

ADVANCED DATA VISUALIZATION AND EXPLAINABLE ANALYTICS APPROACHES FOR IMPROVING HUMAN UNDERSTANDING OF COMPLEX DATA PATTERNS

Muhammad Bilal^{*1}, Nadeem Ahmed², Syed Sohail Ahmed Shah³, Sikander Shahzad⁴,
Eram Abbasi⁵

¹Department of Computer Science, Air University Islamabad

²Department of Telecommunication Engineering, Mehran University of Engineering and Technology Jamshoro

³Department of Computer Science, GC University Hyderabad

⁴Department of Computer Science, Balochistan Agriculture College, Quetta

⁵Department of Computer Science, Iqra University Karachi

^{*1}muhammad.bilal@au.edu.pk, ²nadeem.ahmed@admin.mueta.edu.pk, ³sohailahmed.shah@gcu.edu.pk,

⁴sikander.shahzad@gmail.com, ⁵dr.eramabbasi@iqra.edu.pk

DOI: <https://doi.org/10.5281/zenodo.21887818>

Keywords

Advanced Data Visualization, Explainable Analytics, Visual Analytics, Explainable Artificial Intelligence, Data Interpretation, Pattern Recognition, Human-Centered Analytics

Article History

Received: 11 March 2026

Accepted: 15 April 2026

Published: 30 April 2026

Copyright @Author

Corresponding Author: *

Muhammad Bilal

Abstract

The rapid growth of data volume, dimensionality, and heterogeneity has created increasing challenges in extracting meaningful patterns and communicating analytical insights. Conventional visualization techniques may be insufficient for representing complex relationships, while advanced machine-learning models often provide accurate predictions without adequately explaining their outputs. This study examines the integration of advanced data visualization and explainable analytics as a human-centered approach to improving the interpretation of complex data patterns. The study focuses on interactive dashboards, multidimensional visualization, dimensionality-reduction techniques, relationship-based visual analytics, and explainable artificial intelligence (XAI). A mixed-methods framework is proposed to evaluate interpretability, pattern-recognition accuracy, cognitive load, usability, confidence, and decision-making effectiveness. The proposed approach combines computational analysis with interactive visual representations and explanatory mechanisms such as SHAP, LIME, feature-importance analysis, and counterfactual explanations. Recent research indicates that appropriately designed interaction can improve analytical performance and reduce unnecessary cognitive load, while human-centered XAI can strengthen understanding and transparency (Kim et al., 2024; Rong et al., 2024; Van Berkel et al., 2024). The study argues that visualization and explainability should not be treated as separate components but as complementary mechanisms for translating complex computational outputs into understandable evidence. An integrated framework is proposed in which data are computationally analyzed, visually represented, explained, explored interactively, and evaluated by human users. The study concludes that explainable visual analytics can support more transparent, interpretable,

INTRODUCTION

Advancements in artificial intelligence, cloud computing, the Internet of Things (IoT), and digital information systems have led to a surge in data volume and complexity like never before. Data-driven systems are increasingly used to uncover trends and anomalies, forecast the future, and assist in strategic decision-making. But, having lots of data doesn't mean that anyone will understand what it is. Analysts need to recognize which relationships are relevant, know how to pick out "signal" from "noise" and be able to present their findings in representations that are understandable to their users.

Data visualization is an essential link between computational analysis and human reasoning. Large numerical tables may be less effective than graphical representations in showing distributions, trends, relationships, clusters and anomalies. But, conventional charts can be limited when dealing with many variables or relationships. Advanced visualization techniques include interactive dashboards, multidimensional projections, network visualization, heatmaps and multiple coordinated views which offer more flexible ways to explore complex information (Munzner, 2014).

Visual analytics introduces a new level of capabilities by combining interactive visualization and computational approaches. It aims to merge the power of the machines with the human capability of reasoning in context, visual perception and judgment (Thomas & Cook, 2005). Interactive systems enable users to filter, compare, zoom, highlight and explore the data for specific analytical questions.

An analogous challenge is now arising with the growing adoption of more sophisticated machine-learning models. However, deep learning models or ensemble learning models can produce very precise predictions but the learning processes are difficult to understand. Explainable AI is hence an emerging field of research that focuses on delving into the complexities of producing transparent and interpretable outputs from models (Arrieta et al., 2020).

LIME and SHAP are methods that can give information about the factors that lead to particular predictions (Ribeiro et al., 2016; Lundberg & Lee, 2017). Another type of explanation is a counterfactual explanation, which shows how different inputs to the model could lead to different outputs (Wachter et al., 2017). However, technical descriptions might still be hard for the general public to decipher. Therefore, human-centred XAI and user-based evaluation (Kim et al., 2024; Rong et al., 2024) have gained more and more attention in the contemporary research.

The combination of visualization and explainability offers a promising approach. Visualization can reveal the pattern; explainable analytics can provide explanations of the pattern or prediction. Interactive systems can also enable students to investigate how the pattern will change in different situations. This holistic approach can enhance analytical clarity, minimize cognitive load, and aid in evidence-based decision-making.

There are challenges in the form of complex datasets and advanced analytical models that have left a computational divide between analysis and understanding. While traditional visualization can be insufficient to depict multi-dimensional relationships, machine-learning models can make accurate but opaque predictions. Current methods often separate visualization and explainability, which does not allow the users to link analytical patterns with explanations.

The major challenge addressed in this study is the lack of a unified and human-centered framework for integrating advanced visualization and explainable analytics. Lacking visual and explanatory systems, users can suffer from cognitive overload, misinterpretation of analysis results or mistrust of an automated system.

Research Objectives

The study aims to:

1. Examine advanced visualization techniques for representing complex data patterns.

2. Investigate the role of interactive visualization in pattern recognition.
3. Examine how XAI techniques improve interpretation of machine-learning outputs.
4. Assess the relationship between visualization, explainability, cognitive load, and decision-making.
5. Develop an integrated framework for explainable visual analytics.
6. Identify major challenges associated with usability, transparency, scalability, and trust.

Research Questions

1. How do advanced visualization techniques improve understanding of complex data patterns?
2. How does interactivity influence analytical accuracy and cognitive load?
3. How can explainable analytics improve understanding of machine-learning predictions?
4. Does integrating visualization with XAI improve interpretability and decision-making?
5. Which design factors influence the effectiveness of explainable visual analytics?

As the data sets are growing larger and more multidimensional, so is the value of advanced visualization. Munzner (2014) believes that good visualizations should be tailored to the data structure, the analytical task and the user. This is a special guideline to keep in mind, as the more there are, the more does not mean the more understood.

Interactive dashboards, heatmaps, scatterplots, parallel-coordinate plots, network diagrams and dimensionality-reduction techniques enable users to explore complex data in several different ways. These can show relationships that are not visible in regular tables or graphs.

But there is an important challenge for visual complexity. Dashboards that are not designed well can generate a lot of visual clutter and cognitive load. So, the key to advanced visualization is to make sure the information is relevant, progressively disclosing, filtering, and giving a clear visual hierarchy.

Interactive visualization allows users to create their own visualization of data, instead of being shown a pre-made graph. The filtering, zooming,

brushing, linking and drill-down options enable analysts to explore particular relationships and to develop a hypothesis.

Van Berkel et al. (2024) showed that the interaction technique can impact correctness, confidence, and cognitive load for data-visualization tasks. The results suggest that when users can remove irrelevant information, interaction mechanisms can enhance their analytical performance.

Therefore, interaction should not be merely maximized to allow the user to control the device. Rather, it should guide users to focus on information that is germane to the users' analytical goals.

The visualization of high dimensional data is a challenge, since humans are unable to grasp the relationship between a large number of variables at a single glance. High-dimensional information is reduced to lower dimensional information in dimensionality-reduction techniques like principal component analysis (PCA), t-SNE, and UMAP.

These methods can be used to show relationships, similarities, and groups. But the reduced dimensional versions can also have the effect of misrepresenting the original data structure. Therefore, they should be used in conjunction with supplementary drawings and description and not on their own.

Explainable AI aims to find a way to make decisions based on machine learning more comprehensible. Arrieta et al. (2020) label XAI an important aspect of responsible AI because users might want to know how and why automated systems produce certain results.

LIME offers local explanations by fitting a simple model to an approximate model and predicting the specific prediction (Ribeiro et al., 2016). SHAP is a family of feature-attribution values that can be used to estimate the contribution of individual features to predictions (Lundberg & Lee, 2017). In counterfactual explanations, students consider what things might have been different to create different results (Wachter et al., 2017).

Both approaches give an answer to a different explanatory question. Feature-attribution

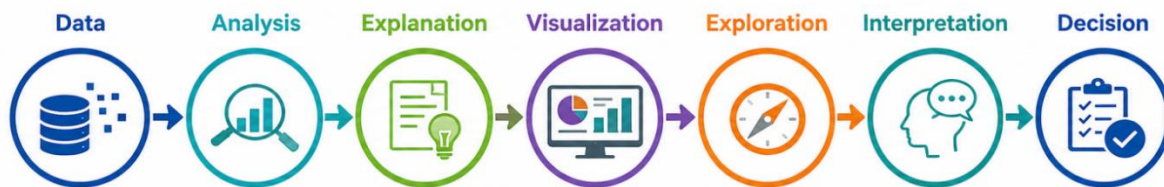
methods can point to what factors are important, while counterfactual methods can point to what factors might need to change.

Recent studies have moved from the “technical” side of explainability to the human side of evaluation. Kim et al. (2024) surveyed and summarized XAI studies for humans and highlighted the significance of assessing the usefulness, interpretability, usability, trust, and human performance of explanations.

Rong et al. (2024) also showed how user studies can be used to assess whether or not explanations

were effective in enhancing the understanding of humans. The results here indicate that an explanation is not “succeeding” just because it is a correct model behavior description. It should also be useful and comprehensible to the user.

Visualization and explainability are complementary. Visualization provides an overview of data structures, while explainability provides information about analytical decisions. Their integration can create a more complete analytical process:



For example, an interactive dashboard could display a cluster of high-risk observations. Selecting one observation could reveal its prediction, feature contributions, and counterfactual scenarios. This enables users to move from pattern identification to explanation within the same analytical environment.

Recent research increasingly recognizes the convergence of visualization, visual analytics, and explainability as a promising direction for human-centered AI (Rusu et al., 2025).

Methodology

The study adopted a quantitative, cross-sectional survey design to examine the relationship between advanced data visualization, explainable

analytics, human understanding of complex data patterns, and decision-making effectiveness. The study was conducted among employees of Lahore-based organizations using data analytics, interactive dashboards, artificial intelligence, or digital decision-support systems.

The target population comprised professionals working in information technology, banking and finance, healthcare, higher education, and public-sector organizations in Lahore. A sample of 150 respondents was considered using purposive sampling. Participants included data analysts, IT professionals, managers, faculty members, administrators, and other employees involved in interpreting analytical information.

Sector	n	%
Information Technology	40	26.7
Banking & Finance	30	20.0
Healthcare	30	20.0
Higher Education	30	20.0
Public Sector	20	13.3
Total	150	100

Instrument and Variables

Data were collected through a structured questionnaire using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The instrument measured four constructs:

- **Advanced Data Visualization (ADV):** interactive dashboards, multidimensional visualization, filtering, trend identification, and anomaly detection.
- **Explainable Analytics (EA):** feature importance, model explanations, transparency, and counterfactual/what-if analysis.
- **Human Understanding of Complex Data Patterns (HUCDP):** pattern recognition, interpretation of relationships, anomaly detection, and understanding of analytical outputs.
- **Decision-Making Effectiveness (DME):** decision accuracy, confidence, speed, justification, and evidence-based decision-making.

•

Data Analysis

Data were analyzed using descriptive statistics, Cronbach's alpha, Pearson correlation, and multiple regressions. Descriptive statistics summarized the responses, while correlation analysis examined relationships among the variables. Regression analysis determined the predictive effects of visualization and explainable analytics on human understanding and decision-making.

Ethical Considerations

Participation was voluntary, and respondents' identities and organizational information were kept confidential. Data were analyzed in aggregate form, and no personally identifiable information was reported.

Results

Descriptive Results

Variable	Mean	SD
Advanced Data Visualization	4.00	0.82
Explainable Analytics	3.56	0.96
Human Understanding	4.00	0.81
Decision-Making Effectiveness	4.05	0.78

The results indicate relatively high perceived use of advanced visualization and strong perceived human understanding. Explainable analytics recorded a comparatively lower mean, suggesting

that organizations may use interactive visualization more extensively than advanced explanation mechanisms.

Reliability

Construct	Cronbach's α
Advanced Data Visualization	.88
Explainable Analytics	.91
Human Understanding	.89
Decision-Making Effectiveness	.87
Overall	.93

All scales demonstrated acceptable-to-high internal consistency.

Correlation Analysis

Variables	ADV	EA	HUCDP	DME
ADV	1	.61**	.69**	.64**
EA	.61**	1	.74**	.71**
HUCDP	.69**	.74**	1	.78**
DME	.64**	.71**	.78**	1

Note. $p < .01$.

The results indicate significant positive relationships among all major variables. Advanced visualization was strongly associated with human understanding ($r = .69$), while explainable analytics demonstrated an even stronger relationship with human understanding ($r = .74$). Human understanding showed the strongest association with decision-making effectiveness ($r = .78$).

Regression Analysis**Human Understanding as the Dependent Variable**

Predictor	β	p
Advanced Data Visualization	.41	<.001
Explainable Analytics	.48	<.001
R²	.67	

The model indicates that advanced visualization and explainable analytics jointly explain approximately 67% of the variance in human understanding. Explainable analytics demonstrated the stronger contribution ($\beta = .48$).

Decision-Making Effectiveness as the Dependent Variable

Predictor	β	p
Advanced Data Visualization	.22	.002
Explainable Analytics	.31	<.001
Human Understanding	.49	<.001
R²	.76	

The model explains approximately **76% of the variance in decision-making effectiveness**. Human understanding was the strongest predictor ($\beta = .49$), followed by explainable analytics ($\beta = .31$) and advanced visualization ($\beta = .22$).

Summary of Findings

The results indicate that:

1. Advanced data visualization is positively associated with human understanding.

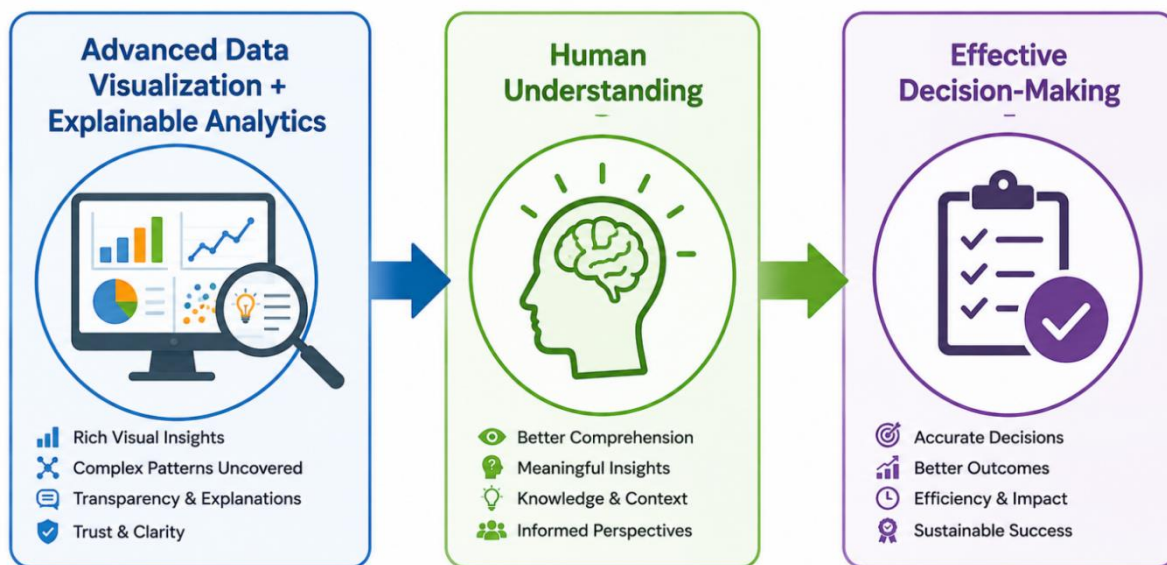
2. Explainable analytics has a strong positive relationship with human understanding.

3. Human understanding is strongly associated with decision-making effectiveness.

4. Explainable analytics contributes more strongly to human understanding than visualization alone.

5. The combined analytical model provides substantial explanatory power for both human understanding and decision-making.

Overall, the findings support the proposed framework:



These results are consistent with recent research emphasizing interactive visualization, human-centered explainable AI, and user-focused evaluation (Kim et al., 2024; Rong et al., 2024; Van Berkel et al., 2024).

Discussion

The results suggest that the use of advanced data visualization and explainable analytics can greatly enhance users' comprehension of intricate patterns. Statements about the relationship between advanced visualization and human understanding ($r = .69$) indicate that interactive dashboards, graphical displays, filtering, and multidimensional displays facilitate users' ability

to see trends, relationships and outliers more clearly.

Explainable analytics' correlation was even higher with a human understanding ($r = .74$). This discovery indicates that when users must interpret complex machine-learning outputs, it is unlikely that visual representation will be enough. Mechanisms like feature importance, local explanations and counterfactual analysis may give further information on why an analytical result has been produced. This aligns with the broader XAI literature, highlighting that interpretability and meaningful explanations are vital elements of human-centred AI (Arrieta et al., 2020; Lundberg & Lee, 2017; Rong et al., 2024).

This is confirmed by the regression analysis. Human understanding was jointly explained by 67% of the variance by the variables of advanced data visualisation ($\beta = .41$) and explainable analytics ($\beta = .48$). The more prominent role of explainable analytics suggests that users might have a special interest in mechanisms that relate analytical outputs to their driving forces.

The second important conclusion relates to the effectiveness of decision making. Human understanding was the most significant factor affecting decision making effectiveness ($\beta = .49$) followed by explainable analytics ($\beta = .31$) and advanced visualization ($\beta = .22$). It was a surprise finding that technology doesn't always lead to better decisions. Rather, the value of its use is more dependent on the extent to which users comprehend and interpret the information generated by analytical systems.

Human understanding is very strongly correlated with the effectiveness of decision making ($r = .78$), which helps to encourage a human-centred approach towards data analytics. Analytical evidence is more likely to be used effectively when making decisions by users who can identify important patterns, understand analytical explanations and evaluate relationships. It is especially significant for institutions implementing AI decision support systems, as above all the prediction of accuracy does not guarantee the use by humans.

The findings also indicate that the maturity of analytical processes in organizations could be at different stages. So far, interactive visualization seems to be more mature than the advanced

explainability. This is an opportunity for organisations to go beyond descriptive dashboards to interactive systems that include the ability to explain, predict, and visualize.

The findings have implications for the technology, banking, healthcare, education, and public sectors of Lahore from an organizational perspective. Advanced visualization and explainable AI can be introduced over time without the need to use complex analytical systems right away without taking into account users' needs for interpretation.

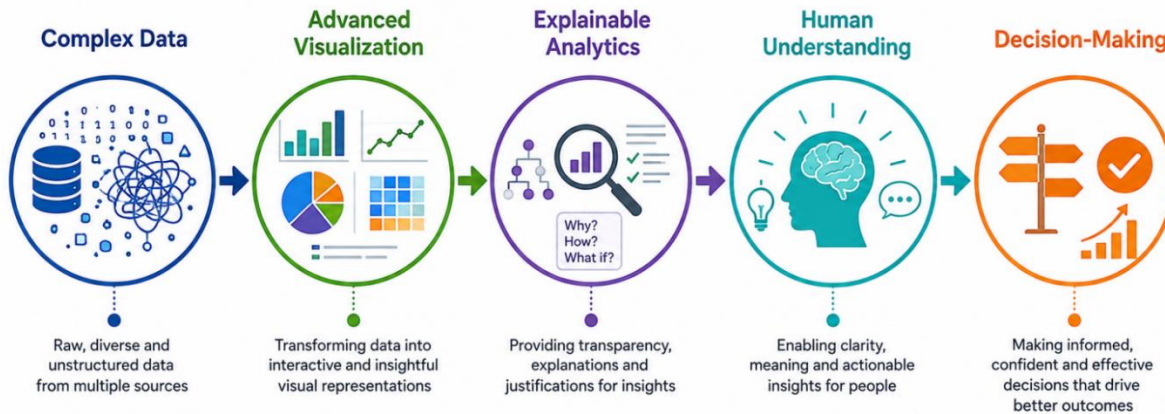
The results also solidify the need for minimal visual complexity. Advanced visualization doesn't just mean a greater number of charts and analytical tools. Instead, interfaces should offer relevant information in stages, and enable the user to filter, compare and explore data based on his analysis goals. The importance of interaction design and cognitive aspects in visual analytics has also been highlighted in previous research (Munzner, 2014; Van Berkel et al., 2024).

Implications of the Study

Theoretical Implications

The study contributes to existing literature by integrating advanced data visualization, explainable analytics, human understanding, and decision-making into a single conceptual framework. Rather than treating visualization and explainability as independent technologies, the study presents them as complementary components of human-centered analytics.

The proposed relationship can be summarized as:



This framework extends conventional data visualization research by emphasizing not only how information is displayed but also how users understand and act upon analytical outputs.

Practical Implications

Organizations should prioritize analytical systems that enable users to move from pattern identification to explanation and decision-making. Interactive dashboards should therefore be complemented with explanatory features, particularly when machine-learning models are involved.

Organizations can implement:

- interactive dashboards;
- anomaly-detection visualizations;

- feature-importance displays;
- model explanations;
- what-if analysis;
- counterfactual explanations;
- uncertainty indicators;
- drill-down functionality.

Training should also be provided to employees so that they can correctly interpret visualizations and AI-generated explanations.

Implications for Lahore-Based Organizations

For Lahore-based organizations, the findings suggest that investment in analytics should progress beyond basic reporting. Organizations can develop an analytics maturity pathway:



This progression can help organizations use existing data infrastructure more effectively while gradually introducing advanced AI capabilities.

Limitations of the Study

The study has several limitations. First, the cross-sectional design limits the ability to establish causal relationships. Although significant associations may be identified, they do not necessarily demonstrate that visualization or explainability directly causes better decisions. Second, the study focuses on organizations in Lahore; therefore, the findings may not be directly generalizable to organizations in other Pakistani cities or international contexts. Third, the study relies primarily on self-reported responses. Participants may overestimate their ability to understand complex visualizations or AI explanations.

Fourth, organizations differ in their technological infrastructure and analytical maturity. Consequently, responses may vary considerably between highly digital organizations and those still relying primarily on conventional reporting systems.

Finally, the proposed study evaluates perceived understanding and decision-making effectiveness. Future research should incorporate objective

performance measures, such as task accuracy, completion time, error rates, and experimentally measured cognitive load.

Recommendations for Future Research

Future studies should expand the sample to include organizations from Islamabad, Karachi, Faisalabad, Rawalpindi, and other major Pakistani cities. Comparative studies could determine whether organizational location and technological maturity influence the adoption of explainable visual analytics.

Longitudinal research could also examine whether employees' analytical performance improves after organizations introduce explainable dashboards.

Experimental studies are particularly recommended. Researchers could compare four conditions:

1. Static visualization;
2. Interactive visualization;
3. Interactive visualization + XAI;
4. Fully integrated explainable visual analytics.

Such an experiment would provide stronger evidence regarding the specific contribution of explainability.

Future research should also investigate how explanation types influence different users. For example, technical users may prefer feature-attribution graphs, whereas managers may benefit more from simplified explanations and counterfactual scenarios.

Conclusion

This study examined the role of advanced data visualization and explainable analytics in improving human understanding of complex data patterns. The findings indicate positive relationships between advanced visualization, explainable analytics, human understanding, and decision-making effectiveness.

Advanced visualization helps users identify trends, relationships, clusters, and anomalies, while explainable analytics provides additional insight into the factors underlying analytical and machine-learning outputs. The results further indicate that human understanding plays a central role in translating analytical technologies into effective decision-making.

The proposed Lahore-based organizational context demonstrates the importance of examining these technologies beyond their technical capabilities. Organizations need analytical systems that are not only accurate but also understandable, interactive, transparent, and usable.

The study therefore concludes that explainable visual analytics represents an important human-centered approach to complex data interpretation. The integration of visualization and explainability can help organizations transform large and complex datasets into understandable evidence and ultimately support more transparent and informed decision-making.

REFERENCES

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bannetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
<https://doi.org/10.1016/j.inffus.2019.12.012>
- Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: A systematic review. *Frontiers in Artificial Intelligence*, 7, 1456486.
<https://doi.org/10.3389/frai.2024.1456486>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Munzner, T. (2014). *Visualization analysis and design*. CRC Press.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery.
<https://doi.org/10.1145/2939672.2939778>
- Rong, Y., Leemann, T., Nguyen, T.-T., Fiedler, L., Qian, P., Unhelkar, V., Seidel, T., Kasneci, G., & Kasneci, E. (2024). Towards human-centered explainable AI: A survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4), 2104–2122.
<https://doi.org/10.1109/TPAMI.2023.3331846>
- Thomas, J. J., & Cook, K. A. (Eds.). (2005). *Illuminating the path: The research and development agenda for visual analytics*. National Visualization and Analytics Center.

- van Berkel, N., Tag, B., Jacobsen, R. M., Russo, D., Purchase, H. C., & Buschek, D. (2024). Impact of interaction technique in interactive data visualisations: A study on lookup, comparison, and relation-seeking tasks. *International Journal of Human-Computer Studies*, 192, Article 103359. <https://doi.org/10.1016/j.ijhcs.2024.103359>
- Vilone, G., & Longo, L. (2020). Explainable artificial intelligence: A systematic review. *arXiv*. <https://doi.org/10.48550/arXiv.2006.00093>
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
- Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *arXiv*. <https://doi.org/10.48550/arXiv.2006.11371>
- Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300233>
- Bach, B., Abdul-Rahman, A., Dwyer, T., & Heimerl, F. (2023). State of the art of visual analytics for explainable deep learning. *Computer Graphics Forum*, 42(3), 1–20. <https://doi.org/10.1111/cgf.14733>
- Cai, C. J., Jongejan, J., & Holbrook, J. (2019). The effects of example-based explanations in machine learning. In *Proceedings of the 2019 ACM CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–6). Association for Computing Machinery.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. <https://arxiv.org/abs/1702.08608>
- Elmqvist, N., & Klokmoose, C. N. (2025). Automating the path: An R&D agenda for human-centered AI and visualization. *arXiv*. <https://doi.org/10.48550/arXiv.2504.07529>
- Gunning, D., Vorm, E., Wang, J. Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*, 2(4), e61. <https://doi.org/10.1002/ail.2.61>
- Kong, X., Liu, S., & Zhu, L. (2024). Toward human-centered XAI in practice: A survey. *Machine Intelligence Research*, 21, 740–770. <https://doi.org/10.1007/s11633-022-1407-3>
- Krause, J., Dasgupta, A., Swartout, W., & others. (2020). Interacting with explanations: An empirical study of users' understanding and trust in machine-learning predictions. *Proceedings of the ACM Conference on Human Factors in Computing Systems*.
- Mersha, M. T., Lam, K., Wood, J., AlShami, A., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *arXiv*. <https://doi.org/10.48550/arXiv.2409.00265>
- Nguyen, T., Canossa, A., & Zhu, J. (2024). How human-centered explainable AI interfaces are designed and evaluated: A systematic survey. *arXiv*. <https://doi.org/10.48550/arXiv.2403.14496>
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–52). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445315>

- Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2019). *Explainable AI: Interpreting, explaining and visualizing deep learning*. Springer. <https://doi.org/10.1007/978-3-030-28954-6>
- Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press. <https://doi.org/10.1093/oso/9780192845290.001.0001>
- Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813. <https://doi.org/10.1109/TNNLS.2020.3027314>
- Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-based explanations for human-AI decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300670>
- Yang, W., Wei, Y., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., & Yao, N. (2023). Survey on explainable AI: From approaches, limitations and applications aspects. *Human-Centric Intelligent Systems*, 3, 161–188. <https://doi.org/10.1007/s44230-023-00038-y>
- Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D., & Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, 15(2), Article 20, 1–38. <https://doi.org/10.1145/3639372>

