

一图胜万言，一目了然

中国知识图谱行业研究报告

©2022.8 iResearch Inc.



前言

研究背景：

2020年，艾瑞已经发布了《2020年中国知识图谱行业研究报告》、《2020年面向人工智能“新基建”的知识图谱行业白皮书》，初步对知识图谱的概念定义、行业场景、整体市场规模与产业链等进行了梳理与分析。

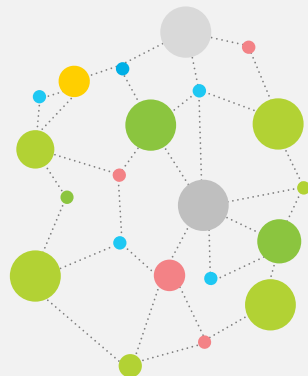
2022年，艾瑞将发布第三篇知识图谱行业报告，深入探讨知识图谱技术在通用知识图谱与行业知识图谱的企业诉求，剖析各行业的核心业务痛点，阐明知识图谱技术与应用对业务痛点的针对性与价值，对比各行业的知识图谱发展现状与潜力，展现知识图谱技术的建设重点与未来趋势。希望通过本报告，为读者呈现现阶段知识图谱行业发展的要点，提供辨析和判断知识图谱行业趋势的方法启发。当然，限于研究周期与行业理解水平，报告仍有不足之处，敬请各界读者指正。

研究对象：

本报告研究对象分为：①通用知识图谱与行业知识图谱，包含互联网、金融、政务与公安、医疗、工业与电力5大行业的行业痛点、知识图谱应用场景及价值、市场规模；②知识图谱现阶段的行业热点与建设痛点；③知识图谱的行业参与者类型及业务特点。

研究方法：

本报告通过业内资深的专家访谈、桌面研究、案例实证研究、行业对比研究、投融资数据统计与行业规模数据推算输出相应研究成果。



报告撰写

艾瑞咨询
产业数字化研究部
人工智能研究组



行业背景

感知到认知的跨越式发展，须引入发展认知技术，**知识图谱**在此形势下成为了破局的关键技术。**数字经济的持续发展**将加速知识图谱产业化进度，推动知识图谱与传统产业融合。**高性能图计算**可为图计算输送更快更准的计算能力，服务于知识图谱运算，**算力规模化部署**也为知识图谱计算的高密度、高功耗要求提供了有利发展条件。深度学习技术、NLP技术、知识图谱技术**协同并进**，尤其是**NLP技术**近几年的快速发展为知识图谱产业化提供了机会。



行业规模

2021年，知识图谱核心市场规模预计达到**107亿元**，而到2026年，相应规模将超过**296亿元**，**2021-2026年CAGR=22.5%**。**金融与公安**两大行业的知识图谱占比较高且增长速度较快，其业务与知识图谱可密切结合，同时具备建设意愿与资金投入，因而成为了市场规模的主要拉力。未来，随着政务数字化建设的完善，**政务**对知识图谱的业务需求会逐渐唤醒，成为未来市场的拉力之一。



建设重点探讨

知识图谱建设需面临的建设难点主要在于**数据治理、行业专家储备、底层图数据库存储、算法生产流程与性能待提升、客户认知待培养以及产品封装形式待优化**。攻克知识图谱的建设重难点将有利于从源头保证知识与智慧真实可靠、可用正确，储备培养深厚行业专家与技术复合型专家，升级底层图数据的存储方式，改善算法性能，为知识图谱建设减少阻碍。



趋势展望

未来，知识图谱厂商、大数据厂商、NLP厂商、互联网大厂与信息化厂商等知识图谱业内参与者将从**强化技术实力与深化行业认知**的角度出发，结合自身原有业务优势，持续深化发展行业知识图谱业务。知识图谱业务场景也将不断迭代，**行业应用场景边界拓宽，垂直应用场景被做深做透**。知识图谱生态也将继续由监管引导方、供给方、需求方、投资方、高校及科研院所融合共建，**汇聚建设合力**，促进产业生态成长壮大。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

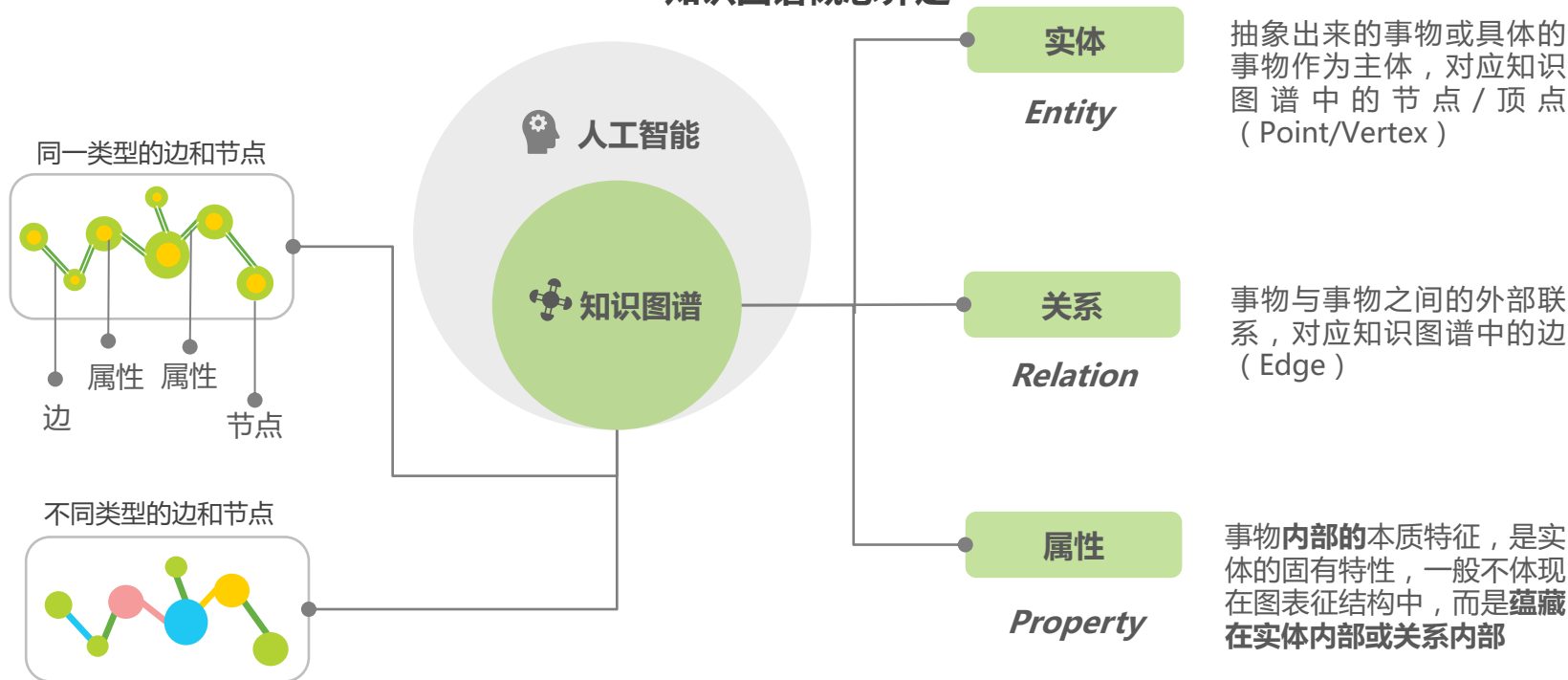
5

概念与研究范围界定

描绘实体之间关系的数据结构语义网络

知识图谱是人工智能的一大底层技术，是描绘实体之间关系的语义网络，自带语义、逻辑含义和规则，通过三元组即“实体×关系×属性”集合的形式来描述事物之间的关系。知识图谱将非线性世界中的知识信息结构化、可视化，辅助人类进行推理、预判、归类。知识图谱中的图并非图像概念，而是类似化学分子式的结构，一个知识图谱往往存在多种类型的实体与关系。知识结构网络化、网络结构复杂、网络由三元组构成、数据主要由知识库承载是知识图谱的四大基本特征。一般而言，知识图谱的数据以文本化数据为主，数据化数据为辅。

知识图谱概念界定



来源：艾瑞咨询研究院根据公开资料自主研究绘制。

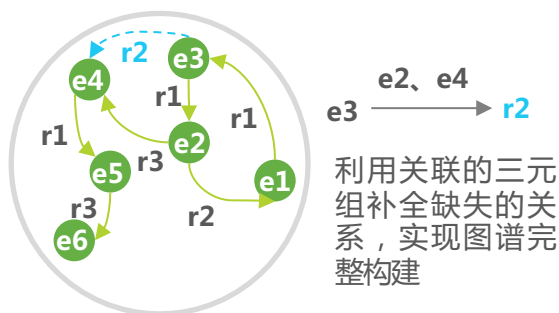
主要任务介绍

五大任务，复杂关联推理为突破重点

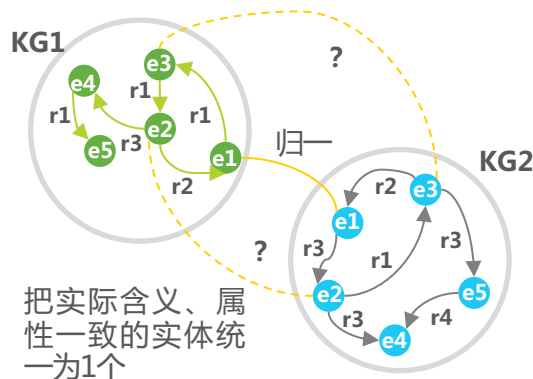
一般而言，知识图谱需执行的主要任务包括知识图谱构建与补全、实体统一（消歧）、实体分类、知识检索问答（简单推理）、复杂关系推理。现阶段的复杂关系推理需要更多依赖人类预测与推断各种可能的情况，并优先推荐可能性大的情况。

知识图谱主要任务

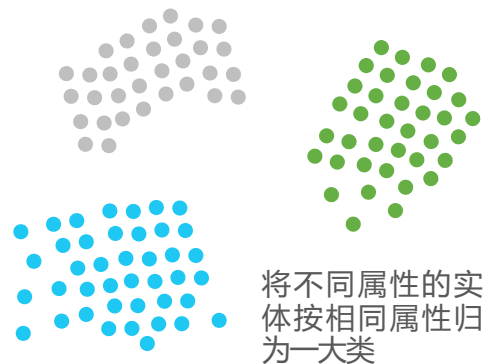
Task1 : 知识图谱构建与补全



Task2 : 实体统一（消歧）



Task3 : 实体分类

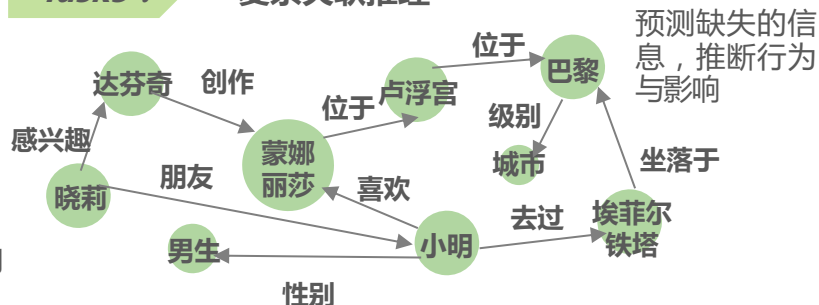


Task4 : 知识检索问答（简单推理）

- 水和二氧化碳反应可以生成什么？
- 《复仇者联盟》黑寡妇的扮演者是谁？
- 并联电路各支路电压的特征？

- 碳酸
- 斯嘉丽 约翰逊
- 各支路电压相同

Task5 : 复杂关联推理

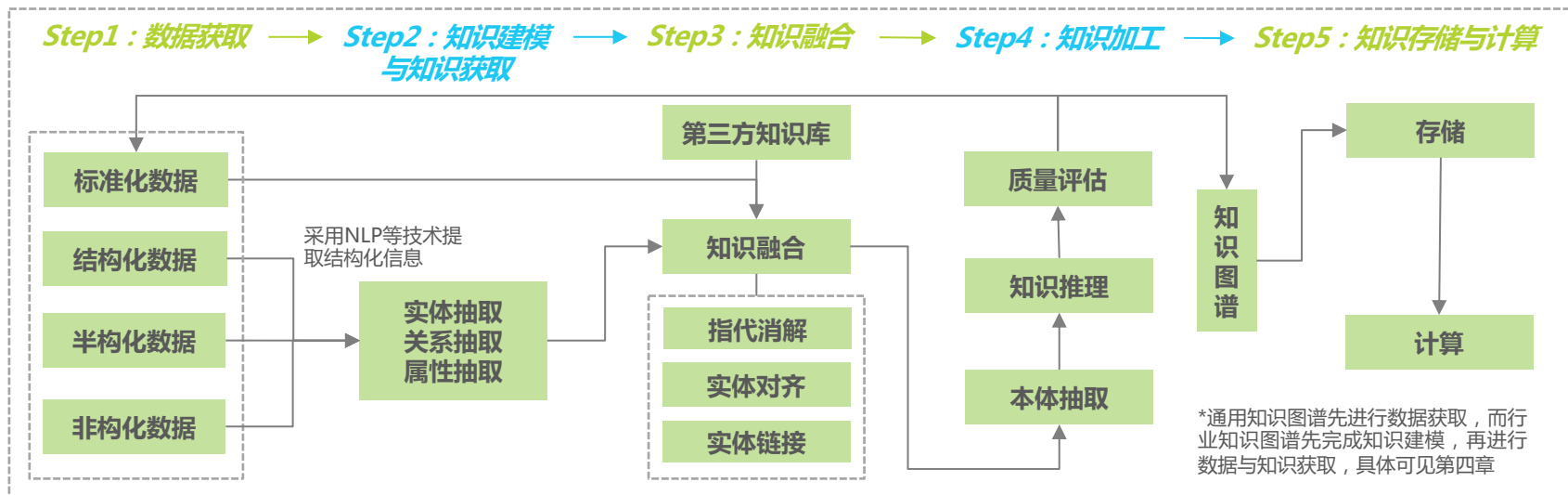


技术架构原理

技术架构分五步走，旨在构建实体语义网络

- **数据获取**：主要获取半结构化数据，为后续的实体与实体属性构建做准备。结构化数据则为数值属性做准备。
- **知识获取**：①从文本数据集中自动识别出命名实体，包括抽取人名、地名、机构名等；②从语料中抽取实体之间的关系，形成关系网络；③从不同的信息源中采集特定的属性信息。
- **知识融合**：①完成指示代词与先行词的合并；②完成同一实体的歧义消除；③将已识别的实体对象，无歧义地指向知识库中的目标实体。
- **知识加工**：①构建知识概念模块，抽取本体；②进行知识图谱推理，并对知识图谱的可信度进行量化评估，评估过关的知识图谱流入知识图谱库中存储，评估不过关的知识图谱返回一开始的数据环节进行调整，而后重复相同环节直到评估过关。
- **知识存储与计算**：存储是为了快速查询与运用知识，需支持底层数据描述与上层计算，有的主体计算包含在存储中。

知识图谱技术架构图



来源：艾瑞研究院根据公开资料、专家访谈自主研究绘制。

主流产品类型

成熟产品：通用互联网知识图谱；起步产品：行业知识图谱

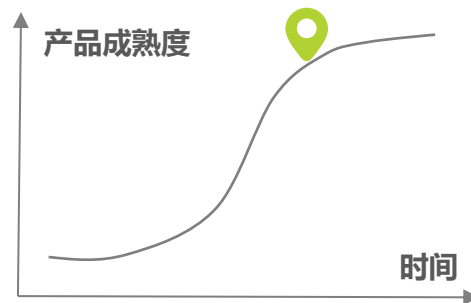
知识图谱的产品类型以通用知识图谱与行业知识图谱为典型代表。通用知识图谱经过开拓性构建阶段后，逐渐演变为通用互联网知识图谱，形成搜索引擎、智能推荐、智能问答三大产品类型，产品发展较为成熟。行业知识图谱处于起步期，但其价值及效果逐渐被客户所认可，是知识图谱当前乃至未来一段时期内的发展热点。

通用知识图谱与行业知识图谱

强调知识广度：通用知识图谱

- 有哪些产品类型？>>> 搜索引擎、智能推荐、智能问答
- 覆盖哪些知识？>>> 通用/垂类泛知识、百科常识、泛学科领域知识
- 用于哪些场景？>>> 互联网的信息检索、商品/服务/视频/交友推荐、KBQA、KGQA等

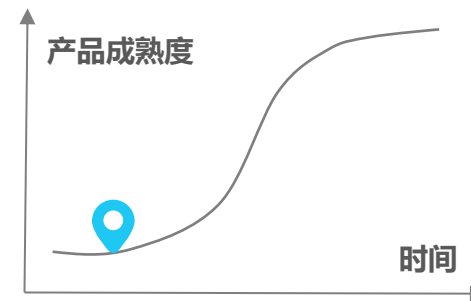
产品处于成熟期



强调知识深度：行业知识图谱

- 有哪些产品类型？>>> 行业知识图谱解决方案
- 覆盖哪些知识？>>> 某一领域内专业的知识，如医学领域的脑卒中诊断依据
- 哪些行业已经投入使用？>>> 金融、公安、医疗、电力、军工等

产品处于起步期



来源：艾瑞研究院根据专家访谈自主研究绘制。

产品形态解析

满足其一即可认定为知识图谱产品，五种形态可相互嵌套

知识图谱的产品形态颇有“盲人摸象”之意。各类厂商分别通过自然语言处理、知识库、数据库、数据平台或中台、机器学习等产品逐步接触到知识图谱，在已有的业务基础上叠加知识图谱产品，或开发出独立的知识图谱产品业务线。就当前的五大产品形态而言，其中的任意一种都可算作知识图谱产品，且知识图谱产品一般为五类产品形态的排列组合复合体。

主流的知识图谱产品形态

起源：网络搜索引擎

- **知识范围**：广域网中的通用知识
- **目标**：形成通用知识图谱，使检索更准确，并进行相关推荐
- **特征**：通用性强，适合生活领域，需强大的爬虫能力
- **举例**：百科词条搜索、关键词相关推荐与问答

点睛：知识图谱可视化

- **目标**：进行知识与关系网络的直观展示，让人感知到图谱的存在
- **特征**：属于知识图谱行业解决方案/大数据产品中的子模块，是产生人机交互的节点
- **举例**：可视化大屏

深化：行业知识图谱

- **知识范围**：某一领域内的数据、专业知识组合形成知识资源
- **目标**：将数据与知识转换为行业知识库，服务于辅助问答、推理、决策、判断
- **特征**：侧重于行业知识应用，具有强烈的行业属性，依赖于领域内专业专家的数据标注，每个行业都要设定特有的数据标准，需要单独构建行业知识库
- **举例**：行业知识图谱解决方案

来源：艾瑞研究院根据专家访谈自主研究绘制。

聚焦：内部搜索引擎

- **知识范围**：局域网中的专业知识
- **目标**：形成某一领域或企业内的知识搜索库，核心以搜索为主
- **特征**：通用性收缩，专业性变强
- **举例**：企业内部的业务知识检索与问答

升级：大数据知识图谱

- **知识范围**：某一领域内的大数据成为知识资源
- **目标**：完成数据治理，进行知识与关系抽取，服务于组织的后续应用与业务分析
- **特征**：是数仓与BI数据分析的形式引申，原来纯知识搜索引擎的知识图谱变为大数据组织形式
- **举例**：数据中台/平台等大数据产品的知识图谱模块

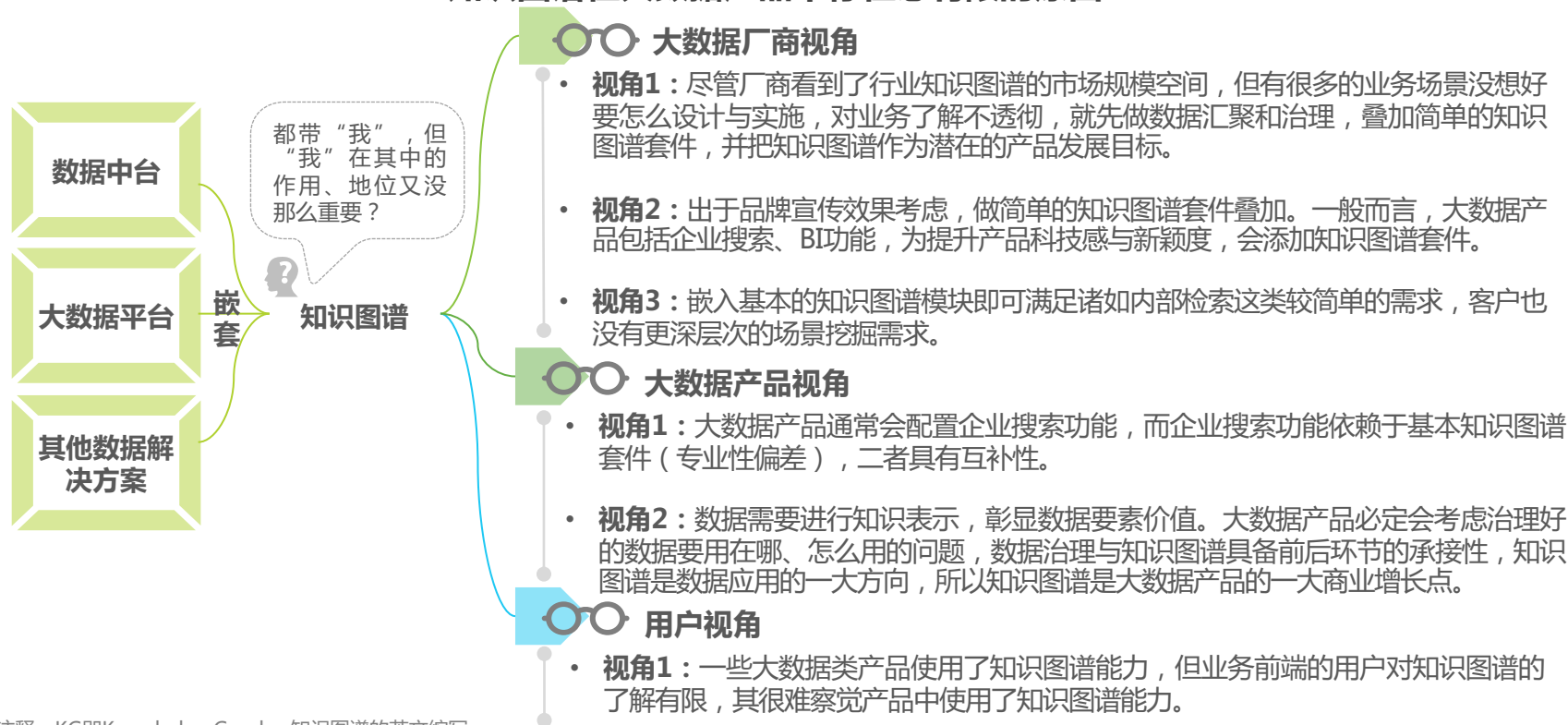
热点探讨一：KG在数据产品中的存在感

iResearch 艾瑞咨询

业务了解不透彻、产品开发承接性等原因诱发的有限存在感

在各类知识图谱的产品形态中，大数据知识图谱的数据产品属性强于知识图谱属性。在数据中台、大数据平台以及其他数据解决方案中，可频繁观察到知识图谱这一画龙点睛的模块或组件，多数大数据厂商也常常自诩自己具备知识图谱能力。然而，一旦深究其中的知识图谱行业能力、底层技术能力，多数厂商的知识图谱工具往往难以覆盖知识图谱本该具备的完整生产流程，并且缺失核心的Schema建模技术，现阶段知识图谱在大数据产品中的地位及作用就显得十分有限。

知识图谱在大数据产品中存在感有限的原因



注释：KG即Knowledge Graph，知识图谱的英文缩写。
来源：艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

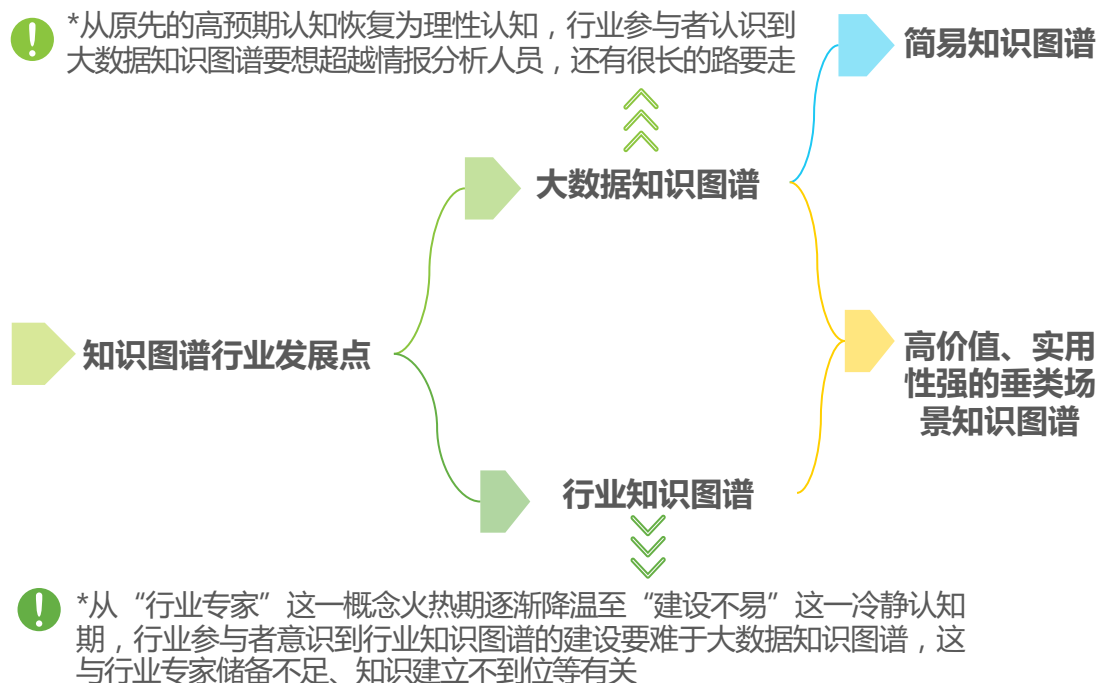
热点探讨二：场景知识图谱为发展方向

iResearch 艾瑞咨询

殊途同归：高价值、实用性强的垂类场景知识图谱

知识图谱产品在当下及未来的行业发展热点主要围绕大数据知识图谱、行业知识图谱两大类产品展开。大数据知识图谱侧，行业参与者主要分化为两大阵营，一类选择做简单的知识图谱可视化展示，不深究场景类产品的开发，另一类则在数据积累的基础上，效仿行业知识图谱厂商做场景的深度挖掘与沉淀，以期实现数据治理能力与场景产品化能力的双融合。行业知识图谱侧，知识图谱作为“行业专家”这一概念的火热潮逐渐退却，参与者在看到B端广大市场空间的同时，也深刻意识到搭建高业务价值、强专业性知识图谱的不易，因而招纳更多的业务端人才，弥补业务短板，促成技术与业务的双融合。

当下及未来行业发展热点



做简单的可视化呈现、信息检索，不深究场景开发与场景价值，此类产品的替代性高，成熟期相较于场景类图谱会较快到来

- 大数据厂商基于大数据产品积累好数据要素，用数据驱动知识工程，挖掘细分场景，聘请行业专家参与其中，做好场景沉淀，提升场景的产品化能力
- 目前行业知识图谱最大的价值领域在B端，B端的业务专家与知识图谱专家融合以后，找到合适的场景模型，高业务价值+大市场空间+强专业性的行业知识图谱不仅能够击中B端客户的需求痛点，也能为供给侧玩家指明产品路线

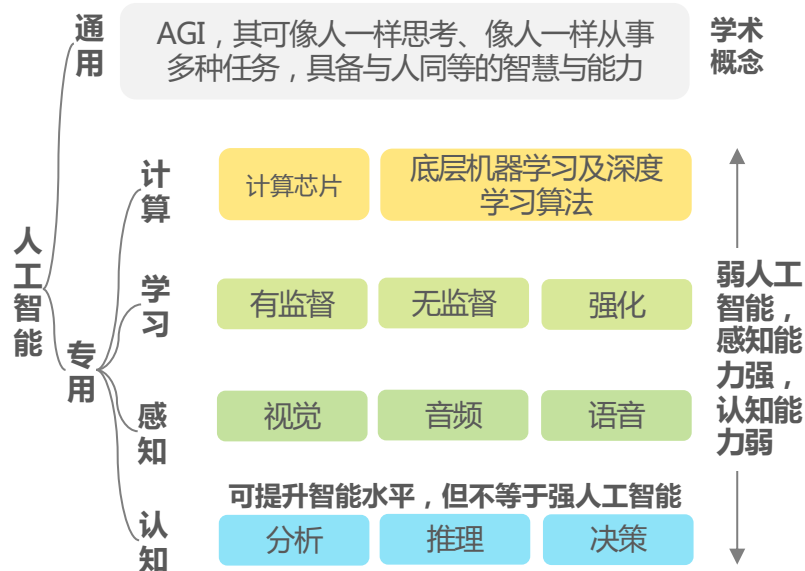
来源：艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

驱动力一：认知智能发展的关键技术

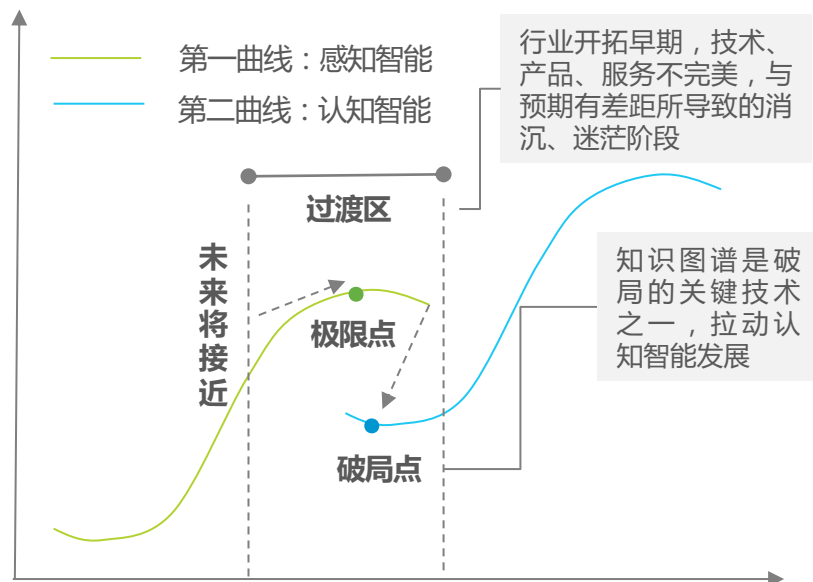
构建知识工程，推动认知类产品商业化，拉动第二曲线

专用人工智能是当前人工智能产业的主流发展方向，主要包括计算、学习、感知与认知能力，是在某个特定领域的单一人工智能，依赖大规模的标注数据进行监督训练，以单模态交互技术为典型特征，呈现出强感知、弱认知的特点。为提升当前人工智能的认知水平，需要对机器赋予知识能力，搭建高质量的知识库供机器学习训练，而知识图谱正是知识库搭建的核心技术之一。知识图谱为机器语言提供丰富的背景知识，使机器能够理解人类的语言与知识，是认知智能落地的关键技术。如果将感知智能作为人工智能产业发展的第一曲线，认知智能则为第二曲线。从感知到认知的跨越式发展，难免会经历探索的阵痛。但若想突破人工智能产业在感知领域的商业增长瓶颈，必须引入发展认知技术，掀起新一轮的人工智能产业革命，知识图谱在此形势下也就顺理成章成为了破局的关键技术。

人工智能分类及用途



感知智能与认知智能市场增量发展曲线探讨



来源：艾瑞研究院根据公开资料自主研究绘制。

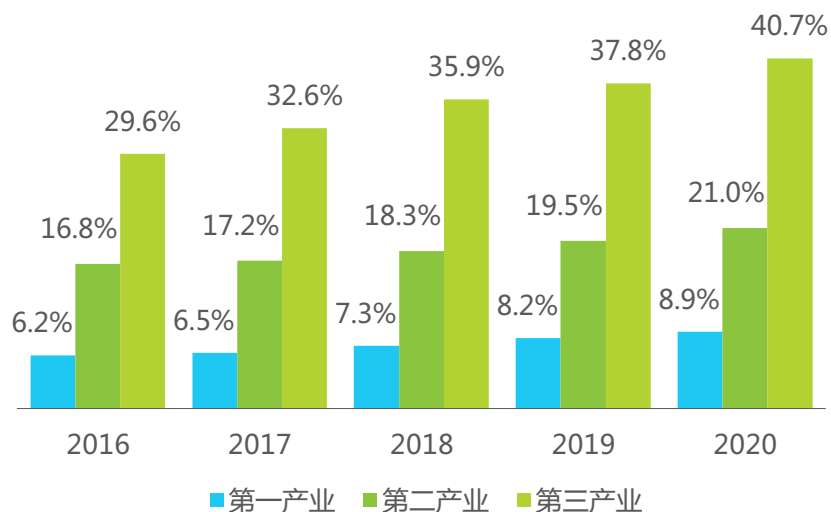
来源：艾瑞咨询研究院自主研究绘制。

驱动力二：数字经济创造发展机遇

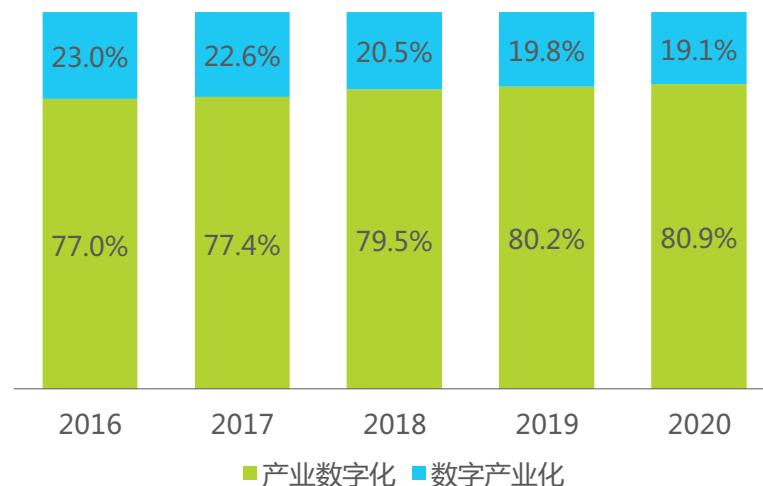
数字经济时代来临，企业数据利用意识觉醒

数字经济是以数据为关键生产要素、以现代信息网络为重要载体、以数字技术应用为主要特征的经济形态。数字经济之下，数字技术的发展与应用，使得各类社会生产活动能以数字化方式生成为可记录、可存储、可交互、可分析的数据、信息与知识，数据由此成为当代社会的新生产资料 and 关键生产要素，推进产业数字化成为了企业顺应时代发展、打造数字化优势的主动选择，而知识图谱作为产业数字化的技术工具，迎来了难得的发展机遇。这一机遇体现为企业的数据利用意识觉醒。诸多企业开始采购数字化与智能化解决方案，盘活自有数据资产，构建内部知识库与行业知识库，开发各类知识图谱场景与应用，辅助企业的产品研发、安全质量控制、风控管理、精准营销等业务，知识图谱产业也由此得到落实。数字经济的持续发展将加速知识图谱产业化进度，推动知识图谱与传统产业融合，而知识图谱产业将逐渐向传统产业横纵拓展，不断催生新场景、新应用、新模式。

2016-2020年我国三大产业数字经济渗透率



2016-2020年我国数字经济内部结构占比



来源：中国信通院《2021年中国数字经济发展白皮书》、艾瑞咨询研究院自主研究绘制。

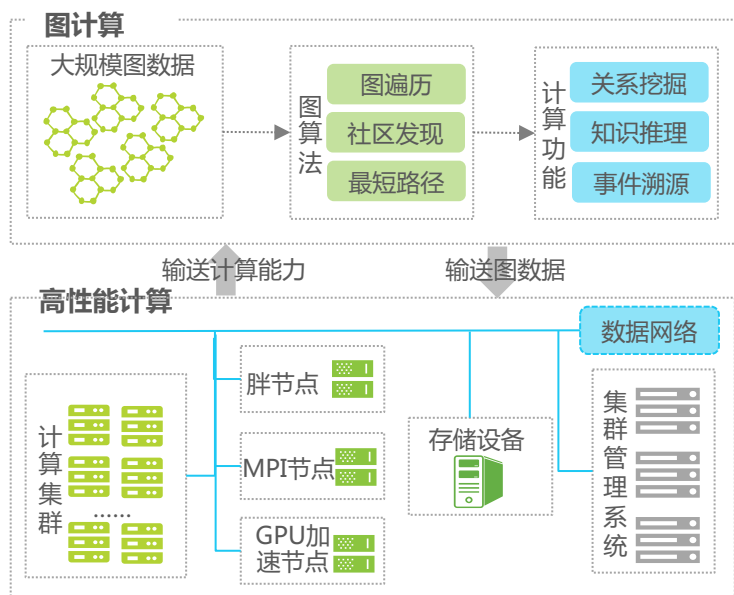
来源：中国信通院《2021年中国数字经济发展白皮书》、艾瑞咨询研究院自主研究绘制。

驱动力三：计算方式与算力规模化推动

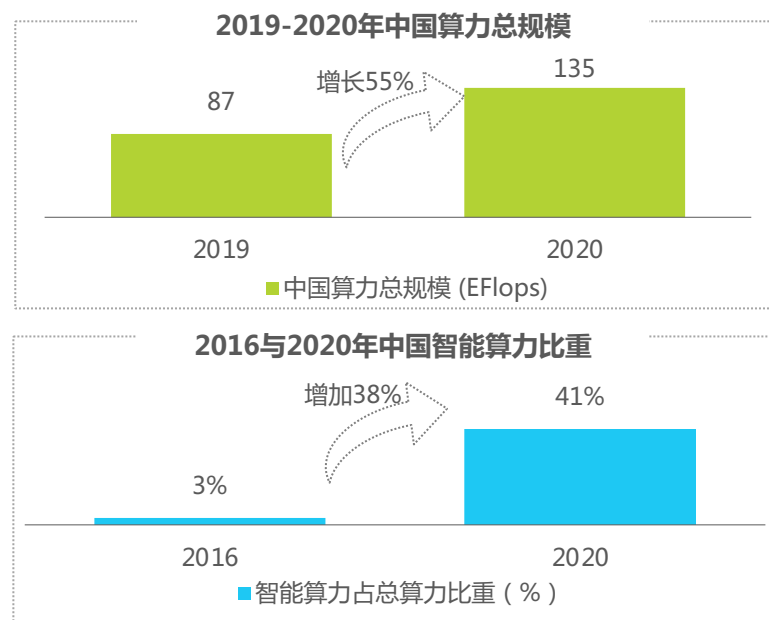
新计算方式，大算力规模，高智能比重

知识图谱为二维链接图，属于图数据结构。图达到一定规模后，对计算的要求会随之提升。若想实现低延迟、高精度的图计算，需引入高性能计算。高性能计算能够加快图的运算速度，满足对大规模复杂图数据的实时处理与存储需求。目前主流的HPC架构能够支持异构结点，性能接近超级计算系统，可为图计算输送更快更准的计算能力，服务于知识图谱运算。算力规模的持续扩大及算力结构的不断演化也助推着知识图谱发展。知识图谱对海量数据的处理与读写操作意味着对内存存取与带宽的高密度、高功耗，知识图谱算法包含大量卷积、全联接计算需求，因而对算力规模与算力结构都具备高要求。2020年，我国算力规模已达到135EFlops，相比去年增长55%；智能算力比重增加至41%，相比2016年增加了38%。

高性能图计算为知识图谱运算赋能



中国算力规模及智能算力增长情况



注释：胖节点，即通过大内存、多网络进行单一结点计算的服务器。
来源：艾瑞咨询研究院结合公开资料自主研究绘制。

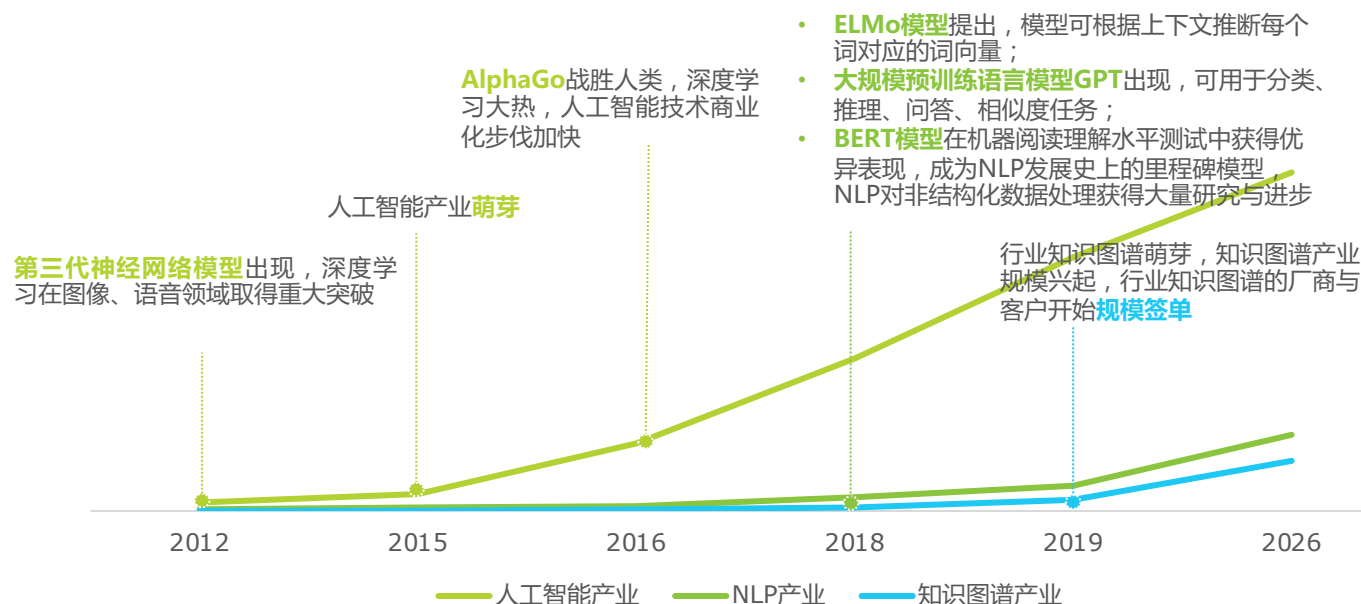
来源：中国信通院《2021年中国算力发展指数白皮书》，艾瑞咨询研究院自主研究绘制。

驱动力四：算法协同并进

算法进步的关联连锁反应

知识图谱的迅速发展与人工智能、NLP近几年取得的巨大进步有很大关系。从发展的关键时间点来看，人工智能在2012年取得了深度学习第三代神经网络的技术进步，此后人工智能技术在我国便开始了技术商业化之路，在2015年人工智能产业正式迈入萌芽阶段。作为深度学习技术进步的受益者，NLP技术随后频繁出现在人工智能的舞台，并携带较多惊喜的成就。ELMo模型、GPT模型、BERT模型三大NLP模型均在2018年出现，标志着NLP对文字语义的处理进入了新的历史台阶。NLP技术与知识图谱发展极为紧密，是知识图谱搭建的前置技术环节，常在知识图谱生产流程中用于自然语言信息抽取，对各类词性进行识别标注。2018年后NLP的泛化能力与自动化能力的增强，降低了NLP模型的训练成本，为知识图谱的知识库构建创造了有利机会。2019年前后，国内不少行业知识图谱厂商与客户开始规模签单，行业知识图谱正式萌芽。

三类AI技术与产业发展关键时间点

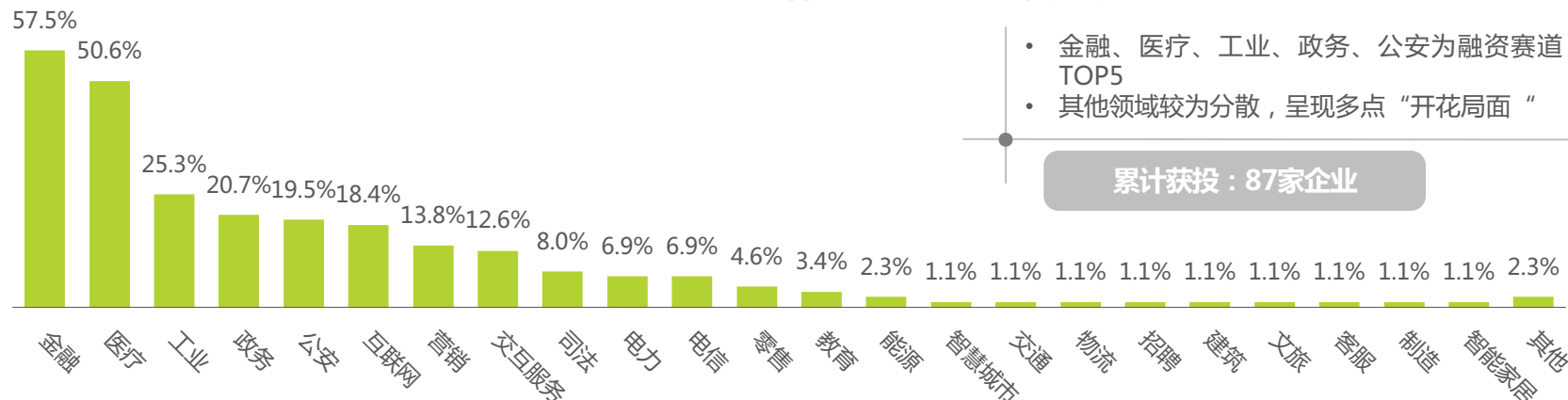


注释：产业发展曲线为定性判断。
来源：艾瑞咨询研究院自主研究绘制。

驱动力五：资本提供资金支持

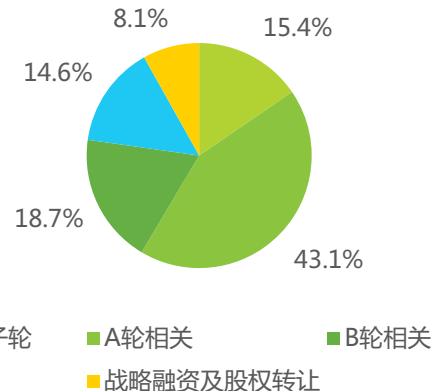
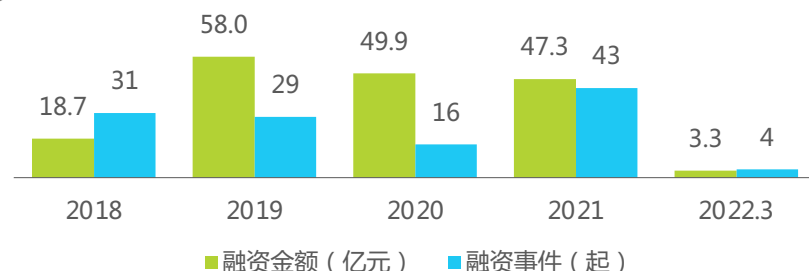
金融、医疗赛道融资热度领先，融资多处于早期

2018-2022年3月中国知识图谱获投企业业务赛道热度统计



2018-2022年3月中国知识图谱融资事件数、金额及轮次情况

- 2021年融资事件数最多，为43起；统计时间内总融资事件数为123起，融资金额为177.2亿元
- 融资集中在天使轮、种子轮、A轮的相关领域，占比高达58.5%，融资轮次靠前；海致、明略、同盾等企业已完成C轮及以后融资



来源：艾瑞咨询研究院根据公开融资数据自主研究整理及绘制。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

5

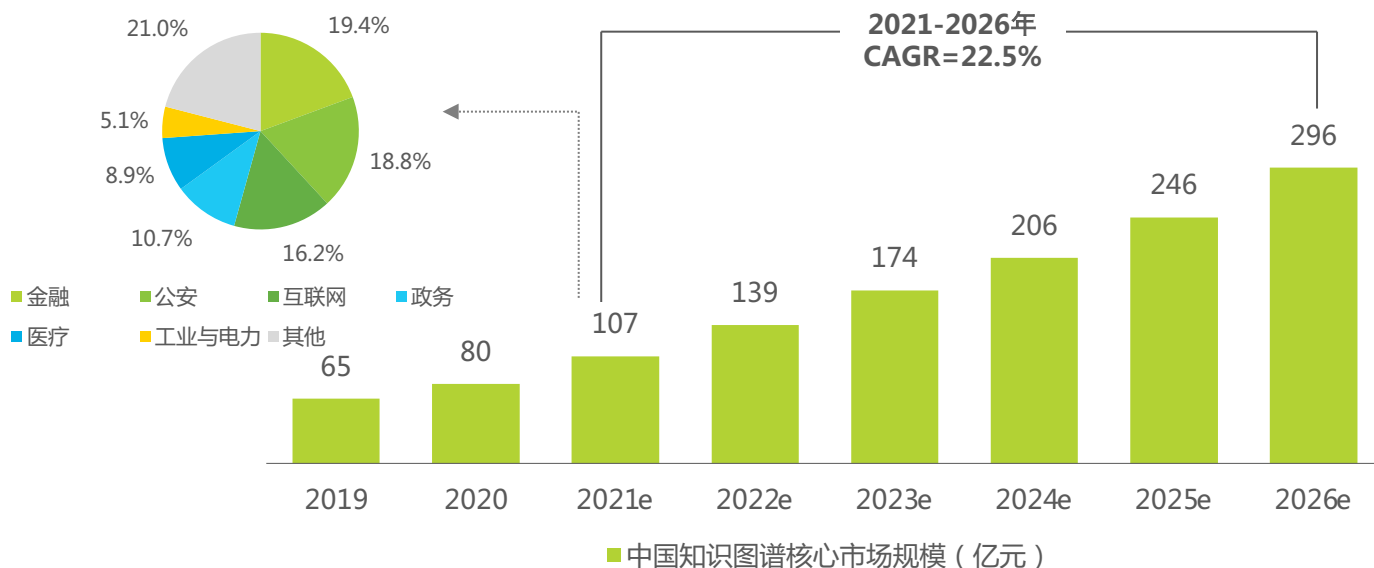
行业总结：市场空间结构与产业链

- **市场空间结构**：所测算规模为TAM。整体市场规模尚处于增量爬坡阶段，以金融、公安为主要行业拉力，行业知识图谱的市场规模潜力要大于互联网通用知识图谱。从供给侧看，行业集中度较低，市场份额分散，尚未形成稳固的市场竞争格局，但在某一行业领域存在市场份额偏高的玩家。整体市场规模方面，2021年中国知识图谱核心产品的市场规模为107亿元，到2026年，核心规模可突破290亿元，2021-2026年CAGR=22.5%。
- **产业链**：上游为硬件基础设施与数据采集及治理服务，中游为知识图谱表示与知识建模技术服务，下游为知识图谱应用。下游应用图谱中，金融知识图谱应用的参与玩家众多，其他行业领域玩家偏少，普遍存在同一玩家跨界耕耘的现象。

金融、公安为市场主要拉力

随着信息化与数字化建设的展开与NLP技术的进步，知识图谱不再局限于网络百科式的搜索，其衍生出了互联网内容与社交、大数据知识图谱与行业知识图谱等多种产品类型，产品专业化与场景化的趋势日渐明显，行业知识图谱已经成为市场开拓重点。金融与公安两大行业的知识图谱占比较高且增长速度较快，其业务与知识图谱可密切结合，具备建设意愿与资金投入，因而成为了市场规模的主要拉力。据艾瑞统计测算，在2021年的行业市场结构中，金融与公安的市场份额合计共占总市场的38.2%。未来，随着政务数字化建设的完善，政务对知识图谱的业务需求会逐渐唤醒，成为未来市场的拉力之一。从市场整体来看，2021年中国知识图谱核心市场规模为107亿元，到2026年，核心市场规模可突破290亿元，2021-2026年CAGR=22.5%。

2019-2026年中国知识图谱核心市场规模



注释：核心规模为软件规模，包含后文金融、政务与公安、医疗、工业与电力的产业规模，以及其他细分行业的产业规模。

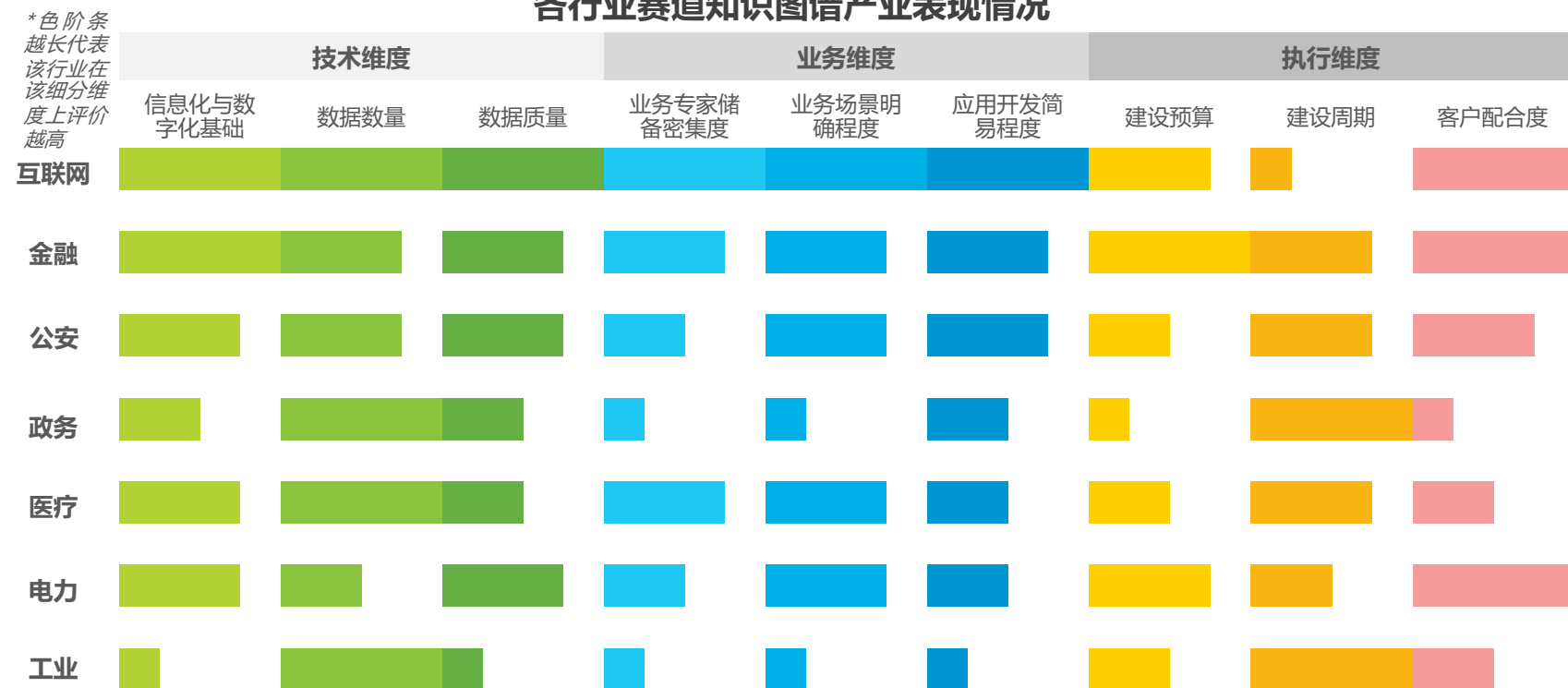
来源：艾瑞咨询研究院根据专家访谈、采招数据、公开媒体咨询、国家统计局、中国卫生统计年鉴等数据自主建模测算。

行业细分赛道表现总览

互联网最成熟，金融为兴起标杆，政务与工业长路漫漫

各行业赛道知识图谱产业情况可从技术、业务、执行三大维度进行评价和对比。互联网是知识图谱产业最成熟的赛道，一般提供轻量级的应用服务，开发难度较低。相比之下，其他行业的知识图谱产业仍处于建设期，以金融的表现作为标杆。政务、工业都有巨大的数据量，但业务专家储备密集度、业务场景明确程度、客户配合度的评价偏低，需要较为漫长的磨合过程。尽管医疗领域的专家储备与业务场景明确度表现较好，但因各级医院资源利益不统一，文本数据理解难度大，其产业发展远不如人们所希冀的高专家水准。

各行业赛道知识图谱产业表现情况

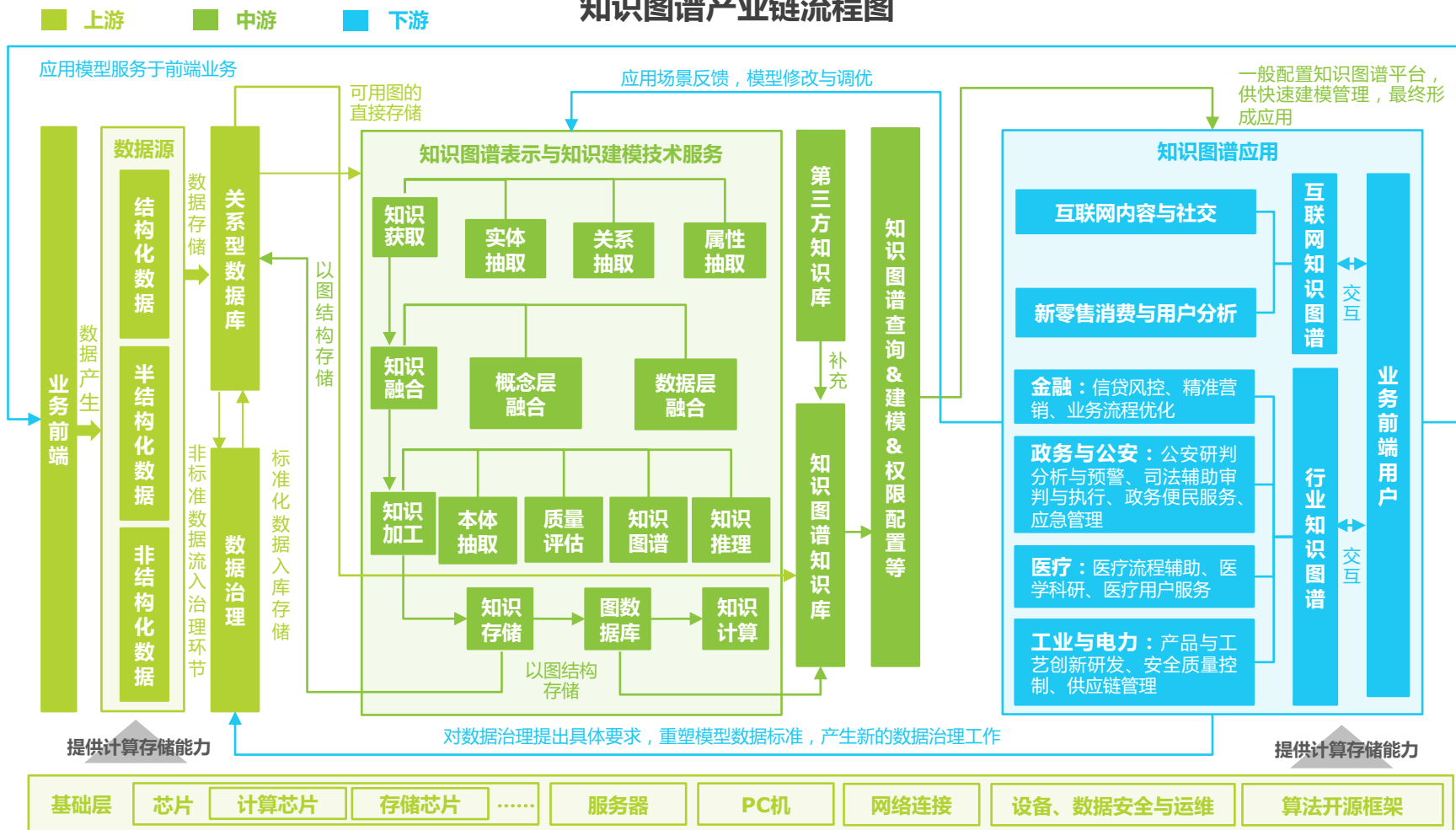


来源：艾瑞咨询研究院根据专家访谈自主研究绘制。

产业链流程

上游着重数据处理，中游搭建模型，应用服务前端形成闭环

知识图谱产业链流程图



来源：艾瑞咨询研究院自主研究绘制。

核心产业图谱

2022年中国知识图谱核心产业图谱



注释：Logo摆放以企业主营业务为主。图谱中所展示的公司logo顺序及大小并无实际意义，不涉及排名。从产品生产角度出发，挑选产业链中的关键模块进行展示。
来源：艾瑞咨询研究院自主研究绘制。

通用知识图谱：互联网篇

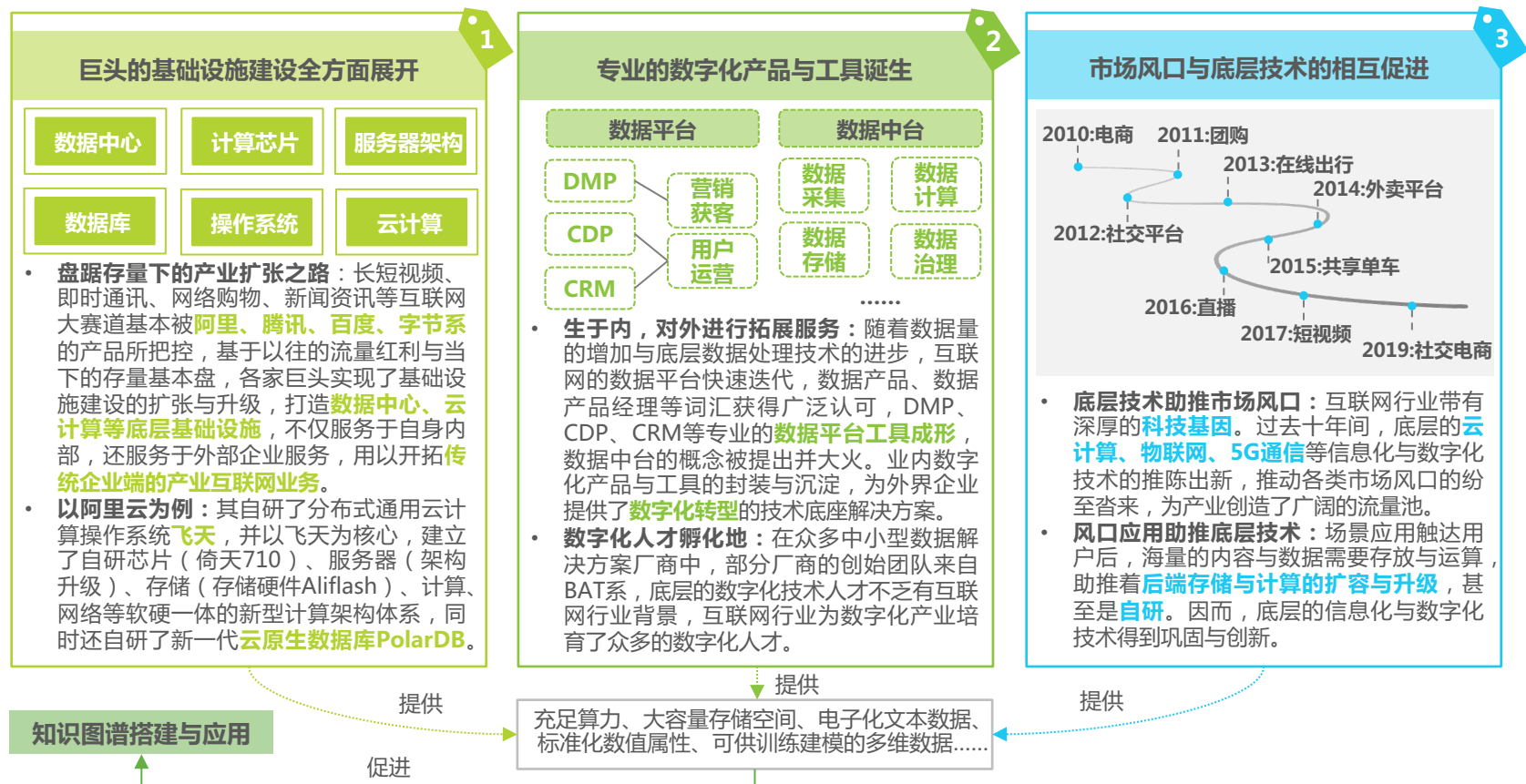
- 每一行业的知识图谱内容包括三个部分：行业信息化与数字化基础观察、场景应用分析、市场规模分析。
- **总体评价：**属于知识图谱应用较为领先的赛道，在底层建模与内容结构化方面具备增量机会，因上层应用开发一般由需求方自主开发，在上层的应用开发方面增量机会小。
- **行业信息化与数字化基础：**行业带有深厚的科技基因，总体信息化与数字化基础良好，大厂具备较为充足的投入预算与建设实力，中小企业主要依托现有基础夯实业务。
- **场景应用：**以智能推荐与辅助决策为核心，推出网络搜索、商品迭代开发、商品精准导购、内容个性化推荐等细分应用。
- **市场规模：**所测算规模为TAM，2021预计中国互联网知识图谱市场规模为17亿元，2026年可达51亿元，2021-2026年CAGR=24.2%。

信息化与数字化基础

基础扎实领先，为知识图谱创造有利基础环境

从巨头的基础设施建设、业内数字化产品与工具诞生、市场风口与底层技术的促进关系三大角度，可窥见互联网行业扎实领先的信息化与数字化基础，这为互联网知识图谱的构建与应用创造了有利的基本环境。

中国互联网信息化与数字化情况

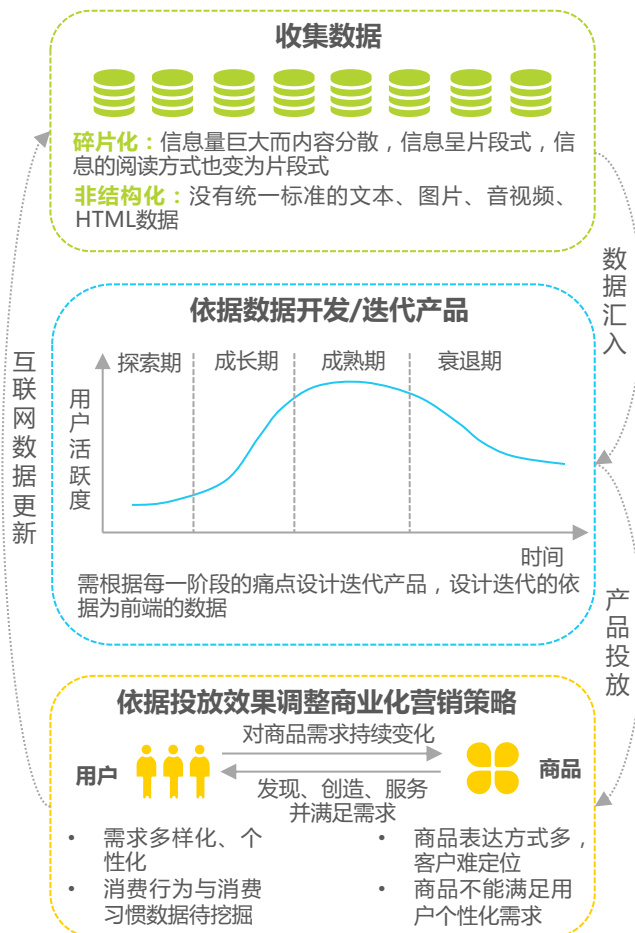


来源：艾瑞咨询《2021年互联网行业挑战与机遇白皮书》，艾瑞咨询研究院根据公开资料自主研究绘制。

行业需求痛点分析

信息碎片化与非结构化突出，产品需持续迭代满足用户需求

互联网行业需求痛点



一、互联网信息碎片化与非结构化的处理需求：

- 互联网的海量信息带有碎片化与非结构化特征。新兴互联网应用的蓬勃发展，让完整信息被分类分解为信息片段，信息被大量简化，从而导致信息本身不全面、内在逻辑不完整。同时，文本、图片、各类报表和音频、视频、HTML等非结构化数据广泛存在于互联网中。互联网企业需要在现有的存量业务中，收集碎片化信息，处理非结构化数据，**挖掘信息间的潜在联系，还原完整的、有价值的信息。**然而，信息的关联关系挖掘是一项工作量巨大且复杂的工程。

二、互联网产品迭代与发展的需求：

- 互联网产品的生命周期路径规划需求**：任何的互联网产品都会经历探索期、成长期、成熟期、衰退期四大阶段，每一阶段各有其痛点。探索期的产品在设计规范与方向上缺乏可参考的产品，难以找到产品的设计方向；成长期的产品往往同质化严重，难以找到产品的创新点，并在竞品竞争中取胜；成熟期的产品基本定型，需要在产品细节上做设计与把控。针对每一阶段的痛点，都需要做好路径规划，持续进行产品迭代，解决产品所处阶段的核心问题。产品迭代设计规划需要依赖合理可靠的决策依据，而决策依据源自将已经结构化的商品表现数据、行业信息等构建网络，发现数据之间的潜在关联。
- 以存量产品迭代为例进行解释**：某APP起初以本地生活服务为主打功能，但该APP已经较为成熟，同时期还存在一些竞争对手。产品设计师考虑在现有功能上叠加旅游服务与电商服务，以增加产品的差异化程度。

三、商品与用户信息的精准化、个性化表达需求：

- 商品方面**：市场管理差异导致的商品信息差，线上商品存在多种表达方式，标准化程度不足；线上商品缺乏个性化设计与表达，对用户引导不到位，难以激发用户兴趣，不利于用户做出消费决策。
- 用户方面**：网页端与移动端都会留下用户的行为数据痕迹，用户的消费观念隐藏在这些数据以及数据间的关联关系中。若要为用户精准推荐个性化内容，将用户切实转化为消费者，需要对用户的数据进行挖掘，刻画用户画像。

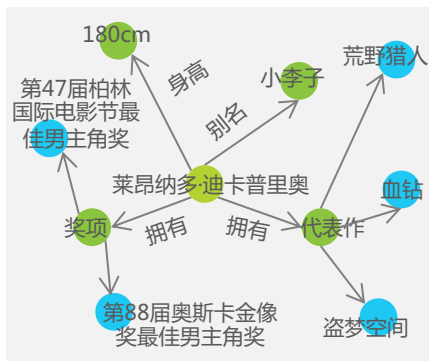
注释：信息是经过收集与整理的数据，为数据的子集，本质还是数据。
来源：艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

痛点下的知识图谱应用

以智能推荐与辅助决策为核心，满足用户与企业决策者需求

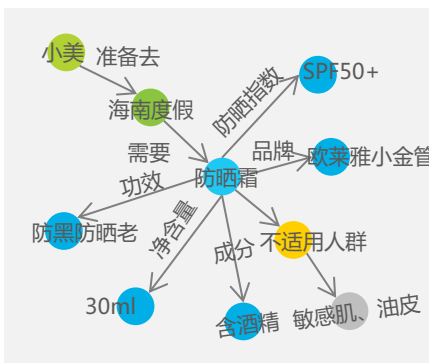
互联网知识图谱应用举例

一、碎片化信息聚合，以网络搜索场景为例



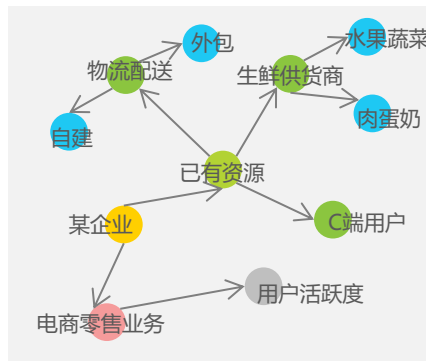
- **预期目标：**推荐与检索词条相关的实体，将碎片化信息聚合为不同维度呈现。
- **核心原理：**通过网络信息与数据搭建通用知识库，将具有相关性的信息联系起来，将完整丰富的信息推荐给用户。
- **核心价值点：**盘活长尾冷门的实体，实现碎片化信息聚合，还原完整信息，补充用户未知知识（尤其是时效性与新奇知识），激发用户搜索兴趣。
- **对应应用：**百度百科、搜狗百科、360百科、微信搜索等。

三、商品精准导购，以新零售网络购物为例



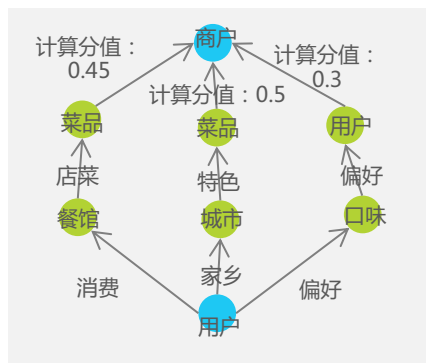
- **预期目标：**通过对客户多维度信息的挖掘，快速筛选出潜在商品需求。
- **核心原理：**基于用户信息库与商品信息库，构建商品决策知识图谱，在导购的具体场景中推送符合用户偏好的商品。
- **核心价值点：**对于用户而言，快速触达所需的商品种类，提高决策效率；对于商户而言，辅助商品导购，实现商品的个性化推荐与精准营销，增加商品收益。
- **对应应用：**淘宝、京东、网易严选、唯品会等。

二、存量产品迭代开发，以开拓新业务市场为例



- **预期目标：**提示决策人员业务的生命周期阶段，辅助其进行业务规划。
- **核心原理：**将存量业务数据反馈构建知识图谱网络，判断业务的生命周期阶段，决定是否要迭代业务。
- **示例：**某公司已有的存量业务为电商零售，基于信息后台的知识图谱网络反馈，业务的用户活跃度基本稳定，旧业务已进入成熟期，公司可以寻找存量业务的突破口。图谱提示决策人员公司已经具有物流配送、生鲜供货商、C端用户的资源优势，可考虑开拓生鲜市场。
- **对应应用：**盒马鲜生、七鲜超市等。

四、内容个性化推荐，以餐饮消费推荐为例



- **预期目标：**寻找用户最可能消费的商户，为用户推荐该商户。
- **核心原理：**将用户的消费记录、家乡、历史口味偏好等信息构建成知识图谱网络，计算出不同维度下的推荐分值，分值最高的路径即为推送结果。
- **核心价值点：**对于用户而言，能够为用户提供个性化、多维度的内容推荐，激起用户兴趣，减少对商户评价内容的依赖，缩短用户的筛选时间；对于商户而言，能够帮助商户吸引客流，增加业务收入。
- **对应应用：**美团、大众点评等。



以上案例均为通用知识图谱范围，在互联网的垂类知识图谱类型体系下，如专业的汽车网站、美妆网站、健康养生知识网站等，知识图谱的搭建原理类似，在此不再赘述。

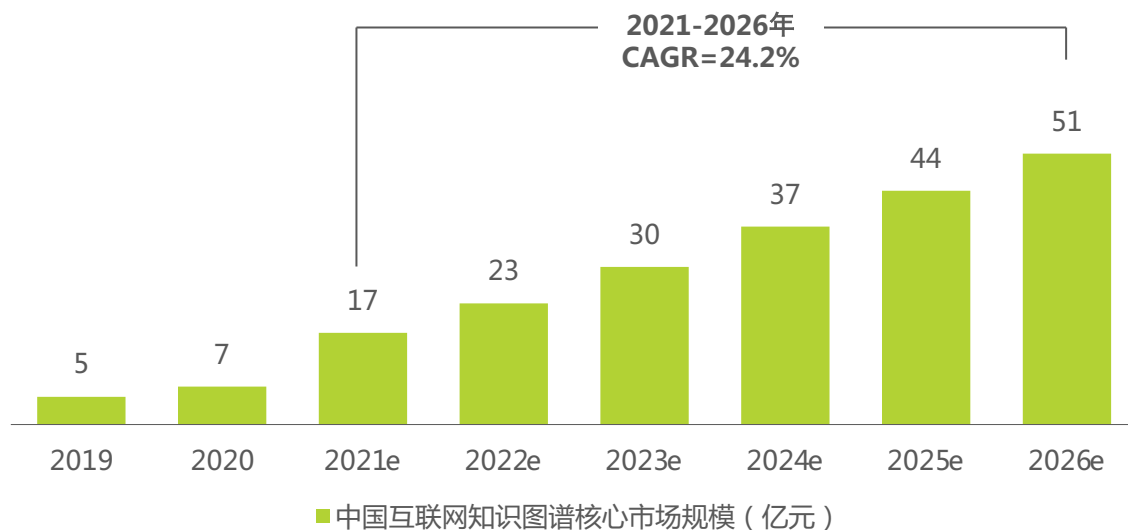
注释：不同颜色的圆代表不同层次的实体。所展示的四类知识图谱均为列举，知识图谱可依据开发需求调整结构。
来源：艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

市场规模分析

中小型互联网企业对底层能力的采购成为市场主要拉力

一般而言，中小型互联网企业为互联网知识图谱的核心需求方，成熟的互联网大厂、部分专业的大数据企业、知识图谱企业为互联网知识图谱的能力供应者。互联网大厂拥有成熟领先的全栈式知识图谱能力，其能力及应用主要服务于自身业务，基本没有知识图谱服务外采需求，故其并非市场的核心需求方。而中小型互联网企业因业务内容需要持续更新但底层开发能力不足，会将知识图谱建模与内容结构化的工作交给外部公司完成，并向这些公司支付授权费与开发费，成为了核心需求方。授权费为外部公司内部数据平台的授权使用费，开发费为业务模型建模费，两类费用为市场规模的重要内容。在上层应用开发方面，主要由需求方内部结合业务内容自主完成。随着中小型企业数量的增加与互联网产品的精耕细作，知识图谱的底层建模与内容结构化服务需求会逐渐释放，拉动市场规模增长。据艾瑞统计测算，2021年中国互联网知识图谱核心市场规模为17亿元，2026年相应规模可达51亿元，2021-2026年CAGR=24.2%。

2019-2026年中国互联网知识图谱核心市场规模



注释：核心规模为软件规模，且不包含企业内部消化的部分。内部消化指企业完全自主独立开发知识图谱并服务于其自身业务，不对外出售服务能力与产品。
来源：艾瑞咨询研究院根据专家访谈、公开媒体咨询等数据自主建模测算。

行业知识图谱：金融篇

- 每一行业的知识图谱内容包括三个部分：行业信息化与数字化基础观察、场景应用分析、市场规模分析。
- **总体评价**：属于知识图谱应用较为成熟领先的赛道，在需求主体分类标准下的主要市场为大中型银行市场，产业增量变的空间与可能性较高。
- **行业信息化与数字化基础**：金融总体的信息化与数字化基础良好，有持续的预算投入建设，科技应用开发意愿强烈。
- **场景应用**：以信贷风控、精准营销、业务流程优化为核心场景，信贷风控应用为重点高频应用。
- **市场规模**：所测算规模为TAM，2021预计中国金融知识图谱市场规模为21亿元，2026年可达68亿元，2021-2026年CAGR=26.6%。

信息化与数字化基础

资金支持下的信息化与数字化率先嬗变造就应用开发条件

知识图谱应用的搭建是建立在信息化与数字化建设基础之上的，金融行业领先的信息化与数字化建设为其后续展开知识图谱应用的开发创造了有利的先决条件。无论是从技术投入看，还是从部分前沿科技的采购情况看，金融业整体的资金投入与采购都呈现持续增长趋势，为行业变革与创新提供了充足的资金支持。基于资金的持续投入，金融业的信息化顺利迈过集中建设阶段，从而快速投身于数字化转型浪潮中，不断进行结构化数据储备，开发各类场景应用。

中国金融业信息化与科技建设现状

建设进展

1 信息化建设迈过集中建设阶段，头部银行进行核心系统升级改造

- 信息化**早期**建设先后经历过**IT电子化与互联网金融**，下级分支行的服务器集中由总行统一管理，而在**后期**，进行由底层分布式计算框架至核心系统的整体升级改造，主要建设主体为**头部银行**。

2 数字化建设如火如荼，银行引领行业数字化转型

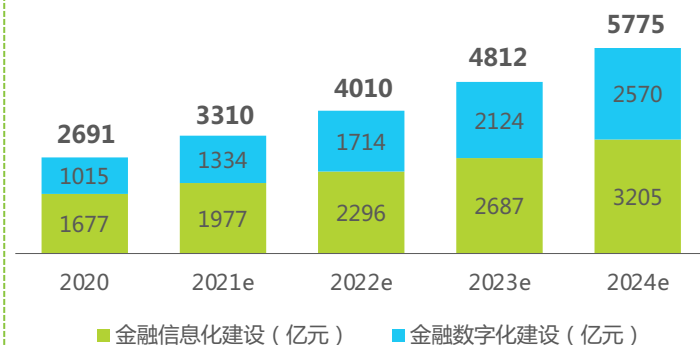
- 数仓、数据湖、数据集市、数据中台等建设在银行业大面积展开，很多银行专门成立了一级**数据部门**以引领自身整个银行系统的数字化转型，承担**基础数据资产的管理**。

3 深入应用场景进行精耕细作，突破业务卡点

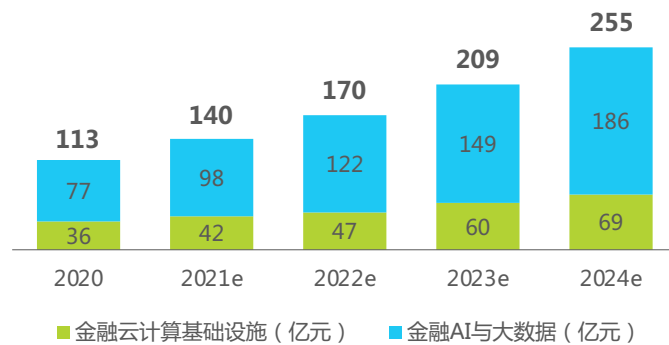
- 基于信息化与数字化建设，使用知识图谱等人工智能技术，针对**信贷风控、精准营销、供应链金融、保险理赔**等场景做精细化应用开发，改善各类业务运营困境。

预算情况

1 2020-2024年中国金融业技术投入情况



2 2020-2024年中国金融业部分前沿科技采购情况



注释：技术投入统计数据含银行、保险、证券三类市场主体的信息化与数字化建设数据；前沿科技采购费用包含银行、保险机构的采购数据。所抽取数据足够代表金融业情况。
来源：《2021年中国金融科技（FinTech）行业发展洞察报告》，银保监会，艾瑞咨询数据统计模型；艾瑞咨询根据专家访谈自主研究绘制。

场景应用分析一：信贷风控业务痛点

信用与欺诈风险加剧，风控手段急需优化升级

信贷，是一种仅以个人信用作为放贷依据的金融业务，具备风险性。传统信贷面临人力依赖性强、数据碎片化、对借款人信息分析不到位等业务痛点，解决方案正是数字化建设。随着金融与科技的融合发展，金融市场正在加速开放，信贷需求也在不断延展，整个金融信贷产业正逐步向无接触信贷转变。基于无接触信贷所展现出的大流量、信息不对称、高可触达特点，信贷业务痛点在原有痛点基础上有所增加：下沉客户质量泥沙俱下，加大了金融业的信贷风险；团伙作案犯罪手段日渐专业化、隐蔽化，欺诈风险提升。引入知识图谱可提升金融风控防范手段，针对痛点2、痛点4、痛点5“对症下药”。

国内信贷业务痛点转变示意图

Before 传统信贷业务痛点

1、依赖风控人员经验与人工审批，放款效率低、管理成本高

贷前、贷中、贷后的工作更依赖风控人员经验，需雇佣较多员工，人力成本高；审批方式以人工审批为主，一般需要2-3周以上方能放款，审批时间长、效率低下、手续与流程繁琐。

2、风控数据海量碎片化，尽调成本高、考证难度大，风险识别监测手段弱

在对公业务中，部分项目信息数据呈海量碎片化，尽调成本高、考证难度大，这一问题对于创新型小微企业更为严重，效果有限的风险识别监测手段不利于评估企业客户及其风险事件。

3、依赖历史财务信息进行静态分析，不利于把握借款人真实的经营状况

评估企业信用情况时，银行更多依赖企业历史财务信息进行静态分析，依据较为单一的评价作出信贷决策，难以完全把握借款人真实的经营状况，存在评估局限。

Now 信贷业务痛点增加

4、金融客户下沉，客户信用能力参差不齐，信用风险加剧

伴随着无接触信贷兴起，线上金融业务日益增多，助推金融服务下沉，客户流量入口多元化，随之而来的是金融客户下沉。部分客户是通过第三方平台推荐引入的流量，这些客户的信用能力参差不齐，信用风险加剧。

5、团伙作案日渐组织化、专业化、隐蔽化，欺诈风险提升

保险、贷款和信用卡申请是一些极容易出现信贷欺诈的领域。信贷团伙为实现骗贷，会有组织、有谋划地开展欺诈，其经手的单笔欺诈金额在几千至上万之间。基于信贷团伙犯罪活动呈现出组织化、专业化、隐蔽化的特点，违反犯罪分子具有相关金融行业的从业经历，犯罪手法相较于过去有了较大程度的提高，而金融一线基层人员对信贷团伙作案的识别与防范存在短板，这为金融业的正常稳定运行埋下了巨大隐患。

信用风险加剧案例



团伙欺诈案例



注释：我国金融市场以银行业为主，故此处挑选银行信贷风控场景为切入点，保险业欺诈骗保场景在此不作讨论。
来源：中国银保监会，艾瑞咨询研究院根据专家访谈、公开资料自主研究绘制。

场景应用分析一：信贷风控原理及应用

iResearch 艾瑞咨询

基于数据信息搭建网络，从网络中识别异常风险

第一章已提到，知识图谱是复杂的关系语义网络，可以从关系角度提供分析问题的能力，这有利于从正常的特征与行为中挖掘出异常的信用风险与团伙欺诈行为，提升金融机构的风控能力。

知识图谱在信贷风控场景中的核心原理及搭建过程



针对信用风险：信用能力评估图谱

1、核心原理：信用能力可依据信息推断预测

- **对私业务**：个人身份、社交关系网络、消费记录、资产负债、年龄等信息可较为客观反映出实际消费能力。
- **对公业务**：企业间关系，以及司法税务、信贷结算、财务与舆情等业务均可被记录与汇总，这些已记录的信息可透露出借款逾期、负面舆论影响、财务资金漏洞等风险，而风险间具有传导关系，概率大小不一，原因可追溯。

2、搭建过程：

• 对私业务



Step1：挑选信息节点

Step2：依据重要节点搭建结构

Step3：依据关系网判断信用能力
图中的甲已经持有信用卡三年，在代理记账公司工作，税前月薪为9000元，年龄为27岁，身边的社交人群较为固定单一，信用记录良好，可考虑放贷

• 对公业务



Step1：描述业务、搭建网络

Step2：输入输出描述

Step3：风险传导识别与预测
图中的A公司偿还招商银行的资金主要来自E公司的还款，E公司因财务困难或违约可能无法偿还A公司债务，从而诱发A公司对招商银行的债务逾期风险



针对欺诈风险：知识图谱助力团伙关系挖掘，实现异常群簇的敏捷识别

1、核心原理：团伙作案带有群体与共性特征，与正常个体存在明显差异

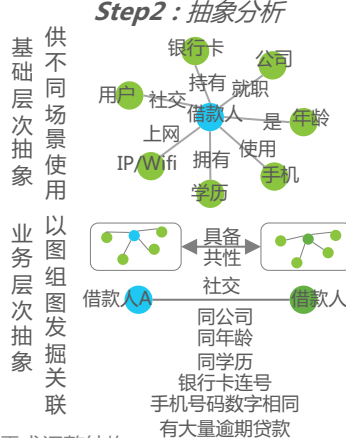
- 团伙在信贷诈骗的作案过程中一定会留下蛛丝马迹。犯罪分子为了节约犯罪成本，会倾向于使用同一部手机切换不同的账号，购买的手机一般为成本偏低的安卓手机，手机中的黑产软件使用共同的网段登录，团伙预留的电话号码更有可能为同一电话号码的多种排列组合。同时，犯罪分子的人员规模具备一定的数量特征，人员信息包装上会具备相似的职位身份，年龄也比较相仿，学历固定在某一水平，团队总负债偏高，男女比例有倾斜等。也就是说，犯罪团伙会呈现出针对某一信贷诈骗场景的群体与共性特征，这些特征正是知识图谱中的关系。
- 在关系网络中，大多数正常个体理应为相互独立的节点，或与另一个节点组成规模为二的团体（多数可能为家人或亲友关系），若出现三个点以上甚至十几个点关系密切时，这些团体可被归为异常。

2、搭建过程：降维——抽象分析——分类——检测

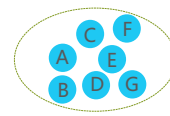


Step1：高维数据降维

- 降低无效、错误数据对建模的影响，提高建模的准确性；
- 少量且具有代表性的数据将大幅缩减挖掘时间；
- 降低存储数据的空间与成本。



Step2：抽象分析



Step3：群体分类

- 选取公司、最大年龄差、学历、网络IP、手机号码、负债等对团伙进行描述分类；
- 潜在风险集群中的个体就职于同一家公司，年龄差相差0-10岁，学历多为本科，使用同一IP上网，同时申请银行卡导致银行卡号码连号，手机号码数字相同，大量个体背有大量逾期贷款。

Step4：检测拦截



与事先设定的量化指标对比，检测相关指标是否异常。若属于异常情况，与公安合作确认为欺诈团伙后，拉黑团伙相关节点，黑节点入库黑名单，依据黑名单规则对关联申请实施拦截

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

来源：艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

场景应用分析二：精准营销业务痛点

产品同质化且创新设计力度不足；客户转化难，增量难拓展

银行与保险的营销业务痛点



场景一：商业银行营销

1

供应链金融从1+N模式向N+N模式转变

- 传统的供应链金融是“1+N”模式，银行围绕1家核心企业开展营销业务，不太看重小微企业的主体信用，而是看重这些企业与核心企业之间的关系，核心企业可提供信用与担保能力。
- 如今，消费互联网向产业互联网纵深演进，越来越多的核心企业在线上开展产业，在销售、供应、生产、经营全流程中聚合了诸多小微企业，供应链金融网络呈现出“N+N”的模式，而商业银行难以发掘出海量企业中的核心企业。

2

虚拟渠道的获客效能极少有实质性发挥

- 现阶段，商业银行在互联网各大平台中植入的虚拟渠道极少能够发挥其低成本获客与低差异化服务成本的优势，仅仅成为物理渠道的延伸。
- 商业银行更侧重已有的银行账户客户，开放度不足，营销力度与服务能力有待加强，未能推动平台用户的流量变现。如何在虚拟渠道搭建好的基础上，根据流量的信息形成精准的客户标签，描绘出客户画像，推荐客户需要的金融产品，是虚拟渠道营销的痛点。

3

产品与服务高度同质化，营销创新能力滞后

- 商业银行渠道承载的产品和服务由于设计理念趋同、流程、服务、收益方式类似，存在高度同质化现象；以粗放的“一揽子”方式在不同渠道间进行产品和服务的简单迁移、单纯叠加、贪多求全，造成资源低效占用和浪费。
- 商业银行的传统营销渠道以交易应答服务为主，其中涉及固定资产投资、市场开拓维护等多方因素，具有极强的延续性，长期形成的路径难以在短期内改变。



场景二：商业保险营销

1

传统的人海战术营销成本高、效率低

- 传统的保险企业在数字化转型前大多采用人海战术的保单销售策略，尽管该战术下的保费收入与代理人的数量呈现明显的正相关关系，但也带来了高人力成本，如代理人培训费、佣金等。
- 某些保险代理人的素质低下导致营销环节低，其更多从自身利益出发推销产品，对客户的需求关注不够，导致其潜在客户的判断不准确、客户转化不够精细化。

2

产品同质化，精算作用未能较好体现

- 当下保险市场中的产品创新较少，客户的细分需求难以被满足，如寿险、重疾险等保险企业的产品在条款设计、病种以及赔付方式等方面同质化严重，缺乏市场竞争力。
- 精算的作用在产品设计中未能得到较好体现，多数保险企业对非标体的承保过于谨慎，大量优秀有需求的客户难以找到匹配自身的产品。

3

理赔不精确降低客户消费体验感

- 理赔属于保险产品中的关键环节，理赔纠纷一直都是保险企业的痛点。不符合理赔条件的事故如果得到赔付，将影响保险企业的正常利润，有的理赔甚至是恶意的欺诈。
- 拒赔带来的纠纷会打击大众对保险公司的信心，进一步加深保险在大众中的负面印象，给保险企业生产经营带来不良影响。

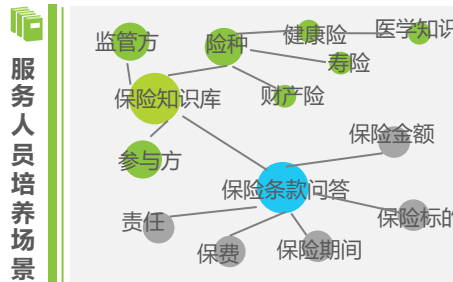


对策：引入知识图谱，搭建营销图结构网络，改善营销业务困局

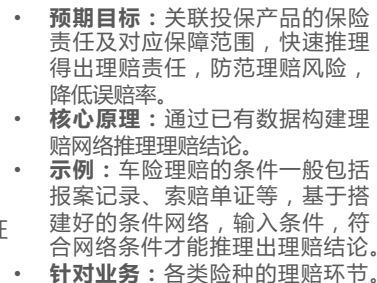
注释：“1+N”的供应链金融模式，1指1家核心企业，N指核心企业上下游的多家小微企业。“N+N”的供应链金融模式则有多家核心企业，每家核心企业背后有多家小微企业。
来源：艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

知识图谱在银行与保险营销业务场景的应用原理

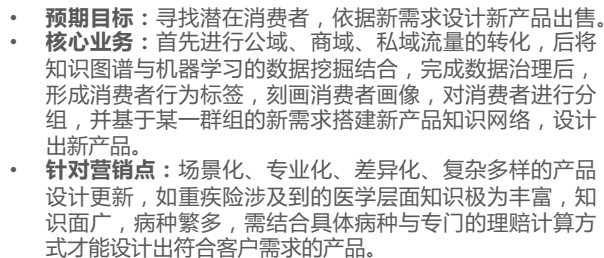
二、保險場景



- **核心原理：**基于已有的百科、学术文献、专业书籍、保险合同条款建立知识库与知识问答网络，形成庞大的知识库，在搜索过程中直接匹配相关答案，便于业务人员快速检索学习，掌握相关知识体系。
- **针对业务：**业务人员素质与运营技能培养。



三、 銀行&保險共同場景



©2022.8 iResearch Inc.

场景应用分析三：流程优化业务痛点

运营难度大、工作效率低、人力成本高、知识体系难构建

除信贷风控与精准营销外，流程优化是金融业的另一高频场景，主要面临来自传统客服与员工培训细分场景所带来的业务痛点。传统客服场景下，银行业在人工客服业务中常面临机械性的高重复咨询服务、高峰时期的响应服务不到位等痛点；保险业在投保到索赔业务中需应对各类用户诉求，客服人员的服务水平不到位影响着客户满意度，以及客户的留存及转化；证券业的痛点在于业务人员长期面临重复性的基础业务咨询。员工培训场景下，银行业因行业监管属性，内外部规定较多，仅依靠数量有限的集体培训难以在短期内培养出熟知规定的人才，员工提出的问题也无法被分类梳理、供日后知识共享；保险业的员工素质参差不齐依然是常态化的问题。两大场景的业务痛点所导致的结果皆为业务流程管理方面的问题，例如业务运营管理难度大，员工工作效率低，人力成本高等，因而引入知识图谱，搭建内外部金融知识库的对策被提出。

金融流程优化场景痛点解析

场景归纳：内外部咨询两大场景

场景一，外部客户咨询：传统客服场景



场景二，内部员工咨询：员工培训场景



场景一 业务痛点：

- 银行业：**主要在于人工客服业务，坐席工作煎熬、机械性的高重复咨询服务、高峰时期无法及时响应且排队时间长、问题数量多与难导致的人员处理压力大、催收可能产生情绪化的争执、无法全天响应客户可能导致客源流失。
- 保险业：**投保到索赔业务的用户诉求千人千面，回答不到位将影响客户的留存与转化率；客户对保险服务进行投诉与刁难导致服务人员负面情绪积压。
- 证券业：**证券类基础问题的咨询重复性高。

场景二 业务痛点：

- 银行业：**重监管属性行业，有诸多外部规定，如人民银行或银保监颁发的条例、管理办法；针对外部规定有相应的内部实施细则、操作规范。内外部知识依赖人工梳理更新，员工遇到业务问题时常咨询老员工，但老员工无法及时响应，且银行内部的知识团队并不充足；共性知识点得不到沉淀，企业内部知识无法共享。
- 保险业：**员工素质参差不齐，人员流动性大；员工短视特征较强，不重视业务深造。

导致结果：业务流程管理问题

- 业务运营管理难度大，员工工作效率低；
- 人员堆积，人力成本高，培训成本高；
- 企业内外部知识点分散，难以构建自有知识库，实现知识的系统化管理；
- 员工价值感与幸福感低，职业稳定性差。

对策选择：引入知识图谱

- 外部知识库：**基于知识图谱技术构建智能客服外部知识库，实现金融客服数据词库积累，辅助人力，在重复性的基础咨询问题上释放人力。
- 内部知识库：**在企业内构建全面而专业的知识框架与体系，结合智能检索问答，用机器推动业务咨询与辅导、员工培训、事后监督检查工作等。

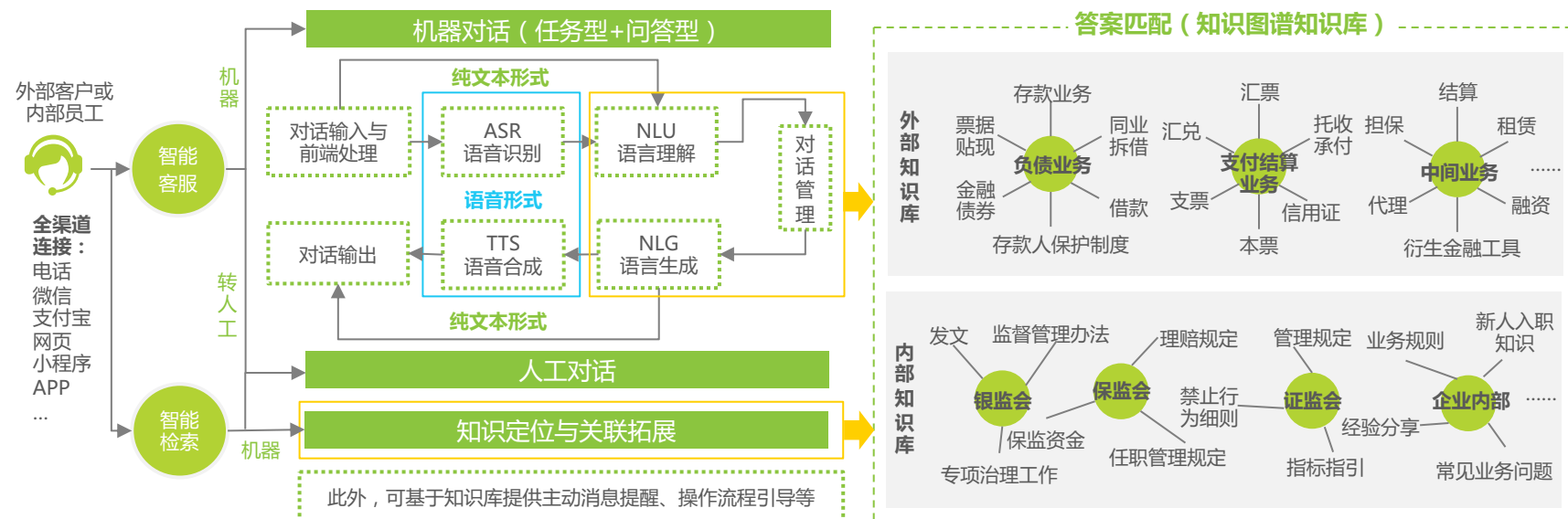
来源：艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

场景应用分析三：流程优化原理及应用

搭建知识网络形式的内外部知识库，快速定位知识点

为快速响应咨询者的问题，提升问答效率，改善金融知识库体系，厂商帮助金融机构在后台搭建好专门答复外部客户咨询业务问题的外部知识库与服务于金融体系内业务培训的内部知识库，将知识图谱知识库嵌入智能客服或智能检索类产品中，与NLP、RPA等技术结合，使得智能客服与智能检索应用能够自动理解咨询者的业务问题，迅速定位知识库中的知识点并输出。若咨询问题超出知识库范围，机器可自动或手动转人工。一般而言，外部知识库将负债业务、支付结算业务、中间业务等银行业务，保单保费等保险业务，以及证券基础知识等证券业务的知识点、客户信息拆解融合为知识图谱网络结构，而内部知识库则将银监会、保监会、证监会等行业监管方的政策规定，以及企业内部的常见业务问题、业务规则等作为知识图谱网络的组成内容，内外知识库以知识网络的形式共同协助咨询业务，使金融业的营销问答、知识构建拓展、工作流程与客户管理等得到优化。

金融知识图谱知识库协助智能客服与智能检索



注释：知识库包含FAQ、文档知识库与知识图谱知识库，此处仅展示知识图谱知识库。

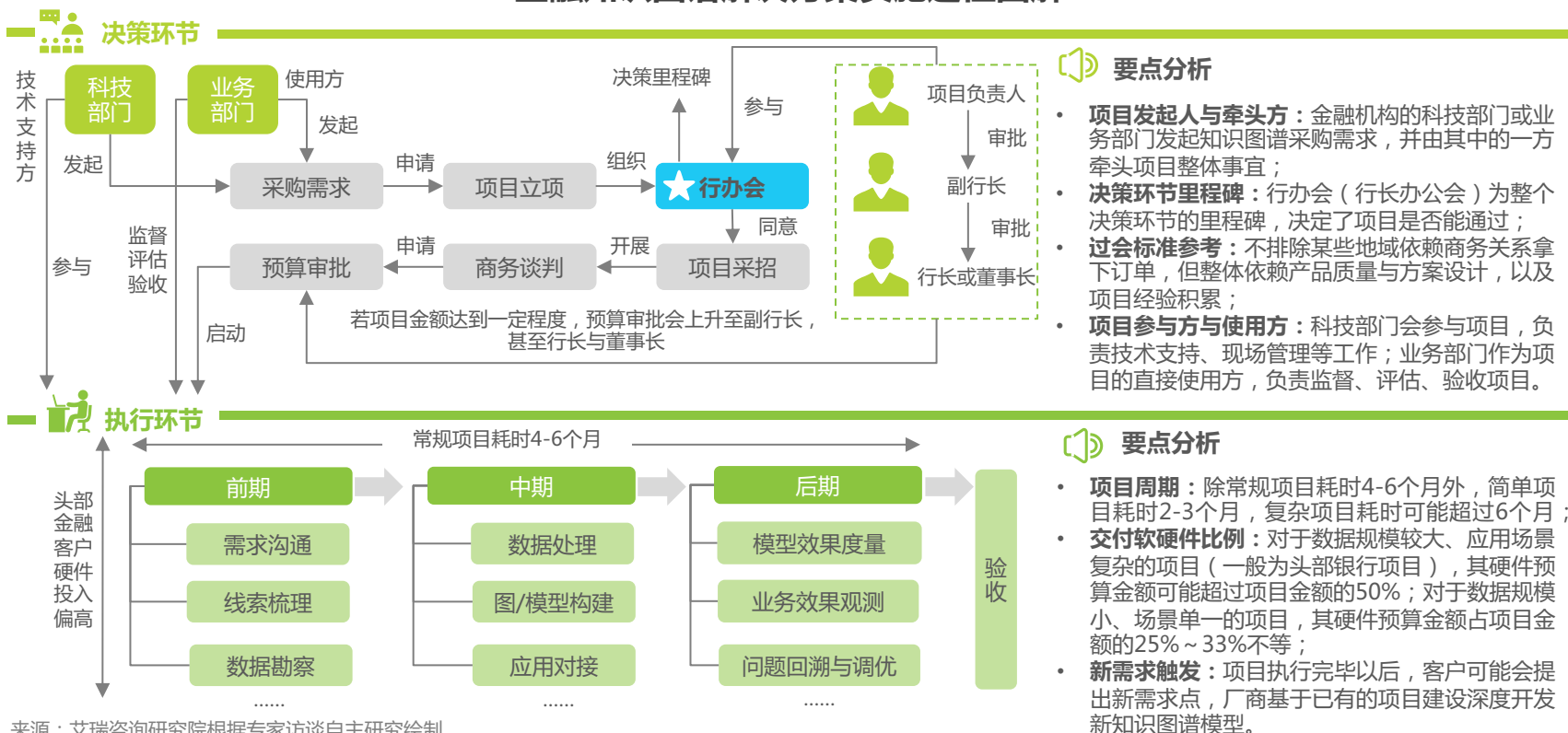
来源：艾瑞咨询《2021年中国对话机器人chatbot行业发展研究报告》，艾瑞咨询研究院根据公开资料、专家访谈自主研究绘制。

解决方案实施路径

决策环节以行办会为里程碑，执行环节类似常规软件项目

金融知识图谱解决方案主要包括决策与执行两大环节。在决策环节，由科技部门或业务部门中的一方发起项目采购需求，之后由其中的一方作为项目牵头方、另一方作为协助方，双方共同申请项目立项。立项后，决策流程先后经过行办会、项目采招、商务谈判、预算审批，项目正式启动后则进入执行环节。整个执行环节分为前期、中期、后期，不同时间段各有不同的执行内容，项目符合业务部门的运行标准后，方可达到验收状态。

金融知识图谱解决方案实施途径图解



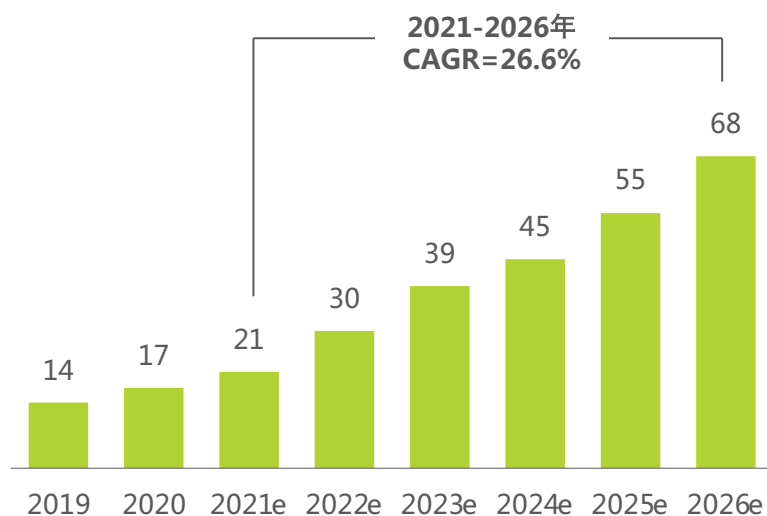
来源：艾瑞咨询研究院根据专家访谈自主研究绘制。

市场规模分析

市场整体维持较快增速，知识图谱解决方案与软件占主导

据艾瑞统计测算，2021年中国金融知识图谱核心市场规模为21亿元，2026年相应规模可达68亿元，2021-2026年CAGR=26.6%，市场整体增长较快。2019年起，知识图谱类的产品逐渐被市场所接受，产品的业务价值在金融领域率先得到认可（**尤其在银行得到采购与实施**），市场规模开始释放。随着知识图谱解决方案与大数据类产品的渗透与加深，金融市场对知识图谱产品的认可度提升而选择深化场景开发，预计在2022年市场规模将迎来新一轮增长浪潮。市场规模结构方面，从产品结构看，知识图谱解决方案类的产品在金融知识图谱市场占据主导地位，2021年其市场规模比例为84.0%；从软硬件结构看，软件规模为金融知识图谱市场的主要部分，2021年软件规模比例为58.9%。

2019-2026年中国金融知识图谱核心市场规模



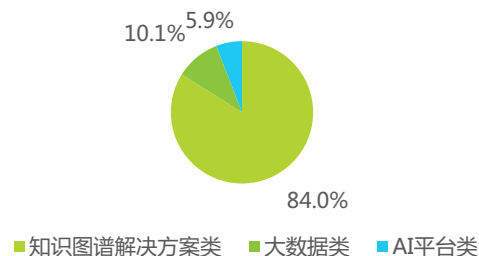
■ 中国金融知识图谱核心市场规模 (亿元)

注释：核心规模为软件规模，包含大数据产品、AI平台产品的知识图谱部分、金融行业知识图谱解决方案的核心软件部分。

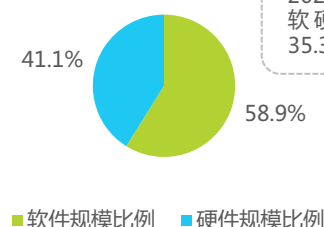
来源：艾瑞咨询研究院根据专家访谈、采招数据、银保监会披露信息自主建模测算。

2021年中国金融知识图谱市场规模结构

核心市场规模，按产品结构分类：



软硬一体总规模，按软硬件分类：



2021年中国知识图谱软硬一体化规模为35.3亿元

来源：艾瑞咨询研究院根据专家访谈自主建模测算并进行市场划分绘制。

行业知识图谱：政务与公安篇

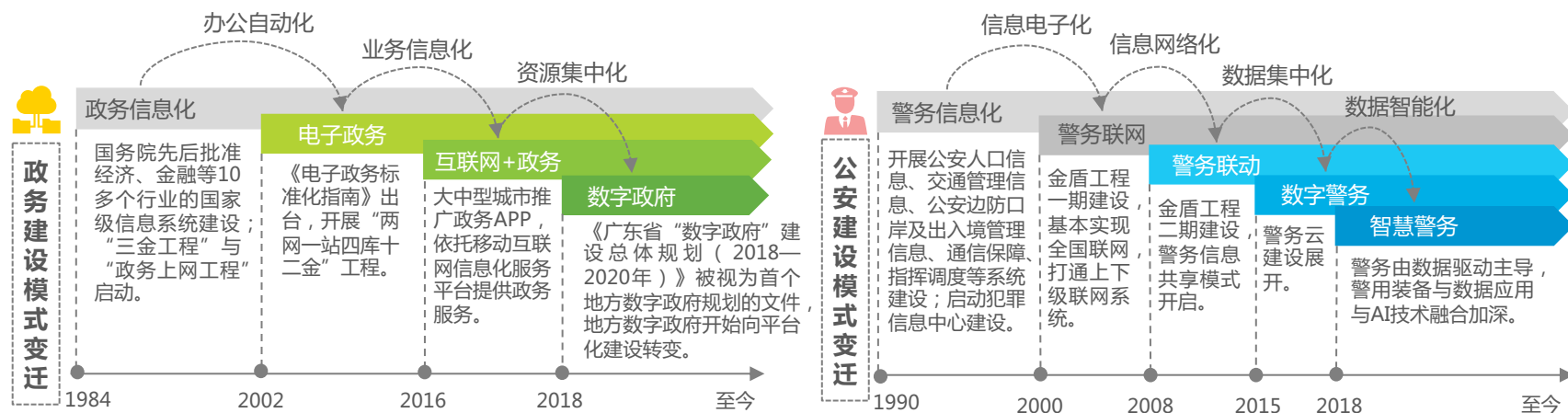
- 每一行业的知识图谱内容包括三个部分：行业信息化与数字化基础观察、场景应用分析、市场规模分析。
- **解释**：因公安的信息化与数字化建设、知识图谱业务独立性高，可脱离其他政务领域存在，故将公安与政务并列，合并为同一章节对比分析。
- **总体评价**：公安知识图谱应用要领先于政务知识图谱，二者潜在需求空间较大，但由于政府项目的属性，产业变现能力不够明朗。
- **行业信息化与数字化基础**：公安信息化与数字化建设总体领先于政务领域，率先进入智慧应用阶段。
- **场景应用**：以公安研判分析与预警、司法辅助审判与执行、政务服务便民、应急管理为核心场景，政务服务便民与应急管理的应用面有待拓展。
- **市场规模**：所测算规模为TAM，2021年中国政务与公安知识图谱的核心市场规模分别为11亿元（政务）与20亿元（公安），2026年相应规模分别为50亿元（政务）与43亿元（公安）。在2021-2026年间，政务的知识图谱复合年均增长率高于公安知识图谱。

信息化与数字化基础

建设模式均已迈向数字化，公安数字化建设领先于政务

整体上看，政务与公安均已实现信息的电子化录入与存储，并于后期开展政务云与警务云建设，争取实现数据上云，走向集中化管理模式。集中化管理模式有利于政务与公安内部开放数据权限、统一数据标准，服务于知识图谱的数据获取与标准化知识库构建。基于领先的信息化与数字化建设，公安目前已经步入智慧警务阶段，率先实现知识图谱的应用。政务涉及的领域众多，涵盖政治与行政、财政经济、科教文卫，其数据种类多且细分场景长尾化严重，数据质量较差，数据治理周期更长，其整体的数智化水平要低于数据与场景聚焦程度高的公安领域，知识图谱的基本库构建与应用开发也更为落后。

政务与公安建设模式变迁史



政务与公安横向对比

公安的建设模式迭代领先于政务，其已经步入**智慧警务阶段**，对知识图谱的应用也更为深入和广泛，线索关联与挖掘已经运用于刑侦、禁毒等多个警种。政务仍在向数字政府转型，其知识图谱还处于探索阶段，场景应用不深，知识图谱的延展性与通用性差。

注释：广义上讲，公安也属于政务领域，但因公安的信息化建设与发展独立性较强，其成为知识图谱最先应用成功的领域之一，所以将之独立分开论述。“两网”指政务内网和政务外网；“一站”指政府门户网站；“四库”指人口数据库、法人单位数据库、空间地理和自然资源数据库、宏观经济数据库；“十二金”指需重点推进的金税、金关、金财、金融监管、金申、金盾、金农、金水、金质、社会保障、办公业务资源系统、宏观经济管理系统的业务系统建设。

来源：艾瑞咨询《2020年政务云行业研究报告》，艾瑞咨询研究院根据公开资料自主研究绘制。

场景应用分析一：公安研判分析与预警

iResearch 艾瑞咨询

依据知识库快速搭建网络分析案情，克服内外部业务痛点

公安知识图谱业务情况

一、主要业务痛点

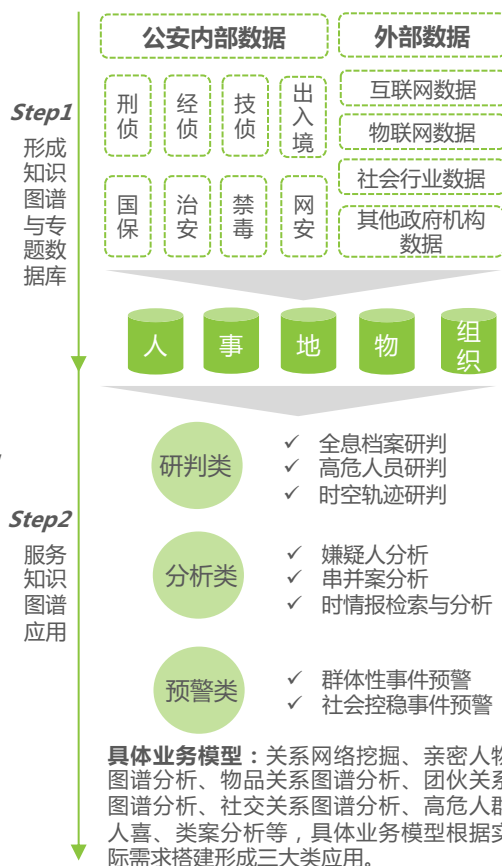
Part1 外部情况：

- 传统犯罪变化：**传统犯罪向动态化、组织化、集团化、专业化发展，并在加速“上网”，在现代化程度高的都市更是呈现出高科技犯罪与犯罪主体年轻化与智能化的特点。
- 新兴网络犯罪丛生：**以互联网为手段的新型网络犯罪不断凸显，呈高发多态势，危害日益严重。新型网络犯罪具有隐蔽性、迷惑性、衍变性，其借助互联网与移动互联等技术使大量犯罪由现实空间转入虚拟空间，以网络诈骗为主要类型。
- 社会人口动态化：**伴随着交通方式进步而来的是社会人口与物品的流动性加强，这为人口跨地域作案提供了便利，提升了社会治理难度。

Part2 外部情况：

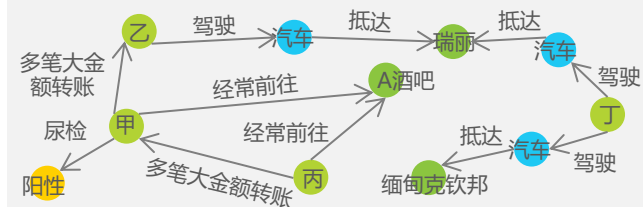
- 打击犯罪新手段亟待提升：**面对外界变化，公安在传统犯罪、新型网络犯罪中难以发现并挖掘人口与案件的关系，打击犯罪新手段急需结合公安大数据与知识图谱等新技术，增强作战实力。
- 编制有限但群众警情需求量大：**因政府编制的特性，公安人员数量有限，但要处理大量的侦查案件，急需使用新技术协助或解放人力。

二、搭建“五要素”知识库与应用



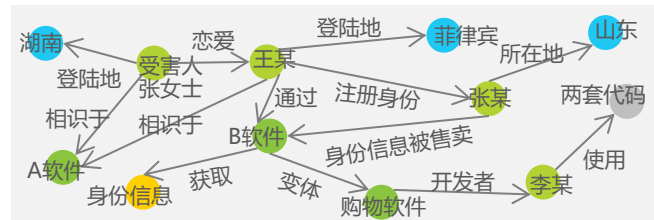
三、知识图谱应用实战举例

涉毒团伙网络挖掘



- 预期目标：**挖掘毒品交易与贩毒团伙网络，快速掌握贩毒证据，制定抓捕方案。
- 核心原理及效果：**通过已有专题数据库与知识图谱技术快速构建贩毒网络，及时发现动态化、多变的团伙关系。
- 示例：**尿检为阳性的甲经常前往吸毒人员经常出没的A酒吧，并收取丙的多笔大额转账，还向乙发起多笔大额转账，并与经常出没于边境的丁有联系，因此甲、乙、丁极有可能是涉毒团伙。
- 服务警种：**禁毒与缉毒警察。

网络诈骗网络挖掘



- 预期目标：**挖掘网络诈骗团伙，打击应对专业化的诈骗犯罪。
- 核心原理：**通过已有专题数据库与知识图谱技术快速构建诈骗关系网络，研判分析犯罪团伙的组织分工关系。
- 示例：**王某使用张女士的身份注册A软件与张女士假意恋爱并骗取其钱财，然而王某的登录地址为菲律宾，且其通过表面为购物软件的B软件购买张女士的身份信息；B软件为李某所开发，其使用两套代码让B软件能够在购物软件与身份售卖软件之间切换。
- 服务警种：**刑侦与经侦等。

注释：网络犯罪主要是利用网络、针对网络和在网络空间进行的各种犯罪。不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。
来源：艾瑞咨询研究院根据专家访谈及公开资料自主研究绘制。

场景应用分析二：司法辅助审判与执行 iResearch 艾瑞咨询

文书自动生成与类案处理为核心应用，解决人力不足的问题

司法知识图谱业务情况

一、主要业务痛点

- 业务核心问题：**政府编制不足且民众诉求过多。案件数量呈现出增长趋势，而政府编制无法快速增加岗位与人员数量，司法审核人员存在考编考证的准入门槛，导致主审案子的法官不足以应对巨量案件需求。
- 建设核心问题：**跨单位/部门数据未打通、业务与数据协调难等。

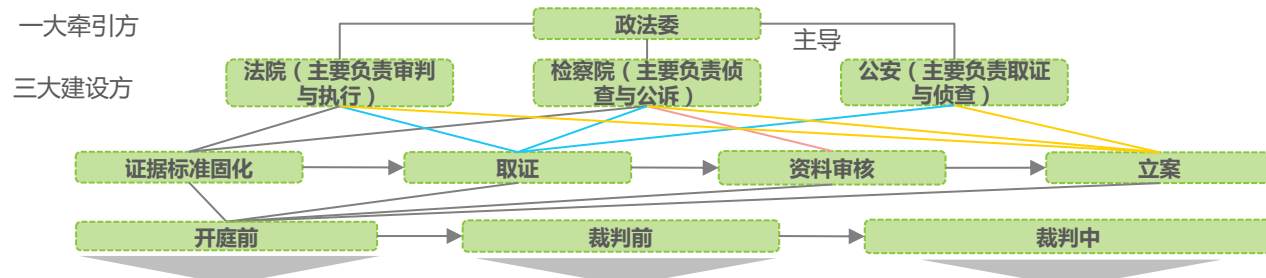
二、建设预算

- 预算充足性：**预算比较紧张。
- 预算紧张原因：**一方面，受新冠疫情影响，政府的首要任务是防疫，防疫需要预算划拨；另一方面，深度调研与项目建设所需预算找不到出资部门。

三、建设情况

- 现状：**简单案件的辅助审判与执行类应用相对成熟，存量市场被政务信息化厂商与AI厂商垄断。
- 未来机会点：**依托数据中台建设向上拓展深度案件应用，以及案件溯源治理（案件调解）。

四、多部门联动的建设机制与场景应用



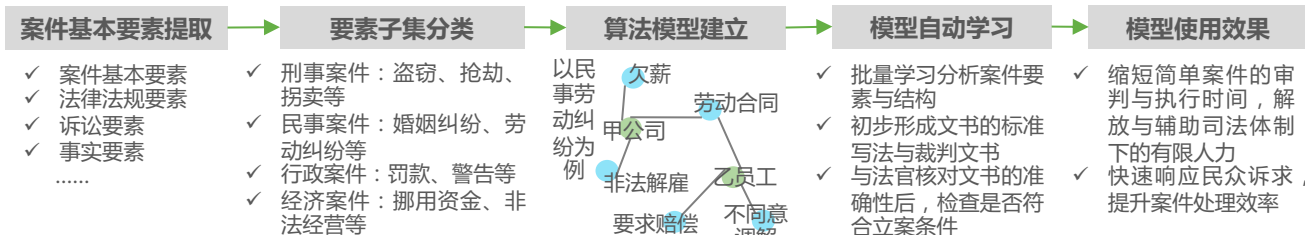
- 核心应用：文书自动生成。**与电子卷宗做捆绑形成智能生成系统。依据公检法三大部门获取的证据与资料，自动生成案件基本要素，形成争议焦点知识网络。
- 核心原理与价值：**通过自动提取和对比公诉意见书、起诉书、答辩状等材料中的诉辩意见与事由，归纳案情与争议焦点，生成庭审提纲，为法官提供智能辅助审判。

- 核心应用：类案检索与推送。**相似案件/法规法条推送。
- 核心原理与价值：**通过对情节、证据、争议焦点、法律适用等关键信息的自然语义识别，形成知识图谱案件知识库，出现目标案件时，通过提取类型案件的关键要素，与系统内的历史案件进行匹配，自动推送类似的案例。

- 核心应用：裁判结果预判。**依据历史案件关联与假设，进行类案推演，预估结果。
- 核心原理与价值：**根据已有类案进行推演，延伸到案件结果的预测，如量刑预测和经济成本预测等。此外，从监察管理的角度，管理者可以对判案结果进行类案判决的偏离预警，对于“同判度”较高的类案，基于预测性判断对法官制作的裁判文书判决结果与之发生的重大偏离情况，进行自动预警从而防止裁判尺度出现重大偏离。

应用范围：无论是在刑事案件还是民事案件中，知识图谱都广泛应用于事由与关系较为简单的案件，如盗窃、婚姻纠纷、金融信用卡、道路交通事故等，而在相对复杂的案件中，主要还是依靠人进行审判与执行。

五、建设流程



注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

来源：中国电子技术标准化研究院《2019年知识图谱标准化白皮书》，艾瑞咨询研究院根据专家访谈自主研究及绘制。

场景应用分析三：政务便民服务

知识库与图谱网络分析服务于咨询问答与分析预警场景

政务便民服务知识图谱业务情况

一、政务便民服务主要业务痛点

Part1 群众：

- 业务办理时间长、手续多且复杂，办证多，办事难；
- 帮助资源不足、引导系统不够、资料无预审、排队时间长、业务变更无通知；
- 大厅或客服坐席上班时有限，但群众业务咨询需求量大，业务诉求反复但得不到及时解决。

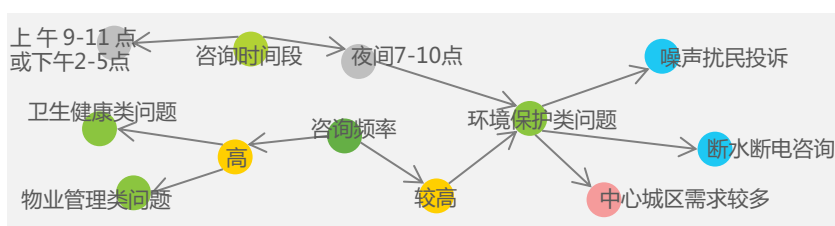


Part2 政务服务人员：

- 办件量大，工作人员负荷重；
- 审批事项多，学习成本高；
- 审批过程仍存在大量重复性作业，人员效率得不到充分利用；
- 办理进度难查询、政策变更难应答、共性问题规律难发现。



三、知识图谱应用举例：12345热线平台中的图谱网络



- 预期目标：**发现群众咨询投诉问题中的共性规律与高频问题，在群众关心的民生问题上提前制定改善措施。
- 核心原理及效果：**围绕居民基础信息和诉求信息，分析诉求趋势、规律、特点，归纳预判居民诉求的周期性、趋势性变化，在群众关心的民生问题上提前制定改善措施。
- 示例：**夜间7-10点时间段内，来自中心城区的环境保护类问题咨询与投诉居多，主要是噪声扰民、断水断电等问题。基于此网络，可提前设置自动引导与问答应用。

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

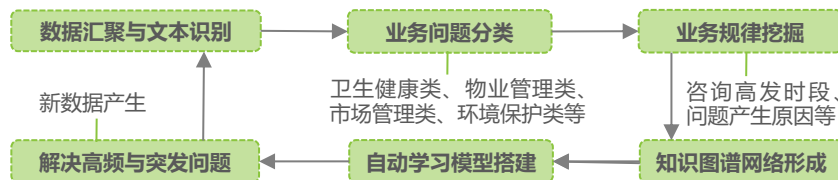
来源：清华大学《2020年数字政府发展指数报告》，艾瑞咨询研究院根据专家访谈及公开资料自主研究及绘制。

二、引入知识图谱切入场景

Part1 咨询问答类场景（结合智能客服直接调用知识库）：



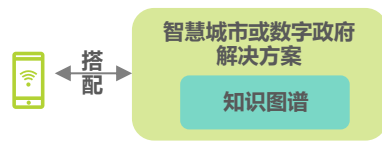
Part2 分析预警类场景（在数据平台/中台中完成）：



四、产品形式及应用区域差异

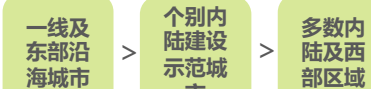
- 产品形式：**嵌套在智慧城市或数字政府解决方案中，鲜少单独存在；与群众端侧的手机App/网页端链接搭配使用。
- 应用区域差异：**以北上广深、天津、山东、江西、浙江、江苏、湖北为应用的领先使用区域，以四川、河南、贵州、海南为应用的起步使用区域，其他地区的应用发展有待提升。

产品形式



应用区域差异

*>代表领先于



场景应用分析四：应急管理

以知识库、关联推理及危险人/物管理服务为核心产品

应急管理知识图谱业务情况

一、应急管理知识相关业务痛点

1 知识体系庞大且信息维度复杂

- 应急管理涉及管理、法律、理科、土木等多个专业学科领域，学科专业性强，整个认知体系具备信息量庞大、信息维度复杂、信息关联紧密等特点。

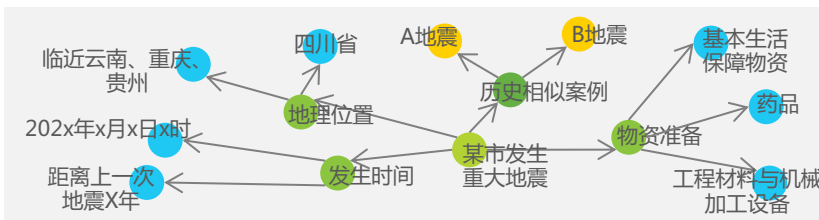
2 应急预案可操作性差且建设不完备

- 部分已制定的各种处置突发性事件预案与应急实践脱节，有的单位对不同层次的预案概念不清，特别是对专项预案、部门预案会存在没有及时补充修订、没有经过专家论证等问题。

3 缺乏智能决策支持

- 尚未充分挖掘并利用紧急事件的数据及数据间的关系，紧急事件缺乏基于数据与知识作支撑的事前、事发、事中、事后4个阶段的决策支持。

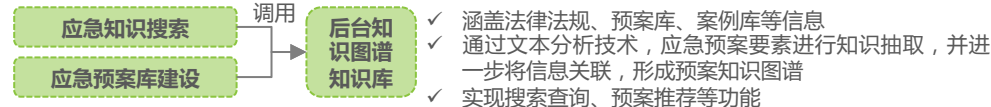
三、应用举例



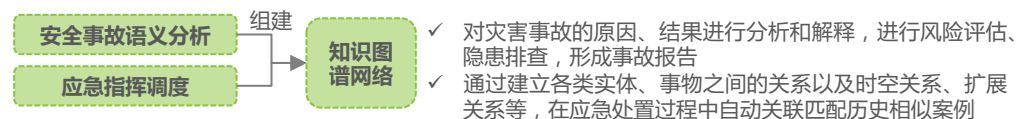
- 预期目标：**及时出具紧急方案，实现抢险救灾。
- 核心原理及效果：**依据地震的即时数据与资料，以及城市自身的区位条件、周围城市物资与人力的响应速度等数据，结合知识中台的知识库与自动化工具，快速构建地震抢险救灾知识图谱，形成第一时间的应急解决方案。
- 示例：**某市发生重大地震，急需制定应急管理解决方案。通过应急管理知识图谱，该市基于数据形成发生时间、地理位置等实体且发散关联关系，寻找历史相似案例，对比出具方案，并依据应急知识库列出所需物资，辅助救灾人员完成方案思考与拟定。

二、核心产品类型及功能

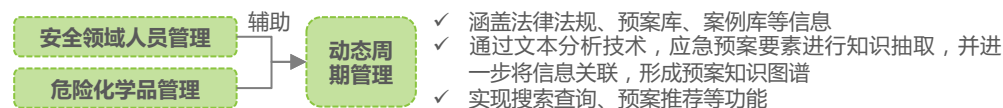
■ Part1：智能知识库类



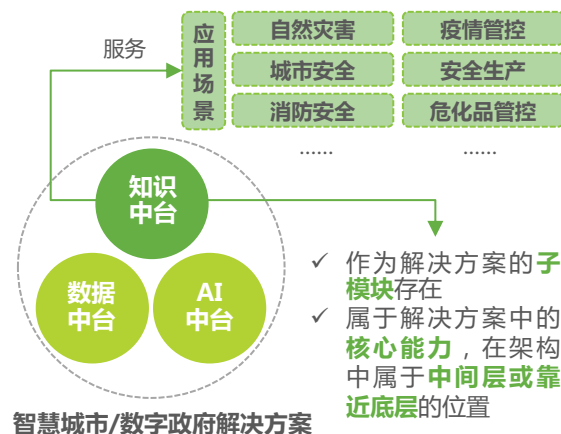
■ Part2：关联推理推荐类



■ Part3：危险人/物管理类



四、产品形式



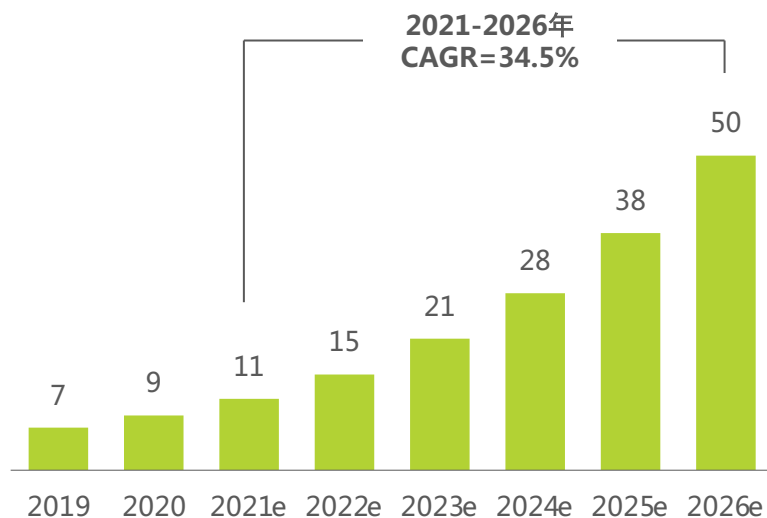
注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。
来源：艾瑞咨询研究院根据公开资料自主研究绘制。

市场规模分析

政务市场走势更快，公安市场以地市级公安为主

因公安的信息化与数字化基础较为领先，知识图谱与公安的业务吻合度较高，公安知识图谱的核心市场规模先于政务领域起量，且起初的规模量级更大。政务的体系庞大复杂，部门与机构众多，尽管其潜在的应用需求远远大于公安市场，但各部门或机构的文本化数据与数据化数据难以共享并统一，政务机构对知识图谱的认知还有待培养，知识图谱应用建设需要一个认知过渡期，同时，政务某一场景的数据只能适用于该场景下的知识图谱，其知识图谱的延展性与通用性差，故市场规模在后期才会得到快速增长与释放。从CAGR（复合年均增长率）对比来看，政务知识图谱的增长速度更快，相应的业务场景与应用将迎来更清晰具体的规划。公安知识图谱的市场规模以地市级公安为主，县级公安的需求与建设投入较低，市场占比偏小，但有个别自治县花费大量投入进行底层基础设施与上层知识图谱应用开发为一体化的重量级建设。

2019-2026年中国政务知识图谱核心市场规模

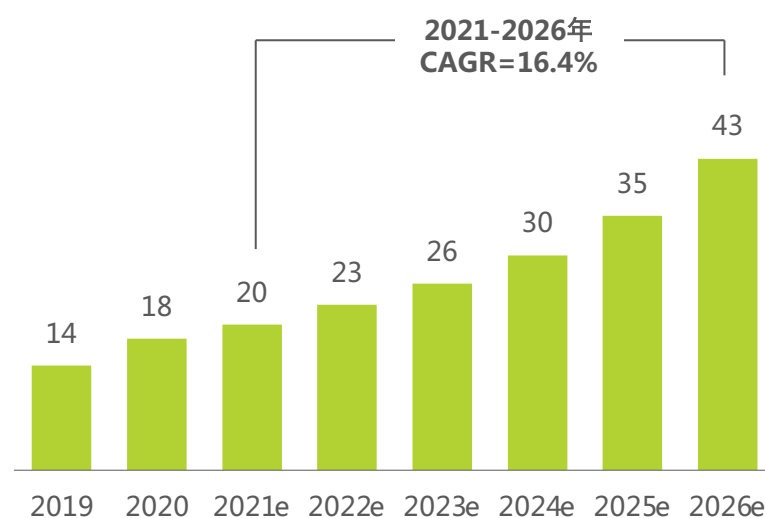


■ 中国政务知识图谱核心市场规模（亿元）

注释：核心规模为软件规模，主要为中国政务云市场规模中的SaaS层、PaaS层中的知识图谱服务与技术工具。

来源：艾瑞咨询研究院根据专家访谈、采招数据、各省/地市政府等数据自主建模测算。

2019-2026年中国公安知识图谱核心市场规模



■ 中国公安知识图谱核心市场规模（亿元）

注释：核心规模为软件规模，包含公安大数据底座产品、AI平台产品的知识图谱部分、研判分析与情报分析等行业知识图谱解决方案的核心软件部分。

来源：艾瑞咨询研究院根据专家访谈、采招数据、国家统计局等数据自主建模测算。

行业知识图谱：医疗篇

- 每一行业的知识图谱内容包括三个部分：行业信息化与数字化基础观察、场景应用分析、市场规模分析。
- **总体评价：**以往的医疗知识图谱被视为“媲美专家，超越专家”的存在，但实际应用建设仍存在诸多问题，科学地将其视为医学辅助工具是业内需要正视的事实。潜在产业空间可期，但产业变现能力将长期受到医院配合度、资深专家认可度、基本库建设、深度NLP技术等因素的挑战。
- **行业信息化与数字化基础：**整体信息化基础良好，三级医院率先进入数智化建设及应用的初级阶段，基本足以支撑知识图谱应用建设。
- **场景应用：**以医疗流程辅助、医学科研、医疗用户服务为核心场景，医疗流程辅助应用的临床价值高，医学科研的创新研发价值高。
- **市场规模：**所测算规模为TAM，2021年中国医疗知识图谱市场规模为10亿元，2026年可达23亿元，2021-2026年CAGR=19.7%。

信息化与数字化基础

整体信息化基础良好，优先进入知识图谱应用阶段

我国医疗信息化从上世纪九十年代开始，经过近30年左右的实践，已取得一定成果。各级医院的信息化水平虽然参差不齐，但除基层医疗机构外的各级医院均已涉足信息化建设。随着医疗业务由粗放向精细过渡，医院内外部的业务规模不断扩张与复杂度提升，数智中心、医联体等新兴的信息化建设兴起与展开，三级医院凭借先天信息化优势以及较为完备的医疗资源储备成为此类建设的先行者。此外，国内的跨国药企已在开展数字化转型，而本土药企仍在持续进行信息化转型。相对良好的信息化基础为医疗知识图谱的构建创造了有利条件，医疗知识图谱得以凭借较快的速度进入应用阶段。

医疗信息化整体建设情况

参与者/信息化分类		医院			医疗政府单位	健康管理机构	药企	
一级分类	二级分类	三级	二级	一级及未定级	卫健委/医保局/药监局	体检机构/药店等	跨国药企	本土药企
医院信息化	HIS							
	HRP							
	数智中心							
临床信息化	CIS							
	EMR							
	CDSS							
	PACS							
区域医疗信息化	医疗集团							
	医联体							
药企信息化	信息化转型							
	数字化转型							

图例



颜色由浅至深表示建设水平由低至高

来源：艾瑞咨询《2022年中国医疗信息化行业研究报告》，艾瑞咨询研究院根据公开资料自主研究绘制。

场景应用分析一：医疗流程辅助

辅助院内外医疗流程，提升服务效率

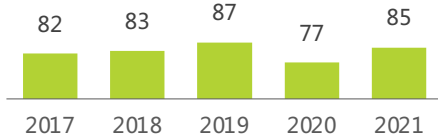
医疗流程辅助知识图谱业务情况

一、主要业务痛点

○ Part1 患者端：

- **诊疗人次基数大：**攀升的诊疗人次意味着医疗卫生需求的增加，诊疗需求缺口逐步扩大。据统计，2021年中国卫生机构诊疗人次已接近85亿人次，已恢复至疫情前水平。
- **人均医疗卫生费用增加：**2021年中国人均卫生费用已突破5300元，相比去年增长4.6%，居民在个人医疗卫生上的花费增加。

2017-2021年中国卫生机构诊疗人次



○ Part2 医院端：

- **三级医院：**资深医师工作超负荷，诊疗方案出具量巨大；患者等待时间长，易产生医患矛盾；年轻医师资历不足，需要持续指导。
- **二级医院：**资深医师数量有限，诊断能力较低，可能存在误诊漏诊情况。
- **一级及未定级医院：**机构数量最多，是国家推进分级诊疗的重点分诊医疗机构层级，但医师与医疗设备稀缺，诊疗经验不足，误诊漏诊现象较为严重。

○ Part3 监管侧：

- **多重因素加剧医保支出，医保控费趋紧：**居民自身的合理的医保消费、人口老龄化的医保支持、疾病谱改变和新病种的出现等不可控因素，以及医保费用增长存在供方诱导需求与需方过度消费等可控因素，都直接或间接地导致医保费用的增长。
- **DRGs付费工作持续推进：**在医保局、卫健委的推动下，DRGs付费的试点区域扩大，控费方式逐步精细化，急需借助数字化与智能化手段辅助DRGs分组。

二、应用建设现状

○ Part1 已嵌入应用：

- **CDSS：**将医学知识库与医疗实例构建知识图谱网络，在诊中服务于单病种或多病种辅助诊断，达到分诊效果。
- **智慧病案与DRGs：**将病历与疾病数据构建知识图谱网络，关联患者疾病与应享有的医保支付范围，控制医保开支。
- **智能医药：**将患者病历的临床诊断与药物做关联，辅助药方出具、提供用药建议，并进行用药审核。

○ Part2 应用水平：

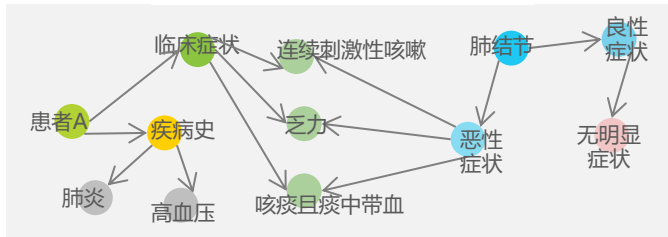
- **CDSS达到初级应用阶段：**针对感冒、肺结节等简单病种可提供辅助诊断服务，在初级智能分诊服务应用比较广泛。
- **中高级应用仍在研发：**针对中级的智能影像辅助分析诊断、高级的DRGs与智能用药，还停留在实验研究阶段。

○ Part3 应用建设不足：

- **急于评级而忽视质量：**受评级相关政策的影响，厂商为搭上评级的“快车”，将CDSS等产品作为评级创收的工具，忽视了知识图谱临床价值深度的开发。
- **基本库建设欠缺：**基本库包括疾病库、治疗库、药品库、个人健康档案、医生档案等，现阶段的基本库并不完整，库中的数据资料也不充足，难以进行调研与Schema构建。

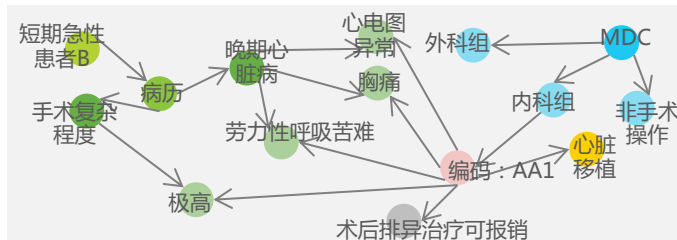
三、应用举例

案例1：CDSS临床辅助决策



- **预期目标：**依据临床症状的关联节点，辅助年轻医师或一级及未定级医院诊断出恶性肺结节。
- **核心原理：**基于疾病库与患者的个人健康档案，将恶性肺结节症状与患者A的临床症状关联起来，并查看患者A的疾病史，提示医师患者A可能患有恶性肺结节。
- **服务部门：**门诊部、检测部。

案例2：DRGs疾病与医保关联审核



- **预期目标：**依据短期急性患者B的临床诊断与症状等关联节点，进行DRGs分组，核实医保可报销的治疗费用范围。
- **核心原理：**基于疾病库、患者的个人健康档案、CHS-DRG目录等知识库，将患者B的临床症状、手术复杂度等的相关内容与心脏移植手术的相关内容做关联，提示患者B的术后排异治疗在医保报销范围。
- **服务机构：**医院与医保局。

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

来源：艾瑞咨询《2021年中国人工智能+医疗与生命科学行业研究报告》，《2021年中国卫生健康事业发展统计公报》，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

场景应用分析二：医学科研

创新科研方法，提升科研创新实力与研究效率

医学科研知识图谱业务情况

一、主要业务痛点

Part1 临床科学研究：

- 文献筛选量大、复杂度高：对顶级三甲医院而言，为进行临床科研创新研究，获取学术排名，其需要自创领先的医学论文作支持，而医学论文中的文献参考、病理假设推断与数据关联等关键工作带有筛选量大、复杂度高特点，研究人员急需借助数智化工具提升科研工作效率与成功的可能性。

Part2 药物研发：

- 居民疾病负担加重：我国的心血管疾病与癌症负担最重，且随着我国人均期望寿命的延长与老龄化加剧，慢性呼吸疾病、糖尿病与肾病、肌肉骨骼失调等慢性疾病负担亦会加重。此外，疫情的出现与变异对整体社会面的健康卫生造成巨大威胁。由此，针对各类慢性病、癌症与传染性疾病，自主研发创新药物，减轻居民疾病负担已成为制药的一大趋势。
- 药企的经济利益与药物研发业务难度驱动：一方面，药企药物研发成功可获取创新药专利，抢占新药专利红利，但国内药企的创新药专利较为稀缺；另一方面，药物研发具有高技术、高投入、高风险、低成功率、审批慢的痛点，药物发现阶段的数据量与计算量巨大。
- 国家控制药物成本的工作持续推进：提升药物创新研发能力，不仅可以控制外国药物进口成本，而且可以改善病人因经济问题无法购药的问题。

二、与机器学习相比，知识图谱在药物研发领域的优势

1 频繁子图挖掘

机器学习的数据特征更多依赖人为筛选，而知识图谱可以做到频繁子图挖掘，即从大量的图中挖掘出满足给定支持度的频繁子图，同时依据算法保证这些频繁图不重复。知识图谱的图特性使其可以更快更自动化地挖掘出药物子图。

2 数据维度丰富

知识图谱可基于多维度数据构建药物知识库，将病理学、历史用药导致结果等多角度的数据进行关联整合分析，以直观的图形形式揭示复杂的药物原理与推理关系，为药物研发提供便捷的工具。

3 可解释性增强

相比于机器学习的黑箱问题缺陷，知识图谱可以结合贝叶斯网络与GNN等算法，过程中的每一参数与节点流程都可展示呈现，可为数据分析、逻辑推理提供更有说服力的证据链与证据流程，提高模型的可解释性。

三、药物研发业务的主要细分应用场景

场景1: 老药新用



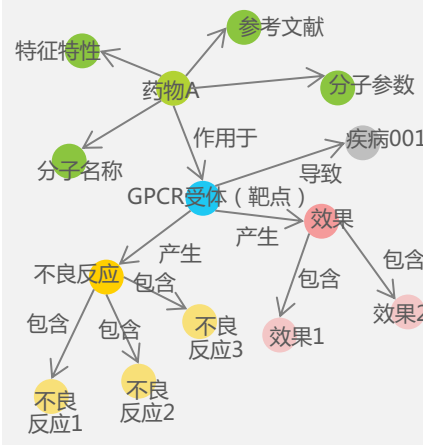
- 核心原理：老药已经具有清晰的拓扑分子结构，可以直接用知识图谱表达出来，与新靶点与新适应症直接做关联，预测老药效果。
- 核心优势：可以快速提取老药的结构与特征，提升研究效率。

场景2: 靶点与药物相互作用预测



- 核心原理：使用新药的分子结构与特性等数据，与患者的靶点做关联，预测新药在靶点的作用结果。
- 核心优势：通过自动分析新药与靶点间的潜在不良反应，快速给出相互作用的结果。

四、应用举例：以药物与靶点之间的相互作用预测为例



- 预期目标：以较快的速度查看新研发的药品A作用于GPCR受体的效果与不良反应，缩短药物研发时间，发现药物潜在风险。
- 原理分析：基于药物数据库与靶点数据库构建知识图谱网络，将药物A与GPCR受体关联，发现在药物A的特性作用下，靶点得到有效治疗的同时，会产生多种不良反应。
- 数据特点：以文本化数据为主，带有部分数据化数据，数据化数据作为数值属性而存在，如某组实验数据是某类药物分子药理实验的结果，可以将该数据赋予到该分子的属性中。
- 服务机构：药企、CRO、AI制药公司。
- 算力需求：因计算量巨大，往往需要配备足够的算力资源，甚至采用超算的方式进行运算。

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

来源：艾瑞咨询《2021年中国人工智能+医疗与生命科学行业研究报告》，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

场景应用分析三：医疗用户服务

应用面积最广，落地化程度最高，助力医师培养与患者问诊

医疗用户服务知识图谱业务情况

一、主要业务痛点

Part1 网络搜索引擎自带缺陷：

- 信息过载，难以聚焦：网络百科式的搜索引擎最大的问题在于信息过载，数据量巨大，难以在纷繁杂乱的网络信息中过滤无效信息，快速定位到精准的结果。
- 缺乏逻辑与常识：网络搜索引擎在检索医学知识时，难以直接在聚焦的医学信息范围内进行逻辑推理，检索出来的某些结果缺乏常识。

Part2 来自医生与患者的巨量检索需求：

- 医生学习训练的检索需求：对于资历尚浅的医生而言，其需要依赖医学知识检索、医学经验检索、患者信息检索等，反复训练与沉淀医学知识经验积累，以支持职业生涯发展。
- 居民问诊与医疗保健的检索需求：患者在入院就诊前，往往借助网络搜索引擎查看自身的症状符合哪类疾病，需要挂哪类科室。此外，出于保健养生的考虑，中老年人群体会检索饮食作息注意事项。

二、应用建设现状

Part1 已嵌入应用：

智能医学知识机器人

- ✓ 将对话知识结构化
- ✓ 关联症状、疾病等实体
- ✓ 精准快速回答问题

智能导诊分诊

- ✓ 诊前信息自动采集
- ✓ 推荐就诊科室与医生
- ✓ 提供用药建议

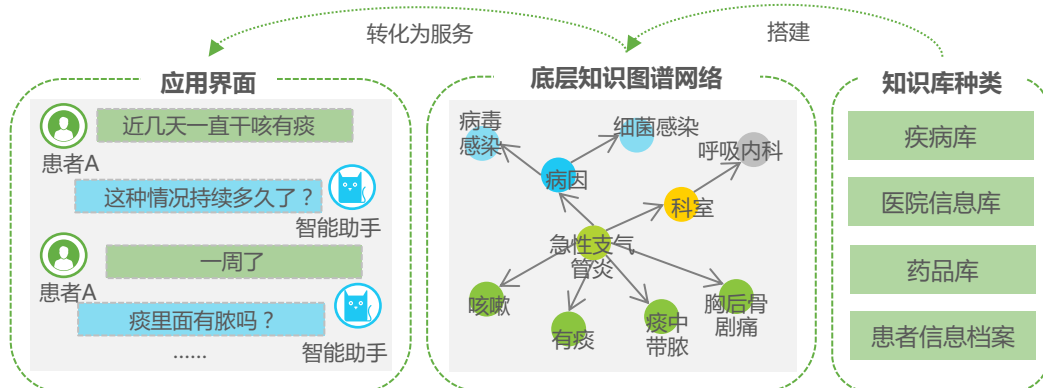
Part2 应用建设不足：

- 底层基本库构建不完善，尚不能满足大量需求：应用真正落地时，会更多地结合FAQ问答对式的问答，知识图谱难以将专业的文献知识完全打散并承载下来，更多是基于网络信息线索去做关联，而关联所需的基本库仍需构建。
- MRC技术有待突破：相对复杂的问题需要借助MRC技术，进行文字阅读理解，但MRC技术难度比较大，需做好完形填空、多项选择等方面的训练。

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

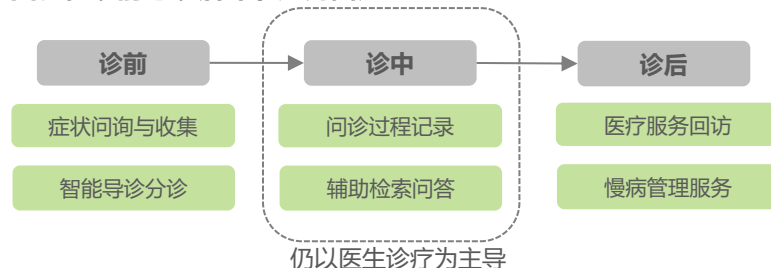
来源：艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

三、应用展示：智能医学知识机器人



- 预期目标：快速准确地询问患者A的临床症状，确定可能的疾病类型，告知患者应就诊科室。
- 核心原理：分析并理解医患沟通对话，抽取对话中患者的疾病、症状、用药等相关内容，关联到后台的知识图谱网络，定位到可能的疾病类型，推理出病因、应就诊科室及可用药物。
- 核心优势：更好地理解语言中的抽象部分，具备逻辑性与常识性。通过智能对话来完成患者病状询问，在检索上能够聚焦于问题本身，更快更精准地定位到相关的医疗知识信息，节约患者与医生时间，并提供专业的导诊分诊建议。

四、在诊前与诊后环节应用面广



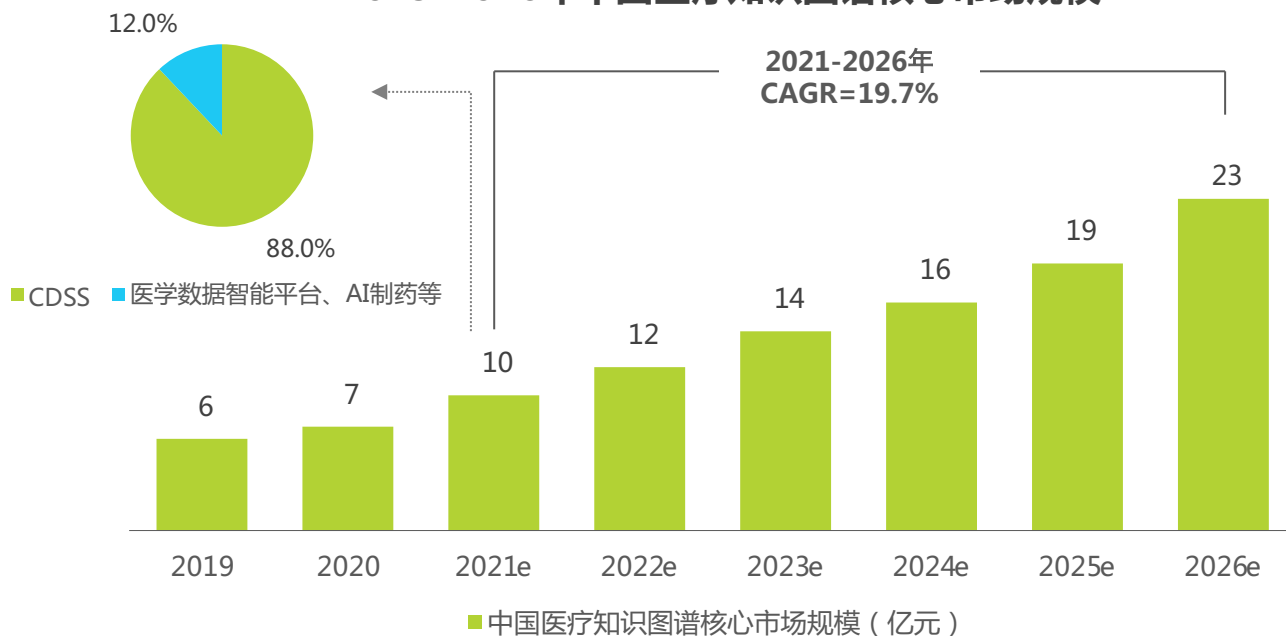
- 诊前与诊后环节均有大量相对简单的重复性基础工作，且业务咨询量大，结合了NLP与知识图谱的应用可以处理该类型的工作，解放人力，让医生投入到业务价值更高的诊中环节。
- 在诊中环节，应用可以做基本的问诊过程记录与辅助性的医学知识检索问答工作。

市场规模分析

院内产品以CDSS为主，院外产品将逐步实现商业化转变

医疗知识图谱主要嵌入院内的CDSS、医学数据智能平台或科研平台，以及院外的AI制药解决方案而存在，目前以CDSS应用最为广泛。在2021年国内医疗知识图谱核心市场规模中，CDSS占比为88.0%。从院内市场看，核心知识图谱市场仍以三级医院（尤其是三甲医院）为主导，未来随着评级与分级诊疗政策的推动，市场会进一步下沉至二级医院。从院外市场看，除AI制药解决方案外，DRGs医保控费或防医保骗保应用、个性化的用户医疗保健应用有望走入市场，逐步实现从技术研发到商业化的转变。据艾瑞统计测算，2021年中国医疗知识图谱核心市场规模为10亿元，2026年相应规模可达23亿元，2021-2026年CAGR=19.7%，市场整体增长速度较为平稳。

2019-2026年中国医疗知识图谱核心市场规模



注释：核心规模为软件规模，主要包含医学数据智能平台、CDSS的知识图谱部分、AI制药的知识图谱部分。

来源：艾瑞咨询研究院根据专家访谈、采招数据、中国卫生统计年鉴等数据自主建模测算。

行业知识图谱：工业与电力篇

- 每一行业的知识图谱内容包括三个部分：行业信息化与数字化基础观察、场景应用分析、市场规模分析。
- **解释**：因电力的信息化与数字化建设、知识图谱业务独立性高，可脱离其他工业领域存在，故将电力与工业并列，合并为同一章节阐述分析。
- **总体评价**：工业与电力知识图谱应用处于起步期，有较高的行业准入门槛，应用效果有待释放，现阶段到未来几年间可明确变现的产业空间较小。
- **行业信息化与数字化基础**：电力行业的信息化与数字化进程较快，领先于工业整体且率先迈入数智化阶段。工业各子领域的信息化水平不一，顶尖的汽车与航空制造、电子制造信息化水平领先。
- **场景应用**：以新产品与新工艺研发、安全质量控制、供应链管理为核心场景，新产品与新工艺研发应用的业务价值最高但开发难度也最大，安全质量控制为落地领先的应用，供应链管理的应用面有待拓展。
- **市场规模**：所测算规模为TAM，2021年中国工业与电力知识图谱市场规模为6亿元，2026年可达17亿元，2021-2026年CAGR=25.4%。

信息化与数字化基础

信息化水平参差不齐，电力信息化水平领先于工业整体

工业整体的信息化水平参差不齐。采用大规模生产的石油化工、能源、钢铁制造行业生产工艺较为成熟，生产过程多采用PCS系统（生产过程控制系统），对连续性流程生产加强管理，信息化水平较为领先，在原料开发与工艺设计方面可结合知识图谱发挥高价值。离散工业对分散物资供应链的依赖性强，在航空、汽车、电子制造等领域的信息化水平靠前，知识图谱在汽车制造与航空制造能够发挥安全质量控制的作用。与工业整体相比，电力行业的整体信息化与数字化水平更为领先，主流的南网、国网信息化已经非常普及与完备，整个行业现已跨入以数智化为建设目标的快速发展阶段。电力行业领先的信息化与数字化水平为知识图谱应用建设创造了有利条件，现已发展出设备检修、内外部知识库检索与问答等应用。

工业信息化水平及知识图谱应用价值情况



根据《“十四五”信息化和工业化深度融合发展规划（2021~2025）》：

- 企业经营管理数字化普及率达**80%**；数字化研发设计工具普及率达**85%**；关键工序数控化率达**68%**
- 工业互联网平台普及率达**45%**，系统解决方案服务能力明显增强，形成**平台企业赋能、大中小企业融通**发展新格局

来源：艾瑞咨询研究院根据专家访谈及公开资料自主研究绘制。

电力信息化与数字化进程

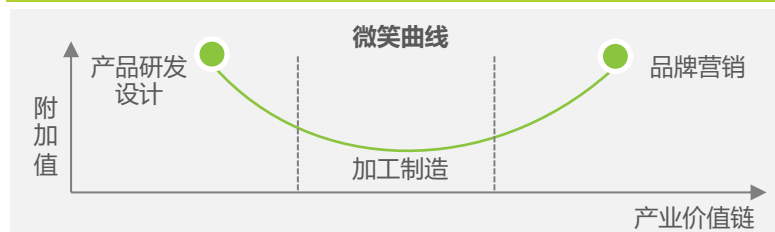


来源：艾瑞咨询《2022年中国电力产业数字化研究报告》，艾瑞咨询研究院自主研究绘制。

场景应用分析一：产品与工艺创新研发 iResearch 艾瑞咨询

高附加值环节依赖国外，需借助知识图谱创新产品与工艺 工业新产品与新工艺研发知识图谱业务情况

一、主要业务痛点：新产品与新工艺研发设计不足



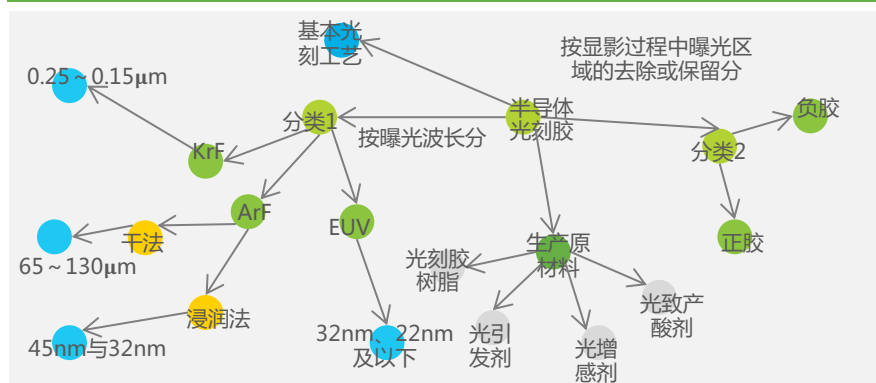
- **高附加值环节依赖国外：**国内在低端产品制造方面产能过剩，但在高端精细化工业产品的生产工艺上仍然依赖国外，高技术、高附加值产品不足。
- **研究积累薄弱：**在化工原理、反应机理、油气勘探、核心零部件设计等基础研究方面，国内的研究积累比较薄弱，需要将专业知识与经验进行体系化梳理，辅助新产品与新工艺研发。

二、知识图谱搭建过程



- **知识梳理：**知识图谱搭建的第一步，需要将业务体系下的行业知识梳理好。
- **边界确定：**如某一细分领域为氨纶，将氨纶的业内专家、化工原理等梳理成知识体，明确研究范围。
- **采集过程：**包含知识清洗、知识对齐、知识存储。
- **管理目的：**建立管理机制，对知识进行增减与更新。

三、知识图谱应用举例：以半导体光刻胶的产品与工艺研发设计为例



四、其他场景应用



- 除化工领域外，在油气、航空、汽车、酸奶等细分领域同样存在产品与工艺的研发问题，可以使用知识图谱技术助力产品与设计创新与改良。

注释：工业知识库问答的知识图谱应用类似前文所述，在工业章节不再描述。不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。
来源：艾瑞咨询研究院根据专家访谈及公开资料自主研究绘制。

场景应用分析二：安全质量控制

检修与缺陷排查耗时费力，知识图谱快速提供原因与措施

工业与电力安全质量控制知识图谱业务情况

一、主要业务痛点

Part1 设备易发多发故障，检修知识与经验需汇聚沉淀：

- 工业生产流程与电力调度过程涉及到多类型、多数量的设备，这些设备都有专业的操作步骤与检修知识，设备故障分析耗费大量人力与时间，而故障的历史数据或检修经验尚未被系统化地汇集、开发与利用。如何围绕设备故障的核心数据与经验，提高故障处理效率，节省成立成本与重复性工作，是设备检修的业务难题。

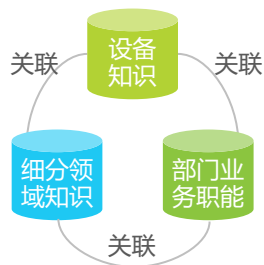
Part2 产品/设备可能存在潜在缺陷，需进行缺陷排查：

- 工业与电力对产品/设备的严谨性与安全性存在高要求，这就需求相关人员对工业/电力整体生产流程所设计设备的安全性、可控性进行严格把关，并进行失效影响分析，完成生产异常溯源，其中涉及大量专业知识与假设，分析检查过程难度大且时间长。

Part3 人力分析可能存在失误或不全面，新手需工具辅助：

- 完全依赖人力劳动带有不可避免的局限性，如因疲劳所导致的失误判断、因认知有限所导致的分析不全面等。此外，新手与经验丰富的专家存在业务能力差距，需要专业简易的工具辅助。

二、知识图谱核心业务价值

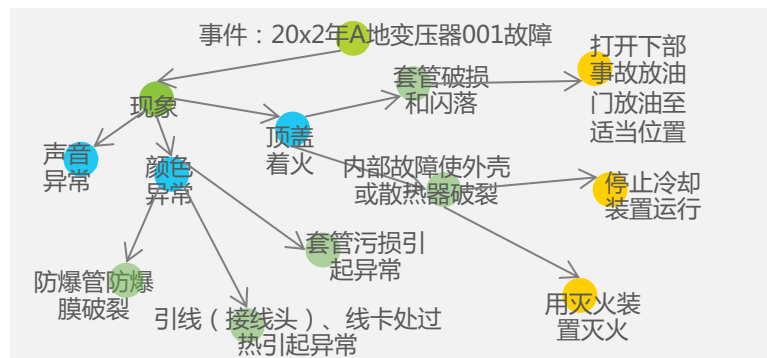


多维度知识关联，快速响应机制建立：

- 知识图谱的广度与维度很多，可以将设备知识、细分领域知识、部门业务职能等多个维度的知识关联起来，从多维全面的角度去分析问题；
- 在设备发生故障时，知识图谱可以快速找到解决方案，如需要什么职位的人去解决、故障的现象可以关联到哪些原因等。
- 在发现设备缺陷时，知识图谱可以溯源至生产厂商、生产环节与流程等，快速分析缺陷原因。

三、应用举例

以电力设备故障检修为例



- 预期目标：**快速发现变压器001的故障原因，并采取相应措施，保证变压器安全并恢复供电。
- 核心原理：**基于过往的变电站故障原因与知识手册等搭建知识图谱网络，根据故障现象快速定位原因与措施。
- 所需数据特点：**以文本数据为主，包含设备知识指南与手册、基本概念、检修经验等。结构化数据可作为数值属性而存在，作为某一故障的具体体现。
- 触发流程：**现场故障设备运行数据发出警告信息——故障文本数据和系统数据构建知识图谱——逻辑分析运算——处置结果产生。

四、其他应用场景

汽车/航空失效影响测试

使用知识图谱协助FMEA，逐一分析构成产品的子部件，找出潜在失效模式并猜想后果，预先拟定紧急措施。

食品安全指标异常溯源

以异常节点为子图，通过知识图谱推理，从根源寻找可能引起食品安全异常的原因。

注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

来源：艾瑞咨询研究院根据专家访谈及公开资料自主研究绘制。

场景应用分析三：供应链管理

通过多维知识图谱网络解决供应链物资供应与成本管理难题

工业供应链管理知识图谱业务情况

一、主要业务痛点

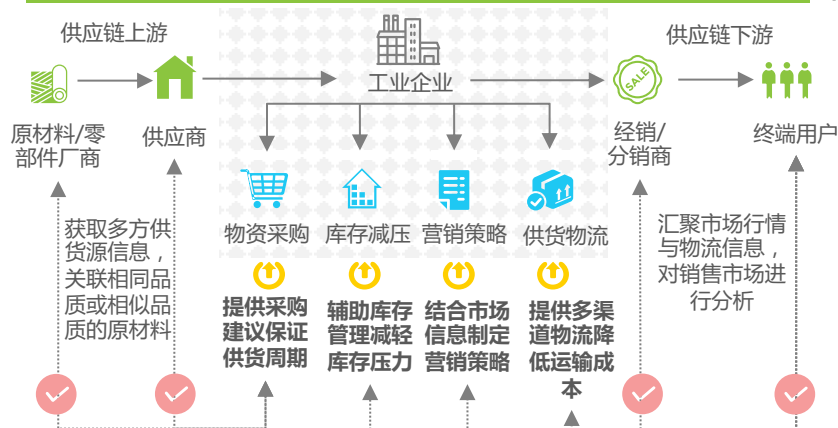
Part1 物资供应难以匹配采购需求

- **采购周期不足导致物资难以到货**：部分工业企业进行采购时，仅考虑市场需求而缺少对库存与采购量之间的衡量，有的领域（如煤炭）采购计划性差，需要随买随用，预留的采购时间紧张，可能带来供应不足或断供的影响。
- **采购价格与物资质量难以平衡**：采购方希望以较低的成本换取优质的物资，但实际采购时由于中间环节多、价格不透明、生产工作量改变等因素的影响，很难找到采购价格与物资品质之间的平衡点。

Part2 供应链成本压力加剧

- **高库存导致的现金流压力**：因地租成本、企业担心潜在的销售损失、客户取消订单、老库存消化不足、产品设计变更或废止等因素的影响，制造业一直存在高库存问题，这对制造企业的现金流造成了不小的现金流压力。
- **其他因素导致的供应链成本压力**：信息流、物流和资金流的信息不透明且多变，以及工业产品设计的的不合理等都会加大供应链的复杂程度和管理难度，最终增加供应链成本。

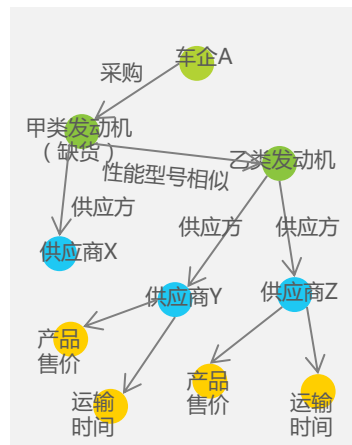
二、知识图谱赋能供应链业务环节，属于较新的应用开发



注释：不同颜色的圆代表不同层次的实体。所展示的知识图谱均为列举，知识图谱可依据开发需求调整结构。

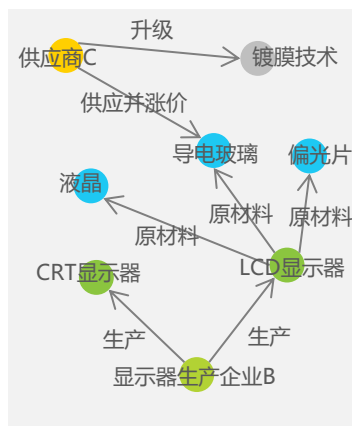
来源：艾瑞咨询《2021年“新基建”背景下中国工业互联网与工业智能研究报告》、《2022年中国供应链数字化升级行业研究报告》，艾瑞咨询研究院自主研究绘制。

案例1: 知识图谱辅助物资补给



- **预期目标**：及时为车企A供应发动机，为车辆提供组装零部件，满足出货需求。
- **核心原理**：基于供应链知识图谱网络，及时发现甲类发动机的缺货状态，寻找性能型号相似的一类发动机进行弥补，并比对多家供应商乙类发动机的产品售价与运输时间，辅助车企A选择性性价比更高、运输时间更短的供货方案。
- **强需求细分领域**：因离散工业需组装来自多方供应商的元件、部件等，供货受到企业与供应商关系、供应商内部生产等多重因素的影响，故对供应链知识图谱有较强需求。

案例2: 知识图谱协助成本分析与营销策略制定



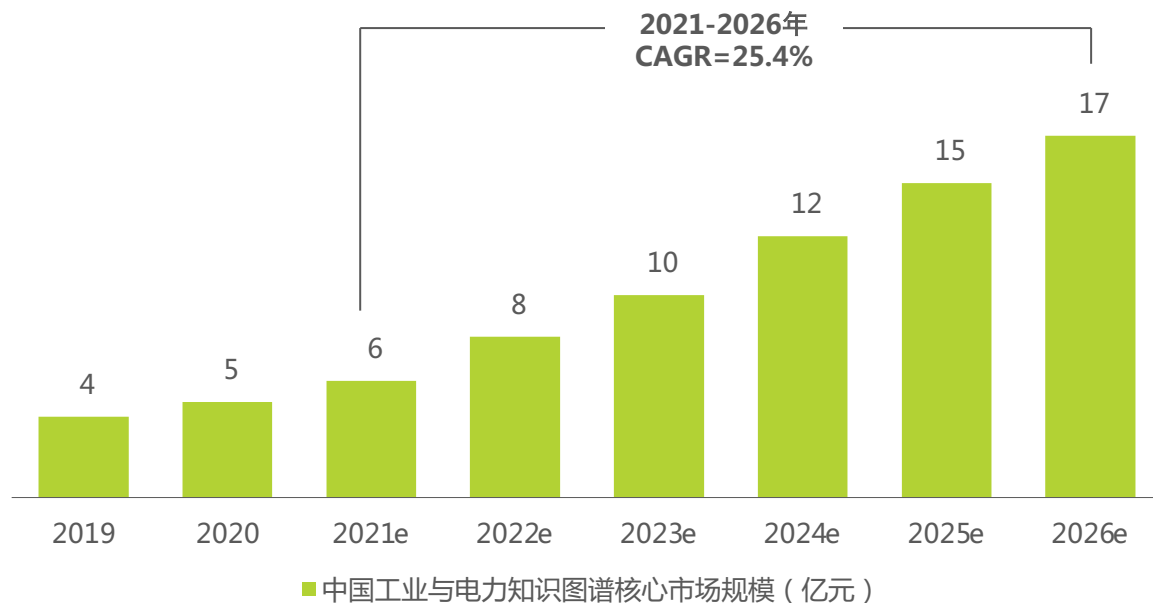
- **预期目标**：帮助企业B分析LCD显示器生产成本上涨原因，并基于此调整营销策略。
- **核心原理**：基于供应链知识图谱网络，发现LCD显示器生产成本提高的原因为导电玻璃的供应商C升级了镀膜技术，提示企业B提前制定相应的显示器促销方案，应对由成本上升所导致的售价上升、销量降低的情况。
- **服务对象**：财务分析人员与市场营销人员。
- **强需求细分领域**：与个人消费密切相关的食品工业、电子产品工业等。

市场规模分析

应用场景开发仍处于探索期，市场空间有待开拓

鉴于工业整体的信息化与数字化基础较差，文本数据、数值数据的归集与治理工作有待展开，底层数据工作的滞后在一定程度上影响了工业知识图谱的搭建工作。同时，工业细分领域众多，应用场景开发还处在探索期，可落地场景相对模糊，知识图谱应用仍待培养。在工业与电力领域，石油化工、能源、航空与汽车制造、电力行业率先树立起知识图谱建设的标杆，采购嵌入知识图谱技术的大数据解决方案，以及以知识图谱开发为主要内容的行业知识图谱解决方案，首先服务于生产流程的设备故障与缺陷排查、专业知识检索问答等业务。据艾瑞统计测算，2021年中国工业与电力知识图谱核心市场规模为6亿元，2026年相应规模可达17亿元，2021-2026年CAGR=25.4%，市场整体处于开拓阶段，增长速度较快。

2019-2026年中国工业与电力知识图谱核心市场规模



注释：核心规模为软件规模，包含大数据产品、工业行业知识图谱解决方案的核心软件部分。

来源：艾瑞咨询研究院根据专家访谈、采招数据、国家统计局等数据自主建模测算。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

5

底层架构形成高服务门槛，平台开发兼顾标准化与差异化

明略科技通过大数据、数据库、图计算与人工智能技术，构建了企业级知识图谱平台，帮助企业及政府实现线索洞察、知识沉淀与管理。明略知识图谱平台由数据源、数据感知层、知识提取层、图计算挖掘层、知识服务、知识应用、基础服务与行业解决方案组成。知识管理层以下均为核心底层架构，是知识图谱服务及应用能否达到好用与高效的关键。在核心底层架构上，明略在数据数量与种类、混合数据结构存储、业务理解、模型沉淀等方面具备丰富而深厚的核心技术与经验积累。此外，在产品标准化程度上，明略将行业Know-How沉淀到基础工具中，辅助数据快速匹配与模型二次开发，兼顾了平台工具的标准化与场景开发的差异化。

明略知识图谱平台



来源：艾瑞咨询研究院自主研究绘制。

基于平台打造解决方案，助力金融风控与快消新品研发速度

针对金融风险稽核场景的业务痛点，明略基于知识图谱平台推出金融风险稽核平台解决方案，可实现十几亿实体、几十亿关系、近百亿事件的构建，服务于贷款资金回流、疑似受控账户、资金流向追踪、员工与资金掮客之间的资金往来关系等场景，及时发现员工异常行为与违规贷款并监控预警，提高了远程风控水平以及非现场审计的效率和能力。针对快消新品研发场景的业务痛点，明略基于知识图谱平台推出快消新品研发解决方案，以数月的研发周期替代2-3年的研发周期，降低调研成本，充分挖掘互联网消费者的需求并分类，指导新品研发，提升产品上市好评度。

明略知识图谱在金融与快消领域的应用

应用一：金融风险稽核知识图谱

业务痛点：商业银行自身的内控与稽核部门有非现场稽核或监管的需求，需要挖掘埋藏在内部的各类数据中的违规违法线索。然而，由于原有技术限制、内控合规体系不完善等原因，商业银行难以发现银行内部员工与外部违规、违法团体或个人勾结的风险行为。

解决方案设想



能够融合所有业务数据、多种外部数据，且能够发现数据隐藏关系，达到监控预警效果的企业级金融风险稽核平台。

应用二：快消新品研发知识图谱

业务痛点：在快消品领域，传统的新品研发需要借助调研公司的力量，通过问卷调研、产品试用/试吃的方法进行测试，通过线下寻找潜在消费群体进行调研，把每个模块的验证积累起来形成产品参考。整个研发测试周期漫长（一般需要2-3年），耗费较高的测试成本，未充分利用互联网消费者数据。

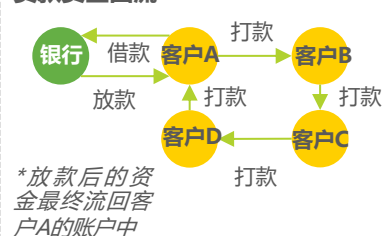
解决方案设想



能缩短产品研发时间、降低调研成本、挖掘网民数据的快消新品研发知识图谱解决方案。

场景预演（可实现十几亿实体、几十亿关系、近百亿事件构建）

贷款资金回流



特点

客户贷款后在一段时间内、经过不同账号将符合一定特征的资金通过资金网络流转回客户本身账号子圈。

疑似受控账户

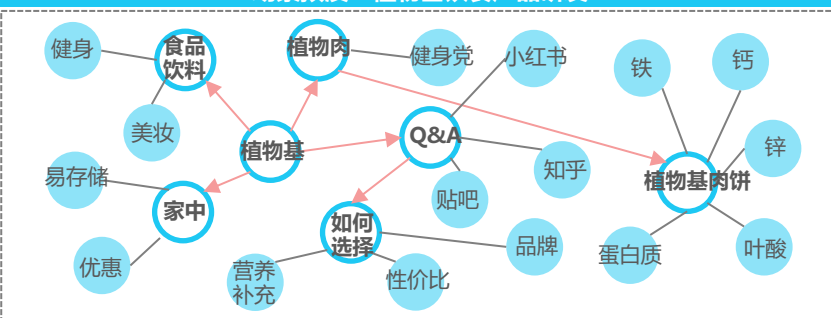


特点

客户贷款后在一段时间内、经过不同账号将符合一定特征的资金通过资金网络流转回客户本身账号子圈。

来源：艾瑞咨询研究院自主研究绘制。

场景预演：植物基饮食产品研发



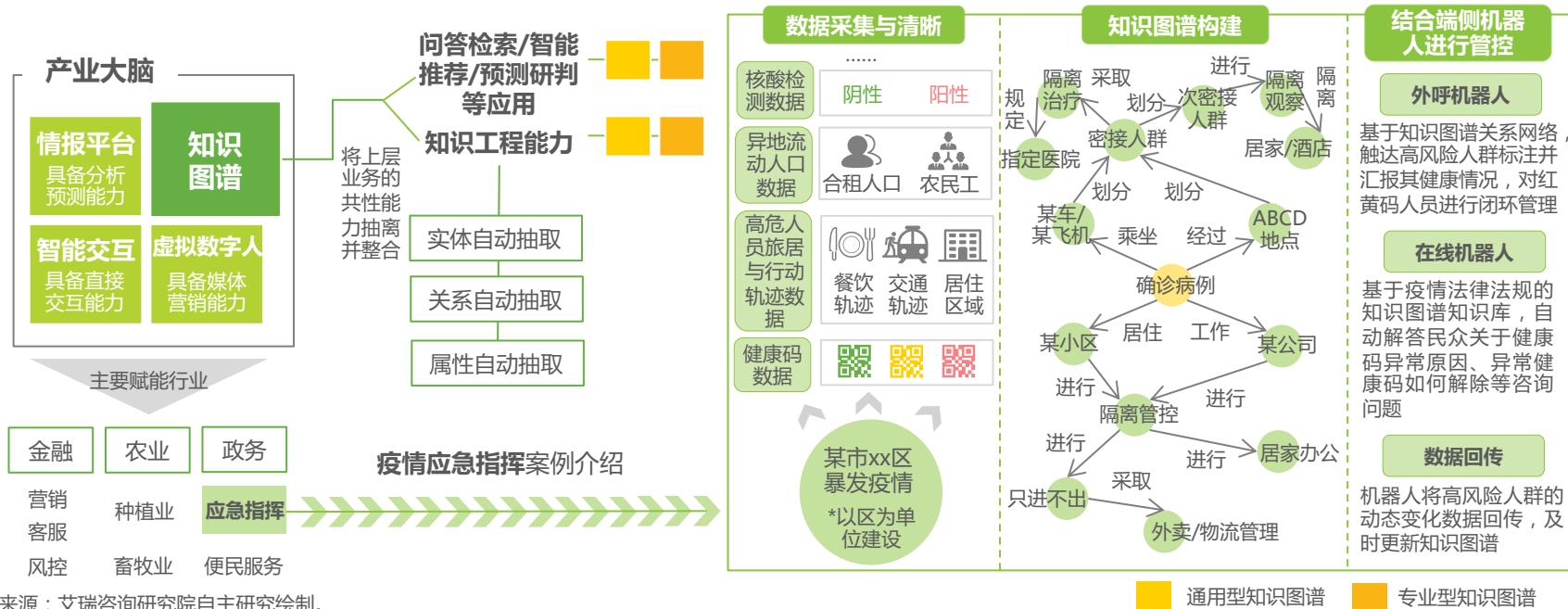
价值点

缩短研发时间，提升产品上市好评度；快速了解消费者对产品的关注情况，深入分析消费者对产品的各类反馈；通过消费者的反馈定位竞争对手，更全面地对比本竞品市场反馈；基于舆情大数据发现市场创新点，评估各创新点对消费者购买驱动的影响。

知识图谱为产业大脑子模块，疫情期间服务疫情应急指挥

远传科技是以知识图谱与自然语言理解为核心技术的高新技术企业，其依托产业大脑，积极推动知识图谱与智能交互产品的研发及应用落地，致力为客户打造智能化与一体化解决方案。远传科技的知识图谱产品作为产业大脑解决方案架构中的一大模块，具备知识工程能力，向上可构建问答检索、智能推荐、预测研判等应用，目前主要服务于金融、农业、政务三大行业。在政务行业，疫情应急指挥是远传科技近期开发的知识图谱场景应用。远传科技的知识图谱产品通过大数据局与企业端获取健康码、核酸检测等数据，构建确诊病例的行动轨迹图谱，划分密接人群与次密接人群，帮助政府相关部门明确各人群分类，从而采取防控措施。基于行动轨迹图谱，远传科技结合智能交互模块的端侧机器人触达高风险人群与有疫情咨询需求的民众，实时监测高风险人群的健康情况，自动解答健康码异常等问题，并实现数据回传与知识图谱及时更新。

远传科技知识图谱产品及疫情应急指挥知识图谱案例介绍

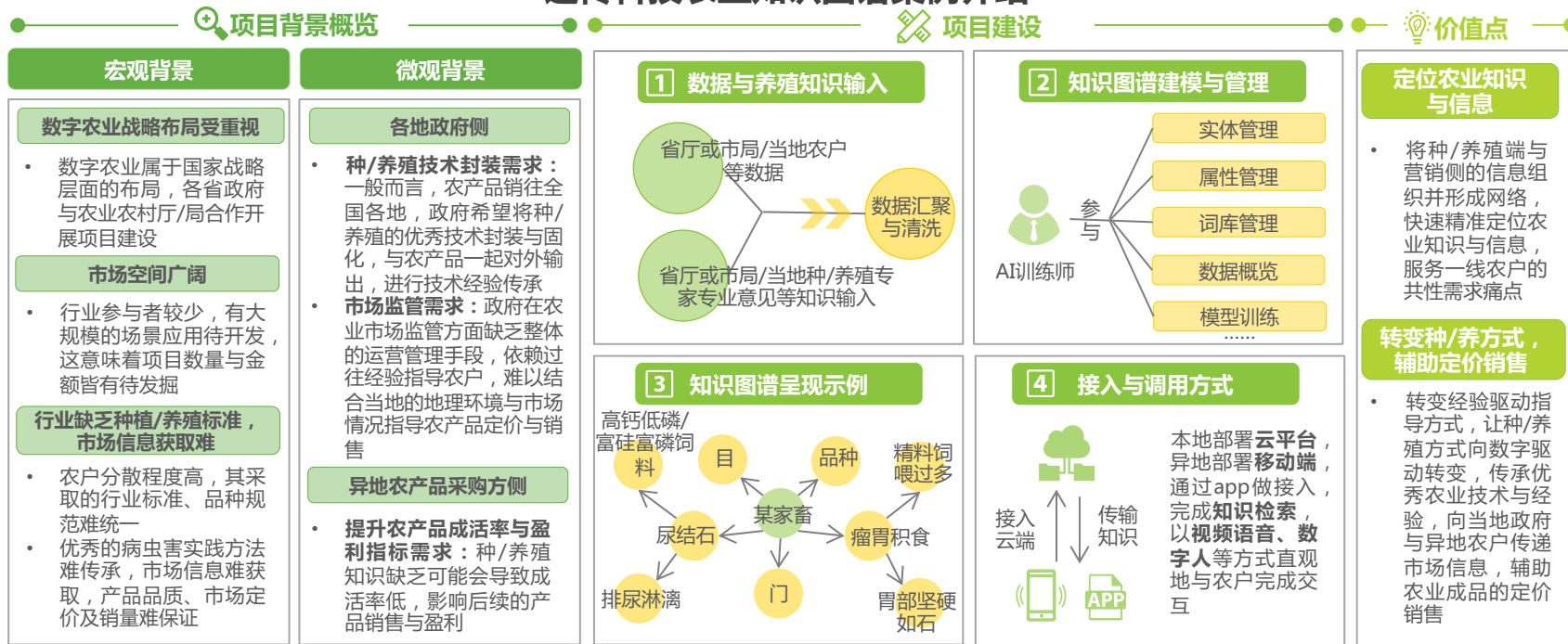


来源：艾瑞咨询研究院自主研究绘制。

封装种/养殖技术经验，传递市场信息，辅助种/养殖与营销

农业知识图谱是远传科技的典型案例应用。从宏观层面看，农业整体面临数字化转型，农业知识图谱市场空间待开拓，且行业整体缺乏种植/养殖标准，市场信息获取难；从微观层面看，政府有养殖技术封装与市场监管需求，异地农产品采购方则有提升农作物/家禽/家畜的成活率与农产品盈利指标的需求。在此背景下，远传科技通过联合省/市农村农业厅/局与当地农户，获取农产品的种/养殖数据与专业知识（如养殖常见疾病症状与病因等），构建农业知识图谱模型，协助当地农户进行优秀技术与经验的封装，帮助异地农户提升存活率与产品品质。同时，知识图谱中的市场信息可辅助当地政府与异地农户进行成品的定价与销售，有效改善市场信息不透明、不对称所导致的营销困局。

远传科技农业知识图谱案例介绍

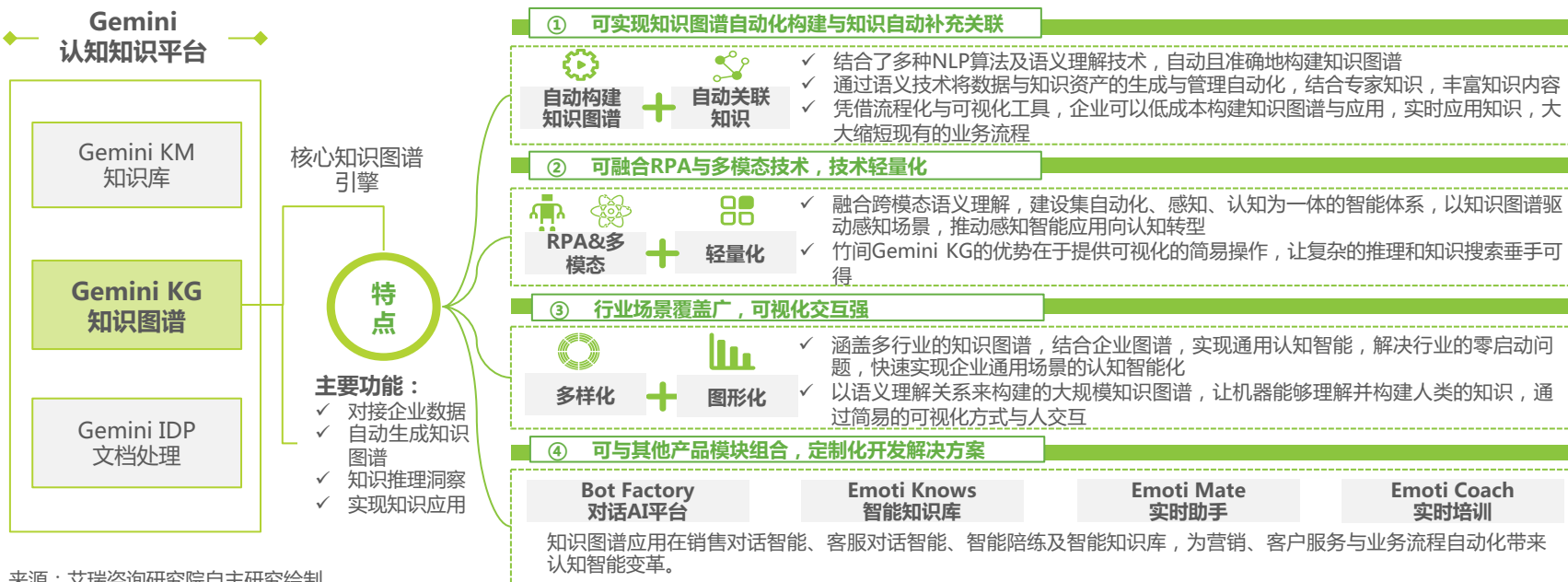


来源：艾瑞咨询研究院自主研究绘制。

技术轻量化，可实现知识图谱自动化构建

竹间智能以知识图谱与自然语言处理技术为基础，将认知能力整合到企业业务中，帮助企业实现智能化转型。**Gemini KG**是竹间智能**Gemini认知知识平台**中的核心知识图谱引擎，作为竹间知识图谱平台的通用知识引擎，可对接企业数据并自动生成知识图谱，创建与管理知识图谱的Schema，无代码构建行业的知识图谱应用。在Gemini KG的四大特点中，技术轻量化是一大差异化亮点。不同于由下至上的重量级解决方案，Gemini KG将认知知识算法和语义理解模型引擎化与模块化，用户使用自然语言即可搜索与探索知识内容，基于语义关系进行洞察和推理，帮助用户进行决策及判断。此外，Gemini KG还可实现知识图谱自动化构建与知识自动关系关联，搭配竹间的Bot Factory对话AI平台、Emoti Knows智能知识库，形成企业的知识工厂，将对话智能、知识智能、智能文档处理(IDP) 结合形成企业的知识认知与管理平台解决方案。

Gemini KG功能及特点

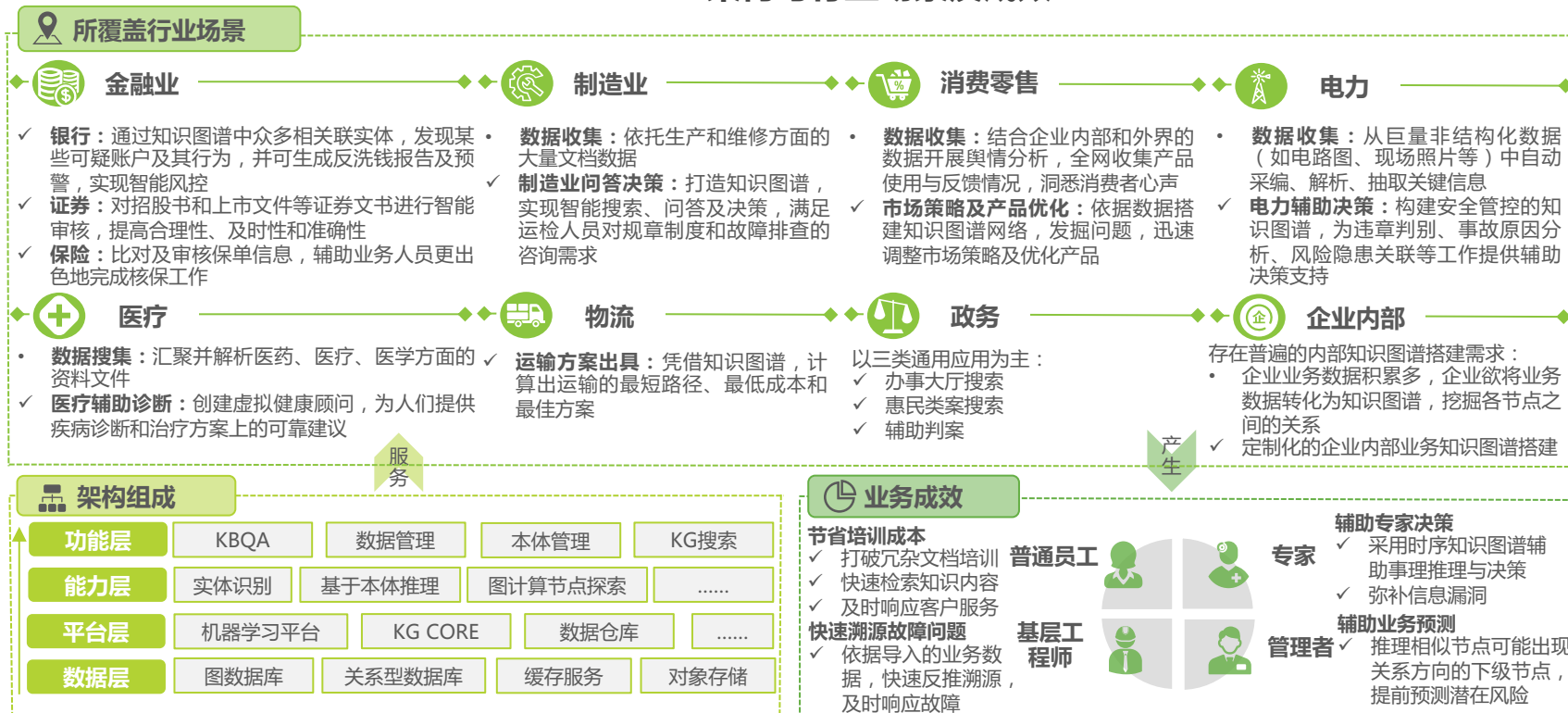


来源：艾瑞咨询研究院自主研究绘制。

行业场景广泛覆盖，助力业务端提质增效

未来的数字化变革将取决于知识图谱技术的应用，竹间Gemini KG涵盖了行业主要的知识图谱应用，完整的API Cloud与企业业务系统无缝结合，搭配底层的自然语言理解技术，形成全链路的完整应用与核心技术的产品解决方案，已服务于银行、证券、保险、制造、医药、消费零售、能源、政务等行业。Gemini KG的主要应用包括智能审核、智能搜索、故障分析与排查、风险洞察、根源分析、知识探索、推理决策及侦查分析。

Gemini KG架构与行业场景及成效



来源：艾瑞咨询研究院自主研究绘制。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

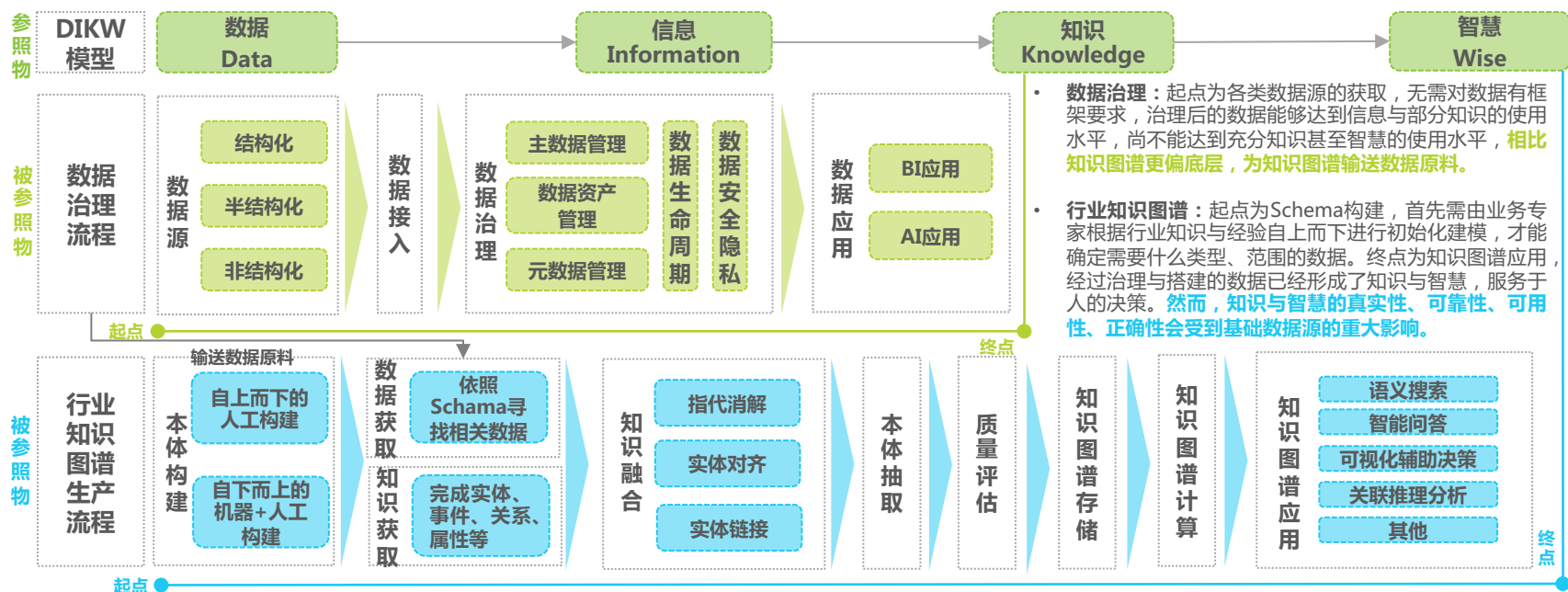
5

持续进行数据治理工程

从源头保证知识与智慧真实可靠、可用正确

数据治理为知识图谱输送数据源，是知识图谱构建的前置环节与基础性工程。完备良好的数据治理不仅能确保知识图谱在搭建过程中获取真实可靠的数据原料，而且能从源头上改善信息质量，提升知识的准确度，建立符合人类认知体系的数据资源池。但是，数据治理在知识图谱（尤其是行业知识图谱）建设卡点中是一个老生常谈的问题。知识图谱应用始终要围绕数据标签、数据清洗、数据归一、数据销毁等数据治理环节展开，应用开发人员往往需要在前期的数据治理工作中投入大量时间和人力，以确保数据源的真实性、可靠性、可用性、正确性。当前，数据标准不统一、数据噪声大、领域数据集缺失、数据可信度异常等数据治理难题依然困扰着知识图谱研发者，持续进行数据治理工程是业内参与者艰巨的使命与职责。

数据治理与知识图谱关系分析



来源：艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

储备培养行业专家与技术专家

缺乏深厚行业专家与技术复合型专家，需做好储备培养工作

目前知识图谱行业整体处于开发资源待完善的局面，行业与技术专家资源稀缺属于其中的一部分情况。一方面，缺少具备深厚行业经验的专家。由于行业知识图谱与行业的关联度高，开发人员需要迅速了解业务与客户需求，在行业专家的指导下完成Schema构建，若涉及到文本抽取工作还需要行业专家进行数据标注，而各行各业中的行业专家往往仅有极少数。对此，供给方企业需要锁定行业业务的强项领域、提前招募培养行业专家、进行内外协作，以完成行业专家储备。另一方面，缺少技术复合型专家。整个知识图谱应用生产流程不仅涉及知识图谱算法，生产流程的靠前环节还涉及到底层的图数据存储与数据治理、NLP文本抽取和语义转换，同时各环节都渗透着机器学习这一底层人工智能技术。这意味着整个生产流程需要多个技术领域的工程师协同合作，而对整套技术均有了解的技术专家数量稀缺。对此，供给方企业需要在项目中让技术实施人员沉淀复合型知识经验，让企业内部多方的技术专家进行错位交流，进行业务培训，以完成技术复合型专家的培养。

知识图谱专家资源紧缺建议



来源：中国电子技术标准化研究院《2019年知识图谱标准化白皮书》，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

研发国产化图数据库

特殊的图存储结构对底层存储技术提出升级需求

由于知识图谱是二维链接的图结构而非行或列的表结构，其需以图数据的形式描述并存储，该方式能直接反应知识图谱的内部结构，有利于知识查询，结合图计算算法进行知识的深度挖掘与推理。满足这一存储要求的数据库为近几年兴起的图数据库。相比于传统的关系型数据库，图数据库的数据模型以节点和边来体现，可大大缩短关联关系的查询执行时间，支持半结构化数据存储，展示多维度的关联关系。高效便捷的新技术往往意味着更高的研发门槛。从时间与归属方面看，全球第一款商用图数据库为2007年诞生的Neo4j，往后十年间的图数据库研发商基本分布于海外，而我国第一款商用图数据库为2017年上线的Galaxybase，比海外布局晚了近十年。从受欢迎度来看，Neo4j以59.4分一骑绝尘，占领着图数据库市场的高地。随着国内各行业知识图谱应用的加深，传统关系型数据库的不足逐渐显现，研发国产化底层图数据库成为了推进知识图谱应用的一大底层技术关键点，同时也从国家战略角度推动国产信创核心自主可控的步伐。

图数据库与关系型数据库对比

1

关联查询性能对比

深度	关系型数据库执行时间 (s)	Neo4j执行时间 (s)	返回记录条数 (条)
2	0.016	0.01 ★	~ 2500
3	30.267	0.168 ★	~ 110000
4	1543.505	1.359 ★	~ 600000
5	未完成	2.132 ★	~ 800000

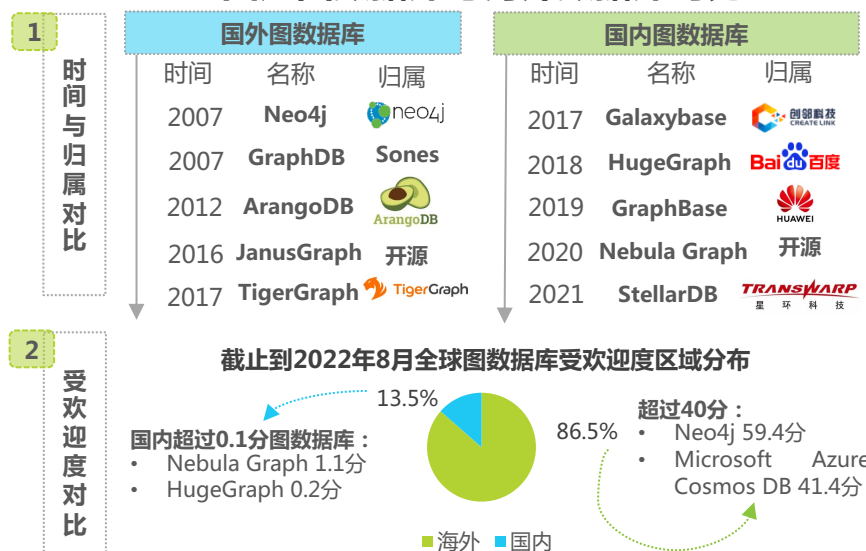
2

关联数据处理对比

分类	图数据库	关系型数据库
数据模型	图 ★	表
存储对象	半结构化数据 ★	结构化数据
2-3度关联查询	高效 ★	低效
6-10度关联查询	高效 ★	低效/不支持
事务性	支持 ★	支持

注释：上星表示效果更好。
来源：O'REILLY《图数据库》，中国信通院《图数据库白皮书》，艾瑞咨询研究院自主研究绘制。

国产图数据库与海外数据库对比



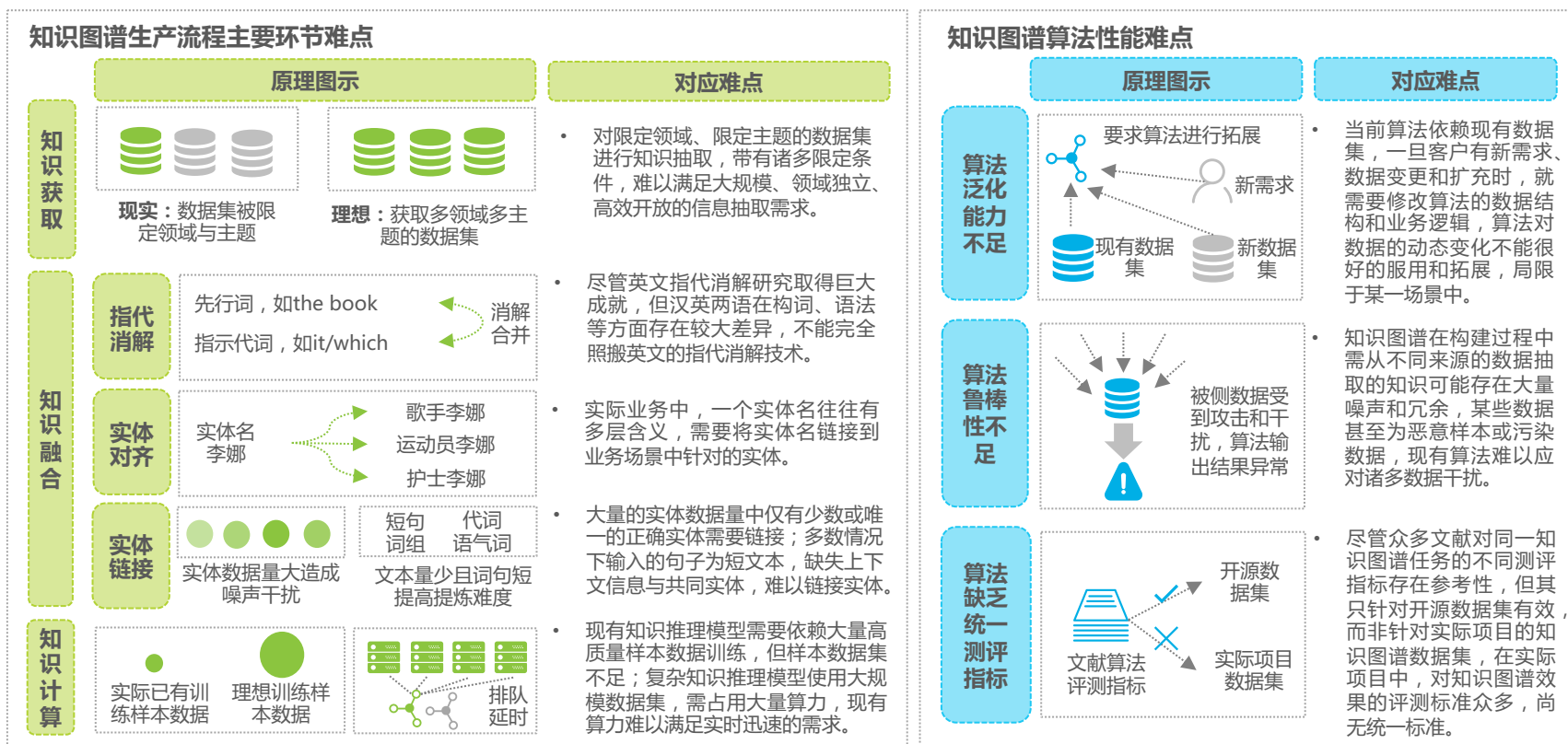
注释：图中图数据库产品与企业均为部分列举；部分有归属的图数据库亦可开源。
来源：DB-Engines，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

不断攻克算法难点

生产流程与算法性能各有难点，有赖于多方协同攻坚

在知识图谱的搭建过程中，仍然面临着各类算法难点，主要难点可归结为生产流程中的算法难点和算法性能上的难点。前者体现为知识获取受数据集限制、知识融合干扰因素较多、知识计算的数据集与算力不足等问题，而后者体现为算法泛化能力不足、鲁棒性不足、缺乏统一测评指标等问题。算法上的难点有赖于供需双方、学术界、政府持续攻坚，而非一方努力即可收获成功。

知识图谱当前算法难点



来源：中国电子技术标准化研究院《2019年知识图谱标准化白皮书》，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

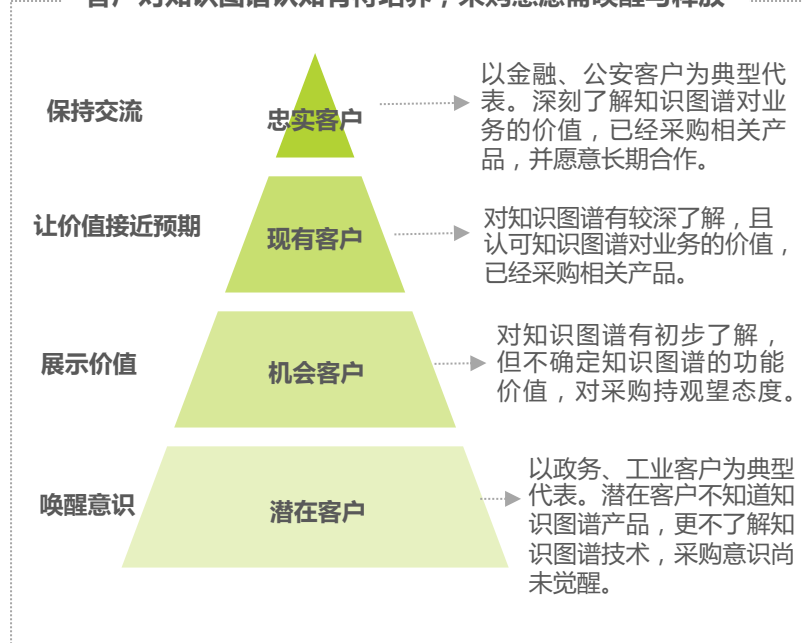
优化市场开拓策略

培养客户内生需求，探索高性价比产品形式

现阶段，知识图谱的市场开拓面临着两大难点：客户认知有待培养与技术产品化能力欠缺。从需求方角度看，广大潜在客户对知识图谱普遍缺乏认知，不知道知识图谱产品与技术的存在，采购意识尚未觉醒，而机会客户对知识图谱多持观望态度，不确定知识图谱能否为自身业务带来价值。从供给方角度看，厂商开发一整套知识图谱解决方案会涉及建模、求解、应用三方面的巨大成本，需要将成本以保证盈利且客户能够承担的思路分摊到各类费用中。各类费用汇总一般会形成售价较高的重量级解决方案，而重量级解决方案的市场受众有限，性价比高的轻量级产品更符合广大客户偏好。如何将高成本开发出的重量级技术解决方案以“小而美”且经济实惠的方式封装并出售，是厂商不得不解决的问题。

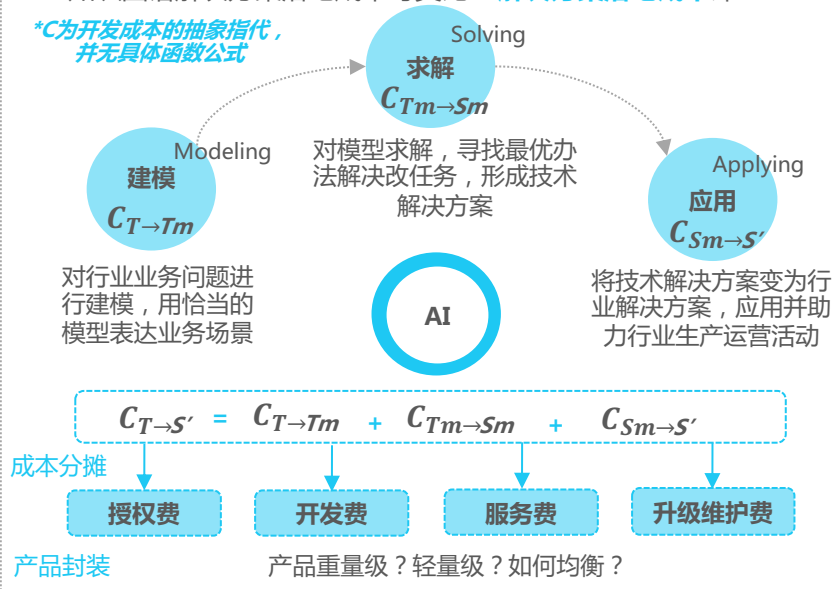
知识图谱市场开拓难点

客户对知识图谱认知有待培养，采购意愿需唤醒与释放



产品开发成本不知以何形式分摊，产品不知以何形式出售

知识图谱解决方案落地成本可类比AI解决方案落地成本即 $C_{T \rightarrow S'}$



来源：人工智能产业联盟、华为云、中国信通院《2022年知识计算白皮书》，艾瑞咨询研究院根据专家访谈与公开资料自主研究绘制。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

5

行业竞争格局走势

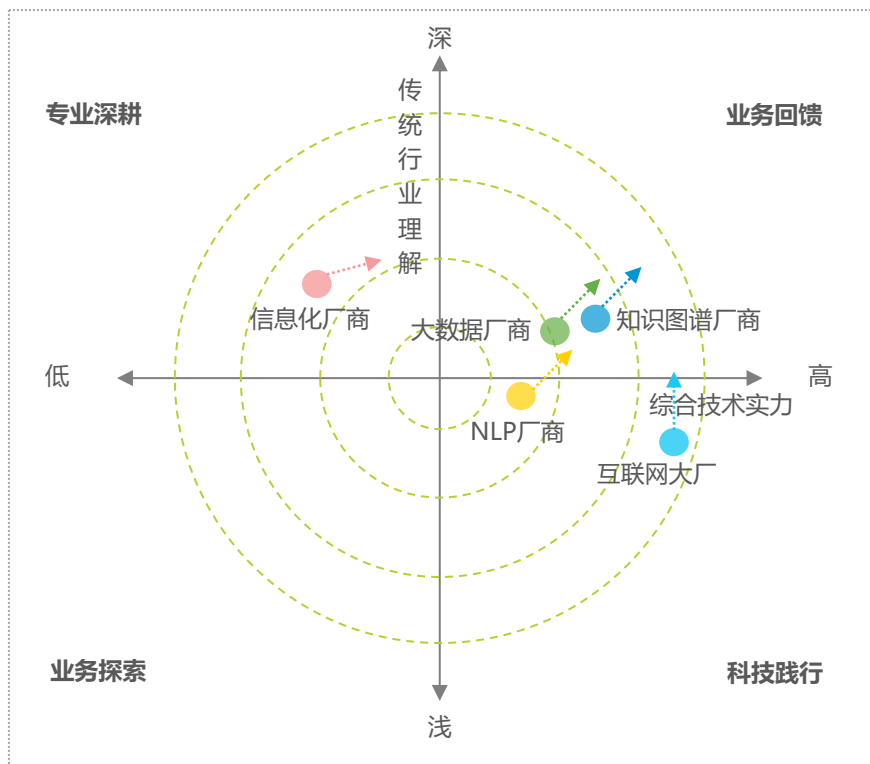
五方盘踞，以强化技术实力与深化行业认知为业务发展方向

我国当前的知识图谱市场主要为五方所盘踞：知识图谱厂商、大数据厂商、NLP 厂商、互联网大厂、信息化厂商。五类厂商优势各异，可将自身特殊优势作为“出牌技巧”，打出打好知识图谱这张业务牌。尽管各方的知识图谱业务开拓方式不一，但未来都以强化技术实力与深化行业认知为发展方向，以寻求良好的知识图谱业务回馈。

知识图谱业务实践雷达

五类厂商优势及知识图谱业务发展方向

- **知识图谱厂商**：具备行业知识图谱先发优势，对传统行业有较为领先的积累，行业认知较强，知识图谱技术资深，行业场景逻辑实力强。未来将朝更多行业拓展业务场景，深化行业理解并提高技术实力。
- **大数据厂商**：拥有数字化客群基础与数据积累，数据治理能力强，数据意识敏感，可在重量级数字化解决方案中捆绑知识图谱能力及产品出售。未来将在数字化业务基础上开发知识图谱行业业务场景，强化知识图谱技术。
- **NLP厂商**：具备NLP技术优势，可结合语音语义产品配套售卖知识图谱产品，知识图谱产品相对轻量化。未来将在NLP业务基础上开拓行业知识图谱业务，实现NLP与知识图谱技术一体化。
- **互联网大厂**：具备知识图谱底层技术实力，可将内部经验与资源作为开拓传统行业市场的利器。未来仍以自身的通用知识图谱业务为知识图谱业务主力，并以产业互联网为方向，深化行业认知与积累，开拓行业知识图谱业务。
- **信息化厂商**：拥有信息化客群基础与数据积累，行业认知较深，可迎合客户使用习惯直接叠加开发知识图谱产品。未来将在信息化存量业务的基础上，提升技术实力，开拓行业知识图谱业务。



注释：图示为定性判断；箭头指向为厂商未来业务趋势；综合技术实力包含数据治理、知识图谱建模与训练、NLP文本处理、图数据存储等关键的知识图谱能力。

来源：艾瑞咨询研究院自主研究绘制。

知识图谱应用场景迭代

行业应用场景边界拓宽，垂直应用场景被做深做透

前文已经提到，高价值且实用性强的垂直场景知识图谱是未来的业务发展方向，这属于知识图谱应用场景迭代发展的内容。在知识图谱行业发展过程中，知识图谱业务的应用场景可持续迭代场景广度与深度。一方面，各行业对知识图谱的认知与需求被唤醒，行业应用场景边界拓宽，知识图谱应用模型能力泛化，多种行业应用同时实现落地。另一方面，随着行业认知加深与技术进步，技术与业务的结合点愈发精准，行业场景颗粒度不断收敛细化。高价值且实用性强的垂直场景得到重视，循序渐进地被做深做透，为传统企业的业务带来显著的业务增效。

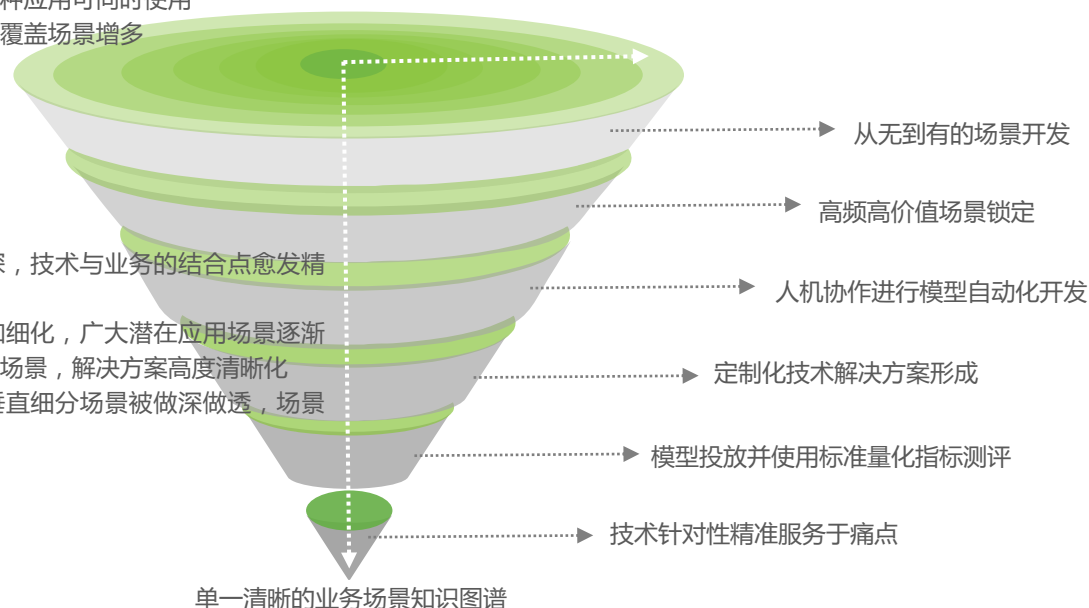
知识图谱应用场景迭代解析图

应用场景广度：

- 知识图谱的行业应用场景边界拓宽，不再局限于现有场景，潜在应用场景日益增加
- 应用场景复合叠加，多种应用可同时使用
- 应用模型能力泛化，可覆盖场景增多

应用场景深度：

- 行业应用场景深度加深，技术与业务的结合点愈发精准，技术逐渐走向深入
- 垂直场景的颗粒度更加细化，广大潜在应用场景逐渐收敛为清晰可见的可用场景，解决方案高度清晰化
- 高价值且实用性强的垂直细分场景被做深做透，场景业务得到显著优化



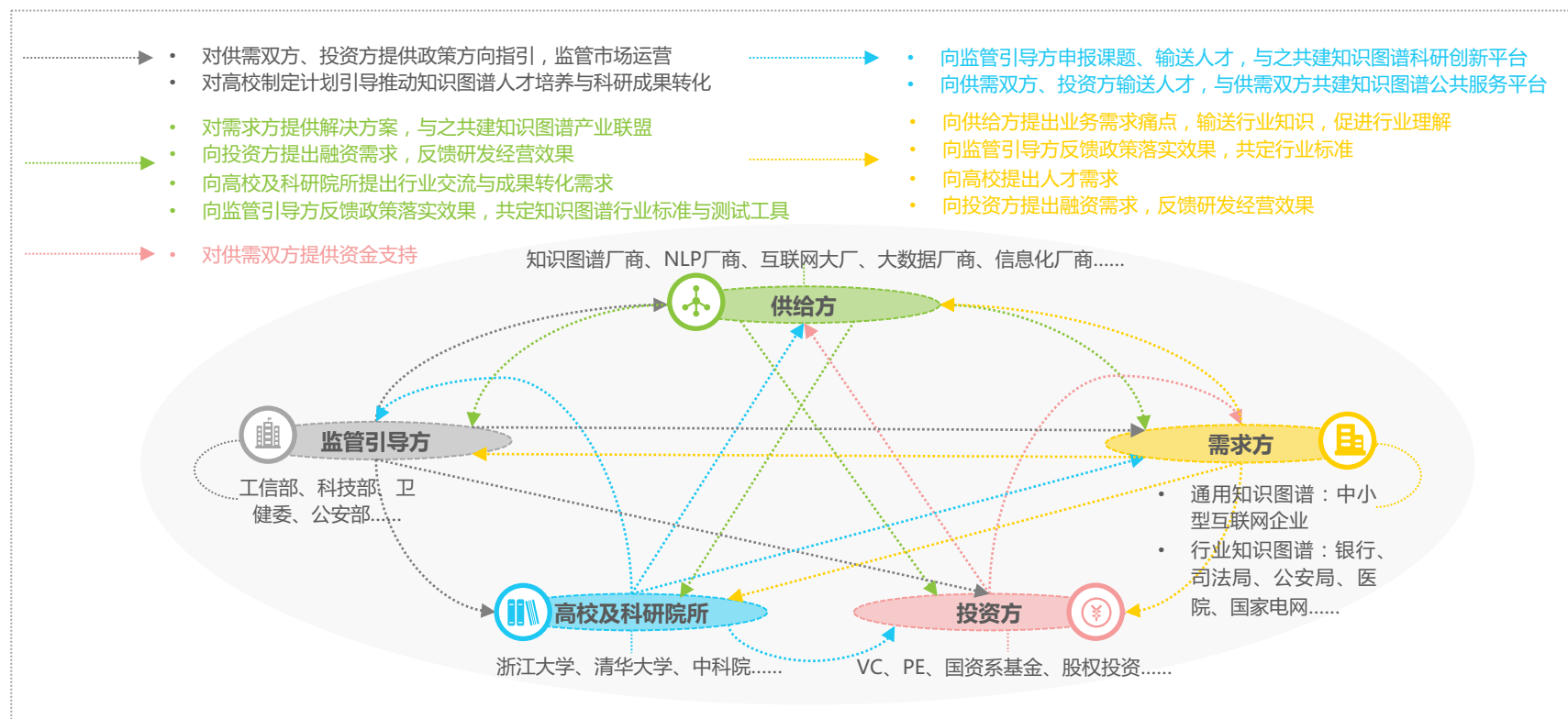
来源：艾瑞咨询研究院自主研究绘制。

知识图谱产业生态构建

产业各方融合共建，产业生态逐步成长壮大

知识图谱生态需由监管引导方、供给方、需求方、投资方、高校及科研院所融合共建，汇聚建设合力，促进产业生态成长壮大。各方相互提供资源支持，促成政策、人才、技术、资本、市场、商业的交互，探索与克服知识图谱行业技术与业务难点，共同收获产业发展价值，实现价值创造与价值分配的有机结合，形成共生共赢的合作体系，推动产业不断向前发展。

知识图谱生态交互示意图



来源：艾瑞咨询研究院自主研究绘制。

特别鸣谢

特别鸣谢以下企业对本报告的行业交流与研究支持：

明略科技



海致星图



同盾科技



远传科技



竹间智能



维智科技



海义知



智通云联



金桥信息



国双



艾瑞新经济产业研究解决方案



行业咨询

- 市 场 进 入 为企业提供市场进入机会扫描，可行性分析及路径规划
- 竞 争 策 略 为企业提供竞争策略制定，帮助企业构建长期竞争壁垒



投资研究

- IPO行业顾问 为企业提供上市招股书编撰及相关工作流程中的行业顾问服务
- 募 投 为企业提供融资、上市中的募投报告撰写及咨询服务
- 商业尽职调查 为投资机构提供拟投标的所在行业的基本面研究、标的项目的机会收益风险等方面的深度调查
- 投后战略咨询 为投资机构提供投后项目的跟踪评估，包括盈利能力、风险情况、行业竞对表现、未来战略等方向。协助投资机构为投后项目公司的长期经营增长提供咨询服务

关于艾瑞



艾瑞咨询是中国新经济与产业数字化洞察研究咨询服务领域的领导品牌，为客户提供专业的行业分析、数据洞察、市场研究、战略咨询及数字化解决方案，助力客户提升认知水平、盈利能力和综合竞争力。

自2002年成立至今，累计发布超过3000份行业研究报告，在互联网、新经济领域的研究覆盖能力处于行业领先水平。

如今，艾瑞咨询一直致力于通过科技与数据手段，并结合外部数据、客户反馈数据、内部运营数据等全域数据的收集与分析，提升客户的商业决策效率。并通过系统的数字产业、产业数据化研究及全面的供应商选择，帮助客户制定数字化战略以及落地数字化解决方案，提升客户运营效率。

未来，艾瑞咨询将持续深耕商业决策服务领域，致力于成为解决商业决策问题的顶级服务机构。

联系我们 Contact Us

 400 - 026 - 2099

 ask@iresearch.com.cn



企 业 微 信



微 信 公 众 号

法律声明

版权声明

本报告为艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。

为商业决策赋能

EMPOWER BUSINESS DECISIONS

