International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

MISSING CONTEXT IN AI EXPLANATION

Ms. CV Nithyasri, Ms. S. Mounika, Ms. G. Hasna

Undergraduate Students BS-MS (Computer Science), Vishwa Vishwani Institute Of systems and management
Hyderabad, Telangana, India | Email:- nchatrati@gmail.com | Mobile no:- 7032971009

ABSTRACT

Artificial Intelligence (AI) systems bolster decisions in the recruitment process, lending, education and public services. Explainable AI (XAI) helps users understand AI generated decisions while leaving out few important context needed to interpret those decisions in explanation. This is a working paper that explores context omission, referring information that might reasonably affect the user's experience to an AI augmentation. The main question is whether an accurate explanation alone is enough to help users make informed decisions. It is a conceptual study, no experiment was conducted, based on literature. Inspired by research about human-centred XAI evaluation and context-sensitive explanation, a prototype for evaluating omissions in four categories: decision relevance, interpretive significance, usability, and contestability.

The paper suggests not only accurate explanations but also context behind it must be provided for the user with the context in which the decision was made. The framework can be used to explore if explanations facilitate informed decision making and review by practitioners, regulators and researchers. It is not validated, so it needs to be tested by users and domain experts in the future empirical studies.

Keywords: Explainable AI (XAI), material context omission, human centred evaluation, decision-making

1. INTRODUCTION

AI-assisted systems are used in decisions about lending, insurance and public benefits. A credit-scoring explanation may identify the factors that influenced a recommendation, but the person affected may still need information to check whether the underlying data is accurate or relevant. Evaluating explanations therefore requires more than checking whether they faithfully describe

a model. It also requires asking whether they provide enough context for a person to understand and respond to a particular decision.

This distinction between explanation fidelity and contextual adequacy matters. For example, an explanation might correctly state that an employment gap lowered an applicant's score. Yet the applicant may need to know whether the employment record is accurate, whether an approved leave period was recorded, and whether the organization's policy allows the information to be corrected. These questions concern the decision context rather than the accuracy of the model description alone.

Research also shows that explaining an AI decision does not always lead to better decisions. Alufaisan et al. (2021) did not find any significant improvements to human decision-making accuracy by including explanations for AI prediction results on these tasks. Outcomes like knowledge acquisition, uncertainty awareness, and trust calibration are considered by Wang and Yin. This paper motivates evaluation of explanation for a user instead of just verifying its existence.

This paper proposes a practical way to identify possible omissions in otherwise accurate AI explanations. Each suspected gap is linked to supporting evidence and assessed against four dimensions: decision relevance, interpretive significance, usability and contestability. The aim is to organize questions that reviewers and affected users can investigate, not to introduce an entirely new theory of context-sensitive explainability.

2. REVIEW OF LITERATURE

2.1 Technical explanation and its limitations

Explainability techniques are used to explain post hoc behavior of a local model. Ribeiro, Singh, and Guestrin (2016) introduced LIME, which approximates a classifier near a particular prediction. These methods may be useful for analyzing features; however, they do not provide an adequate description of all aspects such as organization's policy decisions process or personal effects. Contribution to feature does not imply causality. According to Doshi-Velez and Kim, there is a need for rigorous evaluation of interpretability based on the intended task.

2.2 Human-centred effectiveness

Alufaisan et al. (2021) compare human performance without AI help, with AI prediction, and with predictions accompanied by explanation. The results of this study do not prove an overall statement about how better explanation leads to improved decision making. Understanding, uncertainty awareness and appropriate trust calibration are distinguished by Wang and Yin.

Therefore explanations may be convincing while still hiding some limitations.

Liao, Gruen and Miller (2020) examine the questions people want AI explanations to answer. Ehsan and Riedl (2020) take a reflective sociotechnical approach, emphasizing that explanations operate within social and organizational settings. Both perspectives suggest that an explanation should be evaluated in relation to its users and their circumstances.

2.3 Current studies of context and evaluation

Explainability of models is a part of context for XAI. SAGE model by Mill et al. (2024) is an approach to structuring context around explanation. Kim, Maathuis, and Sent review 73 user evaluations and identify 30 components relevant to meaningful explanation and humancentered evaluation. The study shows that usefulness, sufficiency, and related issues are not new just because they have been renamed.

2.4 Contestability and actionable explanation

Wachter, Mittelstadt, and Russell discuss counterfactual explanations for changes associated with alternative outcomes. This data might be useful for people to know about alternatives available; however, it does not provide guidance on fixing an error or asking for a review. Counterfactual advice might not be feasible or within someone's reach. Verification for corrections and appeals must be done by means of an evaluation procedure instead of being based on a model definition alone.

3. RESEARCH GAP

Existing research recognizes that a technically faithful explanation may not meet the needs of a particular user or decision. A practical question remains: how can reviewers document a suspected omission, identify the evidence for it, explain its possible consequences and decide what response is appropriate? This paper addresses that documentation problem through a

proposed evidence-linked register. It does not claim that earlier research has overlooked context altogether.

4. OBJECTIVES OF THE STUDY

The objectives are to (1) distinguish faithful model explanations from contextually adequate explanations; (2) propose a simple assessment of possible omissions using four dimensions; and (3) outline how the proposed approach could be evaluated in future empirical research.

5. RESEARCH METHODOLOGY

This is a conceptual, literature-based study. It draws on research about interpretability, humancentred explanation, decision context and contestability. Hypothetical recruitment and lending scenarios illustrate how an accurate explanation may still leave an affected person with unanswered questions. The proposed assessment is not presented as an experimentally validated tool, and the literature discussed is not claimed to be an exhaustive systematic review.

To assess a possible omission, a reviewer would compare the explanation with available decision records, relevant policies and other credible evidence. The reviewer would then consider the affected person's circumstances, the potential consequences of the missing information and any available correction or appeal process. Where evidence is absent or conflicting, the register should record that uncertainty rather than treat a suspicion as a confirmed omission.

6. CONCEPTUAL BACKGROUND

Material context omission occurs when an otherwise accurate explanation leaves out information that could matter to someone trying to understand, use or challenge an AI-assisted decision. This paper considers four dimensions. Decision relevance asks whether the missing information could affect a specific decision. Interpretive significance asks whether it could change how the explanation is understood. Usability considers whether the information is accessible and practical for the person concerned. Contestability asks whether the person has enough information to identify possible errors, request corrections or seek review. Adequacy depends on the purpose and circumstances of the decision, not simply on the amount of information provided.

7. ANALYSIS AND DISCUSSION

7.1 Proposed material-context omission assessment

A material context omission is a potentially consequential gap in an otherwise accurate AI explanation. Its importance depends on the decision, the person affected, the evidence available and the actions that person can reasonably take. This is a proposed assessment framework rather than a claim that every unexplained detail is material.

Decision relevance considers whether missing information bears on a decision criterion or condition. Interpretive significance considers whether ambiguity, outdated records or missing background could change the meaning a person takes from an explanation. Usability asks whether the information can be understood and acted on. Contestability considers whether an omission makes it harder to detect an error, request a correction or challenge the decision.

The proposed Evidence-Linked Omission Register records seven elements: (1) the explanation statement; (2) the potentially missing context; (3) supporting evidence and its credibility; (4) the affected user, task and possible consequence; (5) an assessment across the four dimensions; (6) uncertainty and alternative interpretations; and (7) a recommended response. A response may be clarification, additional information, error correction, access to an appeal process or a finding that no further information is needed. The register supports traceable review; it does not automatically detect omissions.

The reviewer first checks whether the explanation is consistent with available model and decision records. Next, the reviewer examines relevant contextual evidence and identifies any apparent gaps. Each gap is assessed across the four dimensions. Materiality is judged by the possible effect of missing information on the decision or the person's ability to respond, with the supporting evidence recorded. An assessment should not require disclosure of another person's confidential information.

The register allows provisional outcomes such as 'material omission', 'contextually relevant but not currently material' and 'insufficient evidence'. These categories are judgments to be reviewed, not proven results. Reviewers should also check the accuracy, lawfulness and clarity of any proposed response.

7.2 Hypothetical applications

Consider a recruitment decision in which an applicant receives a lower score because the recorded employment history contains a gap. A reviewer would check whether the record is accurate, whether the gap reflects circumstances such as approved leave, and whether those circumstances matter under the employer's stated criteria. The reviewer would also identify how the applicant could correct the record or request a review. This hypothetical example illustrates possible questions; it does not establish discrimination.

Consider a loan application declined because the recorded debt-to-income ratio is high. The reviewer could verify the date and accuracy of the debt figure against reliable records. If the figure is outdated or incorrect, a correction route may be appropriate. If it is accurate and current, the explanation may be sufficient. The classification should remain uncertain when supporting records are unavailable.

7.3 Discussion and limitations

The proposal focuses on contextual adequacy and does not replace checks on whether a model explanation is faithful. Evidence-linked records may help distinguish documented omissions from speculation, although judgments can remain difficult when records are unavailable or stakeholders disagree. Providing more information is not always better: excessive detail can make an explanation harder to use.

The main limitation is the absence of empirical validation. Future studies could ask domain experts to classify possible omissions and measure agreement between reviewers. Experiments could compare explanations alone with explanations supplemented by verified context, measuring understanding, error detection and knowledge of review options. Perceived fairness should be assessed separately from demonstrated improvements in decision outcomes.

8. MAJOR FINDINGS

The conceptual analysis suggests that technical fidelity alone does not guarantee contextual adequacy. An explanation's usefulness depends on the decision and the needs of the person affected. Model behaviour, organizational policy and correction or appeal rights are different

kinds of information. An evidence-linked register may make suspected omissions easier to document and review, but its effectiveness remains untested.

9. SUGGESTIONS / RECOMMENDATIONS

Organizations and researchers could assess suspected omissions across the four dimensions and keep an evidence-linked record of their judgments. They should state the limits of available evidence, provide clear correction and appeal routes where applicable, record uncertainty and protect confidential information. Future research should test the framework with affected users and relevant domain specialists.

10. CONCLUSION

An accurate description of model behaviour may not provide everything a person needs to understand or respond to an AI-assisted decision. Drawing on explainable AI research, this paper proposes an evidence-linked approach to reviewing possible omissions through decision relevance, interpretive significance, usability and contestability. The approach offers a structured basis for documenting concerns and possible responses, but it requires empirical testing before claims can be made about its effectiveness.

REFERENCES

Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021). Does explainable artificial intelligence improve human decision-making? Proceedings of the AAAI Conference on Artificial Intelligence, 35(8), 6618–6626. https://doi.org/10.1609/aaai.v35i8.16819

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://arxiv.org/abs/1702.08608

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Towards a reflective sociotechnical approach. In Human and Machine Learning (pp. 449–466). Springer. https://doi.org/10.1007/978-3-030-60117-1_33

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

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3313831.3376590

Mill, E., Garn, W., Ryman-Tubb, N., & Turner, C. (2024). The SAGE framework for explaining context in explainable artificial intelligence. Applied Artificial Intelligence, 38(1), 2318670. https://doi.org/10.1080/08839514.2024.2318670

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

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. https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf

Wang, X., & Yin, M. (2022). Effects of explanations in AI-assisted decision making: Principles and comparisons. ACM Transactions on Interactive Intelligent Systems, 12(4), Article 27. https://doi.org/10.1145/3519266