中国科学技术信息研究所
INSTITUTE OF SCIENTIFIC AND TECHNICAL INFORMATION OF CHINA
万方数据
WANFANG DATA

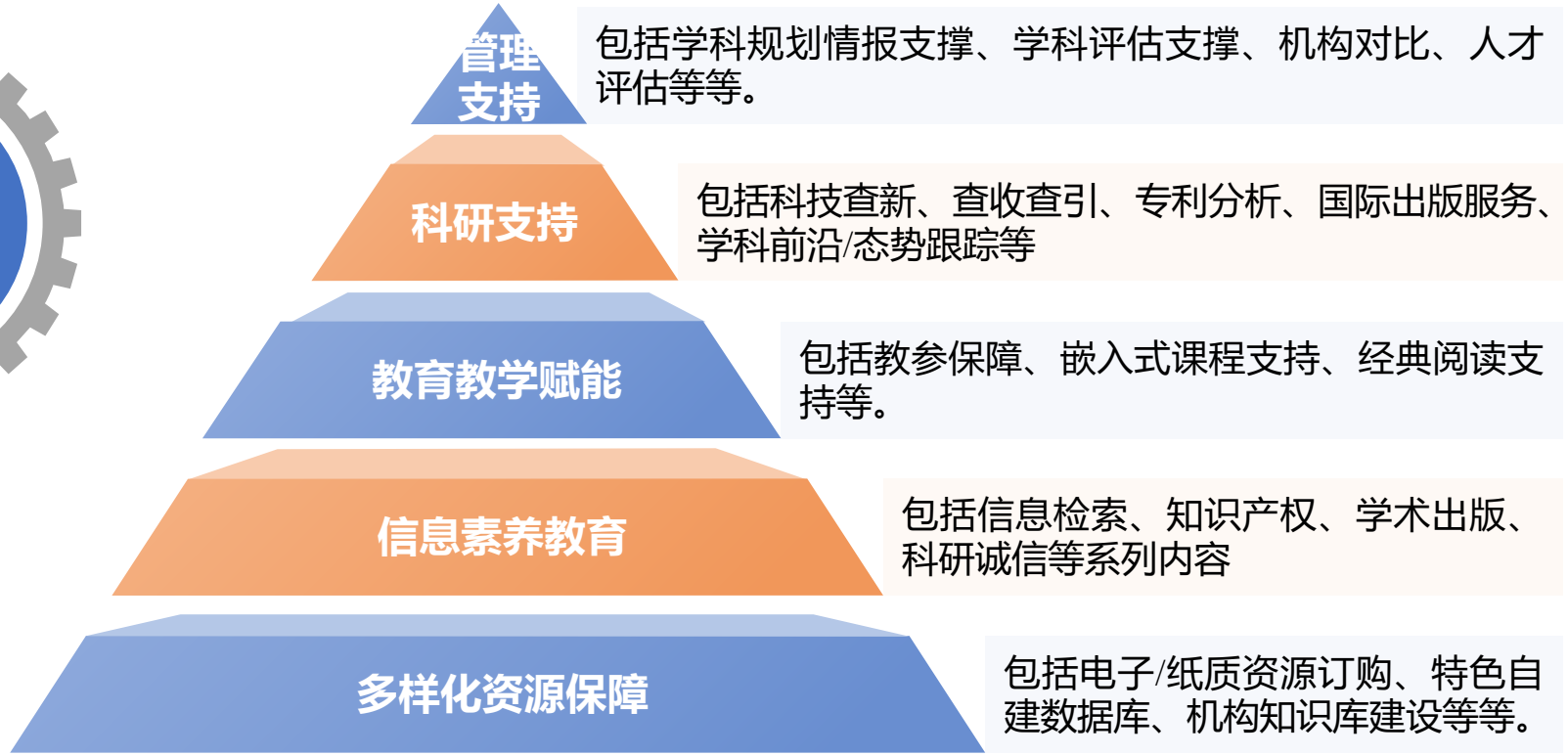
AI与数据融合的学科服务支撑体系探索与实践

万方数据知识产品开发部 刘建华

二零二五年六月二十七日

现有学科服务的主要构成

现有学科服务的主要构成



包括学科规划情报支撑、学科评估支撑、机构对比、人才评估等等。

包括科技查新、查收查引、专利分析、国际出版服务、学科前沿/态势跟踪等

包括教参保障、嵌入式课程支持、经典阅读支持等。

包括信息检索、知识产权、学术出版、科研诚信等系列内容

包括电子/纸质资源订购、特色自建数据库、机构知识库建设等等。

资源工具有使用门槛

分析有规范方法

解读有专门方式

新环境下学科服务面对的一些问题

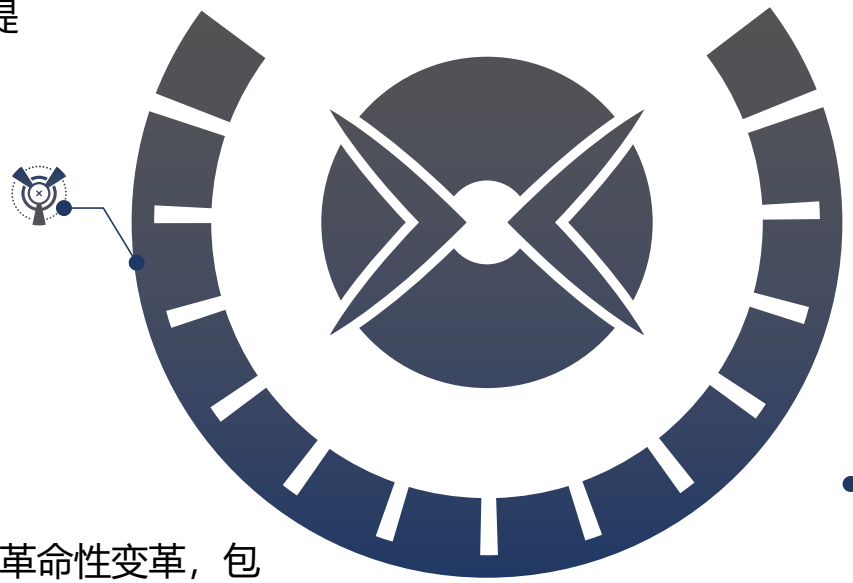
当前学科服务面临的挑战

客观的政策变动及经济原因

逐渐降低对科技查新、简单的查收查引的强制要求；科技评价体系的改革对传统评价提出挑战。普遍经费紧缩。

学术交流生态变化

正式学术交流与非正式学术交流逐渐趋向融合，对于资源组配、资源利用、科学评价等等都带来新的机遇与挑战。



AI+学科、跨学科发展成为常态

如何能在AI融入学科发展、跨学科发展中，发挥图书馆的作用，开拓新的服务领地。

AI推动各类关系变革

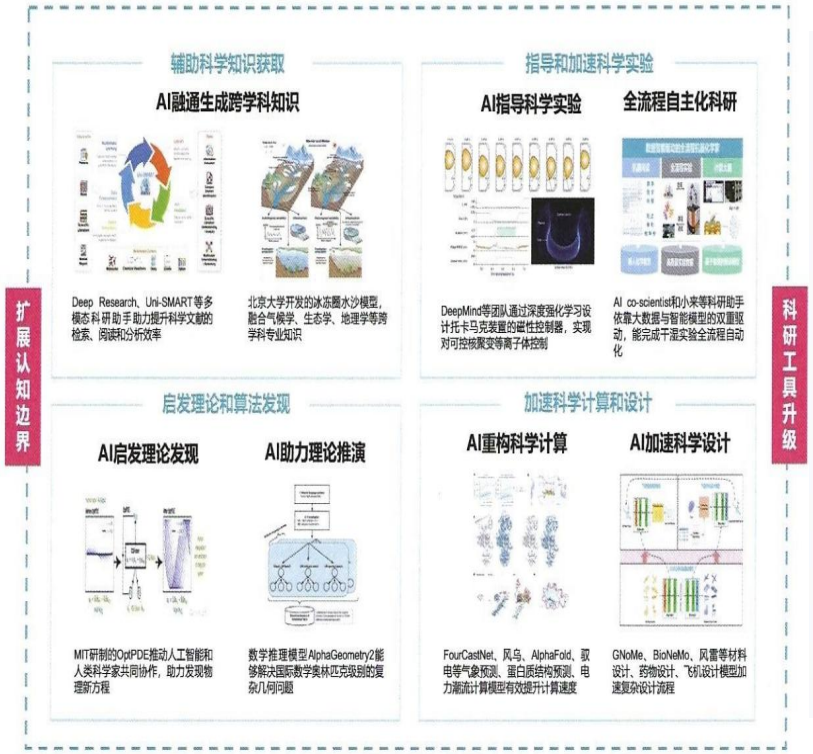
AI引发各类知识服务的生产关系革命性变革，包括但不限于知识学习长链条贯通、知识服务去中心化等。传统图书馆中介服务角色受到AI环境下端到端的服务冲击。

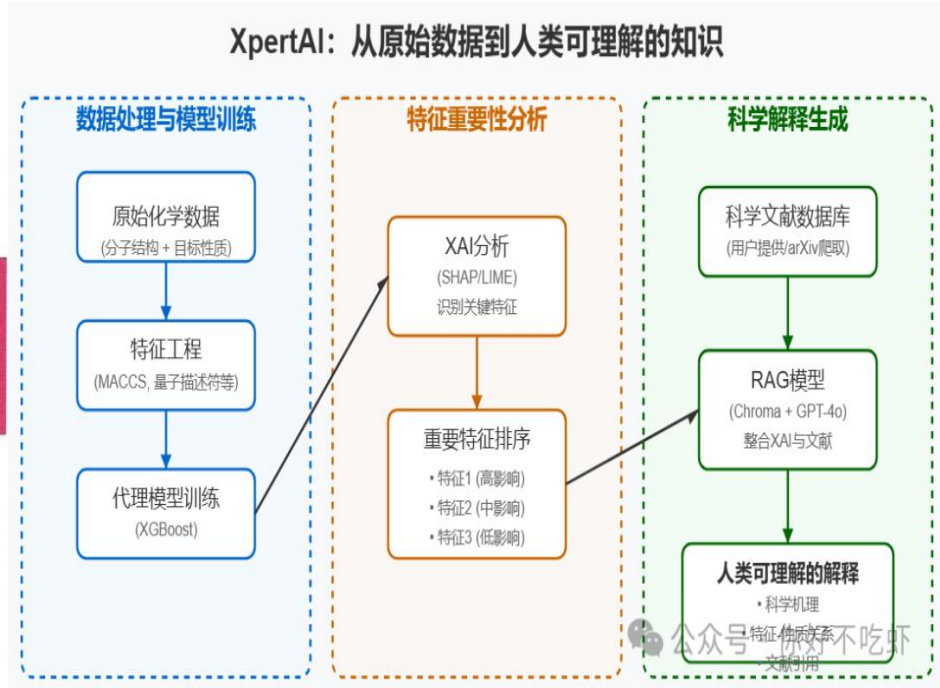
Ludo Waltman.Assessment reform and publishing reform need to go hand in hand.https://www.leidenmadtrics.nl/articles/assessment-reform-and-publishing-reform-need-to-go-hand-in-handEwen

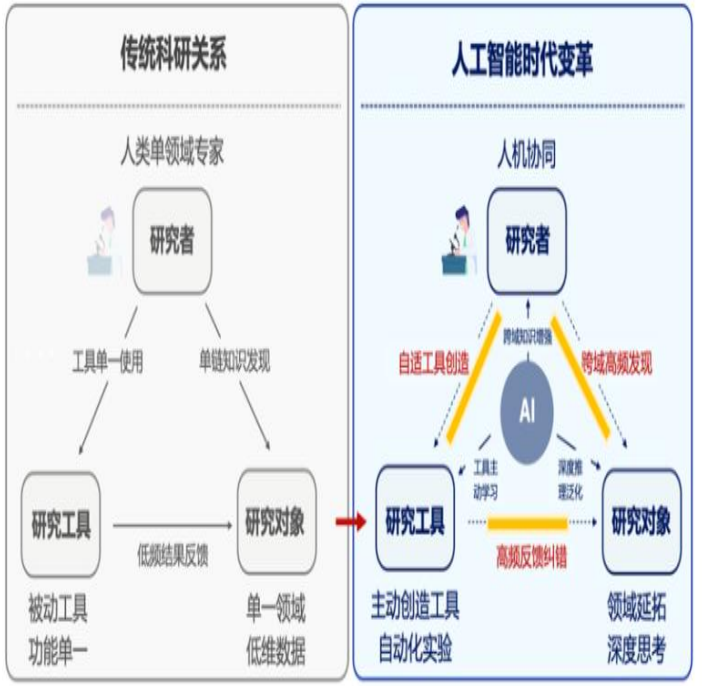
张晓林:Library Inside:AI 赋能图书馆新质生产力的一种基础模型

人工智能与学科、行业的结合，推动AI4S、AI4E等的同时，也在推动着生产关系的变化

- AI背景下，用户需要资源、数据、知识与其本人/其所用AI实现无缝衔接，这既是效率优先的要求，也是AI4S的学科研究中打开AI黑箱的重要保障，端到端的服务服务需求加大，向深度的数据服务、新的服务方式与服务内容提出了需求。







中国科学技术信息研究所.AI for Science创新图谱，2025年3月;
XpertAI如何用科学文献赋能机器学习解释化学结构-性质关系，打破黑盒困境.https://mp.weixin.qq.com/s/HLKpCgrv0VtlsLeDutRA7A
Gottweis, J., Weng, W.-H., Daryin, A., Tu, T, et al. Towards an AI co-scientist. https://arxiv.org/pdf/2502.18864

学术图书馆在积极地应对需求的变化

ACRL Advancing learning Transforming scholarship
Association of College & Research Libraries
A division of the American Library Association

美国大学与研究图书馆协会4月初发布
《学术图书馆环境扫描》

2025 Environmental Scan

ACRL Research Planning and Review Committee

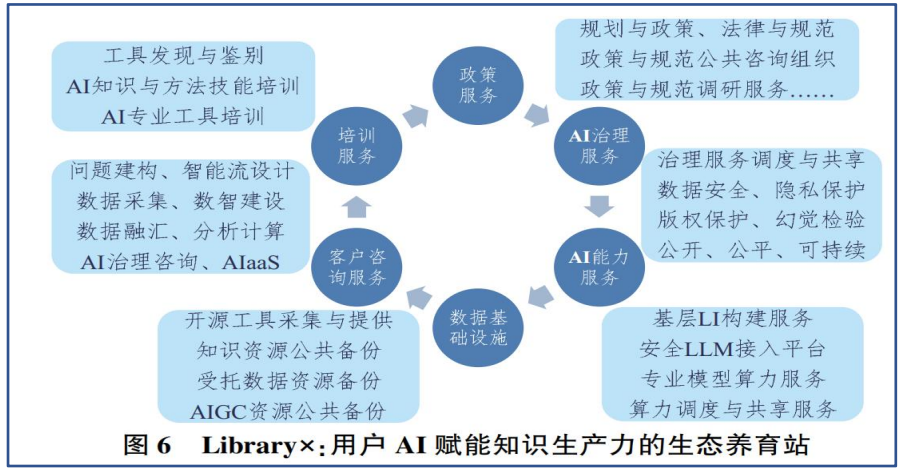


图 6 Library×:用户 AI 赋能知识生产力的生态养育站

IFLA

图书馆与人工智能”的切入点:旨在成为图书馆员评估人工智能伦理使用的辅助工具，列出了关键考量点及应用情境

IFLA
Entry point to
Libraries and AI

Internet Manifesto Annex I

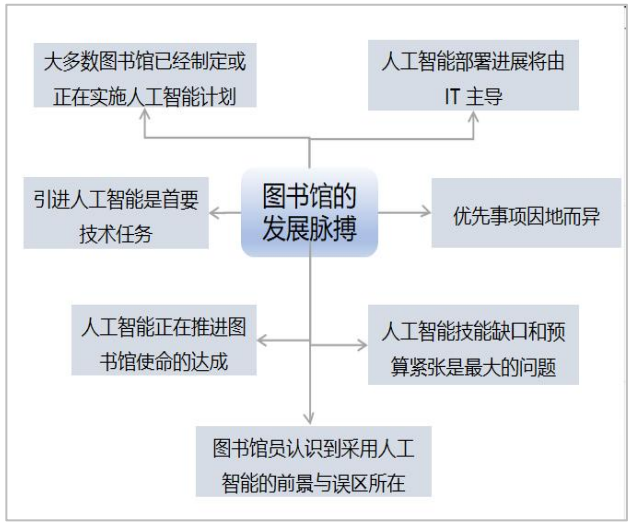


表 5 AI在高校图书馆的实践探索

类型	选项	人数
自动化管理与服务	智能空间	662
	机器人	352
	自动化典藏与定位	344
	自动化编目与分类	288
	管理决策辅助	140
	不了解	132
	未开展相关应用	124
	其他	17
教学与科研支持	智能搜索与学术助手	491
	读者行为分析与预测	485
	定制化知识资源推荐	410
	人工智能素养培养与学术社群支持	340
	不了解	232
	未开展相关应用	137
	其他	3
智能化辅助工具	数据分析与可视化工具	500
	多媒体内容提取与分析工具	385
	智能客服	381
	跨媒体内容生成与创作工具	347
	不了解	247
	未开展相关应用	154
	其他	4

2025 Environmental Scan.https://www.ala.org/sites/default/files/2025-04/2025_ACRL_Environmental_Scan.pdf；图书馆的发展脉搏.https://img02.ma.scrmtech.com/18476/1812/resource/1726741375/图书馆的发展脉搏.pdf；IFLA AI Entry Point for Libraries and AI.https://repository.ifla.org/handle/20.500.14598/4034；张晓林:Library Inside:AI 赋能图书馆新质生产力的一种基础模型
赵兴胜,李小洁,程川生,等.认知与实践：AI技术在高校图书馆应用现状调研分析[J/OL].信息与管理研究,1-16[2025-01-07].http://kns.cnki.net/kcms/detail/31.2118.G2.20250107.1047.002.html.

新的学科服务岗位不断涌现

- 加州大学圣巴巴拉分校“开放获取馆藏策划师”：
  - 评估机会，使学术交流系统向研究开放传播转变参与实施创新的馆藏模式和方法，管理UCSB开放获取（OA）出版基金。
  - 通过eScholarship、ORCID和Dryad促进开放出版选择；监测出版商和供应商提供的信息和数据，以确定开放获取相关的趋势和模式
  - 基于证据的馆藏和内容获取、管理和评估方面进行合作
  - .....

- 芝加哥大学图书馆等发布“人工智能馆员”：
  - 在图书馆开展基于人工智能的研究支持服务开发；
  - 参与全校范围的人工智能与数据科学相关项目。
  - 参与制定促进人工智能合乎道德且负责任使用的战略及资源开发工作，并致力于提升师生数字素养

（围绕科学数据，图书馆新职位嵌入科研全流程，参与数据管理活动）

- 数据图书馆员（基于ALA JobLIST 2006-2018年4月发布的美国学术图书馆相关岗位）：
  - 包括数据素养馆员、研究数据馆员、数据可视化馆员、数据馆员、数据仓储馆员、数据服务馆员、数据与政府信息馆员等各类图书馆岗位名称。
- 斯坦福大学图书馆：提供跨学科数字研究服务、数据服务，并形成人工智能计划。
- 上海交通大学图书馆设置数据服务中心：汇聚研究数据、资源数据和管理数据，支撑科研范式变革；

https://recruit.ap.ucsb.edu/JPF02255；https://news.cci.fsu.edu/jobs-and-internships/cci-job-board/mls-jobs/；https://www.simplyhired.com/search?q=librarian&l=&job=ru7XIWOw-CVi9qBq_xLexwUlOXTiiP3PiYvQ792rny04dtTGt1tBwA

科技信息服务行业中传统信息服务商、科技新贵纷纷涌入，渐成规范

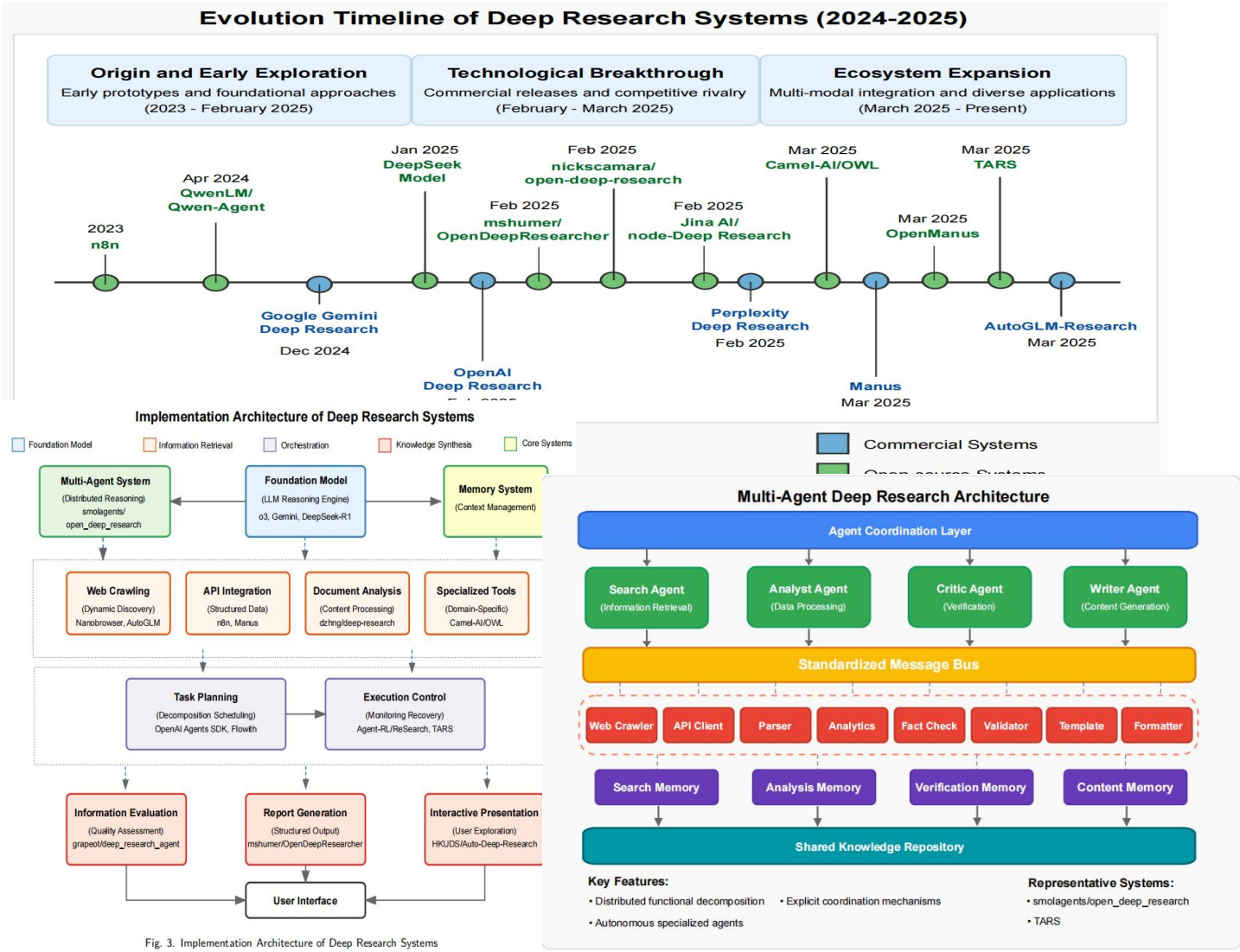


Fig. 3. Implementation Architecture of Deep Research Systems

Generative AI in Higher Education：The Product Landscape. https://sr.ithaka.org/wp-content/uploads/2024/03/SR-Issue-Brief-Generative-AI-in-Higher-Education-03082024.pdf
A Comprehensive Survey of Deep Research: Systems, Methodologies, and Applications. https://arxiv.org/abs/2506.12594

在猛烈拥抱AI的同时需警惕潜藏的风险

1 内容准确性与严谨性

AIGC属于模糊相似中寻找共性，对于同一问题的结果并不完全一致，也加大了严谨性论证的难度。AIGC内容其实是对各种学习资源的重组，在模糊相似中找共性，并不判断对于生成内容本身的合理性。AIGC生成的内容是与用户输入内容语义上关联性最大的，其内容却未必正确，该特质就决定了可能生成大量虚假信息。而且生成式 AI 并不会提供信息的出处，带来的误导性更大。

2 学术失信与公平失衡

1.并非每个人都能使用到AIGC工具，使用权并不平等。2.使用AIGC的抄袭监测难度更高，学术不公现象会只增不减，由此开展的学术不端调查耗费时间与精力也更巨大。

3 隐私与伦理道德

使用AIGC时自身隐私信息无差别被人工智能捕获并学习，导致隐私泄露，被侵犯知识产权。滥用AIGC生成的虚假信息，不经鉴别地传播和出版，导致侵犯他人隐私或知识产权，进而也带来信息欺诈风险。

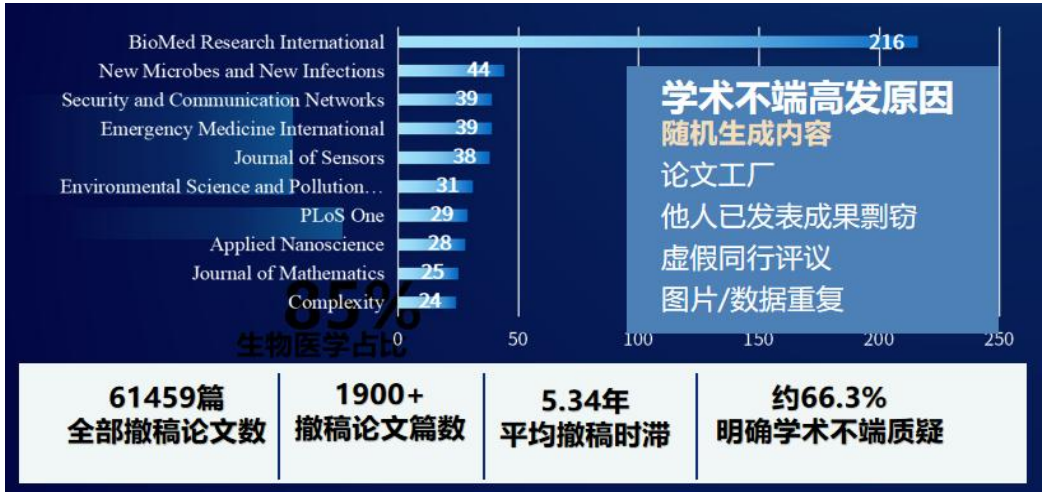
4 算法歧视与偏见

1.训练数据受限和数据偏差会影响输出模型

2.无法区分事实和虚构的缺陷，甚至生成的文本中能提取到充斥着人性弱点的信息。

5 技术成瘾产生创新依赖

AIGC全面的能力和对现实的补偿效应带有强烈的成瘾风险，科研工作者一旦离开AI就不会思考，或者被AI牵着鼻子走。



AI时代的数据建设将尤其凸显重要性和价值

□ 数据建设规范化、丰富化，无论对本身产品及服务的增强，还是试水新的数据要素服务，都是新机遇

高质量数据支撑LLMs

版税分配模式重新计量

AI许可解决方案

针对特定行业的集体版权许可

垂直领域模型的联盟生态共建

表 2 出版行业内容许可交易记录[5][6]

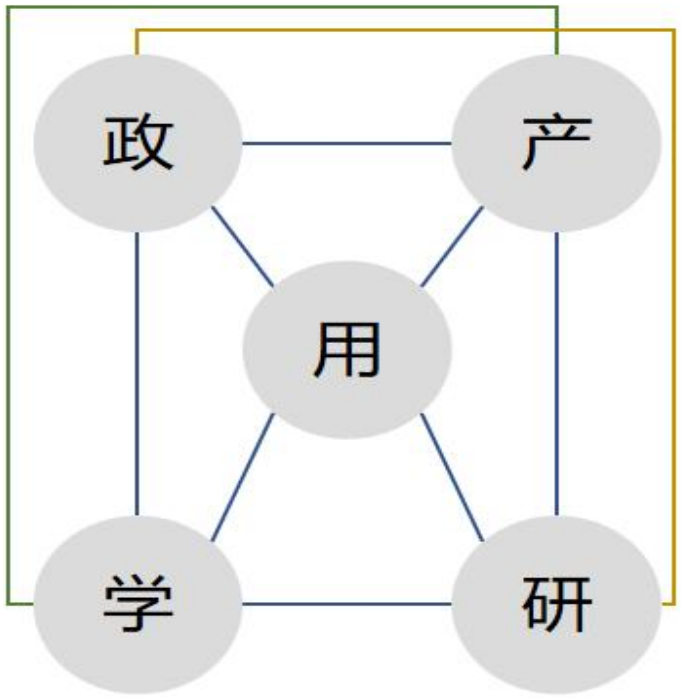
出版商	购买者	交易内容类型	内容用途
Wiley	大型科技公司	学术与专业图书	训练 LLM
	Potato	专注于生物类、经同行评议的文献内容	用于构建 Potato 工具，包括自动论文评审以及实验室指南生成工具
	Perplexity	护理、商业和工程领域的教育资源，包括电子书	订阅 Perplexity 专业版机构用户能访问购买的合作 Wiley 教育资源集。
	Amazon web Services	科技期刊全文	用于针对医疗保健和生命科学的开源工具包。2025 年合作推出科学文献搜索的生成式人工智能代理。
Taylor & Francis/Informa	Microsoft	图书（可能未包含期刊）	用于改进 AI 系统的相关性与性能
	大型科技公司		
OUP	未知	未知	合作开发 LLM
CUP	未知	未知	计划用于训练 LLM
SAGE	未知	出版的高质量科技内容	用于训练 LLM
De Gruyter Brill	未知	未明确，但公布信息特别聚焦图书	训练和测试基础 LLM，将开发工具和应用，并开发授权内容数据库验证 AI 内容准确性
NEJM Group	OpenEvidence	从 1990 年以来的新英格兰医学杂志及相关出版物	RAG 模型支持 OpenEvidence 平台，提升面向医生的服务
AAAS	ProRata	先期包括 Science Advances 期刊同行评议的 OA 论文及科学杂志的新闻故事	用于 Gist.ai 搜索引擎提高透明度与可信度，聚焦于少量高质量的内容

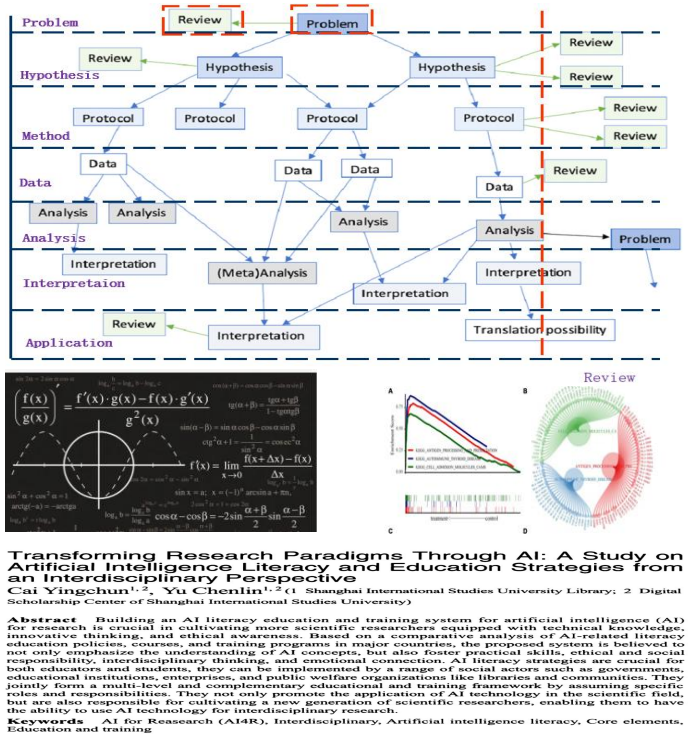
Generative AI Licensing Agreement Tracker. https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker/

始于文献，超越文献

- 覆盖原生资源->解构数据->知识数据的链式数据体系

AI与科技信息服务纵深化需求的结合





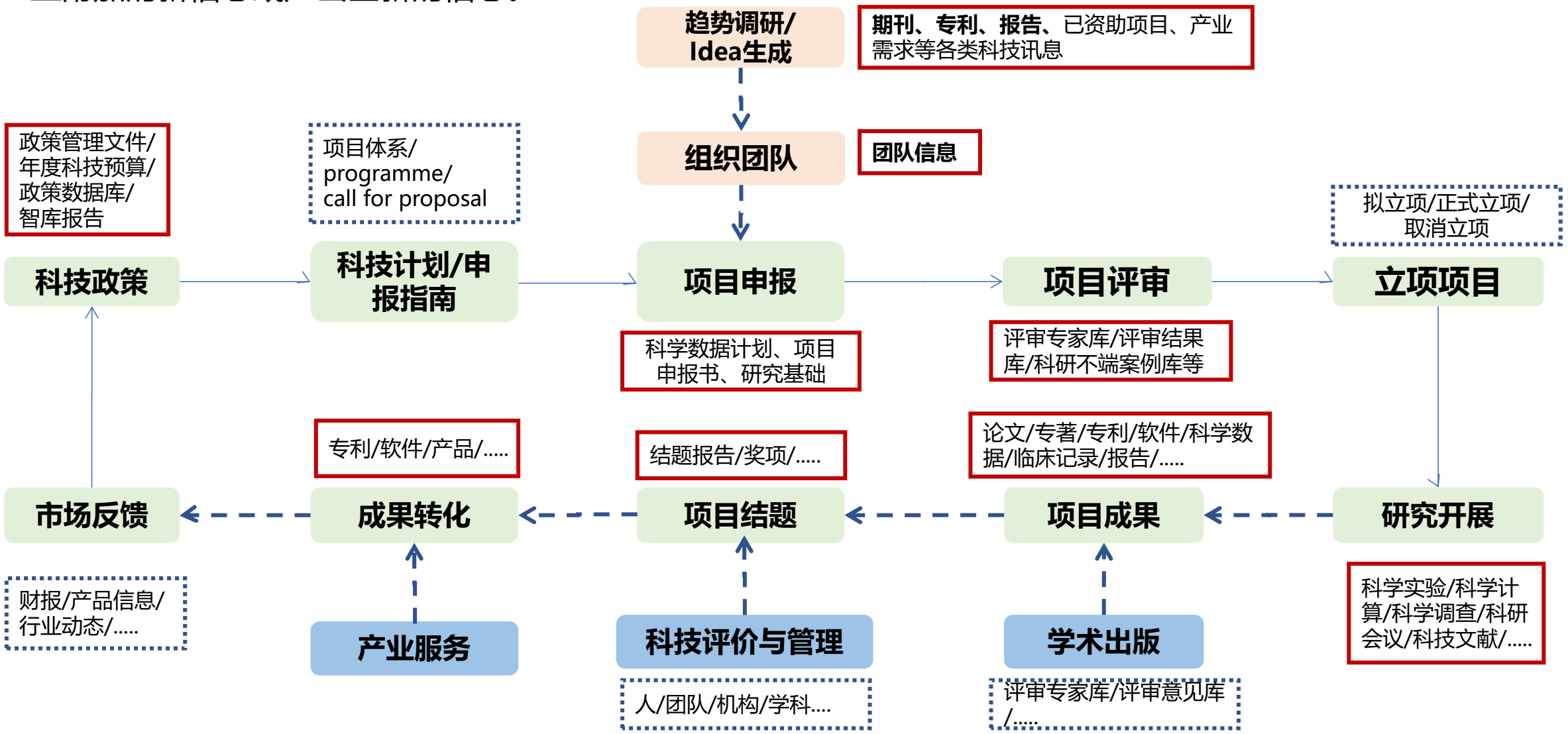
持续推动科技创新全谱段基础资源建设（渠道、类型、模态）

细化多模态原始数据解构重组与知识向量化

纵深数据的知识化挖掘与链式关联

立足文献，向左向右持续扩展

- **原生数据：**从完整的科技创新价值链看原生资源的扩展，并关注各价值环节在利用信息的同时会针对旧信息衍生附加的新信息或产出全新的信息。



立足文献，向下深探

□ 原生数据：文献是AI4S的学科服务中打开AI黑箱的重要保障，需要注意科技文献不仅只有TOC信息，图像、公式、软件、数据、基金信息等等，解析文献的过程实际上是在重现一个研究的过程

所在期刊

标题

作者信息

机构信息

出版变更

www.aging-us.com AGING 2020, Vol. 12, No. 4

Research Paper

Identification of biomarkers related to CD8⁺ T cell infiltration with gene co-expression network in clear cell renal cell carcinoma

Jiaxing Lin[1,*], Meng Yu[2,*], Xiao Xu[3], Yutao Wang[1], Haotian Xing[1], Jun An[1], Jieping Yang[1], Chaozhi Tang[1], Dan Sun[1], Yuyan Zhu[1]

[1]Department of Urology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China
[2]Department of Reproductive Biology and Transgenic Animal, China Medical University, Shenyang 110001, Liaoning, China
[3]Department of Pediatric Intensive Care Unit, The Shengjing Hospital of China Medical University, Shenyang 110001, Liaoning, China
*Equal contribution

Correspondence to: Yuyan Zhu; **email:** yyzhu@cmu.edu.cn

Keywords: clear cell renal cell carcinoma (ccRCC), CIBERSORT, weighted gene co-expression network analysis (WGCNA), CD8+ T cells, CCL5
Received: October 1, 2019 **Accepted:** February 4, 2020 **Published:** February 20, 2020

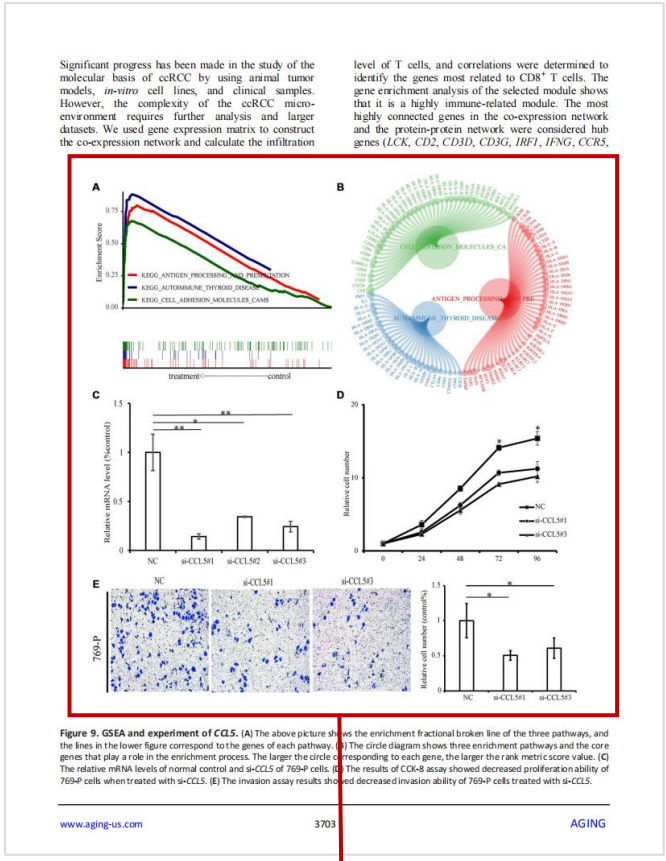
ABSTRACT

Clear cell renal cell carcinoma (ccRCC) is an extremely common kind of kidney cancer in adults. Immunotherapy and targeted therapy are particularly effective at treating ccRCC. In this study, weighted gene co-expression network analysis and a deconvolution algorithm that quantifies the cellular composition of immune cells were used to analyze ccRCC expression data from the Gene Expression Omnibus database, and identify modules related to CD8⁺ T cells. Ten hub genes (*LCK*, *CD2*, *CD3D*, *CD3G*, *IRF1*, *IFNG*, *CCR5*, *CD8A*, *CCL5*, and *CXCL9*) were identified by co-expression network and protein-protein interactions network analysis. Datasets obtained from The Cancer Genome Atlas were analyzed and the data revealed that the hub genes were meaningfully up-regulated in tumor tissues and correlated with promotion of tumor progression. After Kaplan-Meier analysis and Oncomine meta-analysis, *CCL5* was selected as a prognostic biomarker. Finally, the experimental results show that reduced expression of *CCL5* decreased cell proliferation and invasion in the ccRCC cell line. Various analyses were performed and verified, *CCL5* is a potential biomarker and therapeutic target which related to CD8⁺ T cell infiltration in ccRCC.

INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common malignant tumors, ranking seventh among male malignant tumors and tenth among female malignant tumors [1]. RCC accounts for 80% of all kidney cancer, and clear cell renal cell carcinoma is the most common subtype of renal cell carcinoma [2]. Smoking, obesity, and high blood pressure increase the risk of kidney cancer [3]. In recent years, immune checkpoint inhibitors have become the standard for first-line treatment of renal cell carcinoma [4–6]. However, no specific molecular markers have been for immunotherapy of renal cell carcinoma [2]. Therefore, the exploration of immune-related molecular markers is an important focus of renal cell carcinoma research.

RCC are prone to immune infiltration, and the characteristics of tumor microenvironment strongly alter the response to immunotherapy [7]. $CD8^+$ T cells contribute to tumor adaptive immunity. Among the immune cells of ccRCC, $CD8^+$ T cells account for the largest proportion [8]. In most solid tumors, highly infiltrating $CD8^+$ T cells are beneficial to tumor



Significant progress has been made in the study of the molecular basis of ccRCC by using animal tumor models, *in-vitro* cell lines, and clinical samples. However, the complexity of the ccRCC micro-environment requires further analysis and larger datasets. We used gene expression matrix to construct the co-expression network and calculate the infiltration level of T cells, and correlations were determined to identify the genes most related to $CD8^+$ T cells. The gene enrichment analysis of the selected module shows that it is a highly immune-related module. The most highly connected genes in the co-expression network and the protein-protein network were considered hub genes (*LCK*, *CD2*, *CD3D*, *CD3G*, *IRF1*, *IFNG*, *CCR5*,

Figure 9. GSEA and experiment of *CCL5*. (**A**) The above picture shows the enrichment fractional broken line of the three pathways, and the lines in the lower figure correspond to the genes of each pathway. (**B**) The circle diagram shows three enrichment pathways and the core genes that play a role in the enrichment process. The larger the circle corresponding to each gene, the larger the rank metric score value. (**C**) The relative mRNA levels of normal control and si-*CCL5* of 769-P cells. (**D**) The results of CCK-8 assay showed decreased proliferation ability of 769-P cells when treated with si-*CCL5*. (**E**) The invasion assay results showed decreased invasion ability of 769-P cells treated with si-*CCL5*.

Cell culture and siRNA-PTEN construction

The 769-P cells were cultivated in DMEM with 10% of fetal calf serum, 100 U/ml penicillin and 100 μg/ml streptomycin. The cells were cultured at 37 ℃, with 5% CO_2. The CCL5-310 siRNA sequences were as follows: 5'-GCUGAACAAGGGAAGCUUTT-3' 5'-AAGCUU GCCCUUGUUCAGCTT-3'. The CCL5-286 siRNA sequences were: 5'-GCAGGAUUUCCUGUAUG ACTT-3' and 5'-GUCAUACAGGAAAUCCUGCTT-3'. The CCL5-240 siRNA sequences were: 5'-UC GUCCACAGGUCAAGGAUTT-3' 5'-AUCCUUGAC CUGUGGACGATT-3'

RNA extraction and quantitative RT-PCR

RNA samples were extracted from cells with the FastPure Cell/Tissue Total RNA Isolation kit (Vazyme Biotech) and they were reverse-transcribed by HiScript III RT SuperMix for qPCR (Vazyme Biotech). To determine the relative transcript level, PCR was quantified in real-time using a LifeECO PCR machine (BIOER Technology Co, Ltd). SYBR Green was used as the fluorophore. The CCL5 primers sequences were as follows: 5'-CTC ATTGCTACTACTGCCCTCTGCGCTCCTGC-3' and 5'GCTCATCTCCAAAGAGTTGATGTACTC -3'. The PCR parameters were: 95 °C 5min, followed by 50 cycles of 30 s and 1 min at 60 ℃. Each sample was measured in three independent reactions. The threshold values of each sample/primer were determined, and the average error and standard error were calculated. Melting curve analysis was performed, and the mRNA expression levels were normalized against that of β-actin.

Cell proliferation analysis and invasion experiment

769-P cells were cultivated in five 96-well cell culture plates (1500 cells/well) respectively for 5 days. A volume of 10 μl of CCK-8 solution was added to each well in a plate at an interval of 24 h. The cells were cultures for 3 h in 37 ℃, and then a photometry test was performed at wavelength of 450 nm. The 769-P cells were then transferred into the top layer holes in a basement membrane plate with wells containing the medium without serum. The lower cavity was filled with 12% fetal calf serum medium to allow chemical attraction. The medium was removed after 12-hour incubation and the implant in the upper cavity was discarded. Cells invading the lower cavity were fixed with 700 ml of 4% PFA and stained with crystal for 1 h to visualize. The cells in the lower cavity were counted and normalized to the number in control conditions to measure the relative invasion capacity.

Abbreviations

ccRCC: Clear cell renal cell carcinoma; WGCNA: Weighted gene co-expression network analysis; CIBERSORT: Cell-type identification by estimating relative subsets of RNA transcripts; TIICs: Tumor-infiltrating immune cells; PPI: Protein-protein interactions. GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; GSEA: Gene set enrichment analysis.

ACKNOWLEDGMENTS

We want to thank TCGA, GEO, TIMER, and TISDIB for free use.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (grants 81672523, 81472404, 81472403, 81272834, and 31000572), the 2018 Support Plan for innovative talents in Colleges and Universities of Liaoning Province, the 2018 "million talents Project" funded Project of Liaoning Province, and the 2019 Key R & D projects of Shenyang.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016; 66:7–30. https://doi.org/10.3322/caac.21332 PMID:26742998
2. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, Grünwald V, Gillessen S, Horwich A, and ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2019; 30:706–20. https://doi.org/10.1093/annonc/mdz056 PMID:31987338
3. Tahbaz R, Schmid M, Merseburger AS. Prevention of kidney cancer incidence and recurrence: lifestyle, medication and nutrition. Curr Opin Urol. 2018; 28:62–79. https://doi.org/10.1097/MOU.0000000000000454 PMID:29059103
4. Considine B, Hurwitz ME. Current Status and Future Directions of Immunotherapy in Renal Cell Carcinoma. Curr Oncol Rep. 2019; 21:34.

基金资助

引文

文本内容、文本结构

图像、应用的数据、使用的仪器设备/研究软件

立足文献，向下深探-利用AI全面升级技术体系

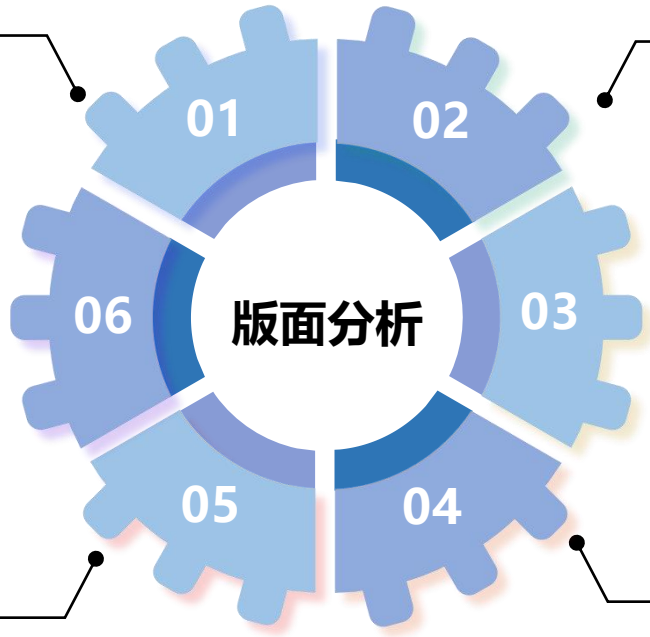
- 解构数据：AI加持的智能版式识别全面提升原生数据的解构粒度与性能，从描述元数据加工生产到知识要素解构

纸本加工提效
自动画框
引文切分

数据提取
大纲
段落
图片

格式转换
DOCX
MARKDOWN
HTML

文档翻译
多语种
版面还原



标题
包含主标题和各级子标题

表格/表注
表格图片、html格式表格

公式
公式图片、LaTex

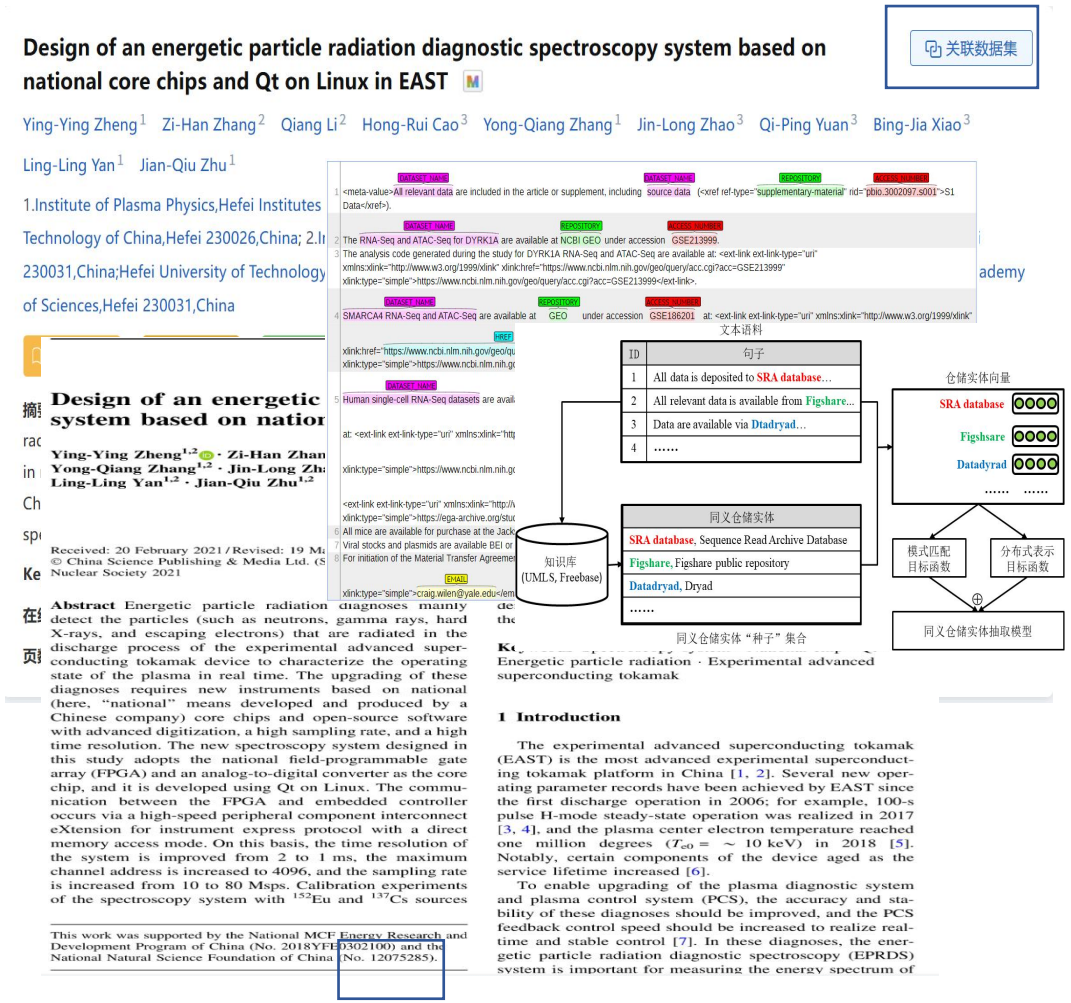
段落
完整语义的段落识别，去乱码

参考文献
切分作者，标题，刊名，年份等

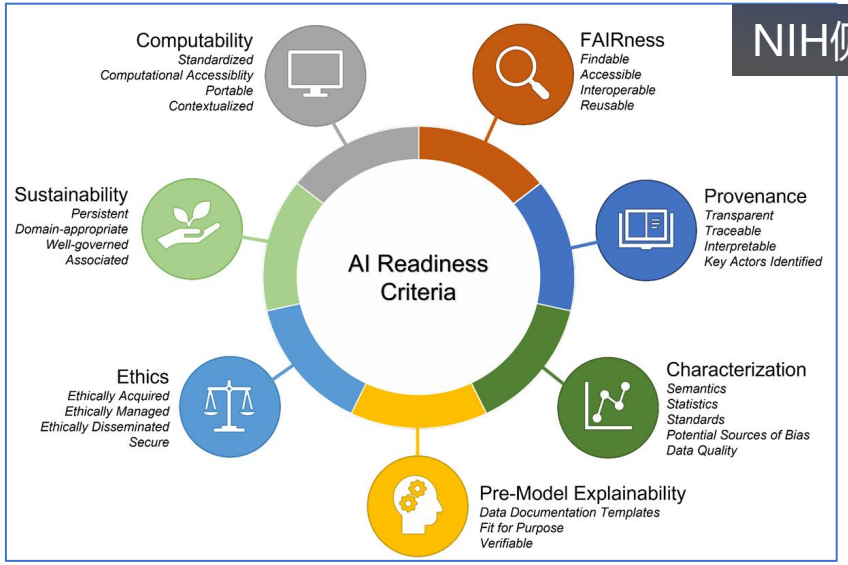
插图/图注
原始图片或截取图片

超越文献，开放互联

- 从单一文本到多元知识关联，支撑知识推理。



标准规范与AI技术同行，推动AI for Data<->Data for AI双循环



NIH侧重数据质量与功能性

对照FAIR标准，AI-Readiness 是指数据具有**机器可读性、可靠性、准确性、可解释性、预测性**，并且对未来的人工智能应用是**可访问**的。并提供了一系列对照工具。

Gartner强调场景适用

AI-Ready是指数据能够充分表征使用场景中的多样化模式、错误、异常值以及特定用途所需的意外情况，用于训练或运行AI模型。这是一种基于元数据的系统性流程与实践，通过数据对齐、验证及治理来实现

data.world侧重数据治理与验证

AI-Ready数据需经历严格的验证过程，如：格式检查、范围检查、引用完整性、名称验证、唯一性检查、一致性测试、问题跟踪系统、安全测试。

NOAA四级成熟度模型

从**数据分级与评估的视角，对照数据质量、访问、文档**三个维度要素形成标准

	数据一致性角度	数据访问角度	元数据角度
级别：0（Not AI-Ready）	未进行内部一致性的检查	仅通过请求或订单系统对公众开放使用	没有数据字典可用，或数据字典不是机器可读格式（例如PDF）
级别：1（Minimal）	手动一致性检查	只有一个非程序化的访问选项，例如文件下载	数据字典是机器可读格式（例如CSV、XML、JSON）
级别：2（Intermediate）	一致性检查部分自动化，部分结果有文档记录	多种访问方式，且至少包括一个可编程访问方法（API）	数据字典使用机器可读的元数据标准
级别：3（Optimal）	完全自动化的内部一致性检查和报告；且考虑了与社区数据集的外部一致性	除多种交付选项（下载、API）外，还有基于云和高性能计算提供的数据即服务（DaaS）	机器可读的元数据标准；参数与行业标准或其他机构数据集标准协调一致

标准规范与AI技术同行，推动AI for Data<->Data for AI双循环

- **解构数据与知识数据：**依托AI技术全面升级数据治理与分析工具，按照工具链思路进行体系化建设与组配，构建数据治理与分析AI智能体。

基于AI技术的数据治理与分析工具链的体系化建设

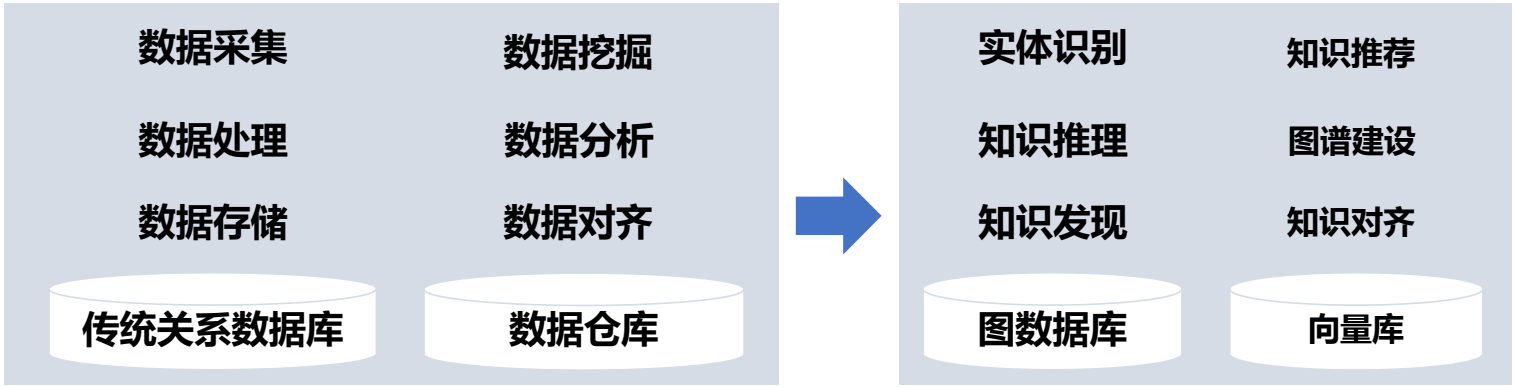
采集->知识抽取与融合->知识存储->知识引擎->知识推理->知识检索与生成							
数据采集	实体抽取	实体链接	多模态对齐	图存储	知识迭代	文本检索	← 强化学习与验证
OCR识别	关系抽取	自动标引	跨语言对齐	向量存储	检索引擎	图谱检索	← 场景任务模型微调
版面分析	事件抽取	自动分类	关系存储	知识推理	推理引擎	向量检索	← 垂域模型训练
内容审查	属性抽取	知识融合	文件存储	知识生成	计算引擎	智能对话	← 通用大模型训练

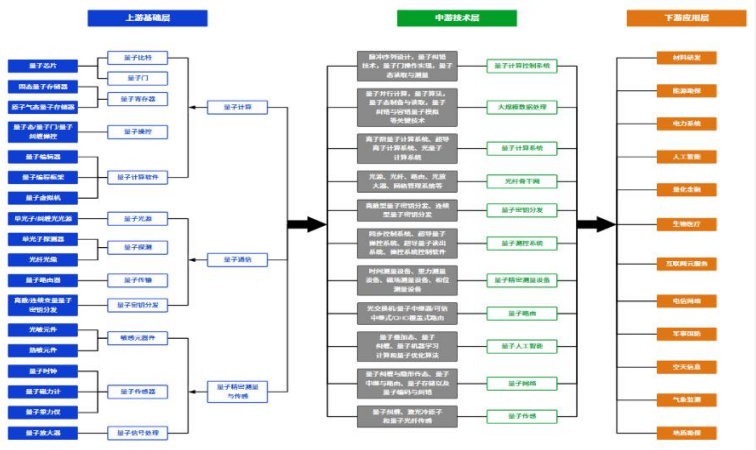
文献采购中心
资源存储中心
信息中转中心

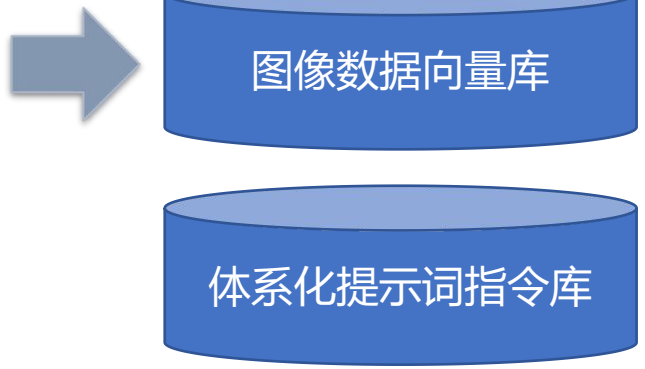
知识生产中心
AI ready数据支持中心
AI模型群智生态中心

标准规范与AI技术同行，推动AI for Data<->Data for AI双循环

- **解构数据与知识数据：**基于“通用大模型+面向场景任务的小模型”策略，形成系列工具，利用向量数据库存储并索引高维空间数据，支撑智能计算和复杂关系计算，提升性能与效果

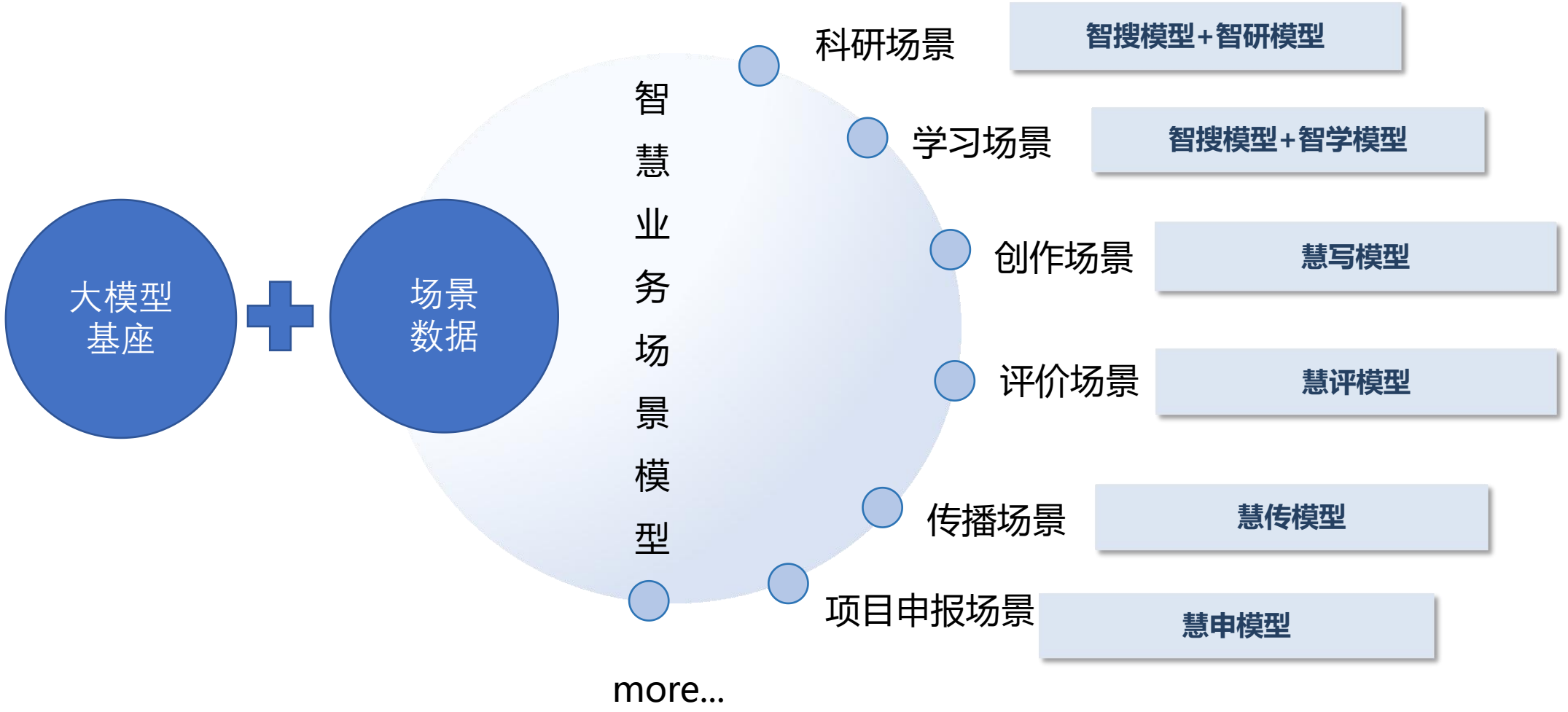






超越文献保障，迈向智能服务

- 基座大模型与场景数据的碰撞，推动文献资源服务转向工具服务、组织服务



场景与业务驱动的智慧服务矩阵重组与创新

面向学生、教师、研究者、科研管理人员等不同创新主体角色，以学习、研究、管理、决策等关键环节的实际业务场景为驱动，打造系列工具集，沉淀不同层次的一体化智能服务能力。

检索发现式的原文服务迈向认知引擎构建

- 在传统检索基础上，深度融合大语言模型，构建「语义理解-知识关联-动态优化」三位一体的认知智能引擎，实现从关键词匹配到语义向量认知的范式跃迁。

AI增强检索

循证式问题解答

智能文献伴读

AI增强检索	答案综述	参考文献快读	可视化知识脉络	文献伴读	全文精炼
术语解释	单篇问答	多篇对比	文献综述	阅读快报	智能翻译

大模型全文标引

科研实体、研究问题、创新点、实验方法、缺点和不足等

语义检索模型

改进检索排序算法，实现真实需求理解与精准检索

检索排序效果监测

首位命中率、前三条结果点击率、首屏效率值、翻页点击率

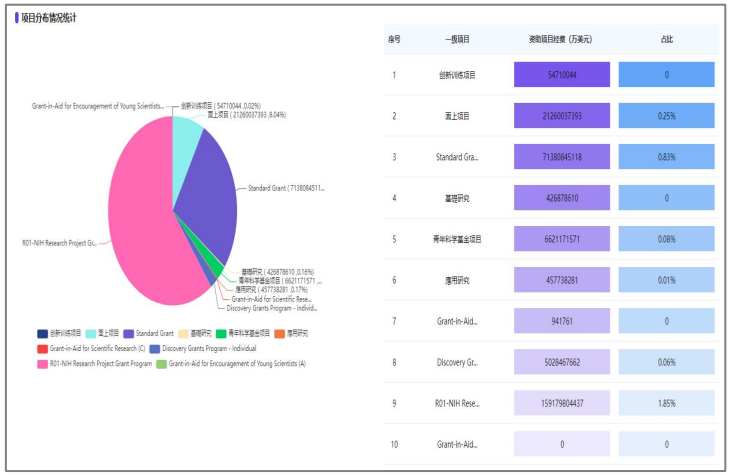
感知终端用户场景，提供链路式工具集自组配机制

□ **智研：**贯穿科研学习全流程，集“搜、读、写、研”为一体的智慧研究服务平台。

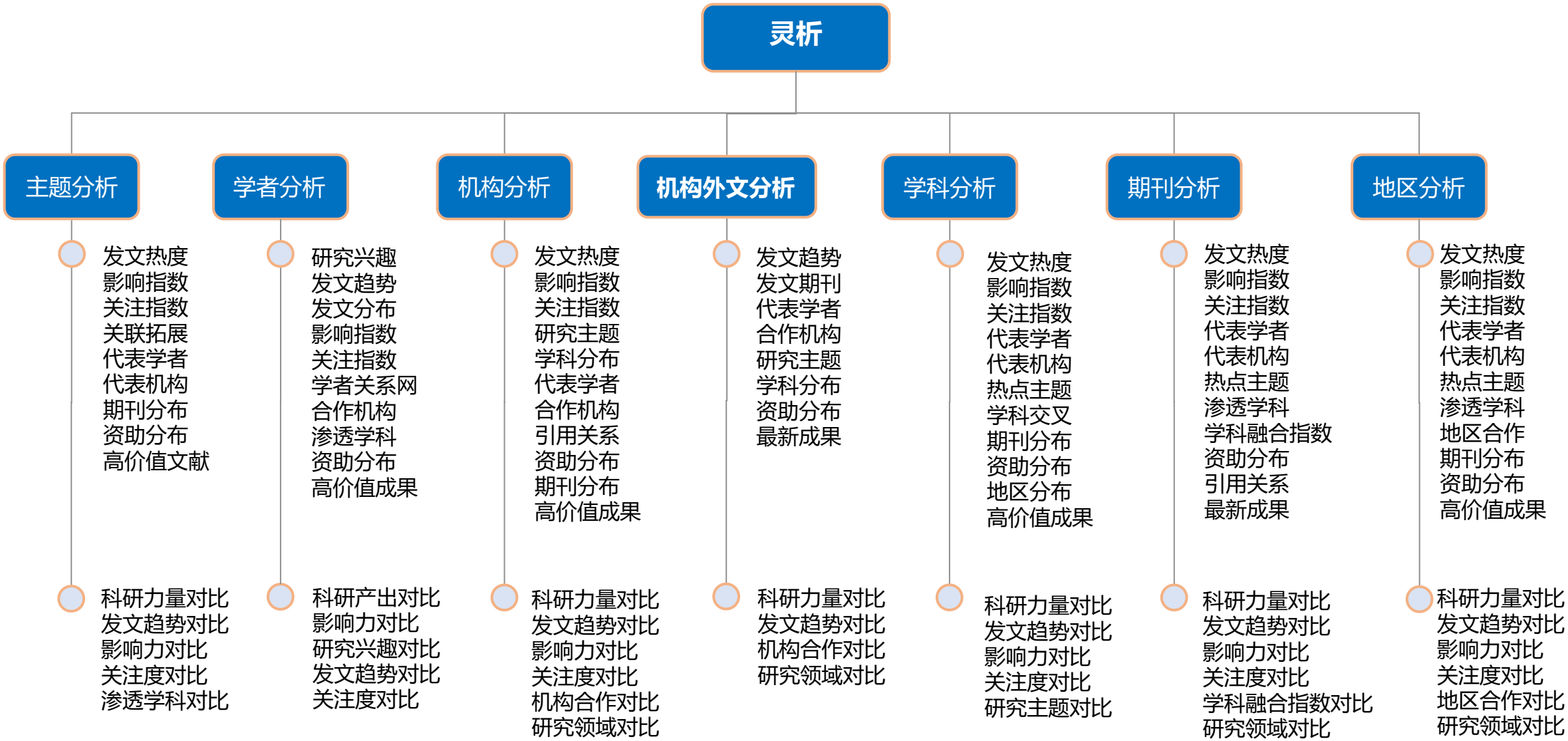
- **研学探知**：文献检索、文献阅读、阅读推荐、科技资讯推送。
- **知识管理**：个人知识库构建，文献收藏、标签管理、阅读笔记、订阅推荐、学术成果管理。
- **创作投稿**：论文选题、智能创作、论文投稿、智能排版、智能审校、论文检测。
- **研习手账**：科研计划个性化生成、日历提醒、科研任务进度提醒等。

从文献发现服务迈向更多业务场景

- **科研项目全周期：**面向科研项目的申报与实施，串联基金申报提醒、选题推荐、资助态势分析、推荐申报代码、立项申请书辅助撰写（国内外研究现状、前期研究基础）、诚信检测、成果汇集、科学数据仓储等全流程服务。



面向创新主体的多维立体评估



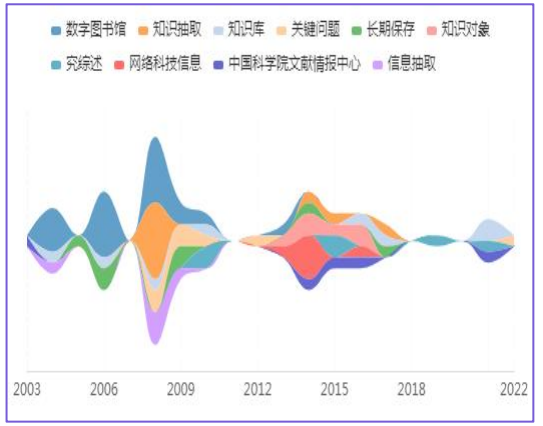
面向学科跟踪与评估需求，分别构建灵析与学科发展支撑服务

找科研方向

找科研合作

做学科建设

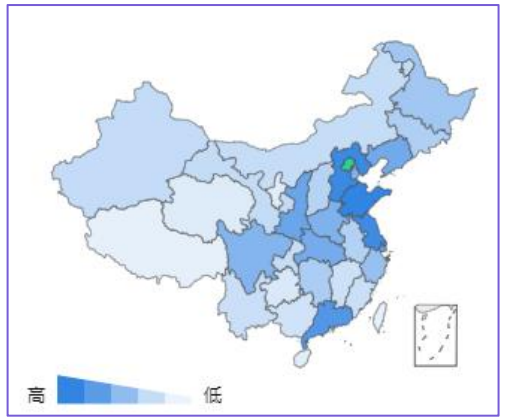
定位领域人才



找发文期刊

查看基金分布

地区科研定位



打通“事前-事中-事后”全链条覆盖的科研诚信建设与学术不端治理体系

文本检测方法、装置、计算设备及计算机可读存储介质：实现文本的有效检测，检测结果科学客观，系统运行稳定。

图像查重方法、装置、电子设备及可读存储介质：本专利实现了在海量样本中高效准确的对目标图像进行查重的目标。

子图分割方法、装置、电子设备及计算机可读存储介质：解决学术论文中图像子图切割与子图题注匹配。

一种论文主题无关引用检查方法及装置：本专利旨在从论文及其引文间的主题相关性角度发现不端行为。

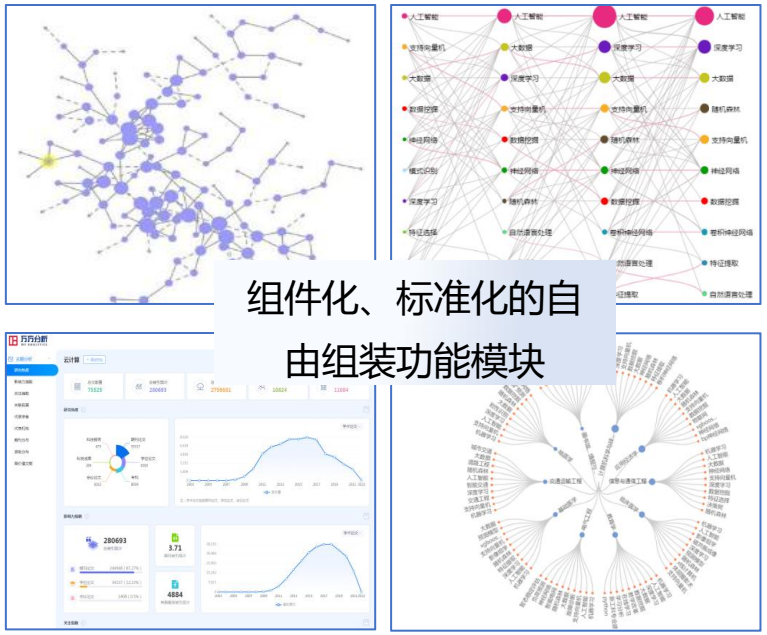
教育-预警-查验-检测全流程覆盖

贯彻通用与本地融合的一体化协同建设

- 构建产品矩阵时，充分考虑产品的模块化、标准化及可自由组装性，一方面支持敏捷组配，另一方面支撑共建应用生态

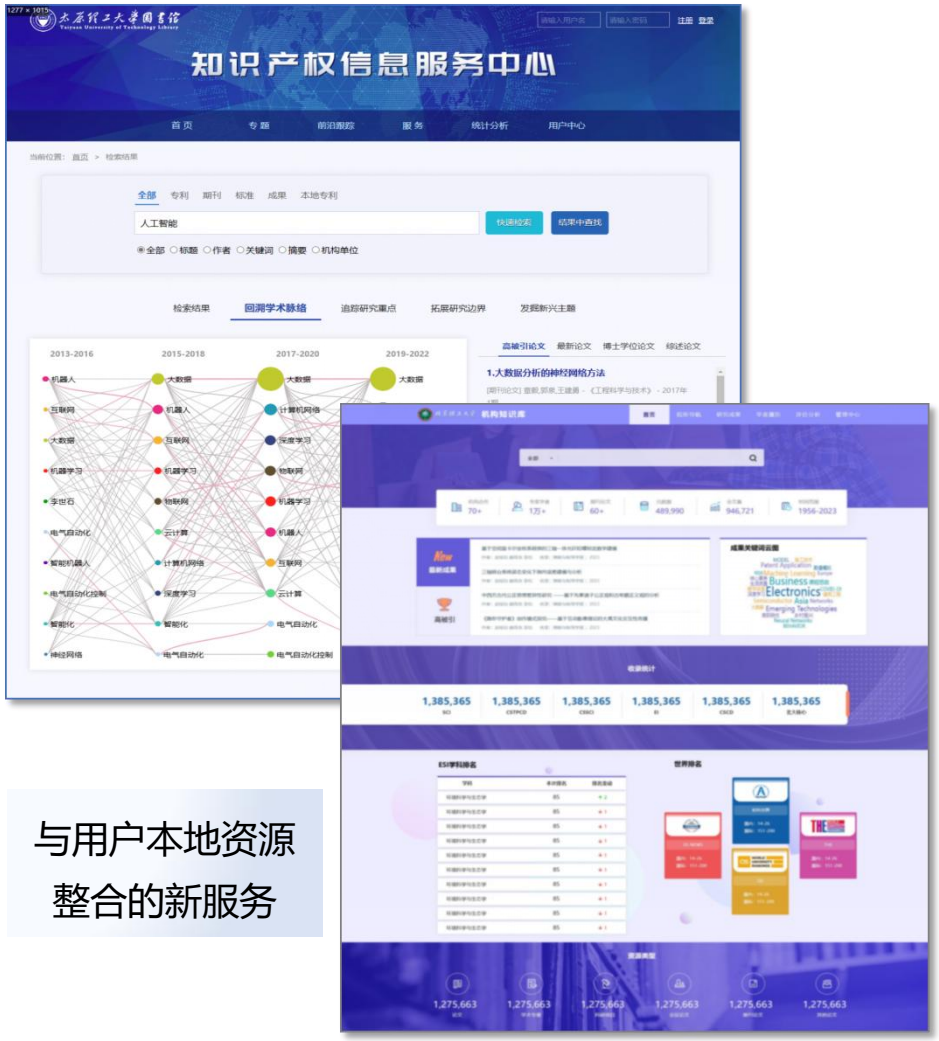
典型服务发展阶段

- 生态
- 服务
- 工具
- 数据



组件化、标准化的自由组装功能模块

资源集合　定制组装　向量集合



与用户本地资源整合的新服务

更多……

- 应对汇率上涨+经费缩减压力，竭力利用国际合作渠道数据+OA资源+馆际互借渠道资源，构建AI支撑的西文文献保障服务，并在探索基于AI的垂直学科领域科技资源服务体系自动构建。

资源订购优化策略咨询服务--方案一

分析维度		数据来源
订购目录	实际订购刊列表	机构提供（需规范提供信息）
订购成本	出版商公开的价目表	需在构建母体刊表过程中，补充单刊订购价
	实际订购价	机构提供
订购效益	订购资源的下载量	机构提供，counter数据
	订购资源中的机构发文	自主计算
	订购资源中的机构引用	
	订购资源中的机构被引	
文献保障成本	OA直达满足率	自主计算
	可使用的历史订阅备份比例	机构提供
	原文传递成本	自主计算

资源订购优化策略咨询服务--方案二

分析维度		数据来源
订购目录	拟订购刊列表	机构高发文、高引用、高被引刊，补充机构重点学科领域顶刊
订购成本	出版商公开的价目表	需在构建母体刊表过程中，补充单刊订购价
文献保障	OA直达满足率	自主计算
	原文传递成本	
订阅永久备份比例	可使用的历史订阅备份比例	需根据出版商订阅政策

持之以恒的价值理念

1、AI-Ready：构建面向科技创新场景，应用驱动的资源体系，形成适配AI利用的从“原生数据”到“知识数据的”全链路数据建设体系（AI Ready Data）。

2、AI In Technical：攻关文本挖掘、知识计算、认知智能等关键技术，加强**智能技术的研究和应用能力。**

3、AI For S&I：聚焦科技创新典型需求场景，加深对用户发展战略的理解，场景驱动产品升级与重构，实现**人工智能+场景服务的深度拓展能力。**

中国科学技术信息研究所 INSTITUTE OF SCIENTIFIC AND TECHNICAL INFORMATION OF CHINA

万方数据 WANFANG DATA

感谢大家！

万方数据将以更开放的心态，与各位精诚合作，建立更密切、多元的合作机制，共同实现高质量可持续发展。

万方数据 刘建华

liujh@wanfangdata.com.cn