

视觉分析挑战与趋势：人工智能时代的 BigVis 社区视角 *

Nikos Bikakis[†]
Hellenic Mediterranean
University &
Archimedes/Athena RC
Greece

Panos K. Chrysanthis
University of Pittsburgh
USA

Guoliang Li
Tsinghua University
China

George Papastefanatos
Athena RC
Greece

Lingyun Yu
Xi' an Jiaotong-Liverpool
University
China

1 介绍

本报告提供了关于人工智能时代人类与数据交互及可视分析面临的挑战、新兴话题和机遇的见解。

BigVis 2024¹ 组织委员会在该领域的专家中进行了一项调查。他们邀请了程序委员会成员和已接受论文的作者分享他们的观点。来自不同研究社区，包括数据库、信息可视化和人机交互的三十二名科学家参与了这项研究。这些科学家代表了业界和学术界，提供了对当前和未来领域形势的宝贵见解。

在这份报告中，我们分析了调查回复，并将它们与四年前进行的一项类似研究 [2] 的发现进行了比较。结果显示了一些有趣的见解。首先，在之前的调查中识别出的许多关键挑战至今仍然高度相关，尽管与人工智能无关。同时，该领域的格局已经显著演变，今天大多数重要的挑战甚至没有在前期的调查中被提及，突显了与人工智能相关的进步所产生的深远影响。

通过总结研究社区的观点，本报告旨在阐明人工智能时代人类与数据交互及可视分析的关键挑战、新兴趋势和潜在研究方向。

2 调查概述

调查分为两个部分。第一部分与挑战（第 3 节）有关，第二部分侧重于新兴研究课题（第 4 节）。

*To appear in ACM SIGMOD Record 2025

[†]Corresponding author.

¹约第 7 位 国际大数据可视化探索与分析研讨会，在第 50 个 国际非常大数据库会议 (VLDB 2024)，中国广州。关于 BigVis 研讨会的更多详情请参见 [7]。

参与者被要求回答六个问题，要么通过填写自由文本字段，要么从提供的选项中进行选择。该调查是匿名的，因为与个人信息相关的问题是可选的，例如姓名、县、所属机构。完成这项调查平均需要大约三到五分钟。

参与者人口统计。我们打算从不同的研究社群（例如，数据库、信息可视化、人机交互）以及产业界和学术界中寻找科学家。为此，调查分发给了 BigVis 2024 程序委员会成员（58 名成员）和 BigVis 2024 录用论文的作者（34 位作者）。最终，有 32 被邀请的科学家中有完成了调查。

如下参与者的特点被收集起来（图 1）：

- **科学领域** (图 1a): 选项为: (a) 数据库; (b) 信息可视化; (c) 数据挖掘; (d) 人机交互; (e) 计算机图形学; 和 (f) 其他。大多数参与者属于信息可视化 (47%) 和数据库 (37%) 领域，而 16% 则属于其他研究领域。
- **职业** (图 1b): 选项为: 学术 (81%) 和行业 (19%)。
- **位置** (图 1c): 选项是: 教授 (59%); 研究人员 (28%); 和分析师/科学家/工程师 (13%)。

3 研究挑战

在本节中，我们概述了与数据可视化和可视分析相关的研究挑战的调查结果。

在第一部分（第 3.1 节），参与者被要求对四年前于 2020 年报告中出现的挑战在今天的紧迫性进行投票，该报告名为“大数据可视化与分析：未来研究挑战与新兴应用” [2]。在接下来的部分（第 3.2

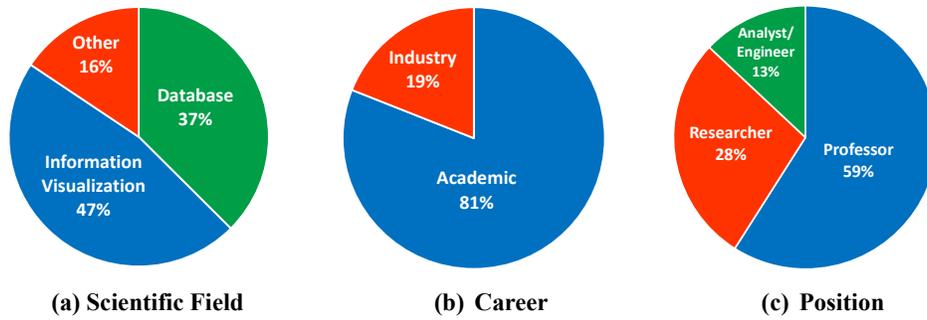


图 1: 参与者人口统计学特征

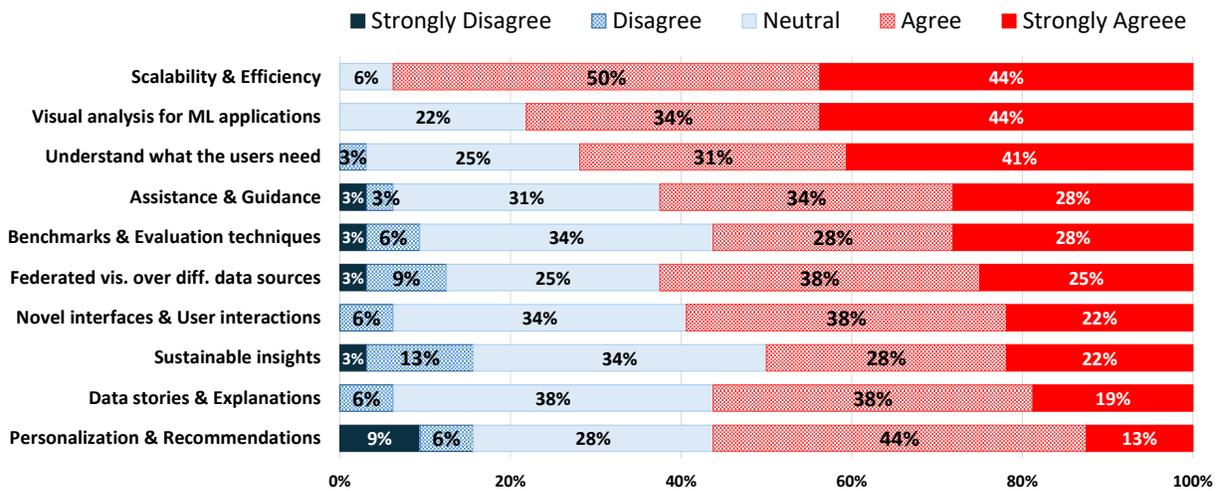


图 2: 2020 年挑战的重要性今天 [“这个挑战今天重要吗?”]

节), 参与者提出他们认为今天最重要的一个挑战, 不论这个挑战是否包含于 2020 年的挑战中。

3.1 2020 年报告的挑战

本节介绍了关于四年前在 2020 年报告中确定的十大挑战今日重要性的调查结果。特别地, 在第三届大数据可视化探索与分析国际研讨会 (BigVis 2020) 的背景下, 组织委员会邀请了 14 位杰出的科学家来自不同领域的专家提供他们对挑战以及他们在未来几年中认为更有趣的领域, 与大数据可视化和分析相关的见解。

2020 年报告中指出的挑战为: (a) 支持可扩展性与效率; (b) 启用可视化分析以用于机器学习应用; (c) 理解用户的需求; (d) 构建新型界面及用户交互; (e) 协助与指导; (f) 生成数据故事与解释; (g) 启用跨

不同数据源的联合可视化; (h) 开发基准与评估技术; (i) 提供可持续的见解; 和 (k) 启用个性化及推荐。

Question 1

参与者被要求对 2020 年报告中基于其重要性/新兴程度所述的十项挑战进行投票。特别地, 参与者使用五级李克特量表 (即强烈不同意、不同意、中立、同意和强烈同意) 对问题 “这个挑战今天重要吗?” 中的每个挑战进行了评分。

问题 1 的回答。 结果通过百分比堆积条形图在图 2 中展示。参与者投票认为 “可扩展性与效率” 是最具挑战性的, 有 94% 的参与者表示同意或非常同意, 而 0% 的参与者表示不同意或非常不同意。第二重要的挑战是 “视觉分析在机器学习应用中的作用”, 有

78% (相应地 0%) 的参与者表示同意或强烈同意 (相应地不同意或强烈不同意)。“可持续见解”是最让大多数参与者不同意 (13%) 或强烈不同意 (3%) 的挑战。最后,“个性化与推荐”被投票为最不重要的挑战。

值得一提的是对于所有挑战中,至少 50% 的参与者强烈同意或同意今天该挑战的重要性。当考虑重要性评分时,可以观察到类似的结果。

3.2 2024 年调查的挑战

本节介绍了参与者认为最重要的挑战,而不考虑这些挑战是否包含在 2020 年的挑战中。

Question 2

参与者被要求在自由文本中提供他们认为对未来几年最重要的挑战,并附上简要说明。

问题 2 的回答。参与者为指示 16 挑战。挑战在表 1 中呈现;某些挑战中出现的括号内的数字表示提到这一挑战的参与者数量。此外,红色字体突出显示了在本调查中首次提到的挑战,即 2020 年调查中未提及的挑战。

最常见的挑战是“大型语言模型在可视化和分析中的应用”(由 16% 的参与者投票),而“公平与可信性”以及“非专家用户的可视化”则是接下来最常见的(各自由 13% 的参与者投票)。请注意,在四年前的挑战中并未提及 LLMs、公平性或可信度这些术语。此外,值得注意的是,作为最常被提到的挑战之一出现的可解释性,在 2020 年的报告中也没有被提及。

其他常见的挑战(至少有两位参与者提到)与以下内容相关:“用户辅助与指导”;“理解用户的需求”;“高维与流数据”;“逐步数据分析与可视化”;和“沉浸式可视化”。在这些挑战中,“高维与流数据”、“逐步分析”和“沉浸式可视化”首次出现。

4 新兴话题

在本节中,我们展示了参与者对于大数据可视化与分析领域最前沿的研究主题的回答。候选名单包括了 BigVis 征稿启事中的感兴趣的课题。

表 1: 调查挑战 *

利用大语言模型⁽⁵⁾, 确保公平与可信度⁽⁴⁾, 启用非专家用户的可视化功能⁽⁴⁾, 提供协助与指导⁽²⁾, 生成解释⁽²⁾, 理解用户需求⁽²⁾, 处理高维及流数据⁽²⁾, 启用渐进数据分析与可视化⁽²⁾, 开发沉浸式可视化系统⁽²⁾, 提供可持续见解, 支持数据抽象, 实现新颖的可扩展接口, 设计上下文特定的可视化, 制定基本的可视化问题, 使用代理建模, 开发基于能耗的解决方案

* c^(x): x 表示在调查中提到挑战的参与者数量。 。红色字体: 当前调查中首次提到的挑战 (仅考虑至少两位参与者指出的)

Question 3

参与者被要求执行从候选主题列表中选择(投票)最多三个他们认为最具前瞻性的主题。

问题 3 的回应。图 3 显示了参与者对每个主题投票的百分比。获得最多选票的主题是“人机交互处理”;“交互式与以人为中心的机器学习”;和“渐进分析”,其中 31% 的参与者选择了这些主题。另一方面,投票较少的主题是:“科学可视化”;“面向设置的可视化”;和“分布式与并行技术”,这些主题获得了 3% 的参与者投票。

5 讨论

首先,调查突出了四年前识别的所有挑战的重要性已被广泛认可(问题 1),至少有一半的参与者强烈同意或同意其今日的重要性。

关于当前挑战(问题 2)的结果揭示了与 AI 相关的问题的重要性。值得注意的是,今天提到的最多的问题四年前完全不存在。例如,与大语言模型、公平性和可信度以及解释相关的难题是一些新出现的领域。此外,“非专家用户”、“高维与流数据”、“逐步分析”和“沉浸式可视化”的挑战也首次出现。

进一步比较响应显示,2020 年最重要的挑战“可扩展性与效率”(问题 1)在问题 2 中未被提及。一个可能的解释是,近 95% 的参与者已经在问题 1 中将其评为(非常)重要,从而减少了再次强调它的必

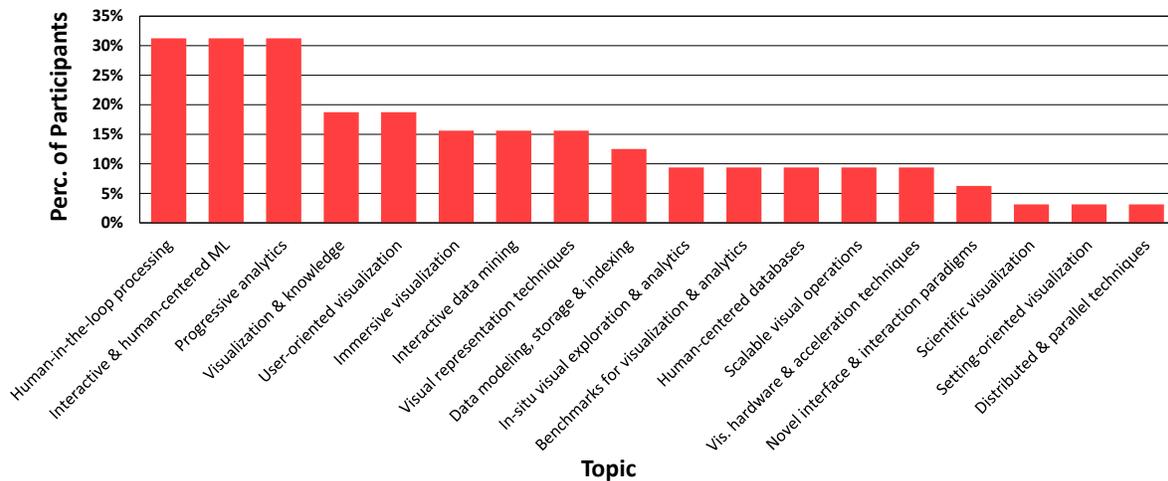


图 3: 新兴话题: 每个话题的参与者投票比例

要性。类似地,“可视化分析在机器学习应用中的运用”,即 2020 年第二重要的挑战,在问题 2 的响应中缺席,尽管它被认定为最新兴的话题之一(问题 3)。

当前挑战及 state-of-the-art 方法的进一步讨论可以在 [1, 3-6, 8-15] 中找到。

References

- [1] Sihem Amer-Yahia, Leilani Battle, Yifan Hu, Dominik Moritz, Aditya Parameswaran, Nikos Bikakis, Panos K. Chrysanthos, Guoliang Li, George Papastefanatos, and Lingyun Yu. 2025. Data Exploration and Visual Analytics Challenges in AI Era. ACM SIGMOD Blog, <https://wp.sigmod.org/?p=3820>.
- [2] Gennady L. Andrienko, Natalia V. Andrienko, Steven Mark Drucker, Jean-Daniel Fekete, Danyel Fisher, Stratos Idreos, Tim Kraska, Guoliang Li, Kwan-Liu Ma, Jock D. Mackinlay, Antti Oulasvirta, Tobias Schreck, Heidrun Schumann, Michael Stonebraker, David Auber, Nikos Bikakis, Panos K. Chrysanthos, George Papastefanatos, and Mohamed A. Sharaf. 2020. Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications. In Workshop on Big Data Visual Exploration & Analytics (BigVis 2020).
- [3] Natalia V. Andrienko, Gennady L. Andrienko, Linara Adilova, Stefan Wrobel, and Theresa-Marie Rhyne. 2022. Visual Analytics for Human-Centered Machine Learning. CG&A (2022).
- [4] Rahul C. Basole and Timothy Major. 2024. Generative AI for Visualization: Opportunities and Challenges. TVCG (2024).
- [5] Leilani Battle and Carlos Scheidegger. 2020. A structured review of data management technology for interactive visualization and analysis. IEEE TVCG 27, 2 (2020).
- [6] Nikos Bikakis. 2022. Big Data Visualization Tools. In Encyclopedia of Big Data Technologies, 2nd Ed. Springer.
- [7] Nikos Bikakis, George Papastefanatos, Panos K. Chrysanthos, Olga Papemmanuil, David Auber, Steffen Frey, Issei Fujishiro, Hanna Hauptmann, Shixia Liu, Kwan-Liu Ma, Tobias Schreck, Michael Sedlmair, and Mohamed A. Sharaf. 2024. Visualizing, Exploring and Analyzing Big Data: A 6-Year Story. ACM SIGMOD Record 53, 2 (2024).
- [8] Muhammad Raees, Inge Meijerink, Ioanna Lykourantzou, Vassilis-Javed Khan, and Konstantinos Papangelis. 2024. From Explainable to Interactive AI: A Literature Review on Current Trends in Human-AI Interaction. J. Hum. Comput. Stud. (2024).
- [9] Junpeng Wang, Shixia Liu, and Wei Zhang. 2024. Visual Analytics for Machine Learning: A Data Perspective Survey. IEEE TVCG 30, 12 (2024).
- [10] Qianwen Wang, Zhutian Chen, Yong Wang, and Huamin Qu. 2022. A Survey on ML4VIS: Applying Machine Learning Advances to Data Visualization. IEEE TVCG 28, 12 (2022).
- [11] Aoyu Wu, Dazhen Deng, Min Chen, Shixia Liu, Daniel A. Keim, Ross Maciejewski, Silvia Miksch, Hendrik Strobel, Fernanda B. Viégas, and Martin Wattenberg. 2023. Grand Challenges in Visual Analytics Applications. IEEE CG&A (2023).
- [12] Aoyu Wu, Yun Wang, Xinhuan Shu, Dominik Moritz, Weiwei Cui, Haidong Zhang, Dongmei Zhang, and

- Huamin Qu. 2022. AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization. IEEE TVCG 28, 12 (2022).
- [13] Weikai Yang, Mengchen Liu, Zheng Wang, and Shixia Liu. 2024. Foundation models meet visualizations: Challenges and opportunities. *Comput. Vis. Media* 10, 3 (2024).
- [14] Yilin Ye, Jianing Hao, Yihan Hou, Zhan Wang, Shishi Xiao, Yuyu Luo, and Wei Zeng. 2024. Generative AI for visualization: State of the art and future directions. *Vis. Informatics* 8, 1 (2024).
- [15] Jun Yuan, Changjian Chen, Weikai Yang, Mengchen Liu, Jiazhi Xia, and Shixia Liu. 2021. A survey of Visual Analytics Techniques for Machine Learning. *Comput. Vis. Media* 7, 1 (2021).