

AI赋能科学研究

WestlakeChat.com 科研智能体

分享人：陈辉

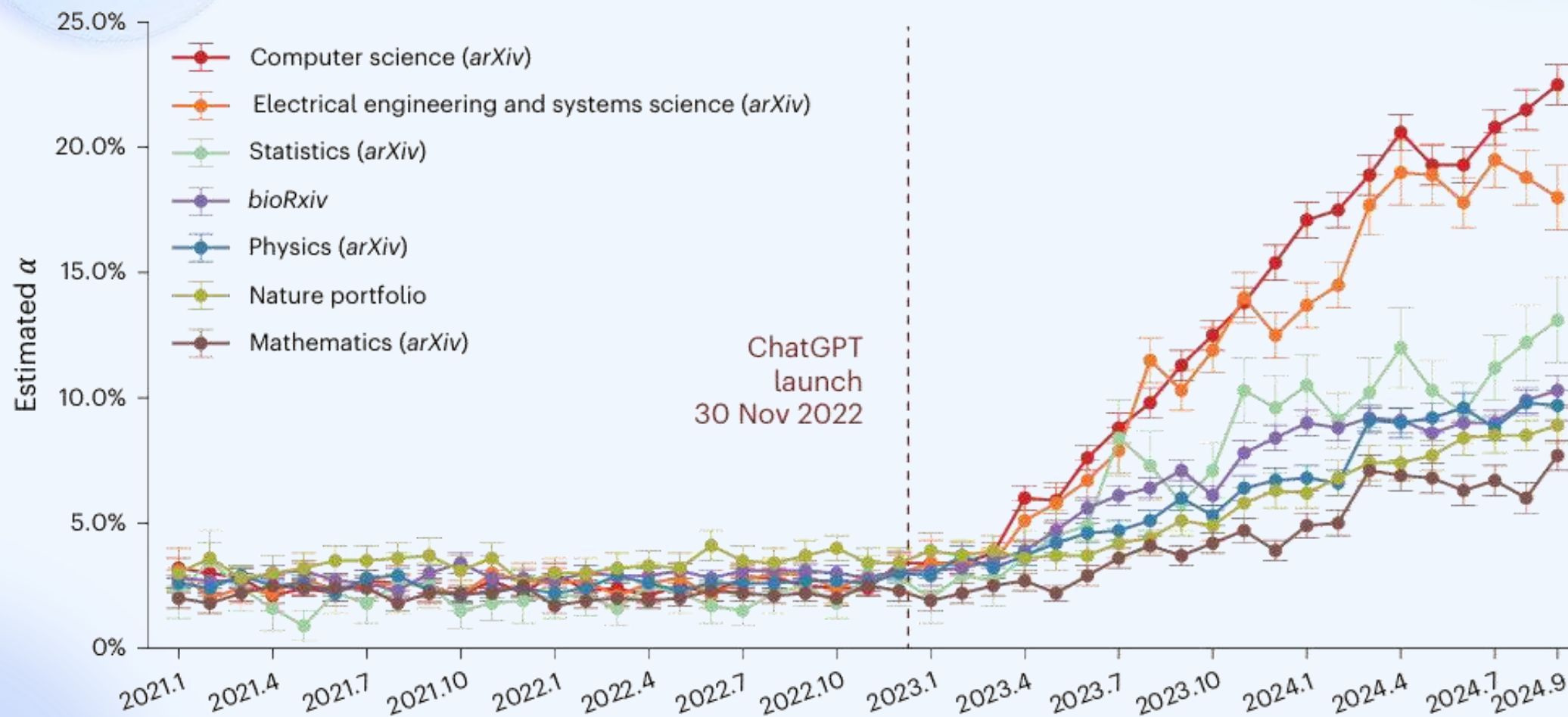
西湖未来基因

讲者介绍

陈辉，西湖未来基因联合创始人兼CEO、人工智能专家，中科大少年班物理学本科、美国德克萨斯大学电子和计算机工程硕士，曾任职谷歌、阿里巴巴、快手等公司，深耕机器学习算法和互联网领域十余年。

AI在研究中的使用越来越普及

AI
For
Science



<https://www.nature.com/articles/s41562-025-02273-8>

AI 助力科研，发展到了哪一步

AI
For
Science



AI应用科研的痛点



信息可靠性（幻觉）

AI系统频繁生成虚假或未经验证的信息，导致研究者必须花费额外时间确认内容真伪，严重影响科研可信度。



文献检索与引用缺陷

AI在文献综述方面表现薄弱，常提供不完整、过时或完全虚构的参考文献，无法满足学术研究对文献准确性和全面性的严格要求。



科研语言理解与输出质量不稳定

科研人员难以将专业学术语言有效转化为AI可理解的查询，同时AI输出质量参差不齐，难以精确捕捉研究者的具体需求和科研目的。



有没有一个工具？

解决这些痛点，并且一站式覆盖科研全流程

AI具体能帮我们做什么

文献检索

- 快速精准定位相关文献
- 处理海量文献，突破“信息过载”困境
- 深度挖掘文献价值，辅助理解
-

文献阅读

- 快速提炼核心信息
- 破解专业壁垒，让复杂内容变易懂
- 辅助深度分析
- 研究启发与拓展
-

文献管理

- 自动抓文献、去重复，解决来源乱、存得多
- 智能分类、打标签，让文献从堆成山到排好队
- 辅助写论文、做研究
-

综述文献写作

- 发表综述论文
- 辅助写研究型论文
- 了解领域知识
- 发现研究空白
-

AI文献检索

【传统方式】： 科研人员通过关键词在 PubMed等数据库手动检索，然后逐篇筛选，耗时费力，且可能因关键词局限遗漏相关文献。

【AI 辅助方式】： AI 工具基于语义理解，自动分析科研人员的研究主题，从海量数据库中精准检索相关文献，并按相关性排序推荐。

传统模式

关键词搜索结果庞杂，筛选耗时

跨学科文献覆盖不全

手动提取核心观点

难以追踪前沿动态

AI 辅助方案 (WestlakeChat 为例)

精准提问（搜索型 / 综述型），定向匹配文献

跨领域知识桥接，整合多学科文献

自动提炼研究结论、方法、数据

实时更新数据库（2 亿 + 论文），捕捉突破

AI的优势

01

使用LLM精准鉴别相关论文，搜索精度远高于Google Scholar等传统搜索引擎

02

模仿人类找论文的过程，一轮搜索后汇总已搜到的相关论文信息，调整方向进行多轮搜索，不遗漏关键文献

03

LLM生成详细报告，从找到的论文中提取和汇总关键结论、时间线等信息，极大提高阅读效率

04

数据库包含超过两亿篇论文，覆盖全学科，并且包含几乎所有Open Access的论文全文

AI文献阅读

科研人员逐字逐句阅读文献，遇到陌生术语需手动查询，理解复杂内容耗时较长，且难以快速把握文献核心观点。

传统阅读

AI辅助方案

- **智能摘要**：快速掌握论文核心（目的、方法、结论）；
- **深度解析**：拆解复杂理论，通俗化解释专业概念；
- **互动问答**：定向查询细节（如“实验样本量”“统计方法”）；

AI的优势



01 效率提升

压缩单篇文献阅读时间从小时至分钟级，大大提升阅读效率



02 专业壁垒破除：

自动拆解复杂理论、跨学科术语，通过通俗化解释降低理解门槛，非专业背景研究者也能快速掌握核心。



03 定向精准问答

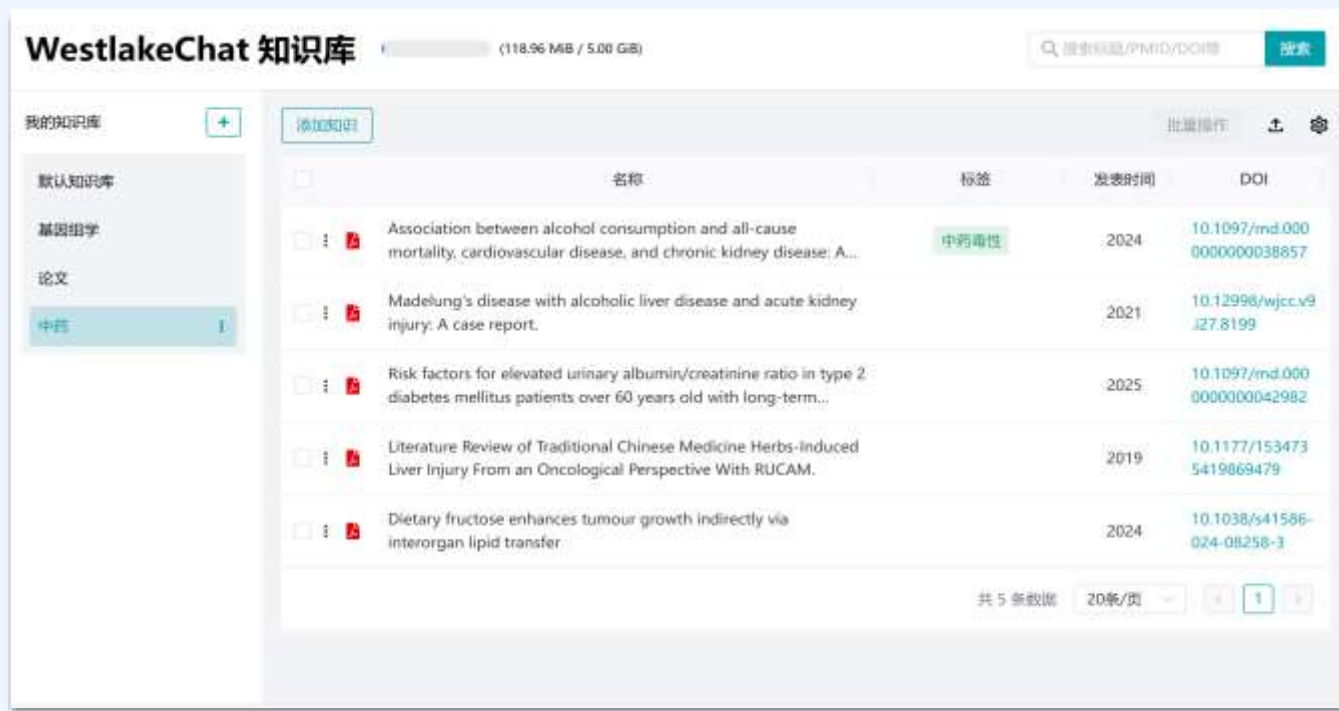
支持针对文献细节（如实验参数、统计方法、结论推导）的自然语言查询，无需反复通读原文即可获取目标信息。

AI文献管理

传统文献管理方式：科研人员手动将文献按主题分类存储在文件夹中，查找时需逐层浏览，且难以建立文献间的关联。本地文件散乱，分类混乱，检索困难，难与团队共享。

AI知识库

- 智能解析：自动按内容分类，减少手动操作。
- 高效检索：多维度快速检索，节省查找时间。
- 关联挖掘：发现文献间的潜在关联，为研究提供新思路。



AI的优势

智能分类：

支持批量上传文献可自动对文献进行解析，支持按主题、作者、时间等检索，替代人工整理耗时

高效检索：

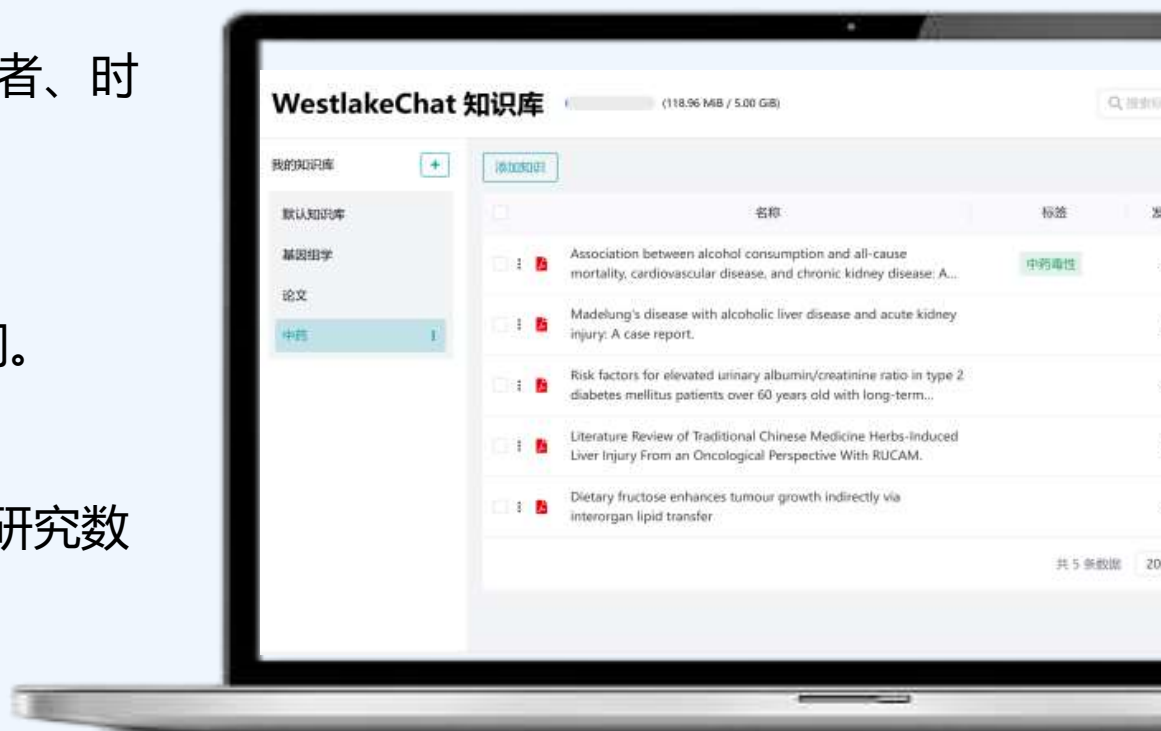
传统 1 天→AI 10 分钟，检索速度提升 90%，节省查找时间。

全量整合能力：

打通内外部文献资源，与 EndNote 等工具联动，实现个人研究数据、团队成果、公开文献的一体化管理。

跨模块协作：

与 AI 写作等模块无缝衔接，写作时可直接调用知识库文献支撑论点，实验设计时自动匹配相关研究的参数参考。



AI综述文献写作

科研人员需要手动整理大量文献，提取关键观点，然后构思框架、组织语言撰写，耗时耗力，且易出现观点重复或遗漏。

传统写作

From Genetic Associations to Genes: Methods, Applications, and Challenges

1. Introduction

1.1. The Challenge of Translating GWAS Findings to Causal Genes

Genome-wide association studies (GWAS) have successfully identified thousands of genomic loci associated with complex diseases and traits, yet the translation of these statistical associations into mechanistic understanding remains a formidable challenge. The vast majority of trait-associated variants identified through GWAS fall within non-coding regions of the genome, substantially complicating the elucidation of causal variants, genes, and underlying biological mechanisms[1]. This interpretative gap arises because GWAS-implicated loci typically span hundreds or thousands of genetic variants that exhibit similar statistical significance, making it difficult to pinpoint the specific causal elements[2]. The resolution limitations are further compounded by linkage disequilibrium patterns that vary across populations, creating ambiguity in variant prioritization. Moreover, even when statistical associations are robust, the functional consequences of identified variants often remain obscure, particularly for those residing in regulatory regions with context-dependent effects[3].

The complexity of this challenge extends beyond mere variant identification to encompass broader biological contexts. Environmental factors significantly modulate genetic associations, as demonstrated by studies showing how intrauterine conditions can dramatically alter the explanatory power of polygenic risk scores for conditions like schizophrenia[4]. Furthermore, the tissue and cell-type specificity of genetic effects introduces additional layers of complexity, with recent investigations revealing that disease-associated loci may exhibit distinct regulatory patterns across different cell populations. For example, Alzheimer's disease risk loci show enrichment in microglial enhancers and associate with specific transcription factors including SPI1, ELF2, and RUNX1[5]. These contextual dependencies necessitate sophisticated approaches that integrate diverse data types to move from statistical associations to mechanistic insights. The methodological landscape for addressing these challenges has evolved substantially, with techniques ranging from statistical fine-mapping and cross-population analyses to integration of expression quantitative trait loci (eQTL) data and application of machine learning frameworks[6] [7]. Despite these advancements, the interpretation of GWAS results continues to require careful consideration of both statistical evidence and biological plausibility to effectively bridge the gap between association signals and causal genes.

1.2. Evolution of Approaches for Gene Prioritization

Gene prioritization approaches have undergone substantial transformation since the early days of GWAS, evolving from simplistic positional mapping to sophisticated multi-layered integrative frameworks. Initial strategies relied predominantly on physical proximity, where variants were assigned to the nearest gene—an approach now recognized as insufficient given the complex architecture of regulatory elements[8]. This oversimplification led to the development of fine-mapping methodologies that leverage statistical evidence to refine association signals. These approaches integrated linkage disequilibrium patterns and functional annotations to prioritize causal variants within identified loci, ultimately narrowing candidates from thousands to a manageable subset[9]. The landscape further expanded with the emergence of expression quantitative trait loci (eQTL) analysis, which established statistical connections between variants and gene expression, though initially limited by tissue specificity and statistical power constraints[10]. As the field matured, integrative frameworks incorporating multiple data types became increasingly prominent, with methods like PrediXcan leveraging built-in SNP-expression models to predict gene expression levels regulated by putative eQTLs, thus enabling gene-based association studies with reduced multiple testing burden[11].

The most recent evolution has witnessed the development of comprehensive predictive models that integrate diverse data modalities. Modern approaches like ENCODE-RED employ iterative perturbations and supervised machine learning to construct increasingly accurate models of enhancer regulation, achieving state-of-the-art performance in predicting enhancer-gene interactions across hundreds of cell types and tissues[12]. These advanced methods systematically integrate chromatin state measurements, 3D genomic contacts, and large-scale genetic perturbation data to create encyclopedic resources of regulatory interactions. Concurrently, the field has embraced functional validation through techniques like CRISPR/Cas9 gene manipulation alongside network-based approaches that

Generated by RestEaseChat.com

AI 辅助

AI 工具可自动整合相关文献的观点，生成综述写作框架，辅助撰写初稿，并对内容进行梳理和优化。

AI的优势

AI生成综述的核心竞争力：

- 1.系统性、深度与全面性更优，适合作为领域研究的“全景地图”。
- 2.学术写作质量媲美真人，显著降低研究者时间成本。

人类书写论文的不可替代性：

在细分领域提供独特视角，图表与工具整合更贴近实践需求。

7 个维度 每项满分 10 分	内容全面性	批判分析深度	组织结构的 条理性和清 晰度	创新性：对 前沿问题的 剖析	实用性：对后 续研究或应用 的指导价值	写作质量	参考文献使用	总分 (满分70)
AI输出	9	9	9	8	9	9	10	63
人工撰写	8	8	8	7	8	8	9	56

效率的巨大提升：将单篇文章的写作时间从平均 27.4 天降低到约 20 分钟

扫码交流，免费试用



给力哥

浙江 杭州



陈辉 - 西湖未来基因

群聊：WestlakeChat交流讨论群



该二维码7天内(8月27日前)有效，重新进入将更新

打开 WestlakeChat.com 获取 14 天免费试用