



ARTICLE



<https://doi.org/10.1057/s41599-024-02759-2>

OPEN

Understanding the dilemma of explainable artificial intelligence: a proposal for a ritual dialog framework

Aorigele Bao^{1,2,3,4} & Yi Zeng^{1,2,3,4}✉

This paper addresses how people understand Explainable Artificial Intelligence (XAI) in three ways: contrastive, functional, and transparent. We discuss the unique aspects and challenges of each and emphasize improving current XAI understanding frameworks. The Ritual Dialog Framework (RDF) is introduced as a solution for better dialog between AI creators and users, blending anthropological insights with current acceptance challenges. RDF focuses on building trust and a user-centered approach in XAI. By undertaking such an initiative, we aim to foster a thorough Understanding of XAI, capable of resolving the current issues of acceptance and recognition.

¹ Department of Philosophy, School of Humanities, University of Chinese Academy of Sciences, Beijing, China. ² Institute of Philosophy, Chinese Academy of Sciences, Beijing, China. ³ Center for Long-term Artificial Intelligence, Beijing, China. ⁴ Institute of Automation, Chinese Academy of Sciences, Beijing, China. ✉email: yi.zeng@ia.ac.cn

Introduction

Progress in Explainable Artificial Intelligence (XAI) has been of great interest, particularly in light of the increasing use of AI systems in various fields, such as Transparency requirements (Szafron et al. 2006; Tamagnini et al. 2017), Case-based explanation methods (Bastani et al. 2018; Keane and Kenny 2019; Sørmo et al. 2005), Natural language explanation (Lakkaraju et al. 2017), and Counterfactual explanation (Guidotti 2022; Verma et al. 2020; Wachter et al. 2017). Drawing from the insights of Arrieta and colleagues, XAI is described as an AI methodology that proffers comprehensive details or rationales, aiming to render its processes transparency and readily understandable to users (Arrieta et al. 2020). Thus, XAI refers to the methodologies and techniques employed to transform the outcomes of AI technologies, such as machine learning, into formats understandable to human cognition during specific applications. The paper conceptualizes an AI model as a computational construct designed to perform specific tasks by processing data and making decisions. Within the paper, “explain” refers to providing clear, understandable reasons for a model’s decisions. “Explainable”, on the other hand, precisely its capacity to offer such transparent rationales in a manner comprehensible to human users.

Previous research on the user-centric XAI conceptual framework endeavored to bridge the gap between user explanation and artificial intelligence (Lim et al. 2019), yet it failed to consider the issue of XAI acceptance from the perspective of both the users and the creators of artificial intelligence. Despite the growing interest in XAI, user understanding of XAI remains a relatively unexplored area. It is crucial that users’ understanding of XAI is multifaceted and complex, and that it can significantly influence the effective social acceptance of XAI approaches. Differences in the understanding of XAI depend on the different outcomes of individuals’ perceptions and explanations of XAI. The complexity of this issue necessitates a deeper exploration of users’ understanding of XAI.

The first point to consider is that explanation and understanding are not the same things, although they are often used interchangeably. The main difference between them lies in the fact that explanation is an active process of disclosure, while understanding is a cognitive achievement that is reached by the users with the help of the AI model (Keil 2006). This distinction is essential because it helps us to see that the aim of XAI is not merely to provide explanations, but to enable the users to achieve understanding, which is the ultimate goal (Lombrozo 2006). It follows from this that the consideration of understanding is critical to the success of the XAI vision, as it is only through understanding that the user’s concerns can be alleviated. Therefore, it is not enough for the AI model to express its own internal structure or explain the results; it must enable the users to achieve the cognitive achievement of grasping the model through understanding. In order to explore this issue in more detail, it is essential to consider the role of analytical philosophy in the scientific discussion of understanding. The focus on cognitive achievement of understanding in analytical philosophy provides a useful framework for this analysis. Thus, this paper will examine the types of understanding implied by existing XAI technologies, in order to evaluate their success in achieving the goal of understanding. This evaluation will be based on the analytical framework that considers understanding as a cognitive achievement. Through this analysis, we can assess the success of the vision of explainability, and explore the implications of this success for the ethical goals of XAI.

The literature on the nature of understanding is multifaceted, presenting various perspectives on what constitutes understanding. Examples include the description of understanding as a

kind of knowledge (Evans 1982; Heck 1995; Higginbotham 1992) or as a belief (Millikan 2004) and the view that understanding is a state of perception in discourse (Barber 2003). The act of understanding involves entering a cognitive state, and this state is invariably linked to communication. To Understand a given discourse or concept is to acquire something, which is then integrated into the state of understanding as a kind of knowledge. Therefore, the communicative nature of understanding implies that it is a relational process that necessitates interaction between two entities, one of which must be a human agent, since understanding ultimately stems from our own cognitive apparatus. As such, understanding is inherently relational, and it follows that there must be a process by which understanding is acquired and developed, which culminates in a comprehensive understanding of the subject matter.

The distinction between internalism and externalism of understanding, as posited within the analytic tradition of philosophy, is a topic of significant philosophical interest (Grimm 2021). Internalism claims that understanding is primarily a relation between one’s object representation and the object one understands, and that this relation takes place in one’s inner world (Kim 1994). Externalism, by contrast, maintains that understanding is a relationship that is acquired in the real world. According to this view, what is grasped is a noumenal relation or a relation of dependence, rather than a purely semantic one (De Villiers et al. 2014; Dellsén 2020; Greco 2014; Kim 1994). The externalist understanding of understanding emphasizes the role of the external world in shaping our understanding of objects and concepts. In terms of the process of understanding, externalism views it as akin to completing a jigsaw puzzle. By grasping the different relations between the elements of a particular object or concept, a picture of the world with its rich content is formed. This process is one that occurs over time and involves a continual refinement and updating of our understanding in light of new experiences and insights. Ultimately, the externalist understanding of understanding emphasizes the importance of the external world in shaping and enriching our understanding of the objects and concepts that make up our world.

According to both internalist and externalist analyses, understanding hinges on the relationship that establishes how one acquires knowledge. The present paper does not intend to engage with the debate between these two positions, but rather to examine the understanding of XAI models by highlighting the central role played by the concept of “relationship”. In this context, the term relationship refers to the connection between things in the world that can be grasped by the human mind.

Within the context of XAI, this dependency can be divided into three distinct modes of understanding, namely, contrastive understanding, functional understanding, and transparency understanding. Contrastive understanding is a type of user understanding implied by XAI techniques such as counterfactual methods, while functional understanding is a type of understanding implied by XAI techniques related to functional disclosure. Transparency understanding is a type of understanding implied by the transparency commitment to the structure of artificial intelligence technology.

Contrastive understanding focuses on addressing “why” and “why not” questions, or hypothetical “what if” scenarios. By juxtaposing factual outcomes with counterfactual scenarios, it articulates to the user the rationale for preferring one particular solution over other possibilities. Functional understanding refers to the understanding of the efficacy of achieving a specified goal. Transparent understanding is an understanding that ensures the user’s right to know, and focuses on the understanding of whether or not it will affect one’s well-being for the user. Functional understanding has something in common with transparent

understanding, as deciphering the intended utility of a model requires a deeper understanding of its internal gears. However, functional understanding contrasts with contrastive understanding, which focuses more on outcomes and less on the overall goals of the model. In terms of goals, contrastive understanding aims to facilitate understanding by articulating the different outcomes produced by an AI paradigm, whereas functional understanding works to elucidate the overall efficacy of such a paradigm.

By analyzing these three paths of understanding, this paper aims to demonstrate the challenges faced by XAI in enhancing the Understanding of their target users. One possible solution to this problem is to add a description of the conversation framework of understanding. Use the conversation framework to drive users' social trust in XAI.

Contrasting understanding

The explanations of AI models, particularly the counterfactual, case-based ones, assume a Contrasting understanding. This understanding involves explaining the results of AI models in a teleological sense, i.e., tracing the causality of the outcome in the reverse direction from the posterior results. The counterfactual question associated with this type of understanding pertains to the possibility of other outcomes generated by the AI model, leading to a distinction between possible worlds (Lewis 1973). This counterfactual 'possible' question aims to answer the query "why output this outcome and not another?" This line of questioning emphasizes the causal traceability of the model results (Lipton 1990; Miller 2019). The explainer and the users discuss and compare possible outcomes around the counterfactual question, examine the impact of the model results on the users, and secure the interests of the users in a new round of model input and output. The process of explanation thus becomes a circular transfer activity (Miller et al. 2017). Contrastive understanding involves questioning the credibility of an AI model's outcomes, with disclosures needed from the AI to address doubts and demonstrate the protection of users rights. Although different possible outcomes are not directly linked, their characteristics reflect the shared understanding of the users (Arrieta et al. 2020).

A salient advantage of contrastive understanding is its non-intrusive approach to the internal mechanisms of artificial intelligence models. In XAI techniques, enhancing explainability often necessitates substantial alterations to the model's architecture, potentially compromising its performance or prediction accuracy. Such modifications might encompass the simplification of model layers, adjustments to weights, or even retraining the model with explainability as a primary criterion. In stark contrast, contrastive understanding circumvents these complexities. It primarily operates at the model's output layer, offering post-hoc explanations for decisions without delving into or altering the foundational algorithmic structure. This superficial manipulation considerably reduces the demands on computational resources and domain-specific expertise, as well as the time required for in-depth model modifications. The paramount quandary confronting conventional XAI pertains to the inability to reconcile the efficacy of explicable techniques with the efficiency of AI models. The dialectical tension between ethical and technical imperatives necessitates that AI model developers make a definitive choice between the two (Došilović et al. 2018). A model that is highly accurate might lack transparency, making its decisions hard to explain. Conversely, easily explainable model might not offer the same level of accuracy as its complex counterparts. In our exploration of AI's contrastive understanding, this tension between technical prowess and ethical responsibility becomes even more pronounced (Zerilli 2022). Contrastive

Understanding-based XAI posits that it can attain its objective of explainability without having to make a binary choice between the aforementioned approaches, by means of comparison.

According to Miller, human need for explanation is selective, and people prefer to choose a few of an infinite number of reasons as explanations (Miller 2019). The concept of contrastive understanding satisfies the public's need for selective reasoning through a selective explanatory framework. The comprehensibility of an AI model's output is directly proportional to its intelligibility. Individuals are not concerned with the technical aspects of how an AI model is generated; rather, they are interested in understanding whether the model's results have implications for their own lives (Miller 2019). As such, contrastive understanding provides a viable solution for decoding the complex "black box" of AI. The contrastive Understanding enables the users to make a direct effort to grasp the inferred outcomes, obviating the need for a comprehensive understanding of the entire AI model, thus mitigating the user's cognitive load. The users need not possess a comprehensive Understanding of AI, but rather solely answers to specific sensitive questions, such as the methodology of the model's facial recognition with regard to safeguarding privacy. Contrastive understanding is viewed as a pre-established, selective explanatory approach and a pathway that satisfies the desire for unencumbered technology. The adoption of contrastive understanding in AI models facilitates the preservation of data confidentiality, thereby preventing disclosure of sensitive information. In particular, when comparing model outcomes, it is unnecessary to extract privacy-related data separately; rather, comparing the output results suffices for achieving the intended objective (Adadi and Berrada 2018).

The counterfactual explanation method is an XAI approach that aims to clarify the differences between predicted and actual results of AI models (Byrne 2007). It is based on contrastive understanding and is considered a typical approach to achieve explainability in AI. The main challenge faced by this method is the explanatory distance between different outcome explanations, which refers to the distance between possible outcomes (Barocas et al. 2020). The possible outcomes of a counterfactual explanation can be represented as a decision tree-like explanative framework. Different inputs may lead to different explanative outcomes, and through contrastive understanding, makers and users can achieve a better understanding of AI models. By exploring various counterfactual outcome explanations, the model can be revised, and the direction of possible outcomes can be determined (Byrne 2019). This explanatory loop is based on distance corrections, which enable the deconstruction of black box outcomes. Contrastive understanding does not require the actual deconstruction of the black box. Instead, it involves comparing different possible outcomes through simulation, approximation, or counter-explanation to achieve an explanation of the outcome (Wachter et al. 2017). This approach can help to achieve explainability in AI models, even in cases where the internal workings of the model are not fully transparency.

However, the limitations of AI explanations based on contrastive understanding are apparent. While such explanations prioritize the output of AI models, they overlook the explanation of features within other AI models (Barocas et al. 2020; Páez 2019). In other words, while it possesses the capability to elucidate "why not B," it may not necessarily expound upon the broader functionalities inherent to artificial intelligence—specifically, the fundamental principles, logic, and interdependencies among various features within a model. This becomes particularly evident when the AI model in question is a sophisticated system endowed with multilayered and nonlinear operations, such as deep neural networks. For instance, should two distinct AI models arrive at identical predictions for a given dataset, a

contrastive understanding can illuminate the reasons underlying each model's specific decision-making process. Yet, this perspective falls short in explicating commonalities or disparities in features, layers, or data processing techniques between the two models. Recognizing these nuances in Understanding is imperative, especially when assessing the robustness, fairness, or potential biases of AI models. As a result, contrastive understanding fails to deliver a comprehensive account of AI models.

Moreover, XAI that employs contrastive understanding acknowledges only the need for selective human understanding, neglecting the importance of complete human understanding (Kulesza et al. 2013). The partial nature of AI explanations based on contrastive understanding renders relevant XAI techniques inadequate to accomplish the objective of unveiling the black box.

Secondly, the problem with contrastive understanding lies in its deterministic attitude, which regards AI models as mechanistic circular processes. As the sole actor involved in the implementation of XAI technology, the users is under pressure to determine the usefulness of the technology, and this pressure may lead to a sense of responsibility to accept the results. Regardless of the success of the explainable approach based on contrastive understanding, the users is pressured to acknowledge that the designer has performed the necessary work, despite their lack of understanding of the AI model. By leaving the black box unopened, AI model designers avoid taking responsibility for their designs, resulting in the black-boxing of design responsibility. The burden of responsibility creates frustration and the demand for further development of XAI. Contrasting paths of understanding create a new imbalance of responsibility between the designer and the users. The inability of the explainable approach to explain itself leads to additional pressures to pursue ethical processes (Gilpin et al. 2018; Shulner-Tal et al. 2022).

Moreover, the contrastive understanding suffers from a fundamental problem, namely its commitment to a relativistic conclusion. When comparisons are made, selecting different models for comparison with reality yields different results. The correctness of each answer is guaranteed by the corresponding contrasting models, such that each answer is correct in its own right. However, the fact that correct answers do not contradict one another results in a lack of verifiable criteria, rendering evaluation criteria arbitrary. Furthermore, the relativist dilemma of the contrastive understanding manifests in the difficulty of further mutual verification of different explainable frameworks. Each explainable approach under the contrastive understanding contains model specificity, making it difficult to discuss and evaluate models across different technical approaches. This renders communication between different frameworks challenging.

Functional understanding

Compared to contrastive understanding, functional understanding has a greater propensity to provide a comprehensive description of the AI model's functions. Functional understanding assumes that the output of an AI model is designed to fulfill a specific function, and that the users can explain the AI's results by accepting a clear definition of the model's function. At the outcome level, functional understanding requires an assessment of whether the output accomplishes the model's function, thereby measuring the model's effectiveness (Páez 2019; Lombrozo 2010, 2016).

According to Lombrozo and Wilkenfeld, functional understanding of an event relies on an understanding of function, goals, and objectives (Lombrozo and Wilkenfeld 2019), and functional understanding is a functional explanation that is future in terms of reverse causality. Reverse causality terminology is expressed as the use of functional terms to explain AI model outcomes

through future expected outcomes (Lombrozo and Wilkenfeld 2019). The reverse dependency feature explains that the "causal account" in the functional understanding of AI models is different from the actual causality, but pursues an understanding of the design's predetermined function, free from the dependency on causal knowledge, to obtain an account of the functional implementation (Lombrozo 2007). Another characteristic of functional understanding is that functional explanations need to take into account the needs of the explanatory users to explain the function of a particular AI model. And based on the needs of the users, the effectiveness of the explanation is enhanced.

Functional understanding provides several advantages. Firstly, an XAI model enables the users to Understand the design function of the AI model and its proper functioning through functional understanding. The users does not need to accumulate knowledge of AI models but only understand the explanation of functional implementation, avoiding the requirement for a high level of technical knowledge to achieve functional understanding. Secondly, the unique causal dependencies of functional understanding are not restricted to systemic requirements but are grounded in functional explanatory models that are more accessible to the users for knowledge transfer (Lombrozo and Wilkenfeld 2019). Thirdly, the strength of functional understanding is demonstrated by the robustness of functional dependencies in the face of external interference. Instead of an explanation of principles, functional regularity is ensured through the realization of causal dependencies, which contributes to the reliability and accuracy of the model. In summary, functional understanding offers benefits to the users by simplifying complex AI models and making them easier to understand, as well as by providing a more robust and accurate understanding of the AI model's functional dependencies.

However, functional understanding also faces problems. Firstly, functional understanding of an AI model may not necessarily result in effective explainability. In the process of functional understanding, the explanation of model functions does not necessarily lead directly to the understanding of model outputs, nor does it necessarily lead to the explanation of model features. If model features are unrelated to model outputs, then functional understanding becomes a constructed understanding that is detached from the model's state. Pursuing explainability through functional explanations may lead to irrelevant answers, which can be unreasonable or dangerous. Gilpin et al. argue that the risk of misexplanation arises from these irrelevant answers, and while the goal of functional understanding is to protect the power of the explaining users, these additional answers can become deceptive and ultimately endanger the user's rights (Gilpin et al. 2018). Therefore, it is important to exercise caution when pursuing functional understanding as a means of explainability.

Secondly, the concept of anticipatory design understanding of reverse causation is based on functional understanding. However, this approach is not without its shortcomings. In particular, it is unclear how functional understanding can be effectively applied to protect users rights when making anticipatory future output judgments. Additionally, the functional understanding approach does not adequately address the internal systemic responsibility of AI models in producing explainable output. By solely relying on a functional account of the model, the factors at stake for the users are not fully discernible, and the explanatory validity is incomplete. As a result, the discussion of responsibility in functional understanding is hindered by the same problem as in contrastive understanding, where accountability cannot be attributed effectively (Páez 2019). This issue also causes the pursuit of accountability to lose momentum and become an exercise in explanation for the sake of explanation.

Third, the steady-state model process is considered a fundamental component in achieving a functional understanding of a system. Exceptions to this process are generally excluded from the explanatory framework, leaving the functional outcome of the steady state as the primary focus. Any exceptions to the steady-state model process are often treated as a black box lacking a clear model process account. It should be noted, however, that exceptional conditions may introduce some level of bias in functional understanding. In the specific context of explainability for AI models, functional understanding may neglect processes that are not part of the steady state, thereby exhibiting a local bias. This bias arises from the tendency of functional understanding to rely on human explanation, which may not always be objective or independent. It is crucial to recognize that functional understanding represents only one explanation of AI models, and independent knowledge explanation beyond the non-biased functional explanation is needed to ensure a complete and unbiased understanding of the system. Therefore, the functional understanding obtained by the users may also be biased, as it reflects the selection of the functional explanation based on human biases (de Bruijn et al. 2022).

The central paradox of functional understanding in the development of XAI lies in the conflict between the public's desire for transparency and protection of their rights. The designer of an AI model holds significant design power, including the ability to determine the level of transparency provided to the users and how the model's impact is communicated to the public. The designer's ethical obligations involve choosing between a public-biased understanding and a public-advocacy understanding, which can lead to accusations of immorality and irresponsibility or require consideration of how to effectively untangle the dilemma of explaining knowledge. Moreover, functional understanding is limited by the problem of reductionist simplification, where the emphasis on function disregards other relevant factors within the system that may influence the user's understanding. This incompleteness of the testimony used to form understanding can lead to a bias in explanation, as the functional testimony does not present a comprehensive description of the AI model's internal complexity. Consequently, simplified understanding cannot be considered a reliable objective representation of the model.

Transparency understanding

The call for transparency in AI applications is linked to responsible behavior, which involves protecting users' rights and explaining the legitimacy of the technology. Ananny and Mike argue that transparency enables us to see how new phenomena create opportunities and obligations, and to hold them accountable (Ananny and Crawford 2018). By opening the black box of AI, transparency can provide a realistic and functional way to ensure users' right to know.

Transparency refers to the level of disclosure about the internal workings of a technology, both for users and makers. The degree of transparency varies, from minimal transparency where users only perceive the outcome, to complete transparency where the underlying code and input data are fully disclosed. In practice, the level of transparency is an indicator of the technology's flexibility. Minimal transparency, which only requires users to understand the output outcome, risks intentional deception and distortion of results. On the other hand, complete transparency raises privacy concerns and may overwhelm users with technical details that are difficult to understand. Therefore, striking a balance between users' understanding, privacy protection, and ethical obligations of XAI is crucial. Achieving this balance requires a level of transparency that meets users' needs for understanding while

also ensuring that privacy is protected and that the ethical requirements of XAI are met.

The user's understanding of transparency is crucial for explaining the AI model. It helps to make sense of the model's uncertainty and address concerns about protecting users' rights. A key issue for users is ensuring cognitive openness in processes that impact them. This involves reducing barriers to understanding AI models and establishing their reliability. A responsible design attitude requires a commitment to transparency and openness, which can be achieved through dialog mechanisms between users and model designers. By facilitating two-way communication and puzzle-solving, dialog mechanisms establish a sustainable mechanism for transparency within the AI model.

Transparency understanding relates to the functional integrity requirements of AI models when geared towards an users. Transparency assessments are directly related to the integrity of the process functional understanding of the explanatory users (Felzmann et al. 2020). Here, the difference between the processual functional understanding pursued by transparency and the functional understanding is that while the functional understanding explains in a reverse causal sense, the processual functional understanding pursued by transparency explains mechanistically in an internal open-ended sense. Transparency assessment is directly linked to the complete representation of the internal mechanisms of the system, and the degree of perfection of the representation affects the outcome of the transparency assessment.

Viganò et al. argue that transparency in the design of AI models should distinguish between the goals of the computer scientist and the goals of the users, and that the goals should be consistent (Loi et al. 2021). This requirement presents the value consistency requirement of a dialog mechanism. The goals of the model designer, as the two parties in the dialog mechanism, are based on direct knowledge of the user's goals, requiring that the dialog between the two is conducted in a broad value alignment. In the relationship between the two parties, the AI model appears to exist as a dialog mediator, but in the specific act of dialog, the AI model becomes a direct mechanism for reflecting the progress of the dialog, and the AI model transparency becomes a realistic feedback to the dialog mechanism. The normative goal of value congruence contributes to users transparency understanding in line with dialog expectations and advances the requirement for AI model transparency as a value pursuit. Broader transparency understanding expectations are also directly related to model descriptions. The model description of a dialog is directly related to the progress of technical translation of AI models as they undergo explainability work. The pursuit of transparency is therefore seen as a distinct virtue in an ethical sense (Felzmann et al. 2020).

Understanding transparency in AI models is a challenging task, primarily due to the diverse needs of users. Each user has a unique explanation of AI models, and to build highly adaptable transparency, it is necessary to consider different purposes. However, due to the variability of user requirements, achieving consensus on transparency is difficult. transparency commitments face dynamic challenges that go beyond simple comparative and functional explanations. While modeling the inputs and outputs of a system, it is necessary to consider their content, and the perception of steady-state transparency should be avoided. Additionally, dynamic exceptions should not be ignored, and the differences in outcomes due to various input attributes should be taken into account. Failure to do so can undermine the promise of fairness and make it impossible to reconcile transparency.

Secondly, transparency is a complex objective that requires specific commitments to XAI models. Such models are perceived to represent an objective and disinterested perspective, which can influence the judgment of users accessing the decision in further

decision-making (de Bruijn et al. 2022). While the credibility of AI models is usually considered to be “for information only,” the presence of a predetermined answer can influence the evaluation of the AI outcome. Consequently, the cognitive responsibility that decision influencers bear is passed on to the AI model, which becomes a “fictitious duty bearer.” This poses a risk as it reduces the likelihood of critical advice for objective, disinterested commitments. In this way, XAI models become explainive power holders and influence individuals involved in a wide range of decision-making scenarios (de Bruijn et al. 2022).

Furthermore, cognitive penetration may affect the process of understanding transparency in AI. Transparency in AI does not function like a mirror that directly reflects the agency within an AI model. Instead, every testimony conveyed introduces a certain top-down bias into the explanation (de Bruijn et al. 2022). Therefore, XAI promised by transparency may not always provide a complete understanding of the AI model, as additional content can influence it. Thus, relying solely on XAI can lead to conclusions that are influenced by personal opinions and biases, rather than being based on a sound understanding of the AI model.

Although transparency is crucial in promoting user trust in AI models, it is important to note that normative social trust and acceptability issues should also be considered in transparency reviews (Felzmann et al. 2020). Communicative knowledge is essential in fulfilling the commitment to transparency, but it is necessary to distinguish between the overall assessment of communicative knowledge and a singular explanation of transparency. The application of transparency is more complex in various contexts, such as specific knowledge and decision-making processes, where biases in results may emerge, potentially impacting the overall evaluation of transparency.

The high level of conflicting opinions undermines transparency commitments and explanations, especially if they are directly linked to public opinion. The existence of preconceptions in public opinion is evident, and they often result in a biased explanation of transparency when it comes to AI models. This biased explanation can have a profound impact on the approach taken by the AI model, as it turns the seemingly objective process of transparency explanation into an external and emotional requirement. Therefore, it is essential to recognize the impact of preconceptions and biases in public opinion and address them accordingly to ensure a more objective and transparency explanation of AI models.

Thus, transparency commitments for AI models entail not only technical requirements but also social and ethical challenges (de Bruijn et al. 2022). These challenges include interventions that can enhance the value of transparency and ethical assessments to determine the degree of transparency required. In contrast to functional and contrastive explanations, transparency commitments involve complete dialogical processes, which makes XAI models more susceptible to social and technical challenges in interactive settings. To evaluate the acceptability of transparency commitments, it is necessary to consider the dialogical relationship between the explaining users and the explainer. Additionally, conducting a valid ethical assessment of transparency presents further challenges, such as cognitive acceptance (Rohlfing et al. 2020). Thus, developing a dialog model is crucial for overcoming these challenges and realizing the potential benefits of transparency commitments.

Ritual dialog framework

The principal technological objective of XAI lies in securing user trust by unveiling assurances that the artificial intelligence model will not engender detrimental effects upon the user. Regrettably,

the current level of implicit Understanding regarding XAI is inadequate to realize this ambition. Consequently, the understand of XAI must be predicated upon a highly sensitive foundation towards user trust, undertaking a thorough assessment thereof (Miller, 2023). This necessitates not only the safeguarding of user rights, but also an evaluation of the impact exerted upon the artificial intelligence model by the societal relations between its creators and its users.

Drawing from the insights of Kaplan, Kessler, and Hancock, trust is invariably described as involving a principal in a vulnerable position and the trustee they depend upon. This characterization underlines a notable asymmetry in the statuses of the two parties central to trust (Kaplan et al. 2020; Lin et al. 2020). In the findings of Lee and See, trust is conceptualized as the expectation that an agent will aid in achieving individual goals amidst uncertainty and vulnerability (Lee and See 2004; Millikan 2004). Their discussions shed light on the multifaceted nature of trust, suggesting it might be perceived as a belief, attitude, intent, or behavior. Furthermore, Kaplan, Kessler, Brill, and Hancock offer a meta-analytical perspective which underscores that antecedents of trust engage with human trustors, technological trustees, and mutual contextual factors collectively (Kaplan et al. 2023; Miller 2023). Herein, it is discerned that beyond the trustee and the principal, a common backdrop, such as the objectives of Explainable Artificial Intelligence (XAI), ascends as a pivotal element of trust. In this context, “trust” refers to the users’ anticipatory stance that XAI will accomplish goals in a manner understandable and acceptable to them. This form of trust is influenced by the dialog between the user and the creators of the AI models. Trust contemplates not only the credibility of XAI itself but also the trustworthiness of the AI model creators.

Inspired by the work of Miller and Hilton, we argue that dialog has an important role in trust in XAI (Miller 2019; Hilton 1990). In order to promote user trust in XAI, a Ritual Dialog Framework (RDF) is needed. An imperative rationale for dialog between the creator and the user is rooted in the difficulty users face discerning between explanations generated by machines and those produced by humans, an issue clearly manifested in studies regarding the transparency Turing test (Biessmann and Treu 2021). The user’s inability to distinguish between these two types of explanations underscores the necessity for interaction between the user and the creator, thus alleviating the user’s apprehension of deception. The three pathways to understanding suggest that an open, wide-area, full life-cycle XAI dialog mechanism should be based on an effective RDF. The significance of the RDF lies in our endeavor to elucidate that Understanding ought to be teleological, regardless of the nature of explanation. Consequently, the RDF acknowledges the societal milieu in which artificial intelligence models are situated, realizing that the credibility of AI systems frequently hinges upon societal assessments grounded in collective experiences with technology, authority, and shared engagements. Hence, the RDF is not merely a communicative strategy but signifies a pioneering sociotechnical structure that draws inspiration from the idea of ritual. In this context, accomplishing a dialog takes precedence over understanding, and understanding here does not allude to the user’s Understanding of the model, but their grasp of the creators. What users obtain from explainable artificial intelligence is the precondition for trust, rather than novel knowledge. This trust necessitates satisfaction via a comprehensive RDF, implying that the framework plays a ceremonial role here. The dialog framework acts as an intermediary from understanding to trust, helping to reach trust between the user and the XAI and the maker.

When delving into the notion of “rituality” within AI interactions, we are inspired by Turner’s insights that rituals symbolize practices undertaken by communities or individuals to either

establish or symbolize specific sociocultural realities or norms (Turner 1967). Turner's analysis of the Ndembu rituals illuminates symbolic communications and performative acts that generate societal meanings. In our proposed RDF we extrapolate this concept to the interactions between humans and AI systems. Within this paradigm, the term "ritual" does not merely denote ceremonial actions, but encapsulates the creation of a structured, symbolic interaction system actions intended to bridge the trust chasm between humans and AI models. In the RDF, dialog embodies this symbolism, signifying more than just an exchange of information—it embodies a shared commitment to transparency, empathy, and the democratization of AI knowledge. Such dialogs are not transient; they are continuous rituals that evolve with AI advancements, ensuring understanding and trust are not forsaken in the face of innovation.

Rooted in the fertile grounds of ritualism, the RDF acknowledges and addresses symbolic representations of trust, cultural nuances, and societal intricacies. This framework carves out a structured arena, fostering meaningful dialogs between AI system developers and users, affirming mutual roles, values, and stature within interactions. Creators are not just technical reporters; they are narrators of the AI model's journey. This involves sharing inspirations behind the model's creation, challenges addressed, and reasons for methodological choices. For instance, developers might elucidate their rationale for choosing specific datasets for training, expound upon challenges encountered, and detail strategies employed to overcome them. Such a narrative approach allows users to perceive the model not as an enigmatic black box but as a dynamic narrative with its own genesis, challenges, and resolutions.

Distinctly, RDF acknowledges that trust in AI stems not just from the system's technical reliability or efficacy but also from users' sense of participation and Understanding of the AI decision-making realm. By reconceptualizing interactions with AI as dialogical rituals, RDF situates human-AI interactions within familiar sociobehavioral frameworks, thereby assuaging apprehensions and striving to cultivate trust. Thus, the rituality of RDF is not mere superficial repetition, but a consistent reassertion of the user's right to understand the AI they interact with—a perpetual commitment to ethical AI practices by developers. Implementing RDF necessitates thoughtful contemplation of the most salient "rituals" for users.

The effective unfolding of the RDF involves a relationship between both the subject of understanding and the subject of explanation, exploring paths of social understanding relationships embedded in social values from the dialog between the two parties. The establishment of fruitful dialog among parties necessitates a fundamental requirement, namely the existence of a shared objective. A dialog can only ensue if there is a clear convergence of goals among the participants. Furthermore, this convergence of objectives serves as a criterion for one's inclination to participate in the dialog. It can be contended that dialog implies the presence of shared rationality (Davies 2007) and the active contribution (Jeremy 2002; Miller 2023) of all parties involved. Hence, a clear willingness to engage in the process is an essential prerequisite for successful dialog.

In order to establish a meaningful dialog, both parties need to take turns (Sacks et al. 1978) to articulate their views through language while being receptive to the other's expressions (Grodiewicz 2021). This exchange of ideas demands a reciprocal exchange of feedback where each participant communicates their thoughts and opinions. This communicative process is underpinned by the internal structure of the dialog, which encompasses the Understanding of language and the cognitive states of both parties (Longworth 2018). Additionally, in most dialogic scenarios, it is imperative to not only attend to the verbal expressions of the interlocutors but also to construct novel beliefs based on the other's discourse. This implies that the RDF we advocate is not to illustrate that the maker's

intervention will render artificial intelligence models more transparency, but to demonstrate that the maker's purpose should be wholly biased towards user benefits. The augmentation of the dialog framework, as a form of ritual, signifies the maker's commitment to trust in their conduct, and is willing, based on this pledge of trust, to exhibit their goodwill over an extended period.

At the onset of cognitive Understanding, the discursive faculty is predominantly focused on the discourse's substantive content. Within a RDF, Understanding connotes the capacity to decode and explain the interlocutor's speech (Hauser et al. 2002). This process of acquisition is not a static outcome; rather, it is a complete linguistic cycle that is consummated through a constant exchange of speech between the two parties (Grodiewicz 2021). Within this cycle, trust and Understanding towards the creator hinge upon the introduction of a novel language, fostering the attainment of this objective through an iterative process. As we traverse through the linguistic delineation, trust ultimately achieves a transformed state. Thus, the trust manifested post-discussion appears as an emergent state of understanding, a condition that surfaced right from the onset of the dialog, advanced through the rotation of speech, culminating in a cognitive realization of trust. The dialog process, enveloped within Understanding, stimulates the cognitive fruits of trust.

A more explicit empirical perception reveals that the process of dialog between two individuals does not allow for ample time for careful consideration of each other's words. As a consequence, trust does not occur in isolation during the exchange, but rather transpires concurrently with the act of listening or reading (Levinson 2016). Hence, the process of understanding is not an entirely unfamiliar process, but rather involves the anticipation of potential language (Kuperberg and Jaeger 2016). Thus, the unfolding of understanding involves a progression of states and ensures that the initial attainment of Understanding is not lost when a new conversational understanding emerges, but rather reaching trust through a sequence of predictions and the integration of new exchanges (Longworth 2009).

The attainment of such trust necessitates a gamut of fundamental prerequisites previously alluded to. Primarily, the establishment of a unified dialog objective is indispensable for maintaining effective discourse between the creator and user. Secondly, both parties must manifest a willingness to align with and participate in the dialog. Thirdly, the capacity to partake in ongoing discourse is of paramount importance for the creator. Lastly, the proficiency to predict discourse and the potentiality for future interactions must be present.

Together, these requirements facilitate the transition of users from implicit three modes of understanding towards a trust understanding of XAI. It is essential to recognize the social significance of these requirements in driving users towards complete Understanding of the technology. Ultimately, the successful implementation of these requirements has led to the formation of a purpose-based trust relationship between the manufacturer and the user.

Conclusion

In this paper, we analyze the contrastive, functional, and transparency understanding of XAI and identify dilemmas in achieving user understanding. We believe that the current XAI dilemma stems from the lack of a framework for a trusting dialog between users and makers.

Thus, this paper proposes an augmentation in the socialization efforts of XAI by introducing a RDF between users and creators. The extant predicament of XAI can be mitigated by instituting RDF the users and the architects of XAI. Such interaction would fuel a comprehensive understanding of XAI, subsequently resolving the prevalent acceptance and recognition issues on both a technical and

social trust spectrum. Furthermore, it will present the creators of XAI with invaluable feedback, guiding them towards refining their systems for easier user acceptance. By undertaking such an initiative, we aim to foster a thorough Understanding of XAI, capable of resolving the current issues of acceptance and recognition.

Received: 19 June 2023; Accepted: 29 January 2024;
Published online: 26 February 2024

References

- Adadi A, Berrada M (2018) Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6:52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Ananny M, Crawford K (2018) Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *N Media Soc* 20(3):973–989. <https://doi.org/10.1177/1461444816676645>
- Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 58:82–115
- Barber A (2003) *Epistemology of language*. Oxford University Press, <http://www.oup.com/uk/catalogue/?ci=9780199250578>
- Barocas S, Selbst AD, Raghavan M (2020) The hidden assumptions behind counterfactual explanations and principal reasons. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 80–89. <https://doi.org/10.1145/3351095.3372830>
- Bastani O, Kim C, Bastani H (2018) Interpretability via Model Extraction. *arXiv*, <http://arxiv.org/abs/1706.09773>
- Biessmann F, Treu V (2021) A Turing Test for Transparency. *arXiv*, <http://arxiv.org/abs/2106.11394>
- Byrne R (2019) Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, p. 6282. <https://doi.org/10.24963/ijcai.2019/876>
- Byrne RM (2007) Precise of the rational imagination: How people create alternatives to reality. *Behav Brain Sci* 30(5–6):439–453
- Davies BL (2007) Grice's cooperative principle: Meaning and rationality. *J Pragmat* 39(12):2308–2331
- de Bruijn H, Warnier M, Janssen M (2022) The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Gov Inf Q* 39(2):101666
- De Villiers J, Hobbs K, Hollebrandse B (2014) Recursive Complements and Propositional Attitudes. In: *Roeper T, Speas M (eds), Recursion: Complexity in Cognition*. Springer International Publishing, vol. 43, pp. 221–242. https://doi.org/10.1007/978-3-319-05086-7_10
- Dellsén F (2020) Beyond Explanation: Understanding as Dependency Modelling. *Br J Philos Sci* 71(4):1261–1286. <https://doi.org/10.1093/bjps/axy058>
- Došilović FK, Brčić M, Hlupić N (2018) Explainable artificial intelligence: A survey. In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 0210–0215. <https://doi.org/10.23919/MIPRO.2018.8400040>
- Evans G (1982) *The Varieties of Reference*. Oxford University Press, Oxford, UK
- Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larriex A (2020) Towards Transparency by Design for Artificial Intelligence. *Soc Eng Ethics* 26(6):3333–3361. <https://doi.org/10.1007/s11948-020-00276-4>
- Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L (2018) Explaining explanations: An overview of interpretability of machine learning. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 80–89. <https://ieeexplore.ieee.org/abstract/document/8631448>
- Greco J (2014) Episteme: Knowledge and understanding. In: *Virtues and Their Vices*, 285–302. Oxford University Press, Oxford, UK
- Grimm S (2021) Understanding. In: *Zalta EN (ed) The Stanford Encyclopedia of Philosophy (Summer 2021)*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/understanding/>
- Grodniewicz JP (2021) The process of linguistic understanding. *Synthese* 198(12):11463–11481. <https://doi.org/10.1007/s11229-020-02807-9>
- Guidotti R (2022) Counterfactual explanations and how to find them: Literature review and benchmarking. *Data Min Knowl Discov*. <https://doi.org/10.1007/s10618-022-00831-6>
- Hauser MD, Chomsky N, Fitch WT (2002) The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298(5598):1569–1579. <https://doi.org/10.1126/science.298.5598.1569>
- Heck RG (1995) The sense of communication. *Mind* 104(413):79–106
- Higginbotham J (1992) Truth and understanding. *Philos Stud* 65(1–2):3–16. <https://doi.org/10.1007/BF00571313>
- Hilton J (1990) Conversational processes and causal explanation. *Psychol Bull* 107(1):65–81
- Jeremy C (2002) *The Liar's Tale: A History of Falsehood*. <http://archive.org/details/jeremy-campbell-the-liars-tale-a-history-of-falsehood>
- Kaplan AD, Kessler TT, Brill JC, Hancock PA (2023) Trust in Artificial Intelligence: Meta-Analytic Findings. *Hum Factors* 65(2):337–359. <https://doi.org/10.1177/00187208211013988>
- Kaplan AD, Kessler TT, Hancock PA (2020) How Trust is Defined and its use in Human-Human and Human-Machine Interaction. *Proc Hum Factors Ergon Soc Annu Meet* 64(1):1150–1154. <https://doi.org/10.1177/1071181320641275>
- Keane MT, Kenny EM (2019) How Case-Based Reasoning Explains Neural Networks: A Theoretical Analysis of XAI Using Post-Hoc Explanation-by-Example from a Survey of ANN-CBR Twin-Systems. In: *Bach K, Marling C (eds) Case-Based Reasoning Research and Development*. Springer International Publishing, pp. 155–171. https://doi.org/10.1007/978-3-030-29249-2_11
- Keil FC (2006) Explanation and Understanding. *Annu Rev Psychol* 57(1):227–254. <https://doi.org/10.1146/annurev.psych.57.102904.190100>
- Kim J (1994) Explanatory knowledge and metaphysical dependence. *Philos Issues* 5:51–69
- Kulesza T, Stumpf S, Burnett M, Yang S, Kwan I, Wong W-K (2013) Too much, too little, or just right? Ways explanations impact end users' mental models. In: *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, pp. 3–10. <https://ieeexplore.ieee.org/abstract/document/6645235/>
- Kuperberg GR, Jaeger TF (2016) What do we mean by prediction in language comprehension? *Lang Cognit Neurosci* 31(1):32–59. <https://doi.org/10.1080/23273798.2015.1102299>
- Lakkaraju H, Kamar E, Caruana R, Leskovec J (2017) Interpretable & Explorable Approximations of Black Box Models. *arXiv*, <http://arxiv.org/abs/1707.01154>
- Lee JD, See KA (2004) Trust in Automation: Designing for Appropriate Reliance. *Hum Factors* 46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Levinson SC (2016) Turn-taking in human communication—origins and implications for language processing. *Trends Cogn Sci* 20(1):6–14
- Lewis D (1973) Counterfactuals and Comparative Possibility. In: *WL Harper, R Stalnaker, & G Pearce (eds) IFS*. Springer Netherlands, pp. 57–85. https://doi.org/10.1007/978-94-009-9117-0_3
- Lim BY, Yang Q, Abdul AM, Wang D (2019) Why these explanations? Selecting intelligibility types for explanation goals. *IUI Workshops*, <http://explainablesystems.comp.nus.edu.sg/2019/wp-content/uploads/2019/02/IUI19WS-ExSS2019-20.pdf>
- Lin Y-S, Lee W-C, Celik ZB (2020) What do you see? Evaluation of explainable artificial intelligence (XAI) interpretability through neural backdoors. *arxiv*, <https://arxiv.org/abs/2009.10639>
- Lipton P (1990) Contrastive explanation. *R Inst Philos Suppl* 27:247–266
- Loi M, Ferrario A, Viganò E (2021) Transparency as design publicity: Explaining and justifying inscrutable algorithms. *Ethics Inf Technol* 23(3):253–263. <https://doi.org/10.1007/s10676-020-09564-w>
- Lombrozo T (2006) The structure and function of explanations. *Trends Cogn Sci* 10(10):464–470
- Lombrozo T (2007) Simplicity and probability in causal explanation. *Cogn Psychol* 55(3):232–257
- Lombrozo T (2010) Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cogn Psychol* 61(4):303–332. <https://doi.org/10.1016/j.cogpsych.2010.05.002>
- Lombrozo T (2016) Explanatory Preferences Shape Learning and Inference. *Trends Cogn Sci* 20(10):748–759. <https://doi.org/10.1016/j.tics.2016.08.001>
- Lombrozo T, Wilkenfeld D (2019) Mechanistic versus Functional Understanding. In: *S. R. Grimm (ed) Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology* (pp. 209–230). Oxford University Press, Oxford, UK. <https://doi.org/10.1093/oso/9780190860974.003.0011>
- Longworth G (2009) Some models of linguistic understanding. *Balt Int Yearb Cognit Log Commun* 5(1):7
- Longworth G (2018) Understanding what was said. *Synthese* 195(2):815–834. <https://doi.org/10.1007/s11229-016-1243-5>
- Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell* 267:1–38
- Miller T (2023) Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven decision support. *arXiv*, <https://doi.org/10.48550/arXiv.2302.12389>
- Miller T, Howe P, Sonenberg L (2017) Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. *arXiv*, <https://doi.org/10.48550/arXiv.1712.00547>
- Millikan RG (2004) *Varieties of meaning: The 2002 Jean Nicod lectures*. MIT press, Cambridge, MA, USA
- Páez A (2019) The pragmatic turn in explainable artificial intelligence (XAI). *Minds Mach* 29(3):441–459. <https://doi.org/10.1007/s11023-019-09502-w>
- Rohlfing KJ, Cimiano P, Scharlau I, Matzner T, Buhl HM, Buschmeier H, Esposito E, Grimminger A, Hammer B, Hüb-Umbach R (2020) Explanation as a social

- practice: Toward a conceptual framework for the social design of AI systems. *IEEE Trans Cogn Dev Syst* 13(3):717–728
- Sacks H, Schegloff EA, Jefferson G (1978) A simplest systematics for the organization of turn taking for conversation. In: *Studies in the organization of conversational interaction*. Elsevier. pp. 7–55, <https://www.sciencedirect.com/science/article/pii/B9780126235500500082>
- Shulner-Tal A, Kuflik T, Kliger D (2022) Fairness, explainability and in-between: Understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics Inf Technol* 24(1):2. <https://doi.org/10.1007/s10676-022-09623-4>
- Sørmo F, Cassens J, Aamodt A (2005) Explanation in Case-Based Reasoning—Perspectives and Goals. *Artif Intell Rev* 24(2):109–143. <https://doi.org/10.1007/s10462-005-4607-7>
- Szafron D, Poulin B, Eisner R, Lu P, Greiner R, Wishart D, Fyshe A, Pearcy B, Macdonell C, Anvik J (2006) Visual explanation of evidence in additive classifiers. In: *Proceedings of Innovative Applications of Artificial Intelligence*, vol. 2. AAAI Press, USA
- Tamagnini P, Krause J, Dasgupta A, Bertini E (2017) Interpreting Black-Box Classifiers Using Instance-Level Visual Explanations. In: *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics*. pp. 1–6, <https://doi.org/10.1145/3077257.3077260>
- Turner VW (1967) *The forest of symbols: Aspects of Ndembu ritual*, vol. 101. Ithaca, NY, Cornell University Press, USA
- Verma, S., Boonsanong, V., Hoang, M., Hines, K. E., Dickerson, J. P., & Shah, C. (2020). Counterfactual explanations and algorithmic recourses for machine learning: A review. *arXiv preprint arXiv:2010.10596*
- Wachter S, Mittelstadt B, Russell C (2017) Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv JL Tech* 31:841
- Zerilli J (2022) Explaining Machine Learning Decisions. *Philos Sci* 89(1):1–19. <https://doi.org/10.1017/psa.2021.13>

Author contributions

AB conceptualized the study, drafted the initial manuscript, and revised the manuscript. YZ conceptualized the study, provided critical feedback, and helped shape the research, analysis, and manuscript writing and revision. All authors have read and agreed to the published version of the manuscript. Correspondence to YZ.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Yi Zeng.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024