家庭大脑白皮书

大模型时代智慧家庭应用新范式

（2024年）

海尔智家 Haier smart home

数字家庭网络国家工程研究中心 National Engineering Research Center of Digital Home Networking

国家高端智能化家用电器创新中心

CHEARI 中国家用电器研究院

中国质量认证中心 CHINA QUALITY CERTIFICATION CENTRE

西安交通大学 XI'AN JIAOTONG UNIVERSITY

北京郵電大學 Beijing University of Posts and Telecommunications

青岛科技大学 Qingdao University of Science & Technology

HoT 和而泰

Microsoft

科大讯飞 iFLYTEK

喜马拉雅

vivo

360视觉云

序 言

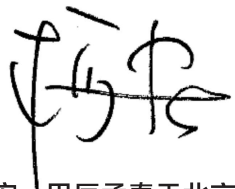
在当前数字经济浪潮下，以生成式人工智能（Artificial Intelligence Generated Content，AIGC）为代表的通用人工智能（Artificial General Intelligence，AGI）技术正在激发全球范围内的科技革命和产业变革。大模型实现了认知智能的技术跃迁，其带来的能力提升和智慧涌现正在向众多垂直领域扩散，更是为智慧家庭行业带来了重要的发展契机。

尽管通用大模型作为技术基础在多个领域表现优异，但由于缺乏专业知识与行业数据，现有通用大模型并不能精准解决某个行业或企业的特定需求和问题，因此很难直接应用于特定行业领域。为此，结合大模型的通用能力和行业的个性化需求，构建高精准、高可靠的垂域大模型成为必然选择。

通过行业知识的积累和专业人员的不断精调，垂域大模型朝着“专业、精细”方向发展，为特定场景提供更精确、更具业务价值的服务，加速行业智能化转型升级。通用大模型与行业垂域大模型交替演进的方式可以有效平衡大模型训练投入成本和边际效益。

在智慧家庭行业，以HomeGPT为代表的智慧家庭垂域大模型，已经率先探索实践，推动了整个行业的发展。HomeGPT不仅继承了通用大模型的自然语言处理、文字处理、图像处理等基础能力，还进一步研发了深度语义理解技术，进行了亿级家庭知识增强训练，并开发了行业首个场景生成引擎。智慧家庭垂域大模型，强调与硬件产品的结合和场景的联动，使场景定制和智慧家庭应用更加智能、普及，开启了智慧家庭AI应用的新时代。

未来，随着AI垂域大模型的落地，将带来智能化生产力的重构，推动智慧家庭行业在新赛道上创新与升级，为消费者提供更智能、便捷、个性化的产品和服务。



梅宏，甲辰孟春于北京

愿景与回顾

海尔智家大脑的愿景是：让冰冷的“House”成为温馨舒适的“Home”，让家变得更智慧、更温馨，时刻主动关怀各家庭成员的不同需求，使房子从“越住越老”进化为“越住越聪明”。

海尔智家给家装上一颗大脑，让体验更有深度。满足的不是单一局部需求，而是复杂场景的多任务需求；不是实现单个产品的功能，而是实现产品之间跨空间、跨系统、跨设备的联动交互。通过跨知识领域智能决策，在衣食住娱各个领域打造更丰富、更有深度的场景体验，让智慧生活的感受大有不同。

回顾以往，海尔智家大脑与行业各领域专家、学者共同探讨技术的发展与应用。2022年发布的《家庭大脑白皮书》1.0，为智慧家庭技术构建和生态平台的搭建提供方向；2023年发布的《家庭大脑白皮书》2.0，为智慧家庭空间计算以及大模型与智慧家庭交互的应用提供方向。我们希望可以与行业一起携手，持续推动智慧家庭行业的能力建设与探索，推进行业健康、快速、可持续地发展。

2023年以来，智慧家庭见证了AI和大模型技术的巨大突破，技术和应用的进步为行业带来了新的发展动力。在大模型技术落地应用到智慧家庭体验方面，海尔智家进行了从“0到1”的探索与实践。Haier HomeGPT，率先将大模型能力落地应用于智慧家庭，通过独有的深度语义理解技术、亿级家庭知识增强、行业首个场景生成引擎三大技术优势，全面推动了行业交互体验、智慧场景能力、生活服务能力全面进阶。

为推动智慧生活品质持续升级，海尔智家联合中国家用电器研究院、数字家庭网络国家工程研究中心、国创中心、中国质量认证中心、西安交通大学、北京邮电大学、青岛科技大学、微软中国、科大讯飞、和而泰、喜马拉雅、vivo、360视觉云等行业优秀高校、单位、企业共同撰写《家庭大脑白皮书（2024）》，旨在通过解析大模型时代下智慧家庭行业发展趋势、技术路线、场景创新，搭建开放生态，为智慧家庭服务企业的创新发展提供新范式。

邓邱伟

海尔智家副总裁

全屋智慧总经理

目 录

01 趋势篇

大模型推动智慧家庭产业快速进入“L4主动智能”发展阶段

自家庭大脑出现以来，随着人工智能、云计算、5G 通信等技术的逐渐成熟，智慧家庭行业迎来了全面爆发期。智慧家庭产品、功能、场景、服务等方面都出现了许多新的方向、趋势和局面，为行业的发展带来了新的机遇和挑战。

——产品方面，智慧家庭产品种类和数量都大幅增加，产品不仅具有基本的智能化功能，如远程控制、语音控制、定时控制、情景控制等，还具备情绪识别、语义理解、自学习、自适应、自优化等更高级的智能化功能，能够更好地满足用户的个性化、多样化和动态化的需求。

——功能方面，全面实现了设备之间的协同和联动，形成了更加丰富和复杂的智能场景。特别是，这些智能场景不仅能够根据用户的预设条件和指令自动触发，还能够根据用户的环境、行为、情绪、偏好等实时感知和理解，实现更加主动、智慧、深度的交互和服务。

——场景方面，相关场景已不仅局限于家庭内部，而是普遍扩展到了家庭外部，形成了更加广泛和多元的智慧场景，如智慧出行、智慧社区、智慧城市等。这些场景不仅能够实现家庭内外的数据和服务的互联互通，还能够实现家庭与其他智慧主体的协作和共享，实现更加开放、共赢和创新的生态价值。

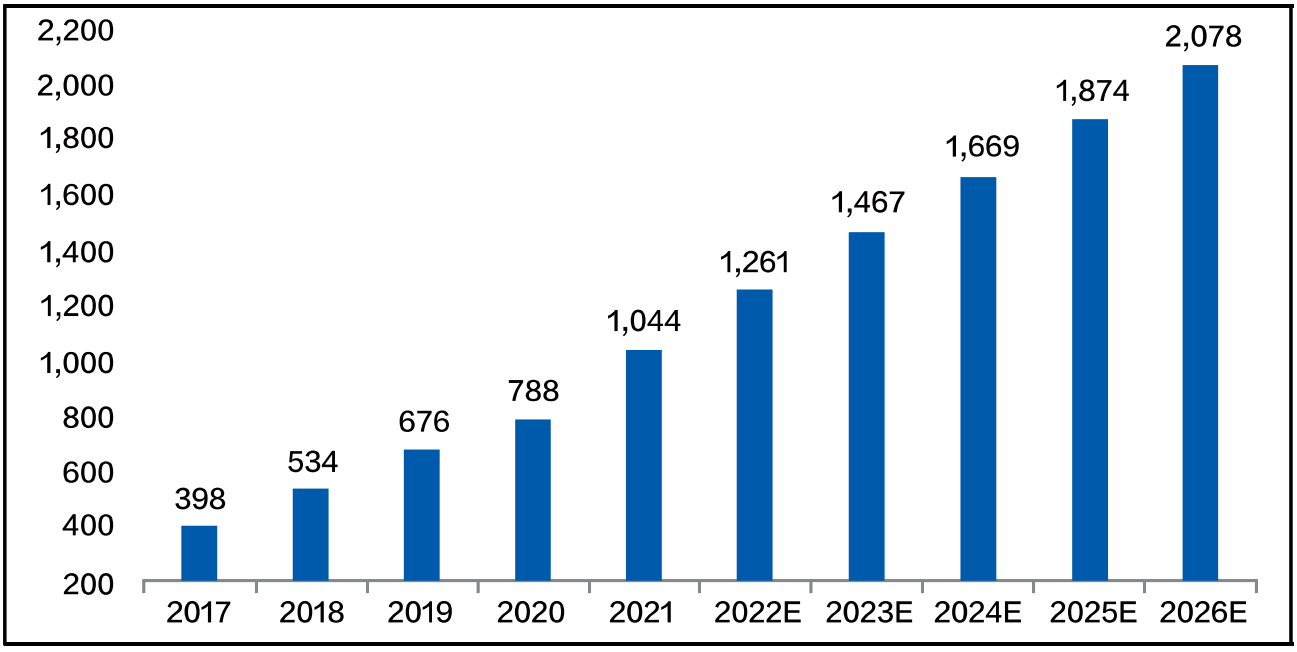
——服务方面，内容推荐、教育培训、健康管理、娱乐游戏、社交互动、生活助理、智能顾问等丰富和高质的增值服务层出不穷，不仅提高了用户的生活品质和幸福感，还同时提升用户的忠诚度和黏性，为智慧家庭的商业模式和盈利模式提供了新的思路和机会。

在这其中，家庭大脑逐步占据了智慧家庭的核心地位，也成为实现家庭智能化、信息化、网络化的关键要素之一。

1.1.大语言模型开启智慧家庭AI应用的全新时代

大语言模型加速智慧家庭产业发展

根据 Statista 的统计，2023 年全球智慧家庭智能家居市场已经突破 1000 亿美元，预计 2024 年将达到 1600 亿美元，2026 年将达到 2,078 亿美元。从发展趋势来看，智慧家庭产业发展空间广阔。



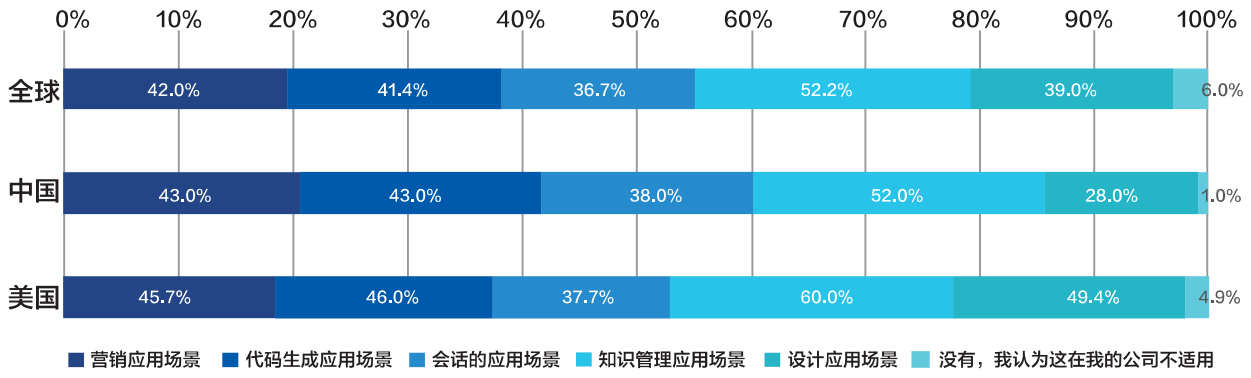
数据来源：Statista

图1 全球智慧家庭市场发展趋势（单位：亿美元）

ChatGPT 的问世，掀起了生成式人工智能的 AI 浪潮，随着知识“涌现”现象的出现，国际、国内出现了非常多的大语言模型（Large Language Model，LLM），包括 ChatGPT、文心一言、星火等众多大模型，以及 Runway、Sora 等文生视频大模型，为用户、开发者、企业打开了新的生活方式。在智慧家庭产业，以 LLM 为代表的 AI 技术也正在加速智慧家庭产品和场景的变革和创新。这些场景的成功落地，不仅在为用户创造更健康、安全、舒适、便捷、低碳的个性化家庭生活环境，也提高了广大用户的消费意愿，为经济发展注入新的动力，进一步提高了全球范围内的智慧家庭渗透率，加速了整个行业的发展。

智慧家庭 AI 应用的全新时代

据 IDC 一项针对全球企业的生成式人工智能调研结果显示，知识管理场景是 AIGC 现在最受组织青睐的应用场景，在数字人、智能对话、推荐等场景中也表现出巨大的潜力。在智慧家庭行业，AI 应用不仅需要更精准、更快速理解家庭的复杂场景，还需要实现更自然的交互，可以提供涵盖衣食住娱等方面的知识解答。可以说，智慧家庭 AI 应用是涵盖知识管理、数字人、智能对话、推荐和设计等多种场景的深度融合。



数据来源:IDC，2023，《2024年AIGC应用层十大趋势白皮书》

图 2 最有希望被企业采用的AIGC应用场景

智慧家庭行业的大模型，更强调与硬件产品的结合和场景的联动，通过重构原有 AI 底层技术，为用户提供更安全、更便捷、更健康与舒适的全新智能化体验，特别是对于用户自然语言的理解、情绪的感知与互动上，实现了颠覆式成长，使得场景定制和智慧家庭应用变得更加智能与普及，开启了智慧家庭 AI 应用的新时代。

在此背景下，海尔智家持续迭代发布了 HomeGPT 家庭垂域大模型，并在家庭大脑场景中得到全面、深度的有效应用。借助海尔近 40 年的高质量领域多模态知识积累，极大提升了语音多模态交互的效果和体验，提供了丰富的图像交互内容和形式，拓展了多模态人机交互的功能和场景，使智慧家庭具备了创造内容与生成场景的能力。

1.2.智慧家庭面临的新挑战

➢ 高质量行业数据的缺乏

智慧家庭的场景服务是多元化的，通常覆盖用户的衣、食、住、行、娱乐、健康等领域，而这些领域缺少大规模公开数据集，互联网相关数据来源分散、结构差异大、质量参差不齐，主要呈现出如下特点：

——数据来源分散： 智慧家庭的数据来源涵盖范围广，不仅涉及衣、食、住、行、娱乐、健康等领域，还包括电商平台、医学网站、娱乐新闻、旅游网站、百科，以及高价值的企业数据和开放知识（如ConceptNet、WordNet）等。

——数据结构多样： 包括结构化、半结构化和非结构化类型数据，不仅包括结构化的设备运行状态数据，还包括半结构化的用户画像、客服工单、电商评论等，以及非结构化的家用电器使用说明书等，其中包含组装结构图（图片类）、功能说明文本（文字类）、表格（图表类）等信息。

——质量参差不齐： 智慧家庭领域在数据质量方面普遍存在一定的问题，例如：家庭产品、用户行为、社区服务、政务服务等多个行业、领域、层次的数据，普遍存在包括数据噪声、数据缺失、数据不平衡、任务无关、冗余过时等问题。

——数据生态不健全： 智慧家庭领域各品牌方私有数据质量高，但获取成本也较高，数据标准难统一，数据流通规则和不同品牌方数据对接机制未能建立，智慧家庭产业尚未形成高效完整的数据产品生态体系。

——数据治理不完善： 智慧家庭领域设备的低激活率、前端数据采集困难等因素造成了家电领域的数据积累碎片化；另外，智慧家庭应用交互能力参差不齐，设备、用户、客服等各类数据管控不一造成了数据的无效沉积，限制了生成式人工智能在家庭场景的落地应用，制约了行业快速发展。

智慧家庭领域数据的这些特点，导致高质量行业数据缺乏、积累速度慢等问题。而近些年来，通用人工智能效果与泛化能力的突破，依赖于大模型在大规模、高质量、多样化数据集的训练，其来源主要包括公开数据集、大规模网络数据以及数据众包方法收集的数据等，但当通用大模型应用于垂直领域时，由于高质量垂域数据集的缺乏，往往导致认知不足、捏造事实（AI 幻觉）等问题，从而限制了通用大模型在垂直领域的直接使用。

碎片化数据和知识的科学治理问题

在智慧家庭领域，行业的碎片化问题是一个不容小觑的挑战。碎片化问题主要体现在四个方面，即数据碎片化、功能碎片化、应用场景碎片化以及相关设备的碎片化。

——数据碎片化： 随着各大企业和研究机构对物联网、人工智能领域持续投入，不断接入的设备和应用产生了大量的数据，然而这些数据却没有被充分利用。由于各个平台、厂商之间无法实现数据共享与互通，无法整合处理这些海量数据，复杂混乱的数据环境导致了数据价值无法被充分挖掘。

——功能碎片化： 目前各家智慧家庭产品功能不一、协议各异，用户需要单独对每个产品进行设置和操作，无法构建对用户友好的智能环境。例如，家中亮度调整、温度管理等基本功能已经可以通过智慧家庭产品实现自动调节，但用户仍需要通过多个不同应用软件进行控制。

——应用场景碎片化： 智慧家庭产品众多，但其应用场景却常常受到限制。因为每一种产品都需要单独安装、使用，不能精准识别用户需求、实现场景拉通并满足用户的特定场景需求，存在产品自由组合及模块通用性难题。

——智能终端产品碎片化： 大多数智慧家庭设备仍然局限于单品牌或者单一设备垂直生态链内。品类不同、形态不同、芯片不同等各类碎片化智能终端，由于存在标准与技术规格不统一的问题，限制了设备扩展性与互动性。

数据安全、隐私与合规等问题

大语言模型的训练和应用需要大量的数据，这就要求数据的质量和安全要有保障。数据质量要求数据的准确性、完整性、一致性、时效性等，数据安全要求数据的严格保密、多层授权、多种加密等安全措施。如果数据的质量和安全出现问题，可能会导致大语言模型的性能下降，甚至产生错误或危害的输出。例如，如果数据中存在错误、偏见、敏感等信息，可能会影响大语言模型的生成效果，或者引发用户的不满或抗议。智慧家庭设备经用户授权后会收集用户的个人信息，包括位置信息、行为习惯、家庭成员信息等，并根据需要在智慧家庭设备之间相互传输数据，还可能将收集到的数据存储在本地或云端。这些数据如果存储不当、被窃取破坏或截获篡改，或被泄露滥用，可能会造成严重的后果。

总体来说，这三条主线交织构成了行业整体发展面临的困难与局限性。如果想要从混乱中建立起秩序，打造出为用户提供真正帮助、便利服务和可完全依赖的高效系统，需要在规范推广、系统架构设计、数据处理、用户体验优化等方面进行深度整合与拓展创新。

1.3.垂域大模型成为生产力工具

在全球 AIGC 产业生态迅速形成与发展的背景下，领域特定的大模型技术已经成为推动生产力革新的关键力量。全球领域各类垂直大模型（法律、设计、客服、代码开发、医药研发等）陆续落地，国内各类垂直大模型（电力、海洋、中医等）的应用数量逐步增长。我们正式步入了一个“模型即服务”（Model-as-a-Service，MaaS）的新时代，这个时代的特征是通过专业化的领域模型来提供深度定制化服务。

垂域大模型专注于特定领域的知识和技能，满足行业纵深需求

垂域大模型是基于领域特有数据，经过精心设计、训练与优化，以服务于特定行业或领域需求的大模型。与通用大模型相比，垂域大模型在特定知识和技能领域展现出更深的专业性和适应性，它们利用先进的深度学习技术进行预训练和微调，从而能够精准理解和处理行业特有的语境、术语和问题，并提供高度定制化的解决方案。同时，垂域大模型的发展致力于满足行业特定需求，提升模型在特定任务上的执行效率、降低运行成本，并加速其在实际应用中的部署。

垂域大模型的发展旨在满足特定行业或领域的需求，提高模型在特定任务上的性能，并加速模型的落地应用。这种模型的兴起使得人工智能技术能更好地服务于特定行业，促使垂域内的智能化发展。

垂域大模型保障数据准确性、隐私保护和知识沉淀

通用大模型采用公共数据进行模型预训练，对于企业或者行业知识的理解和生成存在以下问题：

——数据隐私问题： 以ChatGPT为代表的通用大语言模型，多采用SSE（Server-Sent Events，服务器推送事件）方式提供服务，不支持私有化部署。而企业数据或者行业知识很多属于企业核心数据，可能包含用户和公司的私有信息，直接与通用大模型交互，存在隐私泄露、信息滥用等风险。

——训练成本高昂问题： 通用大语言模型通常包含数百亿、数千亿（甚至更多）的参数，其预训练过程依赖于大规模数据集和大量计算资源，这导致了高昂的训练成本和能源消耗。虽然目前主流的通用大模型都提供了微调模型，但是每次训练的成本都很高。

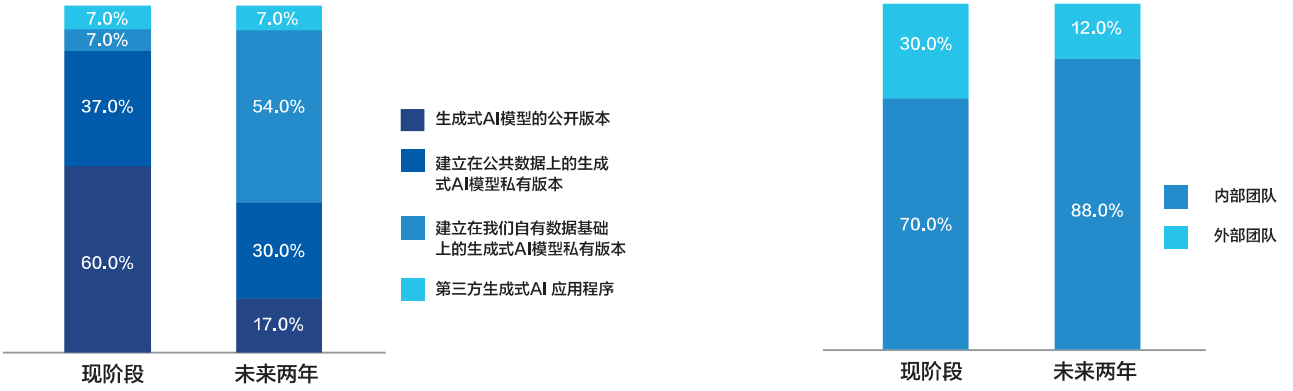
——捏造事实（AI幻觉）问题： 通用大模型对垂直领域知识处理能力相对有限，特别是缺少对垂直领域专业术语和知识的准确认知，导致生成不完整或不准确的答案，会对用户的信任度、决策质量和行业的知识传播产生负面影响。

——知识难沉淀问题： 通用大模型采用实时生成模式，由于这些模型需要处理广泛和多样化的话题，需要实时地吸收最新的数据和信息以保持知识的更新。但是这种持续更新的特性也可能导致一个问题：对于同一个已经确定的问题，模型在不同的时间可能会给出不同的答案，很难保障一致性。

为了解决以上问题，垂域大模型成为典型的解决方案之一。垂域大模型是基于主流的大语言技术框架，运用企业或者行业开发者使用的自有数据集，进行模型训练微调和调优。同时结合企业的信息检索和强化学习、隐私保护等技术手段，并通过私有化部署方式，保障数据准确性、隐私保护和知识沉淀。

垂域大模型加速数字化转型效率，释放数据价值

大模型的未来发展趋势是通用化与专用化并行。IDC 的调研显示：目前有 60% 的企业使用大模型的公开版本，但这一比例在两年后会迅速降至 17%，更多企业会将 AI 应用建立在私有、专属模型基础上；同时，高达 88% 的企业选择通过内部团队开发相关应用。由此可见，垂直领域行业大模型已经成为未来的热点。



数据来源:IDC，2023，《2024年AIGC应用层十大趋势白皮书》

图 3 AIGC模型类型和工作团队的现状及趋势

在腾讯云和中国信息通信研究院联合发布的《2023 年行业大模型标准体系及能力架构研究报告》中，提出了行业大模型构建路线图及相应的标准体系，围绕业务需求分析与资源评估、行业数据与大模型共建、行业大模型微调与优化部署等关键环节，为垂直领域行业大模型构建的标准化流程提供了建议，以更专业、成本更低的方式引导行业和企业建立特有的大模型服务。未来，通过领域化适配，垂域大模型的构建将加速行业的数字化转型效率，从数据的采集、治理到形成领域知识后的问答及应用，都将变得更加高效。在特定场景，垂域大模型将提供更精确、更具业务价值的服务，实现数据价值进一步挖掘和释放，推动行业新商业模式和服务的探索创新。例如，在医疗行业领域，垂域大模型可以帮助研发新药，加速临床试验的数据分析过程；在金融领域，可以精准预测市场趋势，提供个性化的投资建议。这种创新不仅为企业带来新的收入增长点，也为整个行业的发展注入新的活力。

数据来源:《2023年行业大模型标准体系及能力架构研究报告》，腾讯云计算(北京)有限责任公司和中国信息通信研究院云计算与大数据研究所

图 4 垂直领域行业大模型构建路线图

在中短期内，垂域大模型依然是通用大模型无法取代的技术方案。其在高精度、专业化、数据安全、隐私保护等方面的优势使得其成为特定行业内必不可少的智能化解决方案。这种专业化的能力，加上对敏感数据的保护和对隐私的重视，保证了企业和机构能够在遵守法律法规的同时，有效地利用 AI 技术推进其业务和研究。因此，垂域大模型不仅是推动特定行业技术创新的关键，也是实现行业智能化转型的基石。

1.4.智慧家庭垂域大模型的尝试

智慧家庭垂域大模型是针对智慧家庭领域开发的定制化模型，用以提升人工智能在家庭场景中的自然理解和交互能力。这些模型专注于智慧家庭特定需求，解决通用大模型无法满足的领域专业性和个性化问题。

在智慧家庭领域，垂域大模型已经成为智能交互、数据共享、节能增效等方面必不可少的生产力工具。智慧家庭大模型可以根据用户的需求和喜好，生成适合的智慧家庭场景和服务，实现智能化的家庭管理和控制；可以根据用户的语音、图像、手势等输入生成合适的语音、图像、触摸等输出，实现流畅自然的人机交互；还可以根据用户的行为和环境生成相应的效率预测和节能建议，实现智能化的调节和提醒。

降低了入局的门槛条件

智慧家庭中的语音控制、图像识别、文本生成等应用，需要大量的专业知识和资源，形成了一定的行业壁垒。而通过使用语言大模型和多模态大模型，可以轻松地支撑智慧家庭的服务和应用，使得开发智慧家庭语音助手、场景控制、安全防护等 App 的门槛条件大大降低。不仅如此，大模型也展现出惊人的创造力，为智慧家庭产品创新和差异化建设提供强有力支撑，可实现多种类型的内容理解和创作服务，开发出更多的新颖和有趣的产品和功能，满足用户的多样化和个性化的需求，增加自身的品牌影响力和用户黏性。

催生了跨界融合的新业态与增长点

借助大模型的内容生成与逻辑推理能力，泛智慧家庭领域的传统业态正在转型，语音搜索与推荐、智慧教育、影音创作、互动娱乐等跨界融合的业态不断涌现。例如：根据用户的影音需求和喜好，生成相应的影音和创作内容，实现智能化的影音和创作。智能娱乐还可以根据用户的影音素材和主题，生成适合的影音和创作内容，实现个性化和原创的影音和创作。

融合了大模型技术、多模态感知技术及 OTA 技术的智慧家庭垂域大模型 HomeGPT，通过大量领域特定数据的训练，显著提升了智慧家庭对用户需求的理解和响应速度。HomeGPT 垂域大模型在语音交互、图像识别、用户意图理解等方面进行了优化，使得智慧家庭更好地适应用户的个性化需求，通过解决通用大模型存在的领域专业性不足和无法满足特定家庭场景的问题，为用户提供更精准、智能的家庭体验。

1.5.垂域大模型引领智慧家庭进入L4

L4 级应用撬动万亿市场

2022 年 3 月，海尔智家牵头发布《智慧家庭智能家居智能化能力等级评估模型》，构建了完善的智慧家庭智能化等级的评价方法，填补了行业空白。评估模型以智慧家庭系统为评估主体，以人机智能协同理论为基础，通过系统的用户体验作为评估要素，分析用户与智慧家庭系统的关系及业务模式。该标准提出了智慧家庭智能化等级的定义，分别对应智慧家庭发展的 5 个阶段：单机智能（L1）、协作智能（L2）、决策智能（L3），以及即将实现的高度主动智能（L4）和泛在智能（L5）。该标准的发布，从根本上解决了长期困扰智慧家庭行业等级标准缺失的问题，为用户提供了智慧家庭智能化水平判断依据。

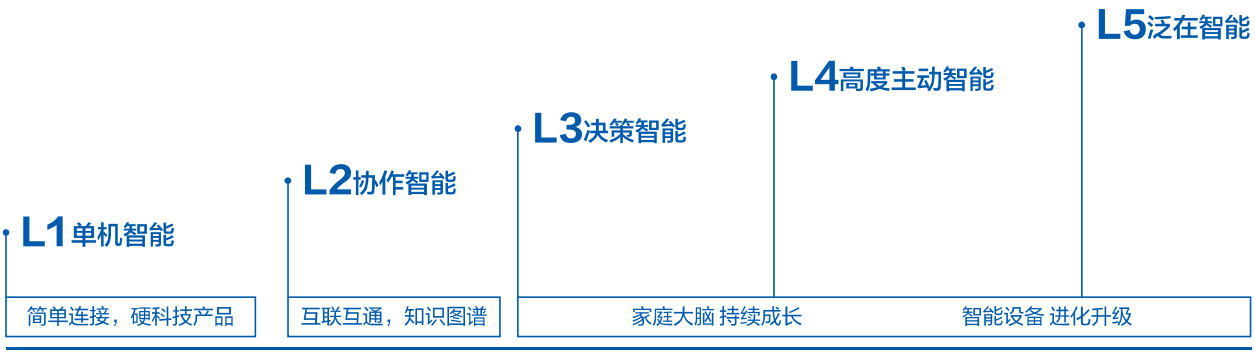


图 5智慧家庭智能家居智能化能力等级评估模型

智慧家庭已经历经了单机智能、协作智能，决策智能阶段，正在逐步跨入高度主动智能阶段。在市场规模方面，根据中国智能家居产业联盟的数据，2023 年中国智慧家庭智能家居市场规模预计可以达到 7157.1 亿元。预计到 2027 年，市场规模有望超过 1.1 万亿元。同时，智慧家庭设备市场出货量也在持续增长，2023 年达到 3.3 亿台。据 IDC 报告，2023 年中国智慧家庭市场面临宏观消费环境和自身发展周期的双重挑战，规模增速有所放缓，但市场并未停止升级调整的步伐。预计 2024 年中国智慧家庭市场需求将逐步回暖，设备出货量将同比增长 6.5%。在垂域大模型的支撑下，L4 级全品类应用将加速落地，越来越多家庭的智慧生活，将因此发生质变，智慧家庭领域进化升级速度将明显加快，智家家庭市场规模将持续增长。

AGI 助力智慧家庭 L4 技术升级

表 1 AGI级别

Levels of AGI		
深度（性能）和广度（通用性）	**Narrow（窄域）**	**General（通用）**
Level 0： No AI（无人工智能）	**窄域非人工智能** 计算软件、编译器	**通用非人工智能** 人机交互计算， 如Amazon Mechanical Turck
Level 1： Emerging（新兴级） 技能相当于或略比没有相关技能的人类要强	**新兴级窄域人工智能** 老式人工智能，简单的基于规则的系统，如SHRDLU	**新兴级通用人工智能** ChatGPT、Bard和Llama 2等大模型属于该阶段，并且已经满足了该阶段要达到的通用性
Level 2： Competent（胜任级） 可以达到正常成年人50%的水平	**胜任级窄域人工智能** 语音助手，如Siri、Google Assistant;视觉问答系统，如Pall、Watson；达到SOTA水平的大模型都属于这一阶段，但都只是在技能指标上合格了，通用性还不足	**胜任级通用人工智能** 尚未实现
Level 3： Expert（专家级） 可达到正常成年人90%的水平	**专家级窄域人工智能** 拼写和语法检查器，如Grammarly；图像生成模型，如Imagen或Dall-E2等可以划为该阶段，在技能水平上达标但通用性不足	**专家级通用人工智能** 尚未实现
Level 4： Virtuoso（大师级） 可达到正常人类99%的水平	**大师级窄域人工智能** 深蓝、AlphaGo等都属于	**大师级通用人工智能** 尚未实现
Level 5： Superhuman（专家级） 在技能指标上，已经可以超越顶尖科学家	**超人级窄域人工智能** AlphaFold、AlphaZero也可划入该阶段，但当前具备超人智能级通用性的AI还没诞生	**ASI（超级人工智能）** 尚未实现

谷歌 DeepMind 团队根据能力的深度（性能）和广度（通用性）提出了“AGI 级别”，该框架认为，发展 AGI 必须遵循 6 个基本原则：关注能力而非过程，同时衡量技能水平和通用性，专注于认知和元认知任务，关注最高潜力、而非实际落地水平，注重生态有效性，关注整条 AGI 之路的发展而非单一的终点。在此原则之上，AGI 将呈现 6 大发展阶段（Level 0 – Level 6），每个阶段如上表所示都有对应的深度（性能）和广度（通用性）指标。

面向领域的 AGI 是从 Narrow AI 到 General AI 的一个过渡阶段，在某些特定的领域或任务上有很强的偏向性。在智慧家庭领域中表现出强大的通用智能，不仅能够完成单一的任务，而且能够跨领域和跨任务地学习和解决问题，能够根据智慧家庭特定应用与场景的偏向性进行优化和调整。面向领域的 AGI 赋能智慧家庭的应用场景，促进形成智慧家庭的闭环，实现智慧家庭的自我学习和自我优化。智慧家庭 AGI 通过收集和分析用户的行为数据、环境数据、设备数据等，不断更新和调整智慧家庭领域知识图谱的参数、策略、内容等，智慧家庭领域 AGI 能够实现多任务规划、多数据融合，并实现对人、设备、空间、环境等多维度信息的智能化决策与评估，自主生成对应控制代码和场景，自主生成动态调整设备控制程序，而且可以综合家庭设备、知识以及实时数据与用户进行语言、语音、屏幕、甚至投影、视频流等形式的互动以适应用户的个性化需求和喜好，提升用户的满意度和忠诚度。

02 技术篇

AGI在智慧家庭领域的探索实践

2023版的智慧家庭大脑白皮书定义了智能感知、智能交互、智能决策、智能连接、智慧生态、安全合规等几个技术模块。随着大模型技术的发展，智慧家庭大脑也结合大模型技术进行了系统性重构与扩充，形成了“一个引擎、两个模块、三个平台”的全新体系架构。

——垂域大模型引擎： 区别于通用大模型，垂域大模型在智慧家庭方向进行了特定微调和优化，更好地为智慧生活提供脑力支撑。

——系统管理模块： 负责系统的整体管理，确保各种服务和功能的安全、可靠运行。同时，系统管理平台还负责数据的ETL（提取、转换、加载）管道，这是数据处理的关键步骤，确保数据的准确性和可用性。

——应用网关模块： 包括API网关和API管理，负责将智能家电（如空调、冰箱、洗衣机等）与用户的需求和环境相结合，实现高度自动化的个性服务。通过应用网关，这些智能设备可以接入系统，实现与智慧家庭大脑的无缝对接。

——场景应用平台： 智慧家庭大脑通过智能场景管理和业务场景编排，根据用户的行为和偏好，自动调整家电的运行状态，实现智能化的生活体验。例如，通过场景库和智能菜谱，可以根据用户的饮食习惯自动推荐菜谱并协调厨房设备的工作。

——行业知识引擎平台： 智慧家庭大脑集成了垂直领域大模型引擎，提供私域知识库和语言理解服务，如NLP（自然语言处理）。这些服务可以提高对用户指令的理解准确性，并通过信心度打分来优化响应。

——生态开放平台： 建立智慧家庭领域生态资源服务的引入、认证、分发、应用、运营的一站式平台。平台提供强大的终端设备接入能力和完善的开发工具，帮助行业从业者快速构建智慧家庭生态服务解决方案。

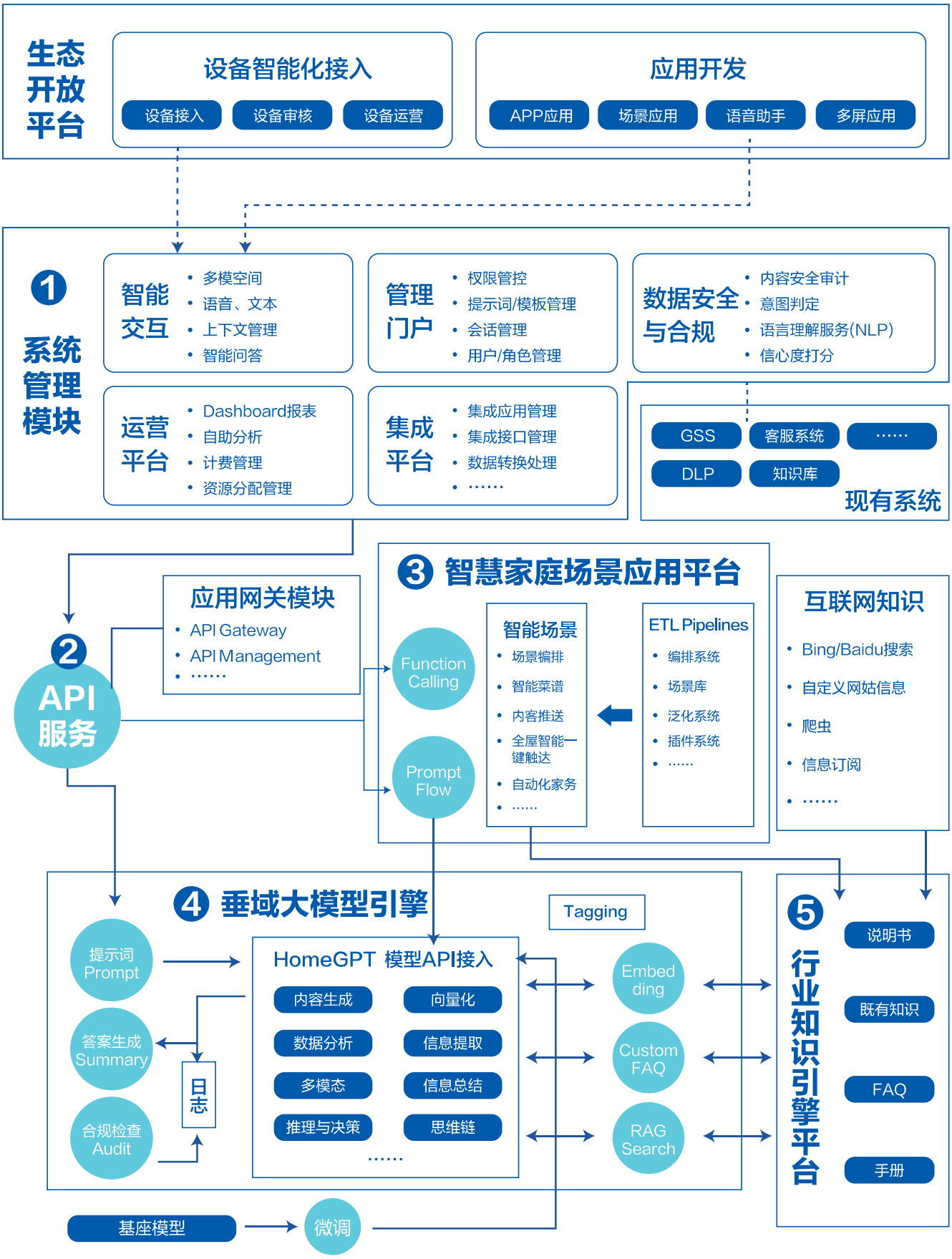


图 6 系统整体架构及规划

2.1.构建垂域大模型的关键能力

在智慧家庭领域内，打造一个功能强大的垂域大模型是一项系统而复杂的工程，涉及基座模型选型、数据处理、模型训练、强化学习以及实时性、安全性等多个方面。首先，需要选择一个具备可适应性和扩展性的基座模型，以便能够满足智慧家庭的多样化需求。其次，要构建对海量数据的高效处理能力，以确保数据质量，以便模型能从中学习到有用的信息，并通过精准的模型训练，让模型掌握智慧家庭中的各项任务。此外，运用强化学习技术使模型能够在与环境的实时互动中不断进步，更好地适应用户的个性化需求。最后，模型要能够即时响应用户的需求和家庭环境的变化，并确保用户数据的隐私和家庭网络的安全。通过这些关键能力的构筑，就能构建起一个智能、高效、自进化的智慧家庭垂域大模型，为用户带来更加舒适和便捷的生活体验。

基座模型选型

目前常见的开源基座大模型如下表所示：

表 2 常见开源基座大模型

序号	模型	简介
1	MOSS基座语言模型	预训练模型在约七千亿中英文以及代码单词上预训练得到，具备多轮对话能力和使用多种插件的能力。
2	GLM-130B基座模型	千亿中英语言模型，具有初具问答和对话功能，在GLM-130B基座模型中注入了ChatGPT的设计思路。
3	GLM-4基座大模型	智谱AI发布的新一代基座大模型，是其大模型事业三年多来积累的技术成果之一，于2024年1月发布。
4	Awesome-Chinese-LLM	一个整理开源的中文大语言模型的项目，包括规模较小、可私有化部署、训练成本较低的底座模型，垂直领域微调及应用，数据集与教程等。
5	Phi2模型	具有约27亿个参数。与其他大模型不同的是，Phi-2并没有通过人类反馈的强化学习（RLHF）进行调整，也没有经过指导性微调。在多个基准测评上，Phi-2的性能表现超越了Mistral和Llama 2的7亿参数版本。
6	Llama2模型	参数规模范围从70亿到700亿不等。这个模型由Meta微调而来，被称为Llama 2-Chat，专为对话场景进行优化。

企业用户需从深度学习框架、微调模型和知识召回等关键模块出发，综合考虑模型的性能特征、与业务场景的任务相似性、可解释性、数据需求、领域适应性、开源性和可扩展性等多个维度，以选择一个既符合当前业务需求又具备未来发展潜力的开源或商用大模型，确保智慧家庭系统能够高效、准确地服务于最终用户。

数据处理能力

构建垂域大模型引擎的关键能力之一是数据处理能力。在家庭大脑中，数据处理能力包含了几个重要环节：①高效的数据采集，能实时采集家庭智能设备数据并支持离线数据采集。②数据预处理，消除噪声和冗余数据，对原始数据进行清洗和归一化，保证数据的有效性。③大数据分析，通过利用最新的大数据技术、算法和模型，从海量的家庭设备数据中抽取有价值的信息进行模型训练。④数据融合，在严格的规范下，对不同设备间的数据进行融合，对智能设备进行多维度分析和了解。⑤实时响应和处理能力，为了支持大规模的智能设备实时响应，需要设计高效的实时数据处理架构和策略。⑥数据安全和隐私保护，遵守国内外相关法律法规，采用最新的数据加密技术和访问控制策略，确保数据存储、处理和传输的安全。⑦数据可视化的实现，通过提供一种方便、直观的方式，使得决策者和普通用户能更好的理解和利用数据。这些组成部分共同构成了强大的数据处理能力，使得家庭大脑能够更好地服务于家庭用户，提供更智能的智慧家庭解决方案。

模型训练能力

模型训练能力是建立家庭大脑的基石。在垂域大模型引擎中，模型训练质量的高低直接关系到家庭大脑的性能表现，决定了它是否能准确理解和响应用户的需求，以及能否提供高效率和高质量的服务。为了实现顶尖的模型训练能力，需要专注于以下几个关键领域。

基座模型微调

家庭大脑依托于最先进的深度学习和机器学习算法。从神经网络架构的设计，到优化算法的选择，始终需遵循最佳实践来提升模型的学习效率和预测精度。通过微调家庭大脑可以适应家庭成员的特定习惯和偏好，优化家庭设备的运行效率，以及更好地理解居住环境中的独特场景和需求。这一过程要求精确地调节模型参数，利用少量的高质量领域数据，达到增强模型对噪声的鲁棒性和对新情况的适应力，而不丧失已学习的知识。有效的微调不仅增强了模型的泛化能力，还大大缩短了学习周期，使家庭大脑能够快速部署，并随着时间的推移在现实世界中持续进化。

高质量训练数据获取

模型训练需要大量的高质量数据。家庭大脑通过先进的数据处理流程确保输入数据的质量与一致性。采用自动化工具进行数据清洗、标注以及增强，保证模型接受到的是准确和多样的训练数据。这些数据不仅覆盖了广泛的情景，也确保了模型能够在复杂多变的真实世界条件下持续有效运行。

高效的资源管理

高效利用计算资源对于模型训练至关重要。家庭大脑借助于强大的分布式计算框架，确保模型训练能够平行地在多个计算节点上运行，显著缩短了训练时间。通过使用自动化资源分配系统来优化计算资源的使用，在保证训练质量的同时最大化降低了硬件成本。

实时监控与调整

在模型训练过程中，家庭大脑的监控系统会实时跟踪每次训练的性能。借助于先进的可视化工具和监控指标，迅速识别并解决训练过程中出现的任何问题。此外，制定调整策略以确保在发现模型性能有所下降时，迅速进行微调，维持训练的最优状态。

持续学习与适应

模型训练不是一次性的任务，而是一个持续的过程。家庭大脑采用连续学习的框架，使得模型能够适应新的数据和用户行为的变化。随着智慧家庭领域的不断演进，模型可通过持续学习保持其先进性和准确性。

结合这些关键要素，针对特定领域构建大规模的领域微调数据集，确保模型在实际任务中具有良好的性能。而在垂域大模型的研发中，通常出现追求模型规模而忽视实际效果的“幻觉”现象，可通过数据的收集、标注和预处理等步骤以减弱或消除该现象，并持续的优化模型训练流程，以确保家庭大脑始终领先于智慧家庭技术的发展前沿。

强化学习能力

强化学习作为家庭大脑的重要组成部分，使系统能够通过与环境的不断互动和试错，自主学习如何做出最优决策。在这一过程中，家庭大脑接收环境反馈作为奖励信号，评估自己的行为并调整策略以最大化长期奖励。强化学习能力的发展涉及到高效的策略探索、风险评估与奖励机制设计，确保家庭大脑的行为能够真实反映居住者的偏好并满足需求。此外，为了使家庭大脑具备更强的泛化与适应性，它将整合模拟环境训练与真实环境中的在线学习，不断提升对智慧家庭场景的理解与响应速度，最终实现提升居住体验与家庭能效的智能自适应系统。

实时性能力

实时性能力是家庭大脑提供高效服务的核心要求。家庭大脑必须能够快速响应环境变化和用户指令，实时处理来自智慧家庭设备的数据流，并做出即时决策，以确保家庭自动化系统的流畅运作和居住者的即刻需求得到满足。在技术层面，这要求系统拥有低延迟的数据处理架构、快速的事件驱动机制和高效的算法优化，确保任务能在毫秒级别执行。家庭大脑的实时性能力还涉及到实时学习的能力，能够持续根据实时反馈调整自身行为，提供更加精准和个性化的服务。通过这种方式，确保在动态变化的家庭环境中，家庭大脑始终保持系统的高效运行和用户体验的最优化。

安全性能力

安全性能力对家庭大脑来说同样至关重要，它确保了智慧家庭系统中的数据保护和隐私维护，以及系统运行的稳定性和可靠性。家庭大脑需要配备先进的安全机制，包括数据加密、访问控制和入侵检测系统，来防御外部攻击和内部泄漏风险。安全性能力还涉及到对异常行为的实时监测和响应，确保在发生潜在威胁时立即采取行动，从而防止对居住者的生活造成影响。通过持续监控系统的健康状态，并针对新出现的安全漏洞迅速更新防御策略，家庭大脑在保障数据和设备安全的同时，提供一个值得信赖的智慧家庭环境。

2.2.行业私域知识引擎平台搭建

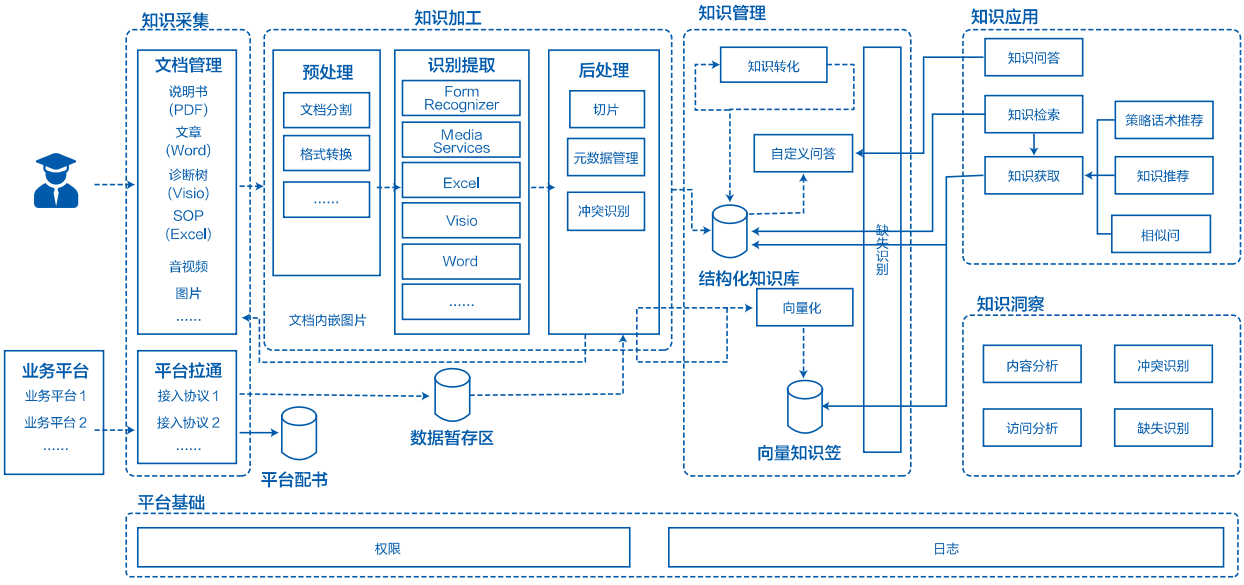


图 7 私域知识平台搭建

在智慧家庭领域，构建领域知识平台是企业垂域大模型建设的核心。这个平台不仅是整合和提炼智慧家庭领域专业知识和经验的基础，也是打造高效、智能的私域知识引擎的关键环节。通过行业私域知识引擎平台，可以有效地管理和利用海量的家庭用户数据，提供个性化的智慧家庭解决方案。

知识采集与获取

将企业既有的数据通过集中化平台进行存储和管理，例如：家电企业运营中积累了大量产品说明书、诊断树、音视频甚至是客户评价等等数据，这些都是私域知识平台的有效数据来源。

知识获取始于多种形式数据的采集，包括 PDF、Word、Excel、Visio 和 SOP 格式的文档。这些原始数据通过 Form Recognizer 和 Media Services 等高级服务进行预处理、识别提取和后处理。这些服务在将非结构化数据转化为结构化知识方面发挥着关键作用，为系统的进一步使用做好准备。

知识存储与加工

对企业私域数据进行格式化处理，以便大模型将数据转成企业知识。知识加工直接影响了大模型回答内容的准确性，以 OpenAI ChatGPT 为代表的大模型训练初期，动用了大量的人力进行知识标注，其实就是知识加工过程。企业私域知识是基于既有数据内容引入大模型基座模型整理而来，相对于冷启动的大模型训练，已经节省很大人力物力。知识加工过程会用到多个 AI 能力，比如语音识别、图片视频、表单识别等等，是一个复合能力集成环节。

结构化的知识存储在安全的数据库中，作为系统在需要时可以调用的中央仓库。知识转化进一步将收集到的信息转换为可行的洞察力，这对智慧家庭响应用户需求和环境变化至关重要。

知识管理与应用

知识管理的核心在于系统能够有效地维护和利用大量信息。这包括知识检索、自定义查询、冲突识别和诊断等能力。系统确保知识库是当前和相关的，允许对用户查询和各种智慧家庭组件的操作做出准确的响应。

在知识应用环节，需要知识平台提供高效、准确的内容定位能力，而搜索技术是目前在海量知识中获得知识内容最快的方式。为了提高搜索准确性，检索增强（Retrieval-Augmented Generation，RAG）技术应运而生。

RAG 提高模型的效率和灵活性

微调(Fine-tuning)和 RAG 都是用于提升机器学习模型性能的技术，但它们在实施方式和成效上存在差异。微调通常是指在大型预训练模型的基础上，通过额外训练来适配特定任务或数据集。这种方法可以显著提高模型在特定任务上的表现，但通常需要较大的计算资源和时间投入。而 RAG 则采用了一种不同的策略。它通过结合检索机制和生成模型，允许模型利用现有知识库中的信息来处理新的输入数据。这种方法允许模型在不同上下文中进行实时学习，而不需要进行代价高昂的微调过程。因此，RAG 能够提升模型在处理新信息时的效率和适应性，使得模型更加经济高效。利用 RAG 增强大型语言模型，主要带来以下优点：

——实时更新的知识： RAG模式可以从最新的文本语料库中检索信息，这使得它能够获取训练后的新知识，处理需要实时更新的问题。

——特定领域的知识： RAG模式可以从特定领域的文本语料库中检索信息，这使得它能够处理需要特定领域知识的问题。

——处理长尾问题： 通过检索机制，RAG模式可以处理一些长尾问题，即那些出现频率低但需要特定知识才能回答的问题。

——灵活性和可扩展性： RAG模式的设计使得它可以很容易地与其他模型结合，或者在不同的任务和领域中使用。

——提高生成的质量： RAG模式在生成回答时，会考虑到检索到的内容，这使得它能够生成更准确和更详细的回答。

知识检索增强与呈现

RAG 主要的工作模式是利用已有的知识来做检索增强并生成最终的结果给用户。它由多个组件协同工作，最终通过 LLM 处理并给出答案，下图展示了主要的工作流程。

这个过程主要包括离线和在线两个部分，离线主要是后台数据的准备，主要步骤包括：数据摄取，文档分割，向量化处理。在线的部分包括：用户提问、问题增强和生成结果这几个主要的步骤。

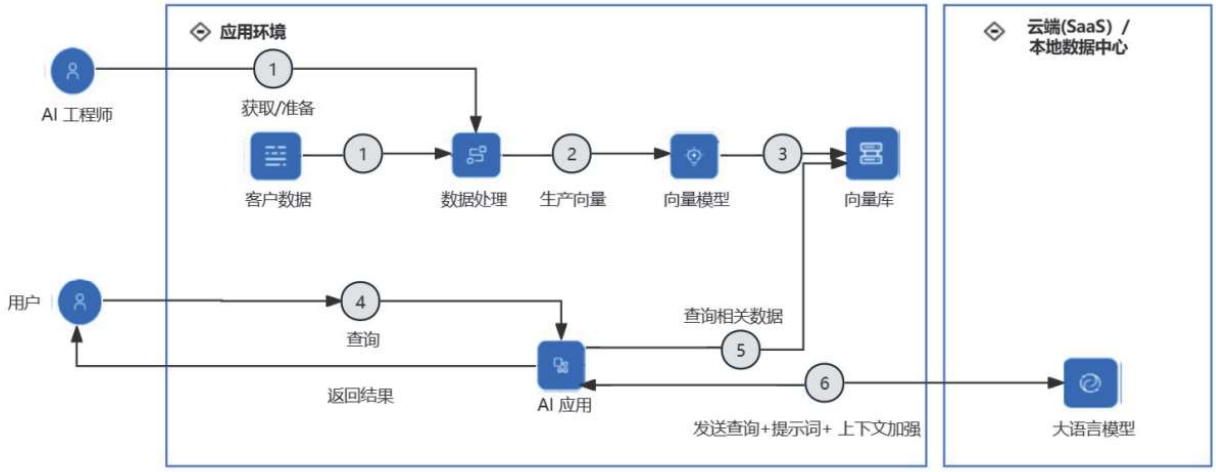


图 8 RAG 流程

数据摄取（Data Ingestion）

需要考虑数据的质量和多样性。数据可能来自不同来源，包括但不限于内部知识库、流程文档、产品说明书、维修手册等等。数据摄取的目的是收集尽可能全面和丰富的信息，以支持模型在后续步骤中进行有效的检索和生成。

在 RAG 系统中，准确且清晰的原始知识数据是提升搜索效率和准确性的关键。为了保证数据的准确性和清晰度，需要对数据进行优化和清洗。通过优化文档读取器和多模态模型并进行基本的数据清洗，来提升 RAG 系统的搜索效率和准确性。

在处理如 CSV 表格等文件时，单纯的文本转换可能会丢失表格原有的结构。因此，需要引入额外的机制以在文本中恢复表格结构，比如使用分号或其他符号来区分数据。这样可以确保原始数据的准确性，提高搜索的准确性。数据清洗是提升搜索效率和准确性的关键步骤，包括以下几个方面：①基本文本清理：通过规范文本格式，去除特殊字符和不相关信息，以及重复文档或冗余信息，减少无关的搜索结果，提高搜索的准确性。②消除实体和术语的歧义，以实现一致的引用，例如，将“LLM”、“大语言模型”和“大模型”标准化为通用术语，可以确保搜索结果的准确性。③合理地划分不同主题的文档：如果作为人类都不能轻松地判断出需要查阅哪个文档才能来回答常见的提问，那么检索系统也无法做到。因此，需要将文档按照主题进行合理的划分，以便于搜索。④使用同义词、释义甚至其他语言的翻译来增加语料库的多样性，提高搜索的覆盖率和准确性。⑤基于现实世界用户的反馈不断更新数据库，标记它们的真实性，以提升搜索效率和准确性。对于经常更新的主题，可以通过实施一种机制来使过时的文档失效或更新，确保搜索结果的时效性。

分块（Chunking）

在这个步骤中，大量的文档会被分解成更小的部分或 " 块 "，这样做的目的是为了使得检索过程更加高效，每个块通常包含一定数量的词或句子，块的大小取决于特定的应用需求，但通常会尽量保持足够的信息，以便在检索阶段能够提供有用的上下文。这个过程需要精细的平衡，因为太大的块可能会导致检索过程变慢，而太小的块可能会失去重要的上下文信息。在 RAG 系统中，文档的向量化嵌入需要进行精细的文本分块。在忽略大型模型输入长度限制和计算成本的前提下，这一做法旨在保持语义连贯性的同时，最大程度地减少嵌入内容的干扰因素，以便更精准地检索出与用户查询最匹配的文档段落。若分块过大，可能会引入过多无关信息，从而削弱检索的精确度。反之，过小的分块可能会遗失关键的上下文信息，导致生成的回答缺乏连贯性或深入性。实施适当的文档分块策略是 RAG 系统的核心，以求找到这种平衡，保证信息的完整性和相关性。理想的文本块应在脱离周边上下文的情况下对人类仍具有意义，从而对语言模型也具有意义。确定文本块大小的参数是一项复杂的任务，需要考虑多个因素。首先，不同的嵌入模型有其最佳输入大小。例如，OpenAI 的 text-embedding-ada-002 模型在 256 或 512 大小的块上表现最佳。其次，文档类型和用户查询的长度及复杂性也是决定分块大小的重要因素。处理长篇文章或书籍时，较大的分块有助于保留更多的上下文和主题连贯性；对于社交媒体帖子，较小的分块可能更适合捕捉每个帖子的精确语义。如果用户的查询通常是简短和具体的，较小的分块可能更为合适；相反，如果查询较为复杂，可能需要更大的分块。在实际应用中，我们可能需要通过不断的实验和调整来找到最佳的分块大小。在一些测试中，128 大小的分块往往是最佳选择，在无从下手时，可以从这个大小作为起点进行测试。

在复杂的信息检索 RAG 系统中，内容的分块会影响查询的准确率。以下将介绍五种常用的分块方法，包括固定大小的分块、内容分块、递归分块、从小到大分块和特殊结构分块，以及如何选择并应用它们以提升搜索效率和准确性。选择合适的分块方法是提升 RAG 系统搜索效率和准确性的关键。根据文档的内容和结构，以及搜索的需求，可以选择最适合的分块方法，以实现最优的搜索效果。

向量化处理

想象一下，如果有一堆无序且格式各异的数据，像是文本、图像或声音文件，它们就如同一堆散乱的积木一样。传统的计算系统往往难以处理这类无结构数据，因为它们依赖于整齐、有序的数据格式。这时，我们就需要一种能够将这些混乱的积木（非结构化数据）转换成有序的积木堆（结构化的数字表示）的方法。这种方法被称为嵌入（Embeddings），它指的是将文本、图像或声音转换为向量的过程。向量是一种数值表示形式，能将概念转换成计算机能理解的数字序列，有助于计算机把握不同概念之间的关系。

嵌入就像是一种魔法工具，它能够将各种形式的信息转化为计算机能够处理的数字格式。然而，这种转换成的数字表示往往非常复杂，宛如一个庞大的数字积木堆，仍旧难以为计算机所处理。因此，嵌入技术的另一个关键作用是将这些庞大的积木堆（高维稀疏向量）简化为更小的积木堆（低维稠密向量），从而让数据变得更易于计算处理。嵌入向量是一系列浮点数，向量之间的距离代表了相应概念之间的相关性 —— 距离越小表示相关性越强，距离越大则相关性越弱。

OpenAI 开发的 text-embedding-ada-002 模型便是一个处理嵌入的专用模型。对于给定的输入信息，它可以输出一个 1536 维的向量数组，有效地将信息编码为计算机可以进一步处理的形式。

用户提问

用户通过各种方式来和知识系统来做交互，例如：通过聊天工具，Copilot，或者语音等方式完成交互。

问题增强

在 RAG 系统中，我们将用户的问题转化为向量形式，然后在向量数据库中进行搜索。很自然地，问题的措辞会对搜索结果产生直接影响，如果搜索结果不尽人意，可通过以下几种策略对问题进行改写，以提升搜索的精确度：

——结合历史对话进行重新表述：在向量空间中，即使两个问题在人类看来非常相似，它们的向量表示也可能并不相近，可以利用 LLM 直接对问题进行改写。另外，在多轮对话中，用户的提问可能会引用前文的信息，因此可以将历史对话和用户的提问一起交给 LLM 进行改写。

——假设性文档嵌入（HyDE）：HyDE 的核心理念是，在接收到用户的问题后，让 LLM 在没有任何外部知识的情况下生成一个假设性的回答，然后将这个假设性的回答和原始问题一起用于向量搜索。虽然假设性的回答可能包含一些不准确的信息，但它包含了 LLM 认为相关的信息和文档模式，有助于在知识库中寻找类似的文档。

生成结果

把用户的问题和从向量数据检索出来的信息，组合成一个新的提示词，输入给 LLM，经过 LLM 的总结和处理，返回给最终的用户，完成一次的对话交互过程。

基于大语言模型的检索增强生成（RAG）方法已经彻底改变了处理数据提取、信息检索和答案生成的方式。特别是在问答系统的背景下，这些技术提供了一种更高效、更准确的方式来提取和分析大量数据。通过利用大语言模型和像 Chroma 和 FAISS 这样的向量数据库，可以创建知识中心和 RAG 流程，自动从大数据集中提取相关信息并生成对复杂问题的准确答案的过程。随着这些技术的持续发展，可以预期在未来的数据分析和信息检索领域会看到更多令人兴奋的进步。

数据合成技术

数据合成技术已经有了飞速的发展，可以成功模仿很多实际数据的基本属性。基于各种数据合成的技术，使得数据合成作为真实数据的替代品，具有广阔的应用前景。数据合成的核心技术如下：

变分自编码（Variational Autoencoder，VAE）

变分自编码器是深度生成模型，与传统的自编码器通过数值方式描述潜空间不同，它以概率方式对潜在空间进行观察，在数据生成方面应用价值较高。VAE 分为两部分，编码器与解码器。编码器将原始高维输入数据转换为潜在空间的概率分布描述；解码器从采样的数据进行重建生成新数据。

生成对抗网络（Generative Adversarial Networks，GAN）

GAN 使用零和博弈策略学习，在图像生成中应用广泛。GAN 包含两个部分：生成器学习生成合理的数据。对于图像生成来说是给定一个向量，生成一张图片。其生成的数据作为判别器的负样本。判别器判别输入是生成数据还是真实数据。网络输出越接近于 0，生成数据可能性越大；反之，真实数据可能性越大。生成器与判别器相互对立。在不断迭代训练中，双方能力不断加强，最终的理想结果是生成器生成的数据，判别器无法判别是真是假。

扩散模型（Diffusion Model）

扩散是受到非平衡热力学的启发，定义一个扩散步骤的马尔科夫链，并逐渐向数据中添加噪声，然后学习逆扩散过程，从噪声中构建出所需的样本。扩散模型的最初设计是用于去除图像中的噪声。随着降噪系统的训练时间越来越长且越来越好，可以从纯噪声开始，作为唯一输入，生成逼真的图片。标准的扩散模型分为两个过程：前向过程与反向过程。在前向扩散阶段，图像被逐渐引入的噪声污染，直到图像成为完全随机噪声。在反向过程中，利用一系列马尔可夫链在每个时间步逐步去除预测噪声，从而从高斯噪声中恢复数据。

Transformer

采用注意力机制（Attention）对输入数据重要性的不同而分配不同权重，其并行化处理的优势能够使其在更大的数据集训练，加速了 GPT 等预训练大模型的发展。最初用来完成不同语言之间的翻译。主体包括 Encoder 与 Decoder 分别对源语言进行编码，并将编码信息转换为目标语言文本。采用 Transformer 作为基础模型，发展出了 BERT，LaMDA、PaLM 以及 GPT 系列，人工智能技术进入大模型参数的预训练模型时代，助力数据合成发展。

Vision Transformer (ViT)

将 Transformer 应用于图像分类任务，Transformer 在图像领域得到广泛应用。ViT 将图像进行分片，并对每个分片进行线性变换，得到固定长度的向量送入 Transformer，使用标准的 Transformer 处理方式。以 ViT 为基础衍生出了多重优秀模型，通过将人类先验经验知识引入网络结构设计，获得了更快的收敛速度、更低的计算代价、更多的特征尺度、更强的泛化能力，能够更好地学习和编码数据中蕴含的知识，正在成为视觉领域的基础网络架构。以 ViT 为代表的视觉大模型赋予了 AI 感知、理解视觉数据的能力，助力数据合成发展。

尽管数据合成不能成为真实世界的标准答案，但不影响其提供高质量数据的发展趋势，未来数据合成的发展趋势如下：

——数据合成的可解释性： 针对垂直领域原始数据的准确建模和深入理解，加强行业数据本身的分析，提高对数据的洞察能力。通过专业工具，研究复杂的模型中洞察数据特征重要程度、预测数据特征的影响程度、特征对预测结果的典型影响程度。从而，获得对模型洞察的诸多优点，使得模型易于调试、构建专业特征工程、指导未来数据合成、启发人为决策并建立良好的信任关系。

——数据合成的可评估性： 数据合成的质量和效果相关的指标评估，如何建立评价共识是检验数据合成质量的前提，需要建立更加严谨的评估体系。为了应对与异构数据和应用相关的众多挑战，将原始数据与合成生成的样本进行比较，可以从五个高级抽象标准评估生成模型，包括代表性、新颖性、真实性、多样性和一致性，并反映在特定行业中对合成数据的要求。

——数据合成的可扩展性： 随着不同行业对训练数据类型需求的增加，需要选择合成通用的数据，可以根据机器学习模型的训练需求进行合成数据的扩展和定制化。无条件生成合成数据扩展到有条件生成数据中，通过给定的上下文，使得有条件生成的合成数据在上下文中具有稳健性，而可扩展性则需要根据上下文进行更细微的区分。

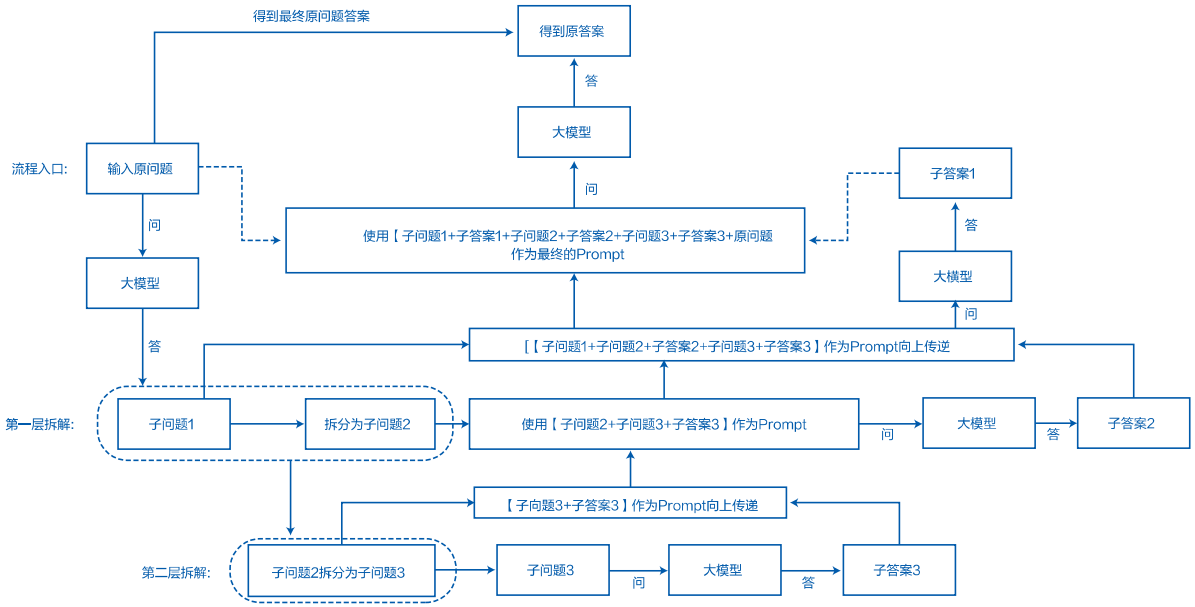
——数据合成的可保护性： 数据合成与数据保护之间矛盾已经逐渐显现，在兼顾数据合成效用的同时确保数据隐私与安全。现有的数据合成技术仍面临隐私保护与信息泄露的风险。差分隐私是未来数据合成重要的研究方向，可以在数据样本中加入随机噪声，有效防止从合成数据中提取敏感信息，既保留原数据统计特性又避免隐私泄露的高质量合成数据，解决数据合成的隐私保护与数据效用之间的矛盾。

2.3.家庭大脑与大模型思维链

2022 年，在 Google 发布的论文《Chain-of-Thought Prompting Elicits Reasoning in Large Language Models》中首次提出，通过让大模型逐步参与将一个复杂问题分解为一步步的子问题并依次进行求解的过程可以显著提升大模型的性能。而这一系列推理的中间步骤就被称为思维链（Chain-of-Thought，CoT）。

在智慧家庭领域，最新的家庭大脑集成了先进的大模型技术和大数据分析的中心处理单元。它能够理解家庭成员的需求，优化家庭资源配置，并提供个性化服务。目前大模型仅针对单项任务处理有很好的效果，但针对多层任务，需要对任务进行自分解、自学习、多任务调度很难达到预期效果。而思维链则是一种新兴的人工智能技术，它通过模拟人类的推理过程，提高了模型在复杂任务中的性能。

智慧家庭中的“思维链”涉及到由“智慧家庭大脑”做出决策后产生的指令或行动链，也就是通过“大脑”与所有连接设备通信的方式，保证操作的协调一致。“思维链”是一个逐步过程，将复杂的任务分解为每个设备可以执行的行动。与传统人工编排方式不同，加入基于大模型的思维链技术可以有效降低人工编排的工作量，大部分用户需求和体验都由“大脑”进行决策，然后由思维链进行执行。



工业级提示流程:
第一阶段:输入原始问题后，识别核心问题，对其进行拆分
第二阶段:从最深的问题层次开始，使用大模型解答每个子问题，并生成相应的答案
第三阶段:下层的问题+答案作为Prompt不断向上层输出
第四阶段:所有子问题+答案逐层累计作为Few-shot提示词，与原问题共同输入到大模型，得到原问题的最终答案

图 9 思维链工作流程

逻辑性与全面性

——“逻辑性”确保智慧家庭大脑的决策过程遵循某种合理且符合规则的推理路径。例如，如果智慧家庭大脑检测到用户已离开住宅，那么它可以通过逻辑推断没必要继续保持某些电器的开启。这种基于逻辑的决策可以最大化资源利用效率，并提高用户的便利性。

——“全面性”确保智慧家庭大脑考虑到所有可能的场景，允许全局管理家庭系统，涵盖安全到娱乐的所有方面。全面性在预防和冲突解决中是必要的，例如，家庭成员间存在睡前习惯与晚间娱乐的偏好冲突，那么智慧家庭大脑需要全面考虑这些矛盾的需求。

智慧家庭大脑有效地结合逻辑性和全面性，确保基于大量数据输入的准确推断，以及智能家电的无缝运作，对现代生活的便利和舒适性大有裨益。

可行性与可验证性

思维链（CoT）技术的引入，使得智慧家庭大脑的决策过程更加可行和可验证。家庭成员可以通过CoT提供的中间推理步骤，理解大脑是如何得出特定决策的，增加系统的透明度和可信度。

——“可行性”确保智慧家庭大脑提出的解决方案和应用能够在实践中实施，会综合包括技术能力、成本效益、用户互操作性等多个因素。如果收益大于成本，并且在现有技术能力范围内，那么一个建议就被认为是可行的。例如：净化室内空气的操作路径不止一条（开启新风系统、开门开窗通风、打开空气净化器等），但如果用户并没有能够自动开合的门窗，这条路径就变的不可行。

——“可验证性”确保智慧家庭大脑做出决策后的结果是可以度量和评估的。思维链的所有反应和行动都应该是可追踪和负责任的。思维链的验证能力提高了家庭大脑的信任度和可靠性，并允许持续改进。例如，由于智能家庭应用，节省的能源或改善的用户体验可以被量化和验证。

本质上，一个有效的智慧家庭系统应该遵守可行性和可验证性原则，确保智能解决方案不仅实用，而且能够被准确地度量和持续地改进，以提供最好的可能的用户体验。

思维链的理念应用

在智慧家庭领域，AI Agent 作为自主智能的实体，可自主的发现问题、确定目标、构想方案、选择方案、执行方案、检查更新。但作为 Agent 主体的大模型是模拟人类智能决策流程的核心，在许多 Agent 需要处理的任务中，Agent 的“先天知识”并不包含解决任务的直接答案，因此，Agent 需要在一系列与外部环境的交互循环中，制定计划、做出决策、执行行动、接收反馈。整个计划、决策与控制的循环过程，是 CoT 的感知、记忆和推理能力与 Agent 结合的典型应用。

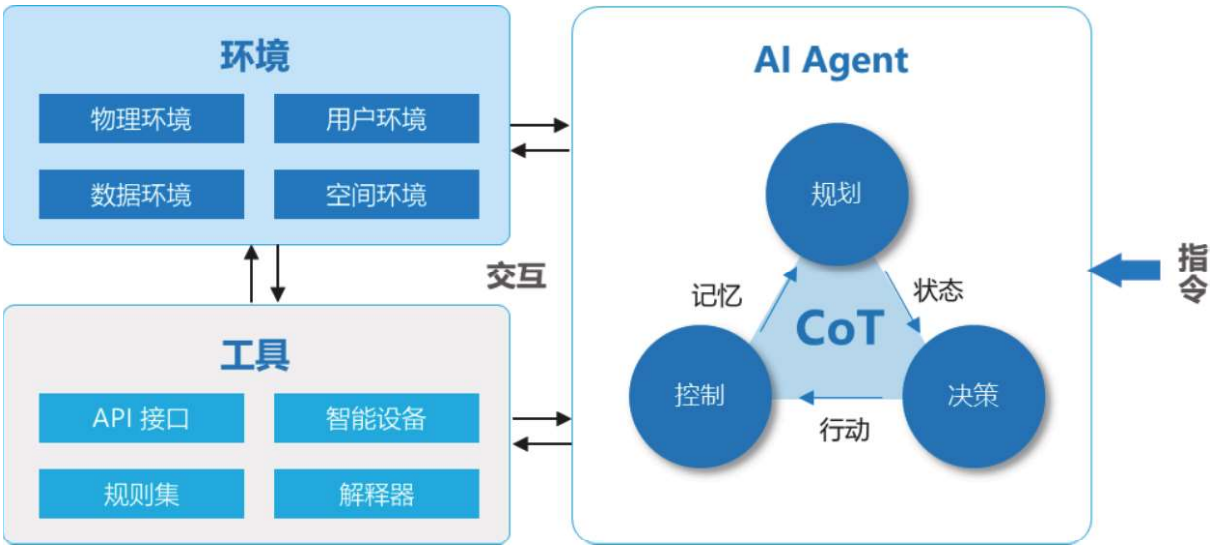


图 10 人工智能体的工作框架

在智慧家庭中，CoT 可以应用于多个场景，如安全监控、健康管理、娱乐推荐等。通过 CoT，智慧家庭大脑能够更准确地预测和满足家庭成员的需求，提升生活质量。在节能管理方面：智慧家庭大脑基于使用者的日常习惯与行为模式进行分析，由 " 思维链 " 控制家用电器如冰箱、空调、照明设备等在合适的时间自动开启和关闭，有效节省能源消耗；在健康监控方面：智能家庭大脑通过连接各种健康设备如智能血压计、健康手环等，通过思维链实时监控家族成员的身体状况，并在必要时通知医疗服务；在智能教育方面：智慧家庭大脑可连接智能教育设备，根据儿童的学习习惯和需要，制定个性化教育方案，并通过思维链激活相关设备提供学习内容；在自动化生活方面：包括自动化烹饪（智能厨具）、自动化植物养护（智能浇水设备）、自动化娱乐（音乐、电视等）等，让家庭生活更加智能化舒适。

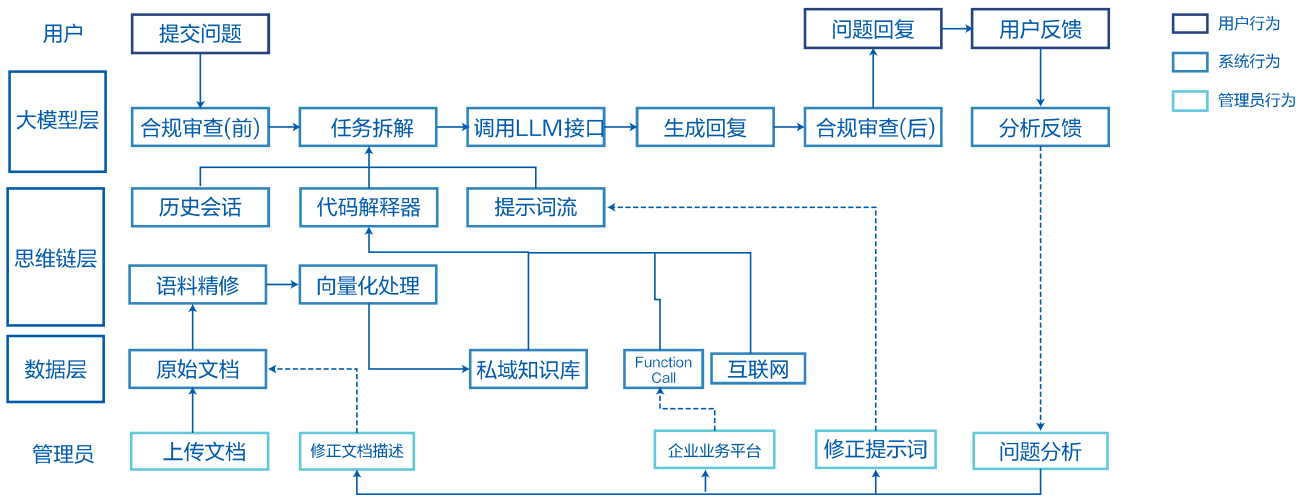


图 11 思维链典型调用场景

随着 CoT 技术的不断发展，未来的智慧家庭大脑将更加智能和人性化。它将能够处理更复杂的任务，提供更精准的服务，成为家庭中不可或缺的智能伙伴。在具备“思维链”技术的家庭大脑的加持下，智慧家庭系统不仅仅是独立设备的集合，而是作为一个有机的、相互连接的生态系统协同工作，创造出真正智能的生活环境。

2.4.多模态联合推理与决策

在 2023 版的家庭大脑白皮书中，有两个非常重要的模块：智能感知和智能决策平台。经过一年的努力，智能决策平台已经借助大模型技术能够达到很好地效果。但是，当前用户仍然较多地使用语音、文字、APP 与智慧家庭助手进行交互，视觉、触觉等交互模态应用较少。因此，引入多模态联合推理与决策提升智能感知 + 智能决策能力，有助于实现用户与智能家电体验全方位衔接。

多模态的定义

在智慧家庭大脑中，多模态是指集成和处理多种类型信息来源的能力，如视觉、声音、触觉、结构光等在内的各种传感器数据。其中包括摄像头捕获的图像，麦克风接收的音频，温湿度传感器、WiFi的数据等。通过合理整合和分析这些数据，智慧家庭大脑能更全面理解家庭环境的实际状态，从而做出正确和高效的决策。

多模态信息处理与联合推理

在多模态信息处理上，关键是要实现各模态数据的高效融合。由于不同模态信息来源的数据特性各异，如图像数据的特征以颜色、形状为主，声音数据以频率、振幅为主等，如何均衡捕获和利用各模态特性就成了问题的关键。通过深度学习等AI技术，可以建立有效的模态融合模型，实现跨模态特征表现的学习。

在联合推理中，我们利用处理过的多模态信息，基于逻辑推理、概率推理等方式进行分析，如：根据家中人员的移动轨迹推断其行为模式，结合气温、湿度等环境条件推测其舒适度，通过声音分析识别家中是否发生异常。

多模态决策制定

在智慧家庭大脑中，计算机视觉、语音识别与理解和无线传感网络等多种信息获取手段使得多模态联合推理成为可能。在此基础上，根据推理结果，系统能进行决策制定，如：系统可自动调整温湿度，或提醒家中某处发生异常，再或是根据家庭成员习惯自动调整家庭日常生活服务等。

综上所述，多模态联合推理与决策在智慧家庭中发挥着重要作用，能为用户提供更为智能、个性化和舒适的家庭环境。在未来，随着科技的不断进步，多模态联合推理与决策的应用将会更为广泛。

2.5.AI技术下内容的安全与合规

AI 技术下内容的安全性、可靠性、合规性与可信 AI 是大模型时代面临最大的挑战，也是技术可持续发展的基础。AI 具有强大的创作能力，可以为人们提供丰富的信息、娱乐和教育资源，同时也带来了一些内容安全和合规性的挑战，如虚假信息、色情暴力、版权侵权、个人隐私等。因此，大模型内容安全需要在模型的生命周期中，采取有效的技术和管理措施，保障内容的真实性、合法性、道德性和可靠性，防止生成的内容对社会和个人造成不良影响。

为实现 AI 模型安全合规的目的，我们提出了大模型生成内容安全的治理方法和安全审核架构：

首先，在大模型的训练和生成过程中，数据是关键的因素，影响着大模型内容的质量和风险。因此，需要对数据进行有效的筛选、清洗、标注、加密等操作，确保数据的真实性、合法性、安全性和多样性，避免数据的偏差、污染、泄露等问题。

其次，在大模型的研发和运营过程中，技术是支撑大模型内容安全的基础，决定着大模型的可靠性和可控性。因此，需要制定统一的技术规范，包括技术标准、技术指南、技术评估、技术认证等内容，对大模型的技术方法、技术流程、技术结果、技术责任等方面进行规范、指导、评价和认证，提升大模型的技术水平和技术信任。

通过建设内容审核平台，可以对用户的输入信息以及生成的内容进行安全审核，确保人工智能生成的内容应当体现核心价值观，不含有颠覆分裂国家、破坏国家统一，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情信息，虚假信息，并采取措施防止出现种族、民族、信仰、国别、地域、性别、年龄、职业等歧视等要求，所有内容的生成和最终的输出必须经过严格的过滤。

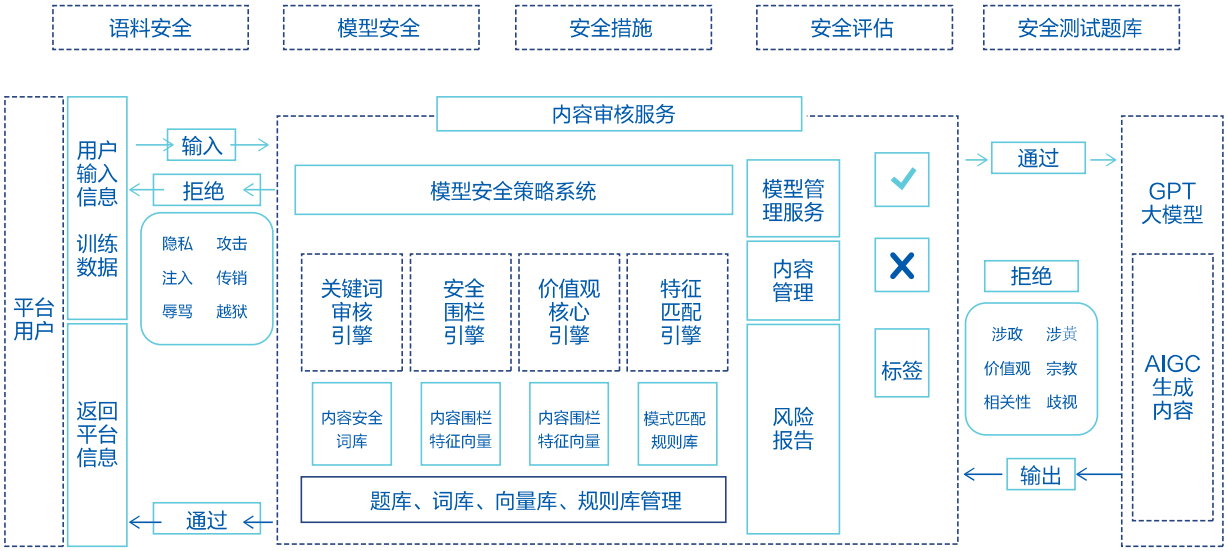


图 12 AIGC 模型内容安全的治理方法和安全审核架构

在大模型的发布和传播过程中，内容是直接面向用户的产品，决定着大模型的社会影响和用户体验。因此，需要建立完善的内容审核机制，包括人工审核、机器审核、用户举报、社区治理等环节，对大模型内容进行实时、全面、多维的监测、评估、过滤和处置，防止大模型内容的不良内容、错误信息、敏感话题等问题。

我们认为，大模型内容安全将从单一的内容审核向多元的内容治理转变，从被动的内容过滤向主动的内容优化转变，从局部的内容管理向全局的内容生态转变，从技术驱动的内容创造向价值导向的内容创造转变。大模型内容安全还面临着技术能力不足、数据资源不均、法律法规不完善、社会认知不高等问题，需要相关方共同参与、协作和负责，建立健全的大模型内容安全体系，实现大模型内容安全的目标。

2.6.大模型时代的AI伦理

与大模型技术的突飞猛进形成鲜明对照的是，大模型仍面临诸多潜在的安全风险。大模型在应用的过程中，可能会产生与人类价值观不一致的输出，如歧视言论、辱骂、违背伦理道德的内容等，这种潜在的安全风险普遍存在于文本、图像、语音和视频等诸多应用场景中，并会随着模型的大规模部署带来日益严重的安全隐患，使得用户无法信赖人工智能系统做出的决策。

面临的问题

在没有借助生成式人工智能技术的情况下，智能家电设备虽然也能与用户交互，但回复给用户的答案或用于控制家电的指令是经过筛选的有限集合。而采用生成式大模型后是否会给出一些违反人类伦理道德的答案或者操作，是值得思考的问题。

传播错误的意识形态

人工智能的目标是模拟、扩展和延伸人类智能，如果人工智能只是单纯追求统计最优解，可能表现得不那么有“人性”；相反，包含一些人类政治、伦理、道德等观念的人工智能会表现得更像人、更容易被人所接受。事实上，为了解决人工智能面对敏感复杂问题的表现，开发者通常将包含着开发者所认为正确观念的答案加入训练过程，并通过强化学习等方式输入到模型中，当模型掌握了这些观念时，能够产生更能被人接受的回答。然而，由于政治、伦理、道德等复杂问题往往没有全世界通用的标准答案，符合某一区域、人群观念判断的人工智能，可能会与另一区域、人群在政治、伦理、道德等方面有较大差异。因此，使用内嵌了违背我国社会共识以及公序良俗的人工智能，可能对我国网络意识形态安全造成冲击。

偏见与歧视

一方面，训练大模型的数据是一定时间前的历史数据，本身往往就具有倒退偏见，没有及时反映后面发生的进步；另一方面，某些训练数据本身就带有人群歧视，而且有可能会被放大。

推动路径

在数据算法安全和伦理规范方面，需要从以下几个方面推动：

制定专门的生成式人工智能安全标准

对于智慧家庭来讲，在应用生成式人工智能技术的过程中，除了要满足国内外行业中的网络安全、数据安全和个人信息保护等方面现有的法律法规和标准外，为应对生成式人工智能算法、数据使用等带来的安全新挑战，以促进生成式人工智能发展为基本目标，统筹发展和安全，亟需针对生成式人工智能的网络安全问题、数据安全和隐私保护问题出台专门标准，包括但不限于生成式人工智能训练数据安全、人工标注过程安全等方面的标准规范。

开展生成式人工智能安全检测评价

基于上述网络安全、数据安全和个人信息保护等方面的现有国内外标准，结合生成式人工智能在智慧家庭的应用，以安全结果为导向，开展检测评价，确保技术应用合法合规。针对检测评价过程中发现的问题，督促厂商及时整改；开展行业比较测试，针对较好的应用开展行业示范，促进技术应用发展。

开展生成式人工智能安全风险监测

《生成式人工智能服务管理暂行办法》自 8 月 15 日期正式实施。《办法》明确提出，“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，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管”。基于这个原则，对生成式人工智能应用过程中遇到的风险进行监测、风险评估，并针对性的动态更新监测方案，推动安全标准不断完善。

2.7.国际性法律法规

各国数据安全法

人工智能治理攸关全人类命运，是世界各国面临的共同课题。今年以来，全球多个国家和组织纷纷出台倡议或规范，一致要求加强人工智能的安全监管。人工智能告别粗放式发展，迎来安全与发展的同步阶段。

2023年11月1日，首届全球人工智能（AI）安全峰会上，28国联署关于人工智能国际治理的《布莱切利宣言》，这是全球第一份针对人工智能这一快速新兴技术的国际性声明。《宣言》鼓励相关行为者采取适当措施，如安全测试、评估等，以衡量、监测和减轻AI潜在有害能力及其可能产生的影响，并提供透明度和问责制。呼吁各国根据风险制定基于风险的政策，包括制定适当的评估指标、安全测试工具，以及发展公共部门的能力和科学研究。并决心支持建立一个具有国际包容性的前沿AI安全科学研究网络，该网络包括并补充现有和新的多边、双边合作机制，通过现有国际论坛和其他相关举措，促进为决策和公共利益提供最佳科学。

2023年6月，欧洲议会通过了《欧盟人工智能法案》授权草案，该法案如正式获得批准，将成为全世界首部有关AI的法规。该法案将人工智能系统根据风险级别分为四个分类，从最小到不可接受。其中，“技术稳健性和安全性”要求人工智能系统在开发和使用过程中尽量减少意外伤害，并具备应对意外问题的稳健能力，以防止恶意第三方非法使用该系统或进行改变其使用方式或性能的行为。此外，该法案禁止通过从互联网或闭路电视录像中无针对性地提取面部图像来建立或扩大面部识别数据库，并禁止使用这种方式将人工智能系统投放市场使用。对于基于这些模型的生成型人工智能系统，法案要求遵守透明度要求，即必须披露内容是由人工智能系统生成的，并确保防止生成非法内容。此外，使用受版权保护的培训数据时，必须公开这些数据的详细摘要。

2023年10月30日，七国集团（G7）发布《开发先进人工智能系统组织的国际行为准则》。这套行为准则共包含11项内容，强调了开发过程中应采取的措施，以确保可信性、安全性和保障性。其中，开发人员需要识别并减轻风险，包括红队测试和缓解措施。同时，开发人员还需要在部署后识别并减少漏洞和误用模式，并促进第三方和用户发现并报告问题。此外，该准则还强调了开发和部署可靠的内容身份验证和来源机制的重要性，例如水印。这些措施将有助于确保人工智能系统的安全性和可靠性，并提高用户对其信任度。

同年，也是10月30日，美国总统拜登正式发布《安全、可靠及可信赖的人工智能》行政命令，这是白宫有关生成式人工智能的首套监管规定。该行政命令要求美国多个政府机构制定标准，对人工智能产品进行测试，寻求“水印”等内容验证的最佳方法，拟定网络安全计划，吸引技术人才，以保护隐私，促进公平和公民权利，维护消费者和劳动者的利益，促进创新和竞争，提升美国的领导地位等。同时，行政命令指出，通过建立检测AI生成内容和认证官方内容的标准，从而保护美国用户免受人工智能欺诈和欺骗。

2023年10月18日，中央网信办发布《全球人工智能治理倡议》，具体措施包括推动建立风险等级测试评估体系，实施敏捷治理，分类分级管理，快速有效响应。研发主体需要提高人工智能可解释性和可预测性，提升数据真实性和准确性，确保人工智能始终处于人类控制之下，打造可审核、可监督、可追溯、可信赖的人工智能技术。同时，积极发展用于人工智能治理的相关技术开发与应用，支持利用人工智能技术防范风险，提升治理能力。此外，倡议还强调逐步建立健全法律和规章制度，保障人工智能研发和应用中的个人隐私和数据安全，反对非法收集、窃取、篡改和泄露个人信息等行为。

2023年7月13日，国家网信办联合国家有关部门公布《生成式人工智能服务管理暂行办法》。要求有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。

数据出海的注意事项

根据各国法规，针对跨境类型的服务，需要遵守以下的基本要求：

- 遵守目标国家或地区的互联网法律法规，包括数据安全、个人信息保护、知识产权、消费者权益等方面的规定。
- 了解目标国家或地区的文化、习惯、需求和偏好，尊重当地的价值观和社会风俗，避免触犯敏感或禁忌的内容提供适合当地市场的服务。
- 建立健全数据安全和个人信息保护的制度和措施，采取必要的技术手段，保障数据的安全和合法流动，防止数据的泄露、篡改、损毁或滥用。
- 获取用户的明确同意，告知用户数据的收集、使用、存储、转移和删除的目的、方式和范围，尊重用户的知情权和选择权，保障用户的隐私权和其他合法权益。
- 配合目标国家或地区的监管部门，按照法律法规的要求，提供必要的信息或协助，履行社会责任，处理用户的投诉或纠纷。

2.8.智慧家庭垂域大模型探索实践

分布式耦合能力构建

在智慧家庭场景下，需要解决交互系统中，超强的用户意图理解与训练数据泛化以及模型可控性的问题：

交互系统中用户意图理解与训练数据泛化能力问题

智慧家庭要求交互系统能在不同的场景和任务中灵活地应用和调整自己的策略和行为，以实现更自然、更智能、更人性化的交互体验。具体来说，以往的交互系统主要是以有监督学习建模为主导的交互系统（如以 BERT 为预训练模型进行有监督微调）。受训练标注数据量不足的影响，往往对自然交互中用户指令的理解不够准确。即针对多种泛化说法，原有系统难以理解用户的准确意图。同样地，受训练数据和先验知识不足的影响，交互系统回复给用户的生成语料也缺乏知识量和智能感。

交互系统的可控性问题

具体来说，以基于无监督学习的大模型为主导的交互系统（如 ChatGPT 等），其在聊天领域拥有强泛化涌现能力的同时，受神经网络不可解释性的天然因素影响，在智慧家庭等应用领域天然存在不稳定、不安全等可控性差的问题，无法有效预估交互系统对用户意图的辨识情况，同时无法控制由模型生成的内容。因此，这类系统下的交互行为往往存在不可预知的风险。另外，系统还存在大模型响应时长较为缓慢、某些特性信息无法单独建模、以及不支持私有化部署等问题。

为了提升交互系统中用户意图理解与训练数据泛化能力与可控性，家庭大脑通过对在有监督学习建模主导的可控性交互系统架构中，对耦合大模型进行多源特征处理、生成数据建库索引，以及动态反馈等机制，解决自监督学习大模型在调用过程中的输出结果不可控、响应慢及私有化部署问题。在以智能交互过程中用户意图为 GPT 模型转换目标，对原始语料自动进行转换及解析，以解决可监督 BERT 建模为主导的交互系统无法满足用户对交互系统中用户意图理解与训练数据泛化的问题。此外，还解决了利用 GPT 技术实现半自动标注工具，解决了降本增效的技术问题。

多节点分布式耦合 GPT 技术架构及引导交互核心算法

在之前的智慧家庭可控交互系统中，存在包括数据预处理、语义理解、上下文控制、对话策略管理、自然语言生成等多个功能节点。如下图所示：

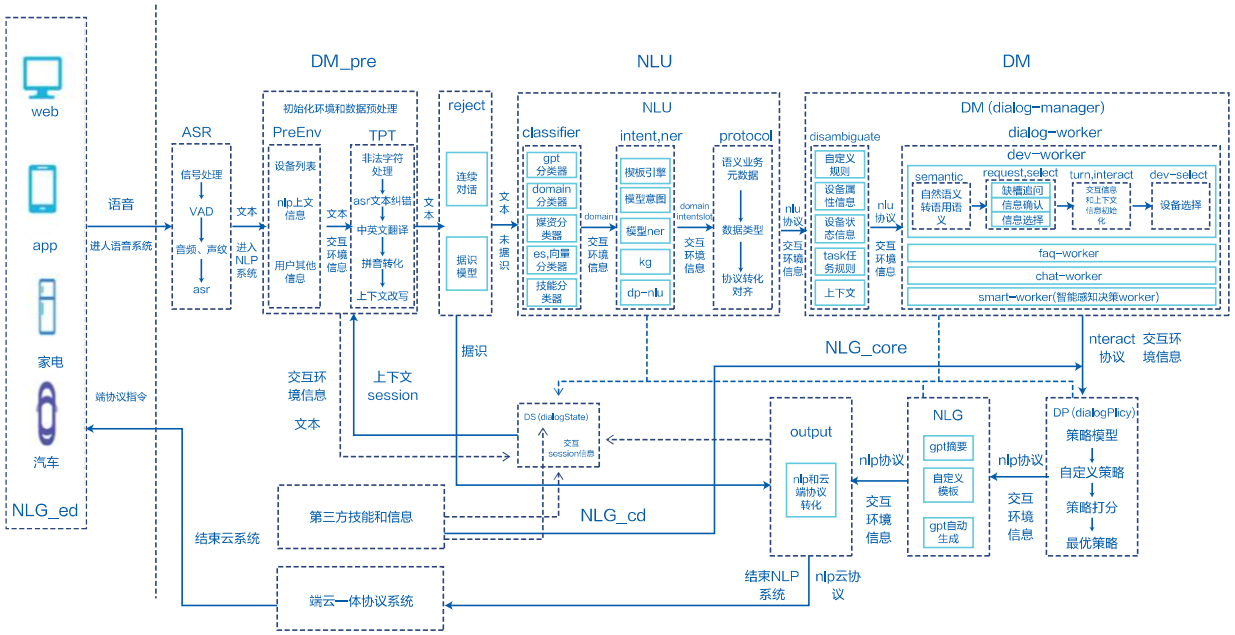


图 13 智慧家庭可控性交互系统技术架构

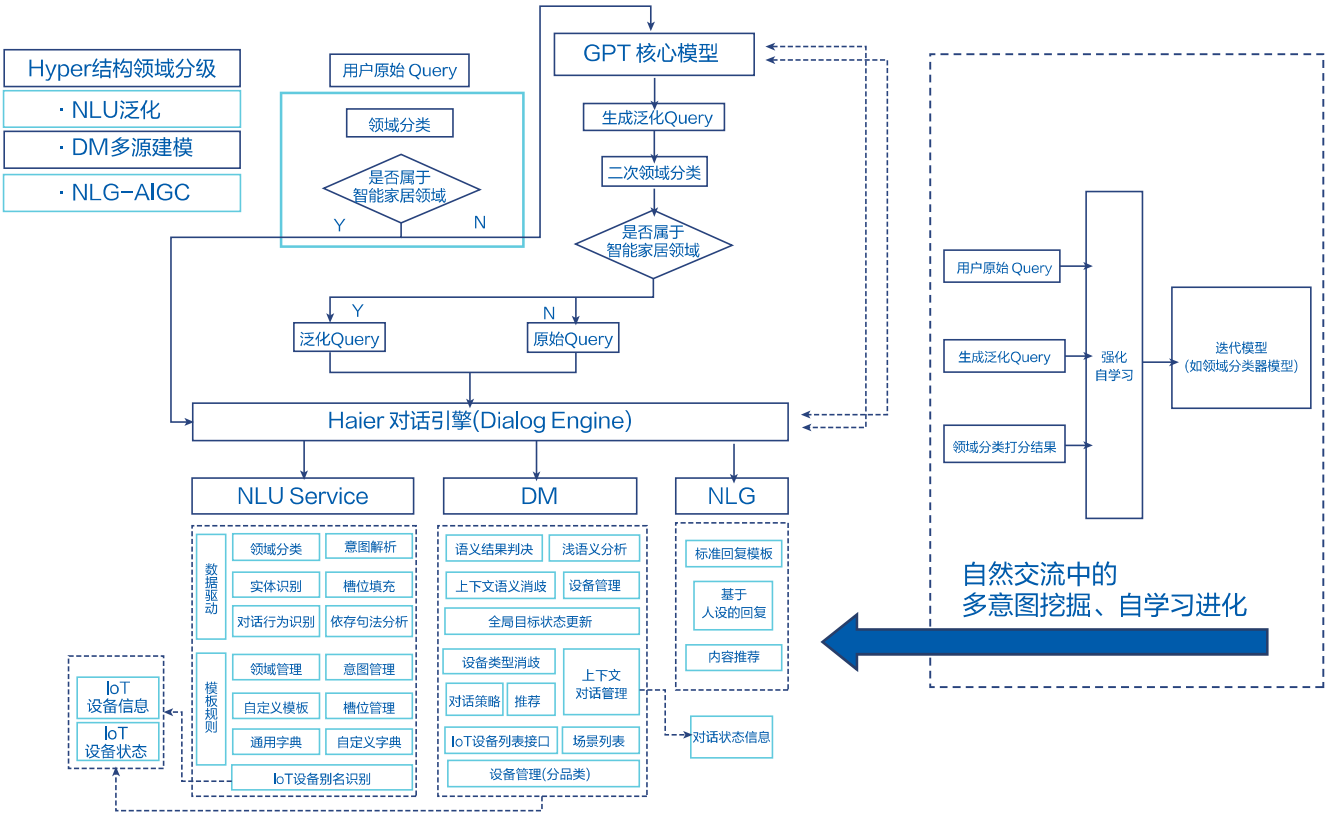


图 14 多节点分布式耦合 GPT 技术架构

为了解决用户意图理解与训练数据泛化和可控性能力并存的问题，需要将生成式模型在多个可控节点上进行有效利用。为此，在传统交互系统中的多个任务节点上，通过对大模型的分布式耦合来提升原有系统的用户意图理解与训练数据泛化和可控性能力。对于可控性部分，通过构建具有自迭代能力的架构优化，实现了端到端全链路闭环反馈及增强学习，并通过优化 NLG 架构形成了真正闭环。对于耦合部分，通过缓存检索机制的构建，以提高未来对类似对话进行解析的响应速度。总体上，在现有的多节点可控架构中，通过对各可控节点的内部进行耦合，并整合大模型输出的生成式信息，以此来保证可控性的同时提升各功能节点的泛化性和理解能力。

➢ Hyper-Structure GPT 模型解耦设计

此外，在上述多节点耦合过程中，通过构建分布式耦合并带有反馈自学习的HyperGPT模型（Hyper-Structure GPT模型），来实现在特征层面上的融合。在引导交互过程中，可用于生成数据缓存/建库/模型转换/信息链索，可拓展使用于多源信息特征融合建模及连续意图理解建模应用。

该模型由编码器和解码器组成。对于编码器，首先输入原始交互文本，并进行分层规范化（一种特征缩放技术，用于修改神经网络中层的输入数据），然后进行多头注意力操作，再输入至层规范化和前馈网络（一种人工神经网络,其结构由多个层次的节点组成,并按特定的方向传递信息），并进行残差连接（指在神经网络中,将前一层的输出直接与后一层的输入相加,从而构成了一种跨层连接的方式），得到编码特征向量并进行多层合并，得到合并后的特征向量；对于解码器，首先输入合并后的特征向量，并进行层规范化和多头注意力操作，对最后一次的层规范化结果进行前馈处理，得到各候选词的预测概率分布，最后输出各个候选词的预测概率分布对应的词语，从而得到解码结果。以此来实现对原始GPT模型拓扑结构的升级。

在HyperGPT模型解耦特征融合实验中，已在系统中的多个功能节点上进行了实践：

（1）在自然语言理解(Natural Language Understanding，NLU)领域分类的相关节点上，在编码层进行特征融合时，通过自注意力机制将领域与原始query进行自学习，将其注意力计算结果强化至相关的单元节点上，从而能够更准确地提升该槽位（用户表达意图的句子中，用来准确表达该意图的关键信息的标识,被称为槽位）的识别效果，并为原始模型注入领域知识。

（2）在命名实体识别（Named Entity Recognition，NER）节点上，通过构建NER Hyper模型来进行通用实体识别模型，来构建多节点耦合结构的应用实践。通过利用上述特征和注意力机制，并引入条件随机场（Conditional Random Field，CRF）层来构建一个设备控制通用NER模型，并基于实体识别结果和意图规则集，来推理出最终的用户意图。

（3）在句法分析节点上，通过构建句法分析Hyper模型来构建另一种多节点耦合结构的应用实践。在该模型中，首先在之前设备控制通用NER模型的基础上，引入带有Hyper结构的句法分析模型并通过解码器生成句子中的语义依存关系，进而对多意图语句进行子句的切分，从而进一步对各子句进行意图和槽位的解析。

综上，通过构建带有特征解耦的模型结构来对原有系统中各任务节点进行强化，即在可控的架构中，对各可控节点的内部进行耦合并引入生成式的信息，来保证可控性的同时提升各功能节点的泛化性和理解能力。

多源建模跨领域全双工连续语音交互技术

此前的传统语音交互中，存在体验不自然、不流畅、不智慧等问题。具体来说，每次交互需要特定唤醒词唤醒，无法自然流畅的连续交互；智能设备连续拾取的声音中，无法有效拒识干扰音频 （如播音员新闻广播等），影响真实用户的真实意图的辨识，难以满足用户足够广泛的、确切意图需要的智慧家庭交互服务。

为解决以上问题，采用了基于多源跨领域建模的全双工连续语音交互技术。针对不自然不流畅问题，基于跨领域全双工连续语音交互能力，实现了多源信息联合建模的自然语言理解拒识算法，并在全双工智能冰箱落地。针对不流畅不智慧问题，基于HomeGPT交互引擎的分发决策、动态反馈、多路GPT模型耦合，缓存建库检索机制、增强学习能力、上下文理解及强弱泛化等可控性大模型技术来进行研发迭代。

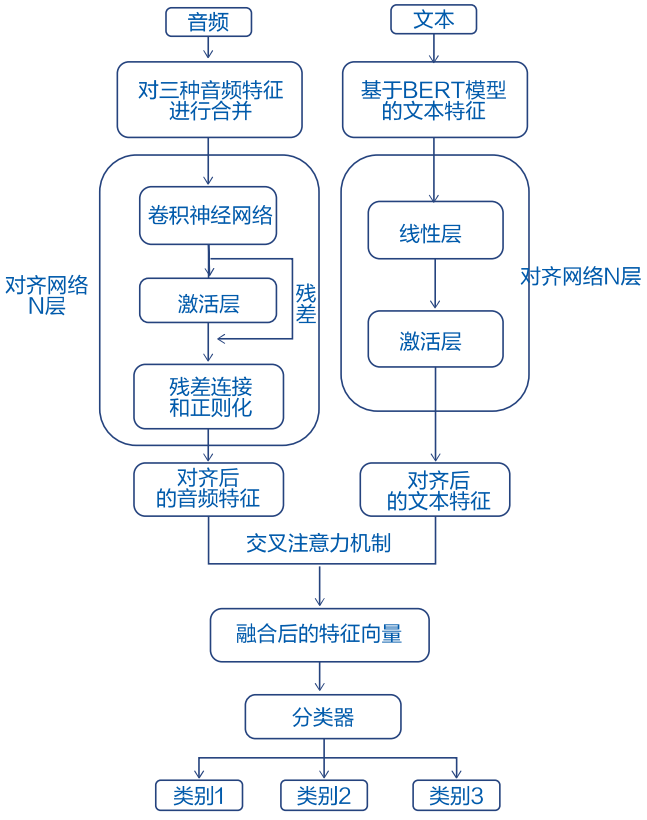


图 15 基于音频和文本多源联合建模的拒识模型结构

提出了基于音频和文本多源联合建模的拒识模型，既可以识别声音来源是否有效，也可以识别文本语义是否合理，有效提升非正常语音请求的拒识精度。其中，多源拒识模型框架设计为开放式结构，未来可以接入更多数据源如温湿度、人感、红外、图像等更多的模态信息，进一步提高模型精度。通过自研的音频与文本联合建模方法，多源融合特征提取算法有效的降低了数据维度，提高了模型训练的速度与推理精度。此外，引入了GPT的耦合决策，利用GPT强大的语义理解能力辅助现有系统更好的理解用户的非家电控制请求，在语义层面上强化对非正常请求的拒识。

提出了基于音频和文本多源联合建模的拒识模型。在该多源拒识模型结构中，多源特征融合模块基于能量谱、mfcc、spectrum3的语音侧特征，并结合BERT Embedding的文本侧特征，融合得到特征向量。此外，通过将音频的特征向量和文本的特征向量分别输入到各自的对齐网络层，得到维度一致的两个向量。使用cross-attention方法将两个特征向量融合成一个向量，最后再经过分类网络识别出音频和文本对应的类别。

上述多源建模跨领域全双工连续语音交互系统，已在智能冰箱上进行了落地实践，该系统的流程如下图所示：

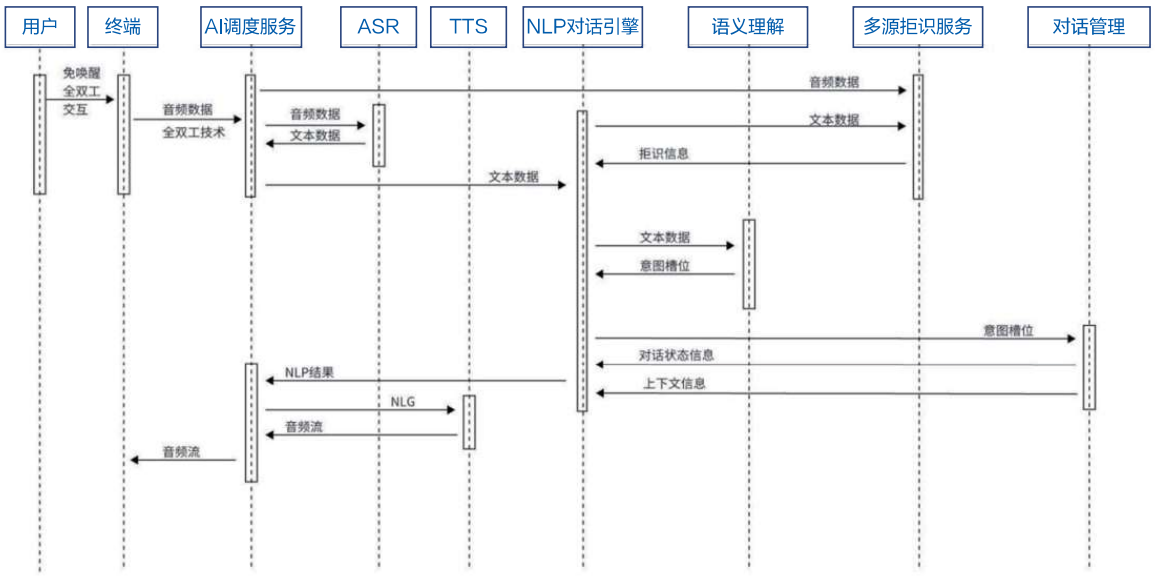


图 16 多源建模跨领域全双工连续语音交互系统

分布式耦合自动标注

此外，在数据标注任务上通过构建分布式耦合自动标注工具，实现了包括词性分析、命名实体识别、句法分析、关键词提取、知识解释等自动标注功能，以及支持同向语义和反向语义的自动泛化能力。通过该自动标注工具，已提升数据标注工作和文本泛化的80%效能。

图 17 分布式耦合自动标注示例

03 应用篇

HomeGPT赋能交互服务与场景的全面升级

随着智慧家庭领域的快速发展，用户所需要的不再局限于简单的远程控制和一成不变的固定场景，而是能够随时满足个性化需求的智能化服务，交互模式也从原来的被动响应方式转向基于家庭大脑的主动感知与主动服务方式。在此背景下，AGI将引领智慧家庭场景变革，HomeGPT以满足用户场景需求为核心驱动力，将重塑智慧家庭AI，带来不一样的家庭生活体验。

3.1.交互的升级

大模型提升了智慧家庭设备和应用的可用性，带来了全新的交互模式和丰富的用户体验，创造出更多想象空间。多模态大模型结合数字家庭知识库可以帮助用户构建出一个更加丰富、友好的交互方式，使大模型应用与人的交互过程无限趋近于人类自身的习惯。同时，多模态大模型还可以与VR/AR、元宇宙、AI Agent等技术体系或场景进一步融合，打造更深层、更多维、更丰满的全新交互体验。

可以想象，在不久的将来，大模型采用数字家庭管家作为总控制端，采用接近自然对话和VR/AR的沟通方式，采用Multi-Agent等技术架构模拟家庭成员活动/模拟家庭空间场景/模拟家电控制关系，采用Single-Agent技术结合本地知识库、本地传感器、本地家电等落地用户需要的服务，这将带来完全不同与传统APP或智能硬件的交互体验。

基于自然语言交互的家电

家庭内语音识别场景将随着大模型的赋能，以下方面将有大幅的提升：

——用户体验： 当前家庭语音识别市场，大部分以离线语音为主，普遍存在离线语音的语音识别指令过于机械化，用户交互复杂，准确度低，而且对用户的普通话水平以及指令的准确度有较高的要求，例如打开灯、关闭电视等。而家庭内在线语音识别的交互方式功能也较少，设备主要依赖于智能音箱等设备，用户体验也有待提升。而大模型赋能之后，离线语音市场份额必将降低，而集成了大模型、语音识别、语音播放的语音模组将让用户和设备的沟通更加人性化，降低对指令格式的要求，降低普通话标准程度的要求，大幅度提升自然语言理解的纠错能力。

——智能化程度： 虽然语音识别技术配合在线语音能力已经取得了很大的进步，但与人类的自然语言理解能力相比，仍有很大的提升空间，当前功能单一，内容扩展门槛高。而随着大模型赋能之后，AI Agent技术将记录人的行为、模拟人的性格、预测人的需求，这将大幅度降低智能化内容接入的要求，配合大模型无穷的知识库和本地用户历史操作信息，语音识别的内容将可大幅度扩展，拟人化程度更高，带有语音识别的家电，配合AI Agent技术，将变得更主动，更智能，更温情，更懂用户。如当用户说“我想看电影”，大模型可以根据用户画像，自动搜索并推荐用户希望看的电影，并将电视打开并切换到电影模式，同时关闭窗帘和灯光，提供更加舒适的观影环境。

——应用场景：目前语音识别技术在家庭领域的应用场景较为有限，主要是智慧家庭控制、语音搜索等基本功能。但是随着大模型的赋能，语音识别技术将告别这些指令型和搜索型的基本能力，而借助大模型“泛化”能力，自动根据用户的习惯，进行主动服务，大幅度扩展应用场景。例如，当用户想要观看某个电影或者电视剧时，只需要说出具体的影片名称或者演员名称，大模型就可以自动搜索并播放相应的影片或者电视剧。此外，大模型还可以学习用户的影音偏好和习惯，提供个性化的影音、硬件协同、随影音内容推出相应的推荐和服务，并提供配套的内容、家庭控制、购物等多场景服务。

基于上下文理解的智慧终端

在大模型的加持下，多数智慧终端（如智能音箱、智能电视、智能手表、手机和平板等），均可升级成为具备支持基于上下文理解的自然交互功能的智能终端。升级后的智慧终端能够根据用户的语音或文本输入，结合家庭环境和设备状态，提供智能化的控制、管理和服务。即便是结果不尽如人意，经过自然交互给出修改意见后，控制效果亦可准确呈现。

——环境上下文：指智慧家庭所处的物理环境和外部环境的信息，如温度、湿度、光照、空气质量、时间、地点、天气、交通等。环境上下文可以通过传感器、摄像头、定位系统等设备获取，也可以通过互联网、物联网等网络获取。环境上下文为智慧家庭提供适应性和智能化的服务，如根据室内外温差自动调节空调、根据天气预报提醒用户出行注意事项、根据交通状况规划最佳路线等。

——用户上下文：指智慧家庭的使用者的个人信息和行为信息，如身份、年龄、性别、健康状况、喜好、习惯、情绪、目标、需求等。用户上下文可以通过人脸识别、语音识别、生物识别、行为分析等技术获取用户身份，通过家电或者内容服务中的历史记录，大模型自动给出用户的用户画像。用户上下文为智慧家庭提供个性化和人性化的服务，如根据用户的身份和喜好推荐节目、根据用户的健康状况和需求提供医疗服务、根据用户的情绪和目标提供娱乐服务等。

——交互上下文：指智慧家庭与用户之间的交互方式和交互状态的信息，如交互设备、交互模式、交互内容、交互历史、交互结果等。交互上下文可以通过智慧家庭的平台、终端、接口等获取，也可以通过用户的操作、指令、反馈等获取，还可以通过AI Agent的存储技术，自动理解并存储交互上下文。交互上下文为智慧家庭提供高效和友好的交互体验，如根据交互设备的特点选择合适的交互模式、根据交互内容的类型选择合适的交互方式、根据交互历史的记录提供交互提示、根据交互结果的评价提供交互改进等。

图 18 智能交互终端

例如：用户想要在智慧家庭中观看电影，那么：**环境上下文**包括了智慧家庭的电视机、音响、灯光、窗帘等设备的状态，以及外部的时间、天气等信息。智慧家庭可以根据这些信息，自动调节设备的参数，如亮度、音量、色温等，以达到最佳的观影效果。**用户上下文**包括了用户的身份、喜好、情绪等信息。智慧家庭可以根据这些信息，为用户推荐合适的电影，如根据用户的年龄和性别推荐适合的类型、根据用户的喜好和情绪推荐适合的主题等。**交互上下文**包括了用户使用的遥控器、手机、语音等交互方式，以及用户与智慧家庭的交互内容、历史、结果等信息。智慧家庭可以根据这些信息，为用户提供便捷和舒适的交互体验，如根据用户的交互方式提供相应的反馈、根据用户的交互内容提供相关的信息、根据用户的交互历史提供快捷的操作、根据用户的交互结果提供满意的服务等。

➢ 基于数字人应用的家庭管家

得益于数字人（Digital Human / Meta Human，指运用数字技术创造出来、与人类形象接近的数字化人物形象）在文娱领域取得了快速发展。更多AI算法和Agent技术让数字人拥有专业的知识技能和服务能力，智慧家庭正是数字人发挥价值的重要落地场景之一，成为未来发展的重要方向。在面对拥有真实形象和完善AI能力的专属数字人时，人们可以投入更多的信任与情感，与传统的终端交互设备相比，沟通更加便捷，交互效率也随之提升。同时，人格化的数字人还能为用户带来独一无二的情感关怀，可以更好地帮助用户构建美好的家庭生活。

在智慧家庭当中，数字人可以是多种形态统一呈现，例如：智能音箱或智能屏的语音助手、智能电视或智能平板的视频主播、智能镜子或智能摄像头的虚拟形象等。数字人可以根据用户的喜好，调整自己的声音、语气、口音、外观、姿态、表情等，提高用户的亲切感和满意度。此外，专属的数字人结合AI Agent，可以长久地记忆用户的兴趣偏好，根据用户的兴趣，定制个性化的服务与内容。

围绕智慧家庭场景，厂商正在打造多模态交互、具备家庭服务专业知识与技能，同时拥有情感交流与陪伴属性的数字人。他们可以成为智慧家庭管家，帮助用户管理家务，根据口味推荐健康菜谱，将家庭环境调整到最佳模式，无论何时踏进家门，都无需操心琐事，只用舒适享受。他们也可以作为母婴管理专家，手把手教学，帮助孕期妈妈安全平稳地度过每一天，指导新手父母进行婴幼儿的抚育。当然，也可以成为用户的健身教练、形象管理专家、理财专家，家庭教师。

与此前语音助手等交互方式相比，数字人能呈现更加自然的人机交互、更加智能的内容生成、更加形象的虚拟展示。数字人也不断学习进化，最终可能成为家庭全生命周期的数字伴侣，数字人成员与生物人成员将共同组成家庭，在数字空间和物理空间共同形成的未来家庭空间中共同生活。

例如，海尔智家通过打造全屋智慧生活管家小优，让家庭大脑平台能力得以更具象化展现。以家庭大脑为核心支撑的小优，能作为贴心的智慧生活“管家”，跟用户像家人一样相处、像家人一样懂用户的所想所需。

图 19 全屋智慧生活管家

对用户而言，智慧生活管家能从多个方面为其提供智慧家服务。它可以一键连接家中所有的设备，把几十种家电合为一个整体，即使对着冰箱发出向电视机的指令，也可以精准识别；空调能分辨与它对话的是老人还是孩子，调整吹风的方向和温度；做饭的时候，家庭大脑不仅能提供菜谱，还会自己掌握火候和时间。

和智慧管家相处的越久，它就越“懂你”，会记得用户的各种习惯和喜好，哪怕用户对它发出一连串指令，它也能有条不紊地按顺序处理这些任务，就像真正的管家一样无微不至，让家越住越聪明，越住越舒服。

3.2.服务的升级

问答式、远程控制等基础智能和服务已很普遍，但真正的智慧生活体验仍需要服务的持续升级。技术发展最终是要为用户体验进阶而服务，在大模型时代，智慧管家式服务已全流程全方位升级，将像家人般主动陪伴，从衣食住娱细节出发，预见并满足生活所需。

知识问答

当下家庭内娱乐教育场景市场很大，但是当前的内容形式和主题相对单一，通常以儿童教育为主，缺乏针对不同年龄段、兴趣爱好的多元化选择。互动性也不足：大部分家庭内娱乐教育内容以单向传授为主，缺乏互动性和参与性，无法激发孩子们的积极性和主动性。所以，发展空间很大。

大模型赋能之后，首先，可以根据用户的兴趣爱好、学习需求等因素，为用户提供个性化的娱乐教育内容推荐，提高用户体验。还可以运用自然语言处理等技术，开发出具有交互性的教育内容，如智能问答、语音识别等，提高用户的参与性和互动性。可以对用户的学习过程和成果进行实时监控和评估，为用户提供及时反馈，提高教育效果。

其次，优质的娱乐教育资源往往集中在一线城市和发达地区，而其他地区和欠发达地区的资源相对匮乏。大模型可以通过数据分析和挖掘等技术，优化教育资源分配，使得优质资源能够更好地服务于更多用户。更何况，大模型通过自己无穷的知识库，可以有效降低对这些教育资源的依赖，降低地区之间的教育差异。

家庭管理

以互联互通带来的全屋智能基本上已经覆盖了，随之而来的服务体验，将随着在AI大模型赋能后逐步提升。大模型不仅在于其可以实现全屋智能、智能控制等基本功能，更在于其可以实现大脑决策、模仿人沟通、根据人的个性提供主动服务等高级功能。

——大模型可以实现大脑决策。大模型可以对家庭设备进行全面的数据分析和预测，从而实现更加智能化的控制。例如，通过对居室温度、湿度、光照等因素的分析，AI大模型可以结合人的习惯和个性行为，自动调整家庭设备的工作状态，实现更加智能化的控制。

——大模型可以模仿人沟通。大模型可以通过语音识别、自然语言处理等技术，实现与人类的沟通。这种模仿人沟通的能力，使得人们可以更加自然地与家庭智能进行交互，不再需要通过机械化的操作来实现控制。

——大模型可以根据人的个性提供主动服务。通过对人类行为、偏好等因素的分析，大模型可以为不同的人提供个性化的服务。例如，对于爱好阅读的人，大模型可以根据其偏好推荐相关的图书，让家庭智能更好地适应人类的需求。

在海尔智慧家庭全场景服务中，基于大模型全屋空气环境可主动净化，时刻清新；观影时灯光随电影画面律动，身临其境；语音洗衣可以自动推荐程序，远程升级丰富面料洗护方案等。

——全屋空气可视可说：通过家庭大脑，全空间全时段6维空气数据实时监测，并可自感知、自调节，同时可开启观影空气场景、瑜伽空气场景、智慧除湿等场景，打造恒温、恒湿、恒净、恒氧的居家生活环境。

图 20 AI 智慧洗

——AI 智慧洗：根据水质、水温、衣物面料等的不同，自动匹配适宜的洗护程序，洗衣液进行智能投放，实现专衣专护。

图 21 全屋空气管理

——智慧舒适家系统：一句话定制好空气，空调会根据用户指令自动定位用户所在地域、季节，结合室内空气质量状况、使用人群，制定到不同地区、不同人群一年四季的空气调节方案，自动进行空气管理。

图 22 智慧舒适家系统

能源管理

家庭内能源管理主要包括水电气管理和光伏等清洁能源管理。在当前市场中，智能化所能带来的功能还非常有限，主要原因在能源管理和人的生活息息相关，需要保证家庭成员的生活体验，这就导致了一些常规化的智能手段很难奏效。

大模型赋能后，将能够更好地应对这些挑战和问题。大模型以及通过大模型管理的家庭系统，更好地识别和处理复杂场景，提高水电气的使用效率。例如，知道了家庭人员的生活习惯，提前设置冰箱制冷、热水器设置、空调频率调整等。还可以和外界能源生态打通，例如，获取电网补贴，通过错峰用电，调整用电，系统光伏和储能手段。更好地适应用户行为和环境因素，提高家电设定、燃气使用的经济型、安全性和便捷性。还可以通过人脸识别和指纹识别技术，因人而异的提供个性化的能源服务。

健康养生

随着人们生活水平的提高，对于健康养生服务需求增长迅速，尤其是中老年人、女性和儿童群体。而且健康市场细分日益明显，家庭内健康养生市场开始出现细分的趋势，比如瑜伽、健身、营养食谱、睡眠等各个领域都有专门的市场。但是这个市场，因为人的体征参数和服务的内容参数没有做到很好的拉通，导致健康市场百花齐放，但是良莠不齐，更何况健康解决方案的试错的成本很高，亟待科技改善。

大模型在医疗领域的应用为智慧家庭的健康养生提供了全新的可能性。结合物联网操作系统，融合分布式感知和多模态大模型，可以全方位检测和分析家庭环境数据。比如智慧家庭系统可以实时监测室内空气质量，通过大模型的深度学习能力准确识别并分析空气质量，在检测到空气质量不佳时，可以自动触发空气净化设备，提高室内空气质量。

通过学习家庭成员的健康数据，可以为每个成员生成个性化的健康建议，甚至及时发现潜在的健康风险。结合大数据和知识图谱，智慧家庭将成为家庭成员健康管理的智能助手。

智慧家庭系统可以将健康信息展现在家电上（如智能电视或智能镜子），能够提供健康建议、疾病预防知识和定期健康检查提醒。通过语音交互，家庭成员可以随时获取专业的健康咨询，使智能家电成为家庭健康的个性化专业顾问。

智能冰箱可以实现更为智能的食品管理。根据家庭成员的健康状况和个性化的饮食需求，系统可以生成合理的饮食建议。冰箱内置的摄像头可以识别冰箱内的食物，并提醒成员关注食材的新鲜度，能够根据食材推荐菜谱，同时智能购物清单的生成也可以依托大模型的学习和生成能力，确保家庭购物更加健康、全面。

全屋用水管理，通过家庭大脑屏，可以实现全屋水质监测、水量统计，水质自洁、设备保养、滤芯更换等主动服务提醒，确保用户用水健康。

图 23 全屋用水管理

3.3.场景的升级

基于家庭大脑垂域大模型开发的场景生成引擎，通过三重融合辅助决策的场景生成技术，能进行精确的指令执行，一句话就能理解用户意图、生成个性场景。另外，场景生成引擎通过技术升级，还将实现更主动的分析和推理决策，根据模糊意图创建场景，甚至根据用户行为和画像，主动创建场景，从而让 AI 更理解用户，让场景技术开发更关注人的感受。

图 24 场景智能生成

智慧场景生成

场景创建难度高，尤其是通过语言无法自动创建家庭智慧场景，一直被用户诟病。传统模式下，用户需要使用类似编程的方法，设置自己想要的智慧场景，但因流程复杂和步骤繁琐，往往让智慧家电又变成了传统电器。

当前以语言大模型作为生成模型，基本具备了支撑一句话自动编排自定义场景的能力。通过AI场景创建工具，只需一句话，就能轻松帮用户创建专属场景。

例如，通过智慧家庭APP，基于场景生成引擎，智慧场景使用更加便捷。不仅是单独的居家场景，在全屋智控解决方案里，只需要用户一句话，HomeGPT自动识别用户意图，自动编排复杂场景，将全屋智能场景交互的编排时长，从分钟级缩短到秒级，真正实现了场景创建零门槛。

即使语言指令比较简单，也可精确的理解用户意图，比如当用户说：“每天早上七点。开灯，打开窗帘，打开空调。”APP 便会为用户生成这个专属的场景，用户还可以为这个场景定义名称或者进行微调；而在复杂语言指令下，同样可以轻松理解用户所需，如“当我说‘我回来啦’，打开客厅灯，再为我播放一首欢快的音乐。”经过对语音指令的深度解析后，APP 便可为用户创建一个特定条件特定空间的联动场景；在主动判断指令情景下，比如当用户说：“当室外温度高于 30℃时，为我主动打开客厅的空调。”场景生成引擎可以为用户获取并解析所在地的相应数据实时变化，为用户生成主动感知、自动触发的智能场景，让用户享受更主动的智慧生活。

HomeGPT 已经可以深度理解语言、理解生活、理解用户，实现更自然的人机交流、更高效的智慧控制，并能依据用户喜好，提供个性化场景定制，为用户带来真正的智慧便捷生活。

图 25 场景自动编排

声音音乐生成

在科技与内容加速融合的背景下，AI技术为音乐内容生产助力正在成为一种趋势。在以前，音乐创作需要用户具备专业的音乐制作能力，用户在家庭中想要制作“独创”的专属音乐是一个很大的挑战。

而现在，用户只需要给出想要音乐的风格、韵律、节奏等信息，AI 音乐生成模型就能够根据用户描述生产用户想要的音乐。比如选择“助眠白噪声”，根据当前用户的内部和外部条件，如地理位置、时区、天气和心率，AI 音乐模型会以此为基础生成对应的高斯分布白噪声，演奏不同的乐曲。比如提示“轻缓音乐”，适配当前用户风格和情绪的音乐会自动生成并上传到云端后，用户就可以在 APP 端收藏和播放生成的音乐。随着音乐轻轻起伏，微调的个性化声音逐渐趋向用户的偏好，抚平心境，带来生活的愉悦。

AI 音乐模型采用的是 transformer 语言技术，将用户描述文件转化成 token 序列，再有音频生产模型将 token 序列转化为波形数据，就是用户听到的音乐。AI 音乐生成解决了长期以来音乐内容制作困难的问题，降低音乐创作的门槛，让更多没有经过专业音乐培训的爱好者们参与进来。

图 26 声音音乐生成

健康菜谱生成

当前市场环境多变，健康生活理念深入人心，而传统厨房给用户提供的大都是由一个个家电拼凑出来的空间，传统模式给用户提供的也是一成不变的标准菜谱，不会聚焦用户真正健康生活下的“使用场景”。

如何满足全家不同口味？如何吃的更健康？面对一日三餐的种种难题，在大模型的助力下，家庭大脑正在通过打造多种美食场景解决方案，提升用户的美食生活体验。

家庭大脑可根据用户的身体状况、口味偏好、饮食习惯等用户画像，主动学习，精准推荐菜谱，并依托系统智能统筹烹饪，可进行任意菜品组合规划最短烹饪路径，30 分钟就能搞定四菜一汤；同时，家庭大脑依托专属 AI 烹饪流算法，能“因材施教”，针对新手全流程语音指导，逻辑可视化，边做边学，百问百答，随时打断，随时提问，好比星级大厨在旁；老手则有极简模式，重点关注流程和设备控制，通过辅助烹饪，管家式服务，全流程辅助，学习用户烹饪习惯，越用越智慧。

基于家庭大脑能力的不断迭代，智慧厨房的场景体验也在不断升级。不仅菜谱不断更新、上万道烹饪技法、全流程可视可说，而且智慧厨房会越来越懂用户的饮食习惯、烹饪习惯等。通过家庭电脑与设备的联动，燃气灶开多大火、烤嫩一点还是脆一点等不用设置即是最佳状态，做完饭菜之后可以自动消毒，一切都被安排的井井有条，饮食健康和厨房卫生问题将不再是用户的困扰。

图 27 智慧烹饪

装修设计生成

“设计是家的灵魂”，在这个越来越注重风格和个性的时代，很多人的新家往往是从一张设计图开始。但是设计一个满意的家并不简单，设计周期长不说，设计风格还要有新意，最难的是，需要把自己的想法和喜好精准传达给设计师，过程中一旦有纰漏，最终的呈现效果往往会不及预期。

而在很长一段时间里，室内设计都是围绕“前装”来展开的，包括水电走线、光线动线、家具家庭等，唯独不包括家电。因为用户一般会先将房子装好、家电再入场，设计师只要简单地预留出电源、空间即可。这也导致一个现状：多数设计师懂家装、懂搭配、懂美学，但不懂家电。而智能家电作为打造“智慧家”的重要部分，不仅需要专业安装、配网来保证体验，更要跟家装家庭风格统一、严丝合缝，为美学服务。这就需要设计师来帮助完成。但摆在眼前的问题有三点：一是没有相关经验，不敢贸然推荐；二是没有资源和工具，无法在图纸上呈现全屋智能的整体效果；三是交付标准不统一，环节冗长繁杂，风险和意外频发。

如今，在智慧家庭垂域大模型 HomeGPT 加持下的设计工具，已在装修设计场景完成了突破。设计工具能够面向用户和设计师提供 AIGC 生成能力及海量智能家电 3D 模型和设计模块、场景方案。用户可以通过拍摄家庭的毛坯图，一键生成维持房屋结构不变的、不同风格的装修效果图，也可以通过户型图一键生成不同风格的 3D 装修效果图，可以像“堆积木”一样尝试、选择、定制整屋方案，家电自匹配。全屋水电管路场景方案设计，所见即所得，在提升设计的人工效率同时，让设计定制方案更丰富，全面打造用户个性化的家。

基于设计工具行业领先的设计能力，还可以将家装、家庭、家电、设计师、门店等全链路的设计资源实现充分整合，让过去单一的硬装和软装，变成了整个家的设计方案，为用户、设计师、门店、生态方带来持续的价值增值。

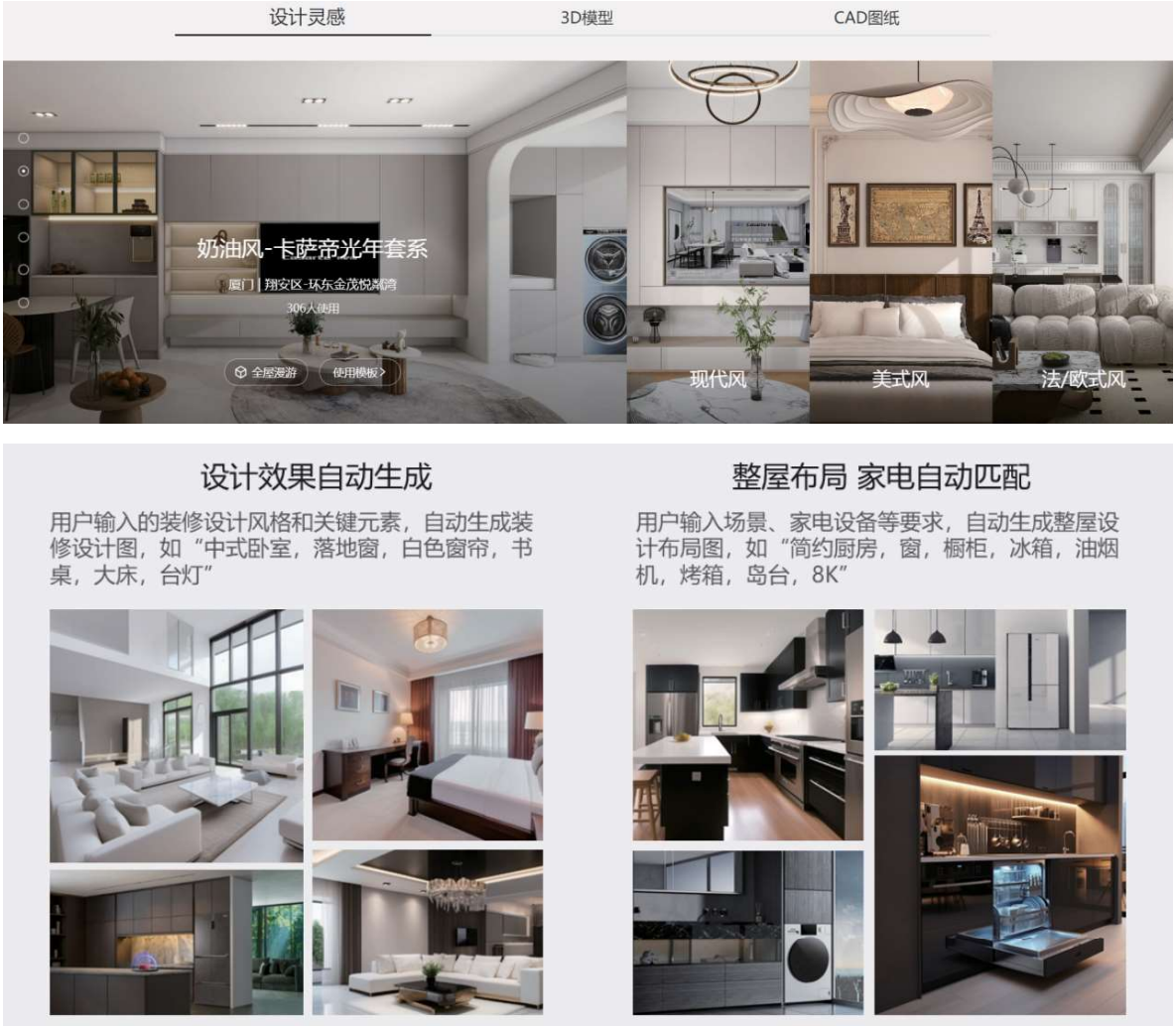


图 28 装修设计生成

04 展望篇

AGI促进智慧家庭全面发展

AIGC技术延伸扩展了人脑智能，降低了使用门槛。而面向家庭领域的AGI则能实现智慧家庭的特定任务，为家庭生活带来更多的便利、舒适和个性化。未来，AGI将与家庭医疗、家庭教育、家庭娱乐、家庭安全等方面巧妙吻合，推动相关技术迅速落地。在产业模式方面，家庭大脑也将由家庭辅助发展到家庭主导，推动智慧家庭的普及和升级，AIGC技术在知识量、信息获取和处理方面的强大能力，将使智慧家庭变革现有的技术进行全面升级、多模态服务机器人等带来全新人机协同生活。在主体方面，实现家庭智能化管理的同时，也挑战传统的家庭结构和价值观：在载体方面，“大模型分析+多模态交互+分布式终端”将使AGI与智慧家庭场景深度适配，决策无感化进程加速演进。在内容方面，跨领域融合能力、高阶内容的生成能力将越来越受到重视，服务主动化和个性化特征愈发明显。

AGI 在智慧家庭中的应用

领域AGI因其特性，能够更好的推动智慧家庭的全面发展。每一个功能具体、独立且精确的AGI都可以在某一方面优化家庭生活。例如，某个领域AGI可能专门负责控制家庭温湿度以提供最优的居住环境，另一个AGI可能在识别和回应家庭成员的语音命令方面表现优异。许多独立的领域AGI可以在特定的任务或领域中，为家庭提供量身定制的智能服务。

AGI 对智慧家庭全面发展的促进作用

AGI的存在可以从以下几方面促进智慧家庭的全面发展。首先，由于领域AGI在某个特定领域的强大能力，可以提高家庭中的效率和效能。然后，它可以提供更加个性化的服务，根据家庭成员的需要调整自己的工作方式。最后，AGI因其自动化程度较高，可以让人们从日常琐碎的任务中解脱出来，释放更多的时间和精力做自己真正喜欢的事情。

总的来说，随着人工智能技术的快速发展，未来智慧家庭的全面发展必将以AGI为重要支撑。AGI将为智慧家庭带来更多的智能化、个性化和自动化的服务，从而大大提高我们的居住环境和生活品质。

结语

随着全球 AI 算力的提升和计算成本的下降以及人工智能技术的快速发展，可以预见在不久的将来，人工智能将在人们生活中无处不在。而从现在类似 Sora 这样的模拟世界大模型的出现与改进，算法与模型框架的进步，人工智能必然会成为改变未来生产力的主要技术，将在全球各行各业掀起基础技术架构的颠覆。智慧家庭行业有望借助人工智能、云计算、世界大模型、5G 通信等技术的发展全面进入智慧家庭 L5 阶段（泛在智慧生活），实现科幻电影中人们对于未来智慧生活的想象。

新一代的智慧家庭大脑具备更高水平和更深度的智能化交互创新，告别被动指令，加强主动服务能力，满足用户个性化需求，实现高效灵活的家庭管理。还能发挥类似人类伴侣的功能，提供陪伴和情绪价值，与用户建立更紧密的情感连接。人工智能的潜力正在逐步释放，预示着未来智能时代的到来。然而，与此同时也需要关注与之相关的社会、伦理和技术挑战，确保 AI 技术的发展能够造福全人类。

希望本白皮书能够为相关从业者提供有益的参考和启示，共同推动智慧家庭领域的发展和进步，进一步探索大模型在智慧家庭领域的应用价值。相信在各方的共同努力下，能让科技的力量造福人类，创造更加美好的生活！