BIG DATA VISUALIZATION THROUGH AI AND ANALYTICAL TECHNIQUES

EMIRANDA LOKA, ALDA KIKA

Department of Computer Science, Faculty of Natural Sciences,

University of Tirana, Albania

e-mail: emiranda.loka@fshn.edu.al

Abstract

The rise of massive, heterogeneous, and rapidly evolving datasets has amplified the need for intelligent visualization systems capable of supporting insight extraction, interpretability, and decision-making at scale. Traditional visualization approaches face limitations when confronted with high-dimensional, unstructured, or streaming big data. Recent developments in artificial intelligence (AI), machine learning (ML), and analytical modeling offer new techniques to automate and enhance data visualization pipelines. This paper presents a comprehensive review of big data visualization augmented by AI and analytical techniques, emphasizing recent advances from 2022–2025, including generative AI for visualization automation, Explainable AI (XAI) integration, human–AI collaboration, and scalable ML-driven frameworks. Challenges and future directions are also identified, linking technological progress with emerging needs in interpretability, scalability, and ethical visualization practices.

Key words: *Big Data, Data Visualization, Artificial Intelligence (AI), Machine Learning (ML).*

Përmbledhje:

Në ditët e sotme, ndeshemi me një rritje të vazhdueshme, shumëllojshmëri dhe evolim me ritme të shpejta të të dhënave, që ka sjellë nevojën për sisteme inteligjente të vizualizimit, të cilat shërbejnë për të mbështetur nxjerrjen e njohurive, për të krijuar mundësinë për interpretim dhe në ndihmë të proceseve të vendimmarrjes në shkallë të gjerë. Metodat tradicionale të vizualizimit përballen me kufizime të konsiderueshme kur bëhet fjalë për të dhëna të mëdha me dimensione të larta, të pastrukturuara ose në rrjedhë të *vazhdueshme (streaming). Zhvillimet e fundit në fushën e inteligjencës artificiale (IA), Machine Learning (ML) dhe të modelimit analitik ofrojnë teknika të reja për automatizimin dhe përmirësimin e zinxhirëve të vizualizimit të të dhënave.*

Ky punim paraqet një rishikim gjithëpërfshirës të vizualizimit të të dhënave të mëdha të pasuruar me teknika të IA-së dhe analizës, duke vënë theksin në përparimet më të fundit të periudhës 2022–2025, përfshirë përdorimin e inteligjencës artificiale gjenerative për automatizimin e vizualizimit, integrimin e Inteligjencës Artificiale të Shpjegueshme (Explainable AI – XAI), bashkëpunimin njeri–IA dhe kornizat e shkallëzueshme të drejtuara nga ML. Gjithashtu, identifikohen sfidat kryesore dhe drejtimet e ardhshme të kërkimit, duke lidhur progresin teknologjik me nevojat në rritje për interpretueshmëri, shkallëzim dhe praktika etike në vizualizimin e të dhënave.

Fjalë kyçe: *Big Data, Vizualizimi i të dhënave, Inteligjenca Artificiale (IA), Machine Learning (ML).*

Introduction

The unprecedented growth of data generated across diverse application domains, such as healthcare, smart cities, finance, cybersecurity, and social systems, has fundamentally reshaped the landscape of data analysis and decision-making. Modern data environments are no longer defined solely by their scale, but also by their complexity, heterogeneity, and dynamic nature. As data volumes continue to increase at an exponential rate, organizations and decision-makers face significant challenges in extracting meaningful, timely, and trustworthy insights using traditional analytical and visualization approaches. These challenges highlight the urgent need for more advanced and intelligent methods capable of supporting effective data understanding in increasingly complex analytical contexts (Neri et al., 2025; Wang et al., 2024).

Within this evolving landscape, data visualization has emerged as a critical mechanism for bridging the gap between complex computational processes and effective analytical reasoning. Visualization enables analysts to explore data interactively, identify patterns and anomalies, and develop mental models that support reasoning and sense-making. However, the growing scale and multidimensionality of big data place substantial strain on conventional visualization techniques, which were primarily designed for static, lowdimensional, and well-structured datasets. As a result, traditional visualization tools often struggle to provide adequate support for understanding complex analytical outputs, particularly when advanced machine-learning and artificial intelligence (AI) models are involved.

Recent research demonstrates a clear and growing trend toward the integration of visualization with AI-assisted systems in order to enhance decision-making, improve model interpretability, and reduce cognitive load for users. In this context, visualization is increasingly recognized not merely as a presentation tool, but as an integral component of the analytical process itself. Neri et al. (2025) emphasize that visualization now plays a central role in AI-assisted decision systems, functioning as a critical interpretability layer that connects computational models with analytical processes. By translating complex model behaviors and predictions into comprehensible visual representations, visualization supports transparency, trust, and informed decision-making in AI-driven environments (Neri et al., 2025).

Similarly, advances in Explainable Artificial Intelligence (XAI) have significantly contributed to the evolution of visual analytics. XAI techniques aim to make machine-learning models more understandable by revealing the underlying logic, features, and decision pathways that drive model outputs. When combined with visualization, these techniques enable users to gain clearer and more intuitive explanations of model behavior, thereby enhancing both interpretability and usability. Yin et al. (2024) demonstrate that visualization-based XAI approaches play a crucial role in supporting user understanding, particularly in complex and high-stakes analytical scenarios where model transparency is essential.

Beyond explainability, AI-driven methods are increasingly being employed to address scalability and adaptability challenges in big data visualization. Machine learning and analytical techniques can assist in tasks such as data abstraction, feature selection, anomaly detection, and automated visual encoding, allowing visualization systems to cope more effectively with large-scale, high-dimensional, and streaming data. These AI-augmented visualization approaches not only improve system performance but also reduce the cognitive burden placed on users by guiding attention toward the most relevant patterns and insights.

Importantly, the integration of AI into visualization systems raises new questions related to collaboration between analytical models and users, system trust, and ethical responsibility. As analytical processes become increasingly automated, it is essential that visualization systems preserve transparency, interpretability, and user control over data-driven decisions (Neri et al., 2025; Yin et al., 2024). Recent scholarly work further emphasizes the importance of designing visualization systems that support clear interaction and analytical reasoning, rather than focusing solely on computational optimization (Beschi et al., 2025; Wang et al., 2024).

Against this background, this paper examines how artificial intelligence and advanced analytical techniques are reshaping the field of big data visualization.

The objective is to provide a comprehensive and up-to-date perspective on the evolving relationship between AI and visualization, with particular emphasis on scalability, interpretability, and effective support for data-driven analysis (Neri et al., 2025; Wang et al., 2024). To achieve this goal, the paper makes the following key contributions:

i. A synthesis of the fundamental requirements and challenges associated with big data visualization in contemporary data-intensive environments

ii. An overview of AI-augmented visualization approaches, including methods inspired by machine learning, XAI, and generative AI

iii. An analysis of analytical techniques that support scalable, adaptive, and interactive visual systems

iv. An integration of recent scholarly developments published between 2022 and 2025, reflecting the current state of the field

v. A research roadmap outlining open challenges and future directions for AI-driven visualization research

By consolidating recent advances and identifying emerging trends, this work aims to contribute to a deeper understanding of how visualization and AI can be effectively combined to support sense-making, decision-making, and knowledge discovery in complex data environments.

Methodology of the Literature Review

To guide the review, the following research questions (RQs) were defined, aligned with the study's objective:

RQ1: Why are traditional visualization approaches insufficient for insight extraction in big data environments?

RQ2: What challenges does big data visualization face today?

RQ3: Which AI-based techniques are currently integrated into big data visualization?

RQ4: In which domains is AI-driven big data visualization most?

These RQs are used for driving the extraction of related studies from different databases such as IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and arXiv. Keywords included: "Big Data Visualization," "AI-driven visualization," "ML4VIS," "Generative AI visualization," "Explainable AI visualization." Inclusion criteria: studies from 2022–2025, peer‑reviewed, English‑language, and directly focused on visualization and AI‑based analytical workflows.

Each included study was reviewed to extract information regarding visualization purpose, AI technique used, scalability characteristics, and application domain. Findings were then synthesized thematically, and results are presented in the coming sections. RQ1, RQ2 and RQ3 are addressed in Section Literature Review: Thematic Analysis. RQ1 and RQ2 in Challenges and Foundational Requirements in Big Data Visualization. Additional cross-cutting evidence related to RQ2 is synthesized in Synthesis and Identified Research Gaps. RQ3 is addressed primarily in AI-Enhanced Visualization Techniques, and is additionally supported by analytical insights discussed in Analytical Techniques in Support of Visualization Conclusions. RQ4 is addressed in Section Applications.

Literature Review: Thematic Analysis

1. Challenges and Foundational Requirements in Big Data Visualization

The challenges in big data visualization stem from data characteristics and limitations in analytical processing. Addressing both dimensions is critical for the design of effective visual systems.

A. Scale and Complexity

The scale and complexity of big data fundamentally reshape the way visualization systems must operate. Unlike traditional datasets, which can often be processed and displayed as a whole, big data arrives in massive volumes and at high speeds, quickly overwhelming conventional visualization tools. As a result, modern systems increasingly depend on distributed architectures that support parallel computation, allowing large datasets to be processed efficiently across multiple resources. In parallel, incremental and progressive rendering techniques enable analysts to begin exploring partial visual results while computations are still ongoing, rather than waiting for a complete and often time-consuming process to finish (Bikakis et al., 2025; Wang et al., 2024).

Maintaining responsiveness in such environments is not merely a technical concern, but a prerequisite for effective data interaction and analysis. To this end, approximation, aggregation, and sampling techniques play a crucial role by reducing computational load while preserving the most meaningful structures and trends within the data. These methods allow analysts to work with simplified yet informative representations, supporting faster interaction without significantly sacrificing analytical accuracy (Cuzzocrea, 2025). In many real-world scenarios, this balance between precision and performance is essential for sustaining analytical flow and user engagement.

Equally important is the ability to process and visualize data in real time. In domains such as finance, healthcare, or urban analytics, insights lose value if they arrive too late. Real-time visualization enables users to interact dynamically with evolving data streams, adjust their focus as conditions change, and make informed decisions when timing is critical (Ye et al., 2024).

Ultimately, the core challenge of scale and complexity lies not in rendering large datasets alone, but in supporting meaningful and continuous interaction with them. Visualization systems must enable users to explore patterns, shift between overview and detail, and respond to emerging insights without cognitive overload or technical friction. By integrating distributed computation, progressive visualization, approximation techniques, and real-time processing, contemporary systems seek to align computational capabilities with analytical requirements. This integration ensures that, even in the presence of massive and rapidly changing data, the analytical process remains interpretable, responsive, and effective (Neri et al., 2025; Yin et al., 2024).

B. Cognitive and Interaction Constraints

Perceptual and cognitive bandwidth is inherently limited, placing natural constraints on the amount of information that can be effectively processed at any given time. As datasets grow in size, complexity, and dimensionality, modern visual analytics systems must be carefully designed to respect these cognitive limits while still enabling deep insight discovery. To address these challenges, advanced systems incorporate strategies such as clutter reduction, which minimizes unnecessary visual elements and highlights relevant information; adaptive abstraction, which dynamically simplifies or aggregates data to match the user's focus; multi-level detail exploration, allowing users to seamlessly drill down from overview representations to granular data points; and guided or intelligent navigation, which provides contextual cues and recommendations that support effective exploration and decision-making (Neri et al., 2025; Wang et al., 2024).

C. High-Dimensionality and Heterogeneity

Bikakis et al. (2025) emphasize that visual analytics systems in the AI era must not only present data efficiently but also actively mitigate cognitive overload. They argue that future systems should incorporate interactive mechanisms that are aligned with analytical reasoning processes, enabling analysts to maintain situational awareness, draw accurate conclusions, and remain in control of the analysis workflow. By designing systems that are cognitively aware and interaction-aware, researchers and practitioners can ensure that even highly complex and high-dimensional datasets remain interpretable, actionable, and suitable for reliable analytical reasoning and decision-making processes (Beschi et al., 2025; Cuzzocrea, 2025; Yin et al., 2024). These challenges are further amplified in the presence of heterogeneous and multimodal data, where numerical, textual, visual, and temporal features must be analyzed and integrated within a unified analytical framework. For example, analytical tasks that combine transactional records, textual reports, spatial information, and temporal signals illustrate how heterogeneity increases both analytical complexity and visualization design challenges.

Ultimately, the goal is to create visualization environments that strengthen decision-making processes, rather than simply display larger volumes of data. By carefully integrating computational intelligence with principles of perception and cognition, modern visual analytics can transform overwhelming streams of information into meaningful, interpretable, and actionable insights, supporting more confident and effective analytical reasoning (Neri et al., 2025; Ye et al., 2024).

2. AI - Enhanced visualization techniques

AI significantly expands what visualization systems can achieve, enabling automation, personalization, and advanced pattern recognition.

A. Machine Learning for Visualization Recommendation

Visualization recommendation systems have traditionally relied on heuristics, grammar-based rules, or pre-defined templates to guide the selection of visual encodings and chart types. While effective in some contexts, these rule-based approaches are inherently limited, as they cannot fully adapt to the diverse, high-dimensional, and dynamic nature of modern datasets (Bikakis et al., 2025; Wang et al., 2024). In contrast, machine learning (ML)-driven visualization recommendation systems offer a more flexible and adaptive solution by learning directly from large repositories of datasets, prior visualization examples, and user interaction patterns (Cuzzocrea, 2025; Ye et al., 2024). By leveraging historical data and observing how users engage with visualizations, ML models can identify patterns that are not explicitly captured by heuristic rules (Neri et al., 2025).

ML-based visualization recommendation systems typically perform a range of advanced tasks, including:

i. Automated encoding selection, where the system predicts the most suitable visual representation for different types of data attributes and relationships (Yin et al., 2024).

ii. Chart type recommendation, suggesting the most effective visualization format based on data characteristics and analytical tasks (Ye et al., 2024).

iii. Perceptual optimization, ensuring that visualizations adhere to established perceptual principles, such as color distinguishability, shape salience, and visual hierarchy, thereby improving interpretability and comprehension (Bikakis et al., 2025).

iv. Task-context adaptation, dynamically adjusting visualizations to the specific analytical goals, user expertise, and cognitive load (Cuzzocrea, 2025; Wang et al., 2024).

As machine learning models continue to learn from interaction data and feedback, they become increasingly capable of generating visualizations that are intuitive, effective, and well aligned with analytical objectives. Unlike rule-based systems, which operate within fixed and predefined constraints, ML-driven approaches support personalization, adaptability, and continuous refinement over time. As a result, these systems enhance efficiency and accuracy in data exploration while also supporting deeper understanding, sustained engagement, and greater confidence in decision-making across complex and high-dimensional datasets (Neri et al., 2025; Yin et al., 2024).

B. *Generative AI for Visualization Automation*

Recent advances in Generative Artificial Intelligence (GenAI) have introduced new possibilities for automating key stages of the data visualization pipeline. Models based on transformer architectures and diffusion processes are increasingly applied to tasks that traditionally required extensive manual design and domain expertise. These models are capable of learning complex patterns from large collections of visual and semantic data, enabling the automatic generation and refinement of visual representations at scale.

GenAI techniques can support multiple stages of visualization construction, including data transformation, visual encoding selection, layout generation, and narrative structuring. Ye et al. (2024) present a comprehensive analysis of how generative models can be integrated into visualization workflows to reduce manual effort and improve consistency across visual outputs. By leveraging learned representations of data semantics and visual conventions, these systems can automatically map data attributes to appropriate visual forms and generate layouts that follow established design principles.

Key capabilities enabled by GenAI include text-to-visualization generation, where natural language specifications are translated into structured visual representations; style transfer techniques that adapt the aesthetic properties of visualizations to specific contexts or design guidelines; and automated data summarization, which identifies salient patterns and trends suitable for visual depiction. In addition, generative models can produce contextual annotations and explanatory elements, supporting clearer interpretation of visual outputs. Conversational and iterative refinement mechanisms further allow visualizations to be progressively adjusted based on feedback, enabling flexible exploration of alternative representations (Neri et al., 2025; Ye et al., 2024). Khan et al. (2025) systematically evaluated the ability of large language models (LLMs) to autonomously generate visualizations from natural language specifications. The study showed that while most LLMs can generate simple charts, they struggle to produce more complex visualizations.

Beyond individual visual artifacts, GenAI also contributes to scalability and reproducibility in visualization systems. Automated pipelines can generate consistent visualizations across large and evolving datasets, reducing variability introduced by manual design decisions. This is particularly relevant in data-intensive domains where visualizations must be frequently updated or adapted to new data conditions. Furthermore, by abstracting complex design operations into high-level specifications, GenAI-based systems expand the range of users who can effectively create and modify visualizations without requiring advanced technical or visualization expertise.

Despite these advantages, challenges remain regarding transparency, controllability, and validation of generated visual outputs. Ensuring that automated visualizations accurately reflect underlying data and adhere to analytical requirements is essential, particularly in decision-critical applications. As generative approaches continue to mature, their integration with interactive and analytical visualization systems is expected to play an increasingly central role in managing complexity and automation within large-scale data analysis environments (Ye et al., 2024; Yin et al., 2024).

C. Explainable AI (XAI) in Visual Analytics

As machine learning models continue to increase in complexity and adoption, the need for interpretability and transparency has become a critical requirement. Many high-performing models, particularly deep learning architectures, operate as black boxes, offering limited insight into how inputs are transformed into outputs. In this context, Explainable Artificial Intelligence (XAI) aims to make model behavior more transparent by exposing internal mechanisms, decision logic, and sources of uncertainty. Visualization plays a central role in XAI by translating abstract model explanations into interpretable visual forms that can be systematically analyzed.

Visual analytics systems support XAI by presenting multiple explanatory elements, including feature importance scores, attention maps, decision paths, and uncertainty indicators. Feature importance visualizations highlight which input variables contribute most significantly to model predictions, enabling comparative analysis across instances or model versions. Attention maps provide insight into which parts of the input data are emphasized during inference, particularly in neural network models applied to text, images, and multimodal data. Decision path visualizations reveal how intermediate rules or conditions lead to final predictions, while uncertainty representations communicate confidence levels and potential risks associated with model outputs.

Integrating these explanatory components within interactive visual analytics environments enables systematic exploration of model behavior across different data subsets, time periods, or parameter settings. Yin et al. (2024) demonstrate that visualization-driven XAI techniques can improve the reliability and acceptance of machine learning–based decision support systems, especially in high-stakes domains such as healthcare and finance, where accountability and risk assessment are essential. By making model reasoning more explicit, XAI-based visualizations support error detection, bias identification, and performance comparison among alternative models.

Beyond individual explanations, XAI-enriched visual analytics contributes to the evaluation and governance of machine learning systems. Visualization allows stakeholders to assess consistency, robustness, and fairness by comparing explanations across multiple predictions and scenarios. This capability is particularly important in regulated environments, where compliance and auditability require transparent documentation of algorithmic behavior (Neri et al., 2025).

Overall, the integration of XAI with visual analytics provides an effective framework for addressing the opacity of complex machine learning models. By combining explanatory techniques with interactive visualization, these systems reduce the gap between algorithmic complexity and analytical interpretability, enabling more informed assessment, validation, and deployment of data-driven decision-making systems (Wang et al., 2024; Yin et al., 2024).

D. User-Oriented AI Interactive Visualization Systems

Recent studies highlight the importance of user-oriented AI principles in visualization systems. Beschi et al. (2025) demonstrate how AI-driven end-user development (EUD) approaches enable analysts to generate visualizations through natural language specifications, iterative refinement, and guided workflows (Beschi et al., 2025).

While ML-based recommendation systems, generative AI, and XAI techniques address different stages of the visualization pipeline, they are largely complementary rather than competing approaches. ML-based methods primarily support encoding selection and interaction adaptation, generative models focus on automation and visual synthesis, and XAI techniques enhance interpretability and model transparency. Understanding the trade-offs between automation, control, and explainability remains an important consideration when integrating these approaches into unified visualization systems.

3. Analytical techniques in support of visualization conclusions

Visualization is only the final layer; robust analytical techniques support every preceding stage.

A. Distributed Data Processing Frameworks

Big data visualization relies on distributed computation infrastructure: Apache Spark, Apache Flink, Dask, GPU-accelerated frameworks

These tools support high-speed aggregation, streaming analytics, and incremental computation.

B. Data Reduction and Sampling Techniques

Maintaining interactivity in large-scale visualization systems requires reducing data complexity while preserving essential structural patterns. Common strategies include density-aware sampling, machine learning–based clustering, hierarchical aggregation, and scalable dimensionality-reduction techniques. These approaches enable efficient representation of large datasets, supporting responsive exploration without excessive computational overhead.

Cuzzocrea (2025) presents a hybrid framework that combines ML-driven aggregation with visual analytics to handle large sequential datasets, such as epidemiological time series. The framework demonstrates how integrated data reduction and visualization techniques can maintain analytical relevance while ensuring scalability and performance (Cuzzocrea, 2025).

C. Predictive Modeling and Pattern Mining

The integration of predictive modeling and pattern mining with visualization enhances the ability to identify trends, anticipate events, and detect anomalies across data-intensive domains. In cybersecurity, predictive analytics supports early detection of anomalous network behavior and potential threats, while visualization translates model outputs into interpretable patterns that can be examined over time. In supply chain management and finance, forecasting models are combined with visual representations to support demand prediction, risk assessment, and market trend analysis. Similarly, in healthcare, predictive diagnostics benefit from visual analytics that present model-driven insights related to disease progression, patient risk, and treatment outcomes. In these contexts, artificial intelligence performs large-scale pattern detection, while visualization serves to communicate and contextualize analytical findings in a clear and actionable manner (Neri et al., 2025; Wang et al., 2024).

Despite growing adoption, standardized evaluation metrics and benchmarking frameworks for AI-driven visualization systems remain limited, highlighting the need for systematic assessment of scalability, interpretability, and analytical reliability across different application contexts.

4. Synthesis and Identified Research Gaps

Although recent studies demonstrate meaningful advances in AI-augmented visualization, the literature remains fragmented and reveals several conceptual and technical gaps. Existing work lacks standardized evaluation frameworks for AI-generated visualizations, multimodal benchmarking datasets, support for real-time, decision-critical visual systems, systematic integration of human–AI co-design processes, and structured consideration of ethics-aware visualization practices. These insights form the foundation for the research roadmap and future directions outlined in the following section.

Future directions

Future work should prioritize multimodal AI visualization pipelines, standardized testing protocols, automated insight‑discovery validation, and ethical AI‑visual governance frameworks. Tighter integration of generative AI with human‑controlled oversight is expected to become essential.

Applications

Recent advances in big data visualization, visual analytics, and AI-driven techniques have significantly expanded the range of real-world applications where large, complex datasets must be explored, interpreted, and acted upon. Visualization systems increasingly serve as a critical interface between advanced computational models and decision-making processes, enabling analytics to be translated into operational insights for end-users (Ozturk & Elmasry, 2024). At the same time, these developments point toward new research directions, emphasizing scalability, automation, interpretability, and responsible AI use, while big-data visualization continues to broaden its application domains, increasingly prioritizing practical deployment and future development trends (Ouyang, 2024).

In healthcare and bioinformatics, visualization plays a central role in supporting data-intensive tasks such as epidemic monitoring, genomic analysis, and patient risk modeling. Biomedical datasets are often high-dimensional, heterogeneous, and temporally evolving, making them difficult to analyze using traditional methods. Visual analytics enriched with explainable AI techniques enables structured exploration of model outputs, helping practitioners examine influential variables, compare patient cohorts, and assess uncertainty in predictive models (Neri et al., 2025; Yin et al., 2024).

In the financial domain, visual analytics systems are widely applied to fraud detection, market surveillance, and risk assessment. Large-scale transaction data and algorithmic trading systems generate complex patterns that require continuous monitoring. Visualization dashboards enhanced with explainability mechanisms support regulatory compliance by exposing model behavior, highlighting anomalous activities, and enabling systematic auditing of automated decisions (Wang et al., 2024).

Visualization is frequently embedded as part of AI analytical pipelines across industries, supporting interpretation and operational decision-making (Xia & Wei, 2024).

Smart city environments represent another prominent application area. Urban sensing infrastructures and IoT networks generate continuous streams of spatiotemporal data related to traffic flow, energy consumption, environmental conditions, and public safety. Real-time visualization systems allow analysts to monitor evolving conditions, identify emerging issues, and evaluate the impact of policy interventions. Scalability and low-latency processing are essential in this context, particularly as data sources and sensor deployments continue to expand (Bikakis et al., 2025). In their study, Kulkarni et al. (2023) demonstrate how information extracted from data collected via roadside Bluetooth scanners can be processed using big-data analytics and subsequently presented through visualizations and tabular summaries within the R Studio environment. Their work illustrates the potential of sensor-based data acquisition, combined with analytical and visualization techniques, to support more effective road-traffic management and inform future mobility planning.

In cybersecurity, machine learning–based anomaly detection techniques are increasingly integrated with visualization to support threat analysis and incident response. Raw detection outputs become actionable through visual summaries that reveal network behavior, attack patterns, and temporal trends. By combining automated detection with interactive visual exploration, analysts can more effectively investigate suspicious events and prioritize mitigation efforts (Cuzzocrea, 2025).

Looking ahead, several directions are expected to shape the future of AI-enhanced visual analytics. One key trend is the development of tighter collaboration between analytical models and interactive visualization systems. Iterative feedback loops, supported by generative AI, are expected to enable dynamic refinement of visual representations as analytical goals evolve, supporting more flexible and adaptive workflows (Ye et al., 2024).

Another important direction involves the visualization of multimodal and heterogeneous data. As analytical tasks increasingly rely on the integration of text, images, audio, temporal signals, and structured records, there is a growing need for unified visualization frameworks that can coherently represent mixed data types without sacrificing analytical clarity (Wang et al., 2024).

Automated insight discovery is also likely to become more prominent. Early-stage research investigating AI-assisted visualization highlights increasing interest in automating visual encoding and insight discovery (Saber, 2024). AI systems are increasingly capable of identifying anomalies, suggesting hypotheses, and selecting relevant visual views automatically. While these capabilities offer substantial efficiency gains, they also introduce the need for rigorous evaluation methods to ensure reliability, relevance, and analytical validity (Neri et al., 2025).

Finally, ethical considerations are emerging as a foundational aspect of future visualization research. Issues related to bias, fairness, transparency, and data provenance are particularly critical in high-stakes application domains. Addressing these challenges requires not only algorithmic solutions but also visualization techniques that make assumptions, limitations, and uncertainties explicit within analytical workflows (Beschi et al., 2025).

Future research should focus on defining evaluation protocols for automated visualization systems, developing robust methods for multimodal data integration, and establishing ethical guidelines for AI-driven analytical workflows. Addressing these directions will be essential for ensuring that AI-enhanced visual analytics systems remain reliable, transparent, and applicable in high-stakes domains.

Conclusion

Big data visualization is entering a stage in which artificial intelligence, machine learning, and advanced analytical techniques have become integral to effective data analysis rather than optional enhancements. Recent research between 2022 and 2025 indicates a clear transition toward visualization approaches that combine automation, interpretability, and adaptive analytical support. AI-driven and generative visualization techniques offer promising mechanisms for accelerating insight discovery by highlighting relevant patterns and reducing the effort required to explore large and complex datasets (Cuzzocrea, 2025; Ye et al., 2024). In parallel, advances in explainable and interpretable visual analytics provide mechanisms for examining the behavior of complex models, supporting transparency, accountability, and informed decision-making (Neri et al., 2025; Yin et al., 2024).

Despite this progress, significant challenges remain. Scalability continues to constrain visualization systems as data grows in volume, velocity, and heterogeneity, particularly in multimodal and streaming contexts (Bikakis et al., 2025; Wang et al., 2024). A recent implementation example is provided by Poornima et al. (2025), who present a scalable architecture for real-time sentiment analysis of YouTube live-chat data using Apache Kafka, Spark Structured Streaming, and Grafana dashboards, illustrating both the potential of real-time analytics and the technical complexity of deploying such systems in practice. Effectively integrating heterogeneous data sources while preserving analytical clarity remains an open research problem, as does ensuring that automated visualization techniques remain reliable and understandable. Ethical considerations, including bias mitigation, fairness, and responsible deployment of AI-driven visual analytics, are also becoming increasingly central to both research and real-world applications (Beschi et al., 2025).

Looking ahead, continued advances in AI-enhanced visual analytics are expected to further reshape data-driven decision-making across a wide range of domains. The integration of scalable computational methods with interpretable and adaptive visualization techniques offers a pathway toward systems that not only present data effectively but also support rigorous analysis and collaborative decision processes. As AI and ML technologies continue to mature and ethical frameworks become more clearly defined, AI-augmented visual analytics is likely to play an increasingly important role in transforming complex data into actionable insights across diverse application areas.

References

Beschi, S., Fogli, D., Gargioni, L., & Locoro, A. (2025). AI-supported EUD for data visualization: An exploratory case study. Future Internet, 17(8), 349.

https://doi.org/10.3390/fi17080349

Bikakis, N., Chrysanthis, P., Li, G., Papastefanatos, G., & Yu, L. (2025). Visual analytics challenges and trends in the age of AI: The BigVis community perspective. arXiv preprint. https://arxiv.org/abs/2504.21500

Cuzzocrea, A. (2025). A machine-learning-based data science framework for visualization and visual analytics. Computers, 14(7), 190. https://doi.org/10.3390/computers14070190

Khan, S. R., Chandak, V., & Mukherjea, S. (2025). Evaluating LLMs for visualization generation and understanding. Discover Data, 3, 15. https://doi.org/10.1007/s44248-025-00036-4

Kulkarni, A., Kumar, N., & Rao, K. (2023). Efficacy of Bluetooth-based data collection for road traffic analysis and visualization using big data analytics. Big Data Mining and Analytics, 6(2), 139–153. https://doi.org/10.26599/BDMA.2022.9020039

Neri, G., Marshall, S., Chan, H. K.-H., Yaghi, A., Tabor, D., Sinha, R., & Mazumdar, S. (2025). Data visualization in AI-assisted decision-making: A systematic review. Frontiers in Communication, 10, 1605655. https://doi.org/10.3389/fcomm.2025.1605655

Ouyang, W. (2024). Data visualization in big data analysis: Applications and future trends. Journal of Computer and Communications, 12(11), 55–76.

https://www.scirp.org/pdf/jcc20241211_51732919.pdf

Ozturk, M., & Elmasry, M. (2024). AI-enhanced data visualization: Transforming complex data into actionable insights. Journal of Data Science & Emerging Technologies, 6(2), 112–130.

https://www.researchgate.net/publication/380712811_AI-Enhanced_Data_Visualization_Transforming_Complex_Data_into_Actionable_Insights

Poornima, S., Shashidhar, K., & Fernandes, R. (2025). Real-time sentiment analysis of YouTube live chat with Apache Spark and Kafka. IEEE Access, 13, 188929–188938. https://doi.org/10.1109/ACCESS.2025.3625785

Saber, M. (2024). Formative study for AI-assisted data visualization. arXiv preprint. https://arxiv.org/abs/2409.06892

Wang, J., Liu, S., & Zhang, W. (2024). Visual analytics for machine learning: A data perspective. IEEE Transactions on Visualization and Computer Graphics, 30(12), 7637–7656. https://doi.org/10.1109/TVCG.2024.3357065

Xia, L., & Wei, Q. (2024). Applications of data visualization technology in artificial intelligence. Frontiers of Business, Economics and Management, 7(3), 44–59. https://drpress.org/ojs/index.php/fbem/article/view/22389

Ye, Y., Hao, J., Hou, Y., Wang, Z., Xiao, S., Luo, Y., & Zeng, W. (2024). Generative AI for visualization: State of the art and future directions. arXiv preprint.

https://doi.org/10.48550/arXiv.2404.18144

Yin, S., Li, H., Sun, Y., Ibrar, M., & Teng, L. (2024). Data visualization analysis based on explainable artificial intelligence: A survey. IJLAI Transactions on Science and Engineering, 2(3), 24–31.