端侧AI上车，从何“下手”？

图为面壁智能在2026北京国际汽车展览会上发布的端侧智能座舱升级版方案

本报记者 杨鹏岳

2026北京国际汽车展览会期间，汽车端侧AI技术迎来从概念验证到量产落地的关键转折。从芯片、端侧大模型到智能座舱方案，产业链各环节集中展示了端侧AI驱动汽车智能化的最新成果。然而，炫酷的演示背后，行业正集体陷入“交互范式仍停留在触屏时代”“端侧模型参数有限却要读懂人情世故”“换车如换秘书，体验无法迁移”等深水区难题。

端侧AI上车，难点在哪？在车展同期举办的汽车行业端侧AI专题研讨会上，多位行业专家与企业家就此展开讨论。专家指出：智能座舱正成为AGI时代定义“AI Car”的核心界面，而端侧AI是实现隐私保护、瞬时响应与断网可用的“雪中送炭”技术。不过，热闹之下，从“可用”走向“好用”，端侧AI上车的考验才刚刚开始。

端侧AI驱动“智能座舱”变革

当前，汽车行业对“智能化”的理解，往往混用智能驾驶与智能座舱，但二者解决的问题并不相同。相较于仍在法规与技术双重约束下稳步推进的智能驾驶，智能座舱凭借更直接的用户感知、更成熟的端侧技术与更灵活的场景创新，成为车企差异化竞争的关键抓手，而端侧AI则是驱动这场变革的核心引擎。

“智能驾驶和智能座舱是两件事，不能割裂来看，但座舱的发展在当前阶段显得更加重要。”清华大学人工智能研究院常务副院长孙茂松给出了一个颇为诗意的定位——智能座舱应成为“一个伴你走天涯的灵动之家”。他用《中庸》里“致广大而尽精微”来进一步阐释：大模型追求规模扩张是“致广大”，是少数大厂的主战场；而对广大产业界而言，深耕垂直领域、把事做透的“尽精微”，才是更具现实意义的路径。智能座舱，恰恰就是这样一个“尽精微”的典型场景，正当其时。

如何定义AI驱动下的下一代汽车？中国汽车工程学会副秘书长郑亚莉给出的定义是“AI Car”——兼具“专业可靠的司机”与“聪明温暖的伙伴”双重角色的智能体。在这一定义中，座舱智能体被明确为“未来人车交互的唯一窗口”，是串联智驾智能体、底盘智能体、动力智能体的自然交互中枢。郑亚莉指出，AI Car的关键特征在于自主性、交互性和适应性，而座舱智能体正是实现跨域融合、从功能集成走向智能协同的核心纽带。

2025年，我国新能源汽车销量渗透率已接近50%，汽车行业就此进入智能化“下半场”。专家指出，智驾已在行业和用户层面达成共识，而智舱则呈现出另一番格局——不同车企技术路线各异，用户需求也高度分化：有人要效率，有人要陪伴，有人要照顾全家出行，有人只希望系统“默默理解我的意图”。

当前座舱正从基于命令的交互，迈向基于AGI变革的更高维度：自然语言将成为综合调用整车能力、连接外部端口的入口，并实现长期记忆。然而，理想图景之下，现实的产品体验仍有明显落差。清华大学车辆与运载学院博士后帅斌从用户视角指出，当前座舱远远不是“真智能”：功能使用率不高，真正有用的功能仍需不断开发；每次坐进车里面对摄像头和语音的采集，用户对数据去向心存疑虑；更关键的是，不同品牌汽车的座舱体验无法平滑迁移，缺乏可延续、可进化的能力。

“现在绝大多数的智能座舱产品可能都不属于‘智能座舱’。”梧桐科技CEO曹斌表示，当前座舱系统大多建立在为手持触屏设备开发的操作系统之上，交互模式需要驾驶者将注意力放在屏幕上点击操作，这种模式在车里其实很不自然、不友好。只是受限于现有技术条件，这仍是目前最主流的产品形态。而大模型的出现，或许能让座舱从这种“过时的”状态中跃迁出来，回归到符合人类直觉的交互模式。

在多位专家看来，智能座舱在汽车智能化中已确立其战略中枢地位，是AGI时代人车关系的核心界面。但当前产品仍处于从“被动命令执行”向“主动智能伙伴”跃迁的早期阶段，面临的挑战集中在三方面：一是交互范式陈旧，亟待从触屏点击走向自然语言与多模态直觉交互；二是用户隐私顾虑和体验可迁移性问题尚未解决；三是开发模式需要重构，从规则驱动转向AI驱动。上汽大众智能座舱总监朱丽敏提出，这不是在原有框架上做加法，而是一次范式级的改造。

“AI Car”是兼具“专业可靠的司机”与“聪明温暖的伙伴”的双重角色智能体。

端侧AI上车面临多重难点

“没有端侧大模型，这件事要打一个大大的问号。”谈及智能座舱的发展时，孙茂松为端侧AI的不可替代性定下基调。他所依据的，是驾驶环境三项基本要求：第一，断网可用——隧道、偏远地区不能中断服务，只有端侧模型能保障；第二，隐私不出车——声纹、人脸等敏感数据必须留在车内，这决定了云端方案天然受限；第三，瞬时响应——“请跟上那辆车”这样的指令，云端无法保证能毫秒级反应。据此，孙茂松提出“宜端则端，宜云则云”的协作原则，更主张“以端为基，以云为缘”——雪中送炭靠端侧，锦上添花可交给云端。

与手机、笔记本电脑等终端相比，汽车为端侧AI提供了相对充裕的算力环境，可以容纳更大体量的模型，这是有利条件。但端侧模型面临的天然难题——“器小易盈”：参数规模有限，能力天花板明显。车规级芯片上能跑的多是数十亿参数级别的模型，与云端万亿参数大模型相比差距悬殊。

因此，端侧模型必须做到“小而强且美”，既要有足够能力覆盖面，又不能牺牲用户体验感。所幸新技术规律正在提供有力支撑——密度定律是让新范式在车端落地的“先决条件和基础”，帮助逾越此前看上去不可逾越的算力障碍。

清华大学计算机系教授、面壁智能首席科学家刘知远分享了团队的“密度定律”发现：模型的能力密度大约每3.5个月翻一番，无须扩大参数规模即可实现能力倍增，增速远超摩尔定律（18个月翻一番）。

虽然方向很明确，但端侧AI上车仍面临从技术到产业的多重难点。

首先，模型“小而强”的技术极限。刘知远将未来大模型趋势归纳为三个方向：输入端全模态、模型本身极致高效、输出端更自主。在车上，感知不能只靠视觉或语音，未来还要融合触觉及各类感知器件的信息，构建物理世界的全模态认知。

与此同时，必须在固定的模型尺寸内持续压缩更强智能。这要求围绕“智能的容器”（模型架构）、“学习的教材”（数据精致度）、“模型风洞”（训练规律把握）和“软硬协同”四个要素不断迭代前沿技术。据刘知远判断：“未来五到十年，但凡能训练出来的大模型，经过两年时间就可以让它运行在终端芯片上。”但孙茂松也提醒，从“可用”到“易用”再到“好用”，每个台阶都有相当难度，仍需持续努力。

其次，安全与隐私的信任构建，远不止“数据不出车”。中国汽车工业协会副总工程师王耀指出，汽车行业对数据安全的监管要求比多数行业更严，没有端侧处理很多功能都无法落地。但仅做到“数据不传出去”还不够，他提出一个更细腻的问题：车内朋友间的玩笑话，端侧模型能不能判断出这是开玩笑而非真实指令？这涉及对上下文的理解、用户习惯的学习和知识库的建立，对参数规模有限的端侧模型而言，单靠模型本身难以完全解决，需要在工程化层面做大量精细设计与处理。

此外，中金资本董事总经理徐萌萌则强调交互设计层面的“信任可视化”——让用户在隐私数据被调用时获得清晰提示和授权感知，实现“我的数据我做主”。

最后是产业链协同与标准化。多位专家认为，端侧AI上车不仅是技术问题，更是产业生态问题。

大模型的介入将模糊传统“应用”的边界，软件构建范式将发生根本性变化。曹斌认为，未来交互形态将演变为“理解意图—理解环境—做规划—调用技能—执行反馈”的智能体闭环，很难再用今天熟悉的“APP”思维去找杀手级应用。这意味着整车企业需要将车辆的各项能力抽象为可被智能调度的“技能”。这对底层架构、接口标准都提出了新要求。

朱丽敏指出，车企最终交付给用户的是一辆车，需要在端侧算力调配与云端超级智能之间找到高价值平衡。

王耀则进一步提出，端侧模型的核心优势在于本地化的上下文理解与知识库沉淀，就像一位熟悉的秘书，“换了就觉得不习惯”。但要实现这种体验的延续，行业必须走向统一的接口标准、共同的数据存储格式和知识库沉淀方式。

高通中国区技术副总裁许迎春呼吁，未来用户隐私数据应有一个标准化的汇聚路径，无论是车、手机还是智能家居，数据统一存在一个受隐私保护的个人“AI Box”中，换车不换体验。这需要国家层面推动标准建设，否则“没有一家企业愿意做，也无法统一”。

端侧模型必须做到“小而强且美”，既要有足够能力覆盖面，又不能牺牲用户体验感。

千里科技发布AI战略 2028年挑战800万辆智驾搭载量

本报讯 记者**许子皓**报道：近日，千里科技“行千里AI相伴”AI战略暨产品发布会在北京举办。千里科技联席董事长赵明在会上表示，截至2026年3月31日，千里科技智驾技术已搭载17款车型共46万辆车，辅助驾驶激活率达92%，里程同比增长215%，月均避险32万次。按照规划，2026年年底ASD智驾搭载量将突破百万辆，2027年达270万~330万辆，2028年挑战800万辆，目标成为全球最大智能驾驶解决方案供应商之一。

赵明表示，AI正加速重构产业价值链，而“AI+车”是AI技术实现商业闭环的第一入口。千里科技从传统制造业向智能汽车领域转型，依托算力、能源与海量用户数据优势，构建起智能驾驶、智能座舱、Robotaxi完整布局，智驾产品迭代速度不断提升，旗下ASD系统从3.0升级至4.0仅用时三个月，首发上车即覆盖30万辆车。

面向L4级自动驾驶，千里科技提出差异化技术路线，其认为端到端仅为入场券，需依托世界模型模拟极限场景，结合基础大模型实现物理世界理解与意图预测。公司宣布与阶跃星辰达成战略合作，联手打造原生智能驾驶基座模型，采用L4与L2统一架构，将L4技术持续下放至L2/L3场景，突破传统模型上限，适配交警手势、特种车辆避让等复杂路况。千里科技董事长、阶跃星辰董事长印奇表示，阶跃星辰与千里科技的协同，是AI模型与终端场景的最优结合，车作为当下最具规模的具身智能载体，将率先实现AGI技术落地。生态伙伴承诺，未来18个月将推动智驾体验下探至经济型车型，保障算力供应，让安心、舒心、省心的AI出行成为普惠体验。

赵明提出，2026年是超级智能体的上车元年。千里科技以基座大模型为核心，打造具备自主规划、长期记忆、个性化交互的超级智能体，形成“大脑+小脑”的具身智能架构，让汽车成为可陪伴、可协作的出行伙伴。该智能体已搭载于极氪8X，可实现模糊语义识别、场景交互、复杂任务处理，大幅提升驾乘趣味性与实用性。

面向未来，千里科技计划2027年推出全套Robotaxi综合解决方案，2030年实现全球超30万辆Robotaxi搭载其技术。同步开拓国内外市场，以L4智驾+超级智能体推动出行产业升级。公司将助力全新汽车品牌PALLADE落地，定价20万~40万元区间，将覆盖城市车型与越野车，搭载千里科技智驾与智能体技术。

海康机器人 提出“具身智造”理念

本报讯 4月22日至24日，海康机器人智造大会2026在杭州桐庐举办。海康机器人首席执行官贾永华首次向业界系统地阐释了“具身智造”这一核心概念，将智能制造的演进逻辑提升至系统能力的全面重构层面。

贾永华表示，传统自动化生产已触及能力天花板，难以应对需求碎片化与熟练工短缺的结构性阵痛。而“具身智造”聚焦两大核心能力：“高柔性智能体”决定设备通用性上限，使其具备多任务、多工艺、多节拍运行能力；“可复制的场景应用能力”决定技术落地效率，让设备快速理解碎片化场景、稳定融入生产流程。两者相结合，最终推动智能制造与物流领域从“人适应机器”向“机器适应环境”跨越，从而让智能设备“下得了车间，复制得了价值”，也让固定产线转化为按需组合的流动产能。

“国内市场每两台工业相机、每三台移动机器人中，就有一台来自海康机器人。”贾永华表示。据悉，海康机器人已形成机器视觉（眼）、关节机器人（手）和移动机器人（脚）三轮驱动的核心业务矩阵。截至目前，海康机器人机器视觉产品累计出货量已超1000万台，移动机器人下线突破18万台。此外，基于完善的硬件产品线，海康机器人自研工业软件授权用户已超60万人次，业务覆盖超1000个细分场景，全球服务客户超2万家；以此作为支撑，海康机器人2025年的全年营收达64.52亿元。

据介绍，当前，海康机器人的解决方案正走向大规模部署态，融贯于整条产业链的各个环节：从长安汽车、蔚来、先导智能的高端制造，到康师傅、泸州老窖、卫龙的产线检查，再到圆通、万德隆的仓储与流通……

值得关注的是，海康机器人聚焦机器视觉核心技术，在新品发布会上推出了涵盖标准视觉（2D/2.5D计算光学）、3D高精度视觉以及AI智能视觉在内的超35款新品及行业解决方案，主要针对产业深水区中复杂场景成像难、高精测量精度低、AI落地难等“卡脖子”难题。

（杨鹏岳）

DeepSeek V4发布并开源 百万字上下文实现普惠

本报讯 记者**陈存**报道：4月24日，DeepSeek V4预览版本宣布正式上线并同步开源，全系列支持100万token上下文。DeekSeek在公告中表示：“从现在开始，1M上下文将是DeepSeek所有官方服务的标配。”

根据DeepSeek的官方介绍，DeepSeek V4系列按大小分为两个版本，分别是DeepSeek-V4-Pro（总参数1.6T、激活参数49B），以及DeepSeek-V4-Flash（总参数284B、激活参数13B）。

其中，DeepSeek-V4-Pro在Agent能力、世界知识和推理性能方面均迎来了巨大提升。DeepSeek官方表示，DeepSeek-V4-Pro据评测反馈使用体验优于Sonnet 4.5，交付质量接近Opus 4.6非思考模式，但仍与Opus 4.6思考模型存在一定差距。同时，V4-Pro在世界知识测评中，大幅领先其他开源模型，仅稍逊于顶尖闭源模型Gemini-Pro-3.1。另外，在数学、STEM、竞赛型代码的测评中，V4-Pro超越当前所有已公开评测的开源模型，取得了比肩世界顶级闭源模型的优异成绩。

V4-Flash版本则主打高性价比。在Agent测评中，就执行简单任务方面与V4-Pro旗鼓相当，在世界知识储备方面略逊于Pro，推理能力与Pro接近；且由于模型参数和激活参数更小，相较之下能够提供更加快捷、经济的API服务。根据DeepSeek官方定价文档，V4-Pro每百万token输入1元（缓存命中）或12元（缓存未命中），输出24元；V4-Flash则分别为0.2元、1元、2元。

值得注意的是，此次DeepSeek-V4开创了全新的注意力机制，结合了压缩稀疏注意力（CSA）和高度压缩注意力（HCA），显著减少了计算复杂度，提升了长上下文处理的效率。具体而言，在1M token的上下文设置下，V4-Pro的单token推理FLOPs只有V3.2的27%，KV Cache只有10%；V4-Flash则分别压缩到了10%和7%。

V4技术报告中还提到，“我们在英伟达GPU和华为昇腾NPU两个平台上均验证了细粒度EP（专家并行）方案。”据悉，昇腾CANN在当天下午4点直播DeepSeek V4在昇腾平台的首发。

日前，成立3年一直未对外融资的DeepSeek传出消息，称将首次开放外部融资。据相关人士透露，腾讯、阿里巴巴等企业正与DeepSeek洽谈相关事宜，可能把其估值推高至200亿美元以上。此次DeepSeek V4预览版的发布，或将影响其融资进程。当天，受V4版本发布作用，多支DeepSeek概念股已迎来涨停。