ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

Explainability, Interpretability, and Accountability in Explainable AI: A Qualitative Analysis of XAI's Sectoral Usability

Yash Mirchandani[1]

[1]Independent Researcher

Publication Date: 2025/08/25

Abstract: The world that we live in today, is dominated by technological advancements. Many breakthroughs dominate our society today. Among them, Artificial Intelligence (AI) has emerged to be a prominent one. It is no longer relegated to being strictly a sci-fi Hollywood blockbuster project of the future. Today, it is a part and parcel of our daily life and human decision-making processes. It is slowly finding its imprint on several sectors with each passing moment. This, in turn, directly affects human well-being, but also poses us a growing question of trust in these AI systems. With time, this inquiry has grown even more urgent, one which requires an immediate addressal.

Artificial Intelligence serves a lot of purposes but has made substantial contributions in crucial sectors like education, healthcare, and finance. Within these, the incorporation of Artificial Intelligence can have direct consequences on individuals' lives. However, despite holding a life-changing potential, there exists an inadvertent issue of public trust in AI and its related technologies. This is primarily due to the "black-box" nature of many models. This makes their decision-making processes opaque. It also results in them being highly difficult to interpret. [1]

In order to tackle this challenge, Explainable AI (XAI) has emerged as a crucial response. The purpose of XAI is to make algorithmic outcomes more transparent, interpretable, and accountable. In simpler terms, Explainable AI focuses on making the Artificial Intelligence technology more comprehensible for humans. [2] The aim of this paper is to explore the role of Explainable AI in building and sustaining public trust. It will focus specifically on the applications of XAI in fields like education, healthcare, and finance. Via it, the paper seeks to demonstrate how enhancing transparency and accountability through XAI can foster greater trust and responsible adoption of AI in these critical sectors.

To achieve the same, the paper will adopt a qualitative approach. It will be informed by published literature, case examples and policy briefings. This will make it possible to critically consider how explainability affects perceptions of fairness, dependability, and liability. In the education sector, the paper will delve into how transparent grading and admission algorithms can enhance acceptance among students, parents, and educators. In the field of healthcare, it will take a look into the significance of interpretability. This allows for enhanced clinical decision support systems. In turn, it impacts life-altering judgements. These not only require accuracy but also human comprehension. [3] Likewise, explainability in finance can lead to higher credit scoring, fraud detection, and robo-advisory systems. This enables streamlined mechanisms to safeguard consumer trust and compliance with regulatory frameworks. [4]

Lastly, the paper will identify cross-sectoral themes. These include themes related to the balance between accuracy and interpretability, ethical dangers of oversimplified explanations, and the role of cultural and social contexts in trust-building. At the end, the paper will also outline future directions. Moreover, it will also emphasise on the need for standardised frameworks, policy interventions, and greater public engagement in shaping trustworthy AI systems. In discussing XAI in relation to the technology discourse, with a focus on ethics and accountability, this paper will further contextualise its significance for responsible innovation and a sustainable public trust in AI decision-making.

Keyword: *Explainable AI, Trust, Transparency, Ethics, Accountability.*

How to Cite: Yash Mirchandani (2025). Explainability, Interpretability, and Accountability in Explainable AI: A Qualitative Analysis of XAI's Sectoral Usability. *International Journal of Innovative Science and Research Technology,* 10(8), 1118-1131. https://doi.org/10.38124/ijisrt/25aug952

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

I. INTRODUCTION

A. The Black-Box Problem and the Trust Deficit in AI

We are in an era of tech-centric revolutions. Computer engineering and technology continue to change the landscape of our world, with new developments coming out every day and leading us towards a digital revolution that impacts everyone everywhere. Among the plethora of tools and technologies that humans have developed to make their personal and professional lives more comfortable, one such tool is Artificial intelligence. AI today holds boundless possibilities. On one hand, it provides a refined avenue for futuristic exploration. On the other hand, it triggers discussions based on ethics and morals. Many threats have arisen from this development. Despite this, AI's great advantages have enabled it to shift from the edge of technological innovation. It now lies at the heart of contemporary decision-making systems. [5]

Today, algorithms have become the new power-brokers. They are used for determining outcomes in education, healthcare, and finance. For instance, with the help of Automated Grading Systems, the academic performance of any candidate could be determined more accurately. Similarly, diagnostic tools allow for better medical judgments, while credit scoring models, enable individuals to gauge the right financial opportunities. Over time, these systems have promised both efficiency and scalability. However, with it comes a pertinent issue of increasing autonomy. This raises an essential question: can people truly trust decisions made by machines whose reasoning often remains hidden? [6]

One of the significant challenges that lies in the public acceptance of AI is the "black-box" problem. It is when highly sophisticated models, such as deep learning systems, produce outcomes that are accurate yet opaque. [1] If a user says they cannot determine why an AI system came to a particular conclusion, it decreases their trust in that system generally. In a low-stakes environment, it is still possible to debate about the situation, meaning that transparent AI may not be required, but in high-stakes settings (e.g. medical diagnosis, loan approval), you need greater transparency. [7] If it is not available, trust might be on the line as well as other ethical and legal quandaries. This trust issue highlights the pressing need to build systems which can be better interpreted and explained as AI becomes stronger. [1]

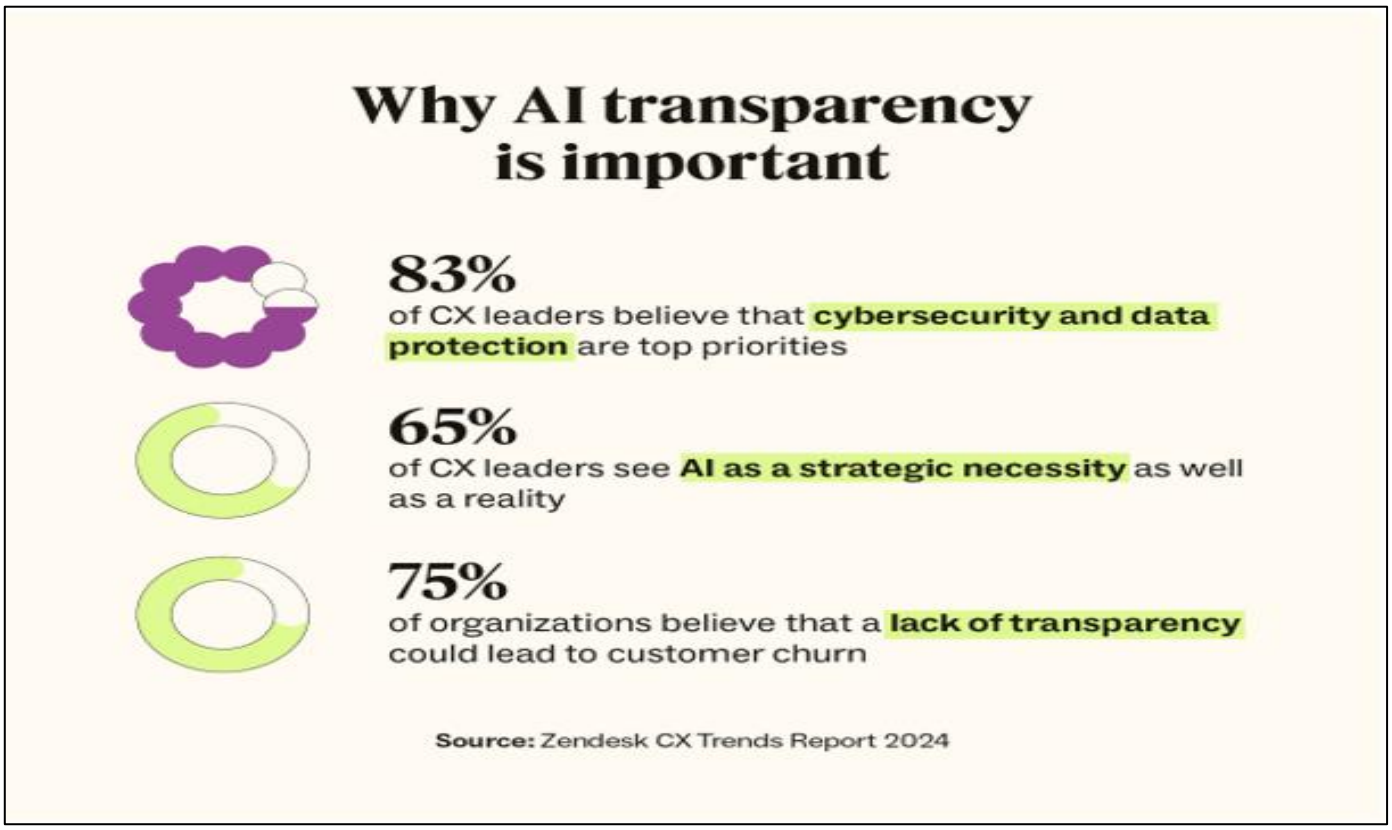

Fig 1 The Need of AI Transparency Across Industries as Suggested by Customer Experience Leaders. Image Taken from Zendesk Blog on 11th August 2025 [8]

Perhaps the most cutting-edge aspect of this technology is that you get a deep introspection about what motivates algorithmic outcomes. This, in turn, allows decision-making processes to be clearer and understandable. Moreover, they become further aligned with human values. Besides the technical aspects, XAI has distinct social dimensions to consider. It is to give people confidence that the systems that they interact with on a daily basis make logical, fair, human-responsive decisions. [2] Thus, XAI contributes in shifting the story away from AI as a shadowy entity implementing automation, to an inclusive instrument when it comes to human decision-making.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

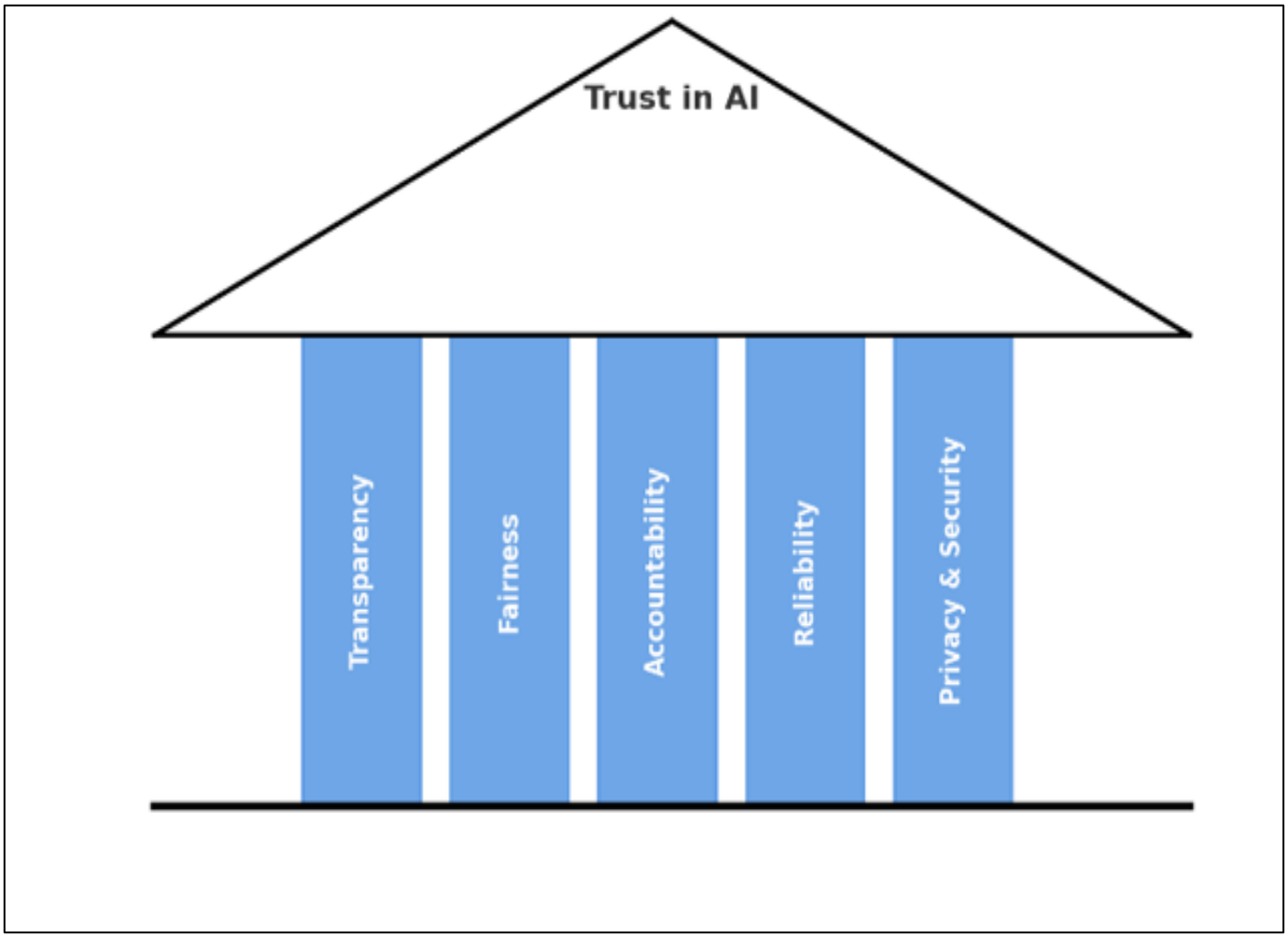

Fig 2 The Key Pillars of Trust in Artificial Intelligence

B. Paper Objectives

The primary objective of this paper is to explore the role of explainable AI in building public trust. To achieve this objective, the paper will undertake a meticulous study of three critical sectors of human development. These include education, healthcare, and finance. The primary basis for the selection of these sectors is that each of them reflects a high-stakes decision-making environment. The outcome of any decision taken within them can profoundly impact human lives. This makes the element of trust indispensable for adoption within these fields.

The paper will deploy a qualitative approach that draws on literature, case studies, and policy insights. It will further examine how explainability shapes trust. Moreover, it will also undertake a detailed study of the challenges that remain and propose future directions that may foster responsible AI integration.

C. Research Questions

The paper poses the following research questions in order to guide further analysis in this study area –

- How does explainability influence public trust in AI across education, healthcare, and finance?
- What challenges arise in different industries when it comes to trust and openness in AI decision-making?
- What ethical and practical problems result when XAI is put into use?
- How can policy, rules, and design principles make AI systems more reliable?

Responding to these questions adds to an emerging conversation about human-centred AI. It further contributes to the importance of disclosing information for advancing technologies ethically.

II. LITERATURE REVIEW & THEORETICAL BACKGROUND

A. Explainable AI (XAI): Definitions, Approaches, and Importance

Explainable Artificial Intelligence (XAI) means a set of techniques and design practices that enable AI systems' operations, transparent and comprehensible to humans. These are completely unlike "black-box" models, which are designed such that, at times, even developers may struggle to interpret outputs. [2] [9] XAI provides clear reasoning behind algorithmic decisions.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

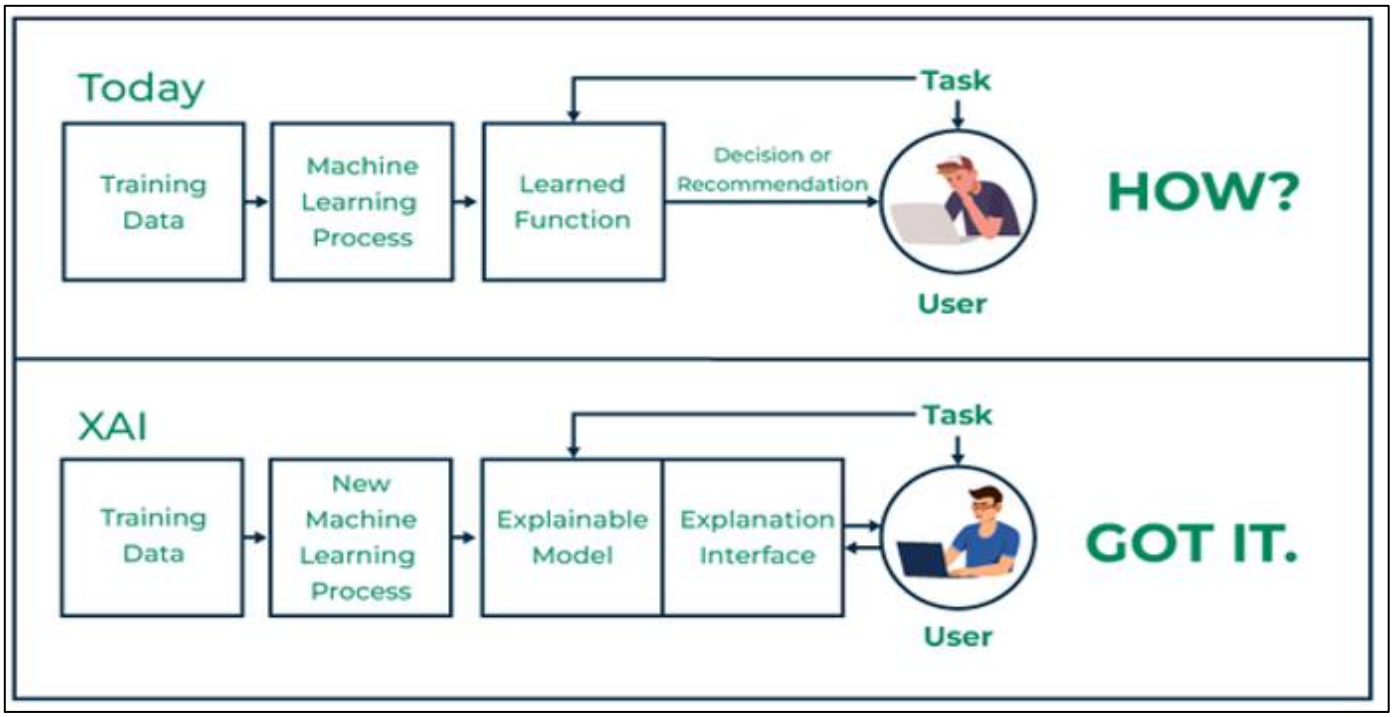

Fig 3 A Diagrammatic Representation of Explainable AI Concept. Image Taken from Geeksforgeeks on 11th August 2025 [10]

One of the clearest distinctions researchers make is between interpretability and explainability. The former relates to how easily humans can understand an AI's function. The latter, however, provides the degree to which an AI can communicate its decision-making process in human-friendly terms. [11]

Approaches to XAI are generally divided into two methods. The first is a model-agnostic method, and the second is a model-specific method. [12] Model-agnostic method, includes approaches such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive Explanations). These generate simplified approximations that shed light on otherwise complex systems. Moreover, model-specific approaches include decision trees or rule-based systems. These are inherently interpretable owing to their transparent structure. [9] [12] [13] While these techniques are used to enhance clarity, they also pose certain challenges. One such being balancing interpretability with accuracy. For instance, deep neural networks, which are high-performing models, are often less transparent. On the contrary, simpler models may sacrifice predictive power for explainability. [14]

The landscape of XAI is not a mere technical understanding. In verticals like healthcare or finance, stakeholders require explanations to certify that the algorithm learned correctly. They also need to ensure the purposes of fairness, accountability and compliance with legal requirements. [3] In this way, explanability becomes a linchpin for responsible AI. It connects technical accuracy to human values and societal norms.

B. Public Trust in Technology: Sociological and Psychological Perspectives

Trust is a core element in how humans relate to technology. It may be viewed as a type of social capital which enables people to become involved in systems that they themselves do not control or fully understand. According to Niklas Luhmann, trust helps to lower complexity by making movement within environments of uncertainty a little easier. [15] In the context of AI, explainability helps reduce perceived complexity. It achieves this by offering transparency. It therefore makes automated systems more approachable and less intimidating. [16]

If we look at the psychological perspectives on trust, we identify that these emphasise cognitive and affective dimensions. Cognitive trust arises from rational evaluation of reliability, competence, and predictability. On the contrary, affective trust is rooted in emotional reassurance and perceived benevolence. Taking these in the context of modern-day AI systems, cognitive trust may stem from technical performance and verifiable explanations. On the other hand, affective trust is majorly an outcome of whether users feel the system is aligned with their interests/values or not. It has been deduced from various studies that individuals are more likely to adopt technology when they believe it is both competent and transparent in its operations. [17] [18]

Trust in AI is derived as much from cultural and contextual aspects of trust as it is shaped by explicit human intentional behaviours. [16] As an example, societies high on digital literacy might show a larger tolerance for automated decision-making, whereas others, with previous deleterious experiences in the form of algorithmic bias, may remain sceptical even if explanations have been provided. This situational aspect of trust moves trust from being a fixed

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

characteristic of technology, and into the realm of dynamics that occur between technology and people. [19]

C. Prior Studies on the Relationship Between XAI and Trust

Various researches/studies have aimed to investigate how explainability influences public trust in AI. Prominent among these is the one conducted by Doshi-Velez and Kim in 2017. This study emphasised that interpretability is not just a technical goal. Rather, it should be seen as a social requirement for accountability. [20] Moreover, studies conducted in the domain of healthcare also reveal similar conclusions. These show that doctors are more willing to rely on AI-assisted diagnostic tools when given transparent reasoning behind predictions. [3] This is a completely different perspective when compared with the recommendations received from black-box. Likewise, in the field of finance, it has been observed time and again that customers are more likely to accept algorithmic credit scoring when models provide understandable criteria for approval or denial. [21]

But the evidence also indicates roadblocks. In fact, explanations might not always lead to higher trust. Some studies show that overly technical or oversimplified explanations can backfire. When explanations are deemed deceptive or veiling inherent bias, users may lose trust.

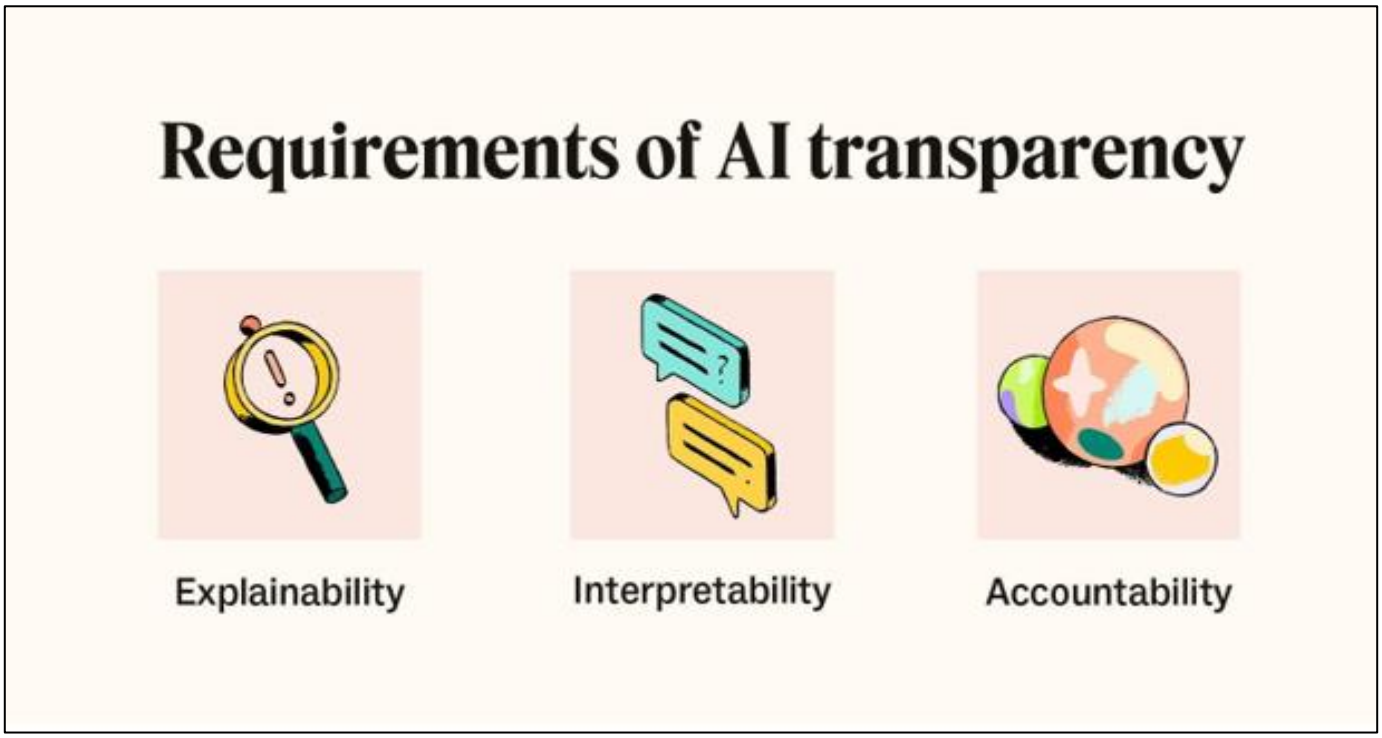

Fig 4 Key Aspects of AI Transparency. Image Taken from Zendesk Blog on 11th August 2025. [8]

Some scholars worry that "explainability" is a tool that can be misused, leveraged selectively to rationalise dubious practices while not actually bringing about more accountability. This tension implies that an explanation does not automatically garner trust, but depends on how well the process elucidates the values at play. [22]

D. Theoretical Lenses: Human-AI Collaboration, Technology Acceptance Models, and Ethics of Transparency

There are several theoretical frameworks that assist us in situating the relationship between XAI and Public Trust.

- *Human-Ai Collaboration:*

As we read through the theories of augmented intelligence, we come across several arguments that state that AI should complement, and not replace, human decision-making. However, trust in such collaborations depends on several factors. The foremost is the ability of humans to understand AI outputs. Secondly, it also depends on how appropriately humans calibrate their reliance on AI outputs. In order to ensure a balanced partnership, explainability lies at the centre of this relationship. [23]

- *Technology Acceptance Models (tam):*

The Technology Acceptance Models, commonly referred to as the TAM frameworks, highlight key determinants of technology adoption. These include the likes of perceived usefulness and perceived ease of use. However,

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

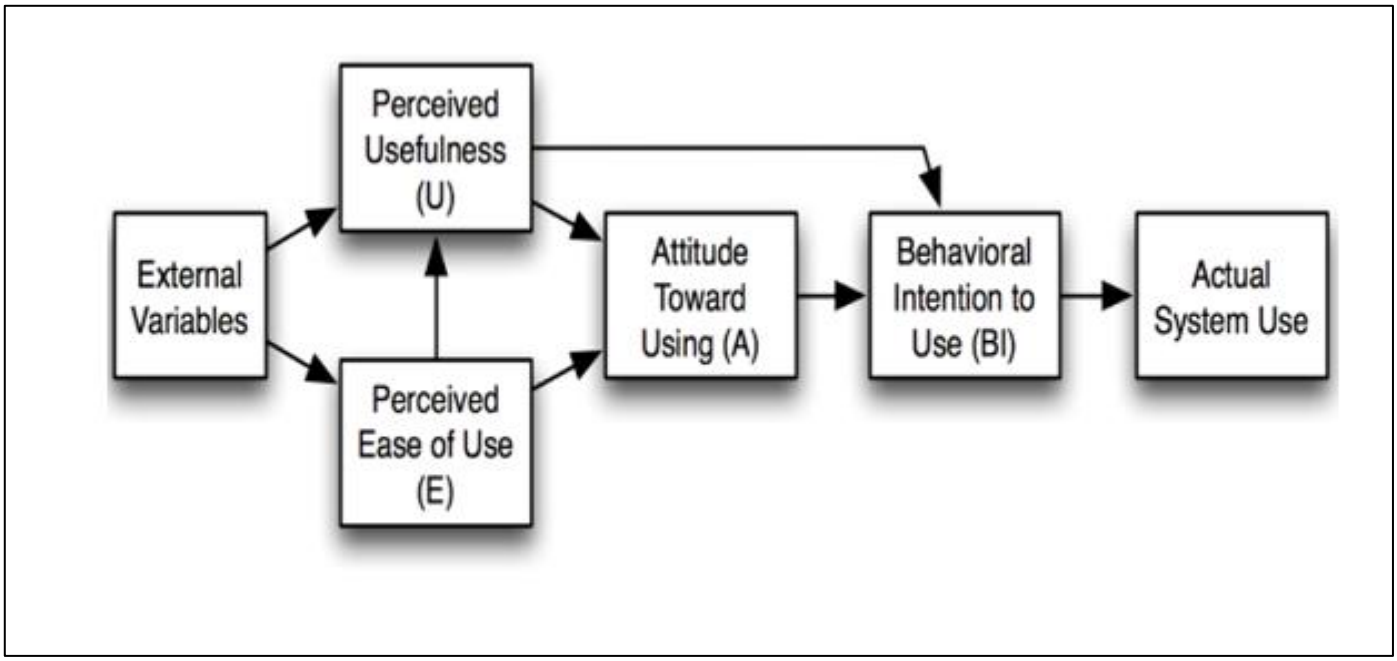

Fig 5 A Pictorial Depiction of the Technology Acceptance Model (TAM). Image Taken From Wikipedia On 11th August 2025 [24]

XAI strengthens both dimensions. On one hand, explanations within the TAM frameworks enhance perceived usefulness by clarifying the rationale behind decisions. On the other hand, TAM frameworks improve ease of use by making systems more accessible to non-experts. [25]

➢ *Ethics of Transparency:*

A key emphasis laid by Ethical Theories is that transparency in any technological system is a cumulative outcome of its commitment to fairness and accountability. Kantian ethics underscores the duty of respecting individuals as rational agents. According to it, these individuals deserve to understand decisions that impact them. However, utilitarian perspectives focus on maximising overall trust and welfare through explainability. These ethical perspectives recalibrate XAI as something more than a technical instrument. Instead, it is a moral necessity in areas where human life is on the line. [26] [27]

E. Summary

The literature points out definitively that explainability is very important for connecting AI's technical capability with its adoption in the society. The trust of the public in AI can only be built on the convergence of technical clarity, psychological assurance, and ethical responsibility. [11] [16] Studies conducted until now have affirmed the positive potential of XAI. However, they also focus on providing a caution that explainability must be meaningful, context-sensitive, and ethically grounded to genuinely build trust. [14] This theory underpins the paper's sectoral analysis across education, healthcare and finance, where trust is both central and subject to dispute.

III. METHODOLOGY

The study will adopt a qualitative research approach. This will be aimed towards the exploration of Explainable AI's (XAI) role in building public trust. The key sectors to consider will be education, healthcare, and finance. For the same purpose, the study would conduct concepts and thematic analysis of literature, case studies, and policy documents. This will provide a more nuanced view of how transparency in AI decision-making affects fairness, accountability, and trust.

The choice to use a qualitative methodology is driven by the type of issues addressed in this study. Trust in AI is not something that can be accounted for as a measurable variable. It is instead a social and moral labyrinth. Its very building blocks depend on human intuitions, cultural frameworks, and institutional preferences. When a qualitative method is deployed, it allows for an in-depth analysis of these dynamics. It highlights how explainability works in theory. Moreover, it also sheds light on how it is received in practice across different sectors. As the study draws on diverse sources, it will capture the interplay between technological design and societal trust. This may, at times, remain unclear in purely quantitative frameworks.

Another important aspect that this study considers is the diversification of literary sources. In order to meticulously analyse the research questions, the paper will take into account scholarly articles, industry reports, and real-world examples. Via it, the research will provide a solid theoretical grounding in concepts such as interpretability, human-AI collaboration, and technology acceptance models. Moreover, by taking a deeper view into the industry reports, the research will be able to draw insights into current trends and policy frameworks that guide AI governance. It will also consider several case studies. Some of the prominent ones include controversies around algorithmic grading in the UK, [28] [29] [30] diagnostic AI in radiology, [31] [32] [33] and biased credit-scoring models. [34] [35] [36] These case studies will offer concrete examples of how explainability, or the lack of it, may impact public trust.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

Towards the end of the research, the methodology will also acknowledge certain limitations. Since the study does not involve primary data collection through focused groups or surveys, it will reflect an inability to directly measure public trust levels or behavioural responses to explainability. Instead, it will, time and again, emphasise on the fact that documented evidence and secondary analysis have been used to deduce the perceived outcomes. This may reflect biases in reporting or limited geographic scope. Moreover, as the modern-day world evolves at a rapid pace, the rate at which AI technologies develop is unprecedented. Owing to it, this piece of literature may quickly become outdated and new research into the field would be then necessary.

Nevertheless, it remains that the research question investigating how explainable AI enhances trust in crucial domains makes the qualitative design particularly apt.

IV. SECTORAL ANALYSIS: TRUST AND EXPLAINABILITY ACROSS CONTEXTS

When we explain explainability in artificial intelligence, we need to keep in mind that it is not just a technical concern. Rather, it is a societal one, which requires immediate attention. Moreover, its urgency becomes clearest when examined across key domains related to human well-being. These domains in question require a higher degree of fairness and transparency, which need to be directly implicated via explainability. [37] This is mirrored in the field of Education, even though healthcare and finance are also ideal contexts to highlight why explainability is critical for social sustainability. Each, at one point or another, due to opaque algorithms, has experienced an implosion in public trust.

A. Education: Meritocracy, Fairness, and the Crisis of Trust

The following section will elaborate upon the role of AI in Education. Ever since the dawn of human civilisation, education systems have carried a heavy responsibility to shape life trajectories. In the age of cut-throat competition, i.e. the 21st century, this responsibility has amplified manifold. As educational institutions make decisions around grading, admissions, and resource allocation, they not only impact students' present but also reverberate through their professional futures.

- *Case Study 01:*

*The 2020 UK Grading Controversy:*One of the most compelling cases of trust and explainability across sectors occurred in the 2020 UK grading controversy. It was during the COVID-19 pandemic, when in-person examinations were impossible to conduct, that Ofqual (Office of Qualifications and Examinations Regulation) introduced an algorithm to predict grades based on the previous years' school performance of the students across the UK.

Fig 6 Protests in Progress Against the Biased Grading Scores Received by Students as Part of Their A-Level Results in the UK in 2020. Image as Taken From Matthew Price's Article on Linkedin.Com on 11th August 2025 [38]

On paper, it seemed like a streamlined system to ensure unbiased and consistent results. However, the actual implementation resulted in widespread downgrading of students from historically underperforming schools. These included often poorer, urban, and ethnically diverse communities. On the contrary, those who belonged to elite institutions saw disproportionately impeccable examination results. Owing to this, there was a swift public outrage. Critics pointed to systemic bias encoded in what was presented as a "neutral" machine-driven process. Within weeks, the government was forced to abandon the algorithm. [28] [29] [30]

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

One of the foremost lessons we can learn from this incident about explainability is the high stakes that technologies like artificial intelligence carry with them. Had Ofqual's model been transparent and interpretable, it would have been open to scrutiny and analysis before being implemented. This would have resulted in its dependence on institutional histories and past exam result patterns being challenged. For one thing, it would have earned the confidence not only of the teachers but of the students and their parents. Each of the individuals involved in the process would have been able to see how individual grades were derived. However, as its purpose was to serve as a temporary grading solution during the pandemic, less focus was laid on deploying explainable AI mechanisms. This turned an on-paper streamlined solution into a legitimacy crisis for the UK state's education policy. [39]

Although, as we analyse this case and look at the broader picture, the lesson here extends beyond the United Kingdom. Today, algorithmic grading, personalised learning platforms, and AI-driven admissions tools have infiltrated global education policies and mechanisms. Trusting them has thus become the foundation on which their mass acceptance rests. As we look at the education sector, explainability is thus not limited to just deriving accurate conclusions. Rather, it is about sustaining belief in education and grading frameworks as a fair system. [40] Without it, students risk perceiving their futures as dictated by inscrutable technologies. This corrodes the very idea of meritocracy.

B. Healthcare: Between Blind Faith and Informed Trust:

When it comes to explainable AI, the stakes in the field of healthcare become naturally high. It is one of the areas where the existential importance of XAI is supremely paramount. The major reason is that within the healthcare domain, the stakes are literally those of life and death. [3] Thus, in this case, trust cannot be demanded. Rather, it needs to be earned. In the current times, we have seen an unprecedented increase in the deployment of AI systems to aid diagnostic imaging, treatment recommendations, and drug discovery. Even though these systems have shown extraordinary promise, they often lag in terms of their clinical integration. The main reason behind this is that their reasoning is opaque.

➢ *Case Study 02: The IBM Watson Oncology Case:*

The above could be clearly understood by discussing the IBM Watson for Oncology case. It is one of the much-publicised cautionary tales in medical field. Watson was marketed as an AI which was capable of providing highly expert cancer treatment recommendations. Owing to this, it attracted partnerships with hospitals worldwide. However, with early use, issues began to crop up in the system. Investigations were put in place, which revealed that the system often generated unsafe or irrelevant suggestions. These were mainly rooted in training data, which was drawn from hypothetical, non-clinical settings. Doctors were unable to interrogate the reasoning behind these outputs. One of the prime examples of the same was Watson's five-year association with MD Anderson Cancer Center, the University of Texas, USA. The partnership that was meant to be lasting ended abruptly after MD Anderson made claims that Watson did not provide correct and safer treatment recommendations. As a result, they quickly lost confidence in the tool. Likewise, similar issues were observed in other hospitals. They scaled back deployments soon after, and Watson Health was ultimately sold off. [31] [32] [33]

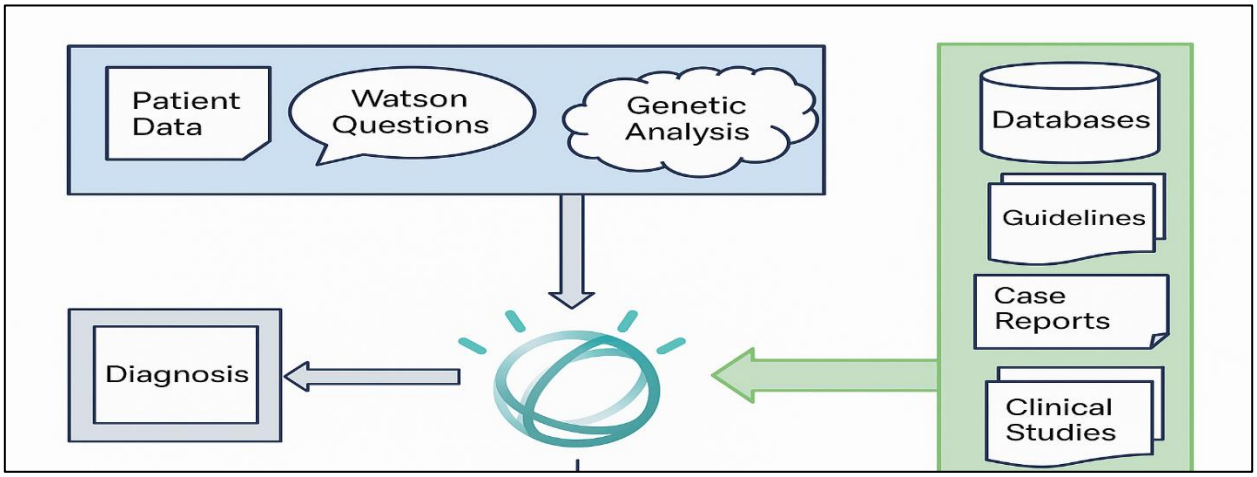

Fig 7 A Pictorial Representation of How The IBM Watson Algorithm Delivered a Medical Diagnosis for Cancer Patients. Image as Taken From the Research Article – A Survey on IBM Watson and Its Services on 11th August 2025 [41]

The story of the collapse of Watson Health reveals a deeper truth. It is that in the field of medicine, patients and doctors will not entrust their lives to a "black-box." [32] In a field that requires immaculate attention to human care, the ethical duty is to ensure that all the decisions undertaken are explainable. [7] This will thus help clinicians to validate them. Moreover, patients can also be assured that they will be able to provide informed consent. [42] It is true that an AI diagnostic model may achieve superhuman accuracy in lab conditions. However, unless it is logically interpretable, i.e. why a tumour is flagged as malignant or why a treatment is recommended, it will struggle to gain legitimacy in real practice.

In the field of medical sciences, explainability also intersects with the cultural dimensions of trust. There have

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

been instances of societies where medical mistrust already runs high. This is solely due to histories of exploitation, structural inequality, or even persistent colonial legacies. In these cases, if opaque AI is implemented, it would only lead to increased suspicions. By contrast, interpretable systems can act as bridges. These would reassure patients that AI supports rather than replaces human judgment. In this sense, explainability is not a mere technical feature but a condition of ethical practice in healthcare. [43]

C. Finance: Trust, Accountability, and Systemic Stability:

Apart from the field of education and healthcare, the role of Explainable Artificial Intelligence in Finance is also indispensable. The field of finance provides yet another lens on the politics of trust. In this domain, opacity in AI-based systems does not impact only individuals. Rather, they hold the power to destabilise entire economies. As tools such as credit scoring, fraud detection, and automated trading systems operate at massive scales, explainability becomes a required mechanism for accountability and compliance. [4]

- *Case Study 03: Apple Credit Card Controversy:*

A striking example of why explainable AI is the need of the hour within the domain of finance comes from the 2019 controversy surrounding Apple's credit card. The card was issued in partnership with leading investment bank Goldman Sachs. The controversy gained fuel when many of its customers, including prominent names like Apple's co-founder Steve Wozniak, reported that women consistently received lower credit limits than men. This was despite sharing finances or even reporting higher income levels. Even though the bank insisted that the algorithm in place that determined the credit limits was unbiased, it refused to disclose the model's logic. This led to increased suspicion of systemic discrimination. It also led to people terming the credit card as 'sexist.' As a result, with increasing public backlash, regulators launched formal investigations in this issue. [34] [35] [36] .

The above case is an apt example of trust and regulatory oversight in the field of finance. This case demonstrates how opacity can erode trust not only between banks and customers, but also between institutions and regulators. [36] What is most prominent in finance, however, is that public confidence is the prime factor that ensures stability. Because of this, explainability in AI systems functions as a protection against reputational and systemic crises. It also has processes for remediation: Customers can challenge a decision; auditors can audit for compliance; and regulators can ensure equity. [21] [4]

Furthermore, aside from the consumer-level applications, interpretability is also important to prevent catastrophic breakdowns in high-frequency trading and risk modelling. [44] The 2008 financial crisis underscored the perils of dodgy financial products that experts struggled to decipher. If AI-based systems are not to make the same mistakes, they must be transparent and interpretable in their decision-making. [4] [16] In this sense, explainability is not only a matter of fairness, but a question of maintaining trust in the most fundamental underpinnings of economic order.

D. Comparative Reflection: The Unifying Thread of Trust:

The fields of education, healthcare, and finance all differ in context and their scope of operations. Yet a common thread that interconnects them is the essence of trust. Moreover, they also showcase the fragility of that trust when it encounters opaque decision-making systems. As we go through the above-mentioned case studies, we witness that each sector illuminates a distinct facet of explainability's role. While in one it marks the need for fairness in education, [40] other requires it for accountability in healthcare systems. [3] [7] Moreover, the third sector involves explainability to ensure compliance frameworks in finance. [21] Despite the different use cases, all converge on a single principle. It is that explainability transforms AI from an alienating black box into a socially embedded partner.

This has led to the emergence of what we know as relational legitimacy. The need of the hour is to devise a grading algorithm that can be questioned, a diagnostic tool that can justify itself, and a credit model that can be audited. All these fosters trust by allowing human actors to see themselves reflected in the system's reasoning. [45] Conversely, opacity turns AI into a force of suspicion. It risks amplifying inequality, undermining professional authority, and destabilising institutions.

Visible in this light, explainability is not ancillary to the future of AI but constitutive of it. Without it, adoption will always be fragile, contested and subject to crises of legitimacy. With it, AI has the potential to excel technically. Moreover, it will achieve the broader social trust on which long-term innovation depends.

V. CROSS-SECTOR THEMES AND ETHICAL CONSIDERATIONS

The sectoral analysis of education, healthcare, and finance reveals multiple things. Firstly, it shows that even though the applications of Explainable AI (XAI) differ in form and context, the core challenges remain the same. These include building and sustaining public trust to converge on common themes. These themes cut across disciplinary boundaries and highlight broader ethical issues. Eventually, these concerns shape how societies perceive, regulate, and ultimately accept AI systems.

A. Common Trust Factors: Fairness, Accountability, and Reliability:

Fairness becomes a common determinant of trust. In education, fairness means fair grading and fair admissions. [40] In health care, it requires that AI diagnostic models must work well in a variety of patient populations and not have systemic biases. [7] In finance, fairness is lending and credit scoring without discrimination by race or gender or socio-economic class. [44] Without demonstrable fairness, the legitimacy of AI systems collapses.

Accountability further underpins trust. Across every industry, stakeholders are eager to understand who is culpable when AI systems make a mistake. [46] Whether it is the designer of the algorithm, the institution using it, or the

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

regulator governing it. Accountability makes sure the weight of error does not rest on the most vulnerable people, who are powerless to question opaque systems. [47]

Lastly, when it comes to reliability, it is defined by whether these technologies can be depended upon on a regular basis. We might have an adaptive learning tool that works in one classroom but fails in another. Or, for instance, there might be a diagnostic AI that underperforms with certain demographics or a fraud detection system that frequently generates false positives. The presence and active deployment of such systems undermine user trust and confidence. It can thus be ascertained that reliability is not merely technical accuracy. Rather, it is contextual robustness. [48]

B. Balancing Accuracy and Explainability:

One of the key issues or challenges that consistently needs to be dealt with across various sectors is the friction that exists between accuracy and explainability. [49] State-of-the-art AI models tend to achieve high performance through complexity. They engage deep neural networks with millions of parameters. Because of their "black box" nature, however, they can be difficult to interpret. [9] Simpler explainable models, on the other hand, may be less accurate in their predictions. In education, it can touch on fairness in admissions; in health care, on clinical decision-making; in finance, on risk assessments. However, the struggle is to achieve balance. The issue lies in giving out explanations that are interpretable enough, yet do not lower the model's performance. [49]

C. Ethical Dilemmas in XAI:

Explainability itself generates new ethical dilemmas. Oversimplification is a key concern. Additionally, as we make AI "understandable," explanations may mislead. [9] This may be an outcome of presenting partial truths or masking deeper complexities. This risks creating a false sense of security. It could also cause stakeholders to be under the presumption that systems are more transparent than they are. [8] Another issue which circulates around the use of AI ethically is privacy.

The deployment of Explainable AI often requires access to training data and decision pathways. However, granting such control to the end user can expose sensitive personal information. [50] For instance, if we look at the domain of healthcare, transparency into model decisions might inadvertently reveal patient histories. Last but not least, the constantly evolving landscape of Explainable AI is also laden with proprietary concerns. This further complicates the field with questions. Proprietary interest in intellectual property may cause private companies to push back on complete explainability. [51] This brings in the clash between what the public demands in terms of accountability, and what corporates seek in terms of competitiveness.

D. Societal Context: India and Global Perspectives:

The societal realm is one of the most important determinants of how trust and ethics will be deliberated in Explainable AI. In a country such as India, one commonly sees that salient challenges around digital literacy, socio-economic disparity and enforcement fall into each other. Owing to such overlaps within the country's social structure, explainability takes on a unique urgency. [52] Moreover, due to the socio-economic discrepancy being widespread in the country, students, patients and consumers struggle to afford challenging the systems that are not transparent. This increases the importance of fairness and transparency. Conversely, resource constraints may motivate quick implementation of AI with little scrutiny, leading to increased risks.

Across the world, notably within the EU and the US, the legislative frameworks focus on accountability and explainability. These include provisions such as the AI Act of the EU or the FDA guidance for AI-related medical devices. The legal infrastructures of these environments are relatively advanced, but innovation and ethics are still a challenge. [53] India's case points to the uneven landscape of global AI governance. As international debates shape aspirational norms, local considerations often play an outsized role in determining what explainability looks like in practice.

E. Towards Ethical and Trustworthy AI:

Finally, the cross-sector analysis highlights that explainability is not merely a technical need. Rather, it is a socio-ethical demand. Trust requires integrity, transparency, accountability, and reliability. It also needs to be balanced by correctness, privacy, and security considerations. [2] With this in mind, as we delve further into the ethical considerations of XAI, we see that neither of the above criteria, alone, makes trust or trustworthiness. [54] Transparency as a provision has to be embedded within broader institutional, cultural or regulatory landscapes. This helps align technology with core human values.

VI. FUTURE DIRECTIONS & POLICY IMPLICATIONS

In the current situation, AI clouds are significantly influencing vital decisions made in the crucial sectors of human progress. These can range from, but are not limited to, education, healthcare, and finance. But if we're imagining the future, the trustworthiness of explainable AI rests on constantly evolving technological work. This is aptly complemented by factors that determine institutional safeguards. As we look at Explainable AI (XAI), we understand that it must evolve from being a research priority. It needs to embrace aspects of practicality, standardisation, and enforceable norms. It has to have a bit of pragmatism, standardisation, and enforced norms. That way transparency, accountability and fairness wouldn't be left to chance. Instead, they're built into the governance from the get go.

A. Standardisation of Explainability Frameworks:

In the context of Explainable AI (XAI), one of the most pressing needs is to develop robust standardised frameworks. These would fine what explainability should look like in practice. As we look at the current approaches, they range from technical visualisations to user-friendly narratives. However, without uniform frameworks and guidelines in

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

place, the explanations delivered to the end user may risk being inconsistent, inaccessible, or at times, misleading. To tackle this, international organisations such as ISO, OECD, and UNESCO have begun to draft AI ethics guidelines. Likewise, the EU's AI Act has made explainability a legal obligation. [53]

When we look from the point of view of India, NITI Aayog has put forth a Responsible AI strategy. This acknowledges the importance of explainability. [55] Despite this, more sector-specific standards are required on an urgent basis. Their need has risen especially in high-stakes domains like healthcare and finance. Thus, it is essential to note that any policies formulated in this direction must ensure a careful balance. They should manage well in terms of promoting innovation. Simultaneously, they should also ensure that AI systems remain interpretable. Additionally, there is a need for such systems to be auditable. Lastly, they should also be aligned with principles of fairness. [4] [2] [37] [40] [43]

B. Human-in-the-Loop (HITL) Systems:

Another hopeful approach is based on HITL systems where algorithmic efficiency is blended with human judgment. These systems help guarantee that AI suggests things. [56] However, they leave the accountability aspect to people at the end of the day. In health, that could be radiologists verifying AI-flagged anomalies; in finance, loan officers inspecting AI-generated credit scores; in education, teachers grounding adaptive learning recommendations. HITL architectures also support public trust by their design. They help in ensuring that AI amplifies human expertise rather than supplanting it. [57]

C. Institutional Roles in Fostering Trust:

One most impactful aspect of ensuring the ethical deployment of AI comes from the proactive role played by governments, regulators, and professional institutions. Taking the example of the three sectors catered in this research paper, we will be able to delve deeper into this aspect better. To begin with, for education, policymakers can make it mandatory to ensure transparency in admissions algorithms. They also need to implement a similar policy for adaptive learning platforms to prevent hidden biases.

Going further, in the field of health, explanations need to be made a central requirement for diagnostic and treatment-support tool by regulators. This is needed before such AI paradigms could earn clinical approval. Equally in the area of finance, regulators such as the Reserve Bank of India (RBI) or the International Monetary Fund (IMF) should introduce more stringent compliance checks. This will allow for credit scoring and fraud detection models to be more auditable and interpretable.

Additionally, as we move beyond the expanse of regulations, institutions also need to invest in capacity building. This can be achieved by training educators, clinicians, and financial professionals in how to critically interpret and apply AI explanations. Without such training, explainability risks remaining a formal checkbox rather than a functional safeguard.

D. Policy Initiatives: Global and Indian Contexts:

On a global scale, the path is toward binding legislation on AI accountability. The EU's AI Act, the U.S. Blueprint for an AI Bill of Rights and UNESCO's AI ethics framework all highlight explainability as a key part of responsible AI. [53] With the diversity in digital literacy and infrastructure in India, there is no one-size-fits-all model that India can copy and paste. Policy should be context-aware. This would ensure that explainable AI is not only for urban elites, but also for rural, multilingual and resource-limited users.

Concrete actions might involve: creating a national AI ethics board and requiring sector-specific explainability standards. It also includes incentivising the development of intelligible models. Lastly, infusing explainability requirements into adherence standards for public-sector AI deployments, can also be a way forward.

E. Towards A Culture of Transparent AI:

Ultimately, the future of XAI is not just technical. It also holds regulatory challenges to deal with. Alongside this, one also needs to look into the cultural aspects that define their implementation. In order to embed explainable AI into regular governing systems, there rests a dire necessity to align technological design with ethical values.

Additionally, it also needs to match public expectations. To achieve this, there needs to be a multi-stakeholder collaboration. This should involve the government, academic institutions, industry, and civil society. This will enable communities to cultivate an ecosystem of transparency. It will also ensure that accountability forms the centrepiece and trust becomes a crucial factor that determines outcomes.

This will allow for Explainable AI to be treated not just as a luxury add-on. Rather, it should serve as a foundational principle of sustainable innovation. Without it, adoption will remain fragile and contested. With it, AI has the potential to evolve into a socially embedded partner in advancing human progress.

VII. CONCLUSION

From the given study, it can thus be concluded that explainability is not a peripheral feature of artificial intelligence. Rather, it is an ethical and functional cornerstone that defines the way humans interact with mechanisms governing organisational functioning. It has also been observed over the course of this study that trust in AI systems cannot be established when interpretability is absent. One needs to keep in mind that transparency is the precondition for legitimacy.

As we analysed the education, healthcare, and finance sectors, we were able to identify a consistent pattern. Each of these was governed by opaque systems, at some point in time or another. Even though their operational processes were highly sophisticated in nature, they failed to inspire confidence and risk amplifying inequities. On the contrary, as explainable systems were kept in place, they accounted for higher degree of fairness and informed decision-making.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

In education, the transparency of algorithmic grading models reveals how unexplainable outputs are unfair. This ultimately undermines institutional credibility. The critical challenge of health illustrates this even more clearly. Clinical dependence on AI hinges on understandability. However, explainability guarantees the protection of patients. Simultaneously, it also ensures professional commitment. Finance exemplifies the systemic effects of opacity, as interpretability is connected with consumer confidence and regulatory certainty in areas that have clear implications for economic stability. In all three fields, we will see that the evidence consistently points to one conclusion. It is that explainability in AI is what makes AI align with human values.

It is not just those sectors that will be affected. AI, if it is to become a socially accepted and an ethically responsible technology, demands an approach to development that emphasizes transparency, interpretability, and responsibility. Explainability is not just a technical protection. Rather, it is a moral requirement. The direction of AI's future will be decided not only by what it permits us to do. However, it will also be determined by how clearly and prudently we understand and react to those possibilities.

REFERENCES

[1]. W. J. Von Eschenbach, "Transparency and the black box problem: Why we do not trust AI.," Philosophy & Technology, vol. 34, no. 4, p. 1607–1622, 01 September 2021.

[2]. R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan and R. Ranjan, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," ACM Computing Surveys (ACM Comput. Surv.), vol. 55, no. 9, p. 33, September 2023.

[3]. T. Hulsen, "Explainable artificial intelligence (XAI): concepts and challenges in healthcare," AI, vol. 4, no. 3, pp. 652 - 666, 2023.

[4]. A. N. A. O. S. T. N. K. A. J. A. I. Anang, "Explainable AI in financial technologies: Balancing innovation with regulatory compliance," International Journal of Science and Research Archive, vol. 13, p. 1793–1806, 30 October 2024.

[5]. W. Ertel, Introduction to Artificial Intelligenc, Springer Nature, 2024.

[6]. T. Mucci, "The future of AI: trends shaping the next 10 years," IBM, 11 October 2024. [Online]. Available: https://www.ibm.com/think/insights/artificial-intelligence-future. [Accessed 07 August 2025].

[7]. R. Hassan, N. Nguyen, S. R. Finserås, L. Adde, I. Strümke and R. Støen, "Unlocking the black box: Enhancing human-AI collaboration in high-stakes healthcare scenarios through explainable AI," Technological Forecasting and Social Change, vol. 219, 2025.

[8]. C. Marshall, "What is AI transparency? A comprehensive guide," Zendesk Blog, 7 August 2025. [Online]. Available: https://www.zendesk.com/in/blog/ai-transparency/. [Accessed 11 August 2025].

[9]. V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud and A. Hussain, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," Cognitive Computation, vol. 16, pp. 45-74, 24 August 2023.

[10]. geeksforgeeks, "Explainable Artificial Intelligence(XAI)," geeksforgeeks, 15 April 2025. [Online]. Available: https://www.geeksforgeeks.org/artificial-intelligence/explainable-artificial-intelligencexai/. [Accessed 11 August 2025].

[11]. S. U. Hamida, M. J. M. Chowdhury, N. R. Chakraborty, K. Biswas and S. K. Sami, "Exploring the Landscape of Explainable Artificial Intelligence (XAI): A Systematic Review of Techniques and Applications," Big Data and Cognitive Computing, vol. 8, no. 11, p. 149, 31 October 2024.

[12]. K. Devireddy, A Comparative Study of Explainable AI Methods: Model-Agnostic vs. Model-Specific Approaches, vol. 1, Cornell University (arXiv), 2025.

[13]. A. Athar, "SHAP (SHapley Additive exPlanations) And LIME (Local Interpretable Model-agnostic Explanations) for model explainability.," Analytics Vidhya, 04 October 2020. [Online]. Available: https://medium.com/analytics-vidhya/shap-shapley-additive-explanations-and-lime-local-interpretable-model-agnostic-explanations-8c0aa33e91f. [Accessed 08 August 2025].

[14]. D. E. Mathew, D. U. Ebem, A. C. Ikegwu, P. E. Ukeoma and N. F. Dibiaezue, "Recent Emerging Techniques in Explainable Artificial Intelligence to Enhance the Interpretable and Understanding of AI Models for Human," Neural Processing Letters, vol. 57, no. 16, 07 February 2025.

[15]. N. Luhmann, Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität [Trust: A mechanism for the reduction of social complexity], Stuttgart: Enke, 1973.

[16]. R. Lukyanenko, W. Maass and V. Storey, "Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities," Electronic Markets, vol. 32, p. 1993–2020, 28 November 2022.

[17]. B. K. Riley and A. Dixon, "Emotional and cognitive trust in artificial intelligence: A framework for identifying research opportunities," Current Opinion in Psychology, vol. 58, August 2024.

[18]. M. Kim, R. Huang and S. J. Lennon, "Understanding the role of cognitive and affective trust in consumer-artificial intelligence relationships," in International Textile and Apparel Association Annual Conference Proceedings, 2022.

[19]. J. Schoeffer, Y. Machowski and N. Kuehl, "Perceptions of Fairness and Trustworthiness Based on Explanations in Human vs. Automated Decision-Making," in Hawaii International Conference on System Sciences 2022 (HICSS-55), Hawaii, 13 September 2021.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

[20]. F. Doshi-Velez and B. Kim, A roadmap for a rigorous science of interpretability, vol. 2.1, arXiv preprint, 2017.
[21]. M. A. FAHEEM, "AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction," Iconic Research and Engineering Journals, vol. 5, no. 3, pp. 177-186, September 2021.
[22]. D. Martens, G. Shmueli, T. Evgeniou, K. Bauer, C. Janiesch, S. Feuerriegel, S. Gabel, S. Goethals, T. Greene, N. Klein, M. Kraus, N. Kühl, C. Perlich, W. Verbeke and A. Zharova, "Beware of 'Explanations' of AI," arXiv.org, April 2025.
[23]. J. Li, Y. Yang, R. Zhang and Y.-C. Lee, "Overconfident and unconfident AI hinder human-AI collaboration," arXivLabs, 12 February 2024.
[24]. "Technology acceptance model," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Technology_acceptance_model. [Accessed 11 August 2025].
[25]. I. Baroni, G. R. Calegari, D. Scandolari and I. Celino, "AI-TAM: a model to investigate user acceptance and collaborative intention in human-in-the-loop AI applications," Human Computation, vol. 9, no. 1, pp. 1-21, 23 May 2022.
[26]. P. Hayes, "An ethical intuitionist account of transparency of algorithms and its gradations," Business Research, vol. 13, no. 3, p. 849–874, 23 December 2020.
[27]. C. Mougan and J. Brand, "Kantian deontology meets AI alignment: Towards morally grounded fairness metrics," 9 November 2023.
[28]. B. Quinn, "UK exams debacle: how did this year's results end up in chaos?," 17 August 2020. [Online]. Available: https://www.theguardian.com/education/2020/aug/17/uk-exams-debacle-how-did-results-end-up-chaos. [Accessed 10 August 2025].
[29]. G. Leckie and L. Prior, "The 2020 GCSE and A-level 'exam grades fiasco': A secondary data analysis of students' grades and Ofqual's algorithm," The University of Bristol, 2023. [Online]. Available: https://www.bristol.ac.uk/cmm/research/grade/. [Accessed 10 August 2025].
[30]. S. Shead, "How a computer algorithm caused a grading crisis in British schools," CNBC.com, 21 August 2020. [Online]. Available: https://www.cnbc.com/2020/08/21/computer-algorithm-caused-a-grading-crisis-in-british-schools.html. [Accessed 10 August 2025].
[31]. henricodolfing, "Case Study 20: The $4 Billion AI Failure of IBM Watson for Oncology," henricodolfing.com, 07 December 2024. [Online]. Available: https://www.henricodolfing.com/2024/12/case-study-ibm-watson-for-oncology-failure.html. [Accessed 10 August 2025].
[32]. H. Faheem and S. Dutta, "Artificial Intelligence Failure at IBM 'Watson for Oncology'," IBS Center for Management Research, 2022.
[33]. L. O'Leary, "How IBM's Watson Went From the Future of Health Care to Sold Off for Parts," Slate.com, 31 January 2022. [Online]. Available: https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html. [Accessed 10 August 2025].
[34]. BBC, "Apple's 'sexist' credit card investigated by US regulator," BBC.com, 11 November 2019. [Online]. Available: https://www.bbc.com/news/business-50365609. [Accessed 10 August 2025].
[35]. W. Knight, "The Apple Card Didn't 'See' Gender—and That's the Problem," Wired.com, 19 November 2019. [Online]. Available: https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/. [Accessed 10 August 2025].
[36]. N. Vigdor, "Apple Card Investigated After Gender Discrimination Complaints," The New York Times, 10 November 2019. [Online]. Available: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html. [Accessed 10 August 2025].
[37]. U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl and J. D. Weisz, "Expanding Explainability: Towards Social Transparency in AI systems," in CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 07 May 2021.
[38]. M. Price, "The 2020 UK exam fiasco has given 'algorithms' a bad name," LinkedIn.com, 13 November 2020. [Online]. Available: https://www.linkedin.com/pulse/2020-uk-exam-fiasco-has-given-algorithms-bad-name-matthew-price/. [Accessed 11 August 2025].
[39]. C. S. Elliot Jones, "Can algorithms ever make the grade?," Ada Lovelace Institute, 18 August 2020. [Online]. Available: https://www.adalovelaceinstitute.org/blog/can-algorithms-ever-make-the-grade/. [Accessed 10 August 2025].
[40]. H. Khosravi, S. Buckingham Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq and D. Gašević, "Explainable Artificial Intelligence in education," Computers and Education: Artificial Intelligence, vol. 3, 13 May 2022.
[41]. A. Kumar, P. Tejaswini, O. Nayak, A. Kujur, R. Gupta, A. Rajanand and M. Sahu, "A Survey on IBM Watson and Its Services," Journal of Physics: Conference Series, vol. 2273, 1 May 2022.
[42]. H. J. Park, "Patient perspectives on informed consent for medical AI: A web-based experiment," Digital Health, vol. 10, 30 April 2024.
[43]. J. Amann, A. Blasimme, E. Vayena, D. Frey and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," BMC Medical Informatics and Decision Making, vol. 20, 30 November 2020.
[44]. A. Kirilenko, A. Kyle, M. Samadi and T. Tuzun, "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," SSRN Electronic Journal, 26 May 2011.
[45]. K. de Fine Licht and J. Licht, "Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy," AI & SOCIETY, vol. 35, December 2020.

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25aug952

[46]. Emerge Digital, "AI Accountability: Who's Responsible When AI Goes Wrong?," Emerge Digital, [Online]. Available: https://emerge.digital/resources/ai-accountability-whos-responsible-when-ai-goes-wrong/. [Accessed 11 August 2025].
[47]. World Health Ogranization, "ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH: WHO GUIDANCE," Geneva, 2021.
[48]. S. T. H. Mortaji and M. E. Sadeghi, "Assessing the Reliability of Artificial Intelligence Systems: Challenges, Metrics, and Future Directions," International Journal of Innovation in Management, Economics and Social Sciences, vol. 4, pp. 1-13, 29 June 2024.
[49]. K. d. Costa, "Practical and Societal Dimensions of Explainable AI," Holistic AI, 2 March 2023. [Online]. Available: https://www.holisticai.com/blog/explainable-ai-dimensions. [Accessed 11 August 2025].
[50]. A. Gomstyn and A. Jonker, "Exploring privacy issues in the age of AI," IBM, 30 September 2024. [Online]. Available: https://www.ibm.com/think/insights/ai-privacy. [Accessed 11 August 2025].
[51]. Intertech, "Risks to Proprietary Data During AI Implementation and How To Protect Your Data in an AI System," Intertech, [Online]. Available: https://www.intertech.com/risks-to-proprietary-data-during-ai-implementation-and-how-to-protect-your-data/. [Accessed 11 August 2025].
[52]. D. D. Vashistha, P. P. K. Chandel and S. Gaur, "Investigating Socioeconomic Disparities in Digital Education Experiences," The International Journal of Indian Psychology , vol. 12, no. 3, July - September 2024.
[53]. Y. Liu, W. Yu and T. Dillon, "Regulatory responses and approval status of artificial intelligence medical devices with a focus on China," NPJ Digital Medicine, vol. 7, no. 1, p. 255, 18 September 2024.
[54]. L. Nannini, M. Marchiori Manerba and I. Beretta, "Mapping the landscape of ethical considerations in explainable AI research," Ethics and Information Technology, vol. 26, p. 44, 25 June 2024.
[55]. Niti Aayog, "NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE," Niti Aayog, June 2018.
[56]. P. Kelly-Voicu, "What is Human-in-the-loop (HITL) in AI-assisted decision-making?," June 2023. [Online]. Available: http://1000minds.com/articles/human-in-the-loop. [Accessed 11 August 2025].
[57]. D. Lukose, "Right Human-in-the-Loop for Effective AI," Medium.com, 13 January 2025. [Online]. Available: https://medium.com/@dickson.lukose/building-a-smarter-safer-future-why-the-right-human-in-the-loop-is-critical-for-effective-ai-b2e9c6a3386f. [Accessed 12 August 2025].