

# Beyond Model Trust: Dual XAI for Adaptive and User-Centric Explainability

Maria Luigia Natalia De Bonis<sup>1,\*</sup>

<sup>1</sup>Electrical and Information Engineering Department, Polytechnic University of Bari

## Abstract

Current approaches to Explainable AI (XAI) often rely on static, one-size-fits-all methods that fail to meet the diverse needs of users. While AI experts require detailed insights to debug, refine, and optimize models, AI non-experts need intuitive, domain-relevant explanations to support decision-making. Consequently, a single explanation format is frequently either too simplistic for expert analysis or, in most cases, too technical to be meaningful for non-experts. This research introduces *Dual XAI*, a multimodal, human-centered framework designed to address these limitations. By integrating complementary explanation techniques with user modeling strategies, Dual XAI aims at providing explanations whose content, format, and level of detail are adapted to the expertise and specific needs of different user types. Grounded in human-centered design principles, Dual XAI enhances both interpretability and usability by providing personalized, context-aware insights. This enables AI experts to perform in-depth analysis of model behavior, while allowing AI non-experts to access accessible and actionable explanations.

## Keywords

Explainable Artificial Intelligence (XAI), Human-Centered AI, Quantitative Evaluation of XAI, AI Debugging and Optimization

## 1. Context and Motivation

Artificial Intelligence (AI) has become a central technology in various domains, including medicine, finance, and law, due to its ability to learn, reason, and adapt [1]. In recent years, advancements in AI models have led to increasingly autonomous and sophisticated systems, enabling them to tackle complex problems with unprecedented performance levels.

However, as AI models grow in complexity, particularly with the rise of Deep Learning (DL), understanding their decision-making processes has become significantly more challenging. Although early AI systems were relatively interpretable, modern architectures, composed of millions of parameters, are often regarded as *black-boxes*, making their internal reasoning opaque and limiting their widespread adoption [2].

This lack of transparency has fueled a growing demand for explainability and interpretability, with increasing pressure from stakeholders, regulators, and end users who require better insights into the reliability of AI model outputs. The absence of detailed explanations for a model's behavior can discourage its use, undermining trust and impeding AI deployment in real-world applications.

A notable example arises in the healthcare sector, where AI is used in high-stakes diagnostic procedures, such as radiological image classification. In these cases, it is a priority for clinicians to understand the reasoning behind a model recommendations before considering its outputs in clinical practice.

Trust in AI depends not only on explainability but also on the overall quality of the model, including its accuracy, robustness, and fairness. A clear explanation alone is insufficient if the system suffers from systematic bias or produces inconsistent outputs. Encouraging responsible adoption, therefore, requires an approach that couples explainability with rigorous assessments of performance, stability of predictions, and absence of systematic errors.

---

*Late-breaking work, Demos and Doctoral Consortium, colocated with The 3rd World Conference on eXplainable Artificial Intelligence: July 09–11, 2025, Istanbul, Turkey*

\*Corresponding author.

✉ marialuigia.debonis@poliba.it (M. L. N. De Bonis)

ORCID 0009-0007-2494-4806 (M. L. N. De Bonis)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The growing need for transparency has led to the emergence of Explainable Artificial Intelligence (XAI) [3], a collection of methodologies and techniques designed to enhance the interpretability of AI models without compromising their predictive performance.

**Challenges in current XAI approaches.** Despite its advancements, current XAI methods face substantial limitations. Many rely on a static *one-size-fits-all* approach, assuming that a single explanation can suit all users, regardless of expertise or context. However, explainability is not a monolithic concept: different users have distinct goals and cognitive needs. A common misconception is that applying an XAI method like SHAP [4] automatically makes a model interpretable for everyone. In reality, uniform explanations often prove too simplistic for experts or too complex for AI non-experts, limiting their practical value.

Overcoming these issues requires a paradigm shift toward flexible, human-centered approaches [5] that adapt explanations based on expertise and interaction context. It is essential to distinguish between two main categories of users: AI non-experts and AI experts, each with specific needs.

AI non-experts (e.g., clinicians, business decision-makers) need intuitive, interactive, and domain-relevant explanations to extract useful insights. AI experts (e.g., researchers, developers, data scientists), on the other hand, require detailed, technical explanations for debugging, bias detection, and model optimization - needs often overlooked in the literature.

Finally, explainability should go beyond trust-building and evolve into a tool for discovery and iterative improvement. To support this shift, it is crucial to develop quantitative, reproducible metrics that assess the stability, fidelity, and usefulness of explanations. These metrics can help AI experts select and optimize XAI techniques while enhancing knowledge extraction for AI non-experts.

**Towards a human-centered and multimodal XAI framework.** In light of these considerations, this research introduces *Dual XAI*, an innovative framework that embraces a human-centered and multimodal approach to overcome the existing limitations of explainability techniques. Dual XAI aims to provide flexible and adaptive solutions that cater to the specific needs of both AI experts and non-experts, ensuring that explainability becomes a practical and effective tool for real-world AI applications. Given the increasing role of AI in healthcare, this research will focus on improving the comprehensibility and clinical relevance of explanations for medical professionals, particularly neurologists, who are a group of users with whom we are collaborating in our research.

## 2. Related Work

**Definitions and conceptual foundations of XAI.** The definitions of interpretability and explainability remain widely debated in the literature [6]. To date, there is no unanimous consensus on what makes a model “comprehensible” in cognitive terms: while some authors view explainability as the ability of a system to provide high-level reasons supporting its decisions [7], others favor approaches more closely tied to the concept of fidelity, meaning the extent to which explanations accurately reflect the model true behavior [8]. This lack of agreement has led to fragmented methodologies and inconsistent evaluation criteria, making it difficult to determine whether an explainability technique genuinely improves interpretability or merely provides a superficial justification for the model outputs. Without a common conceptual foundation, XAI risks evolving in directions that fail to address the practical needs of users. Further complicating the landscape are ethical and legal considerations. In many domains, explainability is not just a matter of clarity; it is also an essential requirement for ensuring responsibility and verifiability of algorithmic decisions.

**Regulatory aspects of explainability.** The regulatory debate has highlighted the fundamental role of explainability in facilitating responsible AI adoption. The *AI Act* has placed legal constraints on the use of AI in high-risk applications, mandating transparency, fairness, and safety in decision-making processes [9]. Similarly, regulations such as the *GDPR* [10] have reinforced the “right to

explanation”, underscoring the obligation of organizations to ensure transparent AI-driven decisions. These considerations demonstrate that explainability is not merely a technical issue and an essential factor for regulatory compliance and societal acceptance of AI.

**XAI Methodologies.** In recent years, a variety of XAI methodologies have been proposed, broadly classified into two categories: *ante-hoc* methods, where the model itself is designed to be interpretable from the development stage, and *post-hoc* methods, aimed at generating explanations for pre-trained models [3]. Within the latter category, there are local approaches such as *LIME* [11] and *SHAP*, which approximate a model behavior around individual input points without relying on its internal structure, making them model-agnostic. In contrast, gradient-based methods like *Grad-CAM* [12] directly depend on the model architecture, leveraging its gradients and activation maps to identify prominent regions in neural networks. Other techniques include global explanations based on feature importance and counterfactual strategies.

**Need for hybrid and multimodal XAI approaches.** Despite this range of approaches, XAI methods are often applied in isolation, without any real integration among them. Each technique highlights specific aspects of the model, so the absence of a multimodal strategy restricts the ability to gain a comprehensive view of the model behavior [13]. Since every technique emphasizes different characteristics of the model, a multimodal or hybrid strategy would be more effective in meeting the varied needs of users [13].

A further distinction in the literature differentiates explanations aimed at building trust for end users, often called “*BLUE XAI*”, from those intended for debugging and model optimization by developers, referred to as “*RED XAI*” [13]. While the former approach prioritizes usability and transparency, the latter provides more technical, detailed analyses valuable for researchers and data scientists. However, most XAI methods fall at one extreme or the other, being either overly simplified and “user-friendly”, e.g. yet unsuitable for advanced analysis, or too technical and thus difficult to interpret for non specialists. Moreover, using standalone XAI techniques does not guarantee that AI-expert users will achieve a deep understanding of model behavior; rather, this level of insight calls for the complementary, multimodal integration of different methods that offer multiple perspectives on the model.

**Challenges in evaluating explainability.** Another unresolved issue is how to evaluate XAI explanations. Currently, such assessments predominantly rely on user studies and subjective questionnaires, which, while useful for gauging human perception, lack standardization. Efforts to introduce quantitative metrics, such as stability and robustness, have not yet led to a broad consensus. Longo et al. propose the “*XAI 2.0*” [14], which integrates metrics addressing precision, consistency, and utility, differentiated by user profile. Meanwhile, Biecek et al. [13] underscore the need for shared benchmarks to compare XAI techniques. The lack of widely accepted quantitative tools hinders standardization and practical adoption of these methodologies.

Concrete examples of the need for more comprehensive explanations are provided by Anders et al. [15], who demonstrate how interpretive methods can identify and mitigate spurious correlations (often referred to as the “Clever Hans effect”), thereby improving models reliability and robustness. Arya et al. [16] introduce *AI Explainability 360*, an open-source toolkit that brings together a variety of interpretive approaches (feature-based, instance-based, and global). Similar toolkits, such as *Alibi* [17] and *Captum* [18], already combine local and global explanations. While these solutions represent an important step toward multimodal strategies, they still lack adaptive and dynamic management of explanations, which would automatically tailor both content and communication style to different user roles and levels of expertise, i.e., AI experts vs. AI non-experts.

Taken together, these considerations highlight the need for a more flexible approach that can provide integrated, customizable explanations. As will be detailed in the subsequent sections, the **Dual XAI** framework proposed in this study aims to move beyond the traditional *one-size-fits-all* model, offering

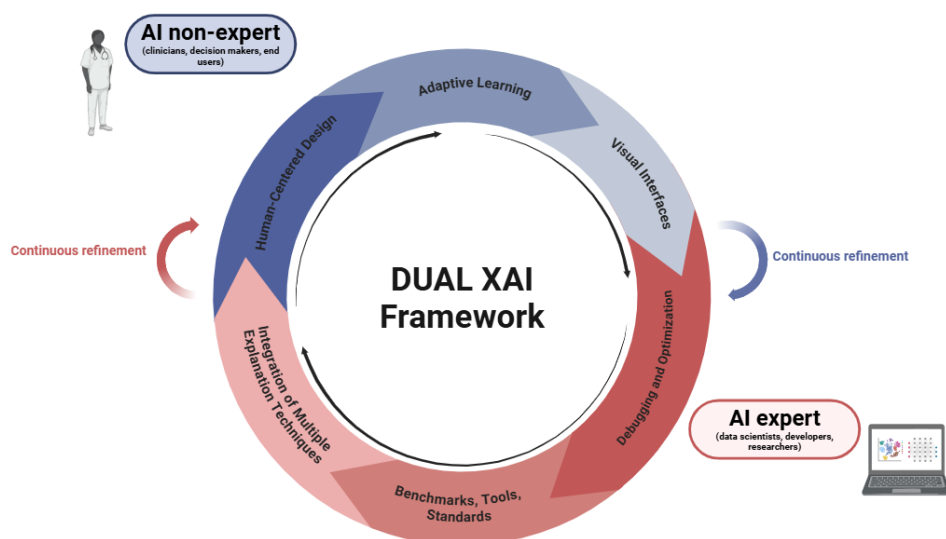
dynamic, multimodal solutions that simultaneously address requirements of transparency, trust, detailed analysis, and model optimization.

### 3. Research Questions and Methodology

In light of the gaps outlined in the previous sections, this research addresses three main questions, designed to fill some of the current voids in the XAI literature:

- **RQ1:** How can AI model outputs be made genuinely usable for *AI non-expert* users, moving beyond mere trust-building to provide domain-relevant, practical tools?
- **RQ2:** How can different Explainable AI techniques be combined to create multimodal and complementary explanations that meet the analytical and optimization needs of *AI expert* users?
- **RQ3:** Which quantitative metrics can be defined to objectively, comparatively, and comprehensively evaluate the effectiveness, robustness, and reliability of XAI techniques, taking into account the diverse requirements of different user groups?

To address these three areas of investigation, this work proposes the development of a framework called *Dual XAI* (Figure 1), structured in three phases corresponding to each of the research questions.



**Figure 1:** Dual XAI leverages a cyclical, user-driven process: insights from AI non-experts inform explanation design, which is optimized by AI experts and refined through tools, standards, and adaptive interfaces.

**Phase 1: requirements analysis and Human-Centered Design** . With the aim of addressing **RQ1** this phase adopts a **Human-Centered Design** (HCD) approach [19], integrating participatory design principles through co-design sessions where clinicians actively contribute to defining the most relevant aspects of explainability. These sessions aim to inform the development of interactive prototypes, ranging from conceptual sketches to functional versions of XAI interfaces. By varying key parameters such as the type of visualization (e.g., bar charts, heatmaps), the level of textual detail, and the possibility of executing *what-if* simulations, the study explores how different configurations impact usability and support clinicians not only in assessing the reliability of model outputs but also in actively engaging with the explanations. Through interactivity and navigable insights, the system would facilitate knowledge extraction, enabling neurologists to explore patterns, investigate alternative scenarios, and refine their understanding of the behavior of the underlying model, as well as their clinical questions.

**Interactive Visualization Techniques** and **Adaptive Learning** mechanisms are used to ensure adaptability to different levels of expertise. The system dynamically adjusts the level of explanation detail based on user expertise, offering more intuitive, high-level representations for residents while allowing specialists to access in-depth analyses when needed. Visual interfaces are designed to enhance interpretability, providing clinicians with the flexibility to navigate between summary-level insights and fine-grained feature contributions, ensuring that the explanations remain contextually relevant and aligned with their diagnostic reasoning processes.

The usability evaluation uses think-aloud protocols and task analysis, measuring both objective performance indicators (such as error rates, help requests, and time to completion) and subjective assessments via standardized questionnaires, including the User Engagement Scale [20], NASA-TLX [21], and AttrakDiff [22]. Beyond usability, the study would assess the impact of explanations on clinical decision-making by examining whether they improve diagnostic accuracy, reduce errors, and enhance the clinician's ability to interact with and interpret the provided insights. Qualitative insights from post-task interviews and focus groups further elucidate how the provided explanations influence reasoning, with a particular focus on how interactivity contributes to deeper exploration and hypothesis generation.

Findings from these evaluations inform an iterative refinement process, updating interface designs based on real-world user interactions and testing subsequent iterations to ensure optimal adaptation to clinical workflows and cognitive demands.

**Phase 2: integration and development of a multimodal framework** RQ2 is addressed through the design and implementation of the Dual XAI framework, which encompasses a platform capable of integrating multiple Explainable AI techniques (feature-based, instance-based, counterfactual, gradient-based) in a complementary and dynamic fashion. In this phase:

- An initial selection of XAI techniques (e.g., SHAP, LIME, Grad-CAM, counterfactual approaches) is carried out by means of a thorough literature review and a preliminary comparative analysis, evaluating their explanatory capacity for different model types (e.g., neural networks for image analysis or tabular models based on clinical features). Once identified, these methods are combined according to how they are most effectively presented, whether through a unified interface with multiple interactive panels, or a mechanism that allows switching between explanatory modes.
- Adaptive algorithms enable a robust **Integration of Multiple Explanation Techniques**, centered on the needs of AI expert users and supporting **Debugging and Optimization** activities (e.g., highlighting potential biases in MRI classification or in demographic clinical variables). The resulting integrated framework is tested with expert users in a controlled experiment, comparing (a) the multimodal approach, (b) a single-method XAI baseline, and (c) no explanation at all. This study assesses the rapidity of *AI expert users* in debugging, detecting model overfitting, and correcting hyperparameters, complemented by a qualitative investigation of whether multimodal visualization genuinely adds clarity or instead imposes excessive cognitive load.

**Phase 3: development and validation of quantitative metrics** In response to RQ3, this phase focuses on defining quantitative metrics to objectively assess the effectiveness, robustness, and reliability of XAI methods. The evaluation framework incorporates both existing techniques and, where necessary, new metrics tailored to specific explainability challenges.

Among the established approaches, the *Model Parameter Randomization Check* tests whether explanations remain stable under controlled perturbations of model parameters, while *Target Sensitivity* assesses whether explanations for different outcomes are truly contrastive. Additionally, *Stability for Slight Variations* measures robustness against minor input modifications. This study also explores novel metrics to address aspects of explainability that remain insufficiently captured.

The development of these evaluation metrics is accompanied by the creation of **Benchmarks, Tools, and Standards** to ensure a structured and reproducible assessment of XAI techniques. Empirical validation involves *AI non-expert users* (i.e., the neurologists) to align these metrics with real-world



analytical needs, ensuring their practical relevance. By integrating well-established evaluation criteria with new tailored metrics, this phase aims to contribute to a more comprehensive and standardized framework for explainability assessment, supporting both model optimization and the development of more reliable interpretability methods. The Dual XAI framework operates through a continuous feedback loop between the two user categories:

- Needs and interactions of *AI non-experts* provide valuable insights to refine XAI techniques used by *AI experts*.
- *AI experts*, in turn, enhance the quality and usability of explanations, ensuring that *AI non-expert users* receive explanations that are relevant, accurate, and tailored to their needs.

This iterative cycle fosters continuous improvement, refining explanations dynamically based on user-specific interactions and feedback.

## 4. Preliminary Results and Contributions

The research carried out in the early period of this doctoral program lays the foundation for the proposed **Dual XAI** framework by targeting two key challenges in XAI: (1) assessing the stability and reliability of model explanations (related to **RQ3**), and (2) integrating Human-Centered design principles to make AI tools truly usable for AI non-expert end users (supporting **RQ1**). In parallel, these efforts contribute to the formulation of a multimodal strategy, central to **RQ2**.

The progress thus far can be grouped into two interconnected lines of inquiry. First, a comparative analysis of multiple XAI methods for brain age prediction pipelines highlighted the variability in explanations, both within individual methods and between different approaches. Second, the development and pilot testing of an interactive tool for neurologists underscored the importance of a human-centered approach to ensure that AI-based systems are interpretable and usable, as well as tailored to the specific user, so that they can be practically relevant in clinical settings.

In the first study, titled “*Explainable brain age prediction: a comparative evaluation of morphometric and deep learning pipelines*” [23], we focused on assessing the stability and coherence of different post-hoc interpretability methods for brain age prediction. Specifically, we examined how changes in the reference background affected SHAP-based explanations and contrasted DeepSHAP with Grad-CAM for CNN-based models. The results showed that varying the background in SHAP led to significant shifts in feature importance, reducing the overall consistency of the explanations. Likewise, comparing Grad-CAM activation maps and feature-attribution methods (e.g., DeepSHAP) revealed key discrepancies, indicating that each approach emphasizes different aspects of the prediction pipeline. We also found that models based on morphometric features (e.g., cortical thickness, gray matter volume) often produce explanations more aligned with known neuro-anatomical markers than purely image-based CNNs. These findings collectively validated the need for a multimodal perspective on XAI, since relying on any single method risks overlooking critical aspects of model behavior.

In parallel, we sought to improve the usability and interpretability of AI models for clinical settings by developing and evaluating a web-based application called *Brain Age Predictor*. This tool integrates the deep learning model that, in our first study, demonstrated the highest stability and consistency of SHAP-based explanations under varying background configurations. Designed primarily for neurologists, the system provides a SHAP-based explanation module along with an interactive interface that enables users to visualize, edit, and simulate the effect of various morphometric features on predicted brain age.

Explanations are presented through two complementary visualizations, both focused on the individual patient level: a Tornado Plot, which displays the most influential features contributing to the predicted brain age for a specific subject, and a custom-designed Glass Brain, an interactive 3D visualization tailored for clinical users. The Glass Brain allows users to navigate and decompose SHAP values across anatomical regions, enabling spatial reasoning and in-depth neuroanatomical analysis. The design, implementation, and formative evaluation of this tool are presented in the paper “*Explainable AI for Brain Age Prediction: Design, Implementation, and Formative Evaluation of an Interactive Tool*”, which was recently accepted at the Hybrid Human Artificial Intelligence (HHAI) 2025 conference.

A formative study conducted with neurology residents reported the interface as generally intuitive, with participants particularly appreciating the “what-if” scenario feature (e.g., adjusting cortical thickness values to see how estimated brain age changes). Nonetheless, the evaluation also revealed the need for clearer SHAP representations and enhanced support for longitudinal patient monitoring. These findings underscore the importance of adopting interactive dashboards and adaptive explanation strategies, to ensure that even AI non-expert users can derive meaningful insights from complex ML models. Together, these preliminary results provide critical evidence that model explanations must be both methodologically robust (e.g., stable across varying reference backgrounds or architectural changes) and contextually tailored to the domain expertise of end users. From a methodological standpoint, the comparative evaluation emphasizes the value of complementary XAI strategies to capture different facets of a model’s decision process. Meanwhile, the usability study demonstrates that delivering intuitive, interactive interfaces is indispensable for facilitating trust and adoption among AI non-expert users, such as medical trainees or clinicians who require clinically actionable insights rather than purely algorithmic details.

## 5. Future Work and Expected Contributions

This research will continue to refine and validate the proposed **Dual XAI** framework, establishing it as a robust, adaptive, and user-centered approach to explainability. A primary focus will be the development of **quantitative evaluation metrics** that go beyond subjective assessments, providing reliable and reproducible measures of stability, fidelity, and contrastiveness to support AI experts in model inspection and optimization. In parallel, **user profiling strategies** will be further developed to ensure that explanations dynamically adapt to varying levels of expertise, task requirements, and usage contexts. Rather than relying on a binary distinction between AI experts and non-experts, Dual XAI will incorporate more nuanced user models to personalize content, format, and interactivity. The framework will also integrate **complementary and multimodal explanation techniques**, combining feature-based, instance-based, and counterfactual methods into cohesive, interactive interfaces. These will be tailored to clinical reasoning processes, with neurology serving as the initial application domain. By grounding explainability in measurable quality, contextual relevance, and personalized delivery, Dual XAI aims to transform explanations from abstract model outputs into actionable tools for decision support, making them effective and meaningful in high-stakes, real-world environments.

## Acknowledgments

I extend my gratitude to Prof. Tommaso Di Noia, my doctoral advisor, and Prof. Angela Lombardi, my co-supervisor, for their guidance and support. I am also thankful to Prof. Carmelo Ardito for his valuable insights. This work was partially supported by the project DEMETRA (CUP D99J22001970006) Missione 6/componente 2/Investimento: 2.1 "Rafforzamento e potenziamento della ricerca biomedica del SSN", funded by European Commission – NextGenerationEU.

## Declaration on Generative AI

The author has not employed any Generative AI tools.

## References

- [1] S. Makridakis, The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms, *Futures* 90 (2017) 46–60.
- [2] M. Christoph, *Interpretable machine learning: A guide for making black box models explainable* (2020).

- [3] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, *Information fusion* 58 (2020) 82–115.
- [4] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, *Nature machine intelligence* 2 (2020) 56–67.
- [5] B. Shneiderman, *Human-centered AI*, Oxford University Press, 2022.
- [6] Z. C. Lipton, The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery., *Queue* 16 (2018) 31–57.
- [7] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* 267 (2019) 1–38.
- [8] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, *ACM computing surveys (CSUR)* 51 (2018) 1–42.
- [9] European Parliament, Council of the European Union, Regulation (eu) 2024/xxxx of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act), <https://eur-lex.europa.eu/>, 2024. Accessed: March 27, 2025.
- [10] E. GDPR, General data protection regulation (gdpr), 2018.
- [11] M. T. Ribeiro, S. Singh, C. Guestrin, "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*, 2016, pp. 1135–1144.
- [12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [13] P. Biecek, W. Samek, Position: explain to question not to justify, in: *Proceedings of the International Conference on Machine Learning, ICML'24, JMLR.org*, 2024.
- [14] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. Del Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, et al., Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions, *Information Fusion* 106 (2024) 102301.
- [15] C. J. Anders, L. Weber, D. Neumann, W. Samek, K.-R. Müller, S. Lapuschkin, Finding and removing clever hans: Using explanation methods to debug and improve deep models, *Information Fusion* 77 (2022) 261–295.
- [16] V. Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović, et al., One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques, *arXiv preprint arXiv:1909.03012* (2019).
- [17] J. Klaise, A. Van Looveren, G. Vacanti, A. Coca, Alibi explain: Algorithms for explaining machine learning models, *Journal of Machine Learning Research* 22 (2021) 1–7.
- [18] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, et al., Captum: A unified and generic model interpretability library for pytorch, *arXiv preprint arXiv:2009.07896* (2020).
- [19] International Organization for Standardization, ISO 9241-210:2019 - ergonomics of human-system interaction — part 210: Human-centred design for interactive systems, <https://www.iso.org/standard/77520.html>, 2019.
- [20] H. O'Brien, Theoretical perspectives on user engagement, *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media* (2016) 1–26.
- [21] S. Hart, Development of nasa-tlx (task load index): Results of empirical and theoretical research, *Human mental workload/Elsevier* (1988).
- [22] M. Hassenzahl, M. Burmester, F. Koller, Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität, *Mensch & Computer 2003: Interaktion in Bewegung* (2003) 187–196.
- [23] M. L. N. De Bonis, G. Fasano, A. Lombardi, C. Ardito, A. Ferrara, E. Di Sciascio, T. Di Noia, Explainable brain age prediction: a comparative evaluation of morphometric and deep learning pipelines, *Brain Informatics* 11 (2024) 33.