式跑通成型。景气度转化为收入利润，关键在于清晰付费通道：一是国家资金保障类（如医保明确支持的医疗服务）；二是学校专项经费采购类（如批阅机约 5 万元/学期/台，学校按规模采购数至数十台）；三是消费者自愿付费类（如作业辅学、校内科学教育、院后患者管理等，通常以县区为单位或直接面向消费者）。这些付费通道要么有明确的资金来源支撑规模化增长，要么是个人自愿付费且逻辑清晰。今年上半年最重要的工作就是把这几类商业模式跑通、把收费逻辑理清楚，不再单纯受限于政府财政投入的紧张程度，并在这些服务中构建起明显的代差优势，这也是下一步的重点增量方向。

问：讯飞在 C 端业务上布局较早、也比较前瞻，公司如何看待未来 C 端产品格局以及相应的战略布局？

答：关于 C 端业务，公司高度重视交互感知能力较强的产品形态。坦率地说，目前公司相关产品在应用效果和性能上仍有较大提升空间，今年下半年到 1024 前后，讯飞 AI 学习机的答疑辅学能力、智能眼镜在交互翻译和会议纪要等场景的表现都会有明显提升。如果一个产品品类的价值发生实质性变化，用户的购买意愿就会被有效激发；公司观察到，端侧模型、多模态等技术一旦应用到眼镜、学习机、办公本这类关键产品上，能够带来价值层面的显著突破——一旦价值被充分释放，对用户购买意愿和相关收入的提升都会有明显带动，因为真正推动行业增长和用户购买决策的，最终还是产品本身的价值。

具体到智能眼镜这款产品上，翻译功能本身是眼镜类产品的一个基础刚需，但更重要的是公司具备的多模态抗噪能力——在公共场所等有背景噪声、多人说话的复杂环境下依然能够准确识别，并进一步与多模态识别能力结合，再叠加最适合在眼镜这类只有约六行显示空间的设备上呈现的模型适配能力，只有真正懂模型的团队才能做好这类适配。基于软硬一体化优势，公司做出了目前业界最轻的智能眼镜，在上海世界人工智能大会展馆的六件镇馆之宝中排名第一，而在众多参展的产品中，讯飞眼镜是具备最强多模态抗噪能力、同时也是最轻便产品的代表。整体来看，C 端业务的方向仍然是聚焦强需求、商业逻辑已经验证的场景。

问：关于研发、管理、销售这三项费用，公司未来的趋势判断是怎样的？

答：

研发费用：预计还会保持增长，与今年上半年占比从去年同期的 20%提升到 26%的趋势基本一致，预计后续会维持在这一水平附近。公司也会把各业务板块——教育、医疗、海外、消费者等板块，以及独立运作、可单独融资的星火大模型板块——更清晰地 toward 投资人呈现，便于大家看清楚，在同行动辄投入一两百亿的情况下，公司投入二三十亿的竞争力体现在哪里、商业逻辑是什么。需要说明的是，研发费用会有所增长，但不会出现不计成本的大规模算力建设、去做那些看不到明确商业价值的“摸高”投入，这也是各企业根据自身定位和发展阶段

做出的不同选择。

管理费用：预计将出现明显下降，目前公司内部 AI 提效已经从研发延伸到各类管理环节，研发效率本身也在大幅提升，“超级团队”和“超级个体”的模式正在讯飞内部逐步显现并深入人心。

市场推广费用：今年下半年会有所增加，但增加的并非广告投放这类支出，必要的品牌宣传会维持在合理水平，更主要的投入方向是此前提到的场景布局，比如批阅机业务两年内要完成对 3 万所学校的覆盖，需要先期完成学校端的部署安装，前期成本投入较大，不过很多情况下渠道合作伙伴也会提前垫付相关费用。医疗方面，诊后管理业务的月收入从年初约 200 万元，到目前已经提升到 700 多万元，年内有望突破 1000 万元，但这并非最终目标，目前突破千万收入的仅是十几到二十家医院。在能够实现全科室整合化管理的领域，目前全国仅讯飞一家具备这一能力，公司的目标是覆盖 93 家头部医院，市场份额 60%以上力争超过 80%，同时覆盖 400 家地市级头部医院。这意味着未来一到两年内，尤其是今年年底到明年上半年，需要优先完成产品部署，再逐步深化合作、面向患者提供服务，进而形成收益，这部分市场投入的商业逻辑已经得到验证，公司会向投资人清晰披露相关节奏。此外，部分海外市场投入也会相应增加，比如与华为共同推进的巴西国家级大模型合作，目标是让大模型服务于该国的公共基础设施和关键战略领域，并逐步延伸到民生和产业层面，为公司在当地形成更大的产业根据地；后续在其他国家落地类似项目后，也会相应增加布局投入。

总体而言，管理费用占比会持续下降，研发费用会保持一定增长，市场费用则会围绕目标明确、收入来源清晰的场景有节奏地增加。

问：公司业务在向央国企等企业端渗透的过程中，是否感受到企业端场景更加复杂，包括整个 workflow 的构建也需要更多人工参与，因此渗透速度可能会比 coding 类应用更慢，公司在这方面有哪些具体观察，又有哪些方式可以提升渗透效率？

答：首先，随着应用场景在央国企中的不断深化，相关场景确实变得越来越

复杂，这是客观事实，这也是我们越来越重视 FDE 模式的原因。

第一，在场景选择上，公司在央国企业务中并非所有场景都会涉入，而是聚焦于未来具备大规模复制潜力的核心场景，比如公司此前在客服领域积累的传统优势，以及供应链、招采以及目前在银行风控等领域的应用——这类场景一旦在单一客户处跑通，向其他客户复制的难度相对较低。因此，场景选择的第一个原则，是优先选择未来规模空间较大、可复制性较强的场景。

第二，在确定关键场景后，公司内部也在持续培育既懂模型、又懂行业本体的复合型人才，梳理形成具备可泛化能力的行业知识体系。以国能集团的招采项目为例，我们按品类梳理形成的知识体系，并不只局限于国能集团内部使用，目前已经拓展到几十家客户，尤其是在能源类企业中，此前在单一项目上投入的大量成本可以被迅速摊薄，整体项目的交付费率也随之大幅下降。这也说明，在行业本体层面沉淀的知识体系，能够在行业内泛化复用中获得更多应用场景。

第三，在与行业业务系统对接的过程中，公司也沉淀出了一批可复用的工具，比如 AI Collection——由于 AI 的介入，这类对接方式本身具备更强的泛化能力，使得公司在深入推进央国企业务的过程中，整体效率也相应提升。所以，虽然企业端场景确实更加复杂，公司应对的方式，一方面是聚焦更关键的场景，同时在与央国企对接过程中，帮助客户培养更多团队去承接更丰富、更细化的场景需求；另一方面则是持续沉淀工具、梳理行业知识库。

支撑上面这几个动作的，其实是一套更根本的判断：企业端场景在不断深化的同时，客户需求也变得越来越明确，不再像过去那样笼统地希望应用大模型、却说不清具体要解决什么问题，或者仅仅停留在“提效”这样一个模糊概念上。基于这个判断，公司在方法论上把握三个关键点。一是要在央国企及重点行业已经积累的良好客户基础上，提供一个能够与客户共创的平台——这个平台需要具备模型路由能力、数据清洗与自训练能力、跨内部系统的数据链路打通能力（也就是通过通用模型自动实现 AI Collection 的对接），以及异构算力的统一调度能力，只有搭建起这样的通用平台，团队才能真正与客户实现共创，让 FDE 模式能够更深入地推进下去，这也是我们判断未来续约率会保持较高水平的原因。二

是要在行业纵深上聚焦重点方向发力，比如能源、汽车、金融等行业。三是要在这个过程中不断提炼可以在全行业乃至泛行业范围内推广复制的智能体和应用场景，比如招投标、风控这类典型应用，这是接下来的整体策略。

问：随着 AI 应用和 Harness 相关能力越来越受到关注，公司本身也具备较好的流量入口优势，未来在应用层和 Harness 层会有哪些布局，如何发挥自身优势？

答：关于应用层与 Harness 层的布局，公司高度重视模型能力与 Harness 工程的结合。单纯依靠模型本身，在实际应用中会出现各种各样的问题，如何把模型能力和行业场景更好地耦合起来，需要依托 Harness 工程——也就是通过 AI 应用平台层面的一系列约束和规范来实现，目前公司内部在这方面的工具也越来越丰富，使得相关对接速度进一步加快。Harness 的价值不仅在于让模型真正在场景中落地，更重要的是，在一些敏感或高风险领域，能够有效降低模型幻觉、使其更稳定地达到专家级水平，比如诊疗领域、教育领域等。

在发挥自身优势方面，尽管公司在国产算力上受到不少限制，在底层算法层面仍在持续进行创新探索。以今年高考为例，在广东高考理科测试中，公司模型与 Claude 并列取得 708 分的成绩，显著领先于国内多个知名模型——部分知名模型成绩约在 620 分左右。这背后涉及的是如何通过注意力机制提升多模态模型的效率和路径选择，以及如何在复杂推理过程中更好地结合评估模型等底层技术问题，公司在这些方向上都在扎实推进，既包括与华为、海光在推理性能和训练算子优化上的合作，也包括在模型算法层面的多项独特创新，这些积累未来会共同构成科大讯飞的核心优势。

日期	2026.08.21
----	------------

附：线上会议接入人员名单

中信证券 潘儒琛、中泰证券 王雪晴、华创证券 吴鸣远、华泰证券 郭雅丽、国泰海通 杨林、国海证券 刘熹、中金公司 袁佳妮、长江证券 郭敬超、国泰海通 刘天雅、兴业证券 陈鑫、国开金融 孟令磊、北汽产投 蔡梦然、上海鑫义创投 陆懿晨、中国经营报 陈佳岚、中金公司 农宇华、每日经济新闻 张宝莲、北工投资 王淑琴、申万宏源 洪依真、太平洋资管 陶韞琦、国开金融 孙超、高盛亚洲

宋婷、中信里昂 王泽阳、时代周报 朱成呈、创金合信基金 王先伟、仁桥资产 钟昕、中国基金报 牛思若、鑫融长弘资管 戴溪、国智投 赵紫盈、国开金融 李岱芸、华强集团 钟源、东兴证券 郑及游、广东润融私募基金 陈可峰、广东润融私募基金 江晖、環球電信資本 李一鳴、光大证券 刘勇、法巴证券 迟琛、武汉证国私募基金 陈阜东、德摩资本 赵妍洁、西部证券 周成、国元证券 李典、耕霖资管 莫莉、聆泽私募 翟云龙、中信里昂 Maggie Ye、广发证券 周源、国元证券 耿军军、国投证券 夏瀛韬、平安证券 闫磊、汇丰前海 张恒、平安证券 黄韦涵、汇丰前海 刘逸然、中信里昂 王泽阳、中泰证券 刘一哲、山证资管 邬敬业、广东比邻投资基金 赖昕、刘旭康