Visual Analytics Challenges and Trends in the Age of AI: The BigVis Community Perspective *

Nikos Bikakis
Hellenic Mediterranean Univ., Greece

Panos K. Chrysanthis
Univ. of Pittsburgh, USA

Guoliang Li
Tsinghua Univ., China

George Papastefanatos
ATHENA Research Center, Greece

Lingyun Yu
Xi'an Jiaotong–Liverpool Univ., China

1. INTRODUCTION

This report provides insights into the challenges, emerging topics, and opportunities related to human–data–interaction and visual analytics in the AI era. The BigVis 2024[1] organizing committee invites the PC members and the authors of accepted papers to participate in a survey. Thirty-two scientists from diverse research communities, including Databases, Information Visualization, and Human–Computer–Interaction, working in both industry and academia, share their perspectives. The 7th International Workshop on Big Data Visual Exploration and Analytics (BigVis 2024) is organized in conjunction with the 50th International Conference on Very Large Databases (VLDB 2024) in Guangzhou, China.

The survey feedback reveals interesting insights. First, several critical challenges identified in a similar survey four years ago remain highly relevant today, even though they are unrelated to AI. At the same time, most of today's vital challenges were not even discussed in the survey conducted four years ago, underscoring the profound impact of AI–related advancements.

The report summarizes community's insights regarding the challenges and the emerging topics related to human–data–interaction and visual analytics in AI era.

2. SURVEY OVERVIEW

The survey is divided into two parts. The first is related to challenges (Sect. 3), and the second focuses on emerging research topics (Sect. 4).

The participants were requested to answer six questions, either by filling free text fields or by selecting among the options provided. The survey was anonymous, since the questions related to personal information are optional, e.g., name, county, affiliation. The survey required, on average, about 3 to 5 minutes to be completed.

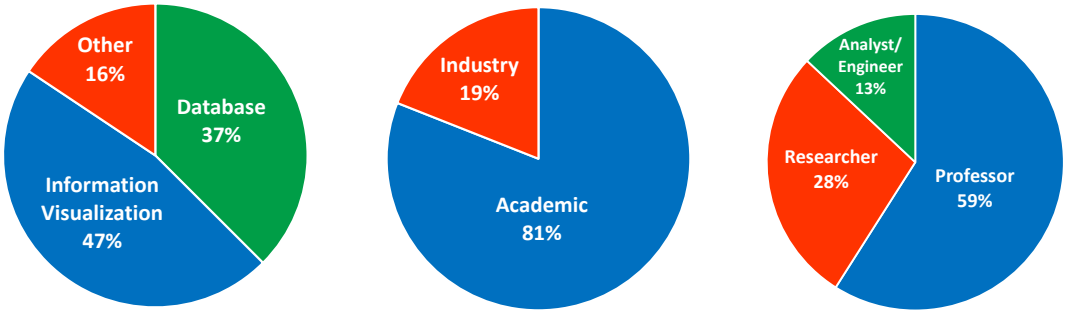
Participants Demographics. We intended to find scientists from different research communities (e.g., Databases, Information Visualization, HCI), and from industry and academia. To this end, the survey has been disseminated to BigVis 2024 Program Committee members (58 members) and to the authors of the BigVis 2024 accepted papers (34 authors). At the end, *32 of the scientists invited completed the survey*.

The following *characteristics of the participants* are collected (Figure 1):

- **Scientific Field** (Figure 1a): The options were: (a) *Database*; (b) *Information Visualization*; (c) *Data Minning*; (d) *Human–Computer–Interaction*; (e) *Computer Graphics*; and (f) *Other*. Most of the participants belong to Information Visualization (47%) and Database (37%) communities, while 16% belong to others research fields.
- **Career** (Figure 1b): The options were: *Academic* (81%) and *Industry* (19%).
- **Position** (Figure 1c): The options were: *Professor* (59%); *Researcher* (28%); and *Analyst/Scientist/ Engineer* (13%).

[1] *https://bigvis.imsi.athenarc.gr/bigvis2024*

*** Under Submission**



(a) Scientific Field **(b) Career** **(c) Position**

Figure 1. Participants Demographics

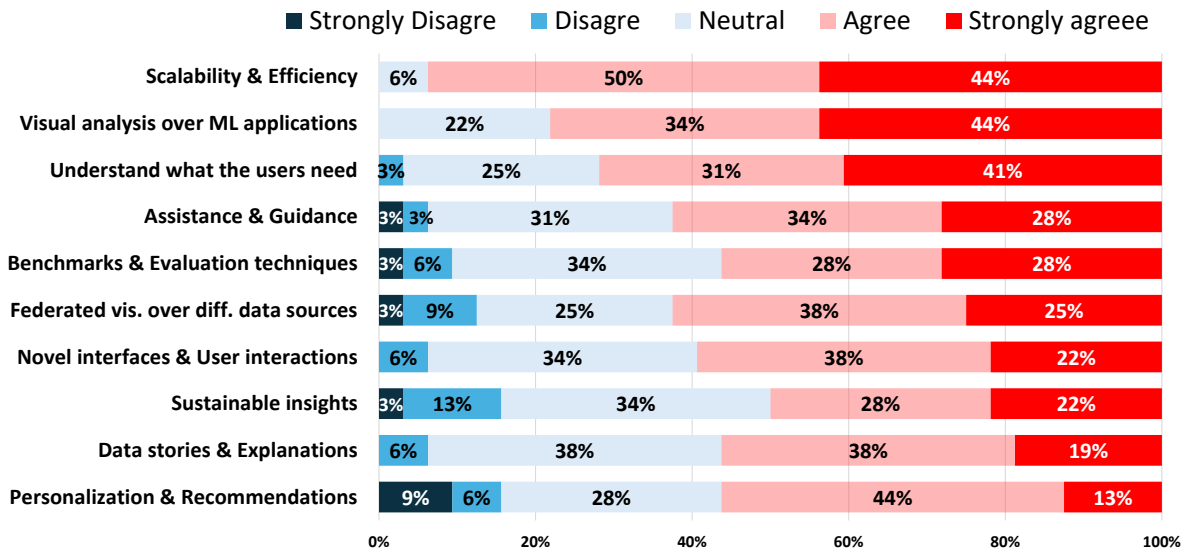


Figure 2. The Importance of the 2020 Challenges Today ["*Is this challenge important today?*"]

3. REASEARCH CHALLENGES

In this section we outline the survey's results regarding research challenges related to data visualization and visual analytics.

In the first part (Sect. 3.1), the participants were asked to vote on today's importance of the challenges stated four years ago in a 2020 report, titled "*Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications*"[2].

In the next part (Sect. 3.2), participants suggested a challenge they consider the most important today, regardless of whether it was included in 2020 challenges.

3.1. The challenges stated in the 2020 report

This section presents the results of the survey regarding today's importance of the ten challenges identified four years ago in the 2020 report. Particularly, in the context of the 3rd International Workshop on Big Data Visual Exploration and Analytics (BigVis 2020), the organizing committee invited 14 *distinguished scientists*, from different communities to provide their insights regarding the *challenges and the applications they find*

[2]Andrienko G., Andrienko N., Drucker S., Fekete J.-D., Fisher D., Idreos S., Kraska T., Li G., Ma K.-L., Mackinlay J.D., Oulasvirta A., Schreck A., Schmann H., Stonebraker M., Auber D., Bikakis N., Chrysanthis P.K., Papastefanatos G., Sharaf M.: "*Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications*", In 3rd Intl. Workshop on Big Data Visual Exploration & Analytics (BigVis 2020)

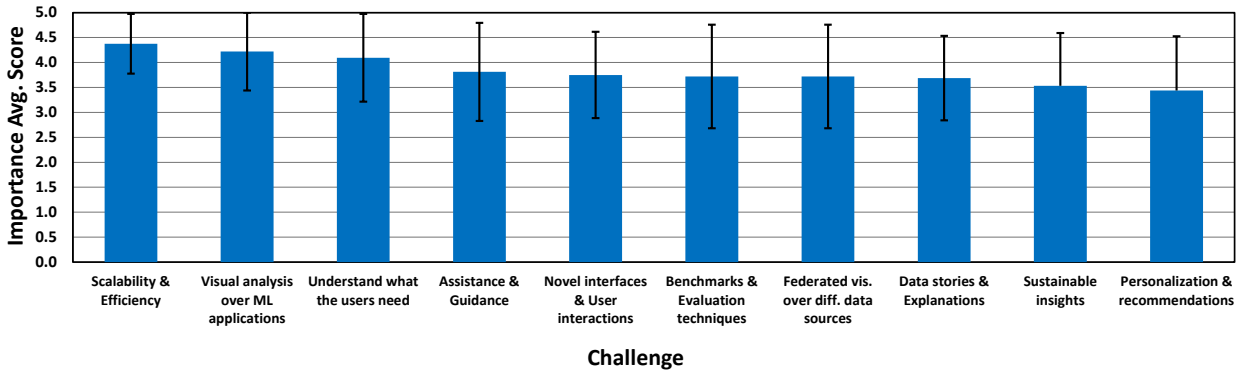


Figure 3. Average Importance Score & Standard Deviation [*Strongly Disagree* score=1 ~ *Strongly Agree* score=5]

more interesting in coming years, related to the areas of Big Data visualization and analytics. *The challenges indicated in the 2020 report were*: (a) *Support scalability & efficiency*; (b) *Enable visual analysis over ML applications*; (c) *Understand what the users need*; (d) *Build novel interfaces & user interactions*; (e) *Assistance & guidance*; (f) *Generate data stories & explanations*; (g) *Enable federated visualization over different data sources*; (h) *Develop benchmarks & evaluation techniques*; (i) *Provide sustainable insights*; and (k) *Enable personalization & recommendations*.

Question 1. The participants were asked *to vote on the 10 challenges stated in the 2020 report, based on their importance/emerge*. Particularly, the participants voted using a five-level Likert scale (i.e., Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree) on the question "*Is this challenge important today?*".

Question 1 Responses. The results are presented in Figure 2 via a percent stacked bar chart. The participants vote as the most important challenge the "*Scalability & efficiency*", with 94% of the participants state that agree or strongly agree, and 0% disagree or strongly disagree. The second most significant is the "*Visual analysis over ML applications*" challenge, with 78% (resp. 0%) of the participants state that agree or strongly agree (resp. disagree or strongly disagree). "*Sustainable insights*" is the challenge in which most of the participants disagree (13%) or strongly disagree (3%). Finally, "*Personalization & recommendations*" is voted as the least important challenge.

It is worth mentioning that *for all challenges, at least 50% of the participants strongly agree or agree on the today's importance of the challenge*.

Similar results can be observed when considering importance scores. Figure 3 shows the average importance score and standard deviation, with voting scores ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Table 1. Survey Challenges ⋆

Exploit LLMs$^{(5)}$, Ensure fairness & trustworthiness$^{(4)}$, Enable visualization for non-expert users$^{(4)}$, Offer assistance & guidance$^{(2)}$, Generate explanations$^{(2)}$, Understand what the users need$^{(2)}$, Handle high dimensional & stream data$^{(2)}$, Enable progressive data analysis & visualization$^{(2)}$, Develop immersive visualization systems$^{(2)}$, Provide sustainable insights, Support data abstraction, Implement novel scalable interfaces, Design context-specific visualizations, Formulate fundamental visualization problems, Use surrogate modeling, Develop energy consumption-based solutions

⋆ c$^{(x)}$: x indicates the number of participants that mention the challenge c in the survey. Red font: Challenges mentioned for the first time in the current survey (considering only those indicated by at least two participants).

3.2. The challenges stated by the survey's participants

This section presents the challenges stated by the participants as the most important, regardless of whether they were included in 2020 challenges.

Question 2. The participants were asked to provide in a free text *the challenge that they consider the most important for the next years*, along with a brief description.

Question 2 Responses. The participants *indicate 16 challenges*. The challenges are presented in Table 1; the number in the parentheses that appears in some challenges indicates the number of participants that mention this challenge. Furthermore, the red font is used to highlight challenges that are mentioned for the first time in this survey, i.e., challenges that were not indicated in the 2020 survey.

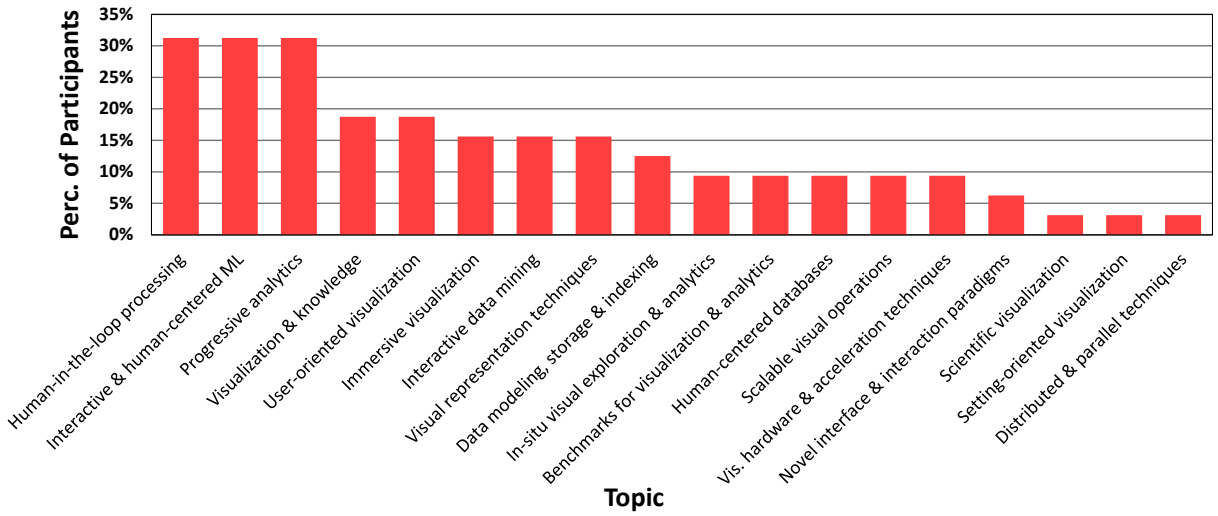


Figure 4. Emerging Topics: The Percentage of Participants that Vote on each Topic

The most commonly suggested challenge is the *Use of LLMs in visualization and analytics*, whereas *Fairness and Trustworthiness*, as well as *Visualization for non-expert users* are the next most common. Note that none of the terms *LLMs*, *fairness* or *trustworthiness* is mentioned in the four years ago challenges. Also note that *explainability*, which is one of the most common challenges, is also not mentioned in the 2020 report.

Other common challenges (mentioned by at least two participants) are related to *User assistance & guidance*; *Understanding what the users need*; *High dimensional & stream data*; *Progressive data analysis & visualization*; and *Immersive visualization*. Among these challenges, High dimensional & stream data, Progressive analysis, and Immersive visualization appeared for the first time.

4. EMERGING TOPICS

In this section we present the participants' responses regarding the most emerging research topics in the Big data visualization and analytics field. The candidate topics list consists of the topics of interest included in the BigVis call-for-papers.

Question 3. The participants were asked *to select (vote) from a list of candidate topics, up to three topics that they consider the most emerged*.

Question 3 Responses. Figure 4 shows the percentage of the participants' vote for each topic. The topics with the most votes are *Human–in–the–loop processing*; *Interactive & human–centered ML*; and *Progressive analytics*, where 31% of the participants were selected. On other hand, the topics with the less votes are: *Scientific visualization*; *Setting-oriented visualization*; and *Distributed & parallel techniques*, which are voted by 3% of the participants.

5. DISCUSSION

First, the survey highlights the broad acceptance of the importance of all the challenges identified four years ago (Question 1), with at least half of the participants strongly agreeing or agreeing on their today's importance.

The results regarding the current challenges (Question 2) revealed the importance of AI–related problems. Notably, the most commonly mentioned challenges today were entirely absent four years ago. For example, problems related to *LLMs*, *Fairness & trustworthiness*, and *Explanations* are some of the newcomers. Furthermore, challenges such as *Non-expert users*, *High dimensional & stream data*, *Progressive analysis*, and *Immersive visualization* appeared also for the first time.

Further comparison of the results shows that the "*Scalability & efficiency*" challenge, which was voted the most important 2020 challenge (Question 1), is not mentioned in Question 2. One possible explanation is that most participants (about 95%) already rated it as (very) important in Question 1 and therefore did not mention it again. A similar case is the "*Visual analysis over ML applications*" challenge, which was voted the second most important challenge of 2020 but is not included in the responses to Question 2. Note that this challenge was also voted as one of the most emerging topics.