

2023 年版

《知识图谱与大模型融合 实践研究报告》

中国电子技术标准化研究院
知识图谱产业推进方阵
全国信标委人工智能分委会知识图谱工作组

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

前言

为推进知识图谱与大模型在企业级的落地应用，分析知识图谱与大模型融合技术路径，研究报告从知识图谱与大模型落地面临的瓶颈出发，分析了知识图谱与大模型的主要特征、知识图谱与大模型擅长的主要场景和核心基础能力，对比了知识图谱与大模型的优劣势，进而从技术演化层面、技术互补层面、知识库建设层面探讨了知识图谱与大模型融合的可行性及收益。

同时，研究报告分析了知识图谱与大模型融合的技术路径及其关键技术，研究了知识图谱与大模型融合系统评测体系，对比了实际融合系统与大模型的性能测试结果。最终，通过梳理已有11个领域的实践案例，给出了技术挑战与发展展望。

转载、摘编或利用其它方式使用本报告文字或者观点的，应注明来源为“中国电子技术标准化研究院”或对应案例提供单位，且不得对本报告进行有悖原意的删减与修改。由于知识图谱与大模型技术发展迅速，研究报告编制时间和作者学识限制，恐有纰漏或不严谨之处，敬请谅解和批评指正。

研究报告编写组

参编单位及人员

中国电子技术标准化研究院	郭楠、韩丽、李瑞琪、李湘、胡成林、陈艳利
中国电信股份有限公司研究院	石晓东、赵龙刚、孙佩霞
南京柯基数据科技有限公司	杨成彪、吴刚、魏爱梅
北京海致科技集团有限公司	瞿珂、李思宇、胡嘉彦
中译语通科技股份有限公司	陈自岩、彭旋
沈阳东软智能医疗科技研究院有限公司	程万军
北京文因互联科技有限公司	张屹、李亚军
中电科大数据研究院有限公司	曹扬、孔德智、熊子奇、尹杨、闫盈盈
北京京航计算通讯研究所	马静、郝创博、白洋、张彤
中科知道（北京）科技有限公司	吴章生、李海英、王海波
北京中企智造科技有限公司	蔡志伟、张燕
浪潮软件科技有限公司	张峰、王珂琛
杭州海康威视数字技术股份有限公司	姜伟浩、赵宏、吴炎、吴鹏亮
广州柏视医疗科技有限公司	刘涛、颜子夜
豪尔赛科技集团股份有限公司	张丰、刘姝、戴聪棋
电科云（北京）科技有限公司	方正、王尚帅
云从科技集团股份有限公司	李军

网智天元科技集团股份有限公司	贾承斌
厦门渊亭信息科技有限公司	洪万福、潘璐阳、朱成忠
国际商业机器（中国）有限公司（IBM）	初德高
青岛海尔科技有限公司	王先庆、鄂磊、鞠剑伟
浪潮电子信息产业股份有限公司	李仁刚、贾麒、范宝余
北京三快在线科技有限公司	黄坤、刘瑾、李轩
深圳市矽赫科技有限公司	洪鹏辉、洪宝璇、林叠守
同方知网数字出版技术股份有限公司	万敏锋、相生昌、周永
中国电力科学研究院有限公司	徐建南、徐会芳、张英强
浙江创邻科技有限公司	周研、马超
湖北汽车工业学院	龚家元
泰瑞数创科技（北京）股份有限公司	刘俊伟、罗伊莎
国电南瑞科技股份有限公司	张万才 石超 施雨
南京航空航天大学	周福辉、袁璐、宋熙
富泰华工业（深圳）有限公司	史喆、张学琴
中国南方电网超高压输电公司	李强

各章节编辑：

第一章	中国电信股份有限公司研究院 石晓东
第二章	网智天元科技集团股份有限公司 贾承斌

第三章	南京柯基数据科技有限公司 杨成彪
第四章	厦门渊亭信息科技有限公司 潘璐阳
第五章	中国电子技术标准化研究院 李瑞琪

更多免费行业报告

并购家

www.ipoiipo.cn

参编单位及人员



中国电子技术标准化研究院
China Electronics Standardization Institute

Wiseweb®
网智天元



Haier 海尔



浪潮信息



GTCOM



美团

cnki 中国知网
www.cnki.net
中国知识基础设施工程



国家电网
STATE GRID

中国电力科学研究院有限公司
CHINA ELECTRIC POWER RESEARCH INSTITUTE

KG DATA

矽赫 SIHERIA
Sensing the real world.



航天赛博
CASICYBER



Neusoft

浪潮科技
Inspur Technology

HIKVISION

CEIC 电科云



MEMECT
文因互联

SmartEarth

泰瑞数创

HES
豪尔赛集团
HES GROUP



创邻科技
CREATE LINK

柏视医疗
PVmed



中企智造

中科知道

NARI



南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

CEIC

中电科大数据研究院有限公司

FOXCONN®

云从科技
CLOUDWALK

更多免费行业报告

并购家

www.ipoipo.cn

第一章 背景

1.知识图谱的定义与发展历程——知识图谱的定义

以结构化形式描述的知识元素及其联系的集合。[1]

知识图谱以结构化的形式描述客观世界中概念、实体及其关系，将互联网的信息表达成更接近人类认知世界的形式，提供了一种更好地组织、管理和理解互联网海量信息的能力。[2]

国家标准
及研究报
告

知识图谱
Knowledge
Graph-KG

百科

学者/机
构

知识图谱本质上是一种叫作语义网络的知识库，即一个具有有向图结构的知识库。[3]

知识图谱旨在建模、识别、发现和推断事物、概念之间的复杂关系，是事物关系的可计算模型。[4]

Farber：知识图谱是一种资源描述框架（RDF）图，可用于描述任何基于图的知识库。[5]

谷歌：知识图谱是一个知识库，其使用语义检索从多种来源收集信息，以提高Google搜索的质量。[6]

维基百科：对事实和数字的组合，谷歌将其用于为搜索提供了上下文意义。谷歌于2012年推出，使用维基百科、维基数据和其他来源的数据。

百度百科：在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。

●[1]GB/T 42131-2022《信息技术 人工智能 知识图谱技术框架》

●[2]中国中文信息学会语言与知识计算专委会，《知识图谱发展报告(2018)》

●[3]漆桂林,高桓,吴天星.知识图谱研究进展[J].情报工程,2017,3(1):004-025

●[4]王昊奋,漆桂林,陈华钧.《知识图谱:方法,实践与应用》[J].自动化博览,2020(1).DOI:CNKI:SUN:ZDBN.0.2020-01-014.

●[5] L. Ehrlinger and W. Wöß, "Towards a definition of knowledge graphs," SEMANTICS (Posters, Demos, SuCESS), vol. 48, pp. 1-4, 2016.

●[6]<https://blog.google/products/search/introducing-knowledge-graph-things-not/>

知识图谱与传统知识库

相比具有的三大特征



图结构化形式

可呈现为有向图结构化的形式

高效的检索能力

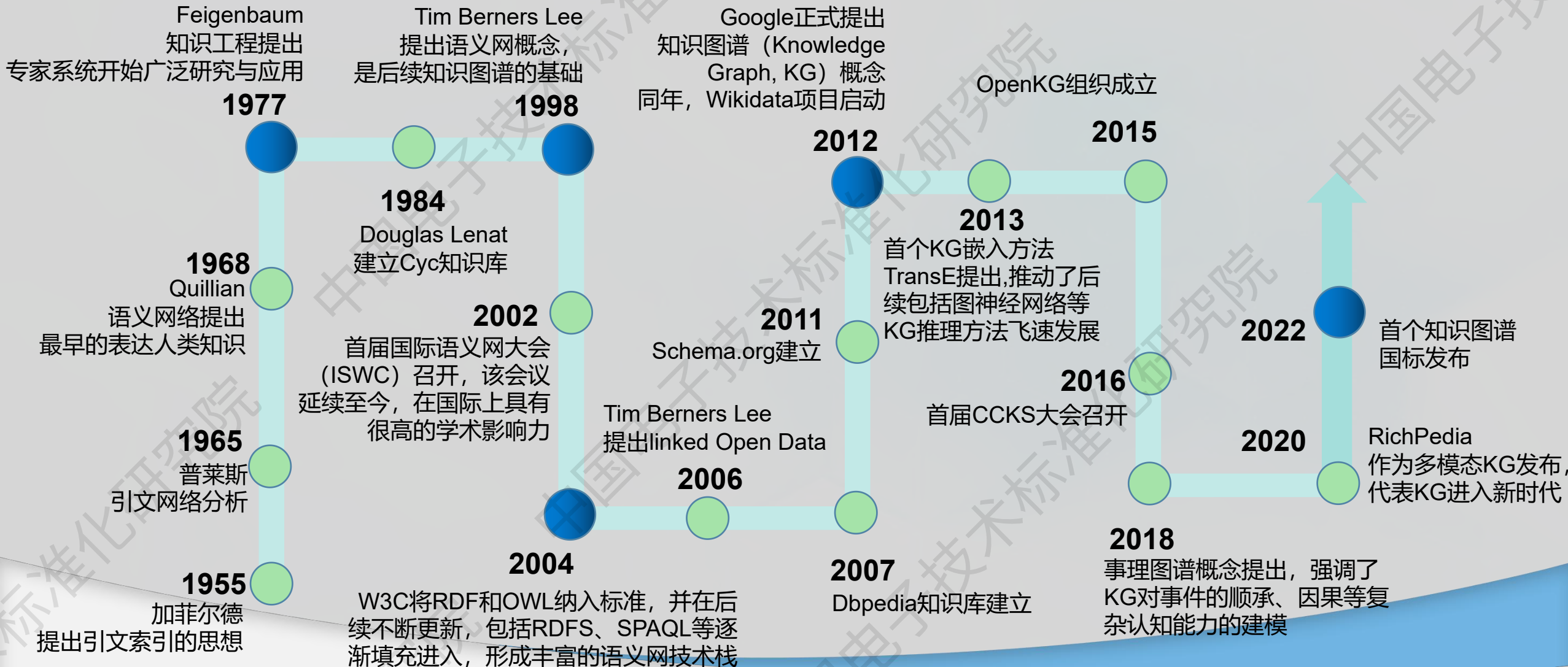
可将概念、实体及其关系结构化组织起来，具有高效检索能力

智能化推理能力

可从已有知识中挖掘和推理多维的隐含知识

附1：海外学者在知识图谱领域相关研究

1.知识图谱的定义与发展历程——知识图谱发展历程



2.大模型的定义与发展历程——大模型的定义

本研究报告中大模型的定义

广义上：

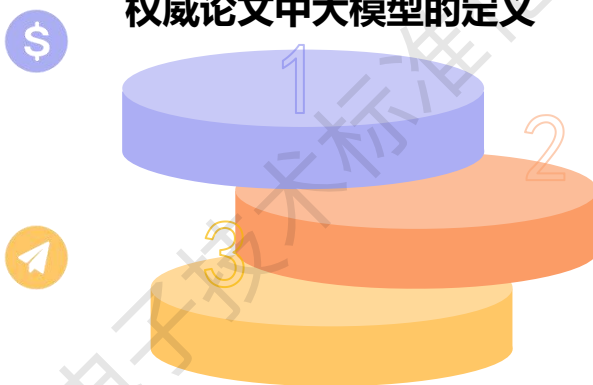
- 大模型是指参数数量大、结构复杂的深度学习模型，具备涌现能力、通用能力，并能够处理复杂的下游任务，如自然语言处理、图像识别等。
- 大模型是一种能够处理图像和文本输入，并生成文本的大规模多模态模型，主要应用于对话系统、文本摘要和机器翻译。^[1]
- 大模型也叫基础模型（FM，Foundation Model），这类模型基于广泛的训练数据，通常采用大规模自我监督学习的方法，通过提示、微调适应广泛的下游任务。^[3]



狭义上：

- 大模型通常是指参数规模在一百亿（10B）以上，使用大规模的训练数据，具有良好的涌现能力，并在各种任务上达到较高性能水平的模型。^[2]

权威论文中大模型的定义



- 大模型又称大语言模型（LLM，Large Language Model，包含数百亿或数千亿个参数，并且具备涌现能力。^[2]

大模型与传统模型相比具有三大特征^[2]

具有涌现能力

在特定任务上，随着模型规模提升
模型性能突然出现显著提升

参数规模庞大

参数规模不少于十亿（1B），严格意义上需超过一百亿（10B）^[2]

具有通用性

能够仅通过提示、微调适应
广泛的下游任务

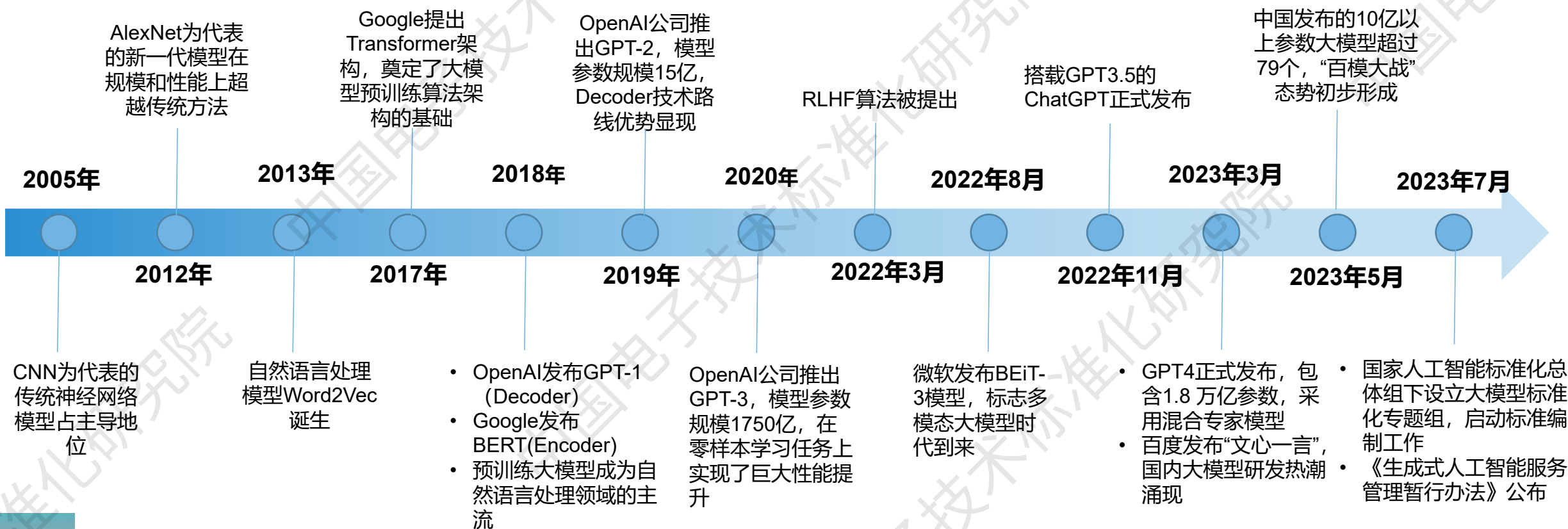
[1] OpenAI, GPT-4 Technical Report, arXiv:2303.08774.

[2] Wayne Xin Zhao, *et al.*, A Survey of Large Language Models, arXiv:2303.18223.

[3] Rishi Bommasani, *et al.*, On the Opportu

注：本研究报告所提及的大模型不仅局限于大语言模型。

2.大模型的定义与发展历程——大模型的发展历程



3.知识图谱落地面临的瓶颈

1.语料数据标注效率低、主观性强

语料数据标注仍大量依靠人工，存在标注效率低、主观性强等问题

2.知识抽取质量,难以保证

知识抽取规则的构建仍主要依赖人工，主观性强，导致可移植性差和误差传播，使得知识抽取质量难以保证

3.语义理解和自然语言处理难度大

知识图谱在面对自然语言中的语义歧义、上下文理解、语言常识推理等问题时，仍缺乏有效的解决办法

4.本体构建难度大

本体构建对领域专业知识和构建经验要求高，实体与关系的标识和对齐、本体扩展和更新、本体评估和质控、不同本体融合等方面仍面临技术挑战

5.知识通用性不足

企业级知识图谱平台及其知识内容具有较强的行业属性和领域专业性，通用性和迁移泛化能力尚有不足，跨行业、跨领域规模化应用有待提升

6.知识完备性不足

企业级知识图谱构建中通常面临领域边界限制、企业内数据规模有限、数据中知识稀疏等问题，导致其知识完备性不足

4.大模型落地面临的瓶颈

1.训练大模型的成本高

大模型的训练和优化需要大量的算力资源和海量的数据资源，涉及高性能硬件设备、强大的分布式计算能力、数据治理与融合等，投入成本巨大



2.训练数据的规模和质量不足

面向特定领域、多应用场景的高质量中文语料规模和质量不足



3.训练过程的可控性差

大模型的黑盒问题使得其推理过程很难得到合理的解释和有效的控制，增加了大模型优化的难度，并限制了其在部分领域的应用



4.输出的可信度不足

大模型的输出结果是根据概率推理而生成，具有随机性和不稳定性，导致其正确性的验证难度大，难以保证结果的准确可信



5.输出的安全性不足

大模型的开放性导致其存在信息泄露、数据攻击的风险，影响输出结果的鲁棒性和安全性



6.知识更新的实时性不足

大模型训练新数据、获取新知识的周期较长，且成本较高，导致其数据更新的滞后和知识时效性的不足



7.领域知识的覆盖率不足

GPT等大模型对各领域专业知识的覆盖仍不足，对专业问题的回答尚无法令人满意



8.社会和伦理问题隐现

大模型的输出可能存在与社会和伦理要求相悖的内容，如：生成内容消极、负面，具有破坏性等





附1：知识图谱领域国内外学者及相关研究

知识图谱国内外研究学者：

- Gerhard Weikum, 德国萨尔布吕肯Max-Planck信息学研究所
- Tom M. Mitchell, 卡内基梅隆大学计算机科学学院最高级别 E. Fredkin 讲席教授
- Ian Horrocks, 英国牛津大学计算机专业教授
- 唐杰, 清华大学教授
- 李涓子, 清华大学教授
- 漆桂林, 东南大学教授
- 陈华钧, 浙江大学教授
- 王昊奋, 同济大学教授
- 刘峤, 电子科技大学教授

知识图谱领域国内外学者的相关研究：

- Gerhard Weikum 研究知识获取表示、分布式信息系统、数据库性能优化与自主计、算、信息检索与信息提取等；
- Tom M. Mitchell 的研究涵盖知识表示、知识库构建、机器学习、人工智能，机器人和认知神经科学等；
- Ian Horrocks 的研究涵盖述逻辑、语义网络、知识表达、知识库、网络本体语言等方向；
- 唐杰研发出研究者社会网络 ArnetMiner 系统，唐杰的高引用论文是 2008 年在 KDD 会议上发表的“ArnetMiner: extraction and mining of academic social networks”对其负责的知识工程实验室 ArnetMiner 系统关键问题进行讨论，整合来自在线 Web 数据库的出版物并出一个概率框架来处理名称歧义问题；



第二章

知识图谱与大模型融合的可行性

1.知识图谱与大模型的对比——典型应用场景层面

场景名称	场景描述	大模型	知识图谱
智能对话	通过自然语言理解和生成，实现人机或机机之间的交互沟通，例如智能客服、智能助手、聊天机器人等。此场景需要能够理解和反馈自然语言，以及维护对话上下文。	😊	
内容生成	通过自然语言生成或音视频生成等技术，根据给定的输入或条件，生成新的文本、图像、代码或者其他形式的内容。此场景需要具备需求理解和相关知识融合的能力。	😊	
内容加工	此场景指通过技术手段处理文本或音视频数据，提取有价值的信息。内容加工的任务包括：内容分析：按标准或规则分组或标记文本或音视频数据；信息提取：从文本中提取实体、属性、关系等信息；摘要生成：从文本中提取主要内容或观点，生成简短的摘要；词云图：根据词语的频率和重要性，生成图形化的词语集合；内容总结：根据文本的主题和结构，生成概括性的总结。	😊	
作品创作	通过自然语言生成或图像理解、生成等技术，根据用户的需求或喜好，创造出具有艺术价值或娱乐性的作品，例如小说、诗歌、歌曲、漫画、电影剧本等。此场景需要生成有创新性和艺术性的内容。	😊	
机器翻译	通过自然语言理解和生成等技术，将一种语言转换成另一种语言，保持原文的意思和风格、并具有一定的跨语言的文化适应能力。	😊	
意图识别	通过自然语言理解或语音识别等技术，识别出用户输入的语言中所表达的目的或需求，例如订票、查询天气、打开应用等。	😊	
智能检索	通过对用户输入的查询进行分析和匹配，从海量的数据中快速返回相关的信息。（通用领域：涵盖各种主题和领域的信息检索，例如百度、谷歌等。专业领域：针对特定的主题或领域的信息检索，例如医疗、法律、教育等。）此场景具备找到和理解查询相关的文本，理解和处理专业领域的知识的能力。	😊	😊
智能推荐	通过对用户的兴趣、行为、偏好等进行分析和预测，向用户提供个性化的内容或服务，例如电商平台、视频网站、音乐软件等。此场景需要理解用户的历史行为和兴趣，预测并推荐他们可能感兴趣的内容。		😊
辅助决策	通过根据用户提供的问题或目标，分析和评估各种可能的方案或结果，提供合理的建议或指导，例如旅游规划、投资理财、医疗诊断等。此场景需要具备专业领域知识，具备一定的分析数据及预测趋势的能力。		😊
知识管理	通过构建和维护一个包含大量知识和信息的系统，支持知识的存储、检索、更新、共享等功能。此场景需要具备知识表、存储、管理以及搜索等能力		😊

😊 代表对此场景有较好的支撑能力。

知识图谱与大模型分别拥有相对擅长的应用场景。

1.知识图谱与大模型的对比——核心基础能力层面

应用场景	大模型的基础能力	知识图谱的基础能力
智能对话	语义理解、指令遵循、思维链、基础常识支持 上下文理解、情感分析、推理规划	语义理解、知识融合、知识查询、知识推理
内容生成	语义理解、指令遵循、思维链、基础常识支持 上下文理解、情感分析、数据可视化	语义理解、知识融合、知识查询 知识推理、知识可视化
内容加工	语义理解、指令遵循、思维链、基础常识支持 上下文理解、语义分割	-
作品创作	语义理解、指令遵循、思维链 基础常识支持、上下文理解、情感分析	-
机器翻译	语义理解、指令遵循	-
意图识别	语义理解、上下文理解支持、情感分析	-
智能检索	语义理解、指令遵循、基础常识 上下文理解、情感分析	语义理解、知识查询、知识推理
智能推荐	语义理解、推理规划	语义理解、知识查询、知识查询
辅助决策	语义理解、指令遵循 基础常识、上下文理解	语义理解、知识融合、知识查询 知识推理、知识溯源
知识管理	-	知识融合、知识存储、知识补全、知识查询 知识推理、知识溯源、知识共享与交换、知识更新与维护

1.知识图谱与大模型的对比——技术特性层面

大模型的优势

通用性：模型具有指令遵循能力，能处理多种任务，并支持多语言、多模态、多领域的应用。

可生成性：模型能生成各种形式和风格的文本，也能生成多模态的内容，如图像、音频等。

学习能力：基于大量语料的训练，能对新输入产生合理的响应，也能从多模态数据中进行学习。

创作能力：能生成新颖、连贯和通顺的文本，也能生成多模态作品，如图片、歌曲等。

常识能力：基于海量通用训练数据中的知识，具有常识理解能力。

语义理解能力：能根据文本、多模态数据中出现的内容，理解其含义和关系。

知识图谱的优势

可解释性：知识图谱可基于明确的语义结构进行查询和分析，具有较好的可解释性。

可信赖性：知识图谱通常是由专家创建和维护，因此其可信赖性较高。

可溯源性：知识图谱中的每个实体和关系都可以追溯到其来源。

可校验性：知识图谱中的信息可以通过专家进行校验。

可评价性：知识图谱的质量可通过查询的准确性和完整性来评价。

领域能力：具有较强的领域知识支持，支撑了其领域服务能力。

推理能力：可根据图谱中的精确知识内容和关联结构，进行高可信度的推理。

大模型的不足

可解释性：模型的决策过程是黑箱的，难以解释。

可信赖性：模型的输出可能存在错误或偏见的信息。

可溯源性：模型的输出是基于训练的数据，而不是特定的数据点或知识点，较难追溯其输出的来源。

可校验性：模型的输出和推理结果有赖于通过人工或者其他系统进行校验。

可评价性：模型的性能和输出可通过一些标准任务进行评价，尚不成熟。

常识能力：无法处理超出训练语料范围的常识问题。

领域能力：缺乏丰富全面的领域知识，领域服务能力一般。

语义理解能力：可能出现理解错误或歧义等问题。

知识图谱的不足

通用性：知识图谱通常面向特定领域，在通用性上可能较弱。

可生成性：知识图谱主要用于查询和分析，而非生成新的内容。

学习能力：缺乏自主学习能力。

创作能力：缺乏自主创作能力。

常识能力：局限于知识图谱中的信息，常识能力较弱。

语义理解能力：语义理解能力主要局限于知识图谱中的知识内容，理解能力较弱。

2.知识图谱与大模型融合的可行性——技术演化层面

由知识工程而提出的语义网络

网络式表达人类知识构造，以此为基础构建专家系统以实际问题

知识高度依赖人工定义，难以进行扩展

由Google提出的知识图谱系统

表达常识知识，补充现有深度学习模型缺乏的认知能力，推理更精准

通过图拓扑建立的隐式的复杂语义以模拟人类认知，但表征能力不足

多模态知识图谱

利用多模态信息补充符号语义表达的不足，强化知识的表征能力，支撑多模态理解、推理和元认知等能力。

知识异构模态语义对齐难，在不同模态间映射关系多样

过去在技术发展中交替演进

未来面临共同的挑战与目标相互支持

AlexNet代表的深度学习出现

由硬件发展推动而产生的新一代AI方法，模型规模和性能超越传统方法

需要大量标注数据支持，完全没有知识建模的能力

Transformer架构推动大模型发展

BERT, Vision Transformer等依靠预训练模型，以参数化形式建模知识，进一步发展为以GPT系列为代表的大模型技术

需要大量数据、大量算力支持，存在幻觉、高层认知能力等缺点

多模态大模型

利用丰富的多模态数据，强化相互之间语义对齐约束，提升高级认知能力，

异构模态之间的数据对齐难，模态间映射关系复杂

大模型和知识图谱是相互依赖的知识处理与应用技术，知识图谱发展激发了深度学习的需求和发展，深度学习和大模型也成为知识图谱构建的基础能力，并共同面对未来多模态知识相关的挑战。

2.知识图谱与大模型融合的可行性——技术互补层面

互补关系

知识图谱

知识图谱能够为通用大模型的工业化应用提供行业领域知识支撑，弥补通用大模型语料里专业领域知识的不足。

利用知识图谱中的知识构建测试集，可对大模型的生成能力进行各方面的评估，降低事实性错误的发生概率。

可以利用知识图谱在大模型中引入指定约束，适度控制内容生成，提高大模型在行业应用场景中的适应能力。

大模型可以利用语义理解和生成等能力抽取知识，提高知识抽取的准确性和覆盖度，也可以抽取出隐含的、复杂的、多模态的知识，降低图谱构建成本。

大模型可以利用其语义理解和指令遵循等能力辅助知识图谱的半自动化构建设计、增加知识的全面性和覆盖度，协助更好的完成知识融合和更新。

大模型可以辅助提升知识图谱的输出效果，生成更加合理、连贯、有创新性的内容，例如文本、图像、音频等。

大模型

融合方向

互补

大模型擅长处理自然语言和模糊知识，而知识图谱擅长表示结构化知识并进行推理。相互结合，可以充分发挥它们的优势，解决更复杂的问题。

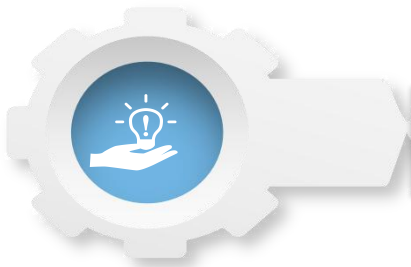
互动

大模型可以用于从文本中提取知识、从而扩展和丰富知识图谱的内容。知识图谱可以为大模型提供结构化知识进行语义补充和生成引导。

增强

知识图谱和大模型融合可以相互增强各自的能力。知识图谱可以提高大模型的语义理解和准确性，而大模型可以为知识图谱提供更丰富的语言知识和生成能力。

2.知识图谱与大模型融合的可行性——知识库建设层面



从知识库角度看，大模型和知识图谱都是存储、处理、利用知识的一种方式。两类知识库虽在知识表示形式上不同，但同作为知识库，二者既有不同又可相互协同。

大模型	知识图谱
动态、概率知识库	静态知识库
参数化知识库，通过网络参数存储知识，不易理解	形式化知识库，通过三元组存储知识，结构清晰，查询简单，易于理解
隐式知识库，隐式的存储知识，决策的过程难归因、解释、溯源	显性知识库，显式地存储知识，有助于归因溯源，提高模型行为的可解释性
更新难度大，忘记特定的知识更加困难	便于更新、修改、迁移知识
知识的通用性更强，适合于高通用知识密度，高专业知识密度（专业语料少）的应用场景	知识的领域性更强，适合于高专业知识密度，低通用知识密度场景
具有上下文感知能力、深层语义表示能力和少样本学习能力	图结构表达能力强。
多模态内容采用模型参数存储，有语义对齐和不可解释性。	多模态知识按照知识表示形式存储。

可以通过prompt，来执行相应信息提取以及思维链的推理任务，形式化成不同形式的知识，例如三元组，多元组或者事件链条。



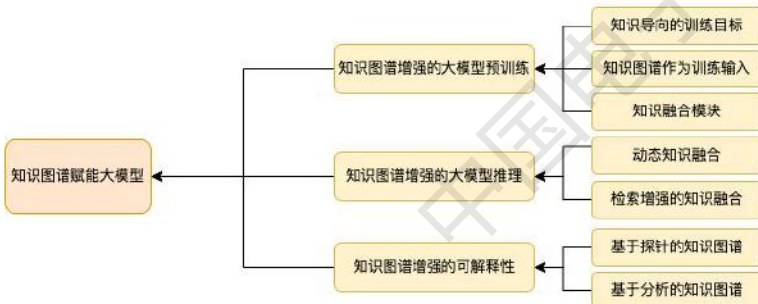
可以利用prompt，参与到大模型的训练前的数据构造，训练中的任务，以及训练后推理结果的约束生成，提升大模型的性能。

两类知识相互融合的桥梁

3.知识图谱与大模型融合的现有研究工作

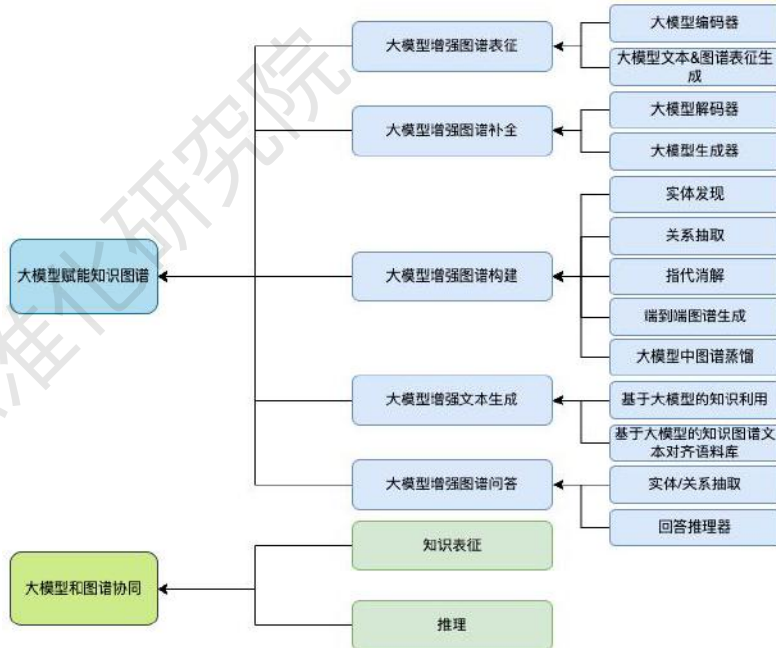
01 知识图谱赋能大模型

通过将知识图谱作为训练目标、模型输入、专门知识融合模块，增强大模型预训练效果；通过动态知识融合、检索增强的知识融合方法，增强大模型推理能力；通过基于知识图谱的探针、分析技术，增强大模型可解释性。



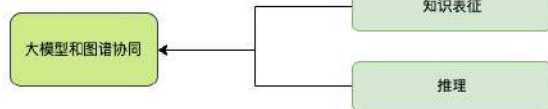
02 大模型赋能知识图谱

通过将大模型作为编码器或者通过大模型的生成能力，增强知识图谱表征；将大模型作为解码器、生成器，作用于知识补全；利用大模型的生成能力，增强图谱构建，对图谱交互、图谱问答等任务提供支持和提升



03 大模型和知识图谱协同

将大模型与知识图谱进行统一表征，增强结果准确性；将大模型和知识图谱结合，运用于推理过程，弥合文本和结构信息之间的差距并提升推理可解释性。



4.知识图谱与大模型融合的收益

知识图谱赋能大模型的收益

- **降低算力**：可减少大模型对无结构化文本的依赖，从而降低大模型的预训练或推理所需的算力和时间。
- **提高知识可信度**：依托知识图谱中经质量评估的知识，可帮助大模型提高信息的质量和可信度，并保障知识的正确性和时效性。
- **增强通用性、领域能力、认知能力**：可帮助大模型获得跨领域和跨语言的知识，并更好地适应不同的领域任务和场景。
- **降低构建成本**：依托知识图谱中的结构化知识，可减少大模型对标注数据或专家知识的需求，从而降低大模型的构建成本和难度。
- **提高可生成性**：可帮助大模型生成更贴近实际、更具有解释性的内容。
- **提高创作能力**：通过知识图谱的知识增强，可帮助大模型创作内容更具逻辑、一致性和创新性等。

大模型赋能知识图谱的收益

- **增强理解能力**：大模型的语义理解能力可帮助知识图谱更好地理解和分类非结构化信息。
- **降低构建成本**：大模型的上下文理解能力、基础常识支持能力等可帮助知识图谱提升非结构化数据的知识获取、知识建模、知识融合等能力，降低其构建和维护成本。
- **丰富输出形式**：大模型的生成能力可帮助知识图谱获得多元化的知识输出和服务形式，增强知识图谱系统的服务效果，并提升人机交互水平。
- **提高知识完备性**：大模型中涵盖的知识及其对新数据的理解能力，可帮助知识图谱进行知识补全和知识校验，提高知识的完备性。

知识图谱与大模型协同的收益

- **提高可解释性**：知识图谱的显性知识与大模型的隐性知识相结合，可提高知识应用的可解释性。
- **实现交叉验证**：知识图谱的输出与大模型的输出相结合，可为知识应用提供交叉验证/比对的手段，提高服务的可信赖性。
- **优化知识存储**：知识图谱的结构化信息存储和大模型的非结构化信息处理相结合，可优化知识存储和检索效率。
- **提高决策能力**：知识图谱推理结果与大模型推理结果的结合，可进一步丰富辅助决策的知识背景，并提供更精确的决策建议。
- **增强隐私保护**：知识图谱中数据加密和保护能力与大模型数据调用能力相结合，可降低大模型对个人隐私数据的依赖，有利于保障隐私安全。
- **确保知识产权保护**：知识管理机制与本地化部署方式相结合，可更好地保护知识产权，防止知识的滥用或盗用。
- **增强伦理边界**：通过优化知识图谱中的知识结构及大模型训练样本结构，构建约束规则类知识并降低数据偏见，强化输出边界。

知识图谱+大模型



• 第三章

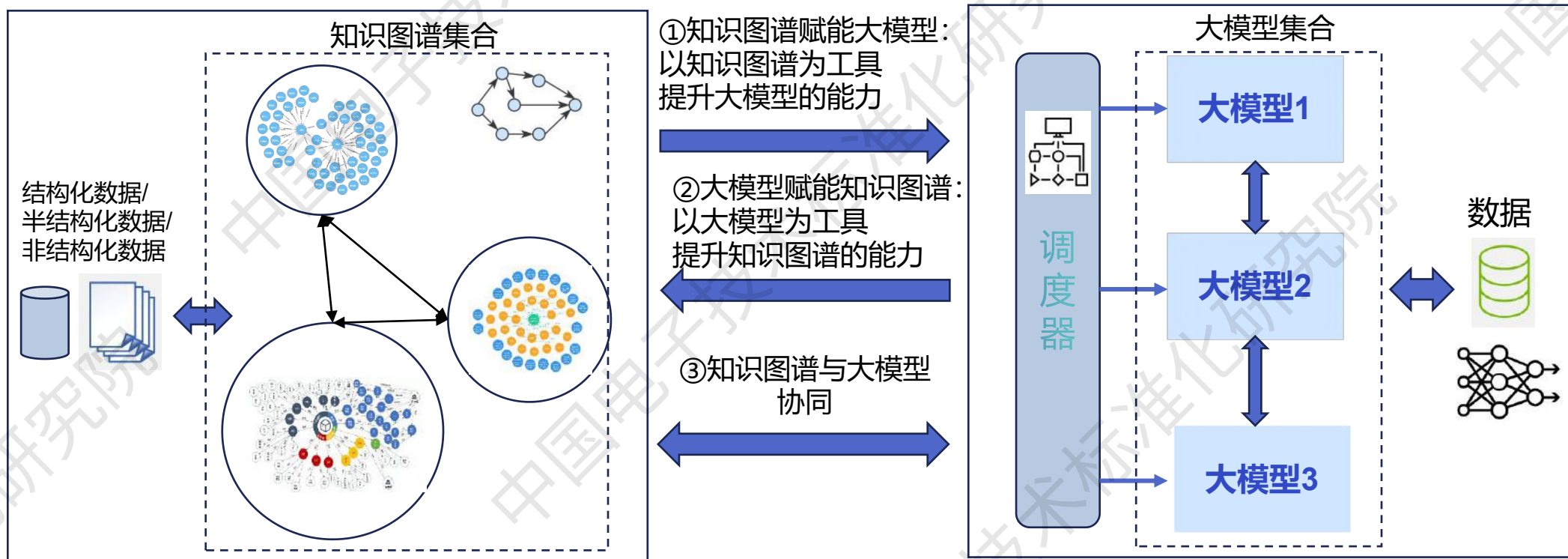
知识图谱与大模型融合的技术路径



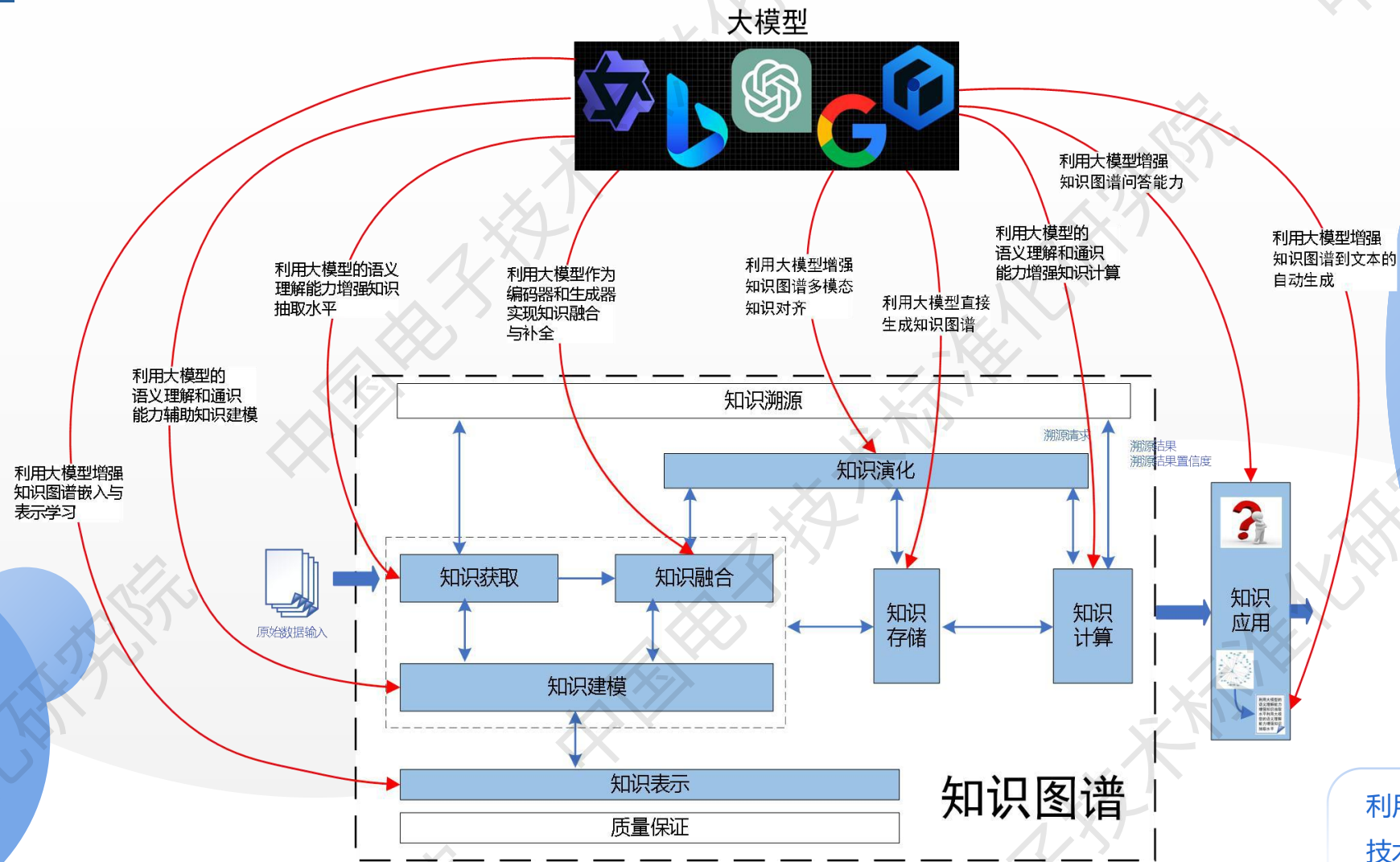
2023

1.知识图谱与大模型融合的总体技术路线

- 利用知识图谱与大模型各自的优势相互赋能（1+1），并结合上层应用集成，实现两者技术的互补。
- 利用知识图谱间的互联互通及大模型间的集成调度（N+N），实现融合后系统能力的持续增强。



2.大模型赋能知识图谱的技术路径

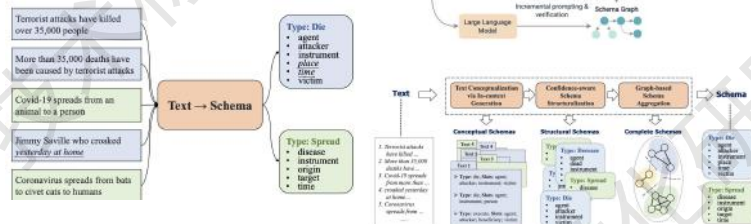


利用大模型在语义理解、内容生成等方面的技术优势，实现大模型对知识图谱构建至应用全生命周期各环节的增强，提升效率和质量。

2.大模型赋能知识图谱的技术路径——关键技术示例

参考文献

2023 Yunjie Ji, et. 《Exploring ChatGPT's Ability to Rank Content: A Preliminary Study on Consistency with Human Preferences》
2021 Shirui Pan, et. 《Unifying Large Language Models and Knowledge Graphs: A Roadmap》
2023 Xiang Wei, et. 《Zero-Shot Information Extraction via Chatting with ChatGPT》



1) 用大模型增强数据标注

- 利用大模型对原始数据进行实体、关系、事件等标注。

3) 用大模型增强知识建模

- 利用大模型进行实体类型提取、关系类型提取、事件类型提取、知识体系提取等。

5) 用大模型增强知识图谱补全

- 利用大模型作为编码器或生成器来补全知识图谱数据，提升知识补全的能力。

2) 用大模型增强知识抽取

- 利用大模型进行实体抽取、关系抽取、事件抽取、因果关系抽取等，例如：DeepKE-LLM。



4) 用大模型增强知识图谱嵌入与表示学习

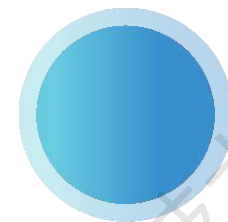
- 利用大模型作为知识图谱嵌入的文本和图结构编码器，解决结构连通性有限的问题，提升知识抽取的能力。



6) 用大模型增强知识图谱构建

- 利用大模型开展实体发现、共指解析和关系提取，构建特定领域的知识图谱结构。
- 采用知识蒸馏等技术实现端到端的图谱构建。





2.大模型赋能知识图谱的技术路径——关键技术示例

8) 用大模型增强知识推理

- 利用大模型进行关系推理、事件推理等



10) 用大模型增强知识图谱文本生成

- 利用大模型自然语言理解方面的优势能够提升从知识图谱中生成文本的质量，提高语言的准确性和在现实场景中的可用性。



12) 用大模型增强知识图谱多模态知识对齐

- 利用大模型的通用性和对多类型数据统一处理的能力，能够增强多模态知识对齐，赋能多模态知识图谱的构建、表示、推理和应用的全流程。

7) 用大模型增强知识融合

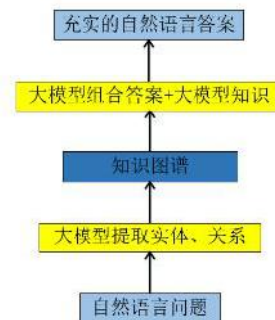
- 利用大模型进行术语定义补全、术语对齐和标准化、实体标准化对齐、同义词提取与融合等

9) 用大模型增强知识图谱可视化

- 利用大模型进行多种形式的知识可视化

11) 用大模型增强知识图谱问答

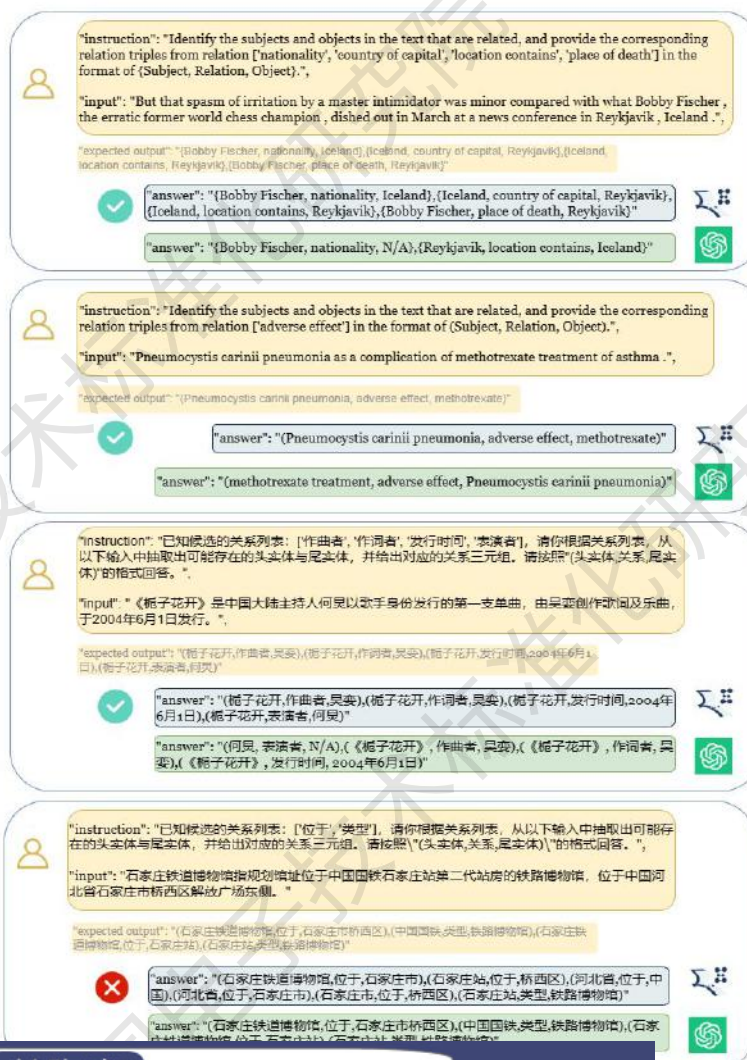
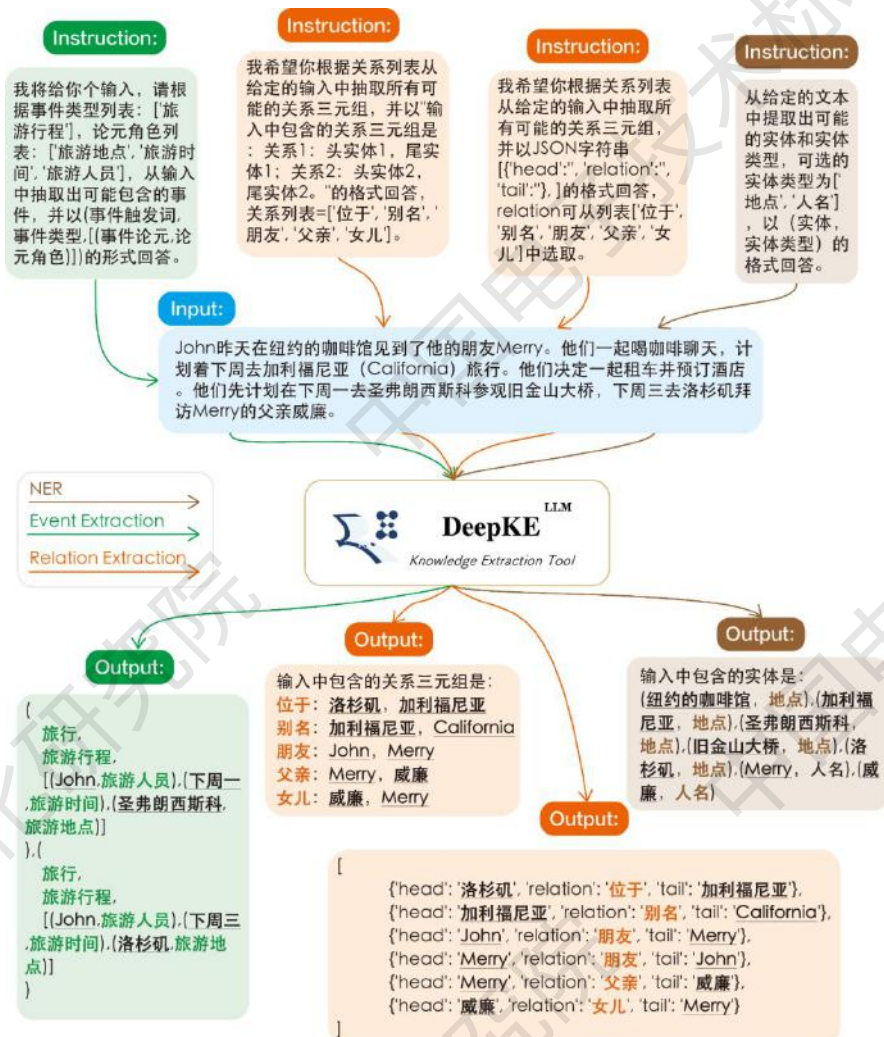
- 利用大模型抽取自然语言问题中的实体、关系，进入结构化的知识图谱寻找问题答案，再通过大模型组合答案并结合大模型自身的知识广度将更充实的答案以自然语言的方式输出，增强知识图谱问答的广度、自然性和准确性。



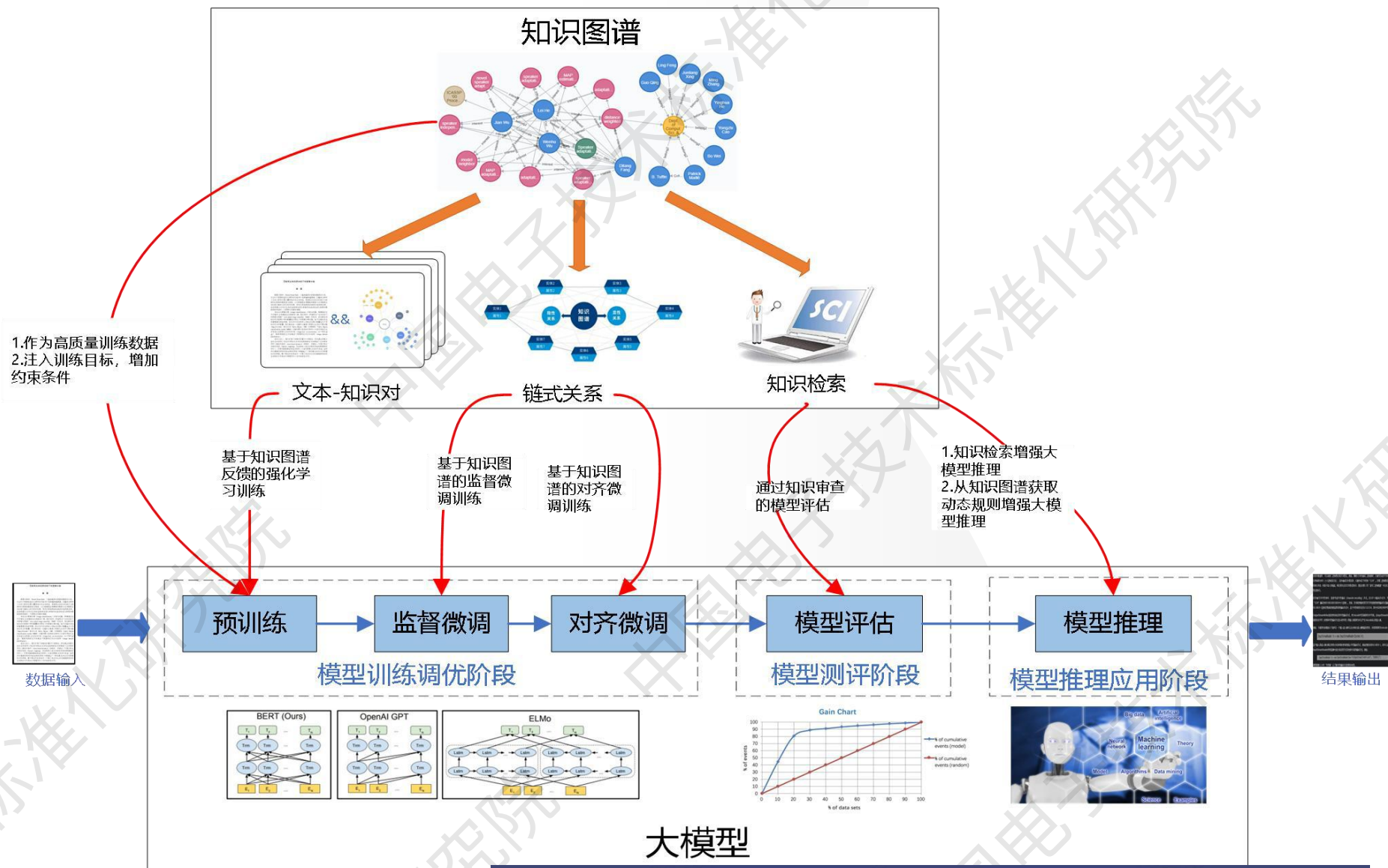


3.知识图谱赋能大模型的技术路径

——应用场景实现示例：基于大模型增强的知识抽取



3.知识图谱赋能大模型的技术路径

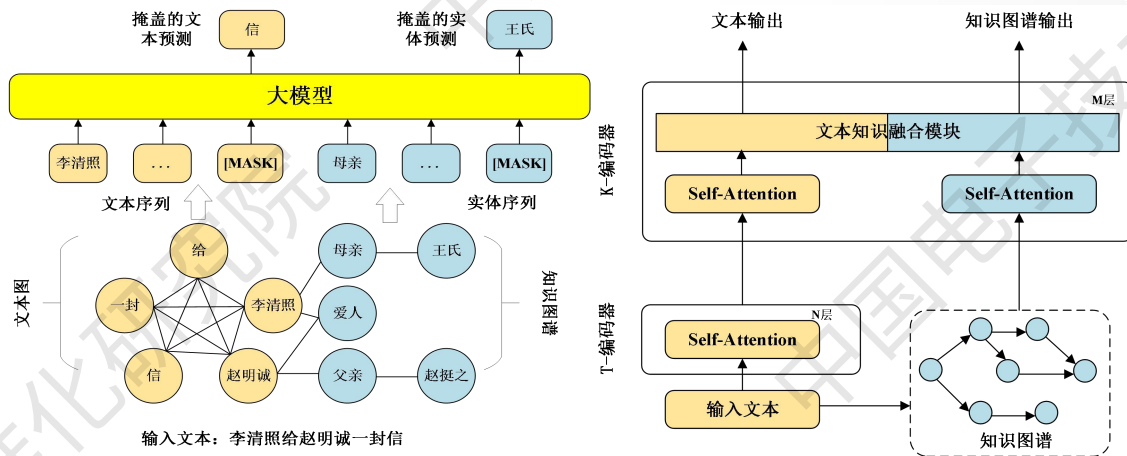


利用知识图谱在知识标准化、可解释性、可信性、可溯源性等方面的优势，通过知识图谱增强大模型从预训练到应用的全生命周期各环节，提升大模型的训练效果和推理结果的可用性。

3.知识图谱赋能大模型的技术路径——关键技术示例

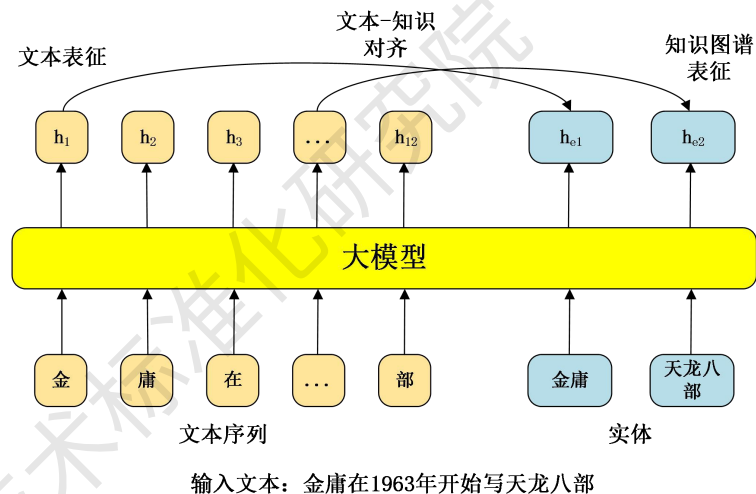
1) 利用知识图谱增强大模型预训练

- 使用图结构将知识图谱信息注入到大模型的输入中，增强大模型预训练能力；
- 通过附加的融合模块将知识图谱注入到大模型，在预训练模型中，可以设计额外的辅助任务，然后通过辅助任务对预训练模型加约束来增强大模型预训练能力；
- 知识图谱的链式关系输入到大模型中，作为大模型的预训练语料。



2) 利用知识图谱增强大模型的监督微调/对齐微调

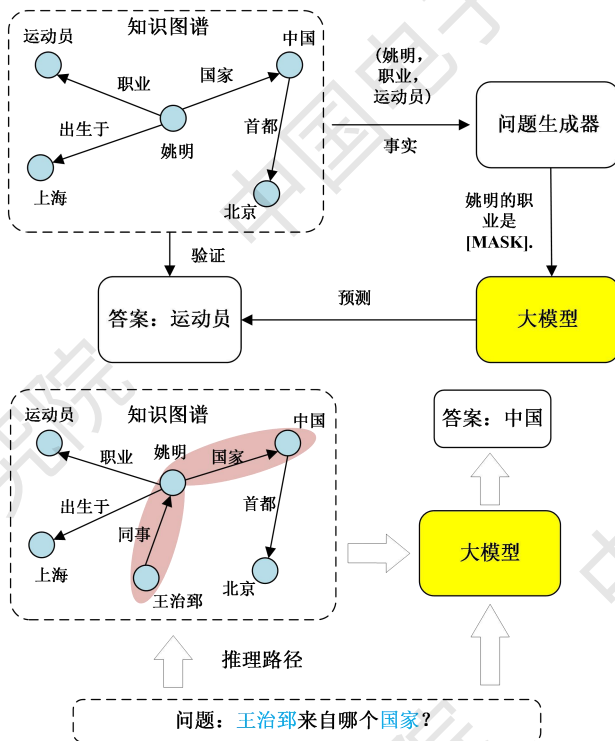
- 通过指令微调训练和基于知识图谱反馈的强化学习；
- 通过文本-知识对齐将知识图谱信息注入到大模型的训练目标中，增强大模型预训练能力。



3.知识图谱赋能大模型的技术路径——关键技术示例

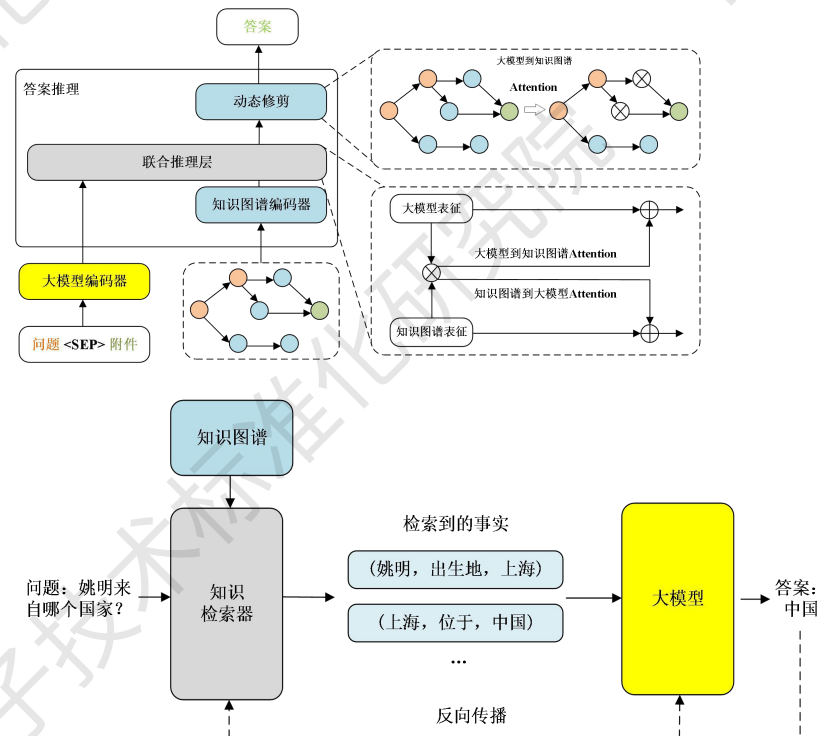
3) 用知识图谱增强大模型的常识和领域知识推理能力

- 把知识图谱当作一个准确的知识库，作为一个外部检索的知识源，提升常识和领域知识方面的推理能力；
- 利用知识图谱的多跳路径作为大模型的输入，提升模型专业性、可信性、真实性与可解释性；
- 使用知识模型根据问题生成知识陈述



4) 用知识图谱增强大模型推理的可解释性

- 通过动态知识图谱融合增强大模型的推理能力，从知识图谱中较容易地获得一系列规则，既从数据中总结得出的特征子图，作为COT、TOT的指令。
- 通过检索增强知识融合增强大模型的推理能力，知识图谱是典型的图结构，有大量的路径。借助这种链式关系可提升大模型推理的可解释性。
- 利用大语言模型对基于知识图谱生成的问题进行预测，验证大模型的可解释性。



3.知识图谱赋能大模型的技术路径——关键技术示例

5) 利用知识图谱增强大模型生成结果的评估与验证

- 把知识图谱当作一个准确的知识库，作为一个外部检索的知识源，解决事实性准确的问题，并进行事实准确性评估

7) 利用知识图谱增强大模型的语义理解能力

- 基于知识图谱，对大模型在输入文本的语义识别过程中进行实体别称补全、实体上下位推理等，提升大模型语义识别的准确性和完整性。

9) 利用知识图谱增强大模型的知识管理和更新能力

- 基于知识图谱，对大模型输出中所依赖的动态数据、隐私数据、事实性知识进行统一的管理，并依托知识图谱的知识编辑能力，保障图谱内知识的实时性和正确性。

6) 用知识图谱增强大模型的对话等应用能力

- 在智能对话等任务中，通过引入知识图谱中的知识对大模型的输出进行约束，提升对话内容的有效性和实时性。

8) 利用知识图谱增强大模型的知识溯源能力

- 基于知识图谱，记录大模型获取知识点和关键数据的来源信息及转化路径，并在内容生成或推理时进行完整呈现，便于使用者评估可信度。

10) 利用知识图谱增强大模型的知识存储和共享能力

- 基于知识图谱，获取和存储大模型中的隐性知识，并通过知识图谱文件或知识图谱间知识共享协议实现知识的交换与流通，提升大模型的知识共享能力。





3.知识图谱赋能大模型的技术路径

——应用场景实现示例：基于知识图谱增强大模型的文档问答

1.离线部分，对文档进行预处理，构建段落级索引，包括全文索引和向量索引

2. 在线部分，使用知识图谱增强大模型的问答效果：在意图识别阶段，用知识图谱进行实体别称补全和上下位推理；在Prompt组装阶段，从知识图谱中查询背景知识放入上下文；在结果封装阶段，用知识图谱进行知识修正和知识溯源

离线

文本提取

文本分段

全文索引

文本转向量

向量索引

在线

用户问题

分类、实体识别、翻译

意图识别

关键词提取

全文检索

文本段落综合排序

文本转向量

向量检索

候选段落

① 实体别称补全
实体上下位推理

③ 知识修正
知识溯源

② 行业背景知识补全

知识图谱

结果封装

大模型调用

LLM api

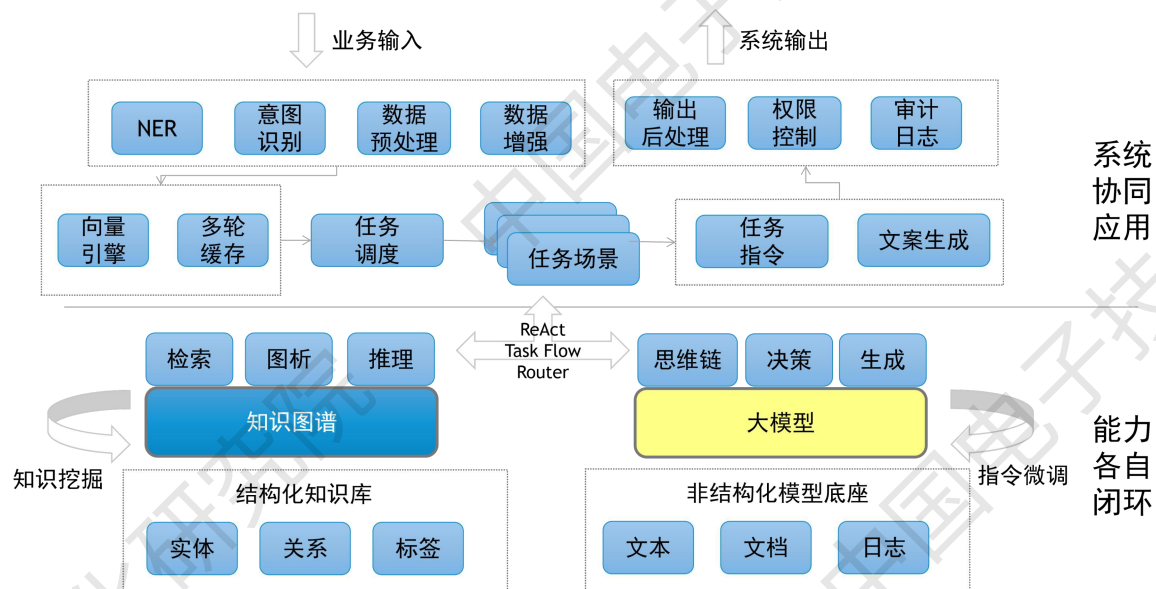
提示组装

Prompt

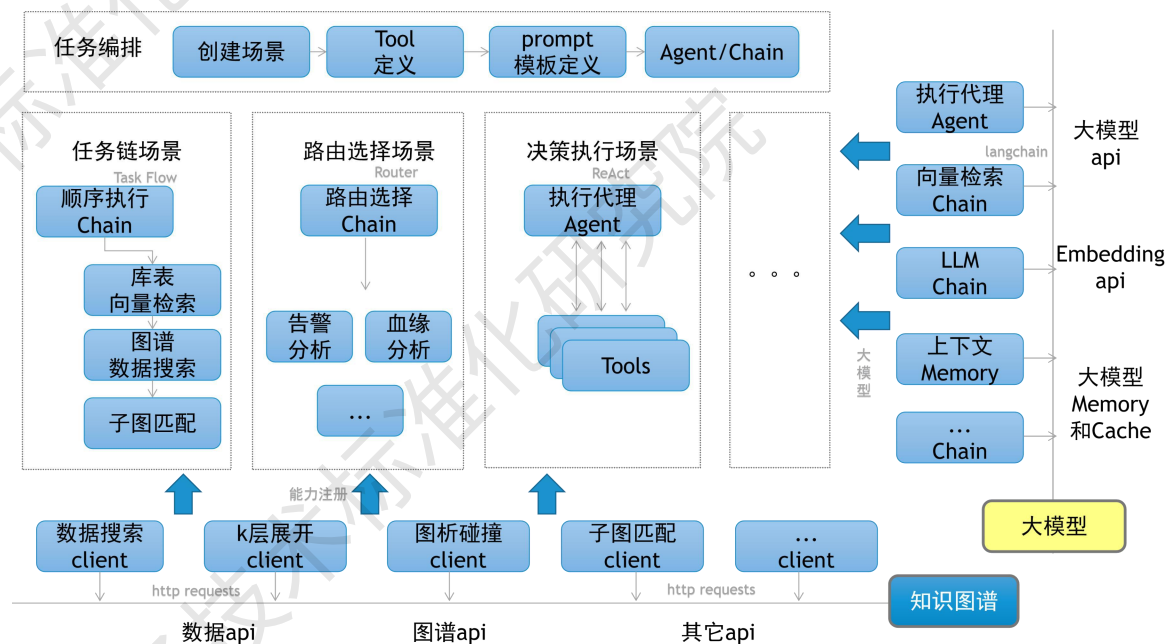
3.知识图谱与大模型协同应用的技术路径

分别发挥知识图谱与大模型两者的技术优势，通过统一知识表征、动态协同知识推理等技术手段，实现企业级认知决策智能水平的升级发展。

动态协同应用



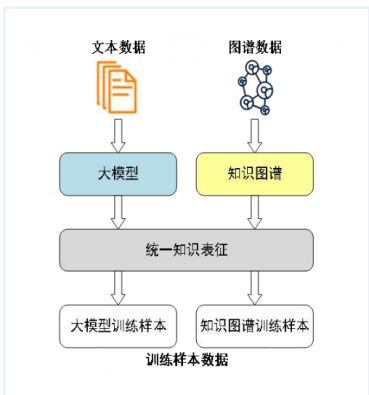
任务流程编排



3.知识图谱与大模型协同应用的技术路径

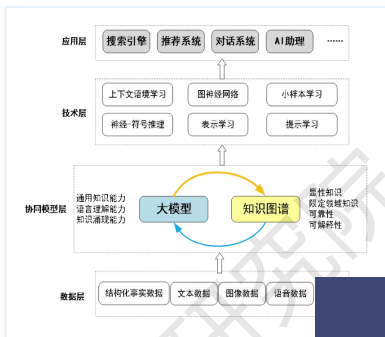
1) 知识图谱与大模型统一表征技术

- 通过对大模型与知识图谱进行知识统一表征,增强结果的准确性。



2) 知识图谱与大模型统一构建技术

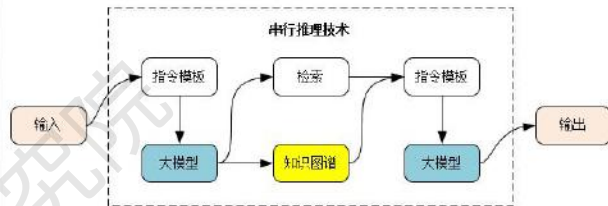
- 通过融合知识图谱的训练目标和大模型的训练目标,构建统一模型,使得统一模型同时具备大模型的通用知识、语言理解、知识涌现能力和知识图谱的显性知识、限定域知识、可靠性、可解释性能力。



参考文献 2021 Shirui Pan, etc. 《Unifying Large Language Models and Knowledge Graphs: A Roadmap》

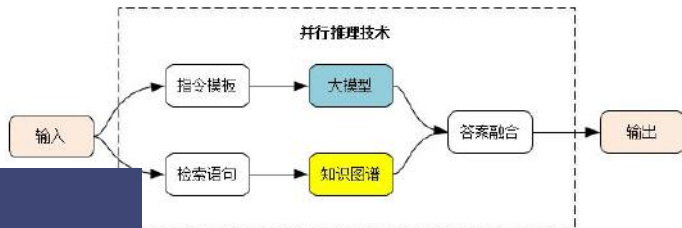
3) 知识图谱与大模型串行推理技术

- 通过知识图谱与大模型的串行应用,原始信息首先经过知识图谱进行结构化抽取关联信息,将检索结果输入大模型进行预测推理,从而提高知识推理预测的准确性。



4) 知识图谱与大模型并行推理技术

- 大模型与知识图谱并行召回答案,动态协同进行知识推理,完成答案融合,即能提高推理结果的准确性,又能拓展推理的知识边界。



3.知识图谱与大模型协同应用的技术路径——关键技术示例

5)



知识图谱与大模型交互接口标准化

- 规定和明确知识图谱与大模型之间交互接口的标准格式，提升不同厂商间产品集成的便捷性。

6)



知识图谱与大模型间任务编排与调度技术

- 知识图谱与大模型协同的过程中，需要基于企业内业务流进行任务的编排和调度，以保证协同过程的流畅性和可操作性。

7)



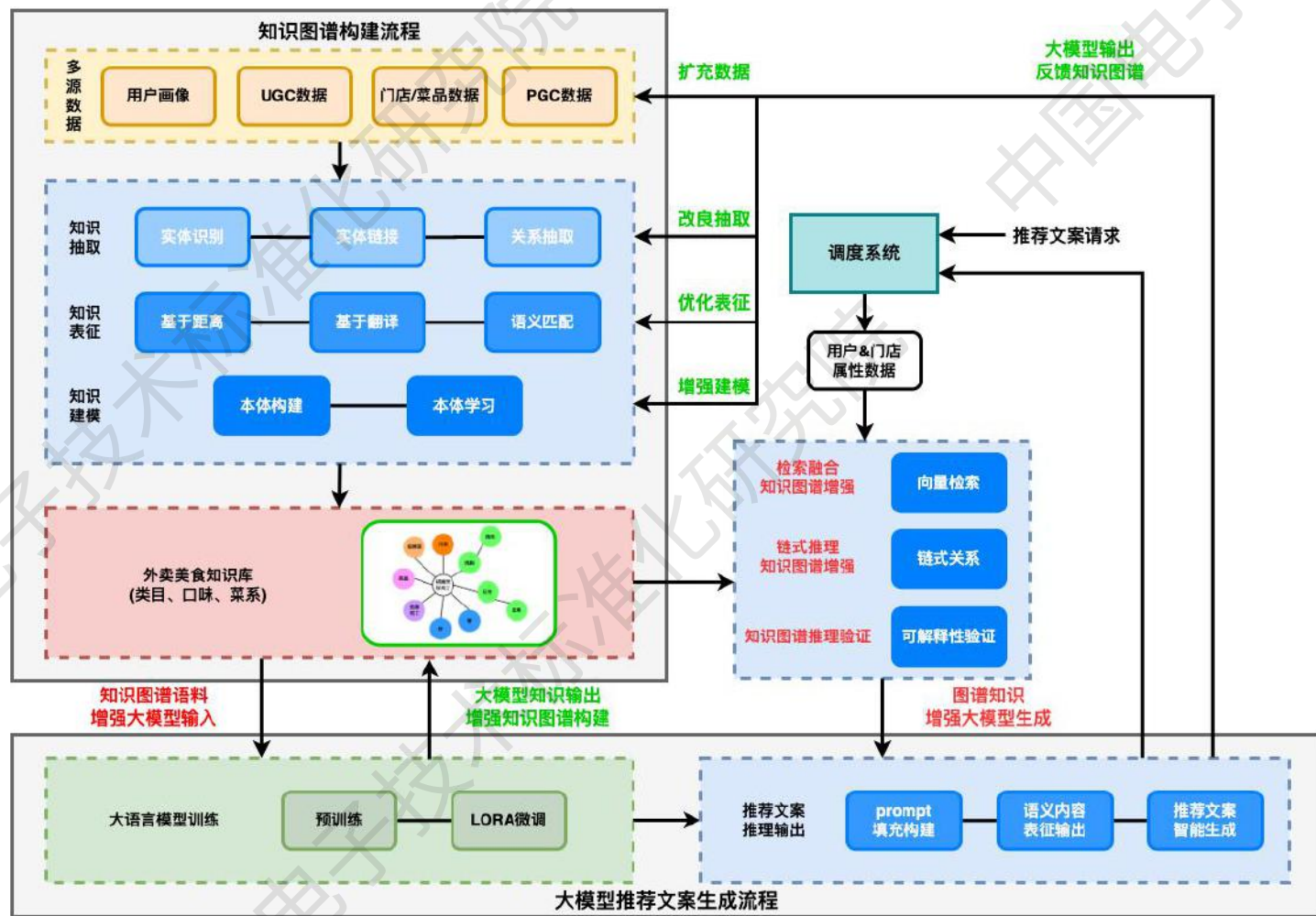
知识图谱与大模型协同中隐私保护技术

- 知识图谱与大模型协同过程中，知识图谱内容仍将被用于大模型的输入或输出中，如何保护知识图谱中的隐私数据不泄漏是系统建设的重要环节。

3.知识图谱与大模型协同应用的技术路径

——应用场景实现示例：基于大模型和知识图谱融合的文案生成

1. 在为用户推荐美食信息的同时，以“知识图谱+大模型”的应用范式智能生成更加触动人心的文案来触达用户。
2. 通过专业领域数据构建行业知识图谱，并通过图谱中知识语料来训练大模型的知识识别与推理能力，训练后的大模型也通过其推理能力来扩充与填补知识图谱，在构建准备阶段形成互补闭环。
3. 在生成推荐文案时，通过向量检索和链式关系等方法利用领域知识图谱中的知识信息增强大模型文案生成的效果，而大模型的生成输出也可以在优化表征和增强建模等方面反馈知识图谱，在文案生成阶段形成互补闭环。



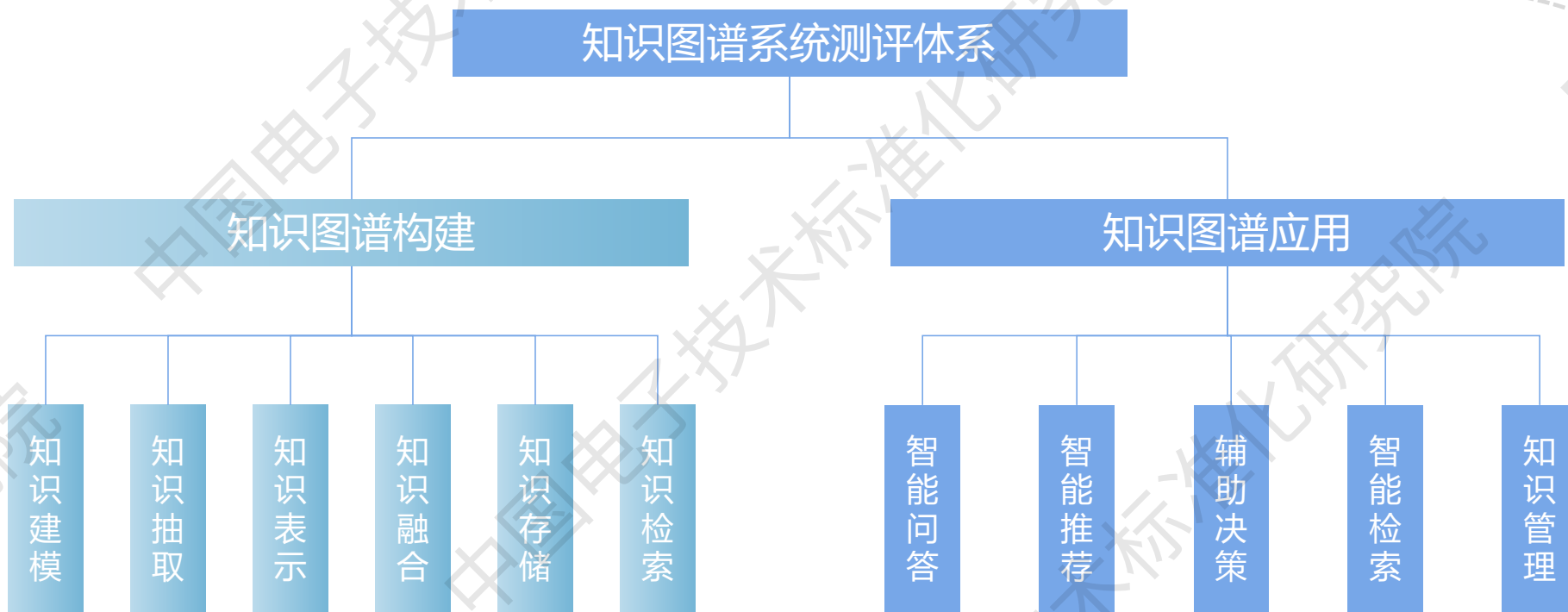
第四章

知识图谱与大模型融合 系统评测



1.知识图谱和大模型系统的测评体系概述

➤ 知识图谱系统测评体系



参考：IEEE P2807.1 《知识图谱技术要求与测试评估规范》

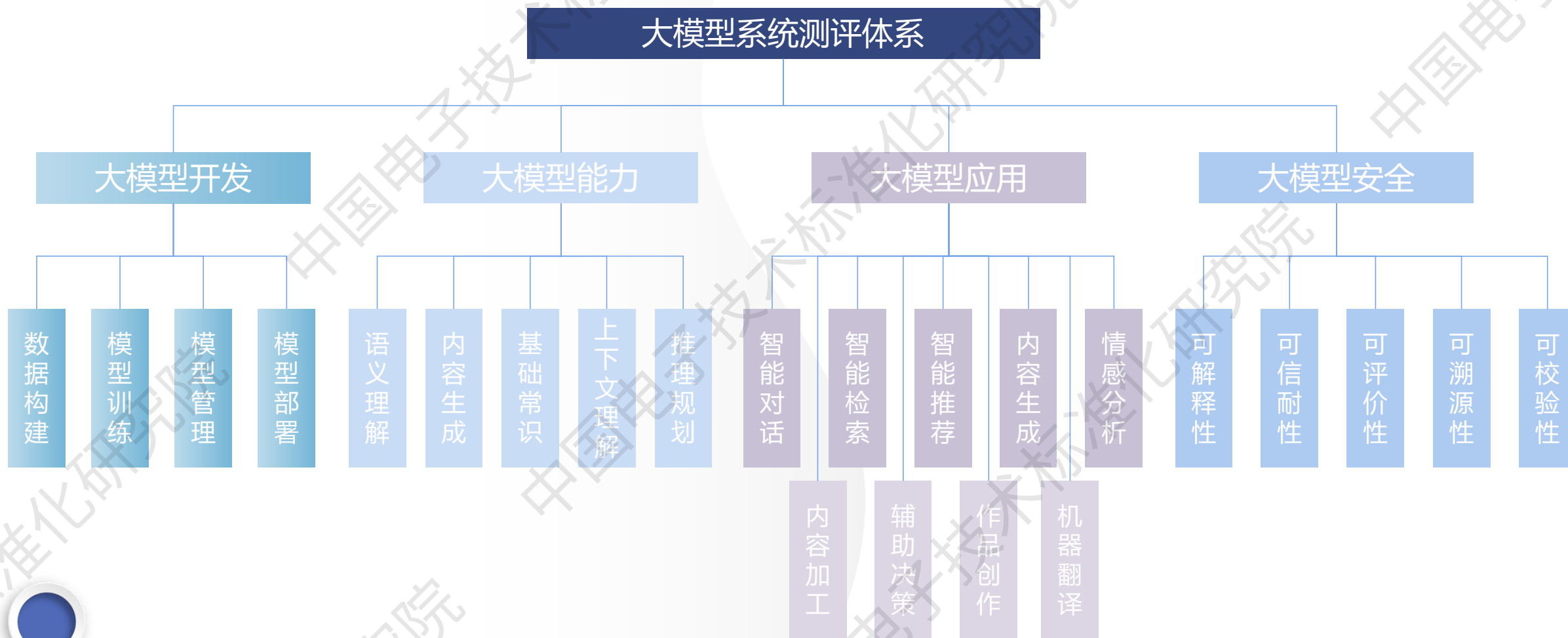
更多免费行业报告

并购家

www.ipoiipo.cn

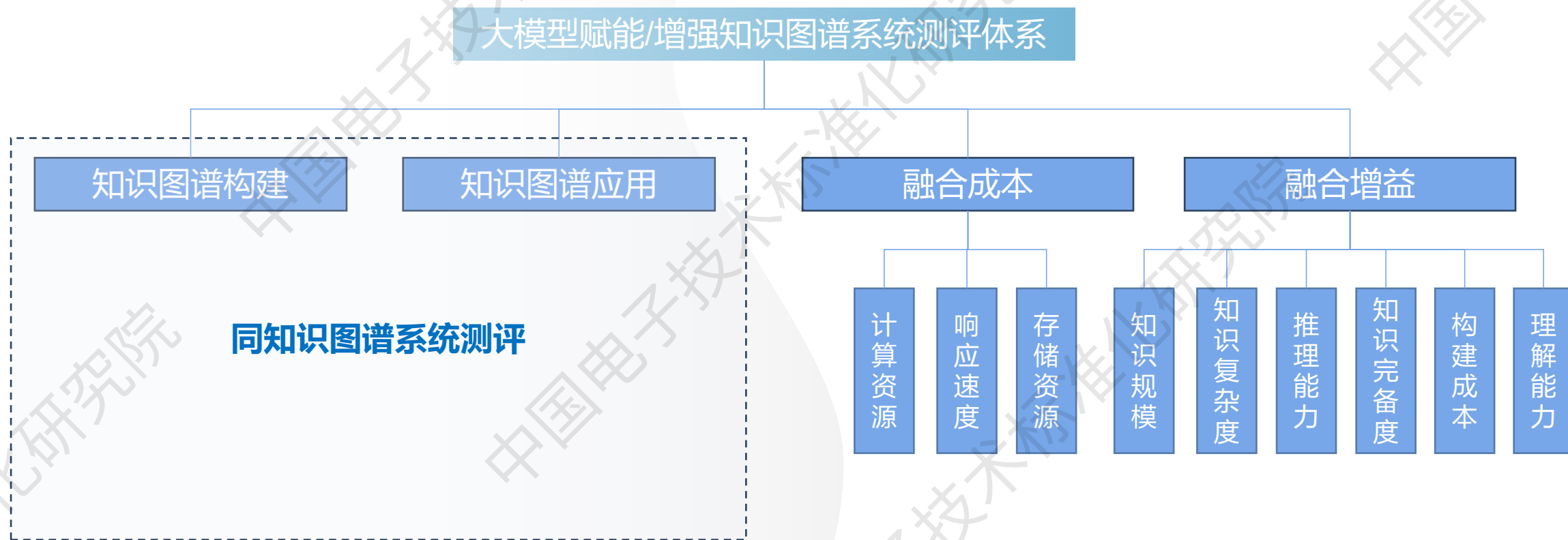
1.知识图谱和大模型的测评体系概述

➤ 大模型系统测评体系



2.知识图谱与大模型融合系统测评体系

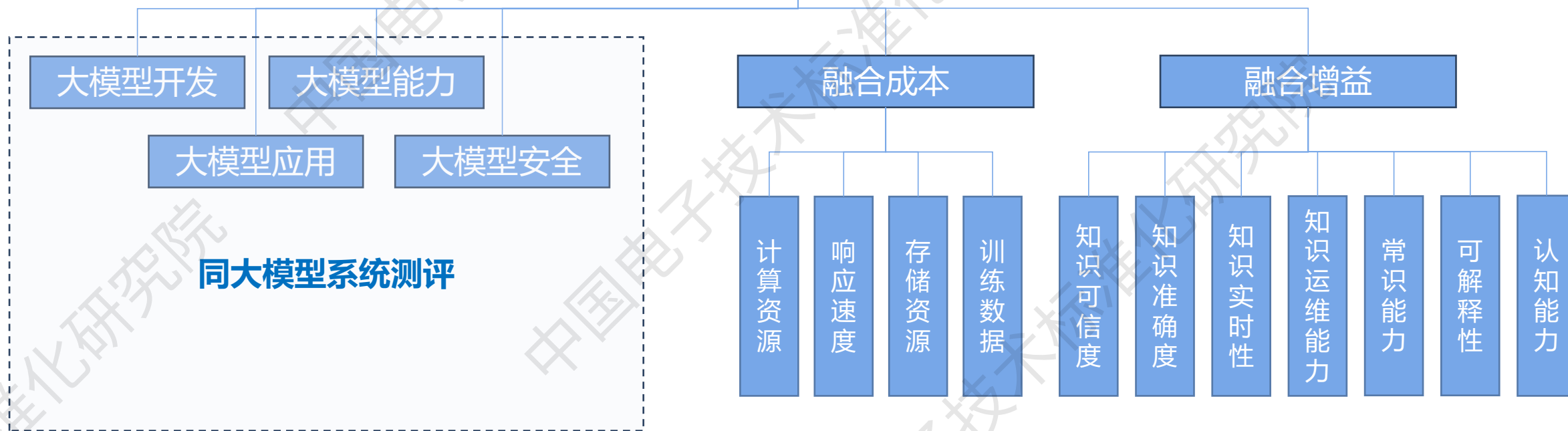
➤ 大模型赋能/增强知识图谱系统测评体系



2.知识图谱与大模型融合系统测评体系

➤ 知识图谱赋能/增强大模型系统测评体系

知识图谱赋能/增强大模型系统测评体系



3.知识图谱与大模型融合系统测评结果

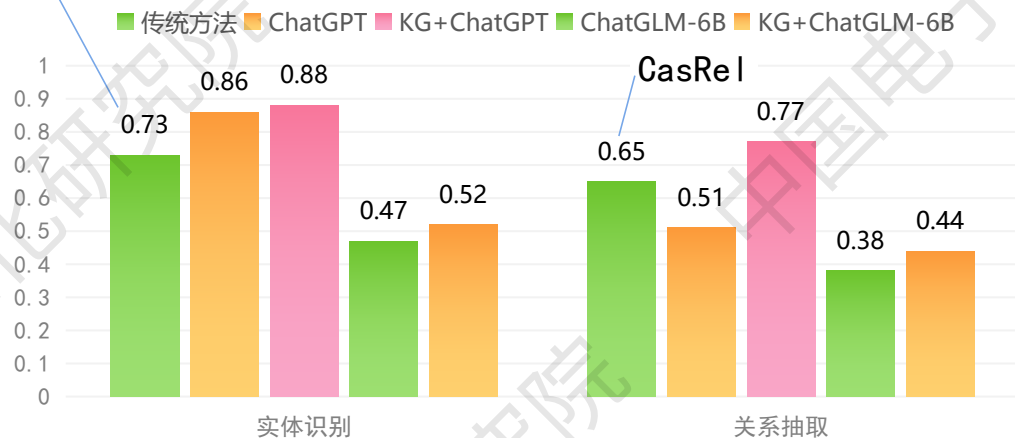
➤ 任务类型：知识抽取

➤ 数据集

数据集名称	规模	子任务	描述
entity-medical-200	200条	实体识别	基于疾病诊疗指南标注的实体识别数据,包含7类实体
relation-medical-200	200条	关系抽取	基于疾病诊疗指南标注的关系抽取数据,包含5种关系

➤ 测评结果

bert+bi lstm+crf



➤ 结果样例

已知 (ALND, 治疗, 早期乳腺癌), 请抽取下文所有实体间的关系: 随着影像技术的发展,越来越多的早期乳腺癌被发现,前哨淋巴结活组织检查 (SLNB) 已迅速取代腋窝淋巴结清扫 (ALND),成为早期乳腺癌腋窝处理的标准方式。

实体之间的治疗关系如下:

- 影像技术 -> 早期乳腺癌
- 腋窝淋巴结清扫 (ALND) -> 早期乳腺癌
- 前哨淋巴结活组织检查 (SLNB) -> 早期乳腺癌
- 前哨淋巴结活组织检查 (SLNB) -> 腋窝淋巴结清扫 (ALND)
- 前哨淋巴结活组织检查 (SLNB) -> 早期乳腺癌腋窝处理
- 腋窝淋巴结清扫 (ALND) -> 早期乳腺癌腋窝处理

KG+ChatGPT显著提升了关系抽取的召回率

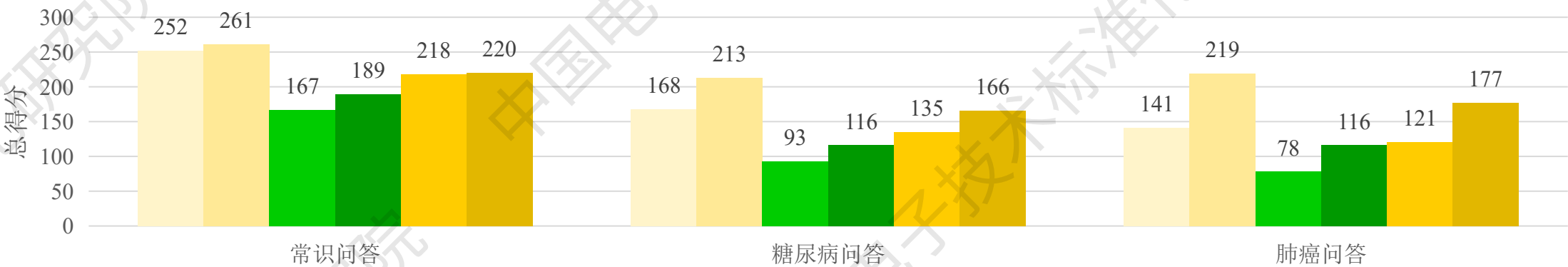
3.知识图谱与大模型融合系统测评结果

➤ 任务类型：智能问答

➤ 数据集

数据集名称	规模	子任务	描述
医药百科图谱	200W三元组		柯基数据基于开源数据构建的医药领域的全科知识图谱
医药常识问题集	100条	常识问答	医学专家人工编辑的常识问题
糖尿病问题集	100条	糖尿病问答	医学专家人工编辑的糖尿病领域的诊疗问题
肺癌问题集	100条	肺癌问答	医学专家人工编辑的肺癌领域的诊疗问题

➤ 测评结果（注：每个问题的答案由医学专家打分，0-3分）



3.知识图谱与大模型融合系统测评结果

➤ 任务类型：智能问答

➤ 结果样例

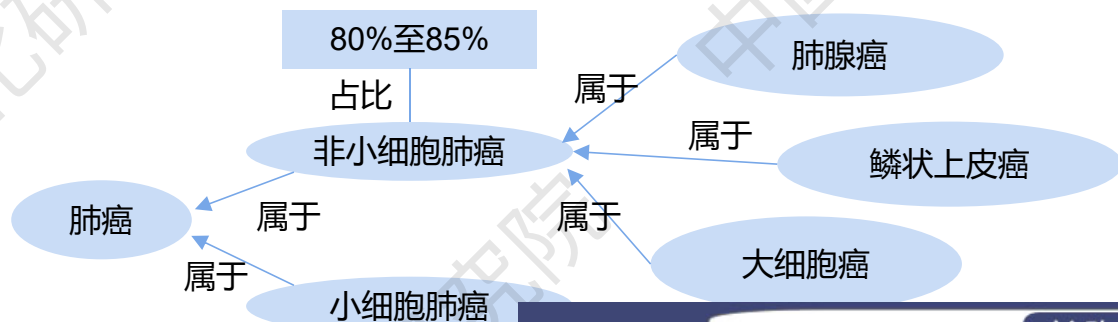
请从一个专业医生的视角回答问题：肺癌包含哪些类型

1分

从专业医生的视角来看，肺癌主要包含以下几种类型：

1. 鳞状细胞癌：这是肺癌中最常见的一种类型，约占肺癌的40%。吸烟是鳞状细胞癌的主要诱因之一。
2. 腺癌：约占肺癌的20%，女性较男性更为常见。吸烟是腺癌的主要诱因之一，但与其他类型的肺癌相比，与吸烟的相关性较弱。
3. 神经内分泌癌：这是一种由神经内分泌细胞发生的恶性肿瘤，是肺癌中较为少见的一类癌。至少有五种不同的病理学亚型，包括类癌、不典型类癌、小细胞癌、大细胞癌和巨细胞癌。

除了以上三种类型，还有其他的肺癌类型，如腺鳞癌、肉瘤样癌、类癌、唾液腺型癌和未分类癌。这些类型在肺癌中较为少见，但也会对患者的健康产生严重影响。因此，及早发现和治疗肺癌对于提高患者的生存率和生活质量至关重要。



已知非小细胞肺癌属于肺癌，小细胞肺癌属于肺癌，肺腺癌属于非小细胞肺癌，鳞状上皮癌属于非小细胞肺癌，大细胞癌属于非小细胞肺癌，非小细胞肺癌占比80%至85%。请从一个专业医生的视角回答问题：肺癌包含哪些类型

3分

从专业医生的视角来看，肺癌主要包含以下类型：

1. 小细胞肺癌：这种肺癌类型起源于神经内分泌细胞或腺体细胞，相对较少见，约占所有肺癌的15%左右。小细胞肺癌的生长和扩散速度较快，通常在早期就可能已经发生转移。
2. 非小细胞肺癌：这是最常见的肺癌类型，约占所有肺癌的85%左右。非小细胞肺癌又包括以下几种类型：
 - 肺腺癌：这种肺癌类型起源于支气管黏膜或腺体，相对较常见，约占所有非小细胞肺癌的40%左右。
 - 肺鳞状上皮癌：这种肺癌类型起源于支气管黏膜上的鳞状细胞，相对较少见，约占所有非小细胞肺癌的20%左右。
 - 大细胞癌：这种肺癌类型起源于支气管黏膜或腺体的大细胞，相对较少见，约占所有非小细胞肺癌的15%左右。
 - 肺泡细胞癌：这种肺癌类型起源于肺泡表面的上皮细胞，相对较少见，约占所有非小细胞肺癌的10%左右。

除了以上几种类型，还有罕见的支气管腺癌和肺转移癌等。

需要注意的是，肺癌类型的诊断需要结合临床症状、影像学检查、病理学检查结果等多方面因素进行综合判断。因此，如果您或您身边的人出现了肺部不适的症状，建议及时就医并听从专业医生的建议进行治疗。

3.知识图谱与大模型融合系统测评结果

➤ 任务类型：文本分类

➤ 数据集

数据集名称	规模	子任务	描述
event-100	100条	文本分类-单层级	警情数据，单层分类的数据
case-1k	1000条	文本分类-多层级	案件数据，有父子三层级分类的数据

➤ 测评结果

子任务	准确率	
	LLM	KG+LLM
文本分类-单层级	67%	93%
文本分类-多层级	31%	56%

➤ 结果样例

T_人工标签	T_大模型标签	T_KG_LLM标签
妨害司法-掩饰、隐瞒犯罪所得、犯罪所得收益-掩饰、隐瞒犯罪所得	妨害社会管理秩序-扰乱公共秩序-开设赌场	治安管理-中华人民共和国治安管理处罚法-故意损毁财物
扰乱市场秩序-非法经营-帮助信息网络犯罪活动罪	扰乱市场秩序-非法经营-帮助信息网络犯罪活动罪	扰乱市场秩序-非法经营-帮助信息网络犯罪活动罪
诈骗-非接触性诈骗-刷单返利类	诈骗-接触性诈骗-其他接触类诈骗	诈骗-接触性诈骗-其他接触类诈骗
治安管理-中华人民共和国治安管理处罚法-吸毒	治安管理-中华人民共和国治安管理处罚法-殴打他人	治安管理-中华人民共和国治安管理处罚法-殴打他人
治安管理-中华人民共和国治安管理处罚法-殴打他人	治安管理-中华人民共和国治安管理处罚法-殴打他人	治安管理-中华人民共和国治安管理处罚法-殴打他人
非接触性诈骗-冒充电商物流客服类-冒充电商物流客服类-其他	诈骗-接触性诈骗-其他接触类诈骗	非接触性诈骗-冒充电商物流客服类-诈骗
诈骗-接触性诈骗-其他接触类诈骗	诈骗-接触性诈骗-其他接触类诈骗	诈骗-接触性诈骗-其他接触类诈骗
治安管理,治安管理-中华人民共和国治安管理处罚法,中华人民共和国治安管理处罚法-侮辱、诽谤、诬告陷害?,故意伤害	治安管理-中华人民共和国治安管理处罚法-故意损毁财物	治安管理-中华人民共和国治安管理处罚法-故意损毁财物
诈骗-非接触性诈骗-刷单返利类	诈骗-接触性诈骗-其他接触类诈骗	诈骗-非接触性诈骗-刷单返利类
妨害社会管理秩序-扰乱公共秩序-帮助信息网络犯罪活动罪	治安管理-中华人民共和国治安管理处罚法-故意损毁财物	妨害社会管理秩序-扰乱公共秩序-帮助信息网络犯罪活动罪
治安管理-中华人民共和国治安管理处罚法-故意损毁财物	治安管理-中华人民共和国治安管理处罚法-故意损毁财物	治安管理-中华人民共和国治安管理处罚法-故意损毁财物
妨害社会管理秩序-扰乱公共秩序-开设赌场	治安管理-中华人民共和国治安管理处罚法-故意损毁财物	妨害社会管理秩序-扰乱公共秩序-开设赌场

• ZHI SHI TU PU YU DA MO XING

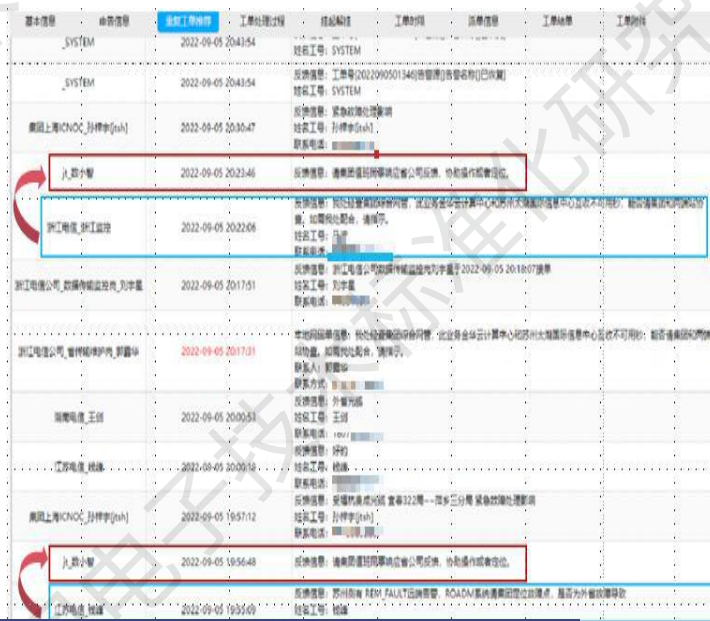
第五章

知识图谱与大模型融合 实践案例

关键技术：1、基于网络大模型和运维知识图谱技术打造智能运维助手；2、基于意图理解和运维知识图谱打造运维智能问答机器人

提升效果：

- 人均维护网元数较部署运维数字员工前增加1倍
- 人工介入处理工作量较部署运维数字员工前减少一半



2.电力行业实践案例：电力智能客服



行业需求

- 1、传统智能客服机器人机械化、条目式的知识检索与问答服务存在用户诉求识别率低、泛化性差等问题，无法满足当前电力客服深度智慧化的需求
- 2、为解决话务量大且座席业务繁重问题，亟需开展智能客服的适应性升级改造，建立智能服务一体化运营管理体系，分流缓解话务高峰，降低客服业务运营培训成本，提升电力客服业务服务水平



关键技术：

- 1、电力客服领域语言大模型微调优化技术
- 2、基于领域知识图谱的大模型知识增强技术



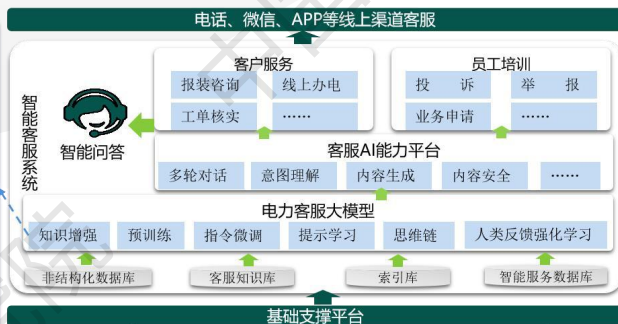
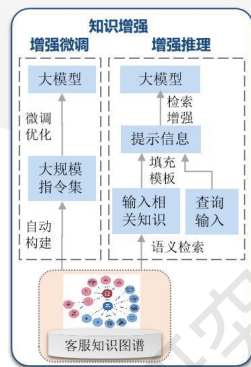
解决方案：

利用客服知识图谱、知识库等语料资源以及LLM大语言模型，构建深度智慧、安全可信的电力客服大模型，满足精准的用户诉求分析、多样化的问答任务响应、实时高效的多轮对话等需求，实现客服问题生成式应答和多样化业务的灵活响应。



提升效果：

提升客服多轮对话内容生成准确率、用户诉求智能客服应答率等性能。



2.电力行业实践案例：信通小数应用

01



行业需求：

1、行业数据量庞大且多样化，数据呈分散态势，难以高效整合和分析；2、行业特点较强，数据包含较多专业术语及领域知识，传统NLP技术难以准确理解分析；3、文本数据存在复杂的结构和语法，对处理系统要求较高。

02



关键技术：

1、自然语言处理；2、领域智能交互；3、语义及情感分析。

03



解决方案：

信通小数应用基于电力领域特性和通用语料训练而成的面向电力行业的智能交互应用，为电力行业安监、营销、基建等八大领域提供文本处理、信息提取和智能决策等多种需求的产品。

04



提升效果：

1、在视频会议的转录及提纲环节减轻记录员相关工作量约90%；2、在综合办公的公文写作及大纲编制环节，提升工作人员60%工作效率；3、应急处理缩短45%处理时间。



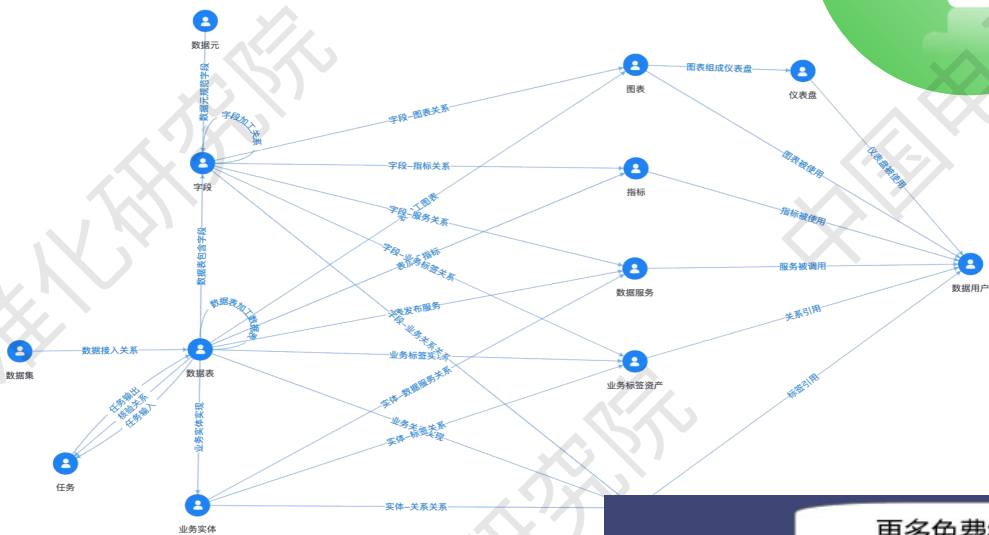
更多免费行业报告

并购家

www.ipoiipo.cn

- ## 关键技术

主动元数据、元数据血缘、特征子图、预训练模型



基于知识图谱，构建数据资产的全链路血缘，将应用级别、资产状态等信息作为属性存储，为特殊数据资产识别提供底层支撑

基于大模型，从图结构信息和节点属性中提取必要特征，智能的为用户进行特殊数据资产的推荐及相关异常预警

已部署于华东某国网，基于大模型和知识图谱的特殊数据资产识别及管理系统，基于用户不同业务场景，推荐不同类别的特殊数据资产（如核心数据资产、边缘数据资产、冗余数据资产等），帮助用户对数据资产进行管理。且在特殊数据资产发生变化时，对受影响的部门或责任人进行自动预警



3.金融行业实践案例：银行智能营销助手



行业需求:

1. 营销领域知识图谱构建费时费力。2.知识图谱的现有展现形式难以快速获取复杂知识和实体关系。

关键技术:

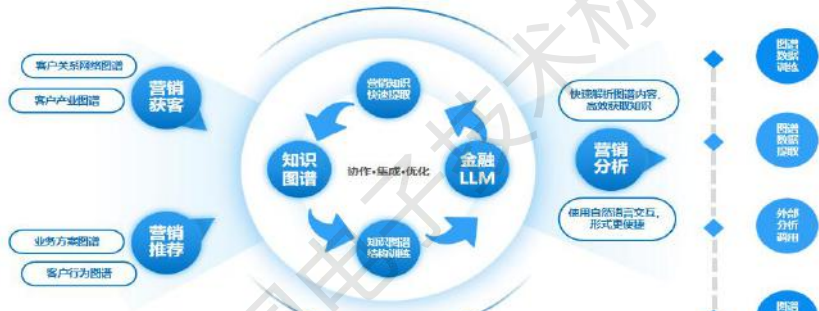
1.利用大模型进行实体、属性、关系等知识图谱要素提取，辅助知识图谱内容生成。2. 训练大模型符合知识图谱内容结构的指令模版。3.利用大模型检索知识图谱进行内容分析。4. 调用外部接口进行进一步的业务分析。5.利用大模型整合内容生成最终的回答。

解决方案:

1.将银行的营销业务知识图谱与大模型相结合，利用大模型实现知识图谱数据的快速提取和分析。
2. 采用便捷的自然语言交互方式，降低传统图谱分析的复杂性，提升分析效率。

提升效果:

实现了基于营销知识图谱的分析问答，助力营销业务高效推进。



3.金融行业实践案例：基于大模型的智能图分析平台反欺诈场景应用

行业需求：

- 1) 根据监管可疑特征构建单规则、复杂规则；规则指标维度较少；预警量大、准确率低；
- 2) 基于涉案名单作为样本构建机器学习模型，提升了召回率、准确率，但可解释性低。

解决方案：

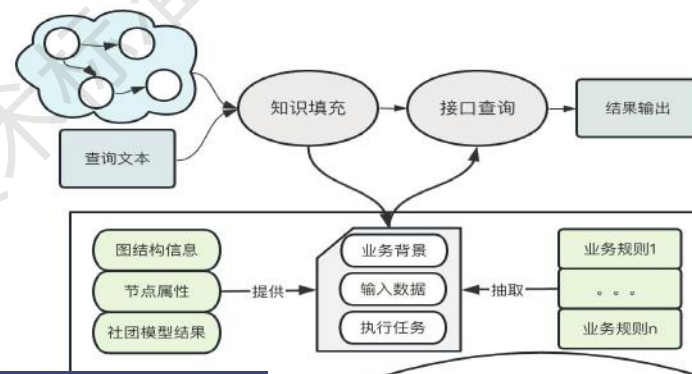
- 1) 基于知识图谱，建立以图算法和机器学习为核心的团伙反欺诈模型，能够挖掘客户关系网络和账户间的隐藏资金链，并提升对可疑团伙的识别能力，无论是静态的还是动态的关系；
- 2) 基于大模型，从图结构信息、节点属性和模型特征中提取关键信息，生成智能风险报告，并通过基于特征的联动图谱可视化展示，使得风险分析更加智能化和直观化。

提升效果：

在银行内反欺诈平台进行了业务可行性评估，智能解读欺诈团伙的行为特征所生成的风险报告，以及提供团伙关系和模型特征的图谱可视化展示，能够提升反欺诈作业人员的研判效率。

关键技术：

图算法、机器学习、图结构信息抽取、预训练模型



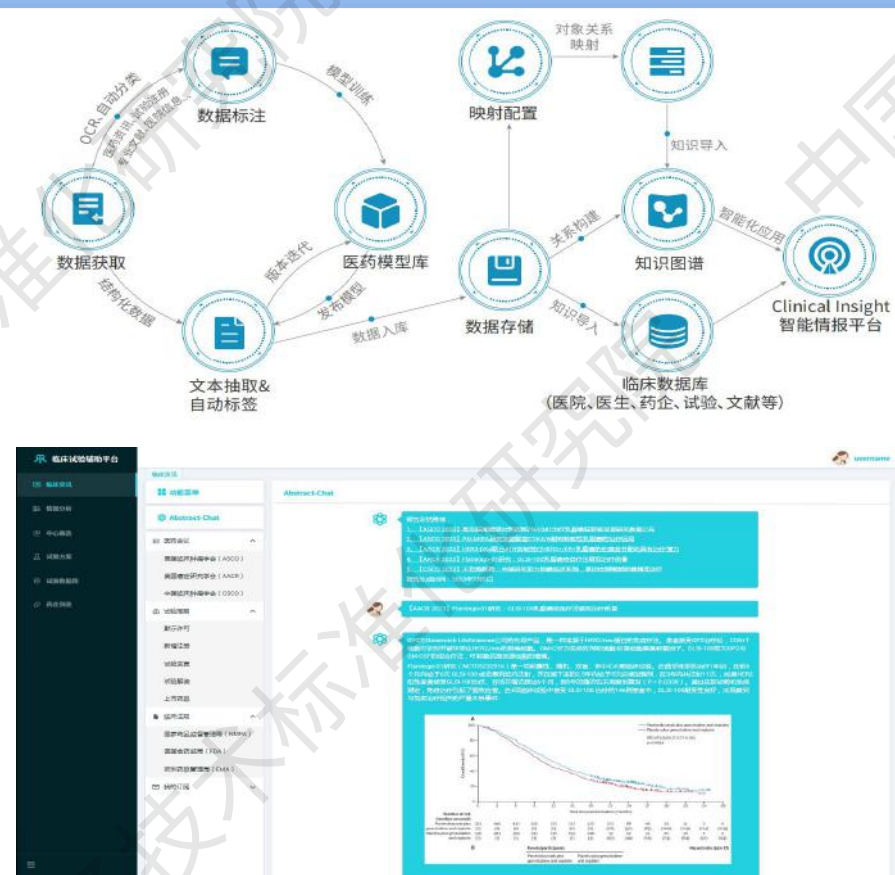
4.医药行业实践案例：Clinical Insight临床试验情报平台

行业需求： 1、加速药物上市前的临床试验设计和临床试验招募，以及上市后的产品上市教育、药品渠道销售、患者全流程管理和数字化诊疗等多种场景；2、整合多源异构信息为医药场景提供高效、客观、科学的循证支持，实现降本增效。

关键技术： 1、医药会议摘要的智能问答；2、临床知识报告生成。

解决方案： 利用知识图谱及LLM大语言模型进行数据的关联分析及内容生成，为企业提供药物试验的潜在竞争情报，并关联临床试验结果，为试验设计提供循证参考。

提升效果： 1、临床试验的入排标准设计和试验中心筛选环节周期缩短60%；
2、实现遵循医学规范，实现医学知识的复用，进一步提高数据的价值和应用。



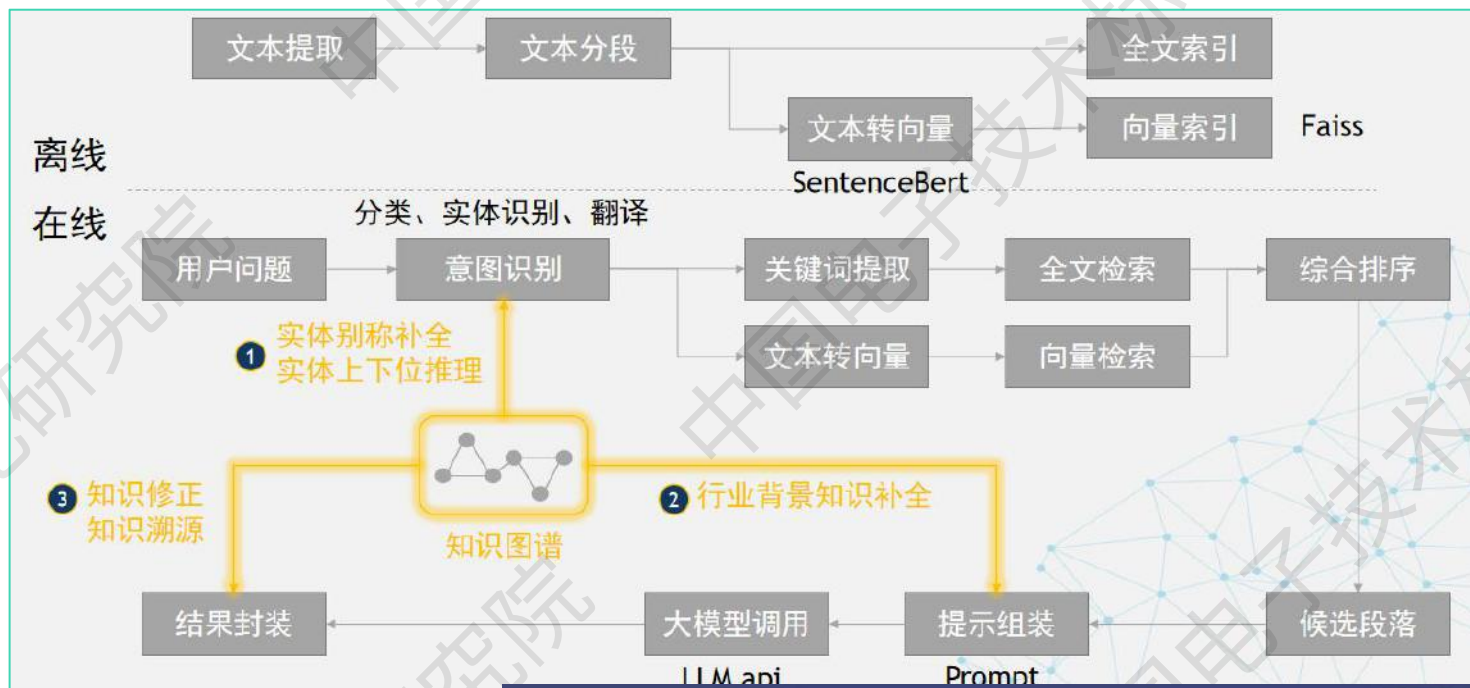
4.医药行业实践案例：医学学术营销平台

行业需求：1、知识分散，没有有效整合，耗费人工去找寻答案；2、医学问询邮件没办法保证立即回复，无法快速地帮助医生/患者等解决问题；3、整合所有资料的知识点，有局限性，还是会出现无回答的情况。

关键技术：1、基于知识图谱的知识增强能力；2、文档解析、问答和自动报告的流程自动化。

解决方案：全球化医学Chatbot平台是一个为医药企业打造的面向外部医生、护士、药剂师等医学专业人士，基于知识图谱和LLM大语言模型能力可循证的疾病用药的应用产品。

提升效果：提升医学部/市场部的效率达到50%



5.汽车行业实践案例：购车攻略平台

行业需求:

- 1) 购车是许多人生活中的重大决策之一，人们希望能够获取针对个人需求的准确且全面的汽车推荐信息，包括车型、价格、性能等方面的细节。
- 2) 提供购车过程中的相关指导和建议，以便做出明智的选择。

解决方案:

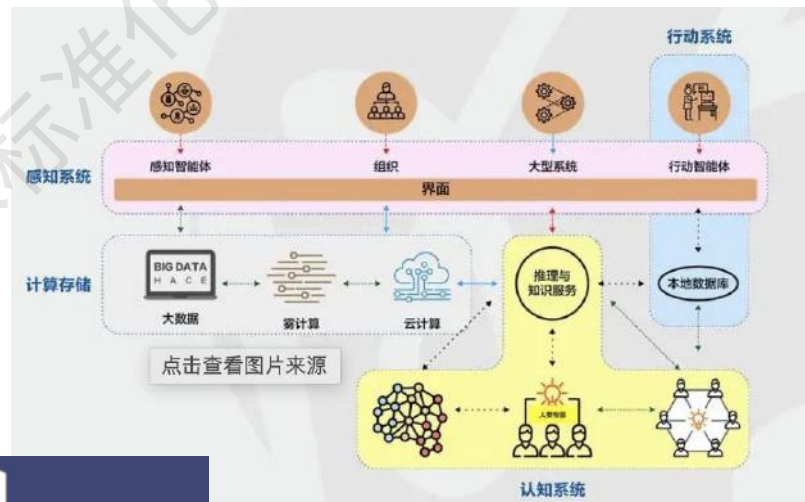
通过智能问答系统，结合知识图谱与自然语言处理技术，为用户提供车型、参数、技术规格、价格、预算、性能和购车推荐和指导。

提升效果:

- 提供个性化的购车推荐和指导，使用户更容易找到适合自己需求的汽车。
- 通过价格预测模型，为用户提供参考的价格范围，帮助他们在合理的预算范围内做出选择。
- 减少用户的购车时间和不必要的试错，提高购车效率和满意度。
- 构建良好的用户体验，提高用户留存和口碑，为汽车销售商带来更多潜在客户。

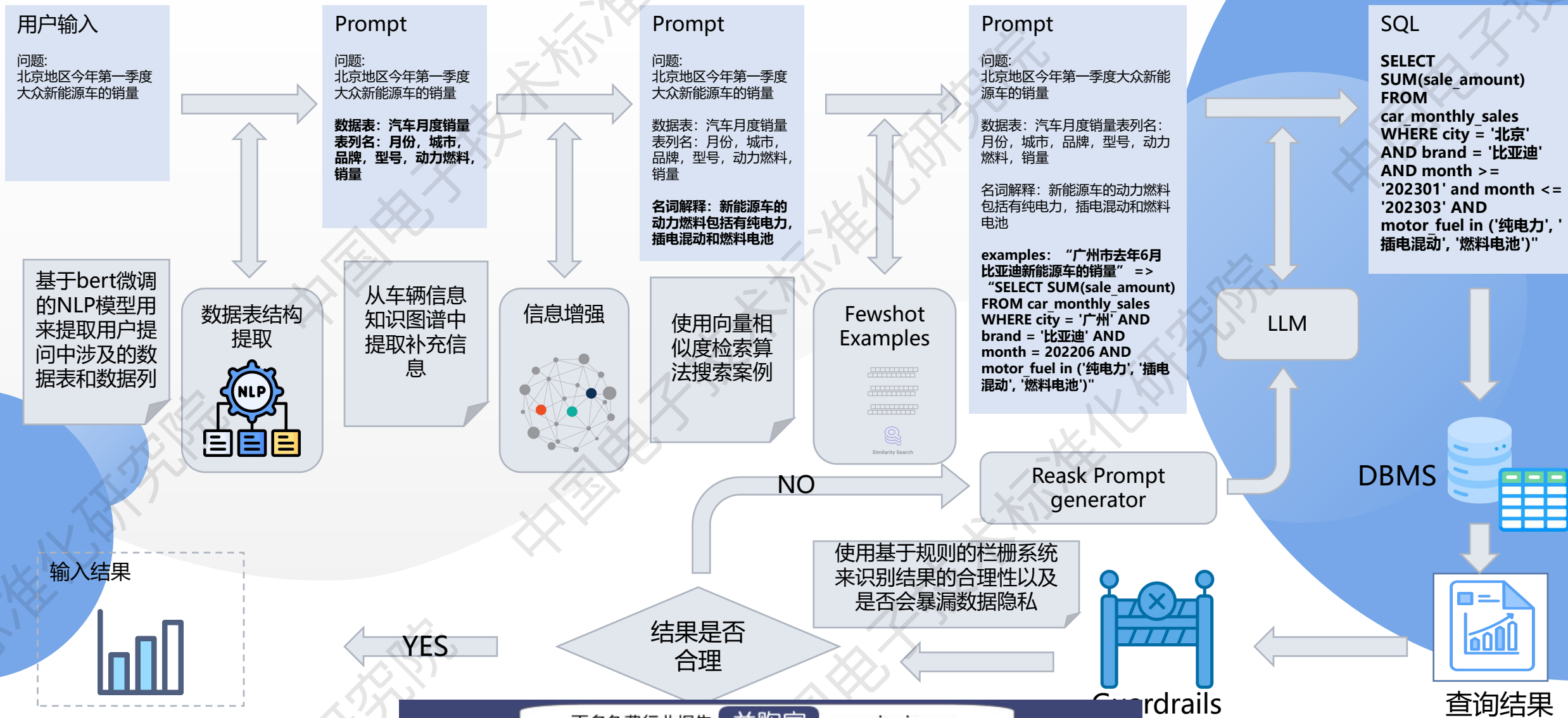
关键技术:

- 自然语言处理（NLP），智能问答。



5.汽车行业实践案例：购车攻略平台

Natural Language to SQL



6.智能家居行业实践案例：智能家居知识泛化及交互提升

行业需求：

1) 进一步提升智能家居用户的交互体验，包括交互过程中的连续对话、语义理解、生成人性化回复；2) 解决研发人员面对的家电知识零散、知识库建设效率等现实问题，实现降本增效。

01

关键技术：

智能家居知识图谱、智能家居行业大模型、安全计算、场景生成等。

02

解决方案：

1) 利用大模型进行知识泛化，解决知识有限、获取难、知识库构建效率低等问题；2) 基于泛化后的语料，实现“任意说”（指令换说法，仍然听得懂）；3) 利用大模型的理解与生成能力，实现上下文理解、连续对话、拟人化回复。

03

提升效果：

1) 智能家居知识图谱的量级从千万提升到亿级，形成高效知识管理平台；2) 用户交互体验大幅提升，从以往控制指令说法受限、回复不精准，进化为连续交互、随意交互和引导交互。

04



7.教育出版行业实践案例：数字教材智能编创与应用系统

01

行业需求

- 1、智能生成内容：辅助编者和教师用户内容生成；
- 2、高效内容处理：通过智能系统辅助翻译、转录、汇集、润饰、评估等内容处理工作，大幅提升编辑们的工作效率；
- 3、智能推荐：用人工智能进行信息推荐，扩大其数字营销能力。

02

关键技术：

- 1、大纲和内容的自动生成；
- 2、精准用户画像自动分析与推荐。

03

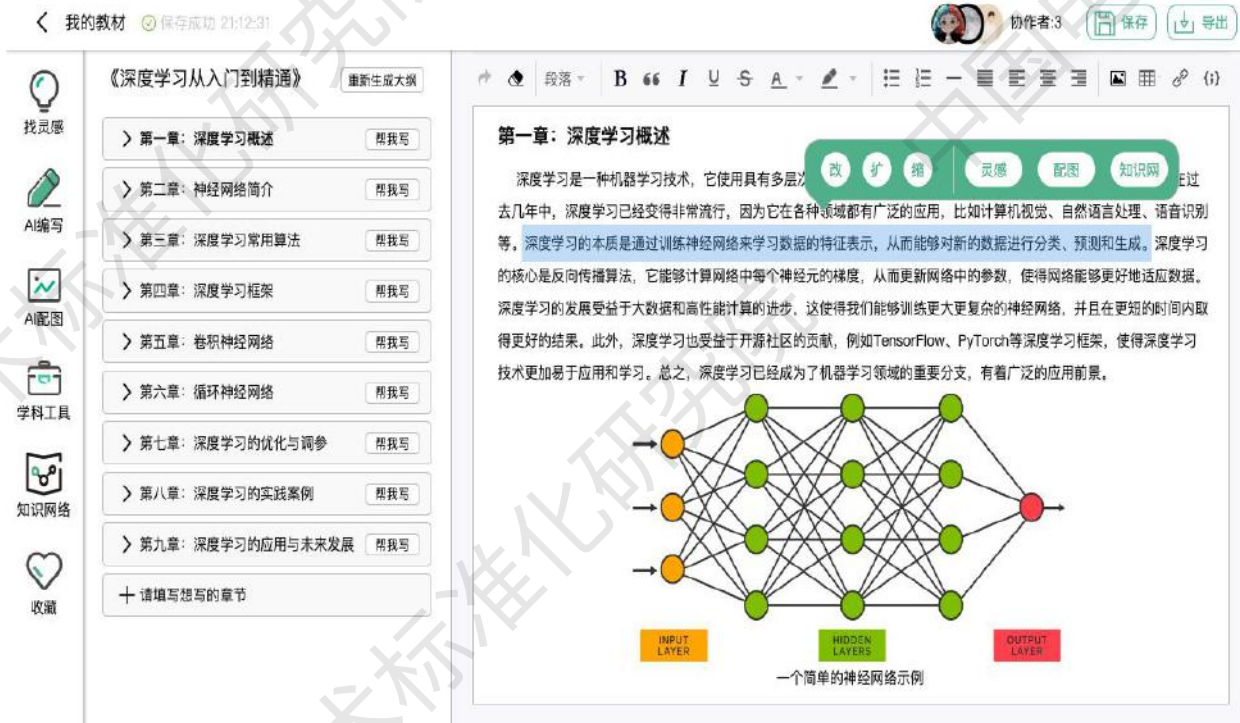
解决方案：

- 1) 基于领域知识等构建跨领域知识图谱，用大模型技术实现知识自动抽取；
- 2) 在生成式大模型提升知识图谱的知识创作能力；

04

提升效果：

通过基于智能AI系统的数字教材编创系统，为编者、编辑、教员、学生提升智能知识服务

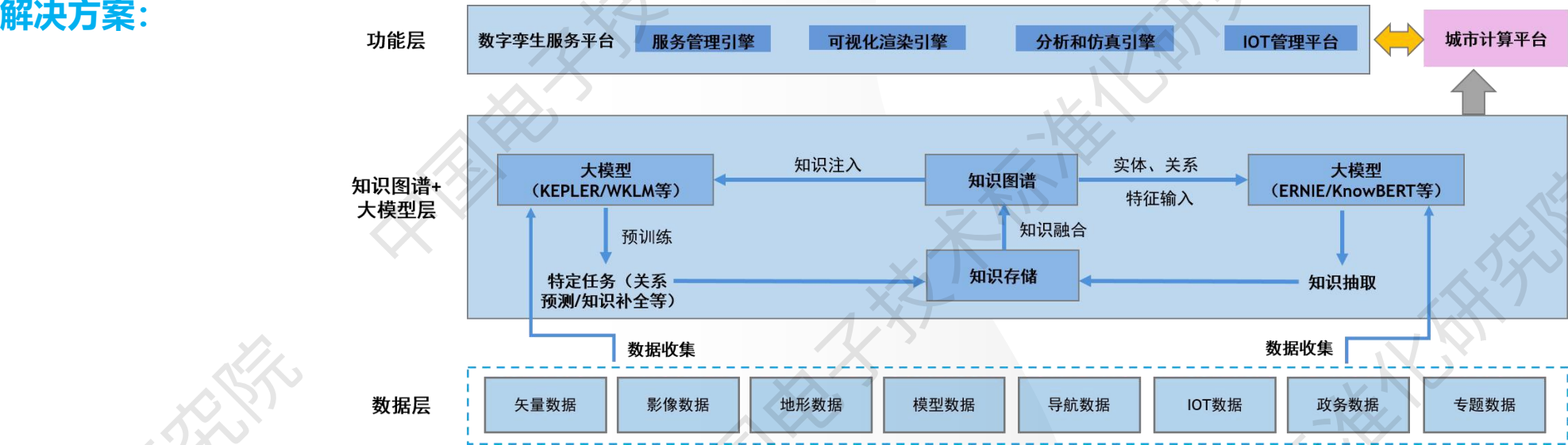


8.智慧城市实践案例：数字孪生城市服务平台

行业需求：

- 1) 在数字孪生城市行业非结构化数据急剧增多的情况下，构建知识图谱需要依赖人工或者半自动方式进行知识抽取和建模，信息利用效率低，数据分析能力不强。
- 2) 现有数字孪生城市知识图谱大部分是针对特定领域或任务定制，扩展性差。

解决方案：



基于矢量数据、影像数据、模型数据、IOT数据、专题数据等构建数字孪生城市知识图谱，结合大模型预训练提升知识图谱的知识抽取和图谱构建能力，并将知识图谱作为大模型输入，提升大模型专业性和可信性，从而利用知识图谱+大模型提升城市运营以及各领域的指挥决策能力以及准确度。

提升效果：数字孪生城市服务平台性能优化，数字孪生城市各领域的信息获取以及利用效率增大，数据分析能力有了很大的提升。

关键技术：知识注入辅助模型预训练、基于大模型的知识抽取能力

9. 社交领域实践案例：成长式个性化社交机器人

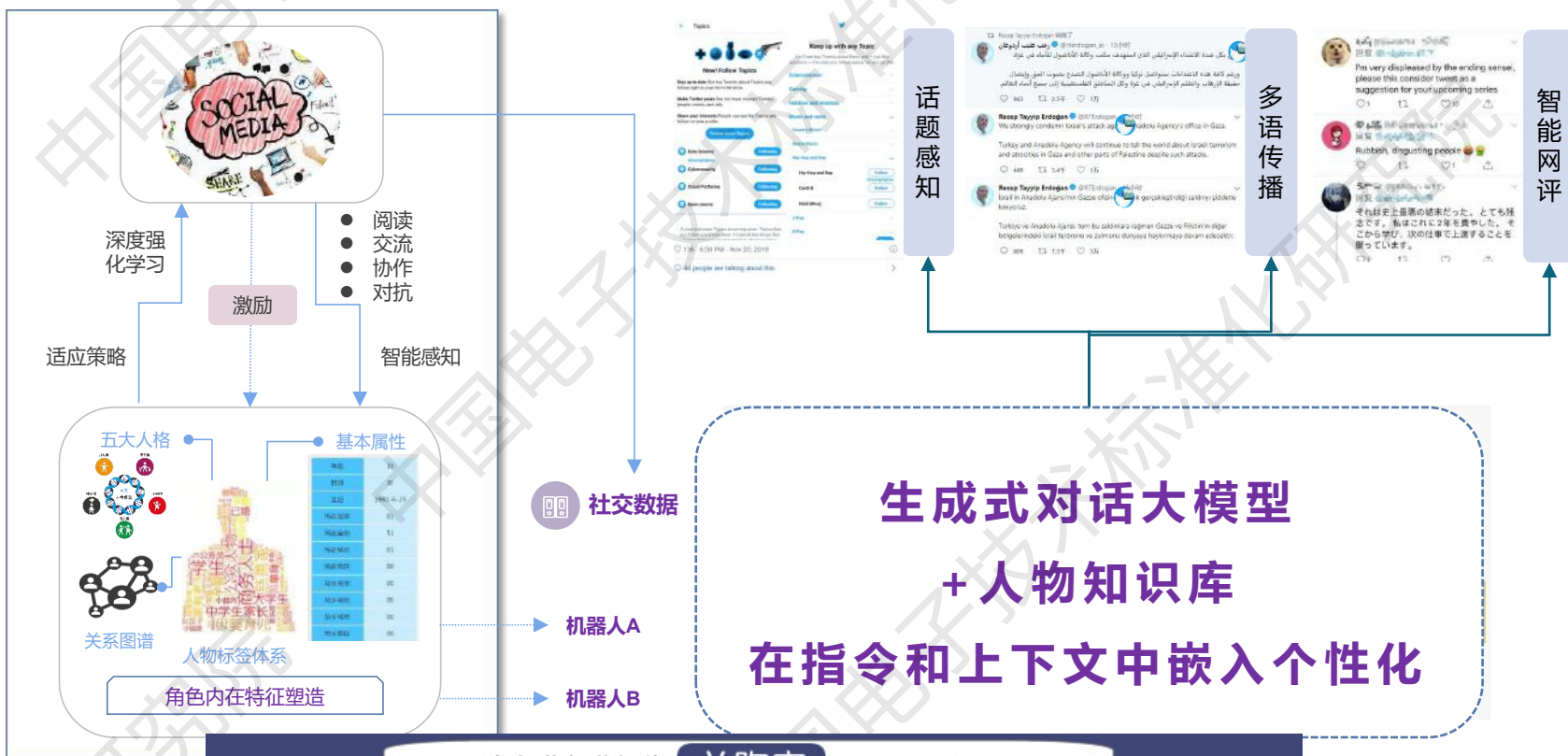
行业需求：

社交领域的智能交互机器人难点在于对社交机器人进行成长式的个性化训练，来生成语义连贯自然、富带感情观点、千人千面的多模态内容。



解决方案：

大模型以百万级人物知识库和社交媒体信息作为个性化指令数据进行精调，具备千人千面的角色学习能力。采用内在特征塑造和强化学习对抗反馈的方式不断加强与人类性格、价值的对齐。



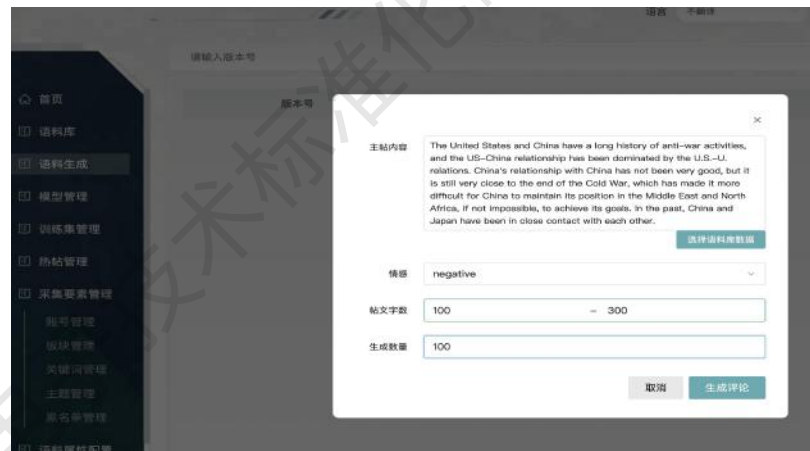
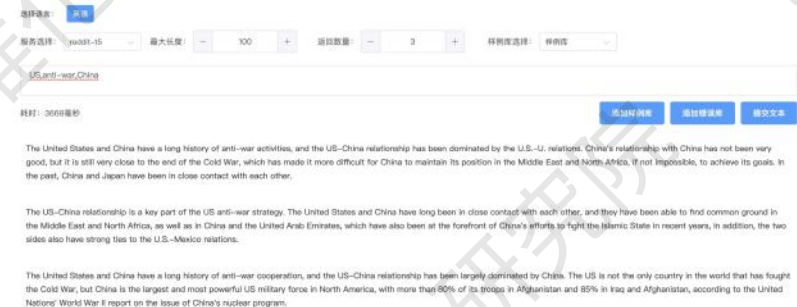
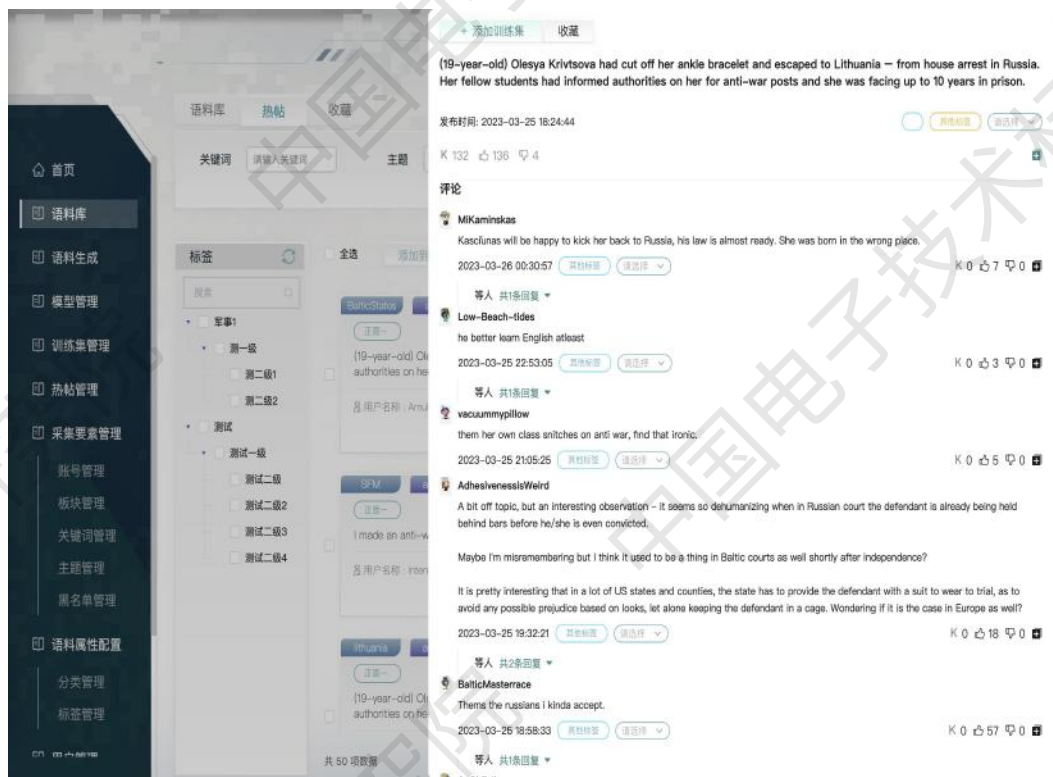
9.社交领域实践案例：成长式个性化社交机器人

提升效果：

采用内在特征塑造和深度强化学习的方式训练社交机器人，能够生成语义连贯自然、富带感情观点、千人千面的多模态内容。以Reddit为媒体平台，实现认知舆论战的贴文生产系统，根据热点、关键词进行流畅的本地化的贴文批量生成，拟人通顺度 $\geq 80\%$ ，连续生成1200条的可用度 $\geq 80\%$ ，重复率 $\leq 20\%$ ，具备根据不断变化的热点进行准实时的模型训练更新。

关键技术：

个性化训练、指令精调、强化学习



10.科学文献行业实践案例：基于大数据的智能检索

01 #ONE

行业需求:

1、搜索是信息时代的通用性刚需，可以提升用户日常行为的效率；2、提高短文本查询Query和长文本Item的语义表达能力与理解能力，给用户提供更好的搜索体验。

02 #TOW

解决方案:

利用知识图谱及LLM大语言模型，识别用户查询意图、生成语义向量，并进行向量检索，同时基于知识图谱进行关联分析，得到关联推荐结果。

03 #THREE

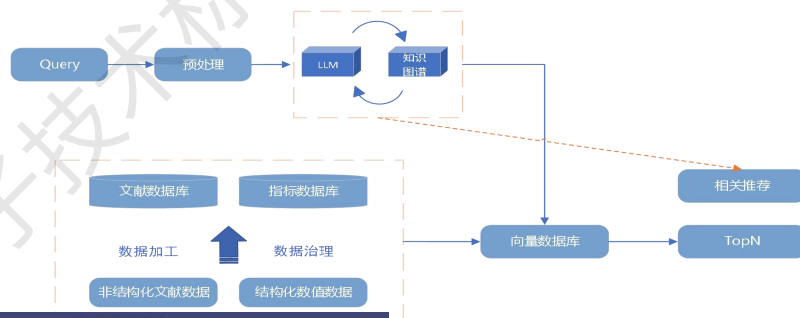
关键技术:

1、面向指标数据、文献数据的查询意图精准识别；2、面向指标数据、文献数据的语义向量检索

04 #FOUR

提升效果:

1、基于大模型的搜索系统的准确率，相比原系统同比提升13%，且大幅降低了人工维护成本；2、大模型赋予搜索更强的自我学习能力，能够持续优化输出结果，更好贴合用户使用习惯，更具个性化。



11.水务行业实践案例：基于大模型和知识图谱的智慧水利知识平台

基于大模型和知识图谱的知识平台是智慧水利的智能支撑，通过构建水利领域大模型，融合知识图谱技术，面向水务领域知识，形成以知识引擎为核心的事理推演，支撑服务及应用场景包括：

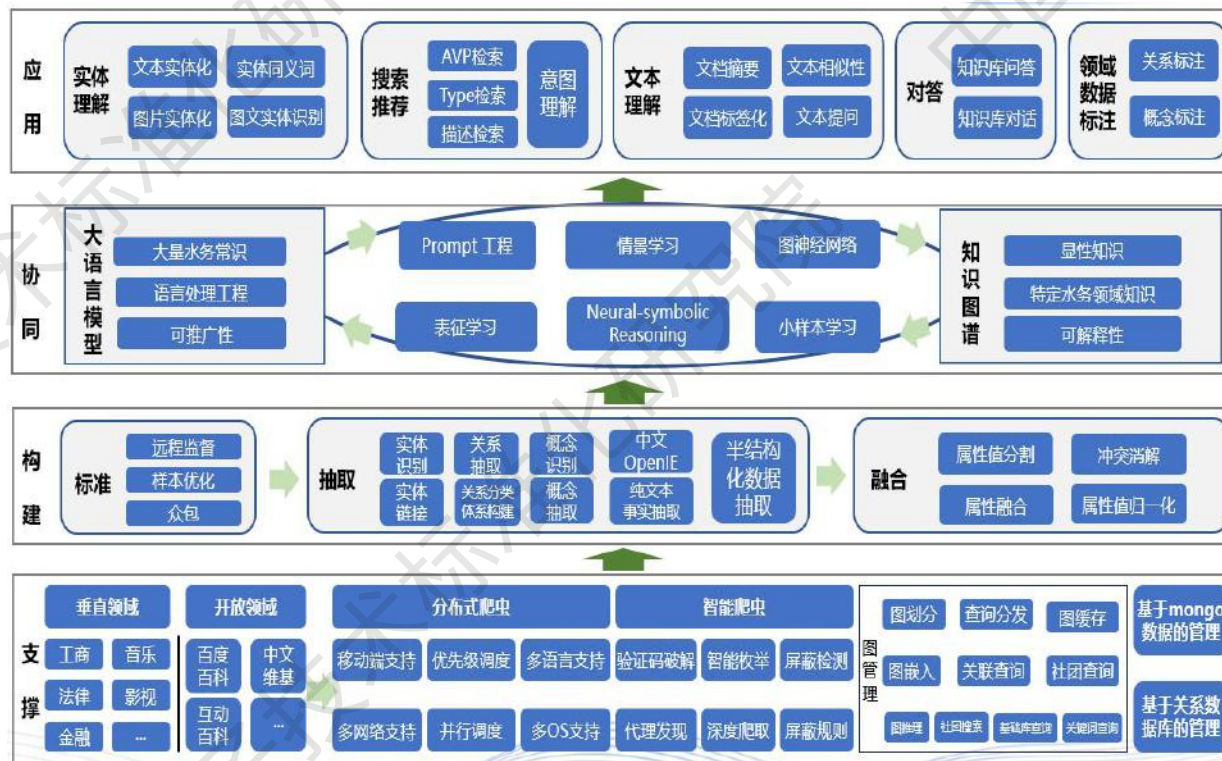
场景一：政务（水务方向）智能问答

行业需求

各种关于水务相关的在线咨询需要人工解答，查找答案时费力，人工客服容易面临相同问题回答不一致或者回答不及时的问题。

关键技术

大模型语义相似度计算、信息抽取、预训练模型语义相似度计算技术。



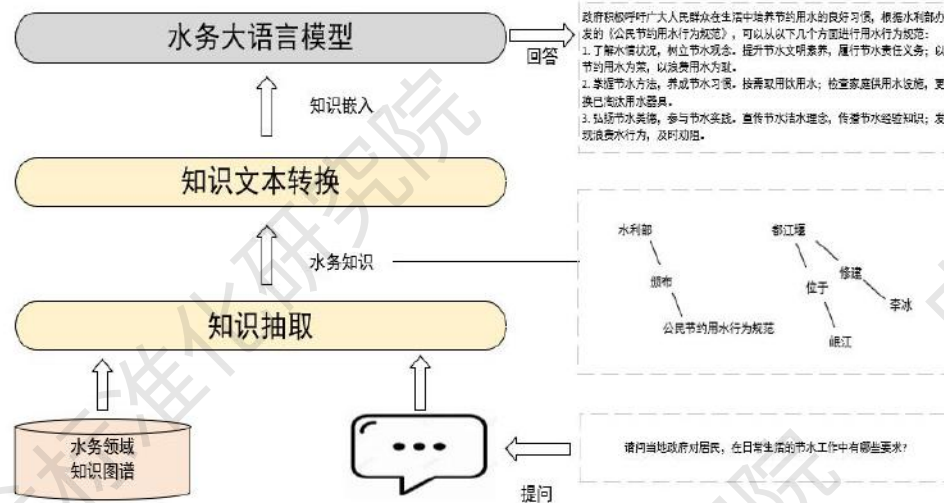
11.水务行业实践案例：基于大模型和知识图谱的智慧水利知识平

解决方案：

基于应急预案、政策等构建水务知识图谱，并构建基于大模型的智能问答系统，从而利用预训练模型语义计算技术智能识别用户的意图，给出针对性的解决思路或答案，并实现从水务知识图谱中快速检索出准确的答案，提升客服服务效率。

提升效果：

基于智能AI机器人（硬件）和大屏的水务方向政务智能问答系统，在线回答时效性提升60%，回答准确率显著提高，且已支持多层问答，语音输入，并基于在线文字及语音理解的生成式多模态图表技术，实现了机器人和大屏的在线联动，数字化大屏展示等效果。



11.水务行业实践案例：基于大模型和知识图谱的智慧水利知识平台

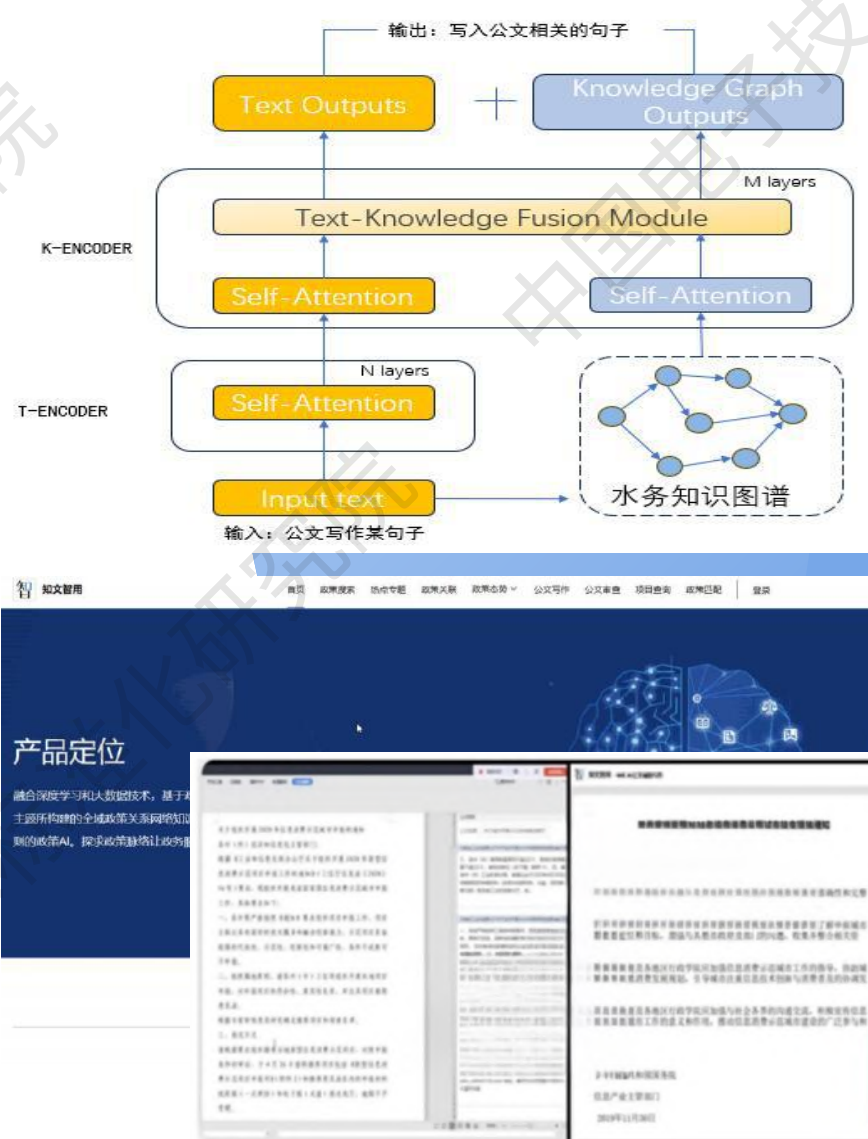
场景二：水务政策公文服务

行业需求：水务相关政策公文面临素材搜寻难、政策发布难、政策宣传难、政策申报难、政策统计繁等问题。

解决方案：基于NLP、知识图谱、大模型技术，构建融合政策、法规、公文、解读、机构、主题等要素构建全域政策关系网络知识图谱，将经验/知识转换为规则政策。

关键技术：政策公文语义搜索、文档解析信息抽取、政策文本关联技术

提升效果：水务政策知识平台（知文智用）智能提供政策语义搜索、公文标引、智能审核等应用，实现公文辅助写作，公文写作联想，相关插件可集成WPS等办公软件，支持公文初稿拟制、河长制日报周报、预警事件处置报告、应急预案等多种文体的自动生成。



11.水务行业实践案例：基于大模型和知识图谱的智慧水利知识平台

场景三：基于大模型的数字孪生水利防洪推演预测系统

行业需求：山洪流域防洪需要：精准的预报预测分析、预警消息及时触发并发布、水利应用场景仿真推演、应急预案快速形成并择优。

关键技术：水利数据演算分析技术、基于仿真引擎及可视化模型双向渲染技术、数字孪生

提升效果：结合大模型技术驱动水利防洪，实现山洪“四预”

解决方案：利用大模型技术驱动水利行业专项业务更精准的预报预测分析，结合数字孪生场景实现水利工程实体及单元部件预警消息的空间关联绑定及消息查看，结合大模型技术实现基于仿真引擎及可视化模型双向渲染驱动下的数字孪生水利应用场景仿真推演，基于场景预演结果，实现以知识平台驱动下的调度方案推送，辅助最优预案决策。



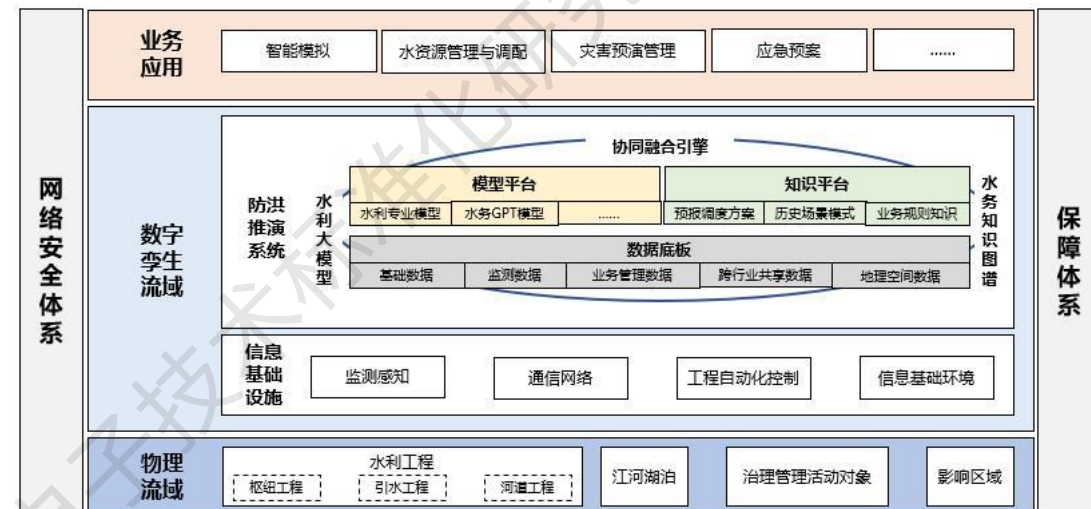
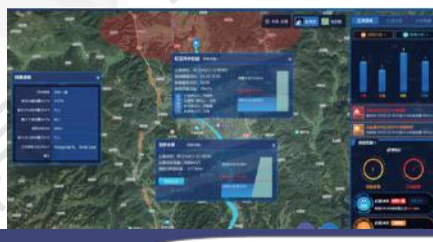
精准超前预报



快速直达预警



前瞻科学预演



第六章

技术挑战与发展展望



技术挑战

01

知识图谱与大模型的增强和效能提升



1. 基于知识图谱与大模型的融合，实现知识图谱的自动构建、架构动态拓展与自动运维。
2. 通过知识图谱与大模型的融合，降低对算力、存储等资源的需求，优化运行效率。
3. 利用知识图谱与大模型的融合，提升知识更新效率。
4. 通过知识图谱与大模型的融合，实现行业大模型的高效构建。
5. 基于知识图谱的结构化知识与逻辑推理能力，增强大模型的可解释性与推理能力。
6. 基于知识图谱增强的大模型，优化解决不确定性问题，提升决策的准确性和效率。

1. 利用知识图谱与大模型的融合，实现对复杂业务场景的深度理解和精准响应。
2. 通过大模型与知识图谱的构建及融合，实现更广泛的多模态应用。
3. 利用知识图谱增强的大模型，实现内容的自动化审查机制。
4. 通过知识图谱与大模型的融合应用，实现面向特定领域的安全保障机制。



知识图谱与大模型的应用和安全保障

02

发展展望

政策层面

- 
1. 建议围绕大模型，加大建设投入与政策保障，纳入国家新型基础设施；
 2. 建议针对大模型，建立国家级的研发中心/基地，提供公开的计算资源、研发资源等，推动中小企业开展研发工作；
 3. 建议围绕知识图谱和大模型融合的数据安全、隐私保护、知识产权保护、伦理等，完善相关法规；
 4. 建议从政策层面，针对国产大模型，开展研发与推广应用的支持。

产业层面

1. 建议针对产业需求，开展知识增强大模型的建设，以促进大模型的产业应用；
2. 建议围绕大模型与知识图谱融合应用，开展行业数据库的打造；
3. 建议根据产业需求，开展开源训练数据集和知识图谱的建设。

标准化层面

1. 建议围绕互操作、数据传输与共享、计算资源等技术领域，开展通用标准制订工作；
2. 建议针对行业应用需求，开展行业标准规范的制订工作。

前期工作介绍

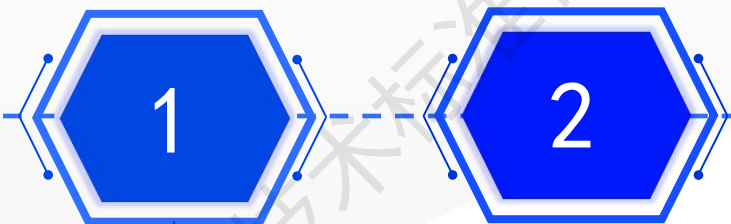
知识图谱产业推进方阵简介

知识图谱产业推进方阵旨在培育和壮大知识图谱领域供应商、集成商、服务商与用户企业，以标准化为纽带，共同促进知识要素在各行业领域的挖掘、富集、流动和应用，推动构建跨行业、跨领域的知识挖掘与应用服务新型基础设施。

方阵成员包括理事长单位、成员单位，并设置轮值主席、专家委员会、秘书处及必要的工作组。方阵将通过供需对接、诊断评估、测试认证、标准宣贯、教育培训、知识交换协议开发等手段服务产业，不定期开展技术沙龙、案例征集、成果发布、专题竞赛、产业峰会等活动，推动知识图谱的技术创新和产业深化应用。

请有意向的单位填写《方阵成员单位申请表》提交至 liq@cesi.cn，经秘书处形式审核及理事长会议审议通过后，将颁发成员单位证书。申请表下载链接如下：

<https://www.jianguoyun.com/p/DXWfUV8Qns-NChil76AE>



知识图谱标准化工作组简介

全国信标委人工智能分委会知识图谱工作组及IEEE知识图谱标准化工作组，由中国电子技术标准化研究院牵头，联合知识图谱相关企事业单位、科研院所、高校、机构，旨在运用标准化的理念、方法和技术梳理分析知识图谱领域核心标准化需求，共同推动知识图谱关键标准的研制等工作，支撑知识图谱技术的高质量推广与应用。工作组现有清华大学、阿里巴巴、联想、华为、百度、腾讯、东软、蚂蚁科技、依图等70余家知识图谱领域相关单位共同参与标准编制工作。目前，已发布GB/T 42131-2022《人工智能 知识图谱技术框架》等国家标准、IEEE标准3项，在研标准10项。

前期工作介绍

序号	标准类型	标准名称	状态
1	国际标准	ISO/IEC DIS 5392Information technology — Artificial intelligence — Reference architecture of knowledge engineering 《信息技术 人工智能 知识工程参考架构》	在研
2	国家标准	《人工智能 知识图谱技术框架》 国家标准号：GB/T 42131-2022	已发布
3	IEEE标准	Framework of Knowledge Graphs 《知识图谱架构》 IEEE标准号：IEEE Std 2807-2022	已发布
4	IEEE标准	Standard for Technical Requirements and Evaluating Knowledge Graphs 《知识图谱技术要求及测试评估规范》 项目号：P2807.1	在研
5	IEEE标准	Guide for Application of Knowledge Graphs for Financial Services 《金融服务领域知识图谱应用指南》 项目号：P2807.2	已冻结
6	IEEE标准	Guide for Electric-Power-Oriented Knowledge Graph 《面向电力行业的知识图谱指南》 IEEE标准号：IEEE Std 2807.3-2022	已发布
7	IEEE标准	Guide for Scientific Knowledge Graphs 《科技知识图谱指南》 项目号：P2807.4	在研
8	IEEE标准	Guide for Medical Clinical Diagnosis and Treatment Oriented Knowledge Graphs 《面向临床诊疗的知识图谱指南》 项目号：P2807.5	在研
9	IEEE标准	Guide for Open domain Knowledge Graph Publishing and Crowdsourcing Service 《开放域知识图谱发布与众包服务指南》 项目号：P2807.7	在研
10	IEEE标准	Standard for knowledge exchange and fusion protocol among knowledge graphs 《知识图谱间知识交换与融合协议》 项目号：P2807.8	在研
11	团体标准	《人工智能 知识图谱 分类分级规范》 项目号：CESA-2020-019	在研
12	团体标准	《人工智能 知识图谱 性能评估与测试规范》 项目号：CESA-2020-020	在研
13	团体标准	《人工智能 医疗知识图谱 构建要求》 项目号：CESA-2023-023	在研
14	团体标准	《人工智能 医疗知识图谱 测试评估要求》 项目号：CESA-2023-024	在研
15	白皮书	《知识图谱标准化白皮书》	已发布
16	案例集	《知识图谱赋能疫情防控与复工复产案例集》	已发布
17	案例集	《认知智能时代：知识图谱实践案例集》	已发布
18	白皮书	《知识图谱选型与实施指南》	已发布
19	白皮书	《知识图谱互联互通白皮书》	已发布
20	研究报告	《知识图谱与大模型融合实践研究报告》	

知识图谱相关成果简介

知识图谱产品认证简介

知识图谱产品认证旨在强化和提升知识图谱相关产品的质量，规范知识图谱在通用及医疗、金融等领域构建与应用过程中的性能指标和功能要求，建立优质知识图谱产品名录，促进行业良性发展，实现技术创新与标准的有机结合。

依托GB/T 42131-2022《人工智能 知识图谱技术框架》等系列国家标准和团体标准，中国电子技术标准化研究院联合北京赛西认证公司等40余家单位研制了《知识图谱构建平台认证技术规范》、《知识图谱应用平台认证技术规范》等基础知识图谱产品认证技术规范，并研制了《金融领域知识图谱构建能力认证技术规范》、《医疗领域知识图谱应用能力认证技术规范》等领域知识图谱认证技术规范，共设置300余项测评指标。

现已有联想、华为、百度、蚂蚁科技、清华大学、中国医学科学院医学信息研究所、科大讯飞等30余家单位的知识图谱系统通过首批、第二批和第三批基础知识图谱产品认证，首批医疗领域知识图谱产品认证。获批使用的认证标识如下：



基础知识图谱产品测评与认证介绍：

<https://mp.weixin.qq.com/s/m9ysbUQu0H0jQ00Je6g14A>

金融与医疗领域知识图谱产品测评与认证介绍：

<https://mp.weixin.qq.com/s/KHy536ZK5je71wBQ7geg4g>

更多内容详见“知识图谱标准化”公众号和视频号。



知识图谱与大模型融合实践研究和评估、知识图谱产业推进方阵、
全国信标委人工智能分委会知识图谱工作组、IEEE知识工程标委会、
知识图谱产品认证相关动态敬请关注微信公众号或与联系人沟通。

中国电子技术标准化研究院

联系人: 李瑞琪

联系方式: 010-64102797

电子邮箱: lirq@cesi.cn

