

DataGraphX：端到端式的 知识图谱交互问答系统

徐州星汉网络科技有限公司

CONTENTS

目录

01 | 摘要

02 | 功能模块解读

03 | 关于DataGraphX

04 | 高性能图智能引擎

05 | 专属知识模板

06 | 企业级性能

07 | 未来展望

01

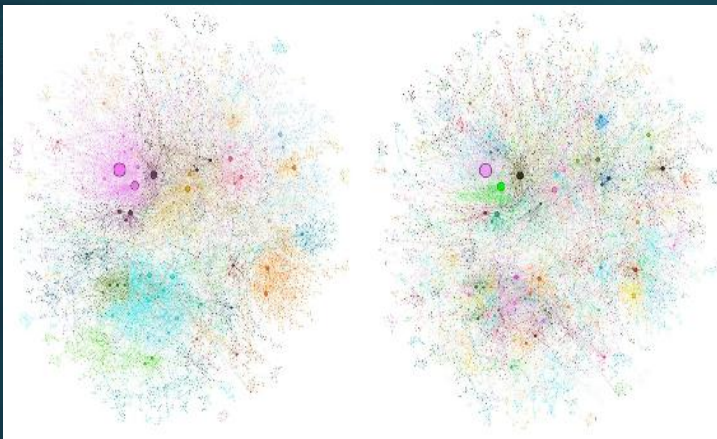
摘要

从“知识图谱”到“交互问答”——让知识更智能地“开口说话”



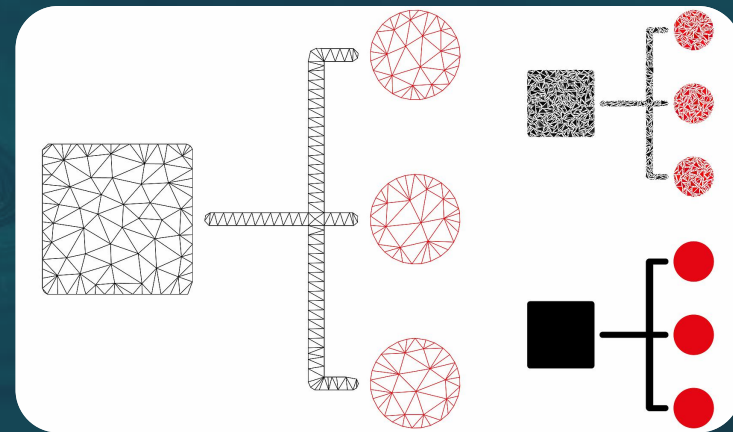
传统RAG的局限

传统RAG就像勤奋的图书馆员，能从海量文档中找到相关文本块总结答案，但它只能“点对点”检索，看不到书与书之间的引用和关联。例如询问“牛顿的理论对爱因斯坦有何影响”，它可能因知识在不同文档中而难以回答。



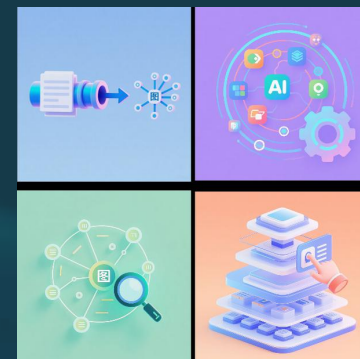
GraphRAG的应运而生

为解决传统RAG的问题，GraphRAG诞生。它将文档知识点构建知识图谱，具有全局视野，能俯瞰连接所有知识的“上帝地图”，还能进行关联推理，发现不同实体和文档间的深层联系，回答复杂问题。



DataGraphX的诞生意义

许多GraphRAG实现构建的图谱临时且有噪声，推理能力受限。DataGraphX基于GraphRAG理念，对各环节深度重构与增强，构建高质量、高效率、高智能的“活体知识大脑”，能为用户生成逻辑严谨、可追溯且有深层洞察的答案。



01

全自动知识管道

实现从原始文档到结构化知识图谱的“一键式”智能转化，极大降低应用门槛。例如，以往构建知识图谱需大量人工处理，现在通过此技术可快速完成，如处理一份复杂的行业报告，原本可能需要数天，现在仅需数小时。

02

异构AI协同引擎

独创性地融合多种顶级AI模型，为不同任务匹配最优算力。在保证极致精准的同时，实现卓越的成本效益。比如在处理大规模图像识别和文本分析任务时，在确保效果的同时，通过智能资源调度大幅提升效率、优化成本结构。

03

深度图谱增强技术

系统不仅能理解文本字面意义，还能通过向量计算主动发现隐藏的“相似关系”。以电商数据为例，可挖掘出看似不相关商品间的潜在关联，为企业带来新的销售机会。

04

企业级架构与交互体验

采用可无限扩展的批量数据处理方案和前后端分离的现代架构。确保系统在大规模应用下高性能，用户探索知识网络时交互流畅。如在处理百万级用户并发访问时，系统响应时间仍能保持在毫秒级。

02

功能模块解读



传统问答的痛点

传统问答是“黑盒式”的信息检索，用户在黑色对话框提问，AI给出答案，但答案来源不明，逻辑是否可靠存疑，难以获得用户信任。



DataGraphX交互问答的优势

DataGraphX呈现“白盒式”的知识探索，其主界面“智能问答”和“图谱交互”并存，答案可溯源、逻辑可视化、知识可探索。

主流开源方案的局限

主流开源方案如Marker、Unstructured主要针对纯文本或简单布局，处理百页级报告需数分钟，且易丢失结构、信息错乱。

DataGraphX的优势

DataGraphX原生支持复杂PDF。可在数秒内完成复杂文档的结构化解析，支持图表、多栏、表格等异构布局。

为交互问答提供数据基座

DataGraphX的PDF秒级解析能力，能让最新行业报告瞬间成为交互系统中可查询、可探索、可视化的鲜活知识，为卓越的交互问答体验提供坚实、快速的数据基座。



ICON 01



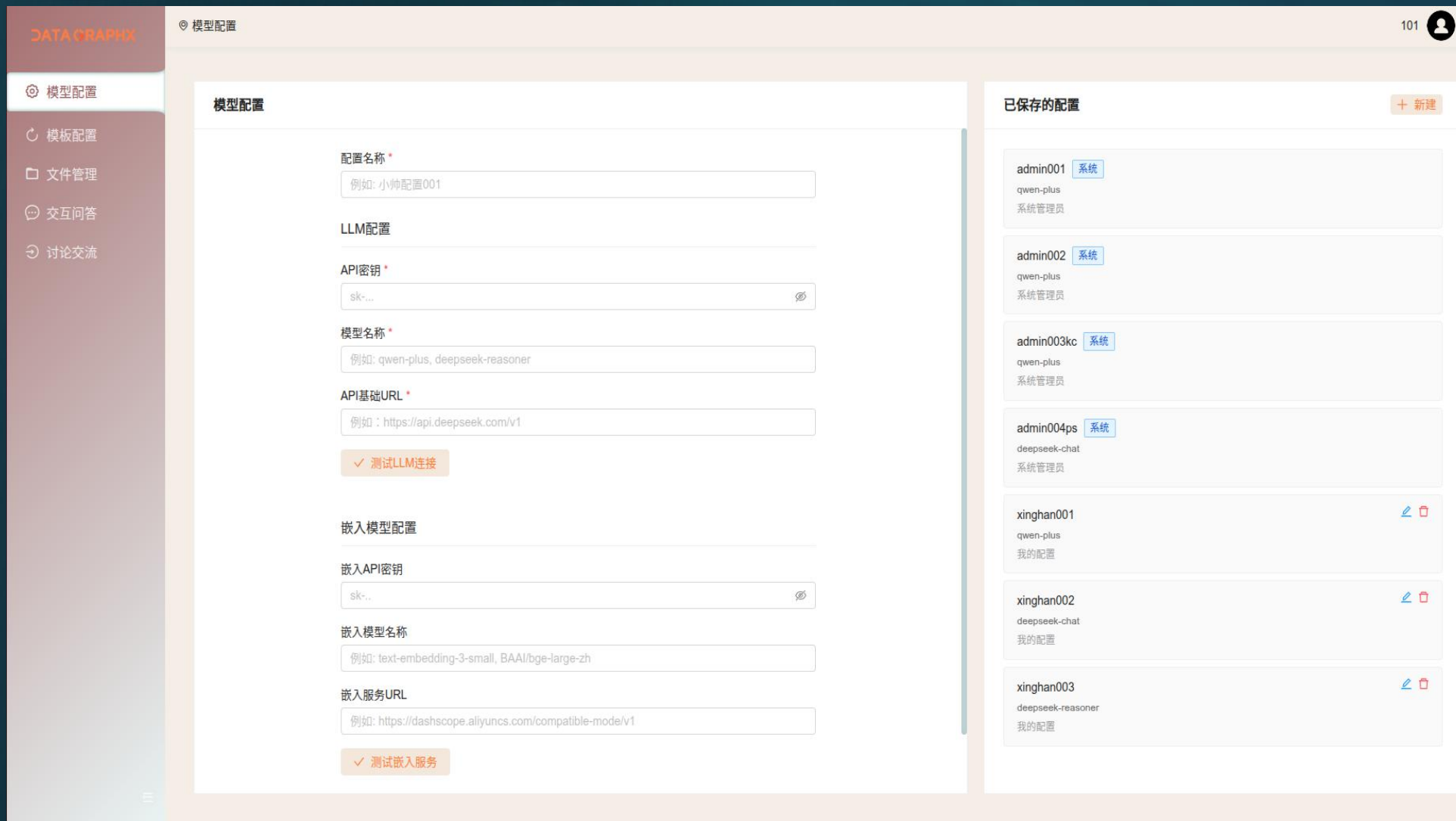
ICON 02



ICON 04



ICON 05



模型配置：

1.LLM配置 (大语言模型):
用于生成答案、进行推理和对话。

2.嵌入模型配置
(Embedding Model):
用于将文本向量化，是实现语义搜索和问答的关键。

3.已保存的配置：
右侧列表显示了所有已保存的配置方案，方便用户在不同项目中快速切换和复用，提高了工作效率。



模板配置：

1.模板的定义：

一个预设的实体和关系框架。

2.核心作用：

为非结构化的文本
(如PDF、Word文档)
提供一个结构化的目标。

3.模板类型：

支持共有模板、私有模板，满足团队协作
和个人使用的不同需求。

DATA GRAPHX

模型配置

模板配置

文件管理

交互问答

讨论交流

文件管理101

书名: 请输入书名搜索重置

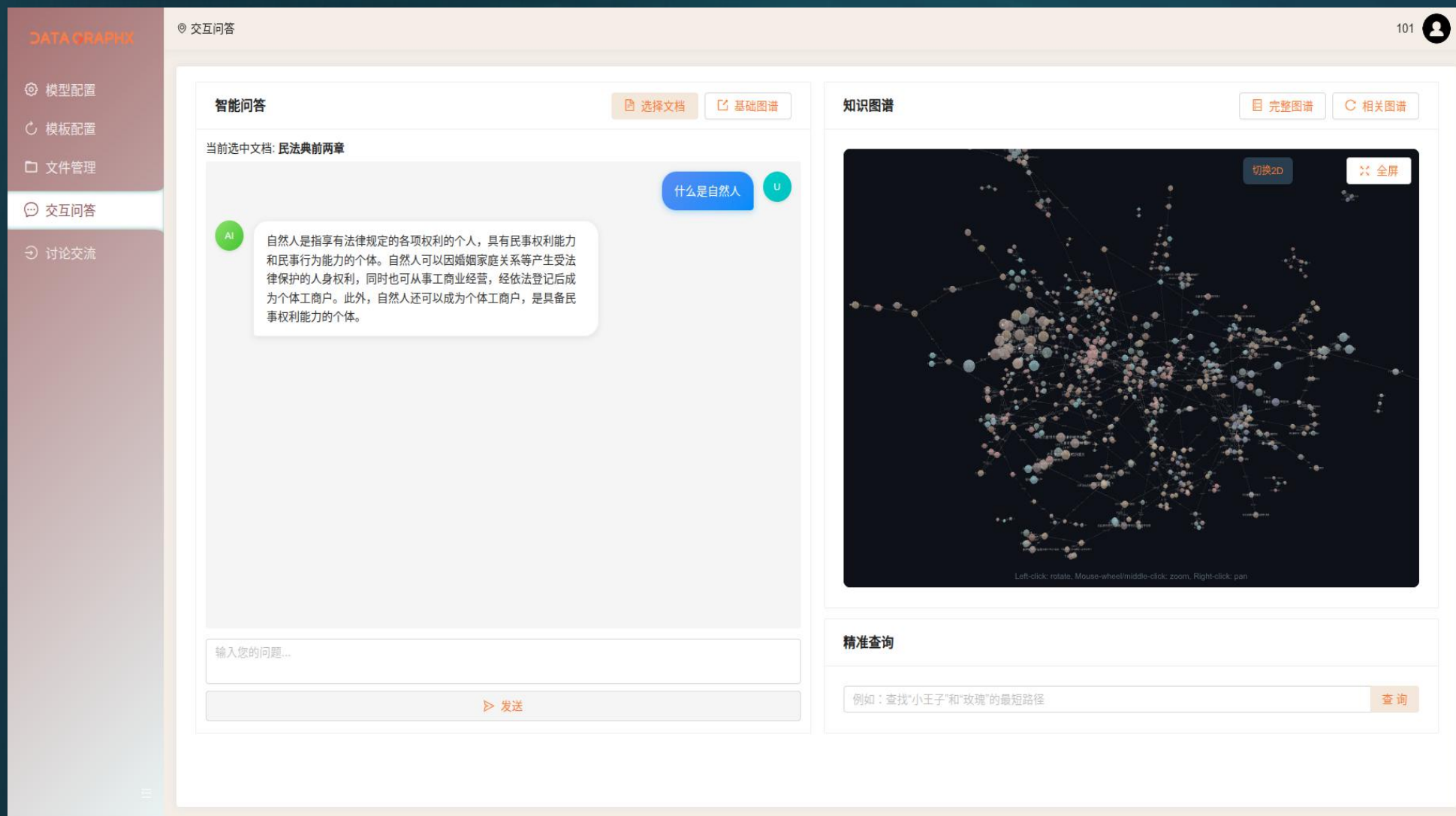
合并追加+ 新增

书名	图谱模板	模型配置	分块大小	重叠大小	分片状态	计算状态	创建时间	操作
☐ 民法典前两章	法律文书模板	xinghan001等	600	120	● 已完成	● 已完成	2025-07-10 09:49:20	↓ ↶ ⏮ ↷ ⏭
☐ 宁德时代	默认模板	xinghan001等	600	120	● 已完成	● 已完成	2025-07-09 15:40:00	↓ ↶ ⏮ ↷ ⏭

共 2 条数据 < 1 > 10 条/页

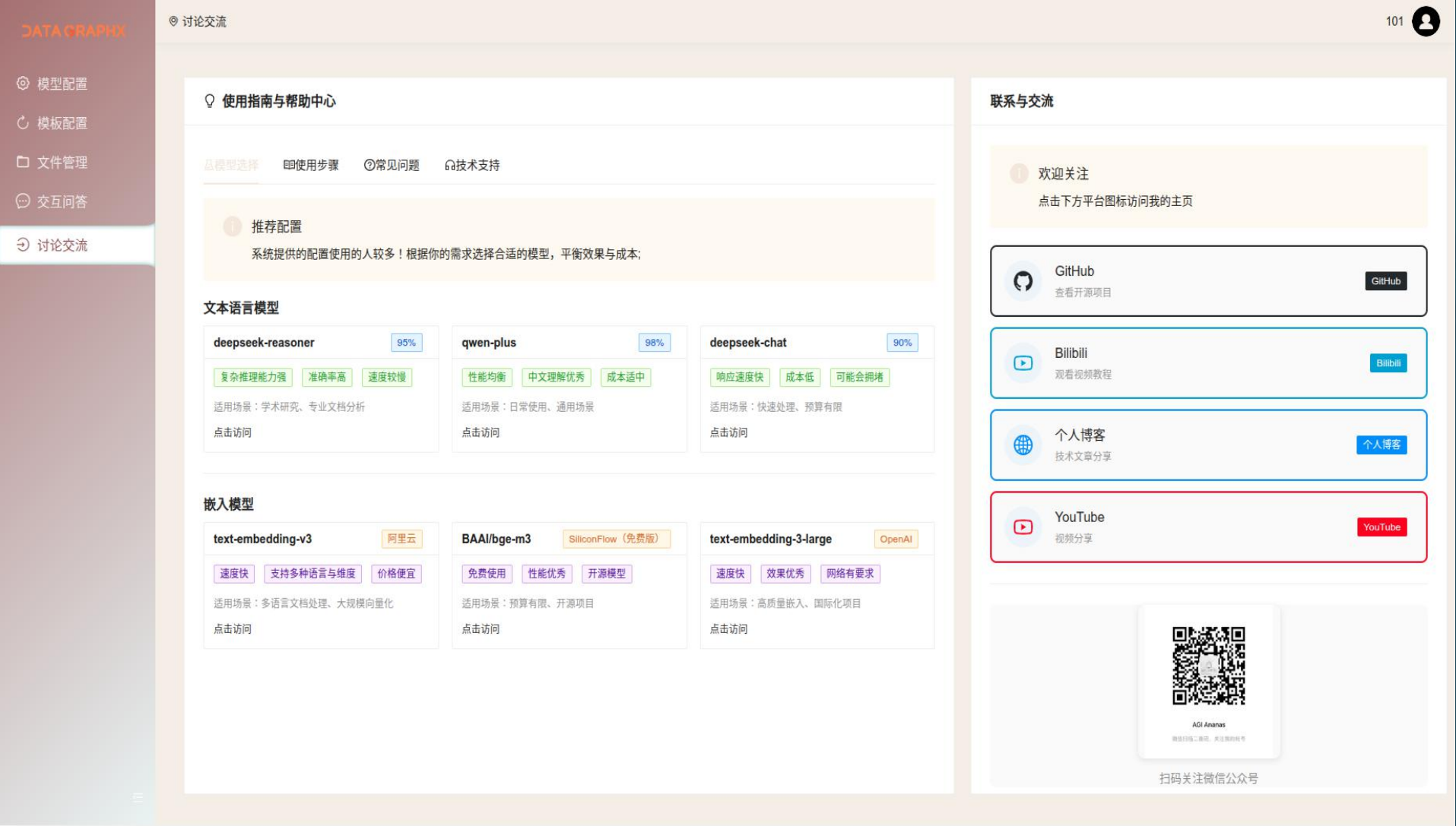
文件管理：

- 1.用户上传自己的文件。为每个文件关联一个“模型配置”（用哪个AI处理）和一个“配置模板”。
- 2.系统会根据所选配置，自动对文档进行分块(Chunking)、计算(Computing)等处理。
- 3.提供下载、重新计算、删除等管理功能，方便用户维护自己的知识库。



交互问答：

- 1. 智能问答 (左侧):**
选择一个已处理的文档。在对话框中输入自然语言问题。AI会基于文档内容，生成精准、可靠的答案，实现了RAG (检索增强生成) 的核心应用。
- 2. 知识图谱 (右侧):**
将文档中提取的实体和关系可视化，形成网络图。用户可以直观地探索知识点之间的联系，发现隐藏的模式和关联。支持查看完整图谱或与问题相关的图谱。
- 3. 精准查询:**
为用户提供结构化查询入口，可直接对知识图谱进行查询。



讨论交流：

1.使用指南与帮助中心：

提供官方文档、常见问题解答，帮助用户快速上手。

2.推荐配置：

针对不同场景（如复杂推理、中文优化），推荐最佳的文本语言模型和嵌入模型组合。帮助用户在性能、成本、效果之间做出最优选择。

3.联系与交流：

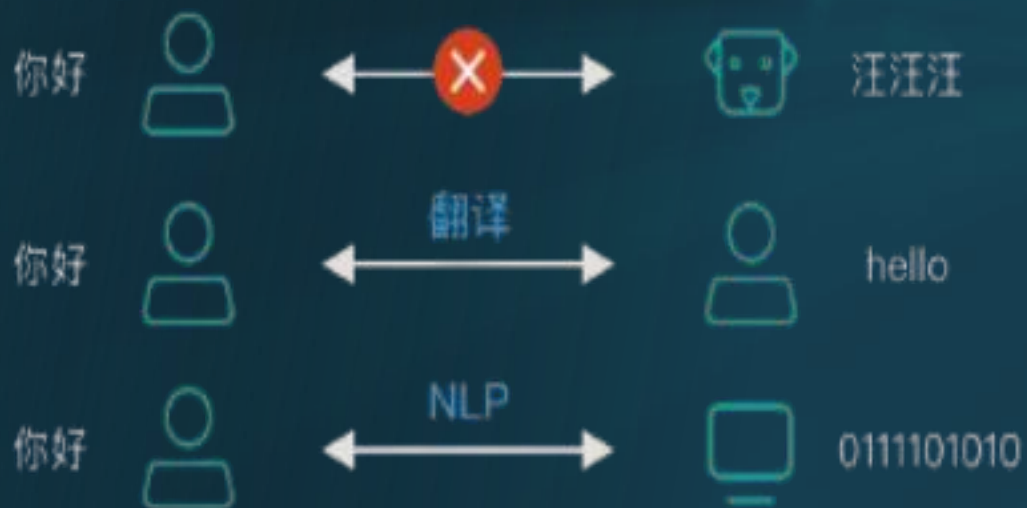
提供GitHub、Bilibili、YouTube以及公众号等社区入口。方便用户获取最新资讯、观看教程、提交反馈或参与社区讨论。

03

关于DataGraphX

我们的垂直领域是什么？

NLP就是人类和机器之间沟通的桥梁



主要属于自然语言处理（NLP）领域，同时涉及知识工程和人工智能应用两个重要分支，具体可以从一下角度细分：

1.核心领域：自然语言处理（NLP）

系统的“知识问答”功能依赖于NLP技术，包括文本理解、语义解析、问答匹配等，是NLP中“问答系统（Question Answering, QA）”方向的具体应用。

2.关键支撑：知识工程

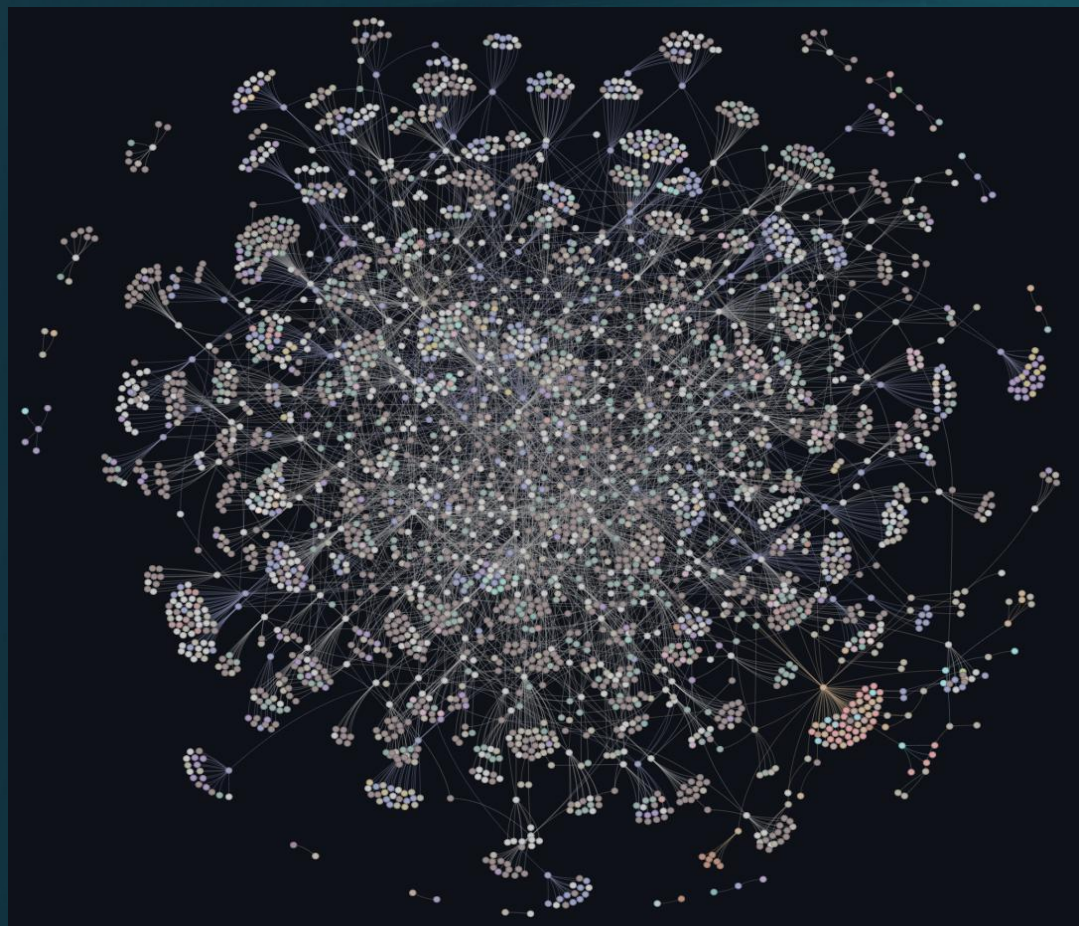
从文件中抽取信息并构建知识图谱，属于知识工程的范畴，涉及知识抽取、表示、存储等技术，知识图谱本身也是知识工程的重要成果。

3.综合应用：人工智能（AI）系统

整体系统通过整合RAG（提升问答的准确性和依据性）与知识图谱（提供结构化知识展示和关联推理），形成了一个综合性的AI应用，可归类到“智能问答系统”或“知识增强型AI系统”领域。

简单来说，它是NLP、知识工程交叉融合的AI应用系统，核心目标是通过结构化与非结构化知识的结合，提升问答的效果和可解释性。

我们的应用面向的人群用户是什么？



我们面向的用户群体主要包括：

- 专业领域从业者：如教师、医生、律师等
- 企业及组织中的决策者与分析师
- 科研人员
- 需要进行知识管理与沉淀的团队及个人

典型应用场景：教育、医疗、法律的知识智能化转型

教育场景【教学评估智能化】

支持从教辅资料、考试试卷中自动构建知识点图谱，识别学生群体中常见薄弱节点，辅助教师设计个性化教学路径与补强计划，实现因材施教。

医疗场景【科研辅助与病因溯源】

支持从临床报告、文献数据库中自动提取‘基因-通路-疾病’的因果链条，快速识别潜在靶点与交互作用，提升科研效率与诊疗精准度。

法律场景【证据逻辑图谱】

支持自动抽取案件材料中的人、物、行为等实体关系，构建证据链路图谱，识别证词矛盾、推理漏洞，辅助律师进行案件审查与逻辑校验。

政务知识管理

构建政策知识图谱，实现政策条文的上下位条款追溯与相似政策匹配，辅助政务人员快速获取政策依据与风险建议。

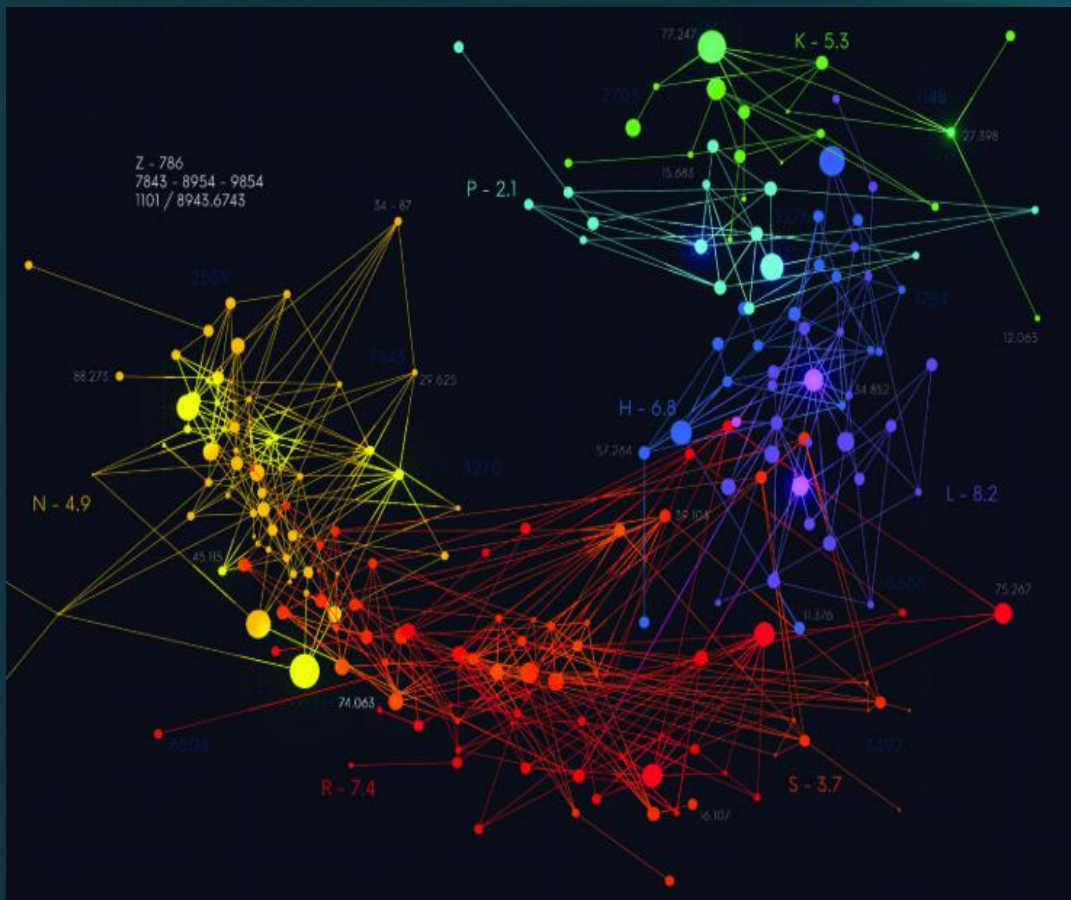
金融风控场景

从公开数据与内网文档中提取企业股权结构、历史变更、司法关系，构建企业图谱，实现穿透式风控与欺诈关联分析。

制造/能源场景

将设备手册、维修日志等非结构化数据转化为‘故障-原因-维修策略’图谱，支持预测性维护与故障辅助诊断。

我们有哪些优势？



1. 知识处理的深度与广度兼具

融合 RAG 与知识图谱技术，既突破传统 RAG “点对点” 检索的局限，具备全局视野和关联推理能力，能发现数据中隐藏的深层关系；又支持从非结构化文档（如复杂 PDF、报告）到结构化知识图谱的一键转化，实现“活体知识大脑”的动态演化，兼顾知识处理的效率与深度。

2. 专业适配性强，门槛极低

支持自定义模板功能，允许不同领域的从业者（如教师、医生、律师）根据自身专业逻辑定义信息提取规则和分析框架，无需编码即可构建专属知识体系，轻松跨越“AI 技术与业务逻辑”的鸿沟，让专业人士能直接掌控知识处理过程。

3. 交互透明可信，体验流畅

采用“白盒式”知识探索模式，问答结果可溯源、逻辑可视化，用户能清晰看到答案的来源和推理路径，解决传统黑盒问答的信任难题；同时具备秒级解析复杂 PDF（含多栏、图表、表格）的能力，搭配“智能问答 + 图谱交互”的并行界面，让知识探索更直观高效。

4. 企业级性能保障，适配规模化场景

针对多用户高并发场景，通过智能任务调度、全局缓存、离线预计算等技术，支持 100 + 用户同时进行复杂交互，95% 查询响应时间控制在 3 秒内，既保障流畅体验，又能通过资源优化降低企业级应用的成本（如 LLM 调用成本降低 70% 以上）。

5. 原生支持知识演化与协作

不仅能构建静态知识图谱，更能通过模板库迭代、联邦知识图谱扩展等功能，让知识体系随业务需求动态生长；同时支持共有 / 私有模板、团队协作式文件管理，方便企业或团队沉淀、复用行业智慧，驱动持续创新。

datagraphx是什么？与传统的rag问答的区别是什么？

- 1.DataGraphX 是一个端到端的、低代码的知识应用构建平台。
它的核心任务是，帮助用户将海量的、非结构化的文本（如PDF、文档、报告）自动化地转化为一个融合了智能问答 (RAG) 和 知识图谱 (Knowledge Graph) 的、可交互的AI应用。
- 2.与传统的rag问答的区别：

特性	传统 RAG	DataGraphX
知识处理	无结构化处理:仅对文本块做向量化。	深度结构化:在向量化之外，通过 模板配置 主动提取用户定义的实体和关系，构建了知识的骨架。
知识存储	单一的向量索引。	双重索引：同时拥有向量索引（用于语义搜索）和图索引（用于关系查询和分析）。
问答质量	容易受检索片段质量影响，对复杂关联问题回答能力弱。	更高质量的答案：不仅能检索相关文本，还能利用知识图谱的结构化信息进行推理，答案更精准、更不易产生幻觉。
能力边界	止于问答：最终产出是文本答案。	超越问答：除了提供文本答案，还提供一个可交互、可视化的知识图谱，让用户可以进行探索式分析，发现隐藏的关联。
核心差异	像一个“文档搜索助手”。	像一个“知识构建与分析引擎”。

04

高性能图智能引擎

DATA GRAPHX



我们基于自主图谱计算引擎，构建高效可扩展的图智能能力体系

精准的符号推理

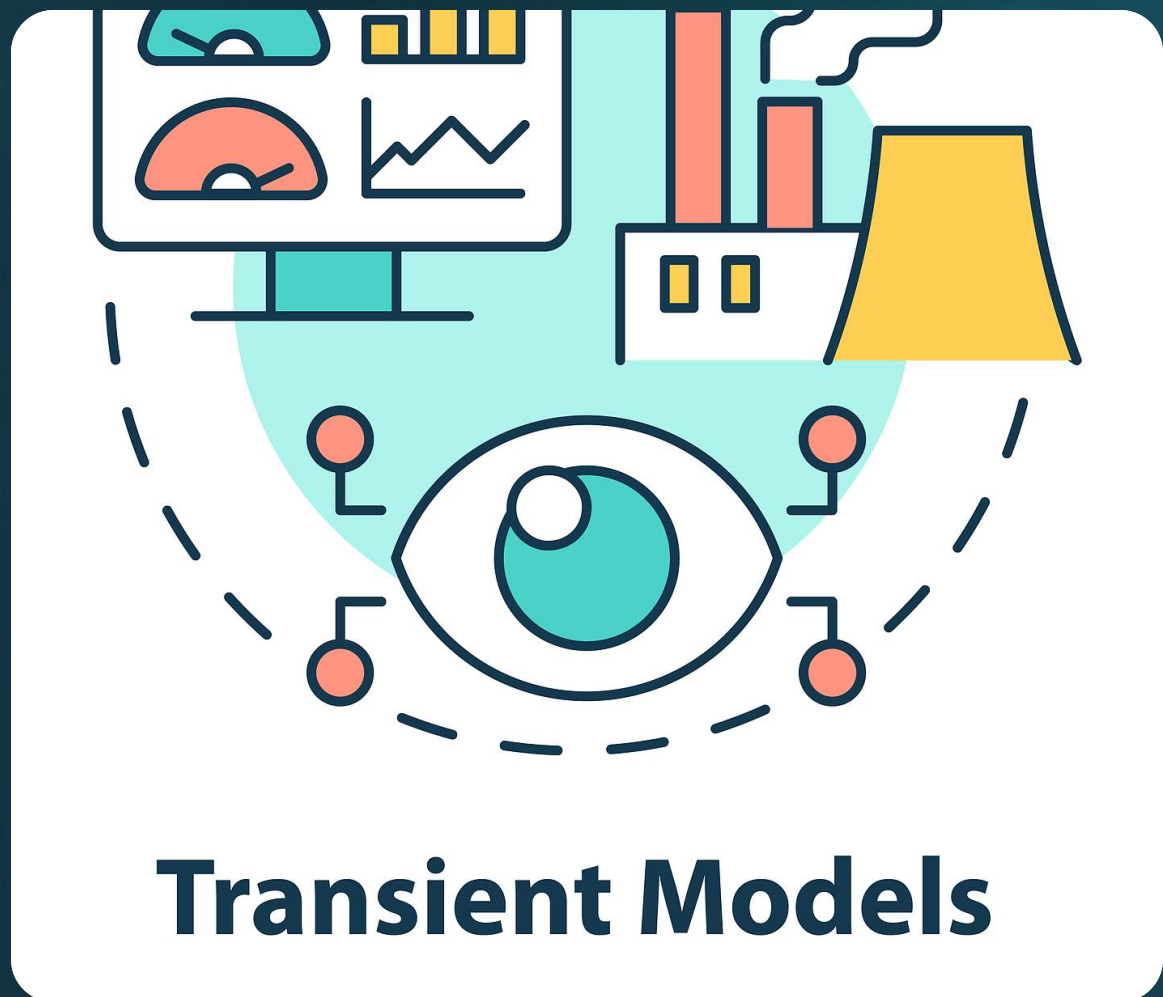
传统AI面对多跳复杂查询往往力不从心，如无法回答“谁是A的朋友的朋友，并且也认识B？”这类问题。而DataGraphX通过图谱引擎的深度路径规划能力，系统可实现任意深度的多跳查询，满足复杂因果溯源需求，像在图谱上进行精确的、毫秒级的路径规划和计算，对供应链溯源、金融风控穿透至关重要。

深度的社区计算

数据中隐藏着诸多模式，例如一群看似无关的用户可能是欺诈团伙。DataGraphX内置社区发现、聚类分析等图算法模块，可在百万级节点中高效运行智能聚类与关系识别任务，发现隐藏的“利益团伙”。

全面的图操作

知识需要不断演化和管理，系统支持图谱的动态管理与实时编辑，包括增删改查、批量处理与可视化操作等。用户可以像操作数据库一样，对图谱进行实时的增、删、改、查，无论是通过API批量更新，还是在界面上进行手动编辑，都能随心所欲地探索与管理图谱。



传统方案的局限

传统方案是原始数据经过LLM处理后得到模糊的答案，属于信息检索层面，难以提供精准、可解释、有深度的洞察。

DataGraphX方案的优势

DataGraphX 方案通过先进的图智能引擎处理原始数据，结合路径推理、模式识别与图谱演化等能力，为用户提供清晰可溯源、逻辑严密、洞察力强的知识输出。

对比带来的价值

通过与传统方案对比，凸显了DataGraphX方案在处理数据和提供洞察方面的显著优势，能帮助用户更好地理解 and 利用数据，做出更明智的决策。

05

专属知识模板



通用AI，如何理解“专业世界”的独特逻辑？

教学领域的独特需求

在教学领域，老师想分析哪些知识点是学生普遍的薄弱环节，以及这些薄弱环节之间是否存在前置依赖关系，核心诉求是知识路径发现。例如，老师想了解学生在学习浮力时遇到困难，是否是因为没掌握好压力和密度的概念，通用AI难以回答此类问题。

医疗领域的独特需求

医疗领域中，医生关注某个基因突变与哪些疾病表征高度相关，并且同时受到哪种环境因素的影响，核心诉求是多因素关联分析。如医生探究基因 - 蛋白 - 通路 - 疾病的复杂网络，通用AI无法满足这类专业需求。

法律领域的独特需求

法律领域里，律师需要梳理案件中的被告人与哪些证人存在间接的社会关系，以及他们的证词之间是否存在逻辑矛盾，核心诉求是隐藏关系挖掘与逻辑校验。像律师关注证据链的完整性和逻辑性，通用AI难以应对。

“智能鸿沟”与解决方案

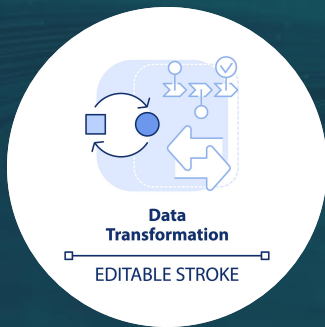
每个专业领域都有自己的“思维模型”，通用AI无法理解这些模型，只能停留在信息表面，这就是“智能鸿沟”。DataGraphX的自定义模板功能正是为跨越这道鸿沟而生。

DataGraphX模板引擎：将您的专业洞察， 转化为自动化分析流程



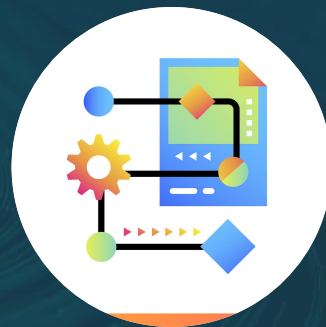
定义“科研模式”

无需编码，医生可根据自己的研究假设，像画草图一样定义想探索的生物学模式，支持用户图形化定义专业逻辑关系，实现可复用知识结构建模。



设定“分析任务”

医生为定义好的模式设定任务，将复杂的科研思路转化为简单的、可执行的指令。如设定“查找所有符合此三角关系的实例，并按药物的上市年份排序”的任务。



生成“智能模板”

将整个流程一键保存为模板卡片，如“靶向药 - 基因 - 疾病关联分析”模板。整个科室的医生、研究员都可以复用这个模板，在最新的数据上快速验证新想法。



赋能智慧教学

在教学领域，DataGraphX的自定义模板如知识点关联路径模板、学生学习行为模式模板，能为教师提供个性化教学路径规划的依据，实现因材施教。



赋能精准医疗

医疗领域中，药物靶点发现模板、患者分群与画像模板等，可加速新药研发和临床研究，辅助医生进行更精准的诊断。



赋能智慧司法

法律领域里，证据链校验模板、涉案人员关系网络模板等，能帮助律师和法官高效梳理复杂案情，发现隐藏线索。



沉淀行业智慧

通过DataGraphX，每个行业都能构建起自己专属的、可演化的智能模板库，沉淀专业知识体系，推动业务与智能融合发展。

06

企业级性能

从“个人跑车”到“公共高铁”：多用户环境下的性能挑战

01

并发请求风暴

当数十甚至上百位用户同时进行复杂的图谱查询和AI问答时，系统面临着巨大压力。高并发环境对响应速度和资源调度提出挑战，需精细化处理

02

资源争抢与隔离

一个用户的超复杂查询可能会占用所有计算资源，导致其他用户的操作“卡死”。比如某用户进行大规模的图谱分析，使得其他用户无法正常使用系统。

03

成本指数级增长

如果每个用户的每次操作都无节制地调用昂贵的LLM API，总成本将难以控制。以频繁查询的场景为例，大量的API调用会使企业在这方面的成本急剧上升。

DataGraphX企业级加速引擎：为高并发而生



智能任务调度器

所有用户请求首先进入智能任务调度器，它就像“交通警察”。会根据查询复杂度、用户优先级和系统当前负载，进行动态资源分配和排队，确保关键任务优先执行，防止单个“重查询”拖垮系统。



全局智能缓存

不仅有全局缓存，还有用户级缓存。不同用户查询相同问题的结果可以共享，极大提升了热门查询的响应速度。例如A同事刚查过的问题，B同事再查，答案瞬间可得，还为公司降低了LLM调用成本。



查询优化与索引

为所有用户的高频查询路径建立共享索引，确保数据库的每一次操作都处在最优状态，从容应对并发查询的压力，提高系统整体运行效率。



离线预计算

对于全公司共享的、计算量巨大的洞察，如全网用户社群划分，在夜间进行统一预计算。白天所有用户能瞬时享受到这些深度分析结果，无需消耗实时计算资源。



ICON 01



ICON 02



ICON 04



ICON 05

在DataGraphX，无需为协作而牺牲性能

高性能承诺

面向多用户并发场景设计，保障流畅体验并具备高可用性，保障了用户的使用体验。

高性价比承诺

通过多层缓存和共享机制，多用户环境下的人均API成本相比单用户独立调用降低70%以上，有效控制了企业成本。

高稳定性承诺

保障单个用户的“重查询”不影响其他用户的正常使用，并且系统能够7x24小时可靠运行，为企业提供稳定的服务。

07

未来展望

今天，我们重新定义了 企业知识交互

光速解析：突破知识获取瓶颈

DataGraphX实现了秒级文档解析，如110页的复杂PDF仅需5秒就能完成高质量结构化解析，极大解决了知识获取的效率难题。

自定义模板：融合业务与AI智慧

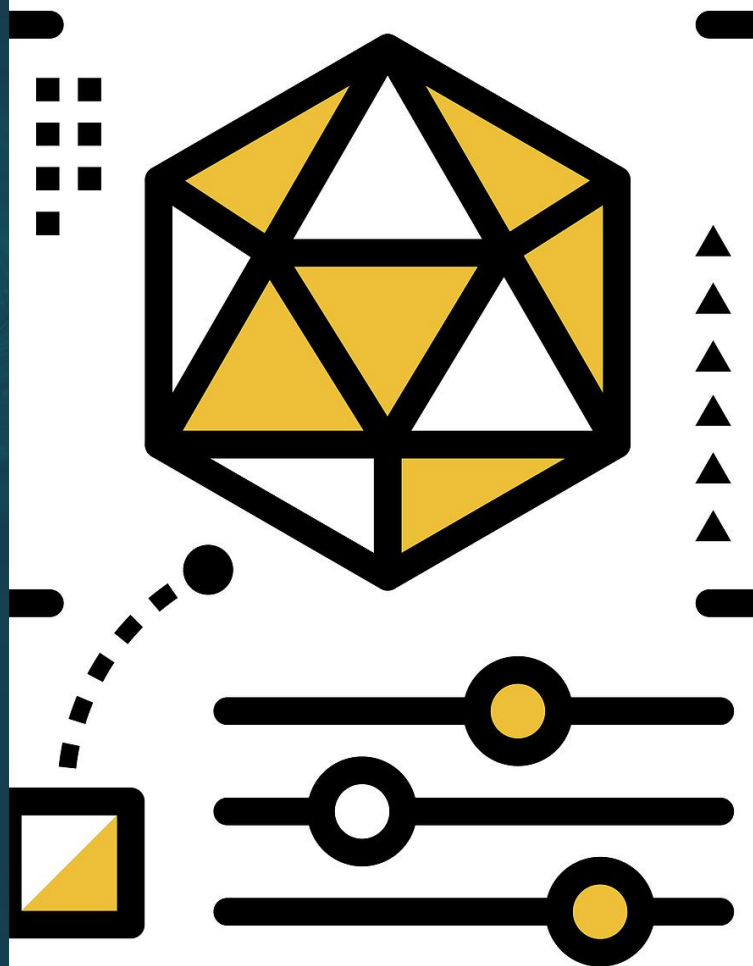
允许业务专家创建自定义模板，如医疗领域医生可创建“靶向药 - 基因 - 疾病关联分析”模板，实现AI与业务逻辑的深度融合。

活体图谱：释放深层洞察价值

通过深度图谱计算，DataGraphX能主动发现隐藏的“相似关系”，像从海量数据中挖掘出意想不到的商业洞察，释放知识网络的深层价值。

闪电引擎：保障多用户流畅体验

具备企业级高性能，加速引擎支持100+用户同时在线进行复杂交互，95%的查询响应时间在3秒以内，保证了多用户环境下的流畅体验。



DataGraphX的下一步： 从“平面图谱”到“沉浸式知识宇宙”

01

从“图查询语言”到“自然语言”：消除技术门槛

未来DataGraphX将上线自研图查询引擎，支持自然语言提问，用户只需用自然语言如“帮我找出所有与‘项目X’相关的风险传导路径，并用红色高亮显示”，系统就能自动完成复杂图谱查询和分析。

02

从“2D平面”到“3D/VR立体交互”：开启沉浸式探索

引入3D和VR/AR可视化引擎，用户可戴着VR/AR眼镜在悬浮的立体知识星团中，通过手势缩放、旋转、点选节点查看详细信息，获得前所未有的沉浸式洞察体验。

03

从“单一图谱”到“联邦知识图谱”：连接一切知识

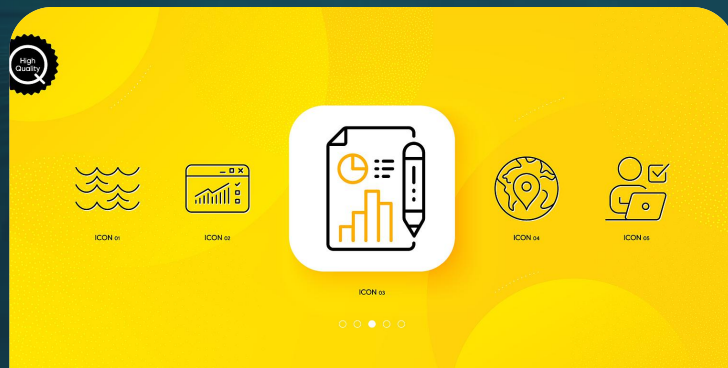
构建联邦查询能力，能跨越多个、异构的知识源，如将财务图谱、人力图谱、供应链图谱等通过虚拟“联邦层”连接，实现“连接企业一切知识”的目标。

DataGraphX：不止于工具，更是您通往未来智能的伙伴



愿景：构建下一代知识交互平台

DataGraphX致力于构建下一代沉浸式、自然语言驱动的知识交互平台，为企业知识交互带来全新变革。



近期路线图：稳步迈向未来

Q3 2024 推出结构化图查询语言自动生成能力 Beta 版，Q4 2024推出实验性3D可视化模块，2025年启动联邦知识图谱项目。



号召行动：共创知识无限可能

欢迎加入DataGraphX社区，试用产品，关注Github并Star，一同探索知识的无限可能。

DATA GRAPHX

谢谢