

# AI in the social sciences, a review based on trend analysis

## Abstract:

随着人工智能技术的飞速发展，其在社会科学领域的交叉应用日益广泛，成为学术界关注的前沿方向。本文基于 Web of Science 核心数据库中 1970 至 2024 年间共 7011 条相关文献，结合主题建模、DataMapPlot、Voronoi Treemap 和 Parallel Tag Clouds 等多种可视化方法，系统梳理了人工智能与社会科学交叉研究的主题演变趋势与热点变迁。研究发现，自 2015 年起该领域研究热度迅速攀升，并于 2022 年达到高峰。通过 LDA 模型，提炼出涵盖技术演进、智能体认知、心理健康、媒体传播与信息治理等七个核心主题。本文构建了一个多维度、系统化的可视化分析框架，为人工智能与社会科学交叉领域的研究趋势提供了量化依据与理论支持，具有重要的学术价值与实践指导意义。

*Keywords:* 人工智能; 社会科学; 可视化分析; 主题建模; LDA; 文献计量

## 1.Introduction:

人工智能的高速发展为社会科学中的许多领域之中都带来了极大的便利，如今已经实现了许多社科与人工智能交叉下的应用场景，激增的应用场景带来的不仅仅是效率的提高，还有许多风险以及潜在的隐患需要我们尤为重视。本文想要从时间的角度入手，探究在社会科学与人机交叉研究中随时间变化的发展趋势，分析随着发展涌现的新兴主题并且结合数据对未来的研究方向进行预测，为未来的交叉研究提供方向参考。

过去的综述类文章多数着重于对应用以及风险进行观点的陈列，虽然有着较大的涵盖面，但是无法实现对趋势的分析，本文侧重于对主题演变的角度进行分析，通过分析主要主题来更好的了解交叉学科领域，通过分析主题的发展趋势进而预测未来的潜在研究方向，通过对主题进行细致的分析来挖掘新兴主题，并且预测这些主题可能对社会产生的影响，为未来的研究提供基于数据的多角度引导。

## 2. Method:

过去的文献综述中多数侧重于社会科学与人工智能交叉领域下应用以及风险进行整合，我们则进行了侧重于对趋势发展的研究，对文献进行了定量和定性分析，具体研究问题如下：

**RQ1:**在社会科学与人工智能的交叉研究中，有哪些子主题的发展使得社会对 Detection 子主题的关注度发生了转变，为什么会产生这些转变？有哪些子主题与 Detection 呈现出较高的相似性？

**RQ2:**在社会科学与人工智能交叉领域中，人工智能的发展如何影响了社会科学研究的文献内容，这些内容的转变是因为什么？

**RQ3:**在社会科学与人工智能交叉领域的发展过程中，Science Technology 主题与 Agents' Cognitive 主题之间是如何互相影响的，为什么 Agents' Cognitive 主题的热度在 2017 年后就开始下降并且在 2023 年开始回升，是由于哪些因素的影响产生了这种趋势？

为了对社会科学与人工智能交叉领域进行系统综述，我们在 WOS 核心数据库中进行了

特定词条的文献检索，检索后的数据包含了 1970-2024 年的文献，在这段时间之中，人们对人工智能与社会科学交叉下的重视性不断增强。将获取到的数据进行清洗后我们根据研究问题选择了几种数据可视化方法，分别是：DataMapPlot, Voronoi\_Treemap 和 Parallel Tag Clouds。通过可视化数据，我们可以更直观地发现人工智能与社会科学交叉下不同领域的影响力以及主题随时间产生的演化结果。

## 2.1 LDA 主题建模

清洗过检索数据后，我们首先利用主题建模对人工智能与社会科学交叉下的研究主题进行了分析。主题建模可以使文档通过计算的方式分类进不同的主题之中<sup>[1]</sup>，我们选用的模型为 Latent Dirichlet Allocation (LDA，潜在狄利克雷分布)<sup>[2]</sup>，LDA 是一种无监督的概率主题模型，能够从大规模文档集合中自动发现潜在主题结构，这种主题模型可以将单词依据共现将单词进行分组，每一个分组都会对应着一个主题，通过分析分组下的单词就可以提取出对应的主题，进而探究主题下的模式和趋势。我们进行的 LDA 主题建模使用的数据为：Web of Science 中依据检索词条 TS=(Social Science) AND TS=(Artificial Intelligence) and Preprint Citation Index (Exclude – Database) and Web of Science Core Collection (Database)导出的 1970-2024 文献共计 7011 条，进行数据清洗后保留了 6526 条数据，数据处理过程中，将清洗后数据进行分词并向量化保留文本特征，进行 LDA 分析和主题困惑度分析，依据困惑度分析的结果（图 1），最后确定的主题数为 22 个。在 LDA 建模基础上，我们通过 TF-IDF 提取特征词，并据此为每个主题命名。

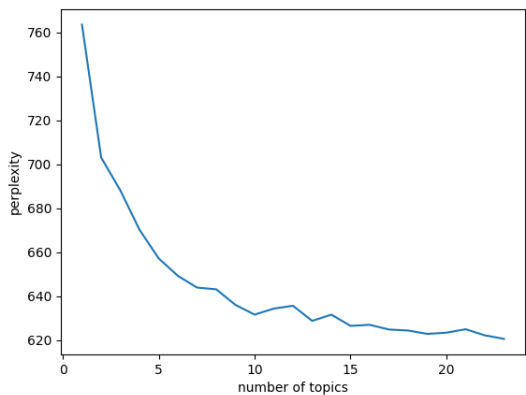


图 1：困惑度分析结果曲线

## 2.2 DataMapPlot 可视化

DataMapPlot 是一种基于密度的二维可视化方法<sup>[3]</sup>，能够有效展示高维数据的聚类结构。本研究首先使用 UMAP 算法将 LDA 主题模型的输出结果降维至二维空间，然后利用 DataMapPlot 生成可视化图表。该方法的优势在于能够直观显示不同主题的影响力大小（通过颜色区域面积表示）以及主题间的关联程度（通过颜色重叠程度表示）。我们可以从 DataMapPlot 中直观地发现人工智能与社会科学交叉下不同领域的影响力以及不同主题之间的关联。根据可视化分析的结果，我们发现了几个热门的主题：Human Agents Cognition , Healthcare , Ethical and Legal , Online Media 等，从主题交叉的角度进行分析也可以大致分为

两大类，分别为：技术发展和人文社科。

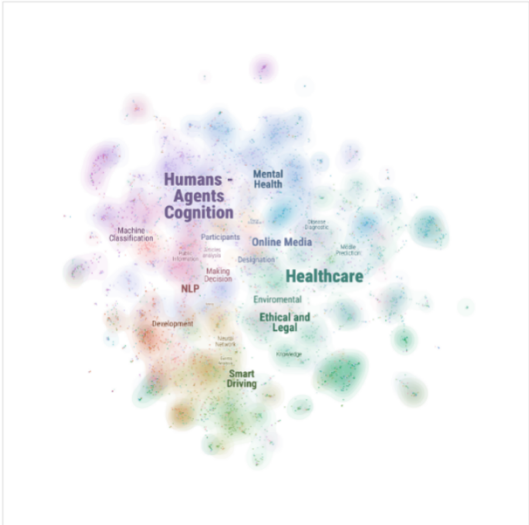


图 2: DataMapPlot 可视化结果

### 2.3 Voronoi\_Treemap 分析

针对研究领域分布分析，本研究采用 Voronoi\_Treemap 可视化方法我们保留了 WOS 数据库中 Research Area 列表中按照权重排序的前 150 个 Research Area，结合 Research Area 的数量以及其重复情况最终保留了前 25 个 Research Area，并且按照其权重分配进行了 Voronoi\_Treemap 的可视化分析（图 3），Voronoi\_Treemap 弥补了传统 Treemap 数据占比较小小时难以可视的问题，Voronoi\_Treemap 可视化可以直观的体现不同领域在交叉研究下的热度。



图 3: Voronoi\_Treemap 的可视化分析结果

### 2.4 Parallel Tag Clouds 分析

为了分析主题的发展趋势以及主题集的演变趋势，我们需要选择一种可以将时间和主题的分析方法，我们选择的可视化分析方法为 Parallel Tag Clouds，这是一种可以将标签权重，以及主题集内容进行可视化的方法，将横轴设计为时间可以更好的分析交叉领域下主题随时间的发展趋势。

在确定了研究领域后，我们缩小了检索的范围，对检索后的词条进行数据清洗后进行分词以及 LDA 建模，确定了建模结果随后对数据进行处理与整合，为数据分配对应的年份窗口并且基于 TF-IDF 为每一个窗口下出现的词分配权重，并为每一个词附加对应的 TopicPros（主题概率），将数据处理后进行 Parallel Tag Clouds 可视化分析，依据分析的结果重新设计停用词词典后得到最后最终的可视化结果（图 4、5）。

图 4 为主题集按照时间变化趋势的可视化结果，子主题下的颜色横柱代表着其下对应的每一个主题集的近似度，子主题会被分到相似度最高的主题集下。每一个主题集下有着若干个子主题，同一个主题集下的主题有着相同的颜色，我们可以根据每一个主题集下子主题的变化来分析出该主题集的演化趋势，进而进行对未来可能的研究趋势进行预测。

图 5 为主题集按照 TF-IDF 分析下的权重进行排序的可视化结果，年份下的子主题按照 TF-IDF 的分析结果进行排列，靠上的子主题为在横轴年份下有着较高影响力的主题，通过观察不同年份下子主题的排列顺序以及子主题的位置变化实现对交叉领域整体趋势的分析。

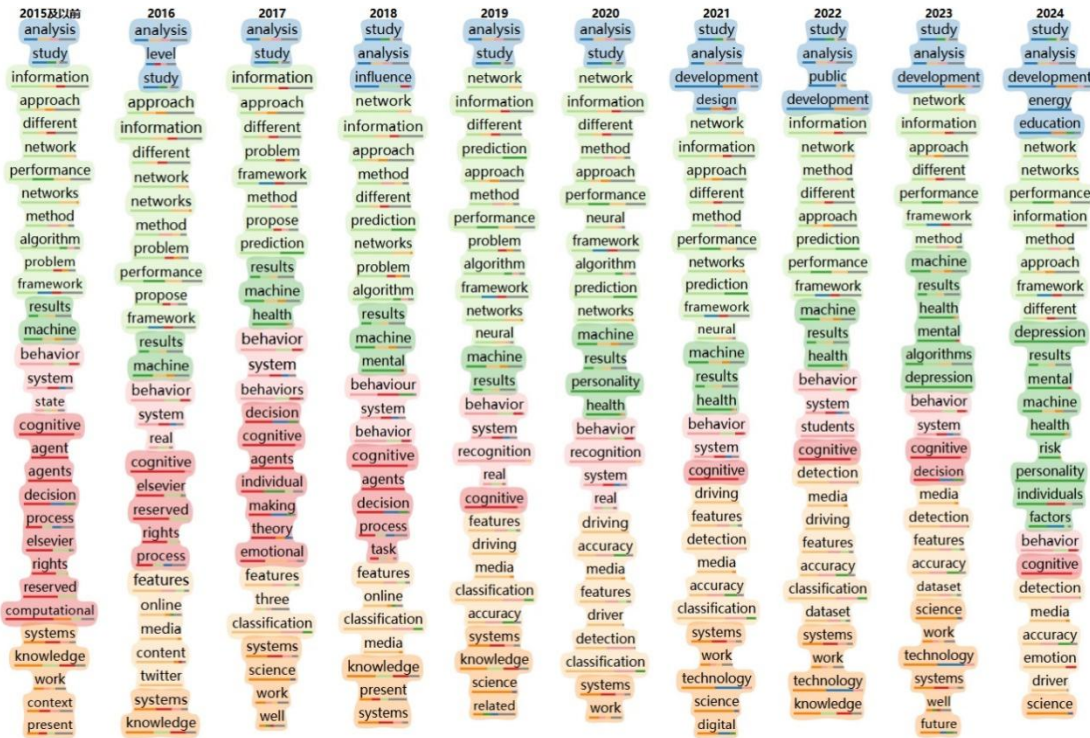


图 4：年份窗口下的主题排序（基于 Topic）





图 5：年份窗口下的权重排序（基于 TF-IDF）

### 3.Result:

人工智能的出现对社会产生了极大的影响，这些方面包括但不限于教育、医疗、经济等领域。要想了解生成式人工智能作为一种新兴技术如何影响我们的生活，我们首先要考虑到技术和社会之间的关系，技术的定义在不同的时期都有着不同的定义，Rudi Volti 在其撰写的书中提到对技术的定义：“A system created by humans that uses knowledge and organization to produce objects and techniques for the attainment of specific goals”<sup>[1]</sup>这个定义很好的阐释了技术与人之间的关系，技术是应人的需求而产生的，技术应着人的技术热情（technological enthusiasm）和应用需求而产生，技术的发展又会影响着人们的动力水平、工作方式、思维模式<sup>[5]</sup>。人们对自动化追求和对类人智能体的好奇心驱使了当今的人工智能热潮的产生，生成式人工智能作为一种技术工具，通过自然语言处理（NLP）和深度学习，人工智能可以进行复杂的数据分析、语言理解和生成任务，这些能力在社科中的研究有着巨大的潜力，这些能力使得社科中的许多领域有了进行更多种类、更高效研究的可能。

为了更好的探究人工智能与社会科学交叉领域下发展趋势，挖掘新型的主题，我们将结合数据展开论述。

#### 3.1 人工智能与社会科学交叉下的主题演变趋势

### 3.1.1 主题挖掘与分析

在主题挖掘与分析方面，我们首先进行了困惑度分析（图 7），在困惑度分析的结果下我们选择的主题数为 7，在主题建模的基础上，我们结合了针对时间窗口的分析，以便追寻数据在时间维度下产生的变化，进而探寻数据在时间下的趋势。

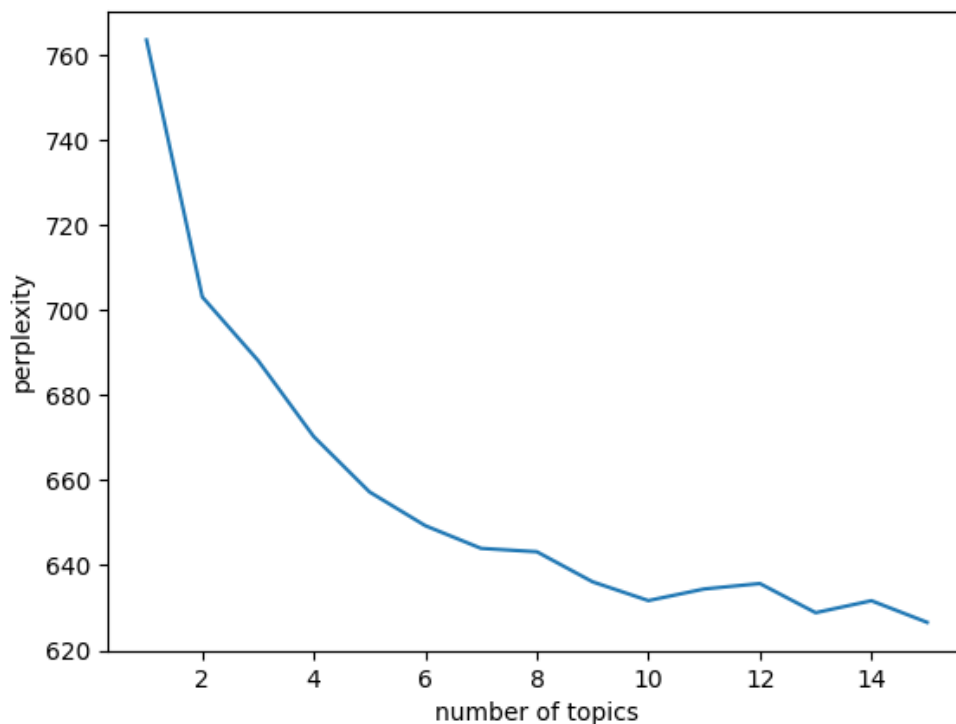


图 7：困惑度分析结果

通过对社会科学与人工智能交叉领域文献的可视化分析，我们发现 2015-2022 年间相关研究的发文量显著增长，并在 2022 年达到高峰。TF-IDF 分析揭示了“behavior”一词的持续高权重，预示着对人工智能行为的深入探讨和政策制定可能成为未来的焦点。同时，“study”、“detection”、“mental”和“depression”等词汇的权重上升趋势，反映了学界对心理健康、信息监管等领域的日益增加的关注度。

在进行过 LDA 主题建模之后，我们为社会科学和人工智能交叉下的七个主题分别进行了命名（表 1），主题按照主题强度依次被定义为“Science Technology”“ Agents’ Cognitive” “ Network information” “ Media and driving” “ Mental health” “ information Recognition(detection)” “ Development and Environment”，通过对这几个主题的分析，位次同时也象征着社会对于主题的重视程度，我们可以发现，在社会科学和人工智能领域下对技术层面的重视程度较高，并且对一些危机有着较高的警惕性，尤其是主题“Agents’ Cognitive”，这个主题在交叉领域下有着较高的发文量占比，除了偏向于技术的主题和智能体伦理道德相关的内容，还有着侧重于临床心理医学的主题“Mental health”和“Media and driving”，人工智能的出现为临床心理医学提供了全新的治疗辅助方式，为针对“depression”以及其他的心理疾病的治疗提供了有力的辅助。人工智能的出现也对媒体带来了极大的冲击，如何识别社交媒体上人工智能产生的虚假信息成为了急需解决的问题。

Topic	Name	Key Words
Topic1	Science Technology	science, technologies, systems, future, development, ethical
Topic2	Agents' Cognitive	cognitive, agent, decision, interaction, behavior
Topic3	Network information	network, information, method, prediction, neural, algorithm
Topic4	Media and driving	media, driving, detection, classification, twitter, machine
Topic5	Mental health	health, mental, depression, personality, risk, medical, clinical
Topic6	Information Recognition(detection)	behavior, recognition, activity, crowd, analysis, accuracy, video, real
Topic7	Development and Environment	Development, public, economic, value, energy, impact, sustainable

表 1: LDA 主题建模后的主题命名以及主题下的关键词

### 3.1.2 主题演化趋势分析

为了更好的探寻主题的演化过程，我们将数据分配到时间窗口下并进行 Parallel Tag Clouds 可视化，

通过分析按照主题的 Parallel Tag Clouds 可视化结果，我们可以发现一些随着年份而变化的主题趋势。在 1970-2024 年的时间范围内，Network information 主题一直稳定地占据着主要部分，且主题的影响力依旧在不断地扩大。“Media and driving”主题的占比在 2016-2024 年之间相对稳定，但是其内容有着一定方向的变化，早期的主题主要构成为“online media”，“classification”，到 2020 年开始构成逐渐转变为“detection”，“accuracy”，这意味着随着时间的演化，主题“Media and driving”重点关注的内容也发生了变化，因为人工智能本身的输出可能会包含偏见、歧视、毒性、虚假的信息<sup>[6][7]</sup>，而且人工智能在进行多模态转化的过程中还可能会出现信息失真的问题，Newsguard 在 23 年底利用 AI Tracking Center 标记发现了 614 个由 AI 生成的虚假新闻信息网站，这些网站扩散着有关政治、宗教、技术等方面的虚假信息<sup>[9]</sup>，类似的问题使得社会对媒体本身的关注转化为对媒介信息真实性的关注，这种趋势反映着社会在人工智能时代对信息真实性的担忧，也极大概率会成为“Media and driving”主题的一大段时间的重点所在。

在 1970-2024 年间有两个主题呈逐步发展的趋势，分别为“Development and Environment”和“Mental health”，“Development and Environment”主题的发展与当今世界可持续性发展的主题相一致，可以预测的是，在未来的交叉发展下，“sustainable”和“energy”会一直作为关键词围绕着主题“Development and Environment”。主题“Mental health”的发展趋势也与当今社会中逐渐突出的心理问题相符合，将人工智能融入临床心理学可以弥补过去临床治疗效率低，个性化难以实现的问题。Xiao-Wei Wang 等人实现了使用机器学习方法从脑电图数据中对情绪状态进行分类的研究<sup>[10]</sup>，Johanna Habicht 基于多个机器学习模型开发 AI 个性化自动聊天机器人 Limbic Access 可以进行预测和在诊断<sup>[11]</sup>，人工智能的出现从许多方面为临床心理学带来了前所未有的强力辅助。

“Agents' Cognitive”这一主题在 2015 年以前占据着相当大的主题比例，人们在 2015 年前对人工智能出现可能带来的道德伦理以及认知问题展开了广泛的讨论，在 2015 年之后该主题的关键词逐渐趋于统一，从 2019 年开始“cognitive”就开始作为该主题的代表词，与“Agents' Cognitive”发展趋势类似的还有“Information Recognition(detection)”这一主题，该主题的代表词为“behavior”，这也寓意虚假信息的检测与智能体的行为规范息息相关，归根结底问题在于如何实现智能体与人类的对齐，目前的一些主流模型已经可以从其输出中

体现其价值观倾向,我们应明确对齐目标,实现的对齐方面应包括指令遵从、理解人类意图、理解人类偏好、理解价值等方面,在“Information Recognition(detection)”主题下如何规范智能体的行为进而抑制生成虚假信息将成为未来的一大研究主题。

## 4. Discussion

### 4.1 主要发现与洞察

本研究通过对 1970-2024 年间 7011 篇相关文献的系统分析,揭示了人工智能与社会科学交叉领域的发展演化规律和未来趋势。研究发现具有以下几个重要特征:

时间演化特征显著:研究热度在 2015 年后呈现爆发式增长,并在 2022 年达到峰值,这一趋势与深度学习技术的突破和生成式 AI 的兴起密切相关。这种增长模式反映了学术界对 AI 技术社会影响认知的深化过程。

主题分布呈现层次化结构:七个核心主题按重要性依次为"Science Technology"、"Agents' Cognitive"、"Network information"、"Media and driving"、"Mental health"、"Information Recognition(detection)"、"Development and Environment"。这种分布表明技术发展与伦理关切并重,应用领域与风险防范并行的研究格局。

子主题的趋势变化体现社会关注重心:Detection 子主题的关注度持续上升,特别是在虚假信息检测、行为识别等方面,这反映了社会对 AI 生成内容真实性和可靠性的担忧日益加深。

### 4.2 趋势预测与发展方向

基于 TF-IDF 权重变化趋势和主题演化分析,我们预测未来研究将聚焦于以下几个方向:

1. 人机协作与信任机制: "behavior"关键词的持续高权重预示着对 AI 行为可解释性和人机交互信任机制的深入研究将成为重点。

2. 心理健康数字化干预: "mental"、"depression"等词汇权重的上升趋势表明, AI 在心理健康领域的应用将迎来快速发展期,个性化治疗和早期预警系统将成为关键突破点。

3. 信息真实性治理: 随着生成式 AI 的普及,虚假信息检测和内容认证技术的研究需求将持续增长,跨模态检测和实时监控技术将成为发展重点。

4. 可持续发展与社会影响评估: "Development and Environment"主题的兴起反映了学界对 AI 技术社会责任和可持续性的关注,未来将需要更多跨学科的影响评估研究。

### 4.3 对研究问题的回应

针对本研究提出的三个核心问题,我们得出以下结论:

RQ1 回应: Detection 子主题关注度的转变主要受到媒体技术发展、虚假信息传播和社交媒体监管需求的驱动。与之相似性较高的子主题为"Information Recognition",它们共同构成了信息治理的技术支撑体系。

RQ2 回应: AI 发展对社会科学研究内容的影响体现在研究方法的数字化转型、研究对象的扩展(从传统社会现象到人机交互行为)、以及研究范式的变革(从定性描述到量化预测)。



RQ3 回应: Science Technology 与 Agents' Cognitive 主题间存在螺旋式互动关系。技术进步推动了对 Agents' Cognitive 的关注,而伦理担忧又促进了技术的规范发展。Agents' Cognitive 主题在 2017 年后的下降反映了技术应用的成熟化,而 2023 年的回升则与 ChatGPT 等大模型引发的新一轮伦理讨论相关。

## 5. Conclusion

本研究为人工智能与社会科学交叉领域的发展提供了系统性的可视化分析框架,主要贡献包括:

将主题建模、时间窗口分析和多维可视化相结合,为交叉学科研究趋势分析提供了新的方法论工具使用组合,基于大规模文献数据,揭示了该领域的七个核心主题及其演化规律,为理解 AI 技术的社会影响提供了数据支撑。通过基于时间窗口可视化进行的趋势分析为未来研究方向提供了基于数据的预测,这有助于优化未来交叉学科研究的资源分配,并为未来的政策制定提供科学决策支持。

本研究有着一定的局限性,例如数据的来源为 WOS 核心数据库,这可能导致数据有着一定的地域和语言偏向,主题建模的主观性也可能会影响主题解释的准确性,以年份做为时间窗口可能会影响趋势分析的精度。

在未来的深入研究过程中,我们将通过扩大数据来源,纳入更多数据库来均衡数据可能带来的主观偏见,在交叉学科的研究中,我们也将结合最新的社科和人工智能理论研究来提高主题建模的准确性。

通过本研究,我们不仅描绘了人工智能与社会科学交叉领域的发展全景,更为未来的研究和实践提供了基于数据的科学指引。随着 AI 技术的持续演进,我们期待这一交叉领域能够在技术创新与社会责任之间找到更好的平衡点,为人类社会的可持续发展做出更大贡献。

## 6. Reference

- [1] Asmussen C B, Møller C. Smart literature review: a practical topic modelling approach to exploratory literature review[J]. Journal of Big Data, 2019, 6(1): 1-18.
- [2] Jelodar H, Wang Y, Yuan C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey[J]. Multimedia tools and applications, 2019, 78: 15169-15211.
- [3] DataMapPlot: <https://datamapplot.readthedocs.io/en/latest/>
- [4] Volti, Rudi, and Jennifer Croissant. Society and technological change. Waveland Press, 2024.
- [5] Au-Yong-Oliveira M, Gonçalves R, Martins J, et al. The social impact of technology on millennials and consequences for higher education and leadership[J]. Telematics and Informatics, 2018, 35(4): 954-963.
- [6] Goldstein J A, Sastry G, Musser M, et al. Generative language models and automated influence operations: Emerging threats and potential mitigations[J]. arXiv preprint arXiv:2301.04246, 2023.
- [7] Bianchi F, Kalluri P, Durmus E, et al. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale[C]//Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023: 1493-1504.

- [8] Rawte V, Sheth A, Das A. A survey of hallucination in large foundation models[J]. arXiv preprint arXiv:2309.05922, 2023.
- [9] Tracking AI-enabled Misinformation: 739 ‘Unreliable AI-Generated News’ Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools <https://www.newsguardtech.com/special-reports/ai-tracking-center/>
- [10] Wang, Xiao-Wei, Dan Nie, and Bao-Liang Lu. "Emotional state classification from EEG data using machine learning approach." *Neurocomputing* 129 (2014): 94-106.
- [11] Habicht J, Viswanathan S, Carrington B, et al. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot[J]. *Nature Medicine*, 2024: 1-8.