多模态大语言模型技术及应用标准领航研究报告

中国汽车标准化技术委员会
智能网联汽车分技术委员会
车用人工智能标准专项组

2025年7月

前言

近年来，人工智能（AI）技术的飞速发展为智能汽车领域带来了前所未有的机遇。智能座舱作为智能汽车的重要部分，得益于AI技术的支持，有了很大程度上的进步，也逐步向着实际应用靠齐。作为智能交通的重要组成部分，智能汽车正逐步走向商业化，并成为全球汽车产业技术创新和竞争的核心。

在全球智能网联汽车发展的竞争中，中国正处于关键的技术突破期，面临着从传统汽车制造到智能化、网联化转型的巨大挑战。AI技术在智能汽车相关领域的突破为中国汽车产业提供了巨大的发展潜力。为了在激烈的国际竞争中占据一席之地，中国必须紧抓智能驾驶、车联网与共享出行等新兴技术的发展机遇，推动汽车产业的快速转型升级，实现AI驱动的“智能化”成为中国汽车产业的核心竞争力。

本报告涉及了智能座舱AI技术及其应用场景的标准化问题，首先对智能座舱的现状进行了分析，基于此进行了对于AI和大模型的应用场景和技术路线的探索，也对其中最为重的视觉和多模态的交互进行了发展趋势和技术路线的讨论；最后进行了在AI应用上关键技术的讨论，将之前的讨论拉入实际，也对其测试和评价的需求进行了一定的规定。

因此，为了推动智能座舱AI技术的全面应用与发展，推动产业升级与技术创新，中国汽车技术研究中心有限公司联合科大讯飞股份有限公司、小鹏汽车、北京车和家汽车科技有限公司、比亚迪汽车工

业有限公司、清华大学苏州汽车研究院、东软集团股份有限公司、重庆长安汽车股份有限公司、上海临港绝影智能科技有限公司、上汽大众汽车有限公司、厦门金龙旅行车有限公司、中国移动上海产业研究院、一汽解放汽车有限公司、中国汽车工程研究院股份有限公司、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）、中兴通讯股份有限公司、蔚来汽车科技(安徽)有限公司、北京与之科技有限公司、中国汽车战略与政策研究中心、北京汽车研究总院有限公司等19家单位，共同编写完成《多模态大语言模型技术及应用标准领航研究报告》。

在本研究报告编制过程中，各起草单位参阅了大量材料，并借鉴了行业的部分素材，鉴于篇幅有限，这里不一一列举，仅作诚挚的感谢！

在此，再次衷心感谢参与研究报告编写的各个单位和组织：中国汽车技术研究中心有限公司、科大讯飞股份有限公司、小鹏汽车、北京车和家汽车科技有限公司、比亚迪汽车工业有限公司、清华大学苏州汽车研究院、东软集团股份有限公司、重庆长安汽车股份有限公司、上海临港绝影智能科技有限公司　、上汽大众汽车有限公司、厦门金龙旅行车有限公司、中国移动上海产业研究院、一汽解放汽车有限公司、中国汽车工程研究院股份有限公司、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）、中兴通讯股份有限公司、蔚来汽车科技(安徽)有限公司、北京与之科技有限公司、中国汽车战略与政策研究中心、北京汽车研究总院有限公司。

主要编写人：华一丁、刘俊峰、雷琴辉、童鹏、王雪初、王潼、李淑玲、卢俊蓉、陈品瑄、张伟豪、姜彦吉、王欣蕊、郭佳鑫、孔麒、何子豪、贾龙、李天然、苏鹏飞、王路宝、 范亦卿、周泽杨、王和俊、陈艳梅、朱丽敏、王峰、俞海山、刘彦革、丁俊勇、王小亮、卢晶、谢良辉、张志新、李庆庆、吴天舒、常婉渲、王荣、李绍鹏、陈晓、黄程、陈思云、王凯、牟文珺、陈韵巧、赵佳、王金奎、郭璋。

目录

第一章 智能座舱 AI 技术应用现状

自 1956 年“人工智能（Artificial Intelligence, AI）”概念的诞生以来，这一领域已经历了数十年的蓬勃发展。1970 年标志着人工智能的第一个春天，当时推出了首款人工智能软件 Logic Theorist 和第一部神经网络著作《Perceptron》。随着第五代计算机的兴起，1990 年人工智能迎来了第二个黄金时期，其标志性事件包括 1997 年“深蓝”计算机在国际象棋比赛中战胜世界冠军。然而，随后第五代计算机的失败使得人工智能领域进入了一段寒冬。

直到 2006 年，深度学习在语音识别领域的显著突破将人工智能推向了第三次浪潮。在此期间，DNN（深度神经网络）、CNN（卷积神经网络）、GAN（生成对抗网络）、Attention 机制和 Transformer 等人工智能架构不断更新迭代，AlphaGo、Squad 等杰出产品也不断涌现，展现了人工智能技术的巨大潜力和广泛应用前景。

进入 2022 年，ChatGPT 的问世标志着人工智能进入了第四次浪潮，这次浪潮以大模型产业的新变革为特点。ChatGPT 等先进模型展示了人工智能在自然语言处理领域的巨大进步，为未来的智能应用开辟了新的道路。人工智能的第四次浪潮预示着更为智能化、自动化的未来，将深刻影响社会的各个方面。

图 1-1 人工智能发展趋势

1.1 智能座舱 AI 技术应用场景和效果分析

智能化已成为智能网联汽车在电动化之外的另一重要发展方向，AI 在座舱中的应用将重构用户的智能座舱体验。目前在座舱领域，AI 技术主要应用于车

载语音交互、视觉交互、多模态交互、其它开放性任务场景，实现效率、智能、情感、舒适等多方面提升。

1.1.1 车载语音交互

车载语音交互是指车辆内部的交互界面采用语音作为主要的输入和输出方式进行操作和反馈的技术。这种技术通过语音识别技术将驾驶员的语音指令转化为计算机可理解的指令，并通过语音合成技术将系统的反馈信息以语音形式传达给驾驶员。

针对当前车载语音系统智能化、个性化、情感化、交互性不足等问题，自然语言处理类大模型可以赋予车载语音系统以智能和情感，从而使车载语音系统能够处理完整对话并可以保持对前后文的理解，能够记录用户的喜好和习惯，提供更加准确和个性的响应。

图 1-2 车载语音交互示例

1.1.2 视觉交互

座舱视觉交互是指在车辆内部，通过视觉显示技术与驾驶员或乘客进行信息交换和互动的过程。它通常涉及车辆内部的各种显示屏，如中控屏、仪表盘、HUD（抬头显示系统）等，以及与之相关的软件和界面设计。座舱视觉交互旨在提供直观、清晰、易于理解的视觉信息，以便驾驶员或乘客能够迅速了解车辆状态、导航信息、娱乐内容等，从而增强驾乘体验和安全性。

相较于通用的拨杆、按键、触屏等被动人机交互方式及车载语音“问答式”人机交互方式，以 DMS、OMS、儿童监控等代表的视觉 AI 的应用，通过车内

乘员的监控，自动监测和识别乘员的行为、意图和身体状态，提供主动式推荐服务，极大提升座舱智能化体验。

目前 DMS 等视觉交互功能，识别准确性和触发时机都存在较大不足，造成用户体验不佳。搭载 AI 算法模型，在用户使用中不断积累数据，持续训练，能够提升识别准确率，获取用户真实意图，并根据用户习惯和喜好给出最合理的反馈，提升座舱视觉交互体验，真正实现千人千面的座舱极致体验。

图 1-3 广汽传祺超感交互智能座舱

1.1.3 多模态交互

多模态交互是指通过融合多种感知模态（如声音、视觉、动作、环境等）来实现的人与机器或人与人之间的交互方式。这种方式模拟了人类之间自然、多样的交流形式，旨在提供更直观、高效、自然的人机交互体验。

在车载语音上车之前，座舱交互的方式较为单一，主要通过按键和触屏方式实现座舱功能操作。随着语音交互、视觉交互、手势识别、香氛、智能灯光等功能搭载应用，座舱交互方式更加丰富。在座舱复杂应用场景下，单个交互方式受技术本身的限制，交互的效率和准确性会存在一定的技术瓶颈，需要通过多模融合丰富座舱应用场景，提升座舱交互体验。

多模态大模型可以助力融合语音、视觉、手势、文字等多种交互方式，满足用户在不同场景下的不同使用习惯，从而赋予用户良好的人车交互体验。

图 1-4 长城 Coffe OS 智慧座舱系统

1.1.4 开放式任务

开放式任务指的是通过 AI 技术来检索完成某一任务需要的信息，并通过系统传感器、数据库等渠道获取跨领域信息来完成复杂的任务。除了提供卓越的交互体验外，基于先进 AI 技术开发的智能座舱产品还能实现一系列令人瞩目的功能，极大地提升了驾驶的便捷性、安全性和个性化体验。

首先，智能推荐功能能够根据驾驶员的偏好和习惯，自动推荐音乐、电台、导航路线等，使驾驶过程更加愉悦和高效。通过学习和分析驾驶员的日常使用数据，智能座舱能够准确理解并满足他们的个性化需求，从而显著提升每次出行的体验。

其次，行车辅助功能利用 AI 技术，为驾驶员提供全方位的安全保障。通过实时监测车辆周围环境、交通状况以及驾驶员的驾驶行为，智能座舱能够提前预警潜在危险，并自动采取紧急制动、车道保持等辅助措施，有效避免事故的发生。

再者，车辆设置功能允许驾驶员通过简单的语音指令或触摸屏操作，轻松调整座椅、空调、车窗等设备的设置。智能座舱能够记忆驾驶员的偏好设置，并在他们再次上车时自动调整至最佳状态，为驾驶员提供舒适的驾乘环境。

最后，服务支持功能通过智能座舱的联网能力，为驾驶员提供丰富的在线服务。无论是查询天气、路况、加油站等实用信息，还是预约维修、保养等售后服务，驾驶员都能通过智能座舱轻松完成。此外，智能座舱还支持多种支付方式，让出行更加便捷无忧。

总之，基于 AI 技术开发的智能座舱产品通过实现智能推荐、行车辅助、车辆设置、服务支持等功能，为驾驶员带来了前所未有的智能出行体验。这些功能的加入不仅提升了驾驶的便捷性和安全性，还使每一次出行都更加个性化、舒适和愉悦。

1.2 智能座舱 AI 技术应用的问题和难点

1.2.1 AI 大模型部署问题

1) 云端部署难点

硬件兼容性问题：国产化的云端硬件可能与国外主流 AI 硬件平台存在差异，这可能导致 AI 大模型在部署时面临硬件兼容性问题。需要对 AI 大模型进行适配和优化，以适应国产云端硬件的性能和特性。

软件生态与工具链支持：国产云端平台可能缺乏完善的软件生态和工具链支持，如深度学习框架、模型训练与推理工具等。这将增加 AI 大模型在云端部署的难度和复杂性。

算力与资源调度：AI 大模型通常需要大量的计算资源，而国产化平台可能面临算力不足的问题。需要设计高效的资源调度和分配策略，以充分利用有限的计算资源。

2) 端侧部署难点

硬件性能限制：端侧设备（如车机芯片、带宽等）的硬件性能通常有限，可能无法满足 AI 大模型的实时性和高效性要求。需要对 AI 大模型进行压缩和优化，以适应端侧设备的硬件性能。

能耗与散热：端侧设备通常需要考虑能耗和散热问题，以保证设备的持续运行和稳定性。AI 大模型的部署需要在保证性能的同时，降低能耗和散热。

模型更新与维护：端侧设备的 AI 大模型需要定期更新和维护，以保证模型的准确性和时效性。然而，由于端侧设备的分散性和多样性，模型更新和维护的难度较大。

1.2.2 安全和隐私问题

AI 应用于智能座舱，需要本地存储和上传大量的个人信息、视频、语音等数据，如何保证数据存储安全，不被网络攻击和窃取，也是人工智能在汽车广泛

应用必须要解决的问题。

在部署 AI 大模型时，需要确保数据的隐私保护。需要建立严格的数据管理制度和安全防护措施，以防止数据泄露和滥用。端侧设备通常涉及用户隐私问题，需要在 AI 大模型部署时考虑用户隐私和权限管理。需要建立严格的用户隐私保护机制和权限管理制度，以防止数据泄露和滥用。

1.2.3 数据问题

数据是 AI 大模型的核心要素之一，数据质量和数据规模很大程度决定了 AI 的整体效率和质量。高质量训练数据获取面临多个难点，这些难点不仅影响模型的训练效率和效果，还直接关系到模型在汽车应用中的性能。主要包括：

1) 数据多样性

AI 模型通常需要处理各种复杂和多样化的场景，因此高质量的训练数据需要具备广泛的多样性。然而，在实际获取过程中，往往难以覆盖所有可能的场景和情况，导致训练数据的多样性不足。

2) 数据标注准确性

对于监督学习任务，训练数据需要准确的标注。然而，由于标注任务通常需要人工完成，标注质量往往受到标注人员专业能力和主观判断的影响，导致标注数据存在误差和不一致性。

3) 数据隐私和合规性

在获取训练数据时，需要确保数据的隐私性、安全性和合规性。这包括遵守相关法律法规、保护用户隐私、保护数据安全和避免侵犯他人权益。然而，在实际操作中，数据的隐私、安全和合规往往难以同时满足，需要权衡各种因素并制定相应的解决方案。

4) 数据规模和效率

AI 模型需要大量的训练数据才能达到理想的性能。然而，在实际获取过程中，往往面临数据规模不足或获取效率低下的问题。这可能需要采用更高效的数据采集方法、优化数据存储和处理流程以及利用云计算等先进技术来提高数据获取效率。

5) 数据质量评估

对于训练数据的质量评估是一个重要但困难的任务。因为数据质量不仅取决

于数据的准确性和多样性，还受到数据分布、噪声和异常值等多种因素的影响。如何准确评估训练数据的质量，并根据评估结果调整数据获取策略，是一个需要解决的问题。

1.3 智能座舱 AI 技术应用相关标准分析

目前人工智能领域相关标准较多，主要为通用 AI 技术相关标准，但针对车载 AI 应用的标准较少。国内，针对车载座舱 AI 技术相关的标准仍在持续建设中，在汽车信息安全、个人信息安全和智能语音交互系统测试方面，国家标准提出了部分技术要求，能够对智能座舱 AI 应用提供参考。

表 1-1 相关标准及测试规程

标准名称	发布单位&类别	主要内容
《面向行业的大规模预训练模型技术和应用评估方法 第 4 部分：汽车》	中国信通院；团标	聚焦汽车行业高质量发展，结合汽车的研发、生产、销售、使用等全过程，形成汽车大模型应用成熟度评价方法，便于各方衡量汽车大模型的应用能效，助推汽车大模型产品升级优化
GB_T36464.5-2018 信息技术 智能语音交互系统 第 5 部分：车载终端	全国信息技术标准化技术委员会；国标	规范了车载终端智能语音交互系统的术语和定义、系统框架、要求和测试方法。
GB/T41797-2022 驾驶员注意力监测系统性能要求及试验方法	全国汽车标准化技术委员会；国标	旨在规范驾驶员注意力监测系统的设计和性能，确保该系统能够有效地监测和评估驾驶员的注意力水平，从而提高道路安全性。标准的制定有助于统一行业对驾驶员注意力监测技术的要求，推动相关技术的健康发展，并为车辆制造商和系统供应商提供明确的指导。
汽车智能座舱语音分级与测评方法	中国智能网联汽车产业创新联盟	旨在建立一套科学、有效的测评方法，以指导智能座舱语音产品的研发设计，

	（CAICV）；团标	为企业开展智能座舱语音水平测试提供依据。
汽车智能座舱智能水平测试与评价方法	国汽（北京）智能网联汽车研究院有限公司等机构；团标	旨在解决现有智能座舱产品与用户需求之间的不匹配问题，并为汽车智能座舱的智能化水平提供一个统一的测试和评价方法。
汽车智能座舱交互体验测试评价规程	中国汽车工业协会；团标	旨在通过不同交互模态在座舱中执行相应交互任务时，评价用户体验优劣的标准。
道路车辆免提通话和语音交互性能要求及试验方法	中国汽车技术研究中心有限公司；国标	旨在探究车载语音交互系统的交互性能表现及用户体验。

第二章 智能座舱 AI 技术应用场景与技术路线探索

2.1 智能座舱语音交互场景发展趋势与技术路线

在智能座舱之中，语音交互作为最自然、最便捷的人机交互方式，其在智能座舱中的应用和发展趋势，以及所依托的技术路线，成为了推动汽车行业向智能化转型的关键力量。

2.1.1 智能座舱语音交互的现状与重要性

智能座舱技术作为汽车工业与信息技术深度融合的产物，正引领着未来出行方式的变革。它不仅仅是一个驾驶空间，更是一个集成了人工智能、物联网、大数据分析等前沿技术的智能生态系统。在这个系统中，语音交互扮演着无可替代的角色，它不仅改变了传统的驾驶交互模式，也深刻影响着乘客的乘坐体验。

（1） 现状概览

A 技术进展

近年来，智能座舱语音交互技术经历了飞速发展。基于深度学习的自然语言处理技术、语音识别技术以及语音合成技术的不断突破，使得语音助手的理解能力、反应速度和交互体验有了质的飞跃。现在的语音系统不仅能准确识别多种语言和方言，还能理解复杂的语境和指令，支持自然流畅的对话，甚至能够根据上下文进行逻辑推理和预测用户需求。例如，当用户说“我有点冷”，系统不仅能自动调高车内温度，还能贴心询问是否需要关闭车窗。

B 多场景应用

智能座舱的语音交互已广泛应用于多个场景，包括但不限于导航、娱乐、通讯、车辆控制（如空调、车窗、座椅调节）以及紧急求助服务。特别是在导航方面，通过语音输入目的地，系统能快速规划路线，避免了手动输入地址带来的安全隐患。娱乐方面，用户可以通过简单的语音指令切换歌曲、电台或者有声读物，让旅途变得更加轻松愉快。

C 个性化与情感化

为了更好地满足用户的个性化需求，现代智能座舱语音系统支持定制化设置，如语音助手的名字、音色、甚至性格特点，让用户感觉像是与一位熟悉的朋友交

流。此外，情感识别技术的应用使得语音助手能够感知用户的情绪状态，并据此调整回应策略，比如在用户表现出疲劳时，主动提供休息建议或播放轻松音乐，增强了人机交互的情感连接。

（2） 重要性解析

A 安全性提升

在驾驶过程中，任何手动操作都可能分散驾驶员注意力，增加事故风险。语音交互的引入，使驾驶者无需动手即可完成大多数车内功能操作，有效减少了视线转移和手动操作频率，极大提升了行车安全性。尤其是在高速行驶或复杂路况下，这种非接触式的交互方式尤为重要。

B 驾驶体验优化

语音交互不仅简化了操作流程，还为用户提供了一种更加自然、直接的交流方式，使得人车互动如同与真人对话般顺畅。这种无缝的交互体验，不仅减轻了驾驶压力，还增加了驾驶乐趣，让每一次出行都变成一次愉悦的旅程。

C 适应多元需求

智能座舱的语音交互技术还特别关注到了不同用户群体的需求，包括但不限于老年人、残障人士等。对于这些群体而言，语音控制车辆功能尤为关键，它降低了操作门槛，使更多人能够享受科技带来的便利，体现了社会包容性和技术的人文关怀。

（3） 智能化趋势的体现

随着自动驾驶技术的逐步成熟，智能座舱将成为未来汽车的核心。在高度自动化驾驶场景下，人与车的交互将更加频繁，语音作为最自然的人机交流方式，其智能化水平将直接影响到乘客的乘车体验和对车辆品牌的认同感。因此，不断提升语音交互的技术深度与广度，是汽车制造商打造差异化竞争优势的关键所在。

2.1.2 语音交互场景的发展趋势

智能座舱语音交互技术的不断演进，预示着未来汽车内部空间将转变为一个高度智能、个性化且情感丰富的交流环境。随着技术边界的拓展，几个关键的发展趋势正逐渐成形，它们共同勾勒出一个更加人性化、高效且富有预见性的车载交互未来。

（1） 高度自然化与情感化的交互体验

在不久的将来，智能座舱语音系统将借助深度学习算法的深入应用，以及情感识别技术的精进，提升与人类对话更高层次的相似度。这意味着，系统不仅能理解命令式的指令，还能领悟用户言谈间的情绪变化，通过用词、语调、停顿等细微差别，识别用户的高兴、悲伤、急躁等情绪状态，并做出相应回应。例如，当检测到驾驶者心情低落时，语音助手可能会用更加温柔的语气给予安慰，或者播放轻松愉快的音乐来调节气氛。这样的设计，让车辆不仅仅是代步工具，更像是一个懂得倾听和慰藉的伙伴，极大地增强了人车之间的情感纽带。

（2） 多模态融合交互

单一维度的语音交互正逐渐向多模态交互转变，这标志着智能座舱将整合视觉、听觉、触觉等多种感官通道，以更全面的方式理解乘客的意图。通过车内安装的高清摄像头捕捉乘客的面部表情和肢体动作，结合语音指令，系统能更精准地判断乘客的真实需求。例如，当孩子在后座哭泣时，系统不仅听到哭声，还能看到孩子的表情和父母的安抚动作，从而自动调节车内氛围灯至温馨模式，播放柔和的摇篮曲，营造一个更加舒适的环境。这种多维度的信息整合，使智能座舱的服务更加细腻、个性化，也更加符合人类自然交流的习惯。

（3） 预测性与主动服务

基于大数据分析和机器学习算法，智能座舱语音系统将能够学习每位乘客的偏好、行为模式，甚至生活习惯，从而在用户提出需求之前就能预判并主动提供服务。例如，系统会根据用户的日常通勤路线，在早晨自动推荐最佳出行路线并规避拥堵；在经常用餐的时间，主动询问是否需要预定常去的餐厅；甚至在检测到天气变化时，提前调节车内温度和湿度，确保乘客始终处于最舒适的环境中。这种从被动响应到主动服务的转变，极大地提升了用户体验，也让智能座舱更加贴合用户的生活节奏和实际需求。

（4） 跨场景无缝连接

随着车联网技术的不断成熟，智能座舱语音交互的边界将进一步拓宽，实现家庭、办公室与车辆间的无缝连接。无论是家中的智能家居控制，还是办公室的日程安排，都能通过智能座舱的语音系统进行远程操控和信息同步。例如：早上出门前，通过车内语音助手就可以开启家中安防系统，调节办公室的空调温度；

下班路上，简单一句话就能查看家中冰箱存货，提前启动电饭煲烹饪晚餐。这种跨场景的一体化体验，不仅极大地方便了用户的生活，也将个人数字生活空间的概念推向了一个新的高度，真正做到了无论身在何处，都能享受到一致且连贯的智能服务。

2.2 技术路线分析

座舱语音交互技术是一种允许用户通过语音命令与车辆进行交互的技术，通常包括唤醒、识别、理解、播报四个环节。唤醒是启动语音交互系统的初始步骤，通常需要说出特定的唤醒词或短语（语音唤醒），系统通过内置的麦克风识别用户的语音指令（语音识别），系统将接收到的语音信号理解为可识别的指令（语义理解），系统执行完指令后，会通过语音反馈结果给用户（语音合成）。以下是座舱语音交互主要技术定义：

- 声学前端：声学前端处理技术当语音信号被各种各样的噪声干扰、甚至淹没后，从噪声背景中提取有用的语音信号，抑制、降低噪声的干扰。
- 语音唤醒：处于音频流监听状态的语音交互系统，在检测到特定的特征或事件出现后，切换到命令字识别、连续语音识别等其他处理状态的过程。
- 语音识别：其目标是将人类的语音中的词汇内容转换为计算机可读的输入，例如按键、二进制编码或者字符序列。
- 语义理解：是针对人类语言的分析理解过程。它是把自然语言内容，转换为有一定结构的文本数据，使应用能够理解用户的意图。
- 语音合成：通过机械的、电子的方法合成人类语音的过程。

2.2.1 依托算力平台提升基础算法效果

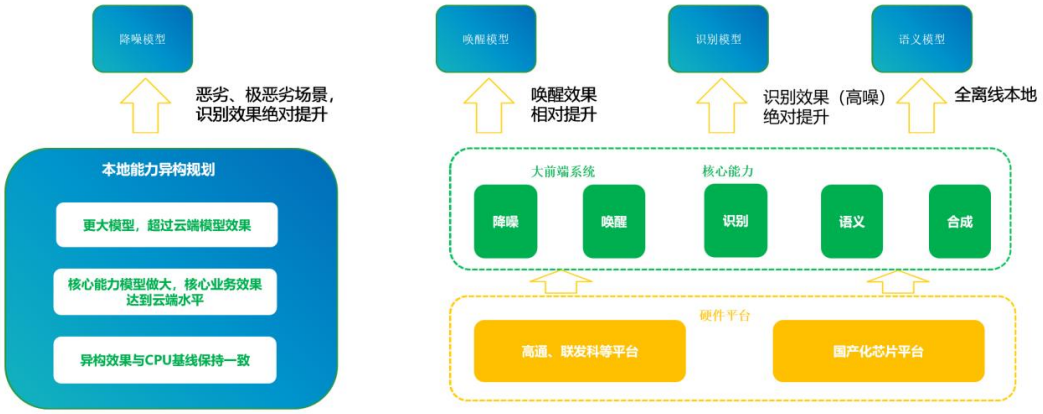

图 2-1 依托算力平台提升基础算法效果图

基于深度神经网络的语音识别技术无疑是智能座舱语音交互系统的重要基础。这种技术通过对大量语音数据的学习和训练，能够不断提升对语音的识别准确率和速度。当用户发出语音指令时，系统需要在极短的时间内准确地将语音转换为文字信息，以便后续的处理和理解。如果语音识别的准确性不高，或者识别速度过慢，将会极大地影响用户的体验，甚至可能导致用户对该功能失去信心和兴趣。

自然语言处理技术在整个语音交互过程中同样起着关键作用。它不仅要理解语音所表达的具体语义，还要深入分析上下文信息。例如，当用户说“我想去北京”，系统需要结合之前的对话历史，判断用户是想要查询去北京的路线、了解北京的天气，还是有其他的意图。此外，情感分析也是自然语言处理的一个重要方面。通过分析用户的语音语调、用词等，可以大致判断用户的情绪状态，从而使系统能够更加智能地做出回应，提供更加贴心的服务。例如，如果系统检测到用户情绪低落，它可以主动播放一些轻松愉快的音乐来缓解用户的情绪。

随着芯片算力不断提升，对语音基础能力的提升有强烈的诉求，研究人员一直在不断探索和创新。一方面，通过使用更先进的深度学习算法和模型结构，如 Transformer 架构等，提高语音识别和语义理解的准确率。另一方面，不断扩充训练数据的规模和多样性，使系统能够适应各种不同的语言场景和用户习惯。

2.2.2 边缘计算与云计算协同

边缘计算与云计算的协同工作在智能座舱语音交互系统中发挥着重要作用。边缘计算能够实时处理大量数据，减少延迟，提高交互效率。在车辆行驶过程中，

很多语音交互请求需要快速得到响应，如紧急刹车预警、变道提示等。边缘计算可以在本地快速处理这些数据，无需将数据上传到云端，从而大大减少了响应时间。

而云计算则提供了强大的计算资源和存储能力。它可以对大量的车辆数据进行分析和处理，挖掘出有价值的信息。例如，通过对大量车辆的行驶数据进行分析，可以优化交通流量预测模型，为用户提供更加准确的导航和路况信息。同时，云计算还可以用于训练更加复杂的语音识别和自然语言处理模型，不断提升系统的性能。

在实际应用中，需要合理地分配边缘计算和云计算的任务，实现两者的协同工作。例如，对于一些实时性要求较高的任务，可以由边缘计算来完成；而对于一些需要大量计算资源和数据存储的任务，则可以交给云计算来处理。同时，还需要建立高效的数据传输通道，确保边缘计算和云计算之间的数据能够快速、稳定地传输。

2.2.3 异构计算技术

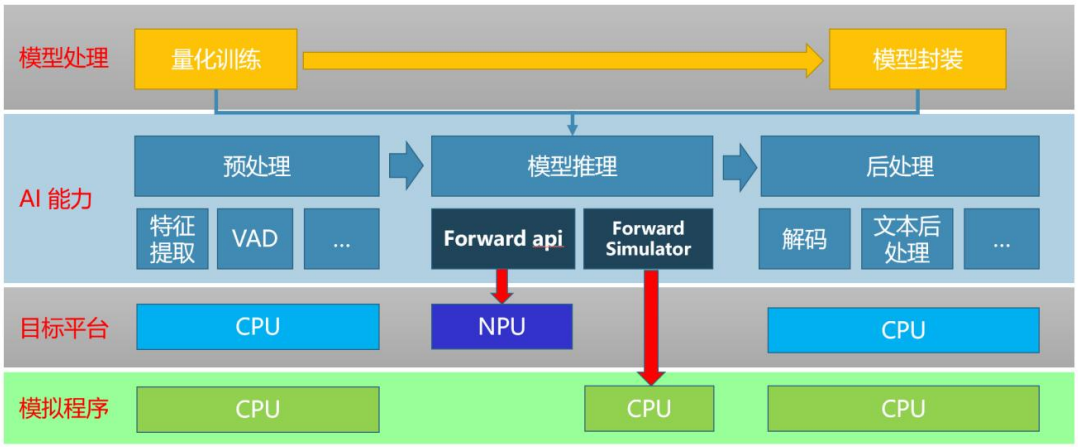

图 2-2 异构计算技术图

异构计算技术是一种创新的计算架构，它通过结合不同类型的处理器或加速器来执行计算任务，以提高效率和性能。在智能座舱人机交互语音交互技术中，这一技术的应用尤为重要，因为它可以加速复杂算法的处理速度，同时降低能耗。异构计算平台通常包括 CPU、GPU、FPGA 和专用 AI 加速器等多种硬件组件，每种组件都有其独特的优势和适用场景。例如，CPU 擅长处理顺序执行的任务，而 GPU 则在并行处理大量数据时表现出色。FPGA 可以根据特定应用进行编程，

以优化特定的计算任务。AI 加速器则专门为深度学习算法设计，能够在保持高能效比的同时提供强大的计算能力。通过合理分配任务到这些不同的硬件上，可以实现最佳的性能表现。在智能座舱系统中，异构推理计算技术的实施方法涉及多个层面。首先，需要进行硬件选择和配置，确保各种组件能够满足系统的性能要求。接着，软件层面的优化也非常关键，包括算法的并行化设计、内存管理的优化以及跨硬件的通信机制。此外，还需要开发高效的调度策略，以确保任务能够在最适合的硬件上运行。异构推理计算技术的优势在于其灵活性和可扩展性。随着智能座舱功能的不断增加，传统的单一计算平台可能无法满足所有任务的需求。异构计算平台则可以通过添加或替换硬件组件来适应新的挑战，保持系统的前瞻性和竞争力。此外，异构计算还能够根据任务的实时性要求动态调整资源分配，从而实现高性能与低功耗的平衡。

2.2.4 安全性与隐私保护

在追求智能化的同时，数据安全和隐私保护是智能座舱语音交互技术发展中不可忽视的一环。用户在使用语音交互系统时，会产生大量的个人数据，如语音指令、位置信息、偏好设置等。这些数据如果被泄露或滥用，将会对用户的隐私和安全造成严重威胁。

为了确保数据安全，需要采用一系列的加密技术。对用户数据进行加密存储和传输，防止数据被非法窃取或篡改。同时，还需要建立严格的权限管理机制，只有经过授权的人员才能访问和使用用户数据。

匿名处理也是保护用户隐私的重要手段。在处理用户数据时，可以将用户的个人身份信息进行匿名化处理，使得数据无法直接关联到具体的用户。这样既可以保证数据的可用性，又能保护用户的隐私。

此外，加强用户教育也是非常重要的。让用户了解数据安全和隐私保护的重要性，提高用户的安全意识和自我保护能力。同时，智能座舱语音交互系统的开发者和提供商也应该承担起相应的社会责任，遵守相关的法律法规，保障用户的合法权益。

2.3 挑战与展望

尽管智能座舱语音交互展现出了广阔的应用前景，但依然面临着诸多挑战，

如复杂环境下的噪声抑制、多语言与方言识别、跨文化沟通障碍、以及如何平衡智能与用户控制权等。未来，通过持续的技术创新与跨学科合作，智能座舱语音交互将不断突破现有局限，向着更加智能、人性化、安全可靠的方向发展，真正实现“人车家全生态”的未来出行愿景。

第三章 智能座舱视觉交互场景发展趋势与技术路线

3.1 发展概述

随着智能座舱和人工智能技术的发展和突破，AI 视觉交互作为一种新兴的视觉交互技术在座舱智能化的过程中扮演着至关重要的角色。该技术利用高级图像捕捉设备与深度学习算法加强了座舱对驾乘人员行为和环境的实时理解和预测能力，使座舱能更好的理解驾乘人员的状态和需求，并通过座舱内的交互终端，达到与驾乘人员互动的目的，从而提供更安全、更舒适、更直观的驾驶和乘坐体验。

AI 视觉交互技术起源于早期的车载视频监控系统，主要目的是增强驾驶安全。随着技术的演进，这些基础设施经历了从简单的图像捕获到复杂的图像处理和分析的转变，逐步融入了深度学习等人工智能算法，以更有效地处理和解释视觉数据。初期系统主要侧重于环境感知，如利用摄像头监测车辆周边的障碍物，提高反应速度和精确性。

随着深度学习和卷积神经网络（CNN）的广泛应用，视觉交互技术实现了质的飞跃，使得系统不仅能够“看到”周围环境，而且能够“理解”驾乘人员的行为和意图。通过分析驾乘人员的面部表情、身体姿势、手势动作和眼动模式等，能够准确地检测疲劳和分心状态，进而采取相应的安全措施，如发出警告或自动调整驾驶模式，从而提升车辆的主动安全性。

随着智能座舱技术的发展，智能座舱视觉交互场景的应用已从单一的驾驶员安全监控，比如疲劳监测、危险行为监测、以及注意力检测等扩展到增强整体乘车体验的多个方面。例如，现代智能座舱可以利用 AI 视觉技术理解驾驶员的情绪，自动调整车内环境（如光线、音乐和温度）以适应乘客的情绪和偏好，还有些座舱可以识别驾驶员的手势动作，以增强驾乘人员的控制距离，提升交互的效率和乐趣，更有些车企可以识别驾乘人员的唇动和视线，探索更广泛的座舱应用。

未来，随着机器学习技术的持续进步和数据处理能力的提升，座舱 AI 视觉交互的使用场景将更加广泛，DMS、OMS 等驾乘人员监控系统将成为现代化智能座舱的标配。车企为了本身的个性化配置将在 AI 视觉交互的场景应用上做更多的探索，这将进一步推动智能座舱从单一的交互模式向多模态、情境感知的交

互模式演进，为驾驶员和乘客提供前所未有的个性化体验。通过不断的技术创新，未来的智能座舱将变得更加智能，能够更好地理解人的需求和行为，成为人车互联的真正实践者。智能座舱 AI 视觉交互场景将向百花齐放的方向演进。

3.2 场景应用

随着 AI 视觉交互技术的不断进化，其应用已经渗透到整个汽车使用过程中，为用户带来全方位的智能体验。例如，用户可在上车前通过人脸识别技术轻松解锁车门并启动车辆系统，这一便捷的入车体验是现代智能技术的直接体现。在驾驶过程中，系统不仅能实时监测驾驶员的疲劳和注意力分散情况，从而显著提高驾驶安全，还能通过精准的手势识别功能，使驾驶员能够通过简单的手势控制车内的娱乐和环境系统，无需分心操作物理设备。此外，智能系统还能识别驾驶员的潜在危险行为，并适时提供反馈以规范驾驶行为，增强行车安全。最新进展如乘客识别、物品监测、唇动识别和视线追踪等功能的加入，不仅进一步丰富了智能座舱的交互层次，也大幅提升了车辆的智能安全配置，展示了 AI 视觉交互技术在现代汽车中的广泛应用和深远潜力。

3.2.1 人脸认证

人脸认证技术允许车辆通过高精度摄像头捕捉和分析来访者的面部特征。该系统与车辆数据库中预存的驾驶员面部数据进行匹配，一旦验证成功，即自动解锁车门并激活车辆内部系统。这种技术不仅提供了无钥匙进入的便利，还增加了防盗安全性，防止未授权的使用。

图 3-1 人脸认证技术图

3.2.2 疲劳与注意力监测

在驾驶过程中，疲劳监测系统通过分析驾驶员的眼睛活动、眨眼频率和头部位置来评估其警觉状态。系统使用机器视觉技术识别长时间不眨眼、头部下垂或频繁打哈欠等迹象，表明疲劳或注意力下降。一旦检测到这些疲劳信号，系统会通过声音或视觉警告提醒驾驶员，并建议休息或开启辅助驾驶模式，以确保行车安全。

图 3-2 疲劳检测&驾驶行为识别预警系统图

3.2.3 手势识别控制

手势识别系统允许驾驶员通过简单的手势来操作车辆的多媒体系统、空调和其他内部功能。利用车内的 3D 摄像头捕捉和解析手势动作，驾驶员可以不必触摸任何物理按钮即可调整音量、接听电话或更换音乐，从而减少驾驶过程中的物理分心。

图 3-3 手势识别控制图

3.2.4 危险行为监测与反馈

视觉 AI 技术同样用于监控并纠正驾驶员的危险驾驶行为，如超速行驶、急剧变道等。系统通过分析驾驶行为与道路交通规则，实时提供反馈给驾驶员，并

在必要时自动调整车速或驾驶轨迹，以维护道路安全。

图 3-4 危险驾驶行为识别与反馈图

3.2.5 乘客识别

乘客识别系统使用高级的视觉 AI 技术来识别车内所有乘客的身份。该系统通过分析乘客的面部特征与预存的数据进行匹配，以确认每位乘客的身份并提供个性化服务。例如，系统可以根据乘客的偏好自动调整座位位置、空调设置和娱乐系统。对于家庭用车，乘客识别还能确保儿童安全座椅正确安装，并启用适合儿童的安全功能，如儿童锁和定制化娱乐内容。

图 3-5 乘客识别系统图

3.2.6 遗留物品监测

物品监测系统专注于识别并追踪车内物品的位置和状态。利用内置的摄像头，该系统能够检测到重要物品如钥匙、手机、手提包等是否遗留在车内，减少物品丢失的风险。此外，对于有小孩或宠物的家庭，物品监测系统也可以确保车辆停车后，没有儿童或宠物被无意留在车内，提高车内安全管理。

图 3-6 遗留物品检测图

3.2.7 唇动识别

唇动识别技术通过捕捉和分析驾驶员的口部动作来辅助语音命令系统，特别在车内嘈杂环境中提升命令识别的准确率。这项技术能够有效区分驾驶员的命令与车内其他噪音，确保系统准确响应驾驶员的需求，无论是进行导航调整、媒体播放控制还是接听电话。

图 3-7 唇动识别技术图

3.2.8 视线追踪

视线追踪技术通过实时分析驾驶员的视线方向和焦点，用来评估驾驶员的注意力分散程度。该系统能够识别驾驶员是否长时间未注视驾驶道路，并通过视觉或听觉警告提醒驾驶员注意驾驶安全。这项技术尤其对于长途驾驶或在复杂交通环境中驾驶具有重要的安全意义，帮助预防因分心导致的交通事故。

图 3-8 实现追踪技术图

3.2.9 AR-HUD

AI 增强的 AR-HUD（增强现实显示）技术是将先进的人工智能与 AR 技术结合，以提供动态的、互动式的驾驶辅助体验。这种系统通过分析从车辆前方摄像头、雷达和 LiDAR 等传感器收集的数据，AI 算法可以实时识别和分类道路标志、障碍物和行人，并将这些信息以增强现实的形式直观展示在驾驶员视野中。例如，它能高亮显示突然减速的前车并提供避碰撞提示，或者根据实时交通状况预测并显示最佳行驶路径。此外，系统还能分析驾驶员行为，提供个性化驾驶建议，并根据驾驶环境调整显示设置以确保信息始终清晰可见。这种集成了 AI 的 AR-HUD 技术不仅提升了驾驶的安全性，还极大地增强了驾驶的便利性和舒适性。

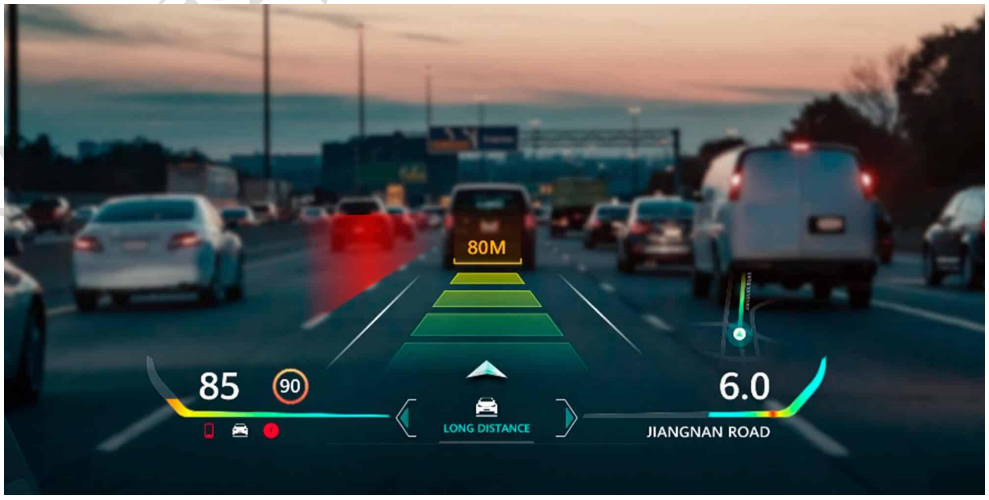

图 3-9 AR-HUD（增强现实显示）技术图

3.3 技术路线

智能座舱视觉技术是一种以机器视角的方式来观察交互者的行为，从而分析判断出用户意图，再通过对应执行机构来满足交互者的使用需求。通常包括感知、识别、和执行三个步骤。感知阶段系统启动后机器视觉传感器实时采集仓内画面将图像输入给检测单元，检测单元通过训练好的算法模型对图像进行逐帧识别，识别到用户意图后，通过通信技术将执行结果传输给执行器执行，激活相应的车辆功能，比如调整座椅位置、播放音乐或调整车内温度。此外，它还可以将驾驶员视线之外的区域，比如车辆的盲区，通过车载显示屏或其他界面呈现给驾驶员，从而增强驾驶安全和便利性。简而言之，智能座舱视觉技术通过感知和分析座舱内外部的视觉信息，实现对车辆内部环境的智能控制和辅助驾驶。

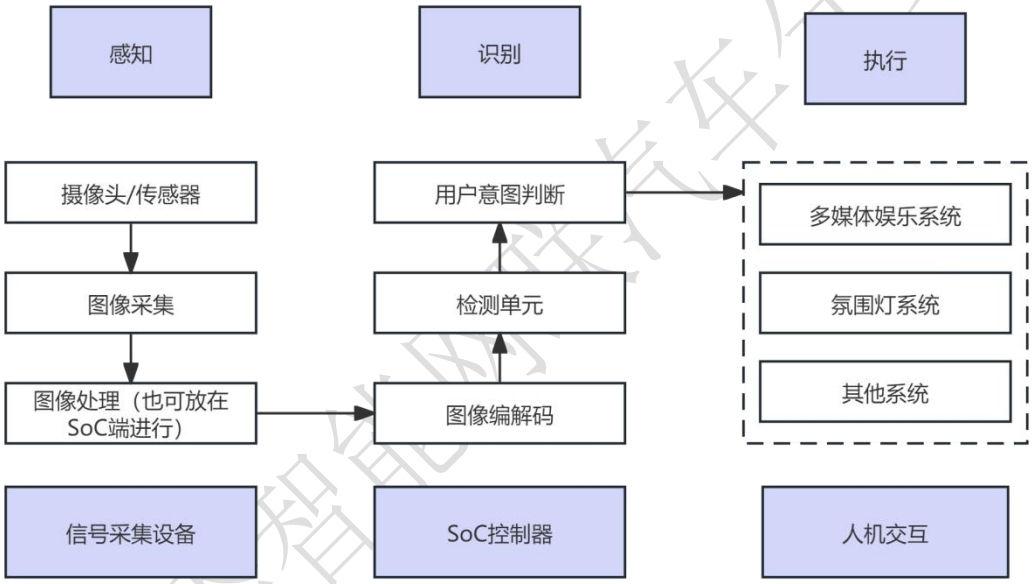

图 3-10 智能座舱视觉技术图

3.3.1 DMS 技术路线

DMS（驾驶员监控系统）是智能座舱内视觉交互技术的一大领域，早期 DMS 的方案有通过监测车辆信息间接判断驾驶员疲劳状态的方法，也有通过生物信号采集装置来监测驾驶员疲劳状态的方法，但二者都准确性较低，而且容易干扰驾驶员驾驶。随着机器视觉技术的进步，通过摄像头和近红外技术，从人眼、嘴部的开合状态来判断驾驶员疲劳状态成为主流方案。

通过大量样本训练好的算法模型对于输入的图像进行人脸检测和关键点检测。将嘴部和眼部关键位置抠图出来，根据基于眨眼行为的 EAR 和 PERCLOS

特征,和基于哈欠行为的MAR和FOM特征,判断是否出现眼睛闭合和打哈欠的状态，当闭眼和打哈欠状态达到阈值设定即可判断为疲劳状态。目前 DMS 功能远不局限于疲劳状态监测，还可以通过深度学习的模型集成吸烟、打手机检测功能，检测手部区域和手部物品的检测，如果驾驶员有手持手机打电话和吸烟的动作并且达到设定阈值，即可识别出吸烟、打手机等不良行为。系统识别到驾驶员有疲劳状态或不良行为后可以通过通信技术，将信号传输给声光报警器或仪表灯执行机构发出提醒，提醒驾驶员谨慎驾驶。

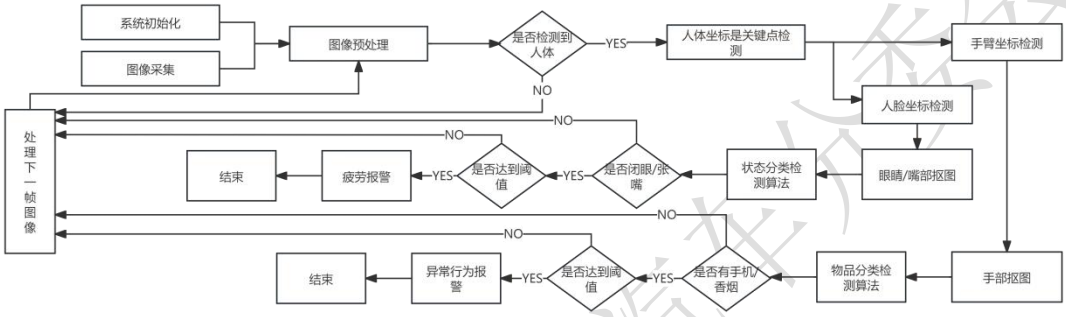

图 3-11 DMS 技术路线图

3.3.2 人脸识别技术路线

人脸识别技术通常分为 2D 人脸识别和 3D 人脸识别，2D 人脸识别是通过 RGB 相机采集一张 2D 的人脸图片，进行特征向量的提取和人脸库做比对，虽然 2D 人脸识别在很多领域都取得了显著的成功，但同时也存在弊端，如光照条件的依赖，使用照片等的数据欺骗。所以 3D 人脸识别将成为主流的发展趋势，常见的 3D 成像方案有双目立体视觉、结构光和 TOF 相机方案，相比于另外的 2 个方案，TOF 相机更具性价比和易用性，在智能座舱领域应用更加广泛。

基于 Time of Flight（TOF）相机的人脸识别技术是一种非接触式生物识别技术，它利用 TOF 摄像头来获取深度信息，以实现高精度的人脸三维建模和识别，主要包括深度数据采集、图像预处理、人脸检测与对齐、三维人脸建模、特征提取、算法识别、比对与验证等步骤。TOF 相机通过发射红外脉冲并测量其反射回来的时间，计算出物体的距离生成深度图，对采集到的图像还需要进行滤波、校准和降噪以便于提取更清晰的人脸特征。使用 Haar 特征或深度学习模型在深度图中定位到人脸区域，并对其进行归一化，使得不同姿态和表情的脸都能在同一空间下对齐。根据深度信息进行 3D 人脸建模，从 3D 模型中提取面部集合结

构、纹理或深度特征，然后进行算法识别，使用 3D CNN 或深度人脸识别网络来比较提取到的面部特征与已知人脸库模版的匹配情况，通过设定相似度阈值来判定是否匹配通过。

智能座舱领域的人脸识别技术可用于车辆解锁，当用户有用车需求的时候主动走进车辆靠近 B 柱传感器，人脸识别系统就可以进行识别是哪位驾驶员要使用车辆，并按习惯的设定自动调节好座椅、后视镜位置，为用户提供方便的用车体验。

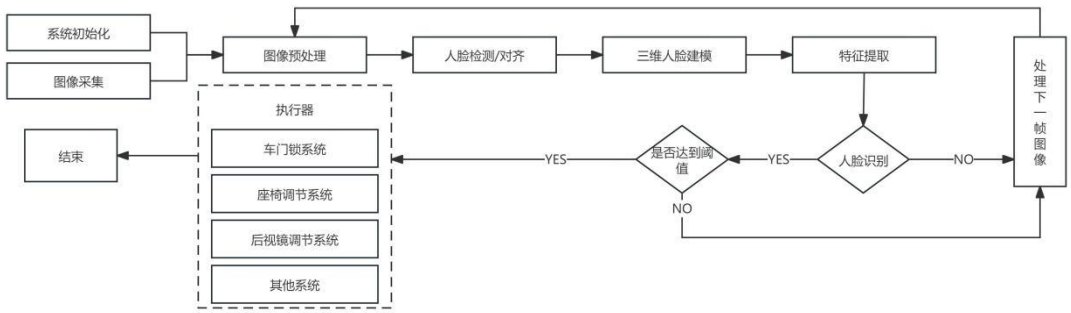

图 3-12 人脸识别技术路线图

3.3.3 手势识别技术路线

手势是人类与生俱来的肢体语言，传统的座舱交互方式多采用实体或虚拟的按键来实现，这往往需要驾驶员主动花费时间去寻找按键位置，增加了不安全隐患，手势识别的交互方式可以通过动态或静态简单的肢体语言来实现智能座舱的控制，可以为智能座舱人机交互增加多样性，娱乐性和沉浸感，而且交互方式更加安全、直观、快捷。

智能座舱手势识别的技术是结合了计算机视觉和人工智能技术的创新解决方案，首先通过图像采集设备，如高清摄像头或深度传感器来感知捕捉驾驶员的手部动作，将每一帧图像进行去噪、增强和色彩校正等预处理，使用机器视觉算法如 Haar 特征或深度学习模型（如 CNN）对手部进行检测，定位关键点并识别出特定手势。再将识别到的手势映射到预先设定好的定义动作类别中，系统通过手势类别来明确驾驶员的需求，最后将用户意图信号传递给车辆的其他系统来执行操作。例如，用户想切换娱乐设备当前的歌曲只需要对着视觉传感器做出左滑或右滑的手势，系统就可以在用户完成特定的手势动作后，识别到用户需求。

机器视觉技术赋予了机器以视觉能力，就像给机器安装了眼睛，使其能够进

行精确的视觉识别和分析，赋予了智能座舱多种多样的人机交互方式。这项技术不仅能够模仿人眼的功能，而且在某些情况下，其性能甚至超越了人类的视觉。机器视觉系统能够以高清晰度和快速度进行检测和控制，同时由于它与被检测的对象没有直接接触，因此操作起来既安全又可靠。

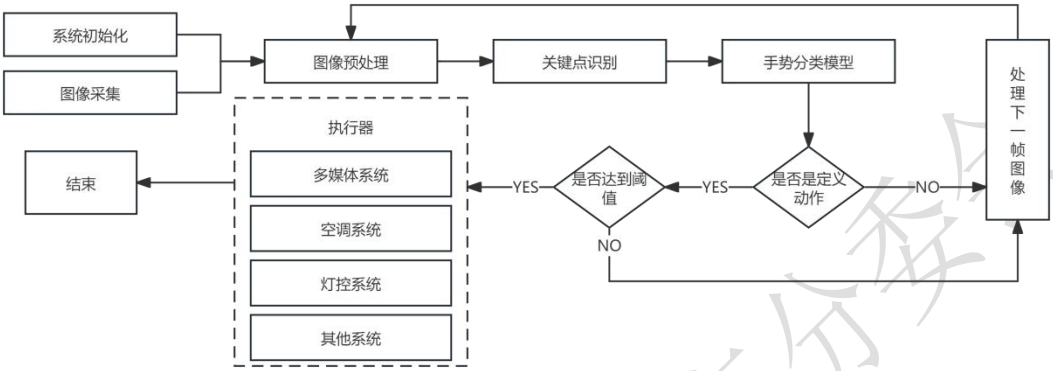

图 3-13 手势识别技术路线图

3.3.4 AR-HUD 技术路线

智能座舱视觉交互技术不光指机器的视觉，还应包含人的视觉。人的视觉主要的交互终端是显示设备，对于汽车的显示设备通常为仪表和 HUD。仪表通常位于中控台，驾驶者观察需要将视线从道路上短暂的移开这样可能会引入一些安全风险，所以这种模式很多车企并不推崇。故引入了 HUD 技术，从最早的 C-HUD，到现在主流的 W-HUD，未来 AR-HUD 将逐步取代 W-HUD 成为趋势。

AR-HUD 技术是将增强现实信息投影到驾驶员前方视线范围内的系统。按照投影技术分类，主流的技术包括 LCD 投影、DLP 投影、激光扫描投影等。LCD 是当下成熟的主流方案，在 AR-HUD 的趋势下 DLP 投影的优势也越来越明显，激光投影目前仍处于探索阶段。

为了更好的实现智能座舱视觉技术，满足数字投影与物驾驶环境的融合，AR-HUD 多采用两个焦面技术来显示不同距离的场景内容，一个焦面显示近景，替代仪表显示车速、里程、车辆状态等信息，另一个焦面显示远景信息，如虚拟导航路径、车外真实环境建模等，通过双焦面技术可以提供更丰富、更大面积的现实增强信息，远近景结合提供更好的融合效果。

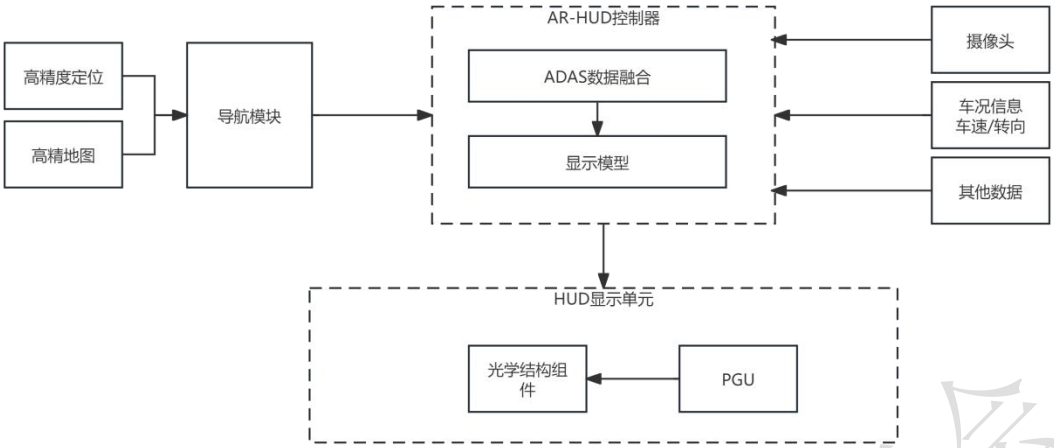

图 3-14 AR-HUD 技术路线图

3.4 未来展望

随着人工智能技术的飞速发展，AI 视觉交互场景在智能座舱中的作用日益增强，将深度整合到我们的日常驾驶体验中，以满足用户对安全、便捷及个性化的持续需求。

从用户需求的角度出发，现代消费者在基本的驾驶需求得到满足之后，转而寻求更高层次的情感需求，更期待通过高度个性化的智能系统获得极致的舒适和便利。未来的汽车将在感知和理解乘客的能力大幅提升，更关注乘客的情感需求，以便提高驾驶体验。

技术层面，域控制器的概念正在逐步取代传统的电子控制单元（ECU）集中式架构，提供了更高效的数据处理和更快的响应速度。这种架构变革使得车辆能够更好地协调各种传感器和执行机构，实现高级的功能，如实时交通调整和环境感知，极大地优化了整车性能和用户体验。同时，随着智能化需求的增加，车载处理器的性能也在显著提升。从高通的 Snapdragon 8155 到更先进的 8295 芯片，车载微处理器和 GPU 的性能飞跃使得车辆能够支持更多实时数据处理和运行更复杂的算法。此外，AI 技术正经历技术变革，大模型的引入进一步增强了座舱的视觉理解和预测能力，最后， 摄像头技术的进步是 AI 视觉系统发展的关键。未来的摄像头将支持更高的分辨率和更广的视角，能够捕捉更细致的图像，从 2D 转向更为精确的 3D ToF（Time of Flight）技术，提供更深层次的空间和运动信息，使得系统更加精确地识别和响应车辆周围的动态环境。

随着用户需求的日益膨胀与技术的进步，未来智能座舱 AI 视觉交互场景应

用将极为精准和广泛，不仅将极大提升驾驶安全性和舒适性，也将将带来全新的用户交互体验。

3.4.1 更深层次的行为和情绪理解

未来的视觉 AI 系统将能够更精准地识别和解释驾驶员及乘客的情绪和行为状态。利用高级的面部表情识别和生理信号处理技术，车辆将能够实时调整内部环境，如气氛光线、温度、甚至香味，以适应乘客的情绪变化，提供更为个性化的舒适体验。

3.4.2 无缝多模态交互

未来的智能座舱将整合触摸、声音、视觉甚至思维控制等多种交互方式，创造无缝的多模态交互体验。例如，驾驶员可以通过简单的手势、语音命令或视线移动来控制车辆功能，而更先进的技术如脑机接口（BCI）可能允许直接通过思维控制车辆系统。

3.4.3 预测性个性化服务

基于大数据和机器学习，未来的 AI 视觉系统将能够预测个人需求并自动提供服务。例如，系统可以根据过去的行为模式和偏好自动规划行车路线，选择音乐播放列表或推荐附近的餐厅。

3.4.4 增强现实（AR）的集成应用

AR 技术将更广泛地应用于智能座舱中，提供更丰富的信息和娱乐选项。驾驶者通过风挡显示的 AR 界面可以看到增强的导航信息、道路安全提示甚至虚拟景观，同时也可以显示关于车辆性能的实时数据。

随着技术的进步，未来的智能座舱将越来越像一个高度集成的移动智能平台，不仅提供交通工具的基本功能，还将成为一个提供娱乐、办公和休息的多功能空间。这将彻底改变我们对汽车的传统认识和使用方式，开启全新的智能驾驶时代。

第四章 智能座舱多模交互场景发展趋势与技术路线

4.1 发展概述

随着汽车行业的智能化进程加速，智能座舱中新技术、新场景和新模式层出不穷。这一发展趋势导致车载交互设备的数量迅速增加，已接近其功能的上限。因此，设计师面临的挑战是如何在确保驾驶安全的同时，提供高效和舒适的人机交互体验。在这种背景下，多模态交互预计将成为未来人机交互的主流方式。

“模态”（modality）这一术语最早由德国生理学家赫尔姆霍兹提出，用于描述生物体如何通过各种感知器官和经验来接收外界信息。在人机交互的语境中，“多模态交互”指的是系统通过综合人类的视觉、听觉、触觉等多种感官输入，利用多种通信通道进行响应，并尽可能地模拟人与人之间自然的交互方式。

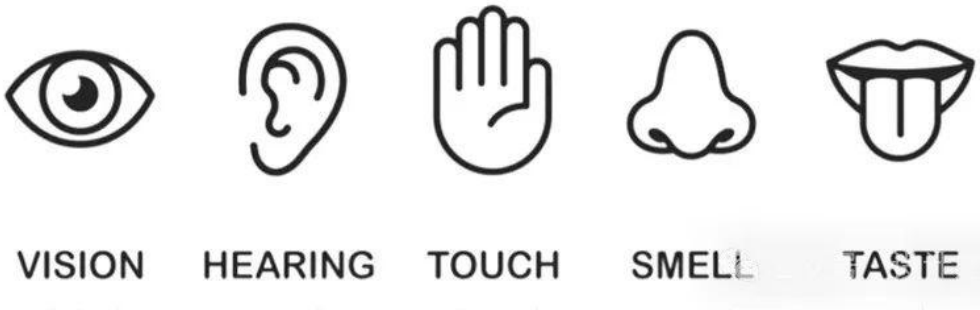

图 4-1 生物体多模态交互图

在实际应用中，汽车制造商已经实施了多种多模态交互方案，涵盖了语音结合唇动识别、语音与面部识别、语音与手势识别、语音与头部姿势、面部表情与情绪识别、面部表情与眼球追踪、以及香氛结合面部表情和语音识别等组合。目前，语音结合其他模式的多模态交互形式已成为主流，如长安启源 A07、极越 01、理想 L7、合创 A06/V09 等车型中均有应用。

这些多模态交互系统的引入不仅极大地丰富了用户的交互体验，也提高了系统的可用性和交互的直观性。随着技术的不断进化和消费者需求的不断提高，我们可以预见未来智能座舱将通过更精细和更智能的多模态交互方式，为驾驶者和乘客创造一个更安全、更舒适、更个性化的驾驶环境。

4.2 场景应用

目前车企已实现的多模态融合包括但不限于语音+唇动识别、语音+面部识别、语音+手势识别、语音+头姿、面部+情绪识别、面部+眼球追踪、香氛+面部+语音识别等。其中语音多模态交互方式为当下主流，应用车型包括上文提到的长安启源 A07、极越 01、理想 L7、合创 A06/V09 等车型。

4.2.1 听觉融合

听觉融合在智能座舱中通过结合语音识别与其他感官输入来增强交互体验和响应准确性。具体实例包括：

表 4-1 听觉融合相关实例

交互模态	功能介绍
语音+手势	驾驶员可以同时发出语音指令和手势，比如通过伸出食指并指向左右或上方，来控制车窗、天窗或遮阳帘
语音+人脸	通过方向盘上方的摄像头识别语音指令的发出者，自动过滤主驾位的非控制语音，如与其他乘员的对话
语音+眼球追踪	通过视线来唤醒语音识别功能，使驾驶员无需物理接触或明确命令即可激活系统
语音+唇动	采用红外光感摄像头智能识别主驾唇形，结合唇语读取和语音特征的双重识别技术，极大提升系统的唤醒和识别准确率
语音+行为追踪	允许驾驶员在车外通过语音指令“跟随我”，使卡车能在安全距离内自动跟随，同时使用传感器和摄像头技术避开障碍物

图 4-2 智能座舱听觉融合技术图

4.2.2 视觉融合

在智能座舱的视觉融合技术中，通过结合人脸识别、眼球追踪以及其他生物识别技术，车辆能够提供更为精准和个性化的驾驶辅助及车辆控制功能。以下是几个具体的应用示例：

表 4-2 视觉融合相关实例

交互模态	功能介绍
人脸识别+眼球追踪	DMS 驾驶员监控系统集成了人脸识别和眼球追踪技术，支持疲劳监测和眼球追踪，实时调整 AR-HUD 投影角度
人脸识别+心率监测	通过搭载 52 个生物传感器，可以识别驾驶员的表情、声音、心率、血氧、血压以及呼吸频率等指标。从而自动调节车内的音乐和温度，以及危险时接管车辆
人脸识别+静脉识别	“Leap In 生物钥匙系统”，结合了人脸识别和静脉识别技术。可实现车门的快速解锁和车辆启动，自动加载个性化的驾驶设置，如座椅调整、后视镜位置及多媒体系统偏好等

图 4-3 智能座舱视觉融合技术图

4.2.3 嗅觉融合

嗅觉融合在智能座舱的多模态交互中代表了一种创新方向，通过结合香氛系统与语音命令、人脸识别等技术，极大地增强了驾驶体验的丰富性和个性化程度。以下是几个实际应用场景，展示了这种融合如何在提升驾驶安全和舒适性方面发

挥作用：

表 4-3 嗅觉融合相关实例

交互模态	功能介绍
香氛+语音	通过语音命令唤醒座舱的场景模式，如 K 歌模式下，香氛系统与音效和氛围灯联动。在醒神模式下，系统自动开启空调、香氛、氛围灯、座椅震动以及动感音乐等多种功能，帮助驱散驾驶疲劳
香氛+人脸	系统能够监测到驾驶员的疲劳状态，并自动启动车内香氛系统释放提神醒脑的香氛，如薄荷或柠檬香味，以确保驾驶安全。
香氛+语音+人脸	能够通过语音唤醒特定的场景模式，如“醒神潮汐”模式下系统会与 DMS（驾驶员监控系统）进行联动。A 柱摄像头监测到驾驶员疲劳时，系统进行语音提醒及释放醒神香氛。

智能座舱的多模态交互系统不仅提高了交互的直观性和自然性，还增强了系统的整体效率和响应能力，进一步提升了驾驶安全和乘坐体验。随着技术的持续进步，未来智能座舱的交互方式将变得更加多样化和人性化，更好地满足用户的各种需求。

4.3 技术路线

智能座舱中的多模态融合技术，通过整合听觉、视觉和嗅觉等多种感官输入，为驾驶员提供更加智能化和个性化的交互体验。实现这一技术的核心在于高精度传感器的集成、高效的数据处理与融合，以及实时响应与优化。首先，需要在座舱内安装高分辨率摄像头、麦克风阵列、眼球追踪器和生物传感器等多种传感器，确保能够全面捕捉驾驶员的各种感官数据。接下来，通过先进的预处理技术，如降噪、滤波和特征提取，确保输入数据的高质量。利用深度学习和人工智能算法，将来自不同传感器的数据进行综合处理和分析，实现多模态数据的有效融合。这种数据融合技术能够生成统一的控制指令，从而实现多感官输入的协同工作。实时处理能力和系统响应速度是确保用户体验流畅性的关键，需要通过优化算法和硬件配置，确保系统能够快速、准确地响应驾驶员的指令。最后，系统集成与优

化至关重要。多模态交互系统需要无缝嵌入车辆的中央控制系统，确保各模块之间的协同工作。通过持续的系统性能测试和用户反馈，不断优化系统功能和响应速度，提升整体用户体验和驾驶安全性。未来，随着情感识别、环境感知和行为预测等技术的进一步发展，多模态融合技术将在个性化和智能化方向上不断进步，为用户带来更加丰富和安全的驾驶体验。

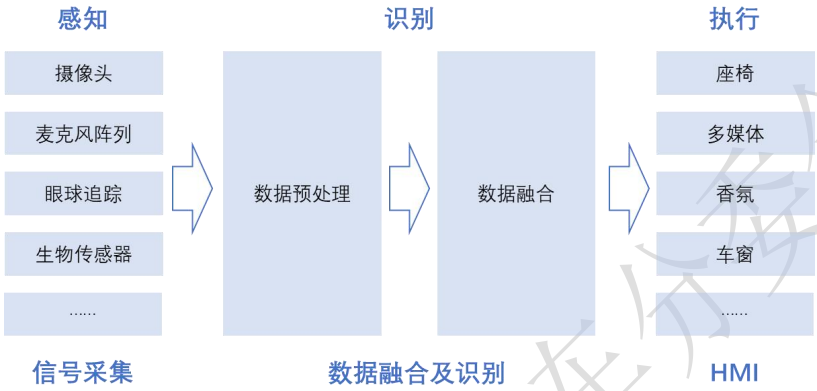

图 4-4 智能座舱多模态融合技术路线图

4.3.1 听觉融合技术路线

听觉融合技术在智能座舱中的应用，通过将语音识别与其他感官输入相结合，实现更加自然和高效的交互体验。实现这一技术的核心在于多模态传感器的集成与高级数据处理。首先，高精度麦克风阵列和手势识别摄像头的部署至关重要，这些传感器能够捕捉用户的语音和手势数据，并通过先进的预处理技术，如降噪和滤波，确保数据的高质量。随后，利用深度学习算法对这些数据进行解析，通过数据融合技术，将语音和手势数据综合处理，以生成精确的控制指令。例如，驾驶员可以通过语音指令并结合手势操作，如通过伸出食指并指向左右或上方，控制车窗、天窗或遮阳帘。实时处理和系统优化是确保用户体验流畅性的核心，通过持续优化算法和硬件配置，提升系统的响应速度和准确性。未来，听觉融合技术将进一步向多感官协同与智能化方向发展，结合环境感知和情感识别技术，实现更加个性化和自然的交互体验。

4.3.2 视觉融合技术路线

视觉融合技术在智能座舱中，通过结合人脸识别、眼球追踪及其他生物特征识别技术，为驾驶员提供更加精准和个性化的驾驶辅助及车辆控制功能。实现这

一技术的关键在于多传感器系统的部署和数据融合技术的应用。高分辨率摄像头和生物传感器如眼球追踪设备和心率监测器，被布置在驾驶舱内的关键位置，以全面覆盖驾驶员的面部特征和生理状态。数据融合技术结合深度学习算法，通过大量驾驶员数据的训练和优化，提升识别模型的准确性和响应速度。例如，通过集成人脸识别和眼球追踪技术，实时监测驾驶员的注意力状态，并动态调整AR-HUD（增强现实头显）的信息显示位置，确保驾驶信息始终处于最佳视野范围内。系统集成与性能优化确保视觉融合系统与车辆中央控制系统的无缝对接，持续的性能测试和用户反馈循环，不断改进系统功能和用户体验。未来，视觉融合技术将进一步融合情感识别技术，通过监测驾驶员的情绪状态，提供更加人性化的驾驶辅助和提醒服务。

4.3.3 嗅觉融合技术路线

嗅觉融合技术在智能座舱中的应用，通过结合香氛系统与语音命令、人脸识别等技术，提升驾驶体验的丰富性和个性化。这一技术的实现路线包括多感官系统的集成和多模态数据处理。香氛释放装置与高精度麦克风阵列和人脸识别摄像头等传感器的集成，确保香氛系统能够根据用户需求和车内环境，精确控制香氛的类型和浓度。例如，通过驾驶员监测系统监测驾驶员的疲劳状态，自动启动车内香氛系统，释放提神香氛，如薄荷或柠檬香味，以确保驾驶安全。多模态数据处理通过数据融合技术，综合分析驾驶员的语音指令、人脸表情和驾驶状态，生成相应的控制指令。实时系统响应和优化是确保交互流畅性的关键，高效的香氛控制算法结合语音识别和人脸识别结果，实时调整香氛释放。用户体验方面，通过用户反馈持续优化香氛系统的交互体验，提供多种香氛选择和个性化设置，让用户能够根据自身偏好，调整系统参数。未来，嗅觉融合技术将进一步结合更多的生物识别和环境感知技术，实现更加智能和个性化的交互体验。随着 AI 和传感技术的发展，香氛系统的响应速度和精确度将进一步提升，为用户提供更为丰富和舒适的驾驶体验。

4.4 未来展望

根据多重资源理论与态势感知理论，多模态交互能在一定程度上缓解驾乘人员的信息过载，提高驾乘人员对环境整体、动态的理解能力，从而提升其对安全

风险的识别，理解和响应能力。另一方面，多模态交互融合了视、听、触、嗅多感官输入输出，极大地提高了驾乘人员的沉浸式体验，使得操作更为直观和响应更为灵敏，因此，多模态交互在智能座舱中的地位越来越重要，是座舱智能化的未来演进路线。

4.4.1 数据融合和上下文感知

未来的多模态交互系统将更加依赖于高级的数据融合技术，这些技术能够整合来自视觉、听觉、触觉和嗅觉等多种传感器的数据。通过上下文感知能力，系统将能够根据当前环境和用户状态调整其响应。例如，系统可以根据驾驶情况和用户情绪动态调整信息的呈现方式和交互的优先级，如在高压环境下优先语音提示而在静态驾驶时通过视觉显示复杂信息。

4.4.2 交互优化算法

随着机器学习和人工智能的进步，未来的多模态系统将使用更复杂的算法来优化交互流程。这些算法能够实时分析用户的反馈，自动调整交互模式以匹配用户的喜好和需求。例如，系统可能会学习到用户在某些特定情境下偏好使用触控而非语音命令，并自动适应这一偏好。

4.4.3 高级用户界面（UI）设计

随着技术的进步，用户界面设计将更加重视多模态交互的集成和无缝体验。未来的 UI 将能够同时支持多种输入和输出模式，提供一致且自然的用户体验。例如，增强现实（AR）技术将与传统显示、触控屏和语音指令更紧密地结合，提供一种更直观且信息丰富的交互方式。

4.4.4 实时性能优化

随着处理能力的增强和算法的优化，多模态系统将能够在实时基础上进行性能调整。这意味着系统将能够即时识别和解决可能的交互冲突，优化资源分配，确保所有模态都能高效协同工作，从而最大化响应速度和准确性。

第五章 智能座舱大模型应用场景与技术路线探索

大模型跨场景通用的基础能力主要包括生成能力、上下文连贯能力、学习能力、理解能力和推理能力。这些能力共同构成了大模型的核心，使其能够在多种场景下提供高效的服务和互动。

生成能力指大模型进行图像生成、自然语言生成（文本生成）以及音乐生成，展现出其创造力和多样性。上下文连贯能力则体现在大模型能够使用短期记忆进行连续多轮对话，保持话题的连贯性和对话的流畅性。

学习能力是大模型对用户行为、偏好进行储存，并根据需求进行提取和更新的能力，这使得模型能够更好地适应用户的个性化需求。理解能力让大模型能够深入文本、视觉、语音信息，实现精准的语义理解和情感分析，从而更准确地把握用户意图和情感状态。

推理能力赋予大模型逻辑思考的能力，无论是在生成内容还是执行任务时，都能够进行合理的推断和决策。除了以上这些基础能力，拟人化和情感化的能力，则进一步丰富了大模型的交互体验。

拟人化能力使得大模型在与用户的互动中，能够模仿人类的行为和反应模式，提供更加自然和亲切的体验。它使用人性化的语言风格，根据用户的情绪状态调整自己的语气和用词，表达同情、鼓励或幽默感，从而增强用户的参与感和满意度。

情感化能力让大模型在处理信息和生成回应时，能够识别和表达情感，不仅识别用户的喜怒哀乐，调整回应的情感色彩，还在生成内容时融入情感元素，使得输出更加生动和有感染力。这种能力还体现在大模型能够根据上下文和用户的历史交互，动态调整情感表达，以适应不同的交流场景，建立起更加深入和个性化的联系。

表 5-1 智能座舱大模型应用场景

	语音交互	视觉交互	多模态交互	开放式任务
行车辅助	基于语音信息辅助导航及驾驶决策 ● 根据用户表达的需求推荐出行路径 ● 语音控制导航功能（途径点）	基于视觉捕捉到的信息辅助导航及驾驶决策 ● 捕捉车外环境信息 ● 道路标志（识别道路管制/ 施工标志更新导航路线） ● 捕捉车内环境信息 ● DMS-OMS-IMS	基于多种模态信息（语音+手势/ 语音+眼动 or 环境信息/ 语音+车辆状态信号）辅助导航及驾驶决策 ● 语音+手势：手势指代 ● 语音+眼动：视线指代 ● 语音+车辆状态信号	主动提醒行程 ● 油量/ 电量不足以支撑完整行程，行程需求超过当前电量余量
车辆设置	基于语音信息控制车辆设置 ● 车辆设置简单指令 ● 一句执行多项设置（单域/ 跨域） ● 自定义复杂任务	基于视觉捕捉到的信息控制车辆设置 ● 捕捉车外环境信息 ● 天气（暴晒天气关闭遮阳帘） ● 捕捉车内环境信息 ● 乘客行为（光线暗看书，打开阅读灯）	基于多种模态信息（语音+手势/ 语音+眼动 or 环境信息/ 语音+车辆状态信号）控制车辆设置 ● 语音+手势：手势指代 ● 语音+眼动：视线指代 ● 语音+车辆状态信号	自动开启/ 关闭/ 调节车辆设置 ● 空气净化 ● 空调温度 ● 屏幕亮度 ● 日间/ 夜间模式 ● 悬架高度 ● 音区
服务支持	基于语音信息提供服务 ● 用车指导 ● 服务咨询 ● 故障原因分 ● 事故处理	基于视觉捕捉到的信息提供服务 ● 捕捉车外环境信息 ● 哨兵模式 ● 捕捉车内环境信息 ● 内饰维护建议（座椅磨损）	基于多种模态信息（语音+手势/ 语音+眼动 or 环境信息/ 语音+车辆状态信号）提供服务 ● 语音+手势：手势指代 ● 语音+眼动：视线指代 ● 语音+车辆状态信号	推荐或提醒保养 /

多媒体	基于语音信息提供媒体播控 推荐音频/ 视频（热门、最新...) ● 查询新闻 ● 语音游戏	基于视觉捕捉到的信息提供媒体播控 ● 捕捉车外环境信息 ● 情景捕捉（看到烟花，推荐播放欢快的歌曲） ● 捕捉车内环境信息 ● 表情识别（播放视频安抚无聊的儿童）	基于多种模态信息（语音+手势/ 语音+眼动 or 环境信息/ 语音+车辆状态信号）提供媒体播控 ● 语音+手势：手势指代 ● 语音+眼动：视线指代 ● 语音+车辆状态信号	/
问答闲聊	基于视觉捕捉到的信息回答问题 ● 生活问答 ● 查询 ● 推荐 ● 闲聊 ● 观点讨论 ● 心情抒发 ● 日常分享	基于视觉捕捉到的信息回答问题 ● 捕捉车外环境信息（包含建筑物、道路使用者、动植物、行车记录仪记录） ● 捕捉车内环境信息（包括人及物件）	基于多种模态信息（语音+手势/ 语音+眼动 or 环境信息/ 语音+车辆状态信号）回答问题 ● 语音+手势：手势指代 ● 语音+眼动：视线指代 ● 语音+车辆状态信号	/

综合这些能力，大模型不仅仅是一个信息处理和任务执行的工具，更是一个能够理解、学习和适应用户需求的智能伙伴，能够在各种场景下提供富有同理心和情感智慧的服务。以下表格中梳理了当前的大模型的典型应用场景。

5.1 大模型应用于语音交互场景

智能网联汽车中的大模型语音交互是指通过将大模型应用于智能网联汽车中的语音交互场景。相比于传统车内语音交互，大模型可以提供更加自然的拟人化交互，以及具有更加智能的理解能力，能够从语音中识别出情感，并作出对应响应。大模型应用于语音交互场景可以提供更加智能化，拟人化，情感化的交互以及更加丰富的场景。下面以语音识别，车内指令控制，情感化拟人化助手，语音合成等四个方面介绍大模型应用于语音交互场景下的技术路线与应用价值。

5.1.1 语音识别

语音识别是将语音转化成对应文本的技术，传统的语音识别使用 DNN（Deep

Neural Network），CNN（Convolutional Neural Network），RNN（Recurrent Neural Network）等方法，传统方法对语音序列进行建模，提取语音信号中的声学特征，再将特征映射到对应的单词以进行语音识别。传统方法对于语音中噪声的过滤以及一些地方方言，多语言识别效果有限。

大模型可以更好的对语音序列中的时间序列信息进行特征提取，从而实现更高纯度的噪声过滤。同时根据大模型的理解能力，可以实现免唤醒，多人语音识别，语音端点检测等。相比于传统方法，大模型在语音识别中的应用优势如下：

1) 多语言，方言识别

由于大模型具有更加复杂的架构以及更多的参数，能够捕捉到更加细微的语音信号特征，同时由于参数量的提升，大模型在语音识别泛化性上也具有明显的优势。通过在训练数据中加入多种不同语言，方言等数据，大模型可以在推理时对一句话中的多语言以及地方方言进行准确识别。

2) 对噪声，语速变化的鲁棒性

大模型具有更强的特征提取能力，对于存在噪音以及语速突然变化的语音识别条件下，大模型通过对语音进行特征提取，依然可以保证语音识别结果的鲁棒性。

3) 多任务学习

大模型可以同时学习多个相关任务，例如在语音识别的同时进行说话人识别，情感分析以及语音唤醒等任务。传统方法则需要对每个不同任务单独训练一个模型来完成。

4) 持续学习

大模型可以通过后续的增量预训练或者微调技术来进行迭代，不断适应新的环境数据变化，例如网络上新出现的流行语识别等，以提高长期性能。

5.1.2 车内指令控制

车内指令控制是指通过用户输入语音转化后的文本对车内一些模块进行控制，传统的指令控制方法通常采用预定义好的语音指令来实现车内模块的控制，存在着识别指令有限且必须按照指令预定义好的规则才能触发的问题，采用大模型进行车内指令控制存在以下几种优势：

1) 多维度分析与全面预测

大模型具备多维度分析能力和全面预测能力，通过对海量数据进行训练分析，同时结合当前车内状态以及与用户的上下文中进行理解，然后在复杂场景下做出准确判断。例如根据分析出的多个指令，控制车窗的具体打开程度，车门开关，各类灯光（远光，近光，雾灯）的开关，后视镜开合，雨刷器运行等功能。根据习惯分析用户的意图，控制各类媒体播放器，播放音频，视频等多媒体信息。理解用户的复杂指令，泛化出用户的详细操作指令，控制各类多媒体应用程序播放音乐，视频和图像。控制音量大小，播放进度，显示大小等。

2) 持续优化与自我学习

通过不断与用户交互，大模型可以持续学习用户的习惯，这是传统基于规则的指令控制所不具备的。能够识别使用者，理解使用者的语言习惯，能够更精准地读懂用户表达不够充分的语言。根据习惯分析用户的意图，调节空调温度风速，座椅展开角度，灯光类型，雨刷器模式等车辆设置。持续学习后，大模型可以更加准确的识别根据用户习惯进行的指令控制。

3) 主动推荐

基于大模型的车内指令控制可以通过当前环境状态主动推荐用户做出某些特定的指令，在用户进行确认后执行该指令。实现更加智能化的指令控制。

5.1.3 情感化、拟人化助手

情感交互助手是指大模型通过分析使用者的情感，生成符合使用者情感的回答内容，实现共情闲聊。能够了解用户的情绪和心理状态，并通过对话完成心理解压和疏导；用户遭遇情感难题时，可以为其提供情感指导；遇到职场问题时，也可以在为其答疑解惑。此外，大模型在与驾驶员对话时，大模型可以根据对话内容和情绪氛围生成更加贴合情感的回复，使对话更加生动和富有同理心。大模型在情感化、拟人化助手中有以下几种应用价值：

1) 情感化导航指引

在提供导航指引时，大模型可以根据驾驶员的情绪状态调整语言风格。如在驾驶员紧张时使用更加温和和安抚的语言。

2) 情感化紧急响应

在发生事故或车辆故障时，系统可以生成富有同情心的语言以及语音来安慰车主，并提供紧急情况下的指导和支持。

3) 拟人化的用车助手

基于大模型的理解能力，帮助用户在车辆参数，零部件操作，功能使用，汽车保养以及故障处理等全链路用车问题上可以随时随地获取专业的用车知识和解决方案，提高用车的便利性和安全性，成为用户贴心的用车助手。

- 拟人化的车内健康顾问

基于大模型，可以为车内配备一个虚拟健康顾问，能够结合与车主对话的情绪状态以及 DMS，OMS 识别到的当前车主的生理状态提供健康建议与情绪支持。

- 拟人化角色定制

基于大模型，用户可以根据个人喜好定制一个或多个虚拟角色，并可赋予角色不同功能，比如旅游规划助手等。这些角色能够展现不同的性格特征，并在与用户的交互中展现相应的情感反应。

- 情感化历史解说

基于大模型，当车辆经过某个历史地标区域时，大模型可以结合驾驶员的情绪状态提供相匹配的历史解说，使历史信息更加生动化，情感化。

- 自由对话

实现人机自由对话，多轮对话。通过语音助手可以对话的内容包括但不限于知识问答，景点推荐，美食推荐，口语教练，信息查询等功能。

5.1.4 语音合成

语音合成是指将文本合成为语音的过程，传统的语音合成方法首先对文本进行语义分析，再将结果进行波形预测后合成得到语音。这种方法虽然可以得到较为自然的音频，但是在波形预测的过程需要大量的音频片段以供后续预测拼接，这使得语音合成存在很大的局限性。

基于大模型的语音合成相比于传统方法存在以下几种优势：

1）自然度与灵活性

大模型通过学习到更丰富的语音特征，能够生成更加流畅的语音。同时大模型可以适应不同的语音风格和情感表达，提供更多样化的语音输出。

2）个性化与多语言支持

大模型可以模仿特定人的语音特征，实现高度个性化的语音合成。大模型同样支持多种语言的语音合成，而不需要为每种语言单独开发模型。

3）上下文理解

大模型具有传统方法所不具备的长上下文理解能力，可以通过上下文生成合适的语音风格以及情感等。

大模型在上述功能或者产品体现出对于强大的理解能力，泛化能力进行语音识别或合成，以及通过推理能力解决用户的复杂需求。大模型将智能网联汽车中若干个小功能模块全部集成到一个大模型内部，不需要对独立模块进行开发。在情感化拟人化场景中，大模型作为情感助手，体现出对于上下文的理解能力，通过对用户的情绪理解来安慰或鼓励用户。

5.2 大模型应用于视觉交互场景

基于大模型的视觉交互技术在智能座舱领域展现了广泛而多样的应用场景。从驾驶员监控到提升乘客体验，从提供安全辅助到实现个性化设置，大模型技术在各个方面都发挥了重要作用，极大地增强了智能座舱的功能和用户体验。尤其体现在驾驶员监控系统(DMS)、乘员监控系统(OMS)、人脸识别 (FaceID)。这些技术的日益成熟，不仅标志着汽车座舱向智能化、个性化空间的转变，也是智能出行时代加速演进的关键驱动力。在此基础上，大模型(LLM)的引入能够对智能车舱内视觉交互场景进行进一步深度优化与革新，以下是大模型应用的视觉交互场景。

1）疲劳检测

通过摄像头监控驾驶员的眼睛和面部表情，检测驾驶员的疲劳状态，并在必要时发出警告。

2）注意力监控

追踪驾驶员的视线，判断其是否在关注道路，若注意力分散（如长时间未看前方），系统会发出提醒。一旦系统发现驾驶员注意力不集中，如频繁眨眼、视线偏离道路过久等行为，会立即触发警示，通过声音、震动或座椅震动等方式提醒驾驶员，必要时甚至自动介入，如减速或开启紧急停车程序。

OMS 关注的不仅仅是驾驶员，而是整个车厢内所有乘客的安全与舒适。OMS 结合大模型和多模态，能够更细致地监控乘员动态，例如乘员的位置、年龄、是否系好安全带，甚至通过面部表情和身体姿态判断乘客的情绪状态。

3）驾驶行为分析

通过摄像头监控驾驶员的行为，如频繁转头、使用手机等，提供驾驶行为分析报告，并在检测到危险行为时发出警告。

4）主动响应与环境感知

结合视觉和其他传感器数据，实时监测车内外情况，做出智能响应，如检测到紧急情况自动报警等、根据车内外环境的变化，自动调整车内设置，如雨天自动关闭天窗、夜间自动开启车内灯光等。

5）人脸识别

系统可以根据识别结果自动调整座椅、镜子、音响等参数，提供个性化的驾驶体验。这种个性化设置不仅提升了舒适度，还增强了用户的满意度和忠诚度。通过大模型的能力，系统能够不断学习和优化用户的偏好设置，提供更贴心的服务，例如提供个性化内容推荐，基于视觉识别的乘客身份信息，推荐个性化的娱乐内容，如电影、音乐等。

总体而言，这些融合了驾驶员监控系统(DMS)、乘员监控系统(OMS)和人脸识别(FaceID)等技术的应用，不仅显著提升了汽车的智能化和安全性，还为驾驶员和乘客带来了更智能、更便利、更个性化的驾乘体验。通过大模型的学习、推理和理解能力，使得人机交互更为自然流畅，能够精准捕捉并响应用户的细微需求与情绪变化，从而定制化推送信息、调节氛围照明、优化路线规划等。

智能座舱利用先进的视觉大模型技术，实现了驾驶员监控、乘客识别、车内环境感知等功能，极大地提升了驾驶安全性和用户体验。本文将重点探讨视觉大模型在智能座舱中的应用场景。

5.3 大模型应用于多模态交互场景

智能网联汽车中的多模态大模型应用是指结合语音、视觉、行车信息，生物信息等多种模态的输入信息，通过模态编码，由大语言模型进行语义化理解和泛化推理，再通过生成式模型输出文本，任务，语音，图像，视频等多种模态信息。多模态大模型极大的改善了智能座舱交互方式。下面从精确感知乘客意图、车辆状态感知与反馈，智能体虚拟人交互体验和多模态生成能力的应用四个方面分析大模型在多模态交互场景中的应用价值。

5.3.1 精确感知乘客意图

多模态大模型可以根据用户的语音信息（说了什么），用户的姿态信息（做了什么），以及车内的环境信息（周围是什么）来精确感知用户的指令意图。避免理解不充分和误解用户的真实意图。相较于传统算法对于复杂，可变，无法预测的环境适应能力较弱，容易受到各类输入噪声和未训练情况影响，多模态大模型可以将麦克风采集到的各个位置上的语音信息，摄像头采集到的视频信息以及其他传感器的信息通过编码向量化作为大模型的输入信息。再凭借大模型的泛化能力，模糊问题理解能力，zero-shot 零次泛化能力可以更准确地感知和理解用户的输入和环境的信息，从而做出更加准确的判断和行动。例如，通过车内语音和视频信息判断说话人的潜在语义，肢体动作，视角朝向，情绪，从而理解对方具体想操作座舱的哪部分功能。不需要用户说非常冗长的话，也不用担心用户一直想不起来这个东西叫什么。

5.3.2 车辆状态感知与反馈

多模态大模型除了可以感知来自用户的信息之外，还可以感知车辆的信息。包括但不仅限于车内的温度，湿度，空气质量，车辆的胎压，发动机温度，车速等信息。相较于传统算法基于设定的模式和规则来判断车辆状态，多模态大模型可以将座舱内温度，湿度，空气质量等各类传感器获得的信息，车辆的速度，制动信息等行驶信息，车辆的发动机温度，胎压等状态信息通过编码向量化作为大模型的输入。再利用大模型的学习能力，泛化能力和推理能力判断车辆状态并自动调整车辆设置或者向乘客预警。大模型可以不拘泥于固定的模式和规则，适应多变的车辆状态和车内环境以及用户的个性化需求。例如，可以综合车内的温度和湿度传感器、空气质量监测器等数据，感知

座舱内实际的温度、湿度和空气状况。基于用户偏好实时调节空调系统来调节座舱温度和湿度，调整通风系统和空气过滤器。也可以实时监测轮胎压力、发动机温度等机械状态，一旦出现异常，大模型可以及时溯源问题来源，并通过智能座舱进行预警提示。

5.3.3 智能体虚拟人交互体验

多模态大模型可以实现具备人类丰富表情，肢体动作，情绪，语言能力的虚拟人 AI 智能体，逐渐接近与真人交流的感受。也可以通过形象复刻，声音复刻

来模拟真实存在的人类，实现交互对象的 3 维数字孪生。通过虚拟人 AI 智能体，可以像老朋友一样理解乘客的需求，模拟人类的语言，行为和情绪和表情来和乘客进行交互。帮助乘客操作座舱中的各项功能，操作各种智能座舱 APP，实现复杂的操作需求。与传统的机械式，模式化，流程化的人机交互方式相比，多模态大模型凭借大模型的推理和泛化能力赋予虚拟人带有情绪的，更贴近于日常语言的语音表达。并且大模型根据表达内容和情绪来生成和控制虚拟人的面部表情和肢体动作。在虚拟人建模方面，也可以根据用户提供的素材，描述来定制有个性化的虚拟人形象。例如，参照多模态 AI 智能体在电脑游戏中扮演 NPC 的应用方式，让智能体虚拟人可以像真人小伙伴一样和后排的小朋友做各种游戏，指导孩子完成作业。也可以体会乘客的情绪，对乘客进行心理辅导，情感交互。在交互的体验方式上，从与传统的人于机器交互的体验转变为与高度拟人化的虚拟人 AI 智能体进行交互，是交互方式的重大升级，也将重塑智能座舱交互的方式。

5.3.4 多模态生成能力

多模态大模型在生成能力方面可以实现文生图，文生视频，文生音乐，图生视频等多样的功能。实现各种模态信息之间的创新创造，为智能座舱附加新的能力。多模态大模型首先将输入的文字，语音，图像等信息通过预训练模型进行编码和特征提取，然后通过扩散模型（Diffusion）以及 Transformers 架构生成符合描述的图像，视频。其中，扩散模型负责从随机噪点图像数据逐步去噪得到清晰图像，而 Transformer 架构则用于降低视觉数据的维度，生成不同分辨率的内容。让大模型能够生成高质量、高分辨率的图像及视频，同时保持较高的处理速度。多模态大模型的生成能力赋予了座舱很多新的应用场景。例如，可以根据语音描述生成车机的壁纸，满足用户个性的多样性的需求；也可以根据孩子的描述信息生成童话绘本，甚至一段儿动画片视频，提升乘客的幸福感和满意度；还可以根据语音，图像，行驶信息识别驾驶员和乘客的情绪和需求，生成合适风格的音乐。

综上所述，相较于其他现有的技术，多模态大模型可以将多种模态的数据作为大模型的输入，利用大模型参数量大，适应力鲁棒性强，泛化能力强的特点，能够理解复杂环境，处理更加复杂的任务。模拟人类的动作，语言和表情使得智能网联汽车中的多模态交互场景更加丰富与智能，更加贴近人类交流的方式。还

可以通过用户的描述生成图像，视频，音乐等多种形式的输出。目前吉利，理想，商汤等多家企业都推出了自研的多模态大模型架构。未来，随着各个细分领域模型技术的进步，多模态模型的感知编码能力，理解和逻辑推理能力，以及生成泛化创新能力都将持续提升。现有的模态将进一步扩充，从现实世界中获取更多类型模态的信息作为输入。各类训练数据的质量和多样性的提升，以及超大数据量训练系统的应用，将提升模型的训练效率以及各项能力和质量。同时，随着模型轻量化部署技术的发展，以及车机端芯片能力的提升，更多的多模态应用场景可以在车机端运行推理，在网络环境不佳的特殊情况下也能运行部分功能。从整车的视角看，借助由多模态大模型塑造的多模态 AI 智能体，汽车从自主行动，拟人化交流，创新创造等几个方面将带有更多的机器人特征。智能汽车将逐渐转变为智能汽车机器人。

5.4 大模型应用于开放式任务场景

大模型开放式任务指的是通过大模型来检索完成某一任务需要的信息，并通过系统传感器、数据库等渠道获取跨领域信息来完成复杂的任务，逐步实现座舱智能体，构建一个立体感知、全域协同、精准判断、持续进化、开放的智能系统。相比传统 AI，大模型具有更强的学习能力、推理能力和理解能力，能够处理更复杂的任务并提供更智能的解决方案。下面从任务规划、生成式交互、健康监测、插件与信源扩展和检索增强生成五个方面分析大模型在开放式任务中的技术路线和应用价值。

5.4.1 任务规划

大模型在复杂任务生成方面的应用体现在三个方面：首先，在对话交互方面，车机系统可以精准识别用户的语音指令，并创建符合用户需求、用户偏好及与场景匹配的复杂任务。例如，主驾可以通过语音指令在车机系统上设置儿童模式：“当后排乘客在下午四点后上车时，就开启后排空调到 25 度并播放英语教学视频。”此一指令覆盖了车辆设置及多媒体等范围，而车机系统可以根据主驾的要求创建快捷的复杂任务，并主动地集成与任务相关传感器的信号在合适的时机点执行对应的任务。这是有别于以往传统单点功能式交互的开放式任务。

其次，在任务规划方面，大模型可以将复杂的任务拆解为更小的子任务。例

如，用户提出“根据日出日落开关氛围灯”，大模型可以将任务拆解为日出关闭氛围灯和日落打开氛围灯两个任务。

最后，在数据管理和记忆的方面，大模型可以作为工具，帮助 Agent 保存和检索过去的数据。例如，在用户的任务大师中，大模型可以帮助记住用户的偏好设置，并在需要时提供相应的信息。总之，大模型在任务大师中的应用可以提高任务的执行效率，增强交互的自然性，并提供个性化的服务。通过合理设计和应用大模型，可以实现更加智能和高效的任务管理。

5.4.2 生成式交互

大模型在生成式 UI（用户界面）方面的应用，极大地提升了智能座舱的能力。通过自动生成脚本和接口调用方式，并进行动态注册，大模型使得车载信息娱乐系统和其他车载功能的集成更为高效和灵活。生成式 UI 意味着界面和交互模式可以根据用户需求和环境变化进行实时调整和优化，从而提供个性化的体验。

首先，大模型利用其强大的自然语言处理和理解能力，可以自动生成与用户需求相对应的脚本。这些脚本能够描述各种交互场景、功能调用和用户操作步骤。举例来说，当驾驶员通过语音指令请求导航到某一目的地时，大模型能够实时生成对应的导航脚本，包括路径规划、交通状况更新和实时语音指导等功能。一旦生成，这些脚本可以自动调用相关接口，实现无缝的功能集成。

其次，动态注册功能使得系统能够灵活应对不同的场景和需求。在智能座舱内，大模型可以通过动态注册机制，自动调整车载 UI 的布局和功能。例如，当检测到车内有儿童时，系统可以自动注册并显示相关的娱乐内容和教育应用。这种高度动态化的界面配置，使得智能座舱能够更好地适应不同用户和场景的需求，提高了使用便捷性和安全性。

此外，通过自动生成与调用接口方式的脚本，系统可以更迅速地整合新功能和应用。大模型在后台自动生成接口调用代码，简化了开发流程，大幅缩短了应用更新和功能拓展的时间。这意味着智能座舱能够更快地响应市场变化和用户需求，保持技术的前沿性和竞争力。

综上所述，利用大模型的生成式 UI 能力，智能座舱实现了更高的灵活性、个性化和便捷性。通过自动生成脚本、动态注册和接口调用，大模型不仅简化了系统开发和更新流程，还提供了更加智能化和人性化的用户体验，进一步提升了

智能座舱的整体能力。

5.4.3 健康监测

近年来，智能座舱在汽车行业中的关注度日益提升，旨在提升驾驶体验和安全性。结合驾驶员心率、血压等生理数据，通过大模型的分析和预判，可以显著增强智能座舱的能力。首先，大模型具有强大的数据处理和分析能力，能够实时监测并解析驾驶员的生理状况，识别潜在的健康风险。例如，如果的心率突然升高，血压不稳定，大模型能够智能预测这些生理变化背后的原因，如疲劳、压力或潜在的健康问题，从而及时给予预警。

其次，大模型可以通过深度学习，对大量的驾驶数据进行训练，形成对驾驶员习惯和状态的理解。这种个性化的分析能够为驾驶员提供更加准确和量身定制的建议，如提示驾驶员进行短暂休息，调整驾驶姿势，或是激活辅助驾驶模式以减轻压力。这样的个性化建议不仅提升了驾驶舒适度，还有效降低了因身体状态引发的交通事故风险。

最后，通过大模型的介入，智能座舱可以实现多模态交提升人车互动的智能化水平。当系统检测到驾驶员状态异常时，可以通过声音提示、座椅震动等多种方式提醒，使得驾驶员更容易注意到并采取相应措施。整体而言，大模型在智能座舱中的应用，不仅提升了车内环境的智能化程度，更强化了安全保障，推动了汽车行业迈向更加智能化、人性化的未来。

5.4.4 插件与信源扩展

大模型结合插件和多种信源的能力，为智能座舱的功能和体验带来了显著提升。通过将大模型与各种插件和信源集成，智能座舱不仅能实现功能的高度定制化和扩展性，还能提供更加详尽、实时的环境感知和响应能力。

首先，大模型与插件的结合使智能座舱具备了高度的扩展性和灵活性。插件可以视作独立的功能模块，通过加载不同的插件，智能座舱能够迅速增加新的功能或优化现有功能。例如，一个智能座舱可以通过加载语音助手插件实现更精准的自然语言处理和对话功能，通过加载健康监测插件实时获取和分析驾驶员的生理数据，从而做出个性化的建议。大模型在这一过程中通过智能分析和决策，指导哪些插件需要优先加载，以及如何高效调用插件的功能，为驾驶员提供最佳解

决方案。大模型也可以通过学习汽车的用户手册，快速引导教学用户熟悉车辆。

其次，信源的多样性和实时性是智能座舱能力提升的重要因素。大模型能够从多个信源获取数据，包括但不限于车内传感器数据、外部交通信息、天气预报、地图服务、驾驶员健康状态等。通过对这些多维度数据的整合分析，大模型能够在车内环境中做出智能决策。例如，通过整合交通信息和天气预报，系统可以提前建议驾驶员选择更加安全的行车路线或合理的出行时间。

最后，大模型在接收信源和调用插件功能的同时，能够不断学习和适应用户的习惯和偏好，实现个性化定制。它可以通过对历史数据的分析，预判驾驶员可能的需求和行为，从而提前准备相关信息和功能。例如，大模型可以根据驾驶员的日常路线和习惯，为其推荐加油站、休息区或餐馆，实现“未询先知”的智能化服务。

总的来说，利用大模型在插件和多信源结合方面的优势，智能座舱能够提供更灵活的功能扩展、更精准的环境感知和更智能的用户互动。这不仅提升了驾驶体验的舒适性和安全性，也推动了智能汽车技术的进步和普及。通过持续学习和优化，大模型将在未来进一步提升智能座舱的综合能力，带来更加卓越和个性化的驾乘体验。

5.4.5 检索增强生成

检索增强生成（RAG）是一种根据用户的查询语句搜索信息，并以搜索结果为 AI 参考从而生成回答。这项技术是多数基于 LLM 工具的重要组成部分，而多数的 RAG 都采用向量相似性作为搜索的技术。在文档中复杂信息的分析时，GraphRAG 利用 LLM 生成的知识图谱大幅提升了问答的性能，这一点是建立在近期关于私有数据集中执行发现时提示词增强能力的研究之上。通过 LLM 构建知识图谱结合图机器学习，GraphRAG 极大增强 LLM 在处理私有数据时的性能，同时具备连点成线的跨大型数据集的复杂语义问题推理能力。普通 RAG 技术在私有数据，如企业的专有研究、商业文档表现非常差，而 GraphRAG 则基于前置的知识图谱、社区分层和语义总结以及图机器学习技术可以大幅度提供此类场景的性能。

第六章 智能座舱 AI 应用的关键技术

随着智能网联汽车技术的快速发展，大数据和人工智能技术在汽车行业中的应用越来越广泛。汽车正迅速从传统的交通工具转变为提供丰富体验的移动空间。用车场景也日益多样化，用车需求的新变化推动了行业向更主动且智能化的服务方式转变。本章将针对智能座舱内 AI 应用的关键技术，从感知、认知、表达三个方面阐述在 AI 算法是如何重塑座舱功能并提供个性化体验的。随后，通过 AI 大模型的异构部署充分释放座舱大模型算力，提供更高效的运算平台。

6.1 智能座舱感知技术

6.1.1 声学前端技术

音频信号处理的最前沿，主要负责对输入的原始音频信号进行预处理，如：噪声抑制、回声消除、语音增强等处理，以消除噪声、回声等干扰因素，对于提高语音识别的准确性和鲁棒性至关重要。且随着技术的不断发展，从传统的信号处理方法到深度学习方法，再到可学习音频前端的出现，声学前端技术正不断向更高效、更智能的方向发展。

（1） 噪声抑制（Ambient Noise Suppression，ANS）技术

噪声抑制是指当语音信号被外界的噪声所干扰时，使用一定的方法提升有用的语音信号质量，抑制外界的噪声。因此降噪的研究实际上是要解决 2 个问题：一是提升有用语音信号的可懂性；二是改善语音的质量、减轻噪声带来的影响。按照研究思路划分，可以分为传统信号处理算法降噪以及深度学习算法降噪。

谱减法是最早提出的语音降噪算法之一，也是最经典的算法。它来源于简单的原理：假设带噪语音信号中的噪声只有加性噪声，且噪声是平稳或缓慢变化的情况下，将带噪语音信号的频谱减去估计出的噪声信号的频谱，就可以得到纯净的语音频谱。

（2） 回声消除（Acoustic Echo Cancellation，AEC）技术

回声的产生简而言之就是自己的声音经过一系列的反射之后，使得讲话人又听见了自己的声音。目前主流的回声消除方法一般可称之为自适应回声消除(Acoustic Echo Cancellation, AEC)，其中自适应滤波器将发挥关键性作用，

且自适应算法的性能也将决定着回声消除的能力。

图 6-1 为 AEC 的基本原理图，由远端房间内讲话人发出的话音信号 x(n)在近端房间内通过回声路径生成回声信号 r(n)，连同近端可能存在话音信号 s(n)和噪声信号 v(n)构成期望信号 d(n)。另一方面，x(n)与自适应滤波器权值 w(n)产生回声的估计信号 y(n)，若将 y(n)从 d(n)中减去则可得到误差信号 e(n)，并将 e(n)传送至远端房间。若 w(n)与回声路径脉冲响应 h(n)达到一致，y(n)即等于实际回声信号 r(n)，则 e(n)中将不包含回声信号。事实上，由于噪声等因素的制约，能够使得估计的回声信号完全等于实际值的自适应滤波器是不存在的，它能做到的只是无限逼近于实际回声信号。因此，AEC 的本质是利用自适应滤波器去逼近回声路径的脉冲响应来产生回声的估计信号，以便在期望信号中减去此估计信号。

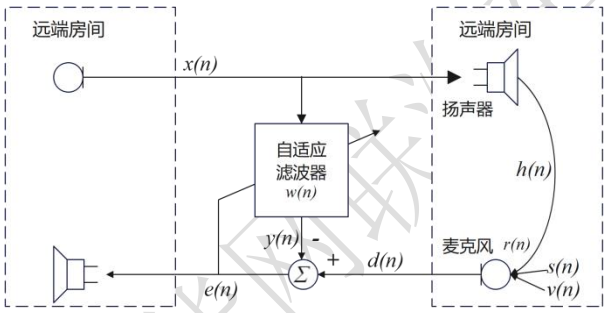

图 6-1 回声消除（Acoustic Echo Cancellation，AEC）技术原理

人工智能在回声消除技术中的应用主要体现在通过深度学习算法来提高回声消除的效果和适应性，尤其是 Transformer 等基于注意力机制模型的广泛应用，借助大量的数据训练，自动学习识别回声信号与原始信号之间的差异，从而更精确地消除回声。如 anyRTC 通过结合传统 AEC 算法和深度学习，改进传统 AEC 技术中的，如噪声抑制效果不明显和双讲效果差等痛点。

（3） 语音增强技术

语音数据由于受到各种复杂环境的影响，接收到的语音会有噪音、混响、人声干扰、回声等各种问题，特别是远场环境，混响会严重影响 ASR 的性能。因此需要通过抑制、降低噪声干扰，从嘈杂的背景环境中提取出有用的语音，即尽可能纯净的原始语音，这就是语音增强技术。

语音增强系统包含数据读取、语音降噪、语音去混响和结果呈现与保存四个模块。数据读取模块支持单条语音读取或多条语音批量读取。数据读取模块由系

统配置文件定义，只需在指定目录下传入待处理的语音数据，即可被系统读取并处理；语音降噪模块对输入的语音信号进行降噪处理去除噪声并将处理后的信号输出至去混响模块，进行去混响处理，最终获得增强的语音数据。

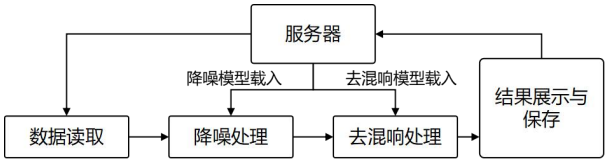

图 6-2 语音增强系统流程图

主流的语音增强方法有很多，经典的方法有谱减法、维纳滤波法、基于统计模型的方法和子空间算法。随着人工智能的发展，神经网络也被应用于语音增强。按照其运用方法的不同进行分类，可大致分为两类：一类是传统的语音数字信号处理的方法，一类是基于机器学习的方法。

在传统的数字信号处理方法中，按照其通道数目的不同，可以分为：单通道的方法和麦克风阵列的方法。

在传统的单通道语音增强方法中，短时谱估计的方法应用最为广泛，具体的算法可以分为三类：谱减法、维纳滤波法和基于统计模型的方法。除了这三类算法，还有一种自适应滤波法，但需要获得噪声语音数据或者纯净语音数据，进而用随机梯度下降向最优解优化，但是大多数情况下，噪声或者纯净语音等先验知识不易获得，因此造成了这种方法的使用难度较大的现象，不过通常手机会收集来自环境中的噪音信号，通过使用一个降噪麦克风来实现，并将这些噪音信号输入，作为降噪参考。此外，还有基于子空间的方法和基于小波变换的方法等。

6.1.2 语音识别技术

语音识别也被称为自动语音识别（Automatic Speech Recognition，ASR），其目标是将人类的语言转换为机器可以识别的指令。语音识别被认为是促进人类与机器、人类之间通信的重要桥梁。其中涉及包括：信号处理，概率论，发生机理，模式识别，人工智能等一系列实用领域。

语音识别主要有语音输入，语音特征提取，识别算法解码，文本输出这四个过程。如图 6-3 所示：

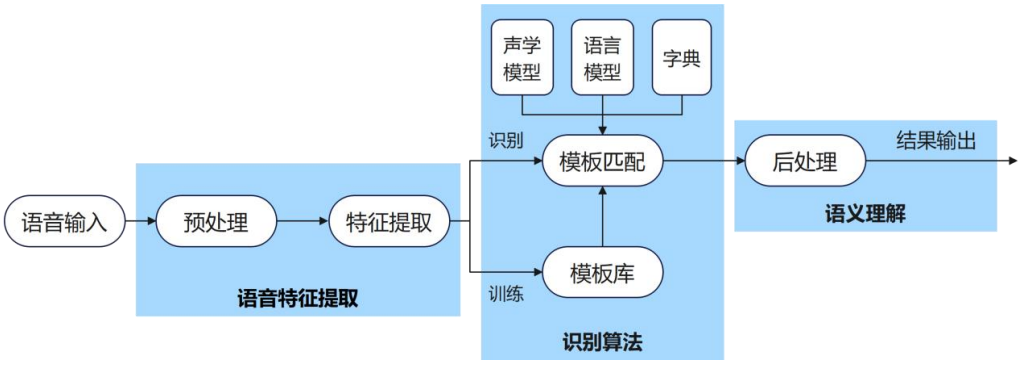

图 6-3 语音识别技术框架图

语音信号预处理是指将语音信号经过预加重、分帧加窗、傅里叶变换、取对数等过程，提取音频的 Fbank 特征或者 MFCC 特征。由于语音数据集的采集过程中，录音设备、噪音以及周围环境都会影响到采集质量，为了获得表示语音本质的特征，需要对音频信号进行预处理操作。

目前语音识别领域对语音信号的预处理分为两种。一种是经过短时傅里叶变换（Short-time FourierTransform, STFT）生成音频特征，其中使用较为广泛的两种音频特征是 FilterBank(FBank)特征和梅尔频率倒谱系数(Mel Frequency Coefficients, MFCC）特征。另一种是直接使用原始语音信号作为声学模型的输入，在神经网络模型中隐含的提取语音信号特征。

识别算法解码无疑是语音识别技术的核心。解码器主要由两个模型组成：声学模型(AM)和语言模型(LM)。声学模型就是将发音图谱转换成文字的对应关系，是对语言学，声学的矢量化表述。而语言模型则是根据人类语言的特征特点结合声学模型的结果匹配语义的对应关系，它是一组对知识化序列的表述。

所谓语音识别其实就是利用声学模型把语音的声学特征分类对应到音素或者字，词这样的最小单元，然后利用语言模型接着把字词解释成一个完整的句子的过程。

6.1.3 语音唤醒技术

语音唤醒任务可以看作是一种小资源的关键词检索任务，主要特点为计算资源小，空间占用资源小。语音唤醒结构图如下图所示。

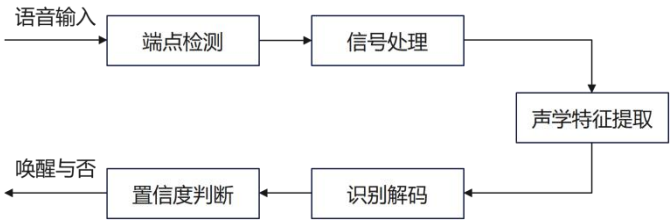

图 6-4 语音唤醒结构图

语音的关键词检索实际上是在连续音频流中检索出关键词数据的过程。与大量词汇连续语音识别(LVCSR)不同的是，语音唤醒不需要将音频流中的所有内容精确识别，只需要该音频流中是否含有关键词即可，有效地降低了对 ASR 系统的要求。关键词检索系统实质上是一个语音的二分类任务，并不需要将音频流转移为文本内容，通过学习关键词和非关键词的音频特征，直接识别音频流是否包含关键词即可。

基于深度学习的神经网络模型，能够从大量的训练数据中学习到语音的特征模式，从而在复杂背景噪声中准确识别目标唤醒指令。此外通过复杂的声学模型来分析语音信号的频谱特征、时间结构和上下文信息，使得机器能够区分唤醒词与日常对话中的相似词汇，减少误唤醒率。不仅提高了识别的准确性和鲁棒性，还增强了个性化、多语言支持和安全性。

6.1.4 人脸识别技术

人脸识别是指使用图像采集设备，采集人脸的动态视频或者静态图像，然后提取有效的人脸特征，通过对这些特这信息进行分析和归类，来实现人脸识别这一目标。通常来说，人脸识别的研究内容包括人脸检测与对齐，面部特征提取，人脸识别与表情分析等多个方面。

典型的人脸识别可以概括为四个步骤：第一步是确定从各种场景获得的目标图像的人脸的区域位置，并实现人脸位置的几何统一，便于后续的进一步处理；第二步是对检测到的面部图像进行图像预处理，包括面部图像的几何校正、面部光照预处理等；第三步是从预处理后的人脸面部特征中提取出有效的人脸特征信息，然后使用可以计算的、具有表达特征性的数值矩阵来表示提取出的特征；第四步是使用合适的匹配算法对特征信息进行分类和划分，以达到人脸识别的目的。

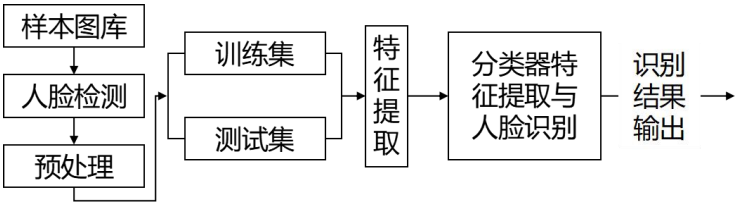

图 6-5 典型的人脸识别过程

人脸识别技术的起源可追溯于 1889 年《Nature》上发表的一篇人脸身份识别论“Personal identification and description”，自那时起到现今，在计算机视觉领域中，人脸识别已经成为最为成功的应用之一。早在上世纪的六十年代，研究人员就开始了把机器用于人脸识别的探索。

6.1.5 动作识别技术

人体动作识别方法多种多样，其分类也有不同的标准。概括来说，动作识别的过程是一个对动作符号进行标签化的过程，其本质是模拟人眼和人脑对动作信息进行理解区分。因此，结合动作信号的特殊性，从流程分类法的观点出发，可以归纳出典型的动作识别问题通常包含三个主要研究内容：运动目标检测、动作特征提取和动作特征理解，其一般流程如下图所示。

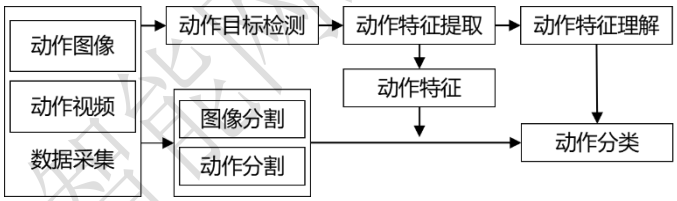

图 6-6 典型人体动作识别流程

传统的动作识别方法是基于手动的特征提取，可分为全局特征方法和局部特征方法。全局特征首先定位出人体位置，然后进行动作识别。但是这种全局的特征包含很多噪音，并且对动作遮挡等情况异常敏感。局部特征是提取一系列的点，然后通过这些点周围图像区域的组合来表示某个动作，这种方法不依赖人体追踪和分割技术，并且对噪音和遮挡等情况不敏感，因此广受欢迎。

随着卷积神经网络（Convolutional Convolutional Network，CNN）的发展，利用深度学习的方法解决动作识别问题已经取得了很大的进展。目前学界利用深度学习解决动作识别主要分为两类方法：双流的网络和 3D 卷积神经网络。

6.2 智能座舱认知技术

6.2.1 语义理解

所谓语义指的是语言的含义，表示在语言中的信息，是一种人对非结构化的语言的抽象的体现。语义不仅仅是对事物的本质进行表述，还包含因果、上下位等多种逻辑关系。

语义理解是自然语言理解(Natural Language Understanding, NLU)所研究领域的一部分。通过建立模型，使计算机能够像人类一样对自然语言进行理解，识别信息中所包含的真实语义。语言的理解与分析的过程是一个由表及深，由内至外的层次化的过程。如图 1-31 所示。

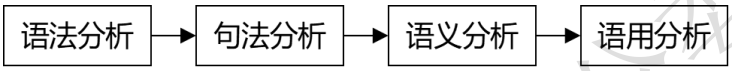

图 6-7 NLU 过程的一般步骤

其中：词法分析的主要目的是从词汇中找到词汇的各个词素，以获得语言学信息；句法分析将从词汇中找到的各个词素按照其线性次序变换成显示其内在联系的结构，并分析它们在结构中的作用。常用的句法分析方法有短语结构语法、扩充转移网络、功能语法、格语法等；语义分析则是建立词法意义、结构意义和任务领域对象之间的联系，从而确定语言所映射的任务指令或概念；语用分析是为了正确理解语言在其特定的语言环境中的含义。

传统的语义理解算法分为基于词的语义理解算法和基于主题的语义理解算法。基于词的传统语义理解算法如词袋模型（Bag of Words，BoW），将文本看作是词的集合。当全部文本有N个词，词袋模型用一个长度为N的向量表示文本，向量的每一个值为对应词的频率。然而，词袋模型中所有的词同等重要，这忽略了关键词的影响。基于主题的语义理解算法，相比于基于词的语义理解算法，更好的克服词的多义、冗余的情况。第一个提出的主题算法是１９９０年的潜在语义分析（Latent Sementic Analysis，LSA）算法，并被用在搜索引擎中。但是基于主题的语义理解算法无法处理文本中上下文信息，导致语义表示向量无法考虑整个文本，制约了语义理解的效果。

随着深度学习的兴起，基于深度学习的语义理解算法，因为可以获得文本的上下文信息，语义表示向量可以考虑整个文本，开始被广泛使用。基于深度学习的语义理解算法可以分为三大类：（1）无监督语义理解模型；（2）有监督语义理解模型；（3）自监督语义理解模型。

无监督语义理解模型通过大规模语料库训练出词表示向量。然后，对文本的所有词的词向量进行求和或加权求和操作，获得该文本的语义表示向量。典型算法有：2013 年谷歌提出的 word to vector（word2vec）、2014 年提出的 Global Vectors for Word Representation（glove）、2017 年的 smooth inverse frequency（SIF）等。

有监督语义理解模型为了能够根据具体场景获得合适的语义表示向量而被提出。有监督语义理解模型使用标注数据作为标签，通过深度学习进行建模，通过端到端学习的学习方式，直接得到文本的语义表示向量。

为了解决标注数据难的问题，自监督语义理解模型被提出。自监督的语义理解模型使用自监督损失函数，摆脱了对标注数据的依赖。

6.2.2 对话引擎

在语音交互场景中语音识别技术（ASR）是为了让机器听见，自然语言理解（NLU）是为了让机器听懂，听懂之后具体的执行则是对话引擎的工作。对话引擎是获得接近甚至等同于人的交互体验的关键，在不同的系统中，对于对话引擎不同的翻译，如：对话引擎 Conversation Engine，或对话管理 DM（Dialog Management）。

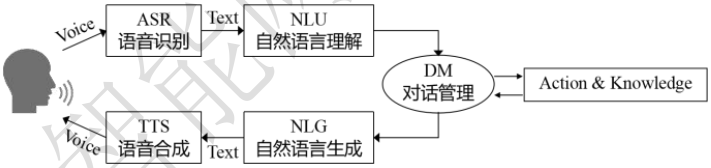

图 6-8 基于规则的对话引擎

传统的对话引擎系统是基于逻辑规则实现的，这种方式可以借助人类专家的知识，为特定的领域快速构建起重要的对话逻辑流程。基于逻辑规则的对话引擎通常是基于有限状态机（Finite State Machine，FSM）或框架（Frame）原理，通过人工编撰的方式定义对话系统的对话状态、状态之间的跳转逻辑，或基于对话状态的对话决策逻辑。这种方式的优点是灵活可控、定制性强，特别适合于对话状态定义及其跳转、决策规划清晰明确的场景；其缺点是解决对话过程中的离题、歧义等异常时需要比较繁琐的配置。

基于机器学习的对话引擎的最典型方法是将对话过程表示为一个部分可观察马尔可夫决策过程（Partially Observable Markov Decision Process，POMDP），其输入是对话理解的结果，输出是下一个时刻的动作策略。POMDP 的内部状态

是针对所有对话状态的概率分布，在每个状态下，系统执行某个动作会有对应的回报，选择下一步动作的策略即为选择期望回报最大的那个动作。该方法的优点是只需定义马尔可夫决策过程中的状态和动作，状态间的转移关系可以通过学习得到，并且可以使用强化学习在线习得最优的动作选择策略；缺点是需要较大规模的，至少是标注的对话训练语料。

6.2.3 场景引擎

随着智能座舱功能的不断发展，座舱功能多样化与用户需求满足度之间的矛盾开始凸显。在传统的竞品对标方法下，更多的座舱功能配置堆积反而使用户更加无所适从。到了智能汽车时代，由于主机厂底层的基础能力越来越相似，突出的差异化不在于外观，而是用户体验。在此背景下，主机厂需要定义自己的智能场景，通过场景将用户需求与智能座舱功能进行对接。

场景即是在某个时间和空间下发生的有始有终的事情片段。用户的日常生活可以分解为一个又一个的场景，将这些场景发生的地点限制到汽车座舱内，也就是汽车智能座舱场景。以场景为中心，基于对用户的理解，用数据和算法驱动主动服务，即可为用户提供高频、刚需、原子化、触手可及的服务。场景引擎融合了人工智能引擎，与车辆、用户、环境、生态、交通数据等深度融合，利用 AI 智能算法对车主在开车期间的下一步需求进行预判并提供服务，增加行车安全，优化智能体验，驱动智行生活。

6.2.4 知识问答

随着大数据、人工智能等技术的发展，网络信息爆炸性增长，人们很难准确、快速地获取有价值的信息。搜索引擎为人们从海量网站中迅速查找有效信息提供途径，然而其使用效果却不尽如人意。基于以上背景，问答系统应运而生。问答系统是指系统接收用户以自然语言形式描述的提问，使用数据处理、查询扩展等技术从大量异构数据源中搜索出能回答提问的精准答案的信息检索系统。

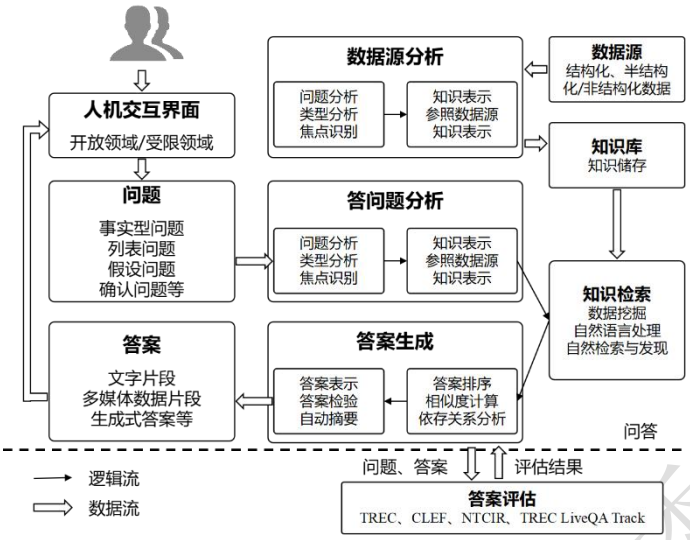

图 6-9 问答系统架构图

问答系统可以看作信息检索的一种高级形式，对用户自然语言提出的问题给予准确、简洁的回答。典型问答系统通常由数据源分析模块、问题分析模块、知识检索模块、答案生成模块和答案评估模块五个部分组成。

其中数据源分析模块负责从多种复杂数据源中使用知识挖掘技术提取有用的信息生成知识表示形式；问题分析模块负责对用户的问句进行语义处理，根据用户的查询意图生成相应的查询语句，此模块涉及到问题分类、问句相似度计算、查询扩展等技术；信息检索模块的主要功能是根据所生成的查询语句检索知识库，获取并解析答案；答案抽取模块则利用信息抽取技术从检索模块所检索到的相关答案中抽取最佳的候选答案返回给用户。答案评估模块是一个补充模块，其对上一步生成的问题答案进行评估，评估结果反馈给其他模块，进而完善和优化整个问答系统。

6.2.5 大语言模型

近年来，随着计算机技术和大数据的快速发展，深度学习在各个领域取得了显著的成果。为了提高模型的性能，研究者们不断尝试增加模型的参数数量，从而诞生了大模型这一概念。

大模型是指具有大量参数和计算资源的机器学习模型，这些模型通常在训练过程中需要大量的数据以保证模型的泛化能力，借助高性能的 GPU 或 TPU 集群提升计算能力，并且具有数百万到数十亿个参数，使得模型能够学习到更为复杂和细腻的数据特征。大模型的设计目的是为了提高模型的表示能力和性能，在处

理复杂任务时能够更好地捕捉数据中的模式和规律。

大模型的原理主要基于深度学习，通过构建具有大量参数的神经网络模型，利用大规模的数据集进行训练，从而学习到数据中的复杂模式和规律。通过不断地调整模型参数，使得模型能够在各种任务中取得最佳表现。通常说的大模型的“大”的特点体现在：参数数量庞大、训练数据量大、计算资源需求高等。

当前流行的大模型的网络架构，更多的还是沿用 NLP 领域最热门最有效的架构：Transformer 结构。相比于传统的循环神经网络（RNN）和长短时记忆网络（LSTM），Transformer 凭借其独特的注意力机制（Attention），给予模型更强的理解力，并对关键词能给予更多关注，同时该机制具有更好的并行性和扩展性，能够处理更长的序列，从而在 NLP 领域中奠定了其作为基础性与通用性模型的地位，并在各类文本相关的序列任务中取得良好的成效。

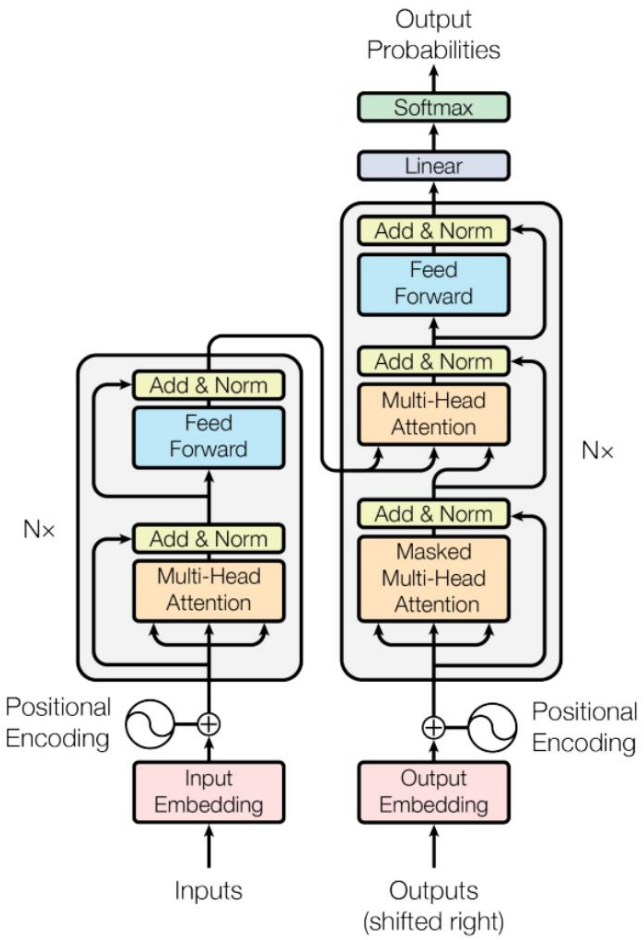

图 6-10 Transformer 结构

而“大语言模型”是大模型的一个特定子集，主要被设计用来理解和生成自

然语言。大语言模型通过在海量文本数据上进行训练，能够学习到语言的复杂结构和模式，从而实现诸如文本生成、语义理解、机器翻译、问答系统等多种功能，充分适配智能座舱内的多种自然语言相关任务。

6.3 智能座舱表达技术

6.3.1 语音合成

语音合成（Speech Synthesis），也称为文本转语音（Text-to-Speech, TTS）技术，是人机交互领域的一项非常重要的任务，旨在合成清晰且自然的音频。通过语音合成，可以解放用户的眼睛，使人能在“眼观”的同时还可以“耳听”，增加信息接收的带宽。语音合成可以分成三个层次：（1）从文字到语音的合成，即文语转换（Text-To-Speech,TTS）；（2）从概念到语音的合成（Concept-To-Speech）；（3）从意向到语音的合成（Intention-To-Speech）。

传统语音合成模型以“源滤波器”模型为基础。早在 1779 年，C.Kratzenstein 制作出了机械式语音合成器，其原理是用风箱模拟人的肺，用簧片模拟声带，并且用皮革制成共振腔模拟声道，通过改变共振腔的形状，可以合成五个长元音。

早期的语音合成方法模型简单，系统复杂。后来语音合成模型包含两个部分：前端和后端。前端主要负责在语言层、语法层、语义层对输入文本进行文本分析，后端主要是从信号处理、模式识别、机器学习等角度，在语音层面上进行韵律特征建模，声学特征建模，然后进行声学预测或者在音库中进行单元挑选，最终经过合成器或者波形拼接等方法合成语音。

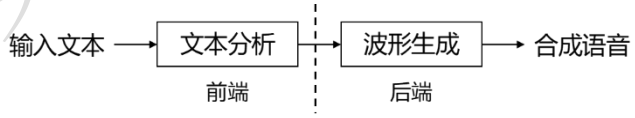

图 6-11 语音合成系统框架图

随着神经网络的快速发展，端到端的语音合成模型逐渐进入人们的视野。端到端神经语音合成模型包含两个组件：声学模型和声码器。这种模型不直接生成音频，而是先根据文字，通过声学模型生成频谱（即音频的一种高度压缩的表示方法）；然后通过一个声码器（vocoder），再将频谱转换成音频。这样做的原因是，频谱的粒度是帧级别（framelevel）的，而音频的粒度是采样点级别（samplelevel）的；一般情况下，频谱的一帧对应了 12.5ms 的音频，在采样频

率为 16kHz 的情况下，那么就有 200 个采样点，且这 200 个采样点的信息是高度冗余的；这样，生成同样一段音频，生成频谱相比直接生成音频，产生的输出长度缩小为 1/200。

虽然本文的焦点为基于神经网络的语音合成模型，但是由于联系紧密，同时结构上也有很多相似之处，仍然需要对基于统计参数的语音合成模型做一些简要的介绍。

虽然基于拼接的语音合成能够得到较为自然的音频，但是也受限于语料库的规模。基于拼接的系统只能生成那些所需音频片段已经被包含在语料库中的音频；因此如果需要生成自然的音频，需要请一个说话人录下大量的音频以供拼接挑选。这对于语音合成的灵活性（生成多位说话人的声音、多种风格、情绪等）存在着极大的限制。

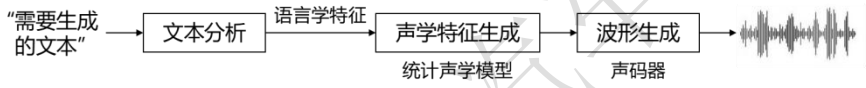

图 6-12 基于统计参数的语音合成模型

因此，基于统计参数的语音合成模型（statistical parametric speech synthesis，SPSS）成为了新一代模型。如图 6-12 就是 SPSS 的流程示意图。这种方法相比于之前的拼接法，最大的优势就是其灵活性——无需预先准备好的语音片段也可以合成需要的音频。SPSS 可以通过机器学习的方法，利用语音数据集进行训练，学习如何让模型将给定的文本输入映射为所需的音频。训练 SPSS 需要用到语音音频及其对应的文本，训练得到的结果可以看作是描述这些语音的统计特征的参数。

6.3.2 声音复刻

声音复刻，是语音合成技术（TTS ，Text To Speech）的个性化应用，通过大量的语音数据训练，使 AI 模型可以学习并精准模仿人类的语音。一方面弥补了传统语音合成技术在数字化人声上的不足，生成的语音纹理更为真实丰富；另一方面，随着算力的提升和大数据的积累，更精细的音调、音色和音量的变化皆能得到还原，让“人声”具有更自然的质感。在很大程度上解决了声音无法“手绘”的问题，为声音塑造带来了无尽的可能性。

6.3.3 基础音效

音效的产生源于物体的振动，当物体振动时，扰动周围的介质如空气，产生声波并向外传播。这些声波的频率和振幅分别决定了听到声音的音调和响度。人耳接收到这些声波后，通过听觉系统将其转化为神经信号，大脑解码这些信号，产生我们所感知的声音。声音具有音调、音量和音色三个基本特性。音调（音频）与频率相关，频率越高，音调越高；音量则与振动的幅度有关；音色则取决于波形的形状，不同物体发出的声音波形不同，因此音色也不同。

音频播放算法主要使用数字信号处理和计算机音频技术，确保音频数据能够被正确解码、处理并转换为模拟信号，从而通过扬声器或耳机播放出来。

6.3.4 独立音区

独立音区技术旨在为汽车座舱内或特定空间内为每位用户创造更加沉浸式和定向的音频体验，通过对不同的声音来源进行独立控制和定位，使每个声源（如乐器、人声、环境音效等）能够在听者周围的三维空间中被精确放置和移动并确保不同区域的音频内容互不干扰。

汽车智能座舱配备的多声道音频系统（包括多个扬声器和声道），为独立音区技术的实现提供了硬件基础。通过独立控制每个扬声器或扬声器组，可以向特定区域播放不同的音频。基于波束形成（Beamforming）原理，结合数字信号处理（DSP）技术，将声音精确地“指向”座舱中的特定区域或座位，将特定的音频流（如音乐、电影音频或导航指令）只播放给某一座位上的用户，而不会显著影响其他区域的听感。

独立音区不止包含音频的分区发出，还涉及音频的独立采集。多音频源定位技术通过声音传播的物理特性和声学原理，并利用信号处理算法，如独立成分分析（ICA）、非负矩阵分解（NMF）或深度学习模型，从混合音频信号中分离出不同的声源信号，对这些信号进行处理和分析，利用阵列信号处理技术，如波束形成（Beamforming）、广义旁瓣相消（GCC-PHAT）、多重信号分类（MUltiple SIgnal Classification，MUSIC）算法等，以推导出音频源的具体位置。让麦克风阵列锁住某个方向的声音，精确识别语音指令发出者并为其提供个性化服务。

6.3.5 主动降噪

外部嘈杂环境（包括车辆座舱）都广泛存在着各种各样的噪声，例如，高度随机的路噪（包括胎噪），相对固定的发动机噪声，高速行驶的风噪，车内的空腔共振噪声，以及空调噪声等。这些大量的噪声干扰不但影响司乘人员的音频体验，引发烦恼和疲劳，还有可能影响到安全驾驶。因此，深入分析各种车辆噪声NVH 性能（包括噪声（Noise）、振动（Vibration）和声振粗糙度（Harshness）等），广泛开展有效的降噪技术研究，实现安静舒适的声环境，已经成为智能座舱领域的关键技术之一。针对不同的实际应用场景，结合 AI 技术的智能感知和降噪处理是当前降噪研究领域的一个重要方向。

6.3.6 车内交流补偿

音频交流补偿技术主要涉及在音频通信或播放中，通过技术手段改善或校正因环境、设备特性或传输过程导致的音频质量下降，以提升听觉体验。这一技术广泛应用于多个领域，包括语音合成、实时通信、音频播放系统优化等。

（1） 跨说话人情感迁移中的韵律补偿

在端到端语音合成领域，韵律补偿策略用于解决跨说话人情感迁移时的问题。它通过一个专门的模块（Prosody Compensation Module, PCM）来增强情感表达，同时保持目标说话人的音色稳定。PCM 利用预训练的自动语音识别（ASR）模型提取的韵律信息来补充情感嵌入中因去说话人化而可能丢失的情感细节，从而在不牺牲说话人相似度的前提下提升合成语音的情感表现力。

（2） 智能音频管理系统中的声源补偿

智能汽车中的飞鱼智能音频管理系统示例说明了如何通过声源补偿技术来优化交流体验。该系统能够针对不同座位的乘客进行声源的精确拾取和播放补偿，确保声音清晰地传递到目标听众，减少交流障碍，比如前排与后排乘客之间的对话，通过技术手段优化声音的定向播放，提升车内交流的自然性和清晰度。

（3）特定频带的扬声器频响补偿

针对扬声器在特定频率响应不佳的问题，有技术通过多速率采样和滤波器优化设计，对扬声器的频响进行补偿，特别是在指定频带内，提高音频播放的质量。这意味着通过技术手段调整，可以使得扬声器在播放音乐或语音时，即使在某些频率响应不足的情况下，也能提供更加均衡和高质量的音频输出。

（4） 音频恢复技术

包括从噪声消除到音频补偿的广泛技术，旨在恢复被噪声干扰的音频信号。这项技术通过算法和模型，如降噪算法，来识别并去除背景噪声，同时尽可能保留原始音频信号的完整性。在计算机音频处理中，这涉及到复杂的数学模型和信号处理技术，以确保在各种环境噪声下都能提供清晰的音频体验。 综上所述，音频交流补偿技术是一个综合性的概念，它涵盖了从语音合成的情感保真度提升，到实际应用场景中的声源定位与补偿，再到扬声器性能优化和噪声消除等多个方面，旨在克服音频交流中的物理和环境限制，提升音频的清晰度、真实感和交互性。

第七章 智能座舱 AI 技术应用测试与评价的流程和要求

7.1 场景交互评测的流程和要求

7.1.1 语音交互场景

当前语音交互在智能座舱的测评规范主要围绕车载语音交互测评，包含语音唤醒、语音识别、语音交互等方面。具体的测评内容涉及了语音、触控以及无线交互等用户高频使用的交互方式。在第一章的法规和标准分析部分，有 7 个我国的语音交互相关应用的评价标准及规范，包括《汽车智能座舱语音分级与测评方法（起草）》、《汽车智能座舱智能水平测试与评价方法》、《智能语音交互质量评价规范》、《中国智能网联汽车技术规程》、《汽车智能座舱交互体验测试评价规程》以及《道路车辆免提通话和语音交互性能要求及试验方法》。

当前，智能座舱的语音交互测评规范已经相对完善，覆盖了识别准确性、交互流畅性、系统响应速度等多个关键测试领域，确保了用户体验的高标准。随着技术的持续进步，尤其是大语言模型的引入，语音交互系统的应用范围和深度都在不断扩大，为智能座舱带来了更加丰富和个性化的交互体验。

然而，尽管现有的测评规范在传统语音交互系统上表现出色，它们尚未充分考虑到大语言模型带来的新特性和挑战。大语言模型的集成，为智能座舱的语音交互系统赋予了更深层次的语义理解、更自然的对话能力以及更广泛的服务范围。这些新特性要求我们重新思考和构建评价规范，以适应未来智能座舱的发展趋势。为了应对这一挑战，构建面向大语言模型赋能的语音技术在座舱中的用户体验评价规范变得尤为关键。

（1） 生成能力

大模型应用于语音交互中，生成能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《语音自然度评测标准》、《措辞多样性评测标准》和《情感表达能力评测标准》，性能层面的标准为《生成流畅性评测标准》。

1）用户体验

《语音自然度评测标准》：语音自然度评测标准旨在建立一套科学、系统的评估体系，以客观、准确地衡量大模型生成语音的真实感和自然程度。它通过对

语音的韵律特征、音质表现、发音准确性、情感传递等多方面进行综合考量和分析，来判断大模型语音与人类自然语音的接近程度。

《措辞多样性评测标准》：措辞多样性评测标准旨在评估措辞多样性的标准，系统性地检验文本或言论中的表达方式是否具备丰富多样性。这一标准不仅关注于内容的准确性和完整性，更注重于表达形式的多样性，包括但不限于词汇选择、句式结构、修辞手法等方面的丰富性。其目的在于确保文本或言论在传达信息的同时，能够引起读者或听众的兴趣，激发其思考和共鸣，提升信息传达的效果和影响力。

《情感表达能力评测标准》：情感表达能力评测标准旨在评估情感表达能力的标准，旨在系统性地检验个体或文本在表达情感方面的能力。该标准注重评估个体或文本在情感传递、情感识别和情感调控等方面的表现，以确保其能够准确、清晰、生动地传达情感信息。通过该标准的应用，有助于提高个体或文本在交流、沟通和创作中的情感表达能力，促进情感共鸣和情感连接，从而增强人机关系。

2）性能

《生成流畅性评测标准》：生成流畅性评测标准从性能的角度系统性地评估生成系统或算法在输出文本流畅性方面的表现。该评测标准考量生成系统或算法在处理大规模数据时的效率、速度和准确性，以确保生成的文本能够在实时性和性能上达到要求。通过该评测标准的制定和应用，有助于推动生成技术在性能方面的进步，提高生成系统或算法的处理能力和效率，从而更好地满足用户对流畅性的需求，并推动生成技术在实际应用中的广泛应用。

（2） 学习能力

大模应用于语音交互中，学习能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《记忆更新能力评测标准》和《学习准确性评测标准》，性能层面的标准为《学习效率评测标准》。

1）用户体验

《记忆更新能力评测标准》：此标准专注于智能座舱系统对用户偏好的长期记忆、对行为模式的学习和适应，以及在新信息出现时的迅速更新和同步。它强调了系统对用户隐私的保护，确保了用户对个人信息的完全控制。同时，它也考察了技术整合性，即不同技术组件如何协同工作以支持记忆更新功能。此外，还

包括了对系统在各种使用场景下可靠性和稳定性的评估，以及用户反馈机制的效率，确保用户可以轻松地纠正系统记忆中的任何错误。

《学习准确性评测标准》：此标准着重系统对用户行为数据的精确收集，以及对这些数据进行有效分析和模式识别的能力。它们评估系统在提供个性化服务时的准确性，以及对用户行为变化的快速响应和持续优化的能力。此外，评测标准还包括对用户隐私保护和数据安全性的严格要求，确保在提升智能化水平的同时，用户的个人信息得到妥善保护。通过这些综合的评测标准，智能座舱的学习准确性得到了全面的考量，进而推动智能座舱技术向着更加智能化、个性化的方向发展。

2）性能

《学习效率评测标准》：专注于评估大模型学习过程中的响应速度和数据处理能力，确保系统能够迅速从用户的语音指令中提取信息，并及时更新其知识库。我们考察系统对用户行为模式的快速学习机制，包括对新指令的敏捷适应和对错误理解的即时修正。此外，还包括对系统在连续交互中持续优化性能的能力进行评估，确保学习过程的动态性和持续性。这些标准旨在推动智能座舱技术在语音交互方面的快速进步，实现对用户指令的即时反应和对用户习惯的准确预测，从而提供更为流畅和个性化的交互体验。

（3） 理解能力

大模应用于语音交互中，理解能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《情感理解评测标准》、《用户意图理解评测标准》和《多语言理解评测标准》，性能层面的标准为《理解效率评测标准》。

1）用户体验

《情感理解评测标准》：衡量大模型在语音交互中识别和响应用户情感的能力，这对于提供更加人性化和富有同理心的用户体验至关重要。这些标准将评估系统如何通过用户的语音语调、语速、音量以及语言表达中的微妙情感色彩来识别其情绪状态，如快乐、悲伤、愤怒或压力。此标准的核心在于系统对情感信号的敏感度和准确性，以及其基于这些情感输入做出恰当反应的能力。这包括系统能否在识别到用户情绪变化时，调整其交互策略，提供更加贴心的服务或反馈。此外，标准还将考察系统在长期交互中学习并适应用户情感模式的能力，以及其

在保护用户情感隐私方面的措施。

《用户意图理解评测标准》：旨在衡量大型模型在语音交互中理解和满足用户意图的能力。评估系统对用户语音或文本输入中意图的准确性，包括对不同意图的区分度，如询问、指示、请求等。标准应包括多意图处理，测试系统处理用户同时提出的多个意图或需求的能力，例如同时询问和请求服务。此外，标准检查系统在用户表达不清晰或意图转换时的处理能力，包括澄清问题或意图，以确保正确理解用户需求。这些标准将有助于评估大型模型在语音交互中理解和满足用户意图的能力，并提供更加智能、个性化和贴心的用户体验。

《多语言理解评测标准》：旨在评估大型模型在多语言环境下理解和处理用户意图的能力。在准确性方面，评估系统对不同语言用户意图的准确识别程度，包括语言特有的表达方式和文化背景的理解。在语言转换和适应方面，测试系统在多语言环境中进行语言转换和适应的能力，包括将用户意图从一种语言转换为另一种语言并保持意图的一致性。这些标准将有助于为多语言用户提供智能、个性化和优质的交互体验。

2）性能

《理解效率评测标准》模型的理解效率，即模型在实际应用中进行理解时的速度和资源消耗。此外，效率评测标准还应考虑模型的泛化能力，即在不同任务和数据集上的稳定性能表现，而非仅在特定条件下的优异表现。综上所述，理解和应用效率评测标准可以帮助我们全面衡量和优化大模型在实际应用中的性能表现和成本效益。

（4） 推理能力

大模应用于语音交互中，推理能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《泛化能力评测标准》、《逻辑推理能力评测标准》、《推理一致性评测标准》和《推理可解释性评测标准》，性能层面的标准为《推理效率评测标准》。

1）用户体验

《泛化能力评测标准》：大模型的泛化能力评测标准至关重要，这涉及到模型在不同数据集和任务上的稳定性能表现。一个具有良好泛化能力的大模型应该能够在训练过程中学习到普适性的特征和规律，而非仅仅在训练数据上过度拟合。

为了评估泛化能力，我们可以采用交叉验证、测试集评估、领域适应等方法。综合运用这些泛化能力评测标准可以帮助我们更全面地了解和评估大模型的实际应用价值和稳定性。

《逻辑推理能力评测标准》：逻辑推理是指模型在处理信息、推断结论或解决问题时所展现的逻辑思维和推理能力。为了评测模型的逻辑推理能力，可以采用一系列标准和方法。首先是逻辑推理任务的设计，包括逻辑规则的定义、问题形式的设定等。这些任务可以涵盖形式逻辑、命题逻辑、谓词逻辑等多个方面，从而全面评估模型的逻辑推理能力。其次是评估指标的选择，比如准确率、推理速度、逻辑一致性等。通过这些指标的量化评估，可以客观地衡量模型在逻辑推理任务上的表现。此外，还可以采用开放式问题或复杂推理场景来评估模型对于真实世界复杂逻辑问题的处理能力。综合运用这些逻辑推理能力评测标准可以帮助我们更深入地了解大模型的智能水平和推理能力，为模型的进一步优化和应用提供重要参考。

《推理一致性评测标准》：推理一致性指的是模型在处理逻辑推理或问题求解时的逻辑连贯性和稳定性。为了评估推理一致性，我们可以采用多种方法和标准。首先是逻辑规则的一致性检查，确保模型在不同推理任务或问题求解中遵循相同的逻辑规则和推理过程。其次是逻辑矛盾的检测，即检验模型在推理过程中是否产生逻辑上的矛盾或不一致性。此外，还可以考虑推理过程中的信息保持和传递性，确保模型在推理链条中信息不丢失或扭曲。综合利用这些推理一致性评测标准可以帮助我们全面评估模型的逻辑推理能力，发现并解决推理过程中可能存在的逻辑问题和不一致性，提升模型的智能水平和推理稳定性。

《推理可解释性评测标准》：推理可解释性评测标准关注模型在推理和决策过程中的解释性。评估方法包括解释性文本生成、可视化展示和推理结果一致性检查，旨在确保模型的推理过程和结果能够被人理解和信任。这些标准有助于提高模型的可接受性和可靠性，增强其在实际应用中的可解释性。

2）性能

《推理效率评测标准》：推理效率评测标准关注模型在推理过程中的速度和资源利用情况。评估推理效率可以采用多种方法，如推理时间、资源消耗、计算复杂度等指标。通过对模型在不同场景下的推理速度和资源消耗进行定量分析和

比较，可以评估模型在实际应用中的效率表现。这些评测标准有助于选择合适的部署方案、优化模型结构或采用量化技术，从而提高模型的推理效率，适应不同应用场景的需求。

7.1.2 视觉交互场景

当前视觉交互在座舱的规范及现状涉及多个方面，包括智能座舱的人机交互设计、眼动追踪技术的应用、虚拟现实技术的利用以及基于视觉感知特性的界面评价等。为了规范智能座舱的交互体验，中国汽车工业协会制定了《汽车智能座舱交互体验测试评价规程》等团体标准，这些标准对术语和定义、评价指标体系、等级划分和试验方法等方面进行了详细规定。

智能座舱视觉交互设计国家规范的制定应综合考虑上述原则和因素，以确保设计出的界面既美观又实用，同时兼顾用户的舒适度、效率和安全性。然而，当前的标准多半为设计规范层面，缺乏通用且权威的推荐设计标准，这使得当前智能座舱中的视觉交互设计随着车企而有显著的差异。随着大模型的使用场景日趋丰富，视觉交互是语音交互之外的核心交互模态。因此，为了确保智能座舱设计符合人体工程学原则，未来应该结合先进的生理监测技术进行相关的人机实验和模拟，集合多学科的专业进行全面的优化与迭代，以期能满足用户需求。

（1） 生成能力

大模型应用于视觉交互中，生成能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《生成准确性评测标准》、《个性化适应评测标准》和《多样性和创造性评测标准》，性能层面的标准为《生成流畅性评测标准》。

1）用户体验

《生成准确性评测标准》：评估大模型在座舱视觉交互中的生成准确性是确保驾驶员和乘员安全的关键因素。信息匹配度、数据处理精度和预测能力是评估生成准确性的重要指标。信息匹配度考量生成信息与实际场景或用户需求的匹配程度，包括驾驶行为分析、乘员安全情况和个性化设置等方面的匹配度。数据处理精度则关注模型对输入数据的处理精度，如人脸识别的准确率和监控系统对驾驶行为的识别准确性。预测能力则评估模型对未来情况的预测能力，如根据驾驶员行为预测可能出现的驾驶风险或根据乘员习惯预测个性化设置。这些评估标准共同确保生成的信息准确无误，提升系统的实用性和用户体验。

《个性化适应评测标准》：评价大模型在座舱视觉交互中的个性化适应能力需要考量多个方面。首先是身份识别准确性，即评估模型对驾驶员和乘员身份识别的准确性。其次是个性化设置效果，需要考察模型根据用户身份进行的个性化设置效果，如座椅、音响等参数的调整是否符合用户的偏好和习惯，从而提供更舒适和个性化的驾乘体验。最后，收集用户的反馈信息也是评价个性化适应的重要指标，通过了解用户对个性化适应效果的满意度和提出的改进意见，可以不断优化大模型的个性化适应能力，提升座舱视觉交互系统的个性化服务水平，进而提升用户的使用体验和满意度。

《多样性和创造性评测标准》：在座舱视觉交互场景中，评估大模型的多样性和创造性是提升用户体验和吸引力的关键。首先是评估模型生成内容的多样性，包括多样化的音乐、视频内容或个性化服务推荐等方面，从而丰富用户的选择和体验。其次是考察模型生成信息的创造性程度，即是否能够超越传统预期，为用户带来新颖、有趣的体验，增强用户的参与感和兴趣。

2）性能

《生成流畅性评测标准》：同语音交互的相关标准。

（2） 学习能力

大模应用于视觉交互中，学习能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《记忆更新能力评测标准》和《学习准确性评测标准》，性能层面的标准为《学习效率评测标准》。这些标准与语音交互的相关标准相同。

（3） 理解能力

大模应用于语视觉中，理解能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《视觉交互-情感理解评测标准》、《视觉交互-用户意图理解评测标准》和《个性化识别评测标准》，性能层面的标准为《理解效率评测标准》。

1）用户体验

《视觉交互-情感理解评测标准》：关注大模型在视觉交互中对驾驶员和乘员情感状态的理解能力。首先，评估情感识别准确性，包括对喜怒哀乐等情绪的识别准确率，以确保系统能准确捕捉用户的情感状态。其次，考察情感反馈效果，即模型是否能根据识别到的情感状态做出相应调整，例如调整座舱环境或提供情

感支持，以满足用户的情感需求。这些评测标准将有助于全面评估大模型在情感理解方面的表现，从而提升座舱视觉交互系统的智能化水平和用户体验。

《视觉交互-用户意图理解评测标准》：旨在评估大模型在视觉交互中对用户意图的理解能力。这包括对用户的语言、动作、表情等多种信号进行准确识别，并从中推断用户的意图和需求。评价用户意图理解可以考虑以下几个方面：首先，评估模型对用户指令的识别准确性，包括对语音指令、手势动作等的准确理解和执行；其次，考察模型对用户行为和反应的分析能力，包括对用户喜好、习惯和情感状态的识别和分析；最后，评估模型对用户隐含意图和上下文的推断能力，例如根据用户历史数据和场景信息推断用户可能的行为和需求。综合考量以上标准，可以全面评估大模型在视觉交互中对用户意图的理解能力，提升系统的智能化水平和用户交互体验。

2）性能

《理解效率评测标准》：同语音交互的相关标准。

（4） 推理能力

大模应用于视觉交互中，推理能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《泛化能力评测标准》、《逻辑推理能力评测标准》、《推理一致性评测标准》和《推理可解释性评测标准》，性能层面的标准为《推理效率评测标准》。这些标准与语音交互的相关标准相同。

7.1.3 多模交互场景

智能座舱产业已逐步融合电子和人工智能领域的先进技术，大屏化、多屏化上车的趋势明显，HUD、VR 和流媒体后视镜等显示技术快速发展。此外，智能座舱的普及率、功能性以及芯片等配置日益多样，消费者的关注度和满意度不高。

多模态交互技术在大数据与人工智能时代得到了广泛应用，其学术和技术发展前沿与图像图形学、人工智能、情感计算、生理心理评估等领域密切相关。然而，多模态人机交互领域仍面临诸多挑战，如如何攻克技术瓶颈、提高系统的准确性和用户体验等。未来智能座舱应由技术导向转为需求导向，技术选择和设计目标应聚焦场景和需求。随着大规模预训练模型、语音语言与视觉技术的日趋成熟，人机交互正在从语音交互演进至多模态交互。

智能座舱多模态交互的标准与现状表明，虽然技术不断进步，但仍然面临许

多挑战与不足，标准的构建赶不及技术的进步。未来制定标准的方向应将更加注重用户体验和实际需求，考虑从多模态交互与融合、大模型与 AI 化、软件定义汽车、车内场景化与网联化、高清大屏与虚拟化以及全员监控系统等应用推动智能座舱向更高水平的智能化迈进。

（1） 生成能力

大模应用于多模态交互中，生成能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《多模态响应生成评测标准》，性能层面的标准为《响应速度与准确性评测标准》。

1）用户体验

《多模态响应生成评测标准》：要求系统能够根据用户的语音、视觉和触觉指令，生成准确且及时的响应。系统应提供自然、直观的交互方式，减少用户的学习成本，并能够根据用户的个性化需求定制响应内容。

2）性能

《响应速度与准确性评测标准》则侧重于系统处理多模态输入的速度和执行任务的准确度，确保系统能够在极短时间内理解并执行用户的指令。

（2） 上下文连贯能力

大模应用于多模态交互中，上下文连贯能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《上下文感知与应用评测标准》，性能层面的标准为《响应速度与准确性评测标准》。

1）用户体验

《上下文感知与应用评测标准》强调系统在多模态交互中保持上下文连贯性的能力，以提供一致且流畅的用户体验。系统应能够记忆用户的操作历史，并根据当前上下文环境，提供恰当的交互响应。…

2）性能

《上下文切换平滑度标准》则评价系统在不同上下文间切换时的稳定性和平滑度，确保用户在不同场景下都能获得无间断的交互体验。

（3） 学习能力

用户体验：《自适应学习与优化标准》衡量系统根据用户行为和反馈进行自我学习和调整的能力，以提升用户的满意度和忠诚度。系统应能够识别用户的偏

好，并主动适应用户的变化。

系统性能：《学习效率与效果标准》则关注系统学习新行为模式的速度和学习后性能提升的幅度，确保系统能够快速吸收新知识并有效应用。

（4） 理解能力

用户体验：《多模态意图理解标准》要求系统能够准确识别和理解用户的多模态指令背后的意图，即使在用户指令不明确或含糊时也能提供有效的帮助。

系统性能：《指令解析精准度标准》则侧重于系统对用户指令的解析精准度，包括对复杂指令的拆分和执行，确保系统能够正确理解并执行用户的每一个指令。

（5） 推理能力

用户体验：《决策推理与异常处理标准》评估系统在面对复杂情境时的决策推理能力，以及在遇到异常情况时的应对策略，确保用户在面对不确定性时仍能获得可靠的支持。

系统性能：《逻辑推理准确性标准》则侧重于系统逻辑推理的准确性和效率，确保系统能够基于正确的推理提供恰当的决策。

7.1.4 开放式任务场景

智能座舱开放式任务的标准与现状正在逐步完善和发展。虽然目前尚无统一的国际标准，但国内已经发布了多项标准和研究报告，推动了智能座舱技术的发展和应用。同时，多个企业也在积极开发开放平台，以适应市场需求和技术进步。2024 年 1 月，由全国汽车标准化技术委员会智能网联汽车分标委组织，中汽中心（CATARC）、东风汽车、德赛西威牵头编制的《智能座舱标准体系研究报告》发布，首次向智能网联汽车产业和智能座舱发展提供了系统化的支撑体系。然而，开放式任务是大模型赋能后，集合模型感知与信息认知的人机交互场景，相关的标准应尽早启动并构建，以完善整个智能座舱产业的规划及打造用户体验的全流程服务。

（1） 生成能力

大模型应用于开放式任务交互中，生成能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《任务整合评测标准》、《个性化内容评测标准》，性能层面的标准为《复杂指令响应能力评测标准》。

1）用户体验

《任务整合评测标准》：该标准的主旨在于系统对用户需求的深入理解，包括对指令背后意图的洞察，以及对不同垂域任务间关联性的把握。它强调了系统在面对用户未明确表达需求时，能够通过上下文信息和先前交互历史来预测并主动提供帮助，例如在用户未意识到的情况下提醒重要的纪念日；在特定乘客上车时主动调整车内氛围及设置。此外，此标准还应包括对系统生成任务解决方案的评估，即创造性和实用性等方面，确保解决方案不仅技术上可行，而且能够满足用户的个性化需求。同时，这些标准也考虑了系统的适应性和灵活性，即在面对用户指令变化或新增时，系统能否快速调整任务计划，以适应新的执行要求。

《个性化内容评测标准》：该标准的主旨在于评估系统如何深入理解用户的个性化需求和偏好，并将这些理解融入到任务生成和执行中。它强调系统在处理多指令任务时，不仅要准确执行每个指令，还要能够根据用户的个人历史和上下文信息，创造性地整合这些指令，提供一站式的个性化服务体验。

2）性能

《复杂指令响应能力评测标准》：该标准的主旨在于系统对复杂指令的解析精度，包括对条件语句、时间表达、序列动作和用户偏好的准确捕捉。它强调了系统在面对用户潜在需求时的预测能力和主动性。此标准还涉及系统在整合不同来源信息、处理并发任务和维护任务执行连贯性方面的能力。此外，标准也考察了系统在用户指令不明确或含糊时的澄清策略，以及在执行过程中对用户反馈的响应速度和调整能力。通过这些标准，智能座舱的大模型技术将能够更加精准地响应用户的复杂指令，提供更加流畅和高效的服务体验。

（2） 上下文连贯能力

大模应用于开放式任务交互中，上下文连贯能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《环境信息识别评测标准》和《情境适应性评测标准》，性能层面的标准为《上下文连贯流畅性评测标准》。

1）用户体验

《环境信息识别评测标准》：旨在评估大模型在开放式任务中对环境信息的识别和利用能力，这对于提供情境相关的服务至关重要。通过衡量系统如何准确捕捉并解析外部环境信息，包括但不限于时间、地点、气候、交通状况以及用户的活动模式。它强调系统在整合这些信息以响应复杂用户指令时的效率和准确性，

例如，系统不仅识别出用户的指令，还要能够根据当前的环境信息来优化任务执行的时机和方式。

《情境适应性评测标准》：旨在评估系统对用户所处的具体情境的敏感度，包括用户的位置、时间、活动状态以及可能的情感需求。当系统在面对用户复杂指令时，如何灵活调整其行为以适应当前的情境，例如，当用户在驾驶过程中提出一系列需求时，系统能够根据实时交通状况、用户的日程安排和个人偏好，智能地安排任务执行的顺序和方式。

2）性能

《上下文连贯流畅性评测标准》：同语音部分的标准。

（3） 上下文连贯能力

大模应用于开放式任务交互中，上下文连贯能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《行为模式识别评测标准》和《学习准确性评测标准》，性能层面的标准为《学习效率评测标准》。

1）用户体验

《记忆更新能力评测标准》：同语音部分的标准。

《学习准确性评测标准》：同语音部分的标准。

2）性能

《学习准确性评测标准》：同语音部分的标准。

（4） 理解能力

大模应用于开放式任务交互中，上下文连贯能力方面的标准建立应在用户体验层面和性能层面。用户体验层面包含《用户意图理解评测标准》和《多指令理解评测标准》，性能层面的标准为《理解效率评测标准》。

1）用户体验

《用户意图理解评测标准》：同语音部分的标准。

《学习准确性评测标准》：同语音部分的标准。

2）性能

《理解效率评测标准》：同语音部分的标准。

（5） 推理能力

大模应用于开放式任务交互中，上下文连贯能力方面的标准建立应在用户体

验层面和性能层面。用户体验层面包含《因果关系推理评测标准》和《决策制定逻辑评测标准》、《异常情况处理评测标准》，性能层面的标准为《学习效率评测标准》。

1）用户体验

《因果关系推理评测标准》：评估大模型如何识别和推理用户指令背后的因果逻辑。此标准旨在衡量系统对用户行为和需求之间因果联系的理解深度，以及它如何利用这些理解来提供更加精准和前瞻性的服务。主要评测系统对用户指令背后原因的洞察力，包括用户提出某项需求的潜在原因和期望的结果。它强调了系统在处理任务时，不仅要考虑当前的指令，还要预测相关行为可能引发的连锁反应，从而为用户提供更全面的解决方案。例如，当用户要求调整车内温度时，系统不仅要执行这一操作，还要考虑到这可能对乘客的舒适度和健康产生的影响。此外，此标准还应考虑系统在面对用户未明确表达的需求时，如何通过因果推理来提供额外帮助的能力进行评估。

《决策制定逻辑评测标准》：此标准的主旨在于系统决策的透明度和合理性，确保用户能够理解系统做出的决策背后的原因。它强调系统在制定决策时，需要综合考虑多个因素，包括但不限于用户的安全、舒适、效率和个性化偏好。例如，当用户提出一系列可能相互冲突的指令时，系统需要能够根据优先级和潜在影响来做出明智的决策。

《异常情况处理评测标准》：此标准的主旨在于系统对异常情况的识别速度和准确度，包括对用户指令中的错误、环境变化导致的服务中断或系统内部故障的快速响应。它强调了系统在面对上述异常情况时的鲁棒性，以及能够提供替代解决方案或恢复计划的能力。例如，当系统检测到用户提出的任务因外部因素无法完成时，应能够及时通知用户并建议可行的备选方案。此标准还包括对系统在用户指令不明确或存在歧义时的澄清策略进行评估，以及系统如何利用历史数据和模式识别来预测和防范潜在的异常情况。此外，由于座舱空间的有限性及特殊性，标准中还应考察了系统在处理用户紧急请求时的优先级排序和快速响应机制，确保关键任务能够得到及时处理。

2）性能

《推理效率评测标准》：同语音部分的标准。

7.2 内容安全评测的流程和要求

国内外对生成式人工智能的关注日益增加，各大企业陆续在各国和地区正在积极探索制定相关的标准和法规，以应对这一新兴技术的快速发展和广泛应用。

欧盟《人工智能法案（AI Act）》是欧盟针对人工智能技术发展和应用提出的全面法规框架。根据 AI 系统对用户构成的风险，将 AI 系统进行分类，并根据风险等级决定监管的严格程度。美国则通过一系列行政命令和政策文件，强调了在促进人工智能技术发展的同时，需要考虑其对社会的影响，并在必要时制定相应的治理方法。

在中国，国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局公布了《生成式人工智能服务管理暂行办法》，该办法自 2023 年 8 月 15 日起施行。该办法旨在促进生成式人工智能的健康发展和规范应用，同时维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益。《办法》提出了国家坚持发展和安全并重、促进创新和依法治理相结合的原则，并采取有效措施鼓励生成式人工智能的创新发展。此外，还明确了技术发展与治理的相关鼓励措施，以及对服务提供者和使用者的基本要求。

在技术标准方面，中国也正在积极推进生成式人工智能服务安全基本要求的制定。例如，国家标准计划《网络安全技术 生成式人工智能服务安全基本要求》由全国网络安全标准化技术委员会（TC260）归口管理，主管部门为国家标准化管理委员会。

7.2.1 现有内容安全场景评测方案分析

近年来，生成式人工智能技术发展迅速，生成能力覆盖文本、音频、视频等多种模态，相关产品和应用不断涌现。针对生成内容安全的评测，在 CCSA TC 1 WG 1 中已有生成式人工智能技术及产品系列标准。技术能力标准中规定了生成式 AI 技术及基于技术形成的相关工具、平台的评估指标，从主观和客观两个维度进行评测。能力标准中规定了具备生成式 AI 技术能力的互联网系统、网页、平台等的评价指标，从基础能力、服务能力和可控能力等三个维度进行评测。这些在做的标准有助于降低生成结果不可控、版权争议、恶意滥用等风险，促进生成式人工智能技术的健康发展，为各行业组织制定本领域的内容安全测评方案提

供参考。

7.2.2 内容生成安全评测

随着生成式人工智能（AIGC）和大模型技术的普及，未来的智能座舱将可能集成更高级的内容生成应用。例如，智能座舱能够根据驾驶员和乘客的偏好、行为模式生成定制化的内容推荐，或根据车内环境和驾驶状况提供即时提醒与警示信息。这些新功能虽然提升了用户体验，但也带来了新的安全风险，需要进行针对性的安全评测。

在目前出台的指导文件中，网信办《生成式人工智能服务管理办法》和信安标委的《生成式人工智能服务安全基本要求》对生成式人工智能具体的安全风险进行了定义和分类划分。这些指导原则有助于制定具体的安全评测标准和方法，以确保评测内容全面覆盖各类潜在风险，提升评测的针对性和有效性。当前的内容安全测评主要由高校和研究机构主导。例如，清华大学的 CoAI 小组研发了安全大模型测评系统，覆盖了仇恨言论、偏见歧视、犯罪违法、隐私泄露、伦理道德等八大核心类别，并细化为四十余个二级安全类别。上海人工智能实验室提出了 SALAD-Bench 大模型安全基准测试框架，旨在提供高效且经济的评测手段，该框架通过三级类别层次分类结构和问题增强过程提升测试难度，并利用大型语言模型的指令跟随能力提供稳定、可复现的评估方法。中国信息通信研究院发布的 AI Safety Bench 涵盖了内容安全、数据安全和科技伦理等三大测试维度，并进一步细分了 20 余个细粒度的测评类别，使用 Responsibility Score(负责度评分)和 Safety Score(安全评分)两个指标进行评价。

7.2.3 内容拒答安全评测

“内容拒答”指的是大型语言模型在面对无法回答或不适宜回答的问题时，能够正确判断并拒绝提供回答或给出无意义、不相关的响应。这一能力对于保障模型的安全性和可靠性具有重要意义。实现内容拒答通常依赖于模型对上下文的深度理解和分析能力，在此基础上提供一种拒绝回答的机制。评估模型的内容拒答能力时，通常会设计一系列无法回答或不适宜回答的问题作为测试数据集，观察模型对这些问题的响应。通过这种方式，可以评估模型在面对不确定或缺乏足够信息时，是否能够做出正确判断并拒绝回答，从而确保其在实际应用中的可靠

性和准确性。

评测方面，中国科学院软件研究所中文信息处理实验室提出重点分析大型语言模型在 RAG（检索增强生成）所需的噪声鲁棒性、拒答、信息整合和反事实鲁棒性四项能力，要求模型对无法回答的问题进行拒答，并将拒答比例作为评价指标。清华大学的 SuperBench 测评平台也将拒答行为纳入安全测评框架中，为了精准评估模型在拒答情况下的性能，基准特别设置了拒答分数和非拒答分数：拒答分数将拒答题目视为回答错误，反映模型在面对不确定性时的保守态度或知识盲区；非拒答分数则排除拒答题目，更纯粹地评估模型在已作答题目上的表现。