

基于 ChatGPT 信息增益的男性不育研究热点和趋势评估*

陆俊宇¹ 王 姣¹ 刘行云¹ 何孟桥¹ 吴蓉蓉¹ 詹超英¹ 陈俐靖¹ 张影波^{1,2}
沈百荣¹

(¹ 四川大学华西临床医学院/华西医院 成都 610041

² 中国热带农业科学院热带作物品种资源研究所 海口 571101)

[摘要] 目的/意义 探索包括 ChatGPT 在内的大语言模型辅助决策系统的应用潜力, 以实现其在临床转化, 分析 ChatGPT 信息增益对男性不育研究热点总结和趋势评估的效果。方法/过程 基于 Scopus 数据库, 运用文献计量学、主题建模、“问-答”咨询等方法, 分析文献证据和 ChatGPT 虚拟数据下男性不育研究的主题差异。结果/结论 男性不育研究主题从单一转多元, “病例发现”“激素诊断”“精子提取”“遗传标志鉴定”等成为研究热点。应用 ChatGPT 可补充文献计量学与主题建模证据, 有效探索男性不育研究热点和趋势。

[关键词] 男性不育; ChatGPT; 文献计量学; 主题建模; 信息增益; 人工智能

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2024.04.009

Research Hotspots and Trends Assessment of Male Infertility Based on Information Gains from ChatGPT

LU Junyu¹, WANG Jiao¹, LIU Xingyun¹, HE Mengqiao¹, WU Rongrong¹, ZHAN Chaoying¹, CHEN Liqing¹, ZHANG Yingbo^{1,2}, SHEN Bairong¹

¹ West China School of Medicine/West China Hospital, Sichuan University, Chengdu 610041, China; ² Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS), Haikou 571101, China

[Abstract] **Purpose/Significance** To explore the application potential of large language models, including ChatGPT, in decision support systems for clinical translation, and to evaluate the effect of information gains from ChatGPT on summarizing research hotspots and assessing trends in male infertility studies. **Method/Process** The Scopus database is used to analyze the thematic differences between the literature evidence and ChatGPT virtual data in male infertility studies by using bibliometrics, topic modeling, “question – answer” consultation and other methods. **Result/Conclusion** The themes of male infertility research have shifted from singular to diverse, with “case discovery”, “hormone diagnosis”, “sperm extraction” and “genetic marker identification” becoming research hotspots. Using ChatGPT can supplement bibliometric and topic modeling evidence, and effectively explore the hotspots and trends of male infertility research.

[Keywords] male infertility; ChatGPT; bibliometrics; topic modeling; information gain; artificial intelligence

[修回日期] 2024-01-11

[作者简介] 陆俊宇, 本科生; 通信作者: 沈百荣, 教授, 博士生导师。

[基金项目] 国家自然科学基金面上项目 (项目编号: 32070671)。

1 引言

近年来, 由于全球环境恶化和局部污染加剧, 男性不育发病率急剧上升, 已成为全球性公共健康问题, 影响 20% 有生育愿望的夫妇^[1-2]。男性不育

的病因和发病机制因精子生成的复杂性而尚不完全清晰^[3]。生物学文献是医学信息挖掘和临床应用的宝贵资源^[4]。高通量测序技术,如基因组测序、转录组测序、甲基化测序等技术已经广泛用于男性不育症的研究和疾病标志物的发现,有效补充了男性不育症相关知识^[5]。因此,挖掘现有医学文献中的数据资源,不仅对构建知识库、知识图谱、聊天机器人和临床决策支持系统等专业知识工程具有重要价值,也对深入了解男性不育症的研究局限和未来发展趋势具有重要意义。

探讨人工智能在医学决策辅助系统中的应用是实现其临床转化的重要途径之一。2022 年 ChatGPT 3.5 作为标志性大型语言模型 (large language models, LLMs) 崭露头角^[6]。其问世对放射学、牙科学、耳鼻喉科等学科都产生深远影响,有学者^[7-9]认为 ChatGPT 可能改变当前医学教育和实践模式。然而,也有观点认为,作为一个广义生成模型,ChatGPT 的所有理解和推断都是基础性的,其主要用于基本知识的查询与推导,而不太可能提供创新性的见解。在某些专业领域,其表现可能无法超越现有的人工智能模型^[8]。

为了深入理解 ChatGPT 在生殖医学文献,特别是男性不育领域的整合与总结能力,本研究选取 Scopus 数据库作为主要数据来源。以“无精子症”“少精子症”“弱精子症”3 个男性不育的核心词为检索关键词,首先采用文献计量学、主题建模探索真实世界下男性不育研究的热点与趋势。其次,利用 ChatGPT 探索虚拟世界下男性不育研究的热点与趋势。最后,通过对比分析,探索 ChatGPT 在文献热点总结和趋势评估中的信息增强能力。

2 研究方法

2.1 文献检索与基础数据收集

Scopus 数据库是 Elsevier 出版集团旗下的引文数据库,因其所收录文献均经过严格的同行评审,受到业界广泛认可与关注。以男性不育研究的高频词“azoospermia”“oligospermia”和“asthenospermia”为检索关键词,筛选出题目中包含这些关键

词的相关文献。本研究的目标专注于英文文献,另外排除通讯、评论、勘误等类文章。同时,聚焦以人为对象的科学文献。导出“*.BibTex”格式引文记录数据,包括引文基本信息,如作者、标题、发表年份、期刊来源等,还包括作者提供的关键词、摘要、第一作者单位、通信作者单位等信息。

2.2 文献计量学分析

首先借助 bibliometrix 包提供的文献计量学函数,分析年度发文量、前 1% 发文机构、前 1% 发文章期刊、前 1% 作者排名 (以 H 指数为度量指标)、前 1% 作者文献产出年度变化等指标^[10]。其次,为了评估期刊对男性不育领域发展的贡献能力,基于影响力指数 (influence score) 计算各期刊贡献度。计算方式如下:

$$\text{Influence Score} = 0.01 \times \text{JCI}_{\text{journal}} \times \text{N}_{\text{record}}$$

其中 $\text{JCI}_{\text{journal}}$ 是期刊引文指数,代表该期刊近 3 年发表文章平均引用次数,其数据来源于 Clarivate 数据库; N_{record} 源于本研究中该期刊的发文记录数目。

2.3 基于隐含狄利克雷分布主题建模分析方法的男性不育研究热点分析

主题建模分析是探索学科研究热点和趋势最常用的方法,尝试以隐含狄利克雷分布 (latent Discriminant analysis, LDA)^[11] 为研究方法探索男性不育研究的热点和趋势。以男性不育医学文献摘要作为数据源构建语料库,首先排除数字、连接符、连词等,使用 quanteda 包构建语料库^[12]。其次,采用 Ponweiser 等描述的基于吉布斯抽样法的 LDA 方法主题建模^[13],其中最佳聚类数目的选择采用梯度上升法,其聚类范围设置 2~100,并以对数转化似然概率为最终聚类数目优选指标。最后,以最优聚类数目进行 LDA 方法主题建模,其参数采用 topicmodels 包默认参数,其中迭代次数设定为 2 000 次,抽样次数为 10 000 次,检验频率设为每 500 代进行一次,同时丢弃前 4 000 次抽样结果^[14]。

2.4 基于 ChatGPT 增益的男性不育研究主题和热点分析

为了探索 ChatGPT 在男性不育文献分析中的增

益效果, 研究对比基于真实世界 (文献) 和虚拟世界 (ChatGPT) 下男性不育研究中的主题和热点差异性。采用“当前男性不育研究的热点”作为查询问题, 将问题输入 ChatGPT 3.5 后提取研究热点关键词。比较基于文献数据和 ChatGPT 提供的关键词间的重合度, 以评估 ChatGPT 在男性不育研究主题和热点分析中的增益效果。为确保分析结果的可靠性, 对问题进行 3 次重复查询。

2.5 数据统计与可视化

上述数据分析和可视化均在 R 语言环境中完成^[15]。使用 bibliometrix 包进行文献计量学基本分析^[10]。主题建模则采用 topicmodels 包的 LDA 方法^[14]。在数据预处理和语料库构建过程中, 还使用了 dplyr 包^[16]和 quanteda 包^[12]等。而可视化结果的呈现主要依赖于 ggplot2 包^[17]和 sf 包^[18]。

3 结果

3.1 基于文献计量学证据的男性不育研究热点分析

截至 2022 年 12 月 31 日, 共从 Scopus 数据库检索到 2 968 篇题目中包含“azoospermia”“oligospermia”及“asthenospermia”的科学文献, 经排除非文稿型、非英语型和不以人为研究对象的文献后, 最后纳入 1 777 篇男性不育关联论文, 包括 1 638 篇研究性论文、118 篇综述性论文、18 篇会议论文和 3 篇书籍章节。

3.1.1 基于计年数据的男性不育发文量计量分析 依据计年数目, 男性不育科学文献数目从 1952 年的每年 2 篇, 提升至 2022 年的每年 119 篇, 其中文献数目年增长率为 6.01%。分析文献年增长速率发现, 20 世纪 70—80 年代和 2010 年左右呈现两次近指数增长期, 推测其分别与核酸检测史上的两次重大科技变革, 即聚合酶链式反应 (polymerase chain reaction, PCR) 检测技术和高通量测序技术发展有关, 见图 1A。对男性不育学科分析发现, “少精子症”“无精子症”和“弱精子”相关男性不育文献主要发表于医学 (67.47%)、生化与分子生物学 (27.01%)、多学科 (1.01%)、药学

(0.90%)、农学 (0.79%), 其他学科如化学、环境科学等占比 2.81%, 见图 1B。

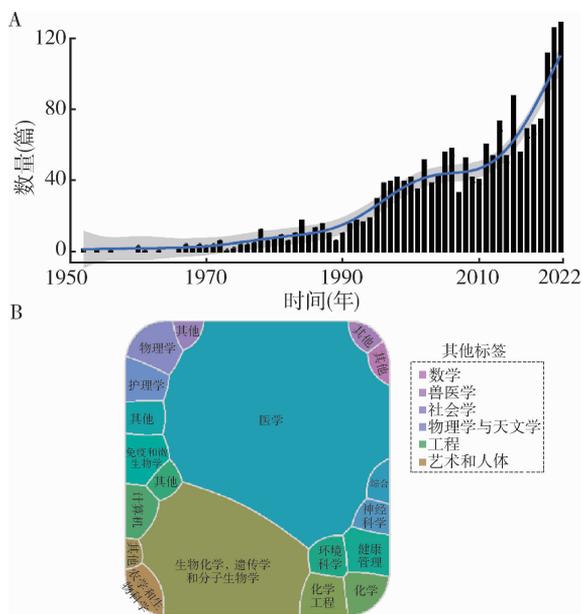


图 1 基于计年数据的男性不育发文量计量分析

注: A 为 1952—2022 年男性不育科学文献量; B 为男性不育总发文的主题分类。

3.1.2 基于计年数据的男性不育顶尖学者分析

对男性不育作者影响力的数据分析发现, 在学术影响力排名前 10 位的学者中有 8 位来自中国, 包括张远伟 (中国科学技术大学)、王沂峰 (南方医科大学) 等。对张远伟发文数量分析发现, 近年其在男性领域相关研究的影响力排名第 2 位, 总发文 22 篇 (均为多作者协作类文章), 其中以第 1 作者发文 4 篇。在男性不育研究领域国际专家中, 以巴西男性生殖中心的 Esteves S C 和康奈尔大学的 Schlegel P N 最负盛名。Esteves S C 的学术影响力为 0.29, 发文总数量为 19 篇, 居首位, 其中 5 篇独作文章和 14 篇多作者文章 (其中 4 篇为多作者文章的第 1 作者)。Schlegel P N 发表男性领域相关学术文章 46 篇 (其中 3 篇独作文章和 43 篇多作者文章), 排名第 6 位。此外, 对比中国和国际学者的发文情况发现, 国际学者如 Esteves S C 和 Schlegel P N 的发文情况中不仅包括多作者文章, 还包括独作文章; 而中国学者则多以团队协作的形式发文, 见表 1。

表 1 基于计年数据的男性不育顶尖学者及其协作网络分析

| 排名 | 作者 | 国家/地区 | 单位 | 影响力 | 文章总数 | 独作 | 多作者 | 多作者文章中 |
|----|--------------|-------|----------|-------|------|--------|--------|------------|
| | | | | 分数(分) | (篇) | 文章数(篇) | 文章数(篇) | 第1作者文章数(篇) |
| 1 | Esteves S C | 巴西 | 男性生殖中心 | 0.29 | 19 | 5 | 14 | 4 |
| 2 | 张远伟 | 中国 | 中国科学技术大学 | 0.18 | 22 | 0 | 22 | 4 |
| 3 | Liu Yu - jun | 中国 | 北京大学 | 0.14 | 22 | 0 | 22 | 3 |
| 4 | 王沂峰 | 中国 | 南方医科大学 | 0.14 | 22 | 0 | 22 | 3 |
| 5 | 李红钢 | 中国 | 北京大学 | 0.13 | 23 | 0 | 23 | 3 |
| 6 | Schlegel P N | 美国 | 康奈尔大学 | 0.12 | 46 | 3 | 43 | 5 |
| 7 | 张欣宗 | 中国 | 广东省精子库 | 0.09 | 23 | 0 | 23 | 2 |
| 8 | 张洪亮 | 中国 | 北京大学 | 0.06 | 34 | 0 | 34 | 2 |
| 9 | 刘见桥 | 中国 | 广州医科大学 | 0.05 | 20 | 0 | 20 | 1 |
| 10 | 李铮 | 中国 | 上海交通大学 | 0.05 | 41 | 0 | 41 | 2 |

3.1.3 基于发文国家/地区差异性的主发文单位分析 对发文的国家/地区分析发现,中国、美国、日本、意大利、英国等占据主要地位,见图2。对发文量排名前10位(总发文277篇,占男性不育总发文量的15.59%)的研究机构分析发现,中国共有5家单位上榜,即南京医科大学、上海交通大学、华中科技大学、香港中文大学和安徽医科大学,其排名分别为第1位、第3位、第4位、第6位和第9位。美国康奈尔大学和克利夫兰诊所基金会分别以40篇和22篇总发文量,占据第2位和第8位。此外伊朗教育、文化和研究学术中心、罗扬研究所和埃及开罗大学表现良好,其分别位于第5位、第10位和第7位,见表2。

表 2 基于发文总量的学术机构排名

| 排名 | 归属机构 | 国家/地区 | 发文数量(篇) |
|----|----------------|-------|---------|
| 1 | 南京医科大学 | 中国 | 43 |
| 2 | 美国康奈尔大学 | 美国 | 40 |
| 3 | 上海交通大学 | 中国 | 32 |
| 4 | 华中科技大学 | 中国 | 27 |
| 5 | 伊朗教育、文化和研究学术中心 | 伊朗 | 26 |
| 6 | 香港中文大学 | 中国 | 25 |
| 7 | 开罗大学 | 埃及 | 22 |
| 8 | 克利夫兰诊所基金会 | 美国 | 22 |
| 9 | 安徽医科大学 | 中国 | 20 |
| 10 | 罗扬研究所 | 伊朗 | 20 |

3.1.4 基于发文数量和影响力指数的期刊影响力分析 对男性不育期刊发文总量和学科影响力分析发现, *Fertility and Sterility*、*Human Reproduction* 和 *Andrologia* 是男性不育研究领域发文数量最高的期刊,其发文量分别为195篇、118篇和55篇。以影响因子、5年影响因子和期刊引用指数为指标, *Fertility and Sterility*、*Human Reproduction* 和 *Journal of Urology* 是男性不育研究领域最重要的期刊。综合分析发文数量和期刊影响力发现, *Fertility and Sterility*、*Human Reproduction* 和 *Journal of Urology* 是男

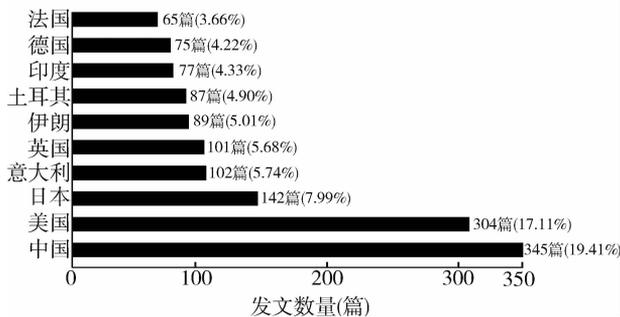


图 2 基于发文国家/地区差异的男性不育发文特征分析

性不育研究领域最重要的期刊，其影响力指数分别 425.1、263.14 和 113.3，见表 3。

表 3 基于发文总量和期刊影响力的男性不育研究

| 排序 | 期刊 | 发文数量 (篇) | 期刊影响力 | | | 影响力指数 |
|----|---|----------|-------|---------|------|--------|
| | | | 影响因子 | 5 年影响因子 | JCI | |
| 1 | <i>Fertility and Sterility</i> | 195 | 6.7 | 7.5 | 2.18 | 425.10 |
| 2 | <i>Human Reproduction</i> | 118 | 6.1 | 7.1 | 2.23 | 263.14 |
| 3 | <i>Andrologia</i> | 101 | 2.4 | 2.5 | 0.86 | 86.86 |
| 4 | <i>Journal of Urology</i> | 55 | 6.6 | 6.2 | 2.06 | 113.30 |
| 5 | <i>Journal of Assisted Reproduction and Genetics</i> | 54 | 3.1 | 3.5 | 0.93 | 50.22 |
| 6 | <i>Asian Journal of Andrology</i> | 53 | 2.9 | 3.1 | 0.63 | 33.39 |
| 7 | <i>Urology</i> | 36 | 2.1 | 2.2 | 0.74 | 26.64 |
| 8 | <i>International Journal of Andrology</i> | 34 | 2.6 | 2.9 | 0.81 | 27.54 |
| 9 | <i>Andrology</i> | 26 | 4.5 | 4.3 | 1.58 | 41.08 |
| 10 | <i>Journal of Clinical Endocrinology and Metabolism</i> | 26 | 5.8 | 6.1 | 1.27 | 33.02 |

注：JCI 表示期刊引用指数，代表该期刊近 3 年文章的平均引用次数。数据来源于 Clarivate 数据库。

3.2 基于主题建模的男性不育研究热点和趋势分析

经模型参数优化后，发现当聚类数设置为 $K = 15$ 时，其呈现的主题分布最符合男性不育研究趋势分布状况，见图 3。其中主题 1 以“vas”“patient”等为主题关键词，表明该主题以病例研究为主要研究方式，探索输精小管阻塞患者中的精子数量或功能变化。主题 2 以“sperm”和“FSH”为关键词，表明该主题主要探索促卵泡生成激素（follicle-stimulating hormone, FSH）在精子数量或功能异常中的作用。主题 3 以“NOA”“gene”和“express”为关键词，表明该主题主要探索非阻塞性无精子症患者中的基因表达模式或标志物。主题 4 与主题 3 的关键词类似，以“NOA”“gene”和“express”为关键词，但是“protein”和“level”等关键词比例明显增加，表明该主题研究方向以“protein”含量的高低探索基因表达变化。主题 5 以“NOA”和“mutation”为主题关键词，表明该主题主要探索脱氧核糖核酸（deoxyribo nucleic acid, DNA）突变对非阻塞性无精子症的影响。主题 6 以“sperm”“NOA”“group”和“rate”为关键词，表明该主题以类群或队列为研究对象，探索精子活性或数量。主题 6 以“delect”“microdelet”“AZF”和

“AZFa”为主题关键词，表明该主题主要探索生精基因 AZF 的缺失或部分缺失对男性不育的影响。主题 8 以“ISCI”“sperm”和“rate”为关键词，表明该主题主要探索活性精子比例对卵泡浆内单精子注射结局的影响。主题 9 与主题 2 类似，但是增加了“serum”“hormon”等关键词比例，表明该主题不仅研究促卵泡生成激素对男性不育的影响，还增加了与其他激素检测比例。主题 10 则以“NOA”“SNP”“SNPs”和“genotype”为关键词，表明该主题主要探索基因型或 SNP 位点对非阻塞性无精子症的影响。主题 11 则以“sperm”“NOA”“testis”和“retrieve”为关键词，表明该主题主要探索非阻塞性无精子症患者中睾丸精子提取术。主题 12 则以“semin”“tubulin”“cell”和“plasma”为关键词，表明该主题以白蛋白数量为指标，探索精浆中免疫细胞对精子数量或功能的影响。主题 13 以“sperm”“male”“infertile”和“cause”为关键词，表明该主题以探索男性不育患者中的精子质量、数量或功能异常为主要内容。主题 14 以“testicular”和“biopsy”为主题关键词，表明该主题以男性不育患者中睾丸活检为主要检验方法，探索睾丸中各生精细胞数量和比例对不育结局的影响。主题 15 以“patient”“group”和“DNA”为关键词，表明

该主题主要探索 DNA 变异在患者类群中的比例。对主题的时间分布规律研究发现，男性不育的主题分布呈现由简单到多样化的变化历程。如 20 世纪 60 年代，男性不育研究的主题主要是“睾丸组织活检”和“非阻塞性无精子症蛋白含量检测”。进入 20 世纪 80 年代，随着 PCR 技术及色谱技术发展，

基因变异检测和激素含量高低成为男性不育患者的常用检测方法，其中主题 2、主题 5、主题 6 和主题 9 等明显增加。进入 21 世纪后，尤其是 2010 年以后，以“转录组”“蛋白组”和“全基因组测序”等为代表的组学技术被广泛应用于男性不育研究，男性不育研究呈现多样化的趋势。

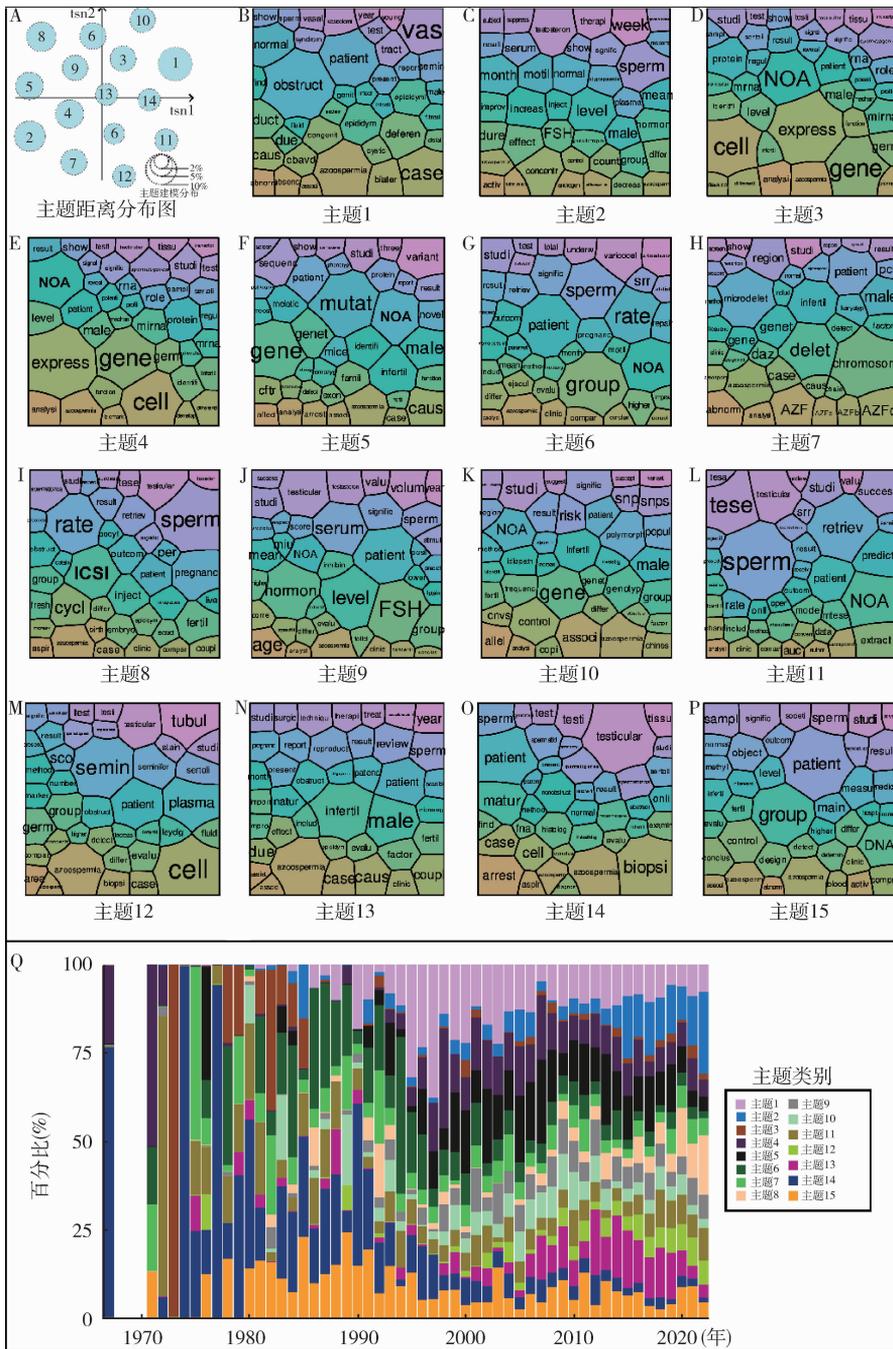


图 3 基于隐含狄利克雷分布的男性不育趋势分析

注：A 为主题距离分布图；B—P 为基于 LDA 分析的主题及其关键词频率；Q 为不同主题随时间分布的变化规律。

3.3 基于 ChatGPT 增益的男性不育研究热点分析

ChatGPT 是人工智能领域的超级工具，被证实具有趋势总结能力。为了探索 ChatGPT 在男性不育研究中的增益效果，尝试以“问 - 答”形式探索 ChatGPT 在男性不育研究热点中的增益效果。结果表明，ChatGPT 共回答了 15 个男性不育研究的热点问题，其中对男性不育的风险因素方面，提及“环

境暴露”“遗传因素”“表观遗传因素”“激素不平衡”“免疫因素”等对精子数量、质量与功能的影响；对于男性不育治疗和预防，提及“干细胞疗法”“非侵入性辅助生殖”“出生控制”和“育性保护”等不育治疗方案和预防策略；对男性不育研究的政策性指导，提及“精准医学”“人工智能”“道德和法律”“男性不育的社会和心理因素”等发展方向和策略，见表 3。

表 3 基于 ChatGPT 增益的男性不育研究热点评估

| 主题 | 主题解释 | 推荐次数 |
|----------|--|------|
| 环境暴露 | 调查内分泌干扰化学物质、生活方式等对男性生殖健康的影响 | 3 |
| 干细胞 | 探索干细胞在男性不育症治疗中的应用潜力 | 3 |
| 激素不平衡 | 探索激素在男性不育症中的作用，发展激素治疗 | 2 |
| 非侵入性辅助生殖 | 发展和改进辅助生殖技术 (assisted reproductive technology, ART)，如胞浆内单精子注射 (intracytoplasmic sperm injection, ICSI) 和睾丸精子提取 (testicular sperm aspiration, TESE) | 2 |
| 男性出生控制 | 研究男性避孕的新方法 | 2 |
| 表观遗传 | 探索精子的表观遗传修饰如何影响生育能力和后代健康 | 2 |
| 社会和心理因素 | 探讨男性不育对个人和夫妇的心理影响 | 2 |
| 精子健康和数量 | 了解影响精子质量的因素以及改善精子质量的方法 | 2 |
| 不育遗传学 | 研究导致男性不育的遗传因素和潜在的基因治疗或干预 | 1 |
| 精准医学 | 根据个人的遗传等特征定制不育症治疗策略，以优化结果 | 1 |
| 人工智能 | 使用人工智能和机器学习算法分析精子质量和生育结果预测 | 1 |
| 育性保护 | 保留男性生育能力的技术研究 | 1 |
| 不育中的免疫因素 | 了解免疫系统在男性不育症中的作用和潜在的免疫治疗 | 1 |
| 道德和法律 | 讨论有关生育治疗、精子捐赠和基因检测的伦理和法律问题 | 1 |
| 男性避孕 | 不断努力开发安全、有效、可逆的男性避孕新方法 | 1 |

4 讨论

近年来，伴随着高通量测序技术发展和不育人数的持续攀升，不孕不育研究相关文献呈现出近指数的速率增长。为了探索和对比基于真实世界（文献）数据和虚拟世界（ChatGPT）数据下男性不育研究中主题和热点的差异性，以“无精子症”“少精子症”和“弱精子症”为题目关键词，共从 Scopus 数据库检索到 1 777 篇男性不育研究相关科学文献。进一步的文献计量学和主题建模分析显示，男性不育相关研究经历了从简单到多样化的变化历程，其中“阻塞性不育患者的病例发现”“激素在不孕不育及单精子注入患者中的价值”“非阻塞性

无精子症患者中睾丸精子提取术”等是男性不育研究领域的热点问题。基于 ChatGPT “问 - 答”系统，获取包括“环境暴露”“干细胞”“激素不平衡”等 15 个男性不育研究热点。对比基于文献证据和 ChatGPT “问 - 答”结果分析发现，ChatGPT 的回答不仅能涵盖“环境暴露”“遗传因素”“表观遗传因素”“激素不平衡”等因素对精子数量、质量与功能的影响，还兼顾了“精准医学”“人工智能”“育性保护”等文献数据报道较少的证据，所以 ChatGPT 是探索和了解学科研究趋势和热点的效果增益工具，其对快速了解学科方向、制订学科发展策略具有重要的应用潜力。

但由于 ChatGPT 训练时的语料库主要来源为网络文本语料库、WebText2、现有图书和维基百科

等, 而非源于专业男性不育医学文献, 其在回答男性不育研究的热点问题时具有泛化性而缺乏专业性, 如 ChatGPT 高频推荐了“环境暴露”“干细胞”“激素不平衡”“非侵入性辅助生殖”“男性出生控制”“表观遗传”和“精子健康和质量”等男性不育研究热点问题, 但对“精准医学”“人工智能”等真实世界(文献)数据的推荐频次较少, 突显出 ChatGPT 医学问题回答的偶然性。因此, 探索将 ChatGPT 与专业知识库相结合的垂直 GPT 将有助于提升相关医学问题回答的专业性和准确性。

5 结语

探索大语言模型在辅助决策系统中的应用潜力是实现其临床转化的重要途径。本研究以男性不育问题为研究对象, 对比真实世界(文献证据)和虚拟世界(ChatGPT 总结)下男性不育的热点。结果表明 ChatGPT 不仅覆盖“环境”“遗传”“激素不平衡”等风险因素对男性不育负面影响的热点问题, 还突显了文献证据下因文献数量较少而无法重点显示的“精准医学”“育性保护”等热点问题, ChatGPT 等大语言模型工具是传统文献计量学的有效补充, 有助于探索男性不育研究热点和趋势。但是在研究中也发现 ChatGPT 在回答医学问题时的随机性和偶然性, 其无法有效提供可实现临床转化的数据证据。此外, 本研究还存在局限性, 即所有证据依赖于对临床文献的总结和大语言模型的系统总结。而在真实世界, 技术的革新、诊断或预测模型的开展、标准或工具的开发将会对学科发展产生极大的促进作用, 后续研究将会重点关注男性不育领域的技术革新、诊断或预测模型的开展、标准或工具的开发, 以期揭示男性不育文献、工具、模型的演化规律奠定基础。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 AGARWAL A, BASKARAN S, PAREKH N, et al. Male infertility [J]. *Lancet*, 2021, 397 (10271): 319–333.
- 2 SUN H, GONG T T, JIANG Y T, et al. Global, regional,

and national prevalence and disability – adjusted life – years for infertility in 195 countries and territories, 1990–2017: results from a global burden of disease study, 2017 [J]. *Aging*, 2019, 11 (23): 10952–10991.

- 3 RAY P F. Deciphering the genetics of male infertility: progress and challenges [J]. *The journal of urology*, 2011, 186 (4): 1183–1184.
- 4 ZHAO S, SU C, LU Z, et al. Recent advances in biomedical literature mining [J]. *Briefings in bioinformatics*, 2021, 22 (3): 1–19.
- 5 张影波, 詹超英, 王姣, 等. 精子表型异常的生物信息学研究: 现状与未来趋势 [J]. *中国科学: 生命科学*, 2023, 53 (2): 274–286.
- 6 BOMMASANI R, LIANG P, LEE T. Holistic evaluation of language models [J]. *Annals of the New York academy of sciences*, 2023, 1525 (1): 140–146.
- 7 CLUSMANN J, KOLBINGER F R, MUTI H S, et al. The future landscape of large language models in medicine [J]. *Communications medicine*, 2023, 3 (1): 141.
- 8 DAVE T, ATHALURI S A, SINGH S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations [J]. *Frontiers in artificial intelligence*, 2023, 6 (5): 1169595.
- 9 THIRUNAVUKARASU A J, TING D S J, ELANGO VAN K, et al. Large language models in medicine [J]. *Nature medicine*, 2023, 29 (8): 1930–1940.
- 10 ARIA M, CUCCURULLO C. Vibliometrix: an R – tool for comprehensive science mapping analysis [J]. *Journal of informetrics*, 2017, 11 (4): 959–975.
- 11 MANGSOR N S M N, NASIR S A M, ABDUL – RAHMAN S, et al. Identifying topic modeling technique in evaluating textual datasets [C]. Singapore: The Data Science and Emerging Technologies, 2023.
- 12 BENOIT K, WATANABE K, WANG H, et al. Quanteda: an R package for the quantitative analysis of textual data [J]. *Journal of open source software*, 2018, 3 (20): 774.
- 13 PONWEISER M. Latent Dirichlet Allocation in R [D]. Vienna: Vienna University of Economics and Business, 2012.
- 14 GRÜN B, HORNIK K. Topicmodels: an R package for fitting topic models [J]. *Journal of statistical software*, 2011, 40 (13): 1–30.
- 15 R Core Team. The R project for statistical computing [EB/OL]. [2023–12–20]. <https://www.R-project.org/>.
- 16 WICKHAM H, FRANÇOIS R, HENRY L, et al. Dplyr: a grammar of data manipulation. [EB/OL]. [2023–12–20]. <https://CRAN.R-project.org/package=dplyr>.
- 17 WICKHAM H. Ggplot2: elegant graphics for data analysis [M]. New York: Springer – Verlag, 2016.
- 18 PEBESMA E. Simple features for R: standardized support for spatial vector data [J]. *The R journal*, 2018, 10 (1): 439–446.