

智能驾驶智算数据平台发展研究报告

2025 年 8 月

发起单位

中国汽车工程学会人工智能分会

中国智能网联汽车产业创新联盟人工智能工作组

联合牵头单位

清华大学

北京工业大学

国家智能网联汽车创新中心

北京万界数据科技有限责任公司

参研单位

中国软件评测中心（工业和信息化部软件与集成电路促进中心）

中国信息通信研究院

重庆大学

比亚迪汽车工业有限公司

同济大学

电子科技大学

国汽大有时空科技（安庆）有限公司

国汽智控（北京）科技有限公司

课题工作支持单位

吉林大学

北京车网科技发展有限公司

蔚来汽车科技（安徽）有限公司

广州汽车集团股份有限公司

智能汽车安全技术全国重点实验室

长城汽车股份有限公司

广州小鹏汽车科技有限公司

浪潮电子信息产业股份有限公司

华为云计算技术有限公司

上海临港绝影智能科技有限公司

蘑菇车联信息科技有限公司

首链（广州）区块链科技有限公司

顾问

李 骏 中国工程院院士
清华大学车辆与运载学院教授
中国汽车工程学会名誉理事长、人工智能分会名誉主任委员
中国智能网联汽车产业创新联盟理事长

指导专家

李 升 波 清华大学车辆与运载学院教授
中国汽车工程学会人工智能分会主任委员
中国智能网联汽车产业创新联盟人工智能工作组组长
公 维 洁 中国汽车工程学会副秘书长
中国智能网联汽车产业创新联盟秘书长
国家智能网联汽车创新中心副主任

编写人员

陈 晨	赵 晓 华	边 扬	陈 桂 华	杨 志 成
孙 宫 昊	王 瑶	裴 世 康	沈 修 齐	白 文 静
张 晋 崇	刘 震	刘 会 会	刘 高 阳	邹 博 松
曾 庆 双	洪 启 安	李 金 华	高 锋	詹 光 伦
朱 福 堂	张 潮	张 伟 伟	余王鹏飞	陈 虹
李 庆 建	胡 宝 宝	魏 文 萱	高 博 麟	赵 睿
路 鹏 飞	周 毅	龚 诚	王 代 涵	苗 成 生
任 祥 云	张 瀛	王 潼	程 智 峰	贺 翔
潘 鹏	杜 君	黄 清 成	侯 大 卫	王 金 兰
何 英 琪	黄 家 威			

目 录

1 智能驾驶智算数据平台定义及预期功能.....	1
1.1 智能驾驶智算数据平台定义.....	1
1.2 行业级智能驾驶智算数据平台预期功能.....	2
2 智能驾驶智算数据平台发展现状分析.....	2
2.1 国外现状分析.....	2
2.2 国内现状分析.....	4
2.3 国内智能驾驶基础资源情况调研.....	6
3 基于调研的行业痛点问题及需求分析.....	8
3.1 行业痛点问题.....	8
3.2 行业需求及解决方案.....	9
4 汽车行业级智算数据平台建设关键问题思考.....	10
4.1 平台建设必要性.....	10
4.2 平台总体定位、建设目标及内容.....	12
4.3 平台建设风险分析.....	14
4.4 平台建设难点.....	15
5 汽车行业级智算数据平台建设模式及规划.....	16
5.1 平台建设模式分析.....	16
5.2 平台阶段性建设规划.....	19
附录 1：国内智算中心建设情况.....	21
附录 2：国内智能驾驶基础资源现状调研分析图.....	23

1 智能驾驶智算数据平台定义及预期功能

1.1 智能驾驶智算数据平台定义

端到端自动驾驶要求海量高价值数据、超大规模算力和专业适配算法作为支持。自动驾驶领域数据类型多样、格式复杂、算力分散，且以 Transformer 为基底的人工智能网络模型结构在智能驾驶领域的适应性尚未被充分验证。建立行业级智能驾驶智算数据平台，有望集中行业力量，建立高质量自动驾驶数据集、集中算力资源、合力研发适用于端到端自动驾驶的算法模型，是推动人工智能技术融入智能驾驶领域，破解我国智能驾驶相对落后局面、实现赶超的最有潜力途径。

“智能驾驶智算数据平台”是指服务于汽车智能驾驶模型开发及测试的数据、算力和算法综合服务平台，通常可分为企业级平台和行业级平台。其中，“智能驾驶模型（智驾模型）”是指可具备感知、预测、决策、规划、控制等功能、可驱动自动驾驶车辆安全、高效运行的分块化或一体化神经网络载体模型。

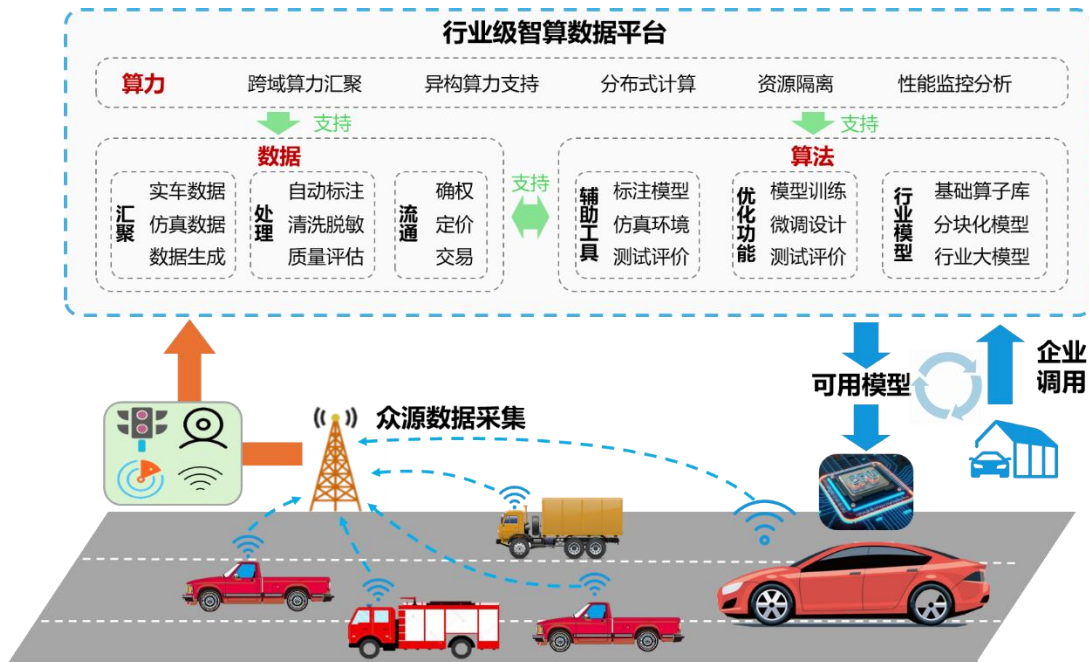


图 1 行业级智能驾驶智算数据平台示意图

智算数据平台一般由大算力 GPU 或专用人工智能芯片并行集群为主要计算硬件，可以服务于人工智能深度学习模型的设计与训练，自动驾驶算法在各类数据基础上不断地重复训练与验证，推动车辆对环境感知、路径规划等的认知水平逐渐趋近于真实驾驶场景。

行业级智能驾驶智算数据平台旨在为行业内整车企业、智驾供应商、研究机构及高校等主体的智能驾驶模型开发和应用提供基础数据、算力和算法服务，并

协调上述单位及相关数据算力供应商开展智能驾驶研发基础资源的汇聚和流通。

1.2 行业级智能驾驶智算数据平台预期功能

行业级智能驾驶智算数据平台预期将在数据、算力、算法三大人工智能支柱方面提供行业级服务。

（1）数据服务。提供行业级数据服务能力，包括但不限于数据的汇聚与传输、确权与鉴权、存储与管理、分类和分级、合规与脱敏、清洗与预处理、质量评估与价值定义、可视化与分析、机器学习平台集成、交易与共享等。

（2）算力服务。提供行业级算力服务能力，包括但不限于算力资源跨地域汇聚、异构算力支持、资源隔离与调度、高可用性与容错性、性能监控与分析等。

（3）算法服务。提供行业级算法服务能力，包括但不限于基础算法库、算法开发与优化工具、数据标注基础模型、智驾模型训练与测试环境、智能驾驶基础模型、剪枝压缩部署工具等。

2 智能驾驶智算数据平台发展现状分析

2.1 国外现状分析

2.1.1 国外企业及项目建设情况

（1）智能驾驶开发商及汽车厂商

智能驾驶技术在国外发展较早，出现了特斯拉、Waymo、Wayve 等智能驾驶开发商。在数据驱动的端到端自动驾驶系统展现强大的驾驶能力后，各自动驾驶开发商以及宝马、奔驰等传统汽车厂商也开始积极进行探索和积累。

特斯拉在数据、算力和算法方面均处于国际领先地位。数据方面，特斯拉通过保持少量标准化的车型并搭载“影子模式”，采集格式统一的道路交通和驾驶行为数据，数年来累积了大量的有价值数据。算力方面，特斯拉 2024 年底平台算力总规模扩大到相当于 30 万块 A100 显卡的总算力，约为 100 EFLOPS(FP16)。算法方面，特斯拉引领了 BEV、占据网络以及数据闭环端到端等自动驾驶技术研发，并且开发了高度真实的仿真平台，兼具 WorldSim 和 LogSim 仿真能力。

Waymo 的主要优势之一在于可依托 Google 在全球的用户网络，通过用户的机器人验证进行道路交通对象数据标注，帮助 Waymo 建立高质量数据集。此外，Waymo 还通过高保真模拟驾驶环境用于补充真实世界的的数据。

Wayve 积极利用人工智能算法开展自动驾驶系统开发。2023-2024 年，Wayve 发布多模态生成式世界模型 GAIA-1、混合模型 LINGO-1、WayveScenes101 数据集、自动驾驶仿真软件 PRISM-1。

奔驰采用谷歌云 Vertex AI 平台用于挖掘客户信息以提供个性化内容服务，同时在构建下一代高级驾驶辅助系统时采用谷歌云平台作为算力中心；采用 NVIDIA DRIVE Sim 仿真软件对智能驾驶功能进行测试验证。

宝马集团与亚马逊云科技开展战略合作，开发定制云软件用于宝马车辆的数据管理和分发，支持包括分析、机器学习、数据库、存储和计算等功能。

英伟达已开发端到端自动驾驶汽车开发平台 NVIDIA Drive、针对深度学习应用的 NVIDIA DGX-1 超级计算机、用于智能座舱的开放和可扩展的软件框架 NVIDIA DRIVE IX、端到端自动驾驶仿真平台 NVIDIA DRIVE Sim。

（2）联盟/行业项目

① GAIA-X

GAIA-X 是一个旨在建立欧洲数据和云主权的倡议，由德法联合牵头，欧盟 27 个成员国参与，覆盖 378 家企业，目标是构建一个真正属于欧洲的开放的数据基础设施，以实现数据和服务的透明度、可控性、可移植性和互操作性，成为欧盟的“母云端”，并创立通用云标准、参考云架构和互操作性要求等。GAIA-X 的架构设计包括数据共享的共同标准、最佳实践、工具和治理机制，以支持数据共享和跨公司合作。

GAIA-X 的核心是提供一个共享的数据和服务平台，使公司更有效地使用资源，并提高价值创造过程的透明度和弹性。它遵循国际数据空间协会（IDSA）的国际数据空间（IDS）标准，确保数据交换基于欧洲价值观，如数据保护和安全、通过联合设计确保平等机会以及为数据创造者保证数据主权和参与者间的信任。

此外，GAIA-X 信任框架（Trust Framework）定义了一套规则，以确保参与者遵守政策规则，如安全、隐私、透明度和互操作性。GAIA-X 标签（GAIA-X Labels）是服务提供商的可选方案，以确保透明度和互操作性的共同水平，同时限制市场参与者的监管负担。

② Catena-X

Catena-X 是欧盟数据战略框架下德国主导的汽车行业共享数据生态系统，旨在创建一个开放的、协作的数据生态系统，以促进整个汽车价值链上的数据共享和合作。成员覆盖汽车制造商、材料与零部件供应商、数字基础设施运营商、公共研究机构等。该项目开发了行业级数据共享平台，制定了数据标准与规则。

Catena-X 基于 GAIA-X 和 IDSA 的欧洲原则，致力于解决汽车行业中的挑战，如供应链韧性、合规性、可持续性和监管需求。通过构建端到端的数据链，使企业能够轻松连接到新的业务应用程序，并实现合规、弥合信息差距。项目的技术基础包括一系列核心服务，如身份和访问管理、数据交换、以及数字孪生注册表等，旨在促进数据的标准化和互操作性，同时确保数据主权和安全性。

Catena-X 得到了德国联邦经济事务和气候行动部资助，并且与 GAIA-X 倡议紧密相连。通过 Catena-X，企业可以提高供应链的透明度，优化质量和物流流程，并实现更高的二氧化碳减排透明度，同时简化主数据管理。Catena-X 的主要应用领域为汽车行业的供应链数据共享与交换，遵循去中心化分散型系统和中立治理原则，为自动驾驶数据的共享流通提供了借鉴价值。

2.1.2 国外现状小结

从已有资料来看，国外智能驾驶开发商中，特斯拉和英伟达是智能驾驶技术开发的引领者。前者长期处于领先并在数据、算力和算法方面全方位累积了大量资源；后者以算力芯片和 AI 架构制造和开发商的身份介入智能驾驶领域开展系统研发，提供计算平台、系统框架、仿真生成等多种智能驾驶基础功能服务，并与奔驰等企业达成合作，竞争优势显著。除此之外，其他国外智驾供应商相对于国内企业，其主要优势仍然在于算力资源获取上，数据方面从公开信息上来看并未具有显著优势，算法层面也未有突出领先优势。行业合力方面，GAIA-X 和 Catena-X 等项目为智能驾驶领域的数据信息沟通和资源交换奠定了良好基础。

2.2 国内现状分析

2.2.1 国内智算中心建设情况

下述智算中心算力规模情况等信息详见附录 1。

（1）整车企业与智驾供应商智算中心

随着智能驾驶技术的快速发展，越来越多的整车企业和智驾供应商开始重视智能驾驶智算中心的建设，算力规模不断提升，为智能驾驶模型的开发和测试提供了强大的算力支持，有力地推动了智能驾驶技术的发展和應用。

（2）互联网科技厂商智算中心

除了智能驾驶智算中心外，互联网科技厂商、大模型厂商等均开展算力投资，以争夺人工智能算力的战略高地，腾讯、百度、阿里、字节跳动、商汤等公司积极部署智算中心。其中，阿里的张北超级智算中心建设规模达到 12 EFLOPS；百度和腾讯也已在广州、上海、北京等多地建立了智算中心；字节跳动则通过与润泽科技等公司的合作，加大在智算中心相关 IDC 领域的投资力度。

（3）政府智算中心

2023 年以来，政府智算中心建设的规模与节奏显著提升，在建及规划项目较多。随着 AI 带动算力需求的提升，单个智算中心的体量提升，2023 年下半年之后建设与投运的智算中心出现较多 1000P 以上的算力规模，大型和超大型数据中心成为主流。

（4）智能网联先导示范区数据算力平台

目前，全国共建设 17 个国家级智能网联汽车测试区、7 个车联网先导区、16 个“双智”试点城市，确定 20 个城市（联合体）为智能网联汽车“车路云一体化”应用试点城市。上述智能网联示范区、先导区、试点城市等通过路侧感知设施构建，汇聚了海量道路车辆运行数据；多地开展云控基础平台建设。

2.2.2 国产生态现状及薄弱环节

（1）国产 AI 算力芯片技术水平快速发展，智算中心国产化率不断提升

近年来，我国国产 AI 算力芯片的设计和制造技术发展迅速，取得了巨大突破。目前，常用 AI 算力芯片以 GPU 为主，国内华为、天数智芯、摩尔线程、中科曙光、燧原科技等企业纷纷布局研发，已有量产产品。此外，其他 AI 算力芯片，如 TPU、FPGA、ASIC、NPU 等也在快速发展，我国企业也有所布局。

但需要指出的是，除华为昇腾外，目前大多数 AI 芯片厂商还无法实现规模化出货。国产芯片在 AI 计算框架方面成熟不足，端到端自动驾驶要求厂商快速构建便捷可用的算力，因此当前的智驾算力平台仍以英伟达相关产品来构建 AI 算力的主体，同时通过加入海光、寒武纪等国产芯片以及加速卡来构建 AI 算力。

（2）国外 AI 计算框架占据市场主流，国产适配性框架处于起步阶段

目前全球 GPU 市场的绝对份额被英伟达和 AMD 占据，其中英伟达之所以能够成为 GPU 全球头号玩家，主要源于其在先发优势下建立的 CUDA 生态。

在 AI 框架方面，华为针对 AI 应用场景，设计了功能类似英伟达 CUDA 的异构计算架构 CANN 以及帮助开发者孵化各种 AI 创新算法和应用的 AI 计算框架 MindSpore。但与主流框架（如 TensorFlow 和 PyTorch）相比，MindSpore 的市占率和社区规模较小，在开发者习惯、工具链兼容性、深度学习算子完备性、技术文档丰富性等方面仍需提升。特别是在软硬件适配方面，MindSpore 对华为昇腾芯片的高度优化虽然是其优势，但也限制了对其他主流硬件的适配能力。

（3）我国智能驾驶数据集开放程度不佳

目前，国外智能驾驶领域拥有丰富的数据集资源，例如 nuPlan、Argoverse 和 Waymo Open Dataset 等。这些数据集不仅涵盖了多样化的驾驶场景，还提供了高质量的标注，广泛应用于感知、决策、规划等关键环节的算法开发。相比之下，我国的数据共享机制尚未充分发展，目前公开的国内数据集少之又少，开放的数据集数量和质量不足，这对智能驾驶算法开发造成一定限制。

（4）数据共享与开放机制不完善

我国在智能驾驶领域缺乏明确且细化的数据共享政策细则，特别是在数据确权、定价、流通等关键环节上存在短板。在实践中，行业仍面临数据产权界定不清、流通规则模糊、共享激励机制不足等问题，数据共享的合规性和安全性保障

不足，许多企业不愿开放自有数据资源，这直接限制了智能驾驶领域高价值数据集的形成，尤其是 Corner Case 和自动驾驶事故数据的汇集，制约数据生态构建。

（5）我国 AI 专业技术人才缺乏，校企合作存在工程化鸿沟

近年来我国加速 AI 人才培养并取得显著进展，2024 年我国 AI 产业人才约 75 万人。然而，我国 AI 人才供需仍存在较大差距。

从人才数量上看，麦肯锡报告显示，预计 2030 年中国对熟练 AI 专业人员的需求将达到 600 万，但人才供应量仅能达到 200 万。专业技术人才数量与产业需求之间的差距较大，制约了我国 AI 技术的进一步发展和创新能力的提升。

从人才结构上看，我国的 AI 专业技术人才集中在算法研发、机器学习和数据处理等核心领域，但在 AI 计算架构优化、跨领域融合等关键环节的人才储备明显不足。此外，高端 AI 技术人才（如深度学习架构设计师、高性能算力优化工程师等）缺口尤为严重，导致企业在技术研发中面临较大的人才瓶颈。

在人才培养方面，校企合作中的工程化鸿沟依然突出，制约了 AI 专业技术人才的培养质量和行业适配度。人才培养与产业需求不匹配，导致校企之间的合作往往停留在表面层次，缺乏针对产业需求的联合培养计划。

2.3 国内智能驾驶基础资源情况调研

为进一步了解我国智能驾驶数据、算力和算法资源现状，以及智能驾驶智算数据平台建设等问题，课题组针对不同类型调研对象分别开展了问卷调研。

调研周期：2024 年 8 月至 2024 年 10 月。

调研对象：整车企业、智驾供应商、数据/算力供应商、超算中心、高校和研究机构，共发放和收集有效问卷 21 份。

以下分析基于调研结果整理得出，仅反映调研周期内受调单位的整体情况。

2.3.1 数据方面

数据闭环机制。大部分整车企业和智驾供应商已建立较为完整和成熟的数据处理流程，并且开发有自动化处理工具链。

路侧数据应用。70%以上整车企业和智驾供应商在智驾模型训练中未使用路侧感知数据；在拥有且使用路侧数据进行智驾模型训练的企业和机构中，几乎全部认为路侧感知数据对提升智驾模型性能十分有效。

高价值数据占比。从所拥有的数据量来看，我国整车企业及智驾供应商的已有数据中，Corner Case 数据的占比在 5%以下。

数据使用成本。总体来看，成本较高环节在于标注以及采集/获取。

数据共享问题。针对限制智能驾驶数据共享的主要原因，受调单位认为最大的阻碍来源于各企业和机构之间的数据格式标准不统一，其次是敏感数据的合规

边界不清晰。智驾供应商还认为数据产权和价值难以确定和追踪也是关键问题。

2.3.2 算力方面

算力中心建设方式。整车企业倾向于采用租赁+自建的方式进行，而智驾供应商更多会选择采用全部租赁的方式。

国产卡使用情况。在拥有自建算力中心的企业或机构中，国产 AI 算力卡的占比低于 10%，并且超过 70%的企业或机构完全未使用国产 AI 算力卡；在使用过程中，主要面临 AI 计算框架适配问题以及大规模算力集群下的算力调度问题。

算力利用情况。大部分整车企业和智驾供应商拥有的算力能够满足其智驾算法训练需求；而对高校和研究机构而言，算力仍是制约其技术研究的关键问题。

2.3.3 算法模型方面

端到端架构建设情况。70%以上的企业和机构已建立端到端架构并开始模型训练。其中，27%自称采用分块式架构，45%自称采用一体式架构。并且，整车企业由于量产压力，更多采用分块式架构以加速技术开发；智驾供应商更倾向采用一体式架构。整车企业已开始端到端系统的实车测试甚至量产车部署；智驾供应商已开始实车测试验证；高校进展较慢，大多数尚未获得可运行测试的版本。

端到端架构发展趋势。在端到端架构是否有可能实现高级别自动驾驶的预测上，71%的企业和机构持肯定态度，并且分块式和一体式架构占比相近。其中，智驾供应商基本全部看好一体式架构；整车企业由于量产压力会两方押注；高校和研究机构持保留态度，67%认为分块式架构更能保持可解释性和优化便利性。

端到端模型训练难点。根据各企业和机构的经验，目前认为阻碍端到端智能驾驶模型训练的最大问题在于模型结构的不确定性，其次为 Corner Case 数据量不足、常规驾驶数据不足。训练算法、算力方面反而并不是目前的关注点。

2.3.4 行业级智能驾驶智算数据平台建设态度方面

平台建设态度。针对是否有必要建立该平台，总体来看业界较为支持，并期望突破标准规范、建立国产 AI 计算生态；同时对建立行业级数据交易平台、辅助服务功能（数据标注、测试评估）的期望较大。而部分车企和智驾供应商认为现阶段不宜投入过大，仅需突破标准规范等来支持行业数据市场的顺畅性即可。

平台建设优先级。几乎所有企业认为数据板块应优先构建，其次是算力。主要原因在于算力方面的获取路径较为直接，并且在数据不足、网络结构和算法路径未打通的情况下，对算力的需求并不迫切；而数据，特别是 Corner Case 数据是现阶段面临的最大短板。

平台功能建设。各类企业和机构普遍认为，平台应重点提供数据格式标准化

工具，支持数据确权，并成为行业级数据共享流通平台。此外提供数据标注和高价值数据筛选功能。对于算力和模型服务，除部分高校外，其他单位需求不高。

平台建设与运作模式。几乎所有企业和机构均认为平台建设需要从国家层面进行推动，主要原因在于数据合规以及流通方面的机制突破需要自上而下进行。平台建成后，大多数企业认为应采取政府补贴与市场运作相结合，采取部分盈利的方式进行运作。平台建成后，多数企业和机构愿意付费购买相关服务。

AI 计算框架选择。在平台建设的 AI 计算框架选择方面，大多数企业重点考虑算子和框架匹配度，因此仅部分机构和企业支持强制采用国产框架；大部分单位认为可采用国外开源框架，或者前期采用国外开源框架，之后逐步替换国产。

平台建设参与意愿。在平台建设的参与意愿中，大多数企业和机构持观望态度，要求具有明确的收益预期才可能愿意投入一部分资金支持平台建设。

平台算力体系参与意愿。若平台建成，70%以上的企业愿意加入行业级智算数据平台主导的算力共享共用体系中，通过通报富余算力，以有偿方式对外共享。

平台建设效果预期。若智算数据平台建成，各企业和机构认为将加速智能驾驶模型的开发及商业化，并扶持中小企业技术研发，但也担心规模性的数据风险。

3 基于调研的行业痛点问题及需求分析

3.1 行业痛点问题

基于上述调研结果，对整车厂商、智驾供应商以及高校和研究机构等受调单位面临的痛点问题进行分析，在一定程度上能够反映行业整体情况。

（1）数据方面

企业数据持续积累，但事故和 Corner Case 数据积累较少。从数据分布看，存量数据中以人类驾驶常规数据和自动驾驶常规数据为主，Corner Case 数据和人类驾驶事故数据较少。从企业需求看，对人类驾驶事故数据和自动驾驶 Corner Case/接管/事故数据的需求较为强烈，甚至部分企业将其视为核心资产且不愿对外流通。从智驾模型训练看，常规驾驶数据（包括仿真数据）能够帮助模型提升在常规场景下的驾驶能力，但进一步性能提升需要大量非常规驾驶数据支持。

常规场景仿真能力较强，但 Corner Case 场景生成仍依赖手动设计。大多数企业和高校均具有自动化场景生成能力，并且兼具规则型方法和学习型方法。但除个别企业外，仅能够大规模生成常规场景，Corner Case 场景生成仍严重依赖手动设计。此外，基本所有企业和高校没有明确的 Corner Case 场景定义，企业也希望智算数据平台能够提供高价值数据筛选功能。

数据采集和标注成本占比较高，自动化水平较低。目前，数据采集和标注占

据较大成本比例，虽然已有大量数据标注供应商，但企业仍以自建标注能力为主，其中人工标注/核验率仍然较高，仅部分企业建立自动标注能力。同时，不同型号车辆、商用车和乘用车不同类型车辆感知数据也无法复用，加剧了数据标注高成本困局。对应地，企业希望行业级平台能够提供数据标注服务。

企业数据流通意愿强烈，但存在格式差异、合规限制及安全等问题。企业具有较强的数据流通意愿，希望以有偿方式共享自有数据以及购买数据。但由于企业之间数据格式不统一以及存在合规性要求，特别是舱内数据、地理信息数据等较之驾驶轨迹数据有更严格的法规限制，导致数据流通受限。企业较为关注的还有数据加密、确权与定价，而当前的相关技术尚未完全成熟。对应地，企业希望对上述问题有所突破，且希望行业级平台具备数据交易平台职能，提供确权和定价功能。此外，需要考虑如何实现数据更新和新旧数据结合；需要考虑地理信息安全问题，包括数据传输安全、云安全等问题；需要考虑数据分发可能导致的数据泄露风险对数据共享造成的影响，以及应采取何种技术措施。

（2）算力方面

国产算力卡占有率极低，AI 框架适配和算力调度是主要问题。在拥有自建算力中心的企业中，其算力中心基本采用国外算力卡，国产算力卡占比低于 10%，主要问题在于 AI 框架适配和算力调度问题，迫于产品压力选择使用更为成熟的国外 AI 框架，如 Pytorch。

（3）算法方面

端到端智能驾驶网络构架尚不确定，技术路径不够清晰。分块式和一体式端到端架构孰优孰劣尚无明确结论，整车企业倾向于分块式架构以满足量产需求，智驾供应商更倾向于一体式架构。虽然部分企业声称已开始实车测试甚至量产，但仍表示模型结构尚未完全确定，技术路线仍然不够清晰。

（4）平台建设方面

支持行业智算数据平台建设，但注重有用性和可持续性。大部分企业和单位支持行业智算数据平台建设，但对于是否愿意付费使用平台功能持审慎态度，认为需要平台表现出必要的技术成熟度和服务有用性才愿意付费，因此各方对平台服务的资金投入和商业模式有所担忧。

3.2 行业需求及解决方案

总体而言，当前行业中对端到端自动驾驶资源要素的需求迫切程度排序为数据>算力>算法，这也是受调单位期望行业级智算数据平台的优先建设顺序。

数据方面，目前企业有一定数据积累，但数据分布覆盖不足、关键数据集中在个别单位手中，需要建立行业级数据交易平台，以帮助解决数据流通问题，为

智能驾驶模型训练提供足够数据要素基础。目前，数据流通的主要阻碍在于标准规范和机制方面，包括需要建立数据格式标准、数据合规及脱敏处理规范、数据确权和追踪机制以及数据定价机制等。

算力方面，当前算力储备能够勉强满足目前的训练需求，但究其原因可能在于智能驾驶模型和训练算法不确定导致的算力利用率低下；同时，随着训练数据量的不断加大以及算法模型的进一步发展成熟，算力需求有可能进一步增加。此外，国产的算力卡及 AI 框架占比极低，存在潜在的安全隐患。另外，高校方面同时受到数据和算力限制，技术研发进度落后于企业。

算法方面，主要面临的问题在于端到端自动驾驶模型结构和训练算法的不确定，但目前由于端到端自动驾驶处于起步阶段，各企业和机构均在摸索前进；而在数据标注、数据筛选、数据生成和仿真测试等通用基础功能方面需求较高，这也是各企业期望行业级智算数据平台能够提供的功能服务。

此外，行业级平台建设也存在巨大的资金风险、技术落后风险以及商业模式风险，大规模的专门性技术开发可行性较差。

因此，建议以数据和服务交易平台为核心，重点突破数据格式等标准规范，开展数据确权、数据定价、数据筛选、模型测试等轻量化技术开发，并开展国产化算力卡和 AI 框架生态建设，以扶持国产算力和 AI 框架发展、支持中小企业及高校和研究机构技术开发，定位为智能驾驶基础功能服务平台，为我国智能驾驶技术开发提供帮助。

4 汽车行业级智算数据平台建设关键问题思考

人工智能技术在汽车行业的深度融合离不开海量高质量专业数据集、大规模集群算力的支持，以及面向汽车行业不同专业领域的适配性神经网络模型和训练算法的设计和开发。不同于自然语言处理领域，智能驾驶相关人工智能技术开发尚属起步阶段，面临高质量数据集获取难度高、算力储备不足、算法模型技术路径未知等挑战。如果建立行业级智能驾驶智算数据平台，有望能够整合行业力量，在政策法规突破、技术创新、应用拓展、生态建设以及资源共享方面提供支持。

4.1 平台建设必要性

行业级智能驾驶智算数据平台的建设旨在着眼于智能驾驶领域的人工智能技术开发和应用，建立一个面向行业的信息互通、资源交换和交流合作的渠道和平台，探索人工智能技术融入下汽车行业新的发展模式和应用场景，推动汽车产业向着更加智能化的方向前进。

当前，智能驾驶领域端到端自动驾驶技术发展如火如荼，各整车企业、智驾

供应商等均开始进行数据、算力资源的累积和算法开发，以在数据闭环驱动的端到端自动驾驶时代取得先机。调研结果显示各企业对于汽车行业级智算数据平台的需求旺盛，并且也迫切需要一个行业平台进行统筹协力，特别是在数据的标准规范和共享交易方面，数据的缺乏已经成为智能驾驶技术开发的严重制约因素。因此，有必要开展行业级数据平台建设，通过行业协力，解决数据规范、数据流通、算力分配、标注评估以及生态建设等难题，加快智能驾驶技术发展。

通过行业级智算数据平台建设，预期将在机制、规范、数据、算力、算法和生态方面为我国智能驾驶技术的研发和落地应用提供支持。

4.1.1 机制方面：应对 AI 技术爆炸，搭建行业协力平台

智能驾驶包含环境感知、认知、预测以及车辆行为决策、规划和控制等功能模块。从感知侧开始，人工智能技术与智能驾驶持续深度融合，预期将迎来爆发式发展。然而，智能驾驶面临复杂的道路环境、多样的交互行为、软硬协同的车辆控制以及数据和驾驶安全等诸多挑战。搭建行业级智能驾驶智算数据平台，将能够为汽车行业的智能驾驶技术开发提供交流合作渠道和载体，通过建立定期的沟通机制和合作框架，设立联合研发项目，集合不同领域的专业知识，共同攻克技术难题，促进技术的快速推广和应用，加速智能驾驶技术的商业化进程。

4.1.2 规范方面：支持标准规范建设，突破机制制度瓶颈

当前智能驾驶技术开发面临数据格式不统一、敏感数据边界不清晰、数据合规要求严格等挑战，限制了智能驾驶模型及算法的开发。依托行业级智能驾驶智算数据平台的建设，将组织相关标准和规范制定，突破智能驾驶数据采集、数据利用、数据流通、模型开发、测试应用等相关机制和制度瓶颈，为智能驾驶模型技术的发展提供一个稳定、可预期的机制规范和环境，加速技术创新和应用落地。

4.1.3 数据方面：开展数据流通试点，疏通资产交换渠道

当前我国企业已开展智能驾驶数据的大规模采集和应用，部分头部企业已积累大量数据。但从数据分布来看，罕见高危场景数据量较小，严重制约了具备应对高复杂、高随机的混杂交通场景能力的智能驾驶模型训练和测试。依托行业级智算数据平台建设，通过建立数据采集、数据标注、数据质量等相关标准，开发适用于驾驶数据的水印、鉴权、追踪等技术，构建数据定价机制，形成安全可信的数据交易第三方平台；开展涉及多方主体的数据流通试点，打通数据资产交换渠道，为高质量驾驶数据筛选和利用积累经验，充分释放高价值数据潜力。

4.1.4 算力方面：实现信息共享互通，优化资源配置使用

受国际环境影响，我国算力芯片及算力卡的获取受到限制，整车企业及智驾供应商较之特斯拉等国外智驾企业算力稍显不足，同时高校和研究机构也因严重缺乏算力支持而技术开发进展缓慢。依托行业级智算数据平台，一方面通过规模化集群建设扶持国产算力卡发展，另一方面通过技术突破和机制创新，实现各地超算和算力供应商的闲置算力集中，形成行业级的算力汇聚，面向算力需求方进行算力租赁和模型托管训练，补充行业各方算力不足，提高存量算力资源利用率，加速智能驾驶模型算法的研发和应用。

4.1.5 算法方面：提供共性基础服务，推动技术发展创新

虽然以大模型为代表的人工智能技术在自然语言处理领域表现出强大的智能涌现能力，但就智能驾驶领域而言，具备高安全、高迁移和高智能的自动驾驶模型技术框架和技术路径尚不清晰，且在数据处理、数据生成、模型测试等方面需要大量辅助支持性工作，单一企业难以囊括所有业务链条。行业级智算数据平台的建设，可着眼于智能驾驶共性服务需求，包括数据标注、仿真生成、测试评估等，集中行业力量进行功能开发形成服务产品，一方面节省研发成本，另一方面可建立标准化测试流程，为智能驾驶技术快速开发和安全应用提供保障。

4.1.6 生态方面：聚焦特定难题和场景，助力国产生态建设

目前智能驾驶领域存在数据、算力资源分布不均，技术路径不确定，兼具汽车和人工智能背景的人才严重短缺等问题；同时，校企合作由于涉及数据、算力及实车测试成本，存在技术开发需求错配、应用渠道不畅等问题；此外，算力方面面临着严重的卡脖子问题，国产算力卡及框架占比极低。依托行业平台，通过聚焦于特定的技术难题或应用场景，开展竞赛活动、交流会、技术研讨会等，鼓励使用国产算力卡和 AI 框架，鼓励企业与高校、科研机构建立长期合作关系，共同培养专业人才，推动产学研深度融合，为国产生态的发展注入强大动力。

4.2 平台总体定位、建设目标及内容

4.2.1 平台总体定位及建设目标

基于当前汽车行业在智能驾驶数据、算力和算法方面的现状和需求，行业级智算数据平台**总体定位**为“汽车行业人工智能关键要素流通的依托服务机构”，旨在构建一个具有高度公信力的行业基础设施，致力于解决行业存在的数据孤岛、资源配置不均等问题，促进数据、算力和算法等关键要素有效流通与合理配置。

具体而言，平台的**总体建设目标**包括：

(1) **数据共享交易。**建立一套完善的数据共建、共享、共用机制，具备数据质检、确权、定价、交易等能力，支持不同主体间数据的安全和高效流通。

(2) **算力资源配置。**建立一套完善的算力资源信息公开和共享体系，提供闲置算力资源信息的发布和供需配对服务，实现算力资源的最大化利用。

(3) **算法基础服务。**建立一套完善的智能驾驶辅助服务算法模型体系，包括数据标注、仿真生成、测试评估等，提供标准化、高性价比的基础服务。

(4) **生态系统构建。**建立一套完善的汽车 AI 生态建设规划，开展周期性竞赛活动、交流研讨等活动，共同打造开放共赢的产业生态体系。

4.2.2 数据方面建设内容

数据方面，行业级智算数据平台定位于“**智能驾驶数据的规范性流通中介**”，重点发展数据标准化、数据质检、数据确权、数据定价等关键技术能力，建立数据信息公开和交易服务系统；平台保有少量典型数据集用于支持生态建设。**主要建设内容包括：**数据标准化服务、数据质检服务、数据确权服务、数据定价服务、数据交易系统、车路云一体化系统协同衔接及数据汇聚、典型数据保有等。

4.2.3 算力方面建设内容

算力方面，行业级智算数据平台定位于“**智能驾驶算力资源信息公开及服务平台**”，重点关注全国范围内算力资源的信息公开和流通共用，充分利用闲置算力，优化资源配置；同时，平台保有少量国产算力资源用于支持生态建设。**主要建设内容包括：**闲置算力信息公开、算力租赁服务系统、国产算力资源保有等。

4.2.4 算法方面建设内容

算法方面，行业级智算数据平台定位于“**智能驾驶开发辅助服务提供者**”，重点开展数据标注、仿真生成、评价测试等智驾模型相关辅助功能技术服务，支持模型训练和测试。**主要建设内容包括：**智能驾驶数据标注服务、智能驾驶数据仿真生成服务、智能驾驶模型标准化评价测试服务等。

4.2.5 生态方面建设内容

生态方面，行业级智算数据平台定位于“**国产生态建设推动者**”，重点开展以我国特色交通环境、国产算力和算法为核心的生态建设，开展竞赛、论坛、产学研合作、资料共享、实车测试、市场推广等活动，助力国产生态发展和完善。

4.3 平台建设风险分析

4.3.1 平台建设风险点

（1）资金风险

- 单一资金来源（单一企业）难以支撑平台建设和运营；
- 数据和算力需求持续增长带来的持续投入风险；
- 融资渠道的不确定性；
- 分期建设带来的财务平衡风险。

（2）技术风险

- 平台架构设计的兼容性；
- 建设进度落后于技术迭代或技术方向转变、平台无法适应新需求的风险；
- 数据确权定价追踪等技术不成熟、分布式算力服务延迟问题导致的风险；
- 高质量标注数据和仿真环境不足，影响平台对复杂交通场景的适配能力。

（3）人才风险

- 兼具 AI 和汽车技术的复合型人才不足，专业技术人才培养周期长；
- 智能驾驶领域国际顶尖技术人才引进难度大；
- 现有人才流失风险；
- 学术研究与产业需求脱节风险。

（4）政策风险

- 数据合规、数据流通限制；
- 高级别自动驾驶相关政策尚不明晰，政策出台或调整时间不确定；
- 补贴和政府支持的波动性；
- 知识产权保护政策不足带来的技术被侵权或盗用风险；
- 标准政策滞后风险，缺乏统一标准可能导致资源浪费和重复建设；
- 国际环境变化，芯片制裁容易导致基于国产芯片的大模型出口受限。

（5）市场风险

- 平台建设初期企业认可度不足的风险；
- 平台运营及商业模式创新带来的可持续发展风险；
- 数据来源的可持续性风险；
- 智能驾驶商业化进程不确定性风险；
- 市场需求与产品匹配风险。

4.3.2 总体风险评估及预期

总体来看，平台建设难度较大，特别是政策、市场和技术等方面的阻碍复杂

而交叉，互相掣肘。具体而言：

数据方面，政策层面存在合规问题，技术层面存在确权问题，市场层面存在定价问题。行业各方是否有动力加入数据流通、平台方能否保证数据安全、运营模式能否保证持续性的多方共赢等问题需要商榷。

算力方面，技术层面存在国产生态较差的问题，在资金层面存在延续性投资风险高的问题，但此方面问题可采取前期小范围试点、待生态和运营模式成熟后再扩大化建设的方案进行规避。

算法方面，技术层面存在数据标注、仿真生成技术不成熟等问题，并叠加智能驾驶技术本身快速迭代、技术路径不清晰等问题，大规模的算法建设容易落后于技术发展进度，而小规模算法服务又难以吸引企业购买。

总体来看，若瞄准建设一个行业级、大规模的数据、算力和算法服务平台，在没有政府力量来协调行业各方资源及诉求的情况下，平台建设存在巨大风险。但在建设模式上，为降低投入成本和风险，可依托某一头部企业或行业协会先期进行少量资金投入，以数据为基础，重点突破数据流通和测评相关的关键技术研发，逐步建立平台生态。

4.4 平台建设难点

4.4.1 确保公信力的技术能力

公信力直接决定平台吸引力和可持续运营能力，平台需在以下关键领域攻克技术难题以确保公信力。

(1) 资源标准化服务能力。平台需要支持异构数据、算力和算法模型的标准化服务，使其可以快速应用于多种场景。

(2) 资源质检能力。平台必须提供高效的数据质量和算法模型性能检查服务，作为第三方中立机构出具权威的检测报告。

(3) 数据安全与隐私保护。平台需建立健全的安全保障体系，利用隐私计算、区块链技术等，确保交易数据在流通过程中的不可篡改性和机密性。

4.4.2 基础资源保有

自有基础资源的储备和规模是平台持续运营的重要基石，资源不足或质量不高将导致平台难以吸引企业使用产品服务，平台需在以下几方面加强资源保有。

(1) 高覆盖率的驾驶场景数据资源。需建设涵盖多种道路类型的数据集，同时覆盖多样化条件；利用仿真生成技术创建无法轻易获取的稀缺数据资源。

(2) 国产算力资源。平台需要部署分布式国产算力中心；优化算力分配机制和支持异构计算模式；开发可供中小企业使用的经济型算力服务。

(3) 基础功能服务资源。平台需要储备丰富的基础功能服务资源，构建支持智能驾驶技术开发的核心工具链。

4.4.3 标准规范突破

为促进行业资源的高效流通，平台需解决法规、技术和行业协作中的关键障碍，形成科学、健全的标准规范体系。

(1) 跨平台互通性标准。需要制定覆盖数据格式、模型接口及算力协作的互通性标准，这些标准需要支持跨厂商数据交易及多平台算力共享。

(2) 数据确权与定价标准。建立一套明确的数据确权流程与评估体系，记录数据生成时间、来源及使用权限；同时制定数据定价规范；可引入智能合约。

(3) 隐私与数据安全标准。平台需建立数据匿名化、分级管理等安全标准；平台还需制定隐私泄露惩罚机制，增强行业对平台的信任。

4.4.4 初始投入及可持续运营

平台建设前期需要初始资金投入，而后为确保平台长期可持续运营，需从商业模式、资源优化及市场拓展多方面入手，构建完整的运营体系。

(1) 减少初始投入。可寻找资金实力雄厚并且本身拥有一定量数据和算力储备的企业支持平台建设，以减少资金投入和硬件的重复建设。

(2) 动态激励机制。建立与行业发展动态匹配的激励机制，根据技术进步调整资源贡献奖励方式；对长期稳定贡献的企业提供优惠或技术支持服务。

(3) 商业化产品矩阵。平台需开发多层次产品服务，满足不同规模企业需求；此外，结合行业需求动态更新服务模块，保持产品竞争力。

(4) 行业联盟与生态扩展。平台运营需与行业利益相关方深度合作，构建多方共赢的产业生态联盟；同时推动与上下游产业的合作，扩展平台影响力。

(5) 创新投资与合作模式。平台可探索与投资机构合作，吸引资本投入；可与科研机构联合开展攻关项目，将最新科研成果迅速转化为商业化产品。

5 汽车行业级智算数据平台建设模式及规划

5.1 平台建设模式分析

5.1.1 平台总体建设模式

汽车行业级智算数据平台需要同时满足资源共享、算法开发支持以及商业化服务的多层次需求，并为行业各方提供统一、高效的技术协作环境。在具体建设模式上，可选择集中化建设模式或分布式建设模式，二者具有不同的优势和适用

场景。为确保平台高效运作和长期发展，需要对两种模式进行深入比较和分析，并结合硬件设施和软件工具链建设的实际需求制定最优方案。其中：

（1）集中化建设模式，是指由核心企业或行业联盟牵头，统一规划、建设和运营平台的基础设施与服务内容。这种模式强调资源的集中管理与统筹调度，适用于大型企业集团或行业联合体。

（2）分布式建设模式，是指通过多节点的存储和算力架构，将平台资源分散部署于多个地点，同时通过网络实现资源的统筹和共享。该模式强调灵活性和扩展性，适合资源分布广泛且需求差异化的行业环境。

5.1.2 硬件资源建设模式

硬件是汽车行业级智算数据平台的物理基础，涵盖存储、计算和网络等关键资源。硬件层面需要充分考虑性能、成本、扩展性和安全性等因素，合理规划存储设备和算力资源的部署模式。

（1）存储资源建设模式

智算数据平台需存储海量驾驶数据、算法模型和仿真结果，同时满足高频访问需求和数据安全性要求。存储资源建设模式如下：

1）集中式存储。通过高性能存储设备（如 SAN 或 NAS）集中管理所有数据资源，适合对数据安全性和快速访问有较高要求的场景。

2）分布式存储。通过多个节点将数据存储在不同物理位置，并利用网络实现数据的统一管理。

3）混合存储模式。结合集中式和分布式存储的优势，可构建混合存储架构：将高频访问的核心数据集中存储，将历史数据、冷数据和冗余数据分布式存储。

（2）算力资源部署模式

算力资源是支撑智能驾驶数据处理和模型训练的核心硬件。针对算力需求，平台可选择自建算力中心或租用云算力等模式。

1）自建算力中心。企业自行投资部署高性能计算资源。

2）云算力租赁。通过第三方云服务按需使用计算资源。

3）混合算力模式。结合自建算力与云算力优点，适用于多样化计算需求。

5.1.3 数据及软件资源建设模式

数据和软件是智算数据平台的核心。软件系统是平台的核心技术基础，负责支撑数据管理、模型训练、资源流通等关键功能。

（1）数据资源建设模式

数据资源是平台的核心资产，其建设模式决定了数据采集、存储和应用的效率。平台需在以下维度进行模式选择：

1) 自建数据采集。由平台建立数据自采能力，如标准化数据采集车等，集中收集驾驶场景数据，确保数据结构一致性和标准化。该模式的优点在于易建立统一标准，数据质量更高；挑战在于经济成本高、耗时长，特别是面对我国复杂的道路交通环境，在数据量难以快速积累的情况下，数据规模难以满足各方需求。

2) 数据贡献模式。通过多企业联合的数据共享协议，各企业按自身需求上传或授权数据。该模式的优点在于资源整合效率高，覆盖场景更广泛；但存在数据标准不一、产权确权难、数据贡献积极性不高等问题，以及存在严苛的合规性限制，使得数据的共享共用存在巨大阻碍。

(2) 软件工具链建设模式

工具链是支持智能驾驶研发的基础技术组件，包括数据采集、清洗、标注、仿真和测试等模块。工具链开发模式直接影响平台的研发效率和使用便捷性。

1) 自主研发。完全掌控核心技术，平台可根据自身需求定制功能，确保系统与业务高度契合。同时，数据安全有保障，所有开发过程在内部完成，避免数据泄露或外部依赖问题。此外，工具链可随企业需求变化进行灵活升级。然而，自主研发也面临高开发成本和技术门槛，开发周期较长，难以快速适应市场变化。

2) 联合开发。融合多方技术优势，通过与科研机构、技术供应商合作，共享研发成果；同时，参与方根据需求分担开发投入，降低单方负担；此外，多方联合开发工具链有助于推动标准统一。然而，该方式也存在利益分配和成果所有权可能引发争议、多方协作可能增加项目管理复杂性、联合研发积极性等问题。

5.1.4 平台建设模式建议

当前，汽车行业级智算数据平台的建设需以行业服务为目标，通过整合数据、算力和算法资源来支持智能驾驶技术的发展。然而，由于现阶段数据合规性要求的严格性、行业对建设效率和经济性的高要求，以及技术路线和市场的不确定性，直接开展大规模平台建设风险巨大。建议采用“前期小规模集中式建设为主、后期持续分布增量式建设”的混合建设模式。

(a) 前期小规模、集中式建设试点

前期的集中式建设可依托智能网联汽车示范区及车路云一体化试点城市，以既有基础设施为支撑，快速搭建智算数据平台的核心功能。具体而言：

第一，车路云系统中的车端、路端和云端存储设备已积累大量高价值数据，应调研并整合现有资源，通过制定统一的存储架构，将路侧边缘计算节点中存储的实时感知数据与云端历史数据结合，实现临时数据与长期数据的高效分层存储。**第二**，整合车路云体系内算力资源，包括路侧边缘计算服务器和云端算力中心资源，通过智能算力调度算法打破算力孤岛，形成有机的算力资源池。**第三**，就数据标注、仿真生成和评价测试开展功能性接入，提供相关的智驾技术开发辅助服

务。**第四**，推动数据及软硬件设施标准化，促进不同设备和系统间的互联互通。

(b) 持续分布式、增量式建设运营

后续的分布式运营建设将重点放在建立一个开放、协作、创新的行业运营机制上，以激活行业中分散的数据、算力和算法资源。在这个阶段，平台的核心任务不再局限于硬件资源的部署，而是通过构建一套完善的运营机制，促进行业资源的高效流动。具体而言：

第一，搭建行业级数据共享与交易系统，实现数据确权、追踪和安全流通，打破行业资源孤岛。**第二**，设计灵活的算力交易和调度机制，鼓励企业贡献闲置算力资源，建立按需分配和收益分成的激励模式，降低行业的整体计算成本。**第三**，鼓励企业将自研算法模型通过平台授权使用，形成算法模型共享与交易市场，推动算法快速应用和迭代优化。**第四**，强化平台的行业核心地位，成为技术发展的“连接器”和“赋能者”，通过与头部企业、科研机构 and 中小企业广泛合作，逐步构建完整的智能驾驶生态系统。**第五**，组织技术论坛、竞赛和联合研究，进一步吸引行业资源参与。**第六**，通过整合资源、协调技术路径，推动数据格式、接口协议、算法评测方法等标准化进程，提升公信力和行业影响力。

5.2 平台阶段性建设规划

5.2.1 短期目标（1-2 年）：核心功能模块建设

短期阶段，平台需以基础功能的快速搭建和资源整合为主要目标，集中完成核心模块的部署和初步运营。

数据层面：整合车路云数据资源，建立标准化数据存储和管理体系；通过对数据的归集与分类，形成高效的多层次数据存储结构。

算力层面：重点整合试点区域内的算力资源，依托路侧边缘计算节点和云端算力中心，构建区域内统一的算力池。通过算力调度算法充分利用闲置算力，支撑简单数据预处理和局部软件服务功能。建立闲置算力信息发布系统。

算法层面：提供基础的模型训练支持，开发智能驾驶初级算法库，包括数据标注算法和测试评价功能，支持企业在平台上进行数据和场景驱动的模式测试。

生态层面：建立基础的数据确权与共享机制，确保数据流通的安全性与合法性。推动试点城市的企业与平台开展合作，形成初步的行业协作网络。同时，依托政策支持与资金补贴，为平台建设提供资源与技术保障。

5.2.2 中期目标（3-5 年）：行业资源整合与机制优化

中期阶段，目标是扩展平台覆盖范围，推动行业资源整合，优化平台运营机制，提升平台的服务能力和影响力。

数据层面：进一步扩展数据来源，接入更多城市及区域的数据资源，通过统一的数据接口标准和流通规则实现全国范围的数据共享。同时，开发高保真仿真数据生成工具，补充稀缺场景数据，丰富平台数据资源池。

算力层面：在原有算力池的基础上，增加更多分布式节点以支持全国范围内的计算需求。引入异构计算资源（如 GPU、FPGA），优化不同场景下计算性能。

算法层面：完善平台算法模型库，支持数据标注、仿真生成和测试评价算法的开发和部署。开发仿真测试工具链，支持复杂场景下的模型验证。

生态层面：进一步完善运营机制，建立数据交易与算法授权平台，推动数据、算法模型的商业化应用。通过举办行业技术竞赛和论坛，吸引更多企业参与，提升平台影响力。积极参与标准制定，包括数据格式、接口协议和评测方法等。

5.2.3 长期目标（5 年以上）：生态系统完善与商业化服务

长期阶段，平台需从单一服务功能转向多元化生态系统，构建覆盖全国的资源协作网络，实现行业内外广泛连接。

数据层面：建立面向全国的高效数据流通体系，推动数据资产化进程。支持实时数据的动态更新和历史数据的高效查询，始终保持数据的时效性和全面性。推动国际化数据合作，探索跨境数据共享机制，扩展平台数据资源规模。

算力层面：构建全国算力调度网络，实现算力资源在跨区域和跨行业场景中的流转与共享。探索边缘计算与云计算相结合的模式，支持实时决策与大规模训练任务的灵活切换。

算法层面：推动大规模智能驾驶模型开发和应用，支持自动驾驶算法的端到端训练与优化。通过与高校和科研机构合作，探索多模态感知和深度学习算法的前沿技术。逐步推出行业级算法评测和认证服务，帮助企业快速验证算法性能。

生态层面：形成以平台为核心的智能驾驶生态系统，构建多层次的商业化服务模式，包括数据订阅、算力租赁、仿真测试等服务。推动平台国际化发展，与海外平台形成技术和资源互补，进一步提升行业影响力。

5.2.4 收支平衡预期

短期阶段，平台需依赖政府补贴和行业融资来支持基础设施建设与研发投入，短期内难以实现盈利。

中期阶段，随着服务能力的提升和用户群体的扩大，平台收入来源将从数据服务、算力租赁和模型授权中增长，预计可实现盈亏平衡。

长期阶段，平台通过多元化服务实现稳定盈利，进一步扩大资源储备与生态覆盖范围，形成可持续发展的商业模式。

附录 1：国内智算中心建设情况

附表 1-1 智能驾驶相关云端算力规模情况

企业名称	算力规模（EFLOPS）	信息公开时间
吉利汽车	1.2	2025E
	23.5（算力联盟）	2025.3
小米汽车	11.45	2025.3
小鹏汽车	10	2025.4
理想汽车	9.3	2025.4
蔚来汽车	1.4	2023.9
长城汽车	3	2025.2
长安汽车	1.42	2023.8
商汤绝影	23	2025.2
华为车 BU	10+	2024 年底

资料来源：根据公开信息整理

附表 1-2 部分科技厂商自建智算中心

名称	地点	建设内容
阿里云张北超级智算中心	河北	总建设规模为 12000PFLOPS
阿里云涿州智算中心	河北	规划 20 万台服务器
阿里飞天云智能华东算力中心	上海	二期 5000 个标准机架数
阿里云乌兰察布智算中心	内蒙	建设规模为 3000PFLOPS
商汤科技人工智能计算中心	上海	一期 5000 个机柜，峰值训练算力 3740PFLOPS
商汤科技前海深港人工智能算力中心	深圳	一期建设算力达 500PFLOPS
万界数据宜昌智算中心	湖北	规划机架 26400 个，一期 4400 个，2500PFLOPS
万界数据鄂尔多斯智算中心	内蒙古	建成 3000PFLOPS 算力
万界数据巴州智算中心	新疆	一期规划 5000PFLOPS，二期规划 15000PFLOPS
万界数据张掖高台智算中心	甘肃	一二期建设 7000PFLOPS，三期建设 15000PFLOPS
万界数据资阳智算中心	四川	一期规划 880PFLOPS
华为云贵安数据中心	贵州	规划百万级服务器规模
华为云乌兰察布数据中心	内蒙古	规划百万级服务器规模
华为云芜湖数据中心	安徽	规划百万级服务器规模
百度智能云-昆仑芯（盐城）智算中心	江苏	算力规模达到 200PFLOPS
百度天津智算中心	天津	-
百度沈阳智算中心	辽宁	规划算力 500PFLOPS，一期 208PFLOPS
百度济南智算中心	山东	-
腾讯长三角人工智能先进计算中心	上海	目前已有 1 万余台服务器投入工作
腾讯智慧产业长三角（合肥）智算中心	安徽	-
润泽（西南）国际信息港项目	重庆	规划 AIDC 高等级智算中心机房 8 栋，单栋机房机柜数量为 4184 架，总机柜容量 3.2 万架以上

润泽（廊坊）国际信息港新型智算中心	河北	容纳 16384 台超性能服务器并联
-------------------	----	--------------------

资料来源：根据公开信息整理

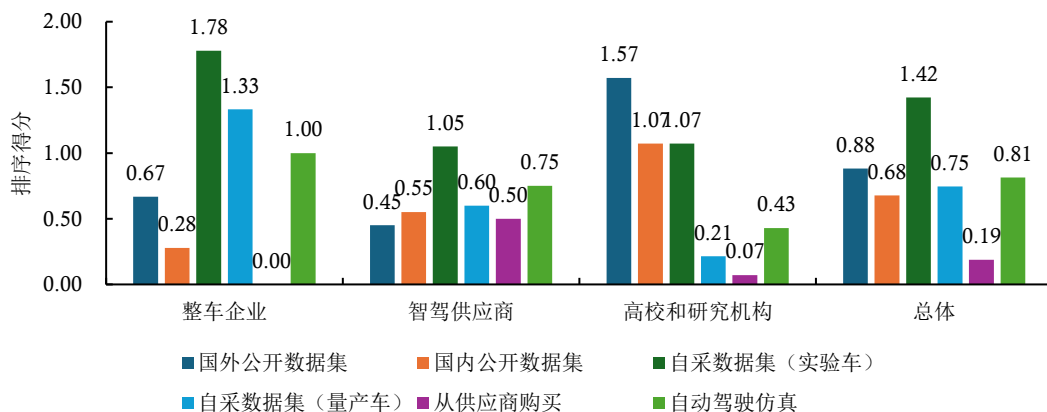
附表 1-3 部分政府智算中心建设情况

时间	名称	城市	算力	芯片
2019	横琴先进智能计算中心	广东	4EFLOPS	寒武纪
2020	深圳鹏城云脑	深圳	1000PFLOPS	一期：英伟达 二期：昇腾
2021	武汉人工智能计算中心	湖北	200PFLOPS	-
2021	南京智能计算中心	江苏	1250PFLOPS	寒武纪
2021	西安未来人工智能计算中心	陕西	一期 300PFLOPS，总规划 1000PFLOPS	昇腾
2021	中原人工智能计算中心	河南	一期 100PFLOPS，总规划 300PFLOPS	昇腾
2021	昆山智能计算中心	江苏	500PFLOPS	寒武纪
2022	青岛人工智能计算中心	山东	100PFLOPS	昇腾
2022	成都智算中心	四川	300PFLOPS	昇腾
2022	广州人工智能融合赋能中心	广东	总规划 1000PFLOPS	昇腾
2022	大连人工智能计算中心	辽宁	100PFLOPS	-
2022	重庆人工智能创新中心	重庆	总规划 1000PFLOPS	昇腾
2023	北京昇腾人工智能计算中心	北京	短期 400PFLOPS，远期 1000PFLOPS	昇腾
2023	天津市人工智能计算中心	天津	规划 300PFLOPS，首批 100PFLOPS	昇腾
2023	福建人工智能计算中心	福建	总规模 400PFLOPS，一期 105PFLOPS	-
2023	长沙昇腾人工智能创新中心	湖南	1024PFLOPS	昇腾
2023	沈阳智能计算中心	辽宁	总规划 500PFLOPS，一期 208PFLOPS	-
2023	甘肃庆阳算力中心	甘肃	3500PFLOPS，计划 6000PFLOPS	英伟达，昇腾
2023	湖南人工智能算力数据中心	湖南	2000PFLOPS	-
2023	河北人工智能计算中心	河北	规划 300PFLOPS，一期 100PFLOPS	昇腾
2023	济南 AI 算力中心	山东	规划 1000PFLOPS	-
2023	宁夏国产千卡智算集群建设项目	宁夏	-	摩尔线程
2023	石家庄人工智能计算中心	河北	规划 500PFLOPS，一期 100PFLOPS	-
2024	北京亦庄人工智能公共算力平台	北京	3000PFLOPS	-
2024	连云港大数据中心	江苏	2000PFLOPS	-
2024	北京石景山智能算力中心	北京	投运 200PFLOPS，规划 610PFLOPS	-

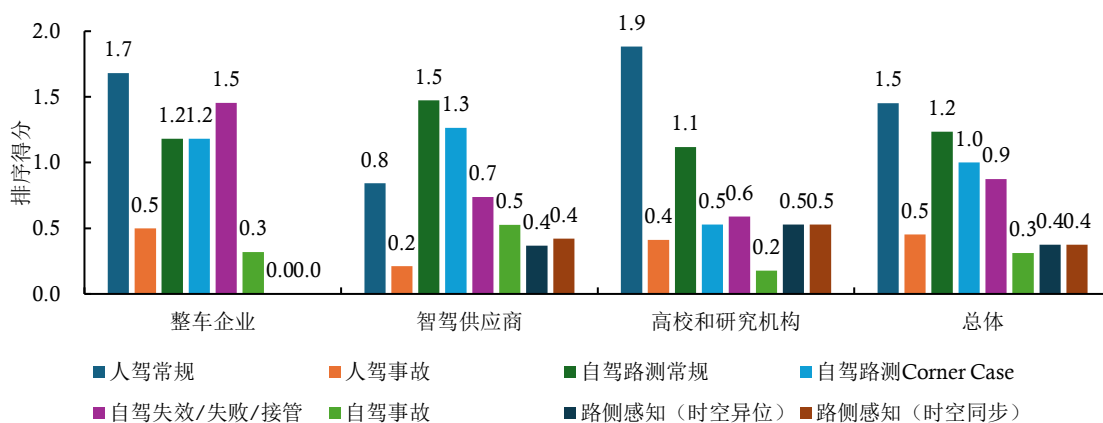
资料来源：根据公开信息整理

附录 2：国内智能驾驶基础资源现状调研分析图

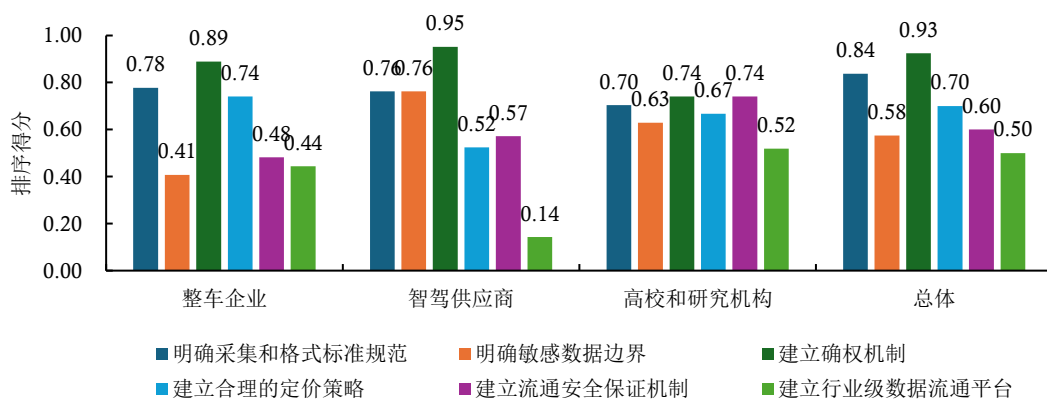
针对排序型问题，纵轴排序得分采用计算方法为：选项平均综合得分 = $(\sum \text{频数} \times \text{权值}) / \text{本题填写人次}$ ，反映重要性。



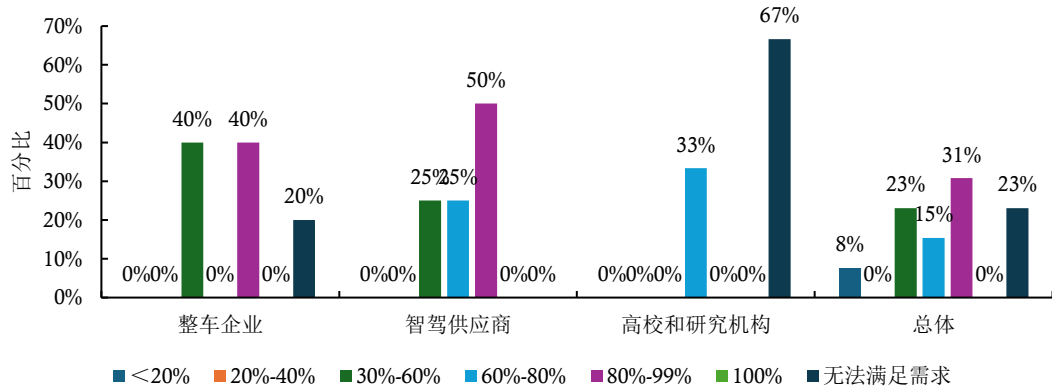
附图 2-1 数据来源



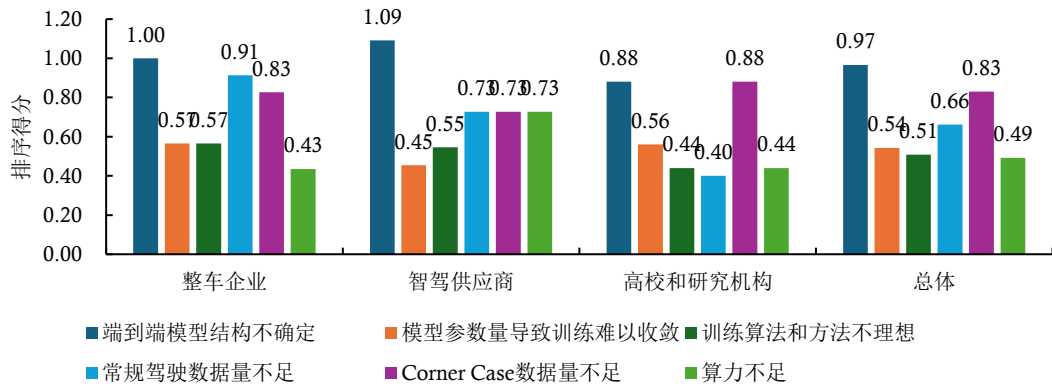
附图 2-2 数据分布



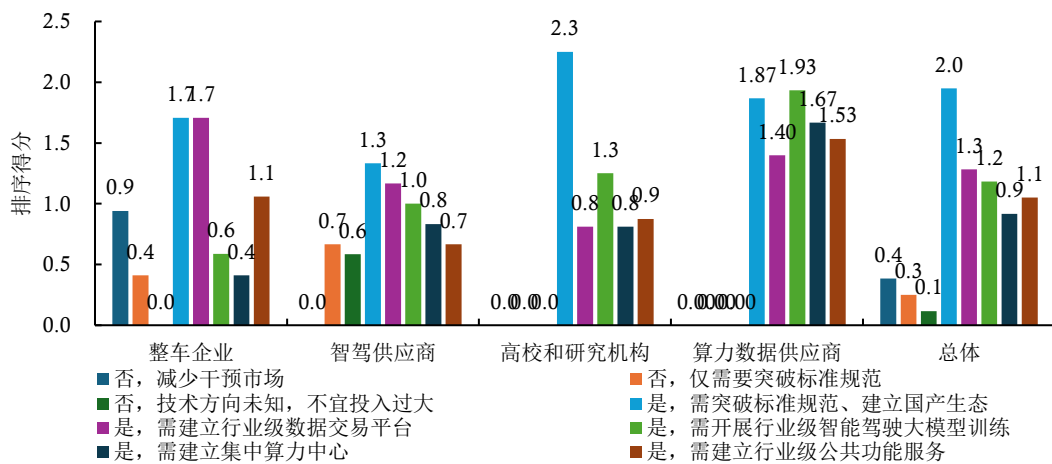
附图 2-3 数据流通需要解决的问题



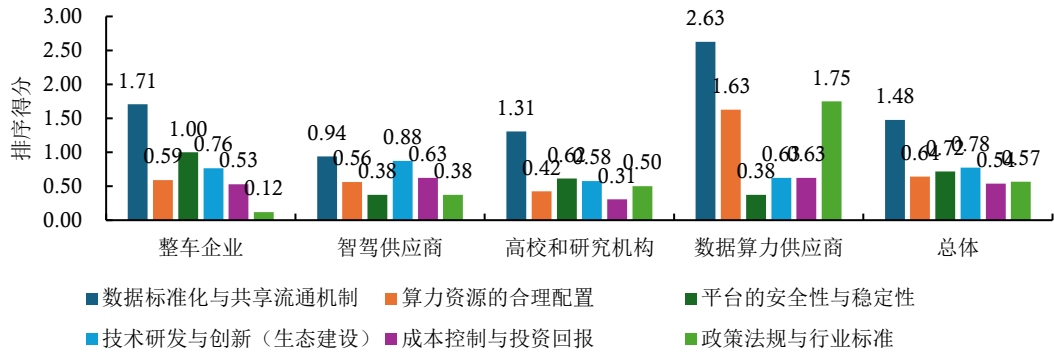
附图 2-4 算力利用率



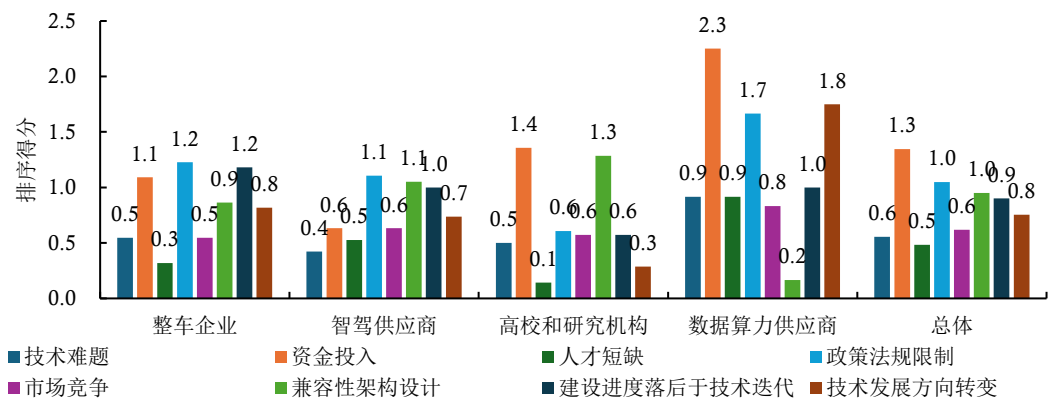
附图 2-5 端到端智能驾驶模型训练难点



附图 2-6 是否有必要建立行业级智能驾驶智算数据平台



附图 2-7 平台建设需要优先解决的问题



附图 2-8 平台建设潜在风险