



**Trustworthy and Responsible AI:
The Integral Foundations of an
AI-Enabled Future**

CONTENTS



01

Introduction



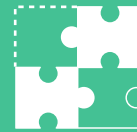
02

AI in
Healthcare



03

Bridging The Ethics
Gap Surrounding AI



04

How Quantiphi
can help :

- Security and Privacy
- Governance and Accountability
- Fairness
- Human-centric
- Transparency and Explainability
- Safety, Robustness and Reliability
- Socially Beneficial
- Scientific Rigor



05

Conclusion



06

References

Introduction

The anticipated benefits of artificial intelligence (AI) in terms of scientific progress, human wellbeing, economic value, and exploring solutions to solve major social and environmental problems seem remarkable. However, such a powerful technology comes with certain concerns like its ability to make important decisions that are free from bias, to align those decisions with an ethical framework that considers relevant human values, and to explain the rationale for the reasoning and decision-making in reaching a conclusion. Since many successful AI techniques rely on large volumes of data, it is important to know how data is handled and managed by AI systems.

The aforementioned concerns by AI users, early adopters, and policymakers are among the obstacles that potentially limit or restrict AI's widespread adoption and use. The use of black-box solutions, the introduction of bias in the formulation of data sets that are used to test AI algorithms, and the lack of disclosure, accountability, and responsibility that lead to incorrect recommendations or decisions are detrimental to establishing trust in the use of AI.

According to a new study by IBM's Institute for Business Value, 82% of all enterprises and 93% of high-performing enterprises are now considering or moving ahead with AI adoption. Reasons for this move include the technology's potential to drive revenues, improve customer service, lower costs, and manage risk. Although businesses realize the benefits of this technology, 60% of those companies fear liability issues and 63% indicated they lack the skills to harness AI's potential.

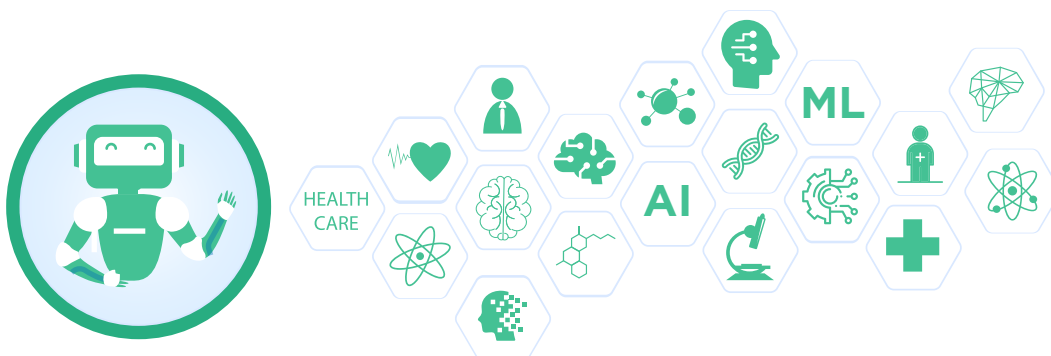


AI in Healthcare

Artificial Intelligence (AI) and machine learning (ML) are powerful technologies that can augment or replace human tasks. While AI/ML methods promise great performance gains in various application domains, the solutions they provide are not readily comprehensible to humans. While a black-box approach may be acceptable in some domains, the acceptance bar in medical applications, which require transparency, appears to be higher. For instance, in medicine, clinicians must understand the decisions of a trained machine learning model (ML), and ultimately validate and accept its recommendations before using the technology. This understanding is needed so the provider can explain the rationale behind a recommendation to the patient. AI can help healthcare providers enhance patient care, render accurate diagnoses, optimize treatment plans, support pandemic preparedness and response, inform policymakers' decisions, and distribute resources within health systems, among other things. To realize this potential, healthcare professionals and systems need a strong understanding of the context in which such systems can operate safely and efficiently, the conditions that must be met to assure dependable, appropriate use, and the methods for continual system auditing and assessment, which are all needed to instill trust in the system.

Trust in AI can be influenced by several human factors such as user education, past experiences, user biases, and perception towards automation. Additionally, properties of an AI system including controllability, transparency, complexity of the model, and associated risks influence our ability to trust AI. It is also important to note that maximizing the user's trust does not necessarily yield the best decisions from a human-AI collaboration. As trust is maximized, the user's acceptance of the AI-generated recommendations increases.

In the healthcare domain, AI may be integrated with other technologies, such as robots, to help with diagnoses and/or to perform surgery. However, many doctors are still skeptical about technologies such as AI assistants, and most patients prefer to trust a human doctor more than a robotic doctor [1]. Further, many aspects of machine learning, such as deep learning, appear to be black-boxes and the lack of explainability and transparency does little to promote trust, which is the cornerstone of effective doctor-patient relationships.



Bridging The Ethics Gap Surrounding AI

As more companies worldwide adopt AI, healthcare leaders grapple with ethical decisions about its design and use. While they seek direction on how to use cutting-edge technologies like AI, they also face ethical issues in its actual development.

Consumers interact with businesses on a daily basis through acts such as perusing web pages, online banking, and calling customer care. These transactions appear to be free, but they are not. The currency being exchanged is consumer data. Customers should be able to trust that the data they contribute to companies and the AI algorithms they utilize will be used responsibly and without bias.



¹ ["AI Can Outperform Doctors, so Why Don't Patients Trust It?"](#) Harvard Business Review, October 30, 2019

In today's world, consumers and investors are more aware that bias can show up in numerous ways, including in technology. Promoting anti-bias measures can set a business apart by establishing greater customer confidence and trust in AI applications. Companies can help secure consumer data and design algorithms that lower the likelihood of bias contamination while simultaneously creating brand equity and customer confidence by concentrating on AI bias, and emphasizing AI ethics. On the flip side, when word gets out that a business has implemented unethical algorithms, particularly in cases that promote bias, research shows that the business is penalized.

The core of the trustworthy AI recommendations can be split into three parts including lawful, ethical, and robust. To implement these three core parts, best practices and recommendations laid out by the WHO and European Union should be incorporated into the AI system for it to be considered trustworthy. Based on these recommendations from the regulatory bodies and our experience in building AI/ML solutions for healthcare customers, Quantiphi has created a framework to ensure that the AI applications we build are trustworthy and responsible.



Security and Privacy

AI systems must be developed and deployed in secure and conducive environments, both for data collection and storage. Additionally, developers must follow best practices while dealing with the security and privacy of data used by the AI systems. For instance, AI developers must collect only the required data with the user's consent and provide information regarding data-sharing with other parties along with an opt-out option.

There are stringent regulations on sharing healthcare data such as medical history of patients, current conditions, and patient reports. Federated Learning helps in this case by allowing each of the medical centers to have a local version of AI systems that interact with their securely stored local data. The individual model updates from each of the centers communicate with a global model, which incorporates the input model updates and returns new model updates. In this way, each medical center can have a better performing local model or ML system, without their data leaving the secure local servers.



Governance and Accountability

AI systems need to abide by the local and federal/national/territory rules and regulations, and need to be answerable and accountable to the specific governing councils for all consequences— both intentional and otherwise. Organizations need to have teams in place with well-defined roles and pointed responsibilities for any actions and outcomes. Detailed technical audits of AI and data policies also help bring out the best practices of building responsible AI systems.



² "Ukanwa, Kalinda. "Algorithmic Bias Isn't Just Unfair - It's Bad for Business - The Boston Globe." BostonGlobe.com. The Boston Globe, May 23, 2021

Fairness

Bias in AI-systems is a phenomenon where the outcomes of the system favor a specific set of attributes in the data. Fairness represents lack of visible or latent bias. Visible biases represent biases that can easily be witnessed and measured, and most often pertain to specific features present in the dataset. For instance, a classification model may be biased to a specific column in the dataset like 'Age' or 'Location' or 'race/ethnicity'. Latent biases are much harder to notice, and are difficult to quantify and measure, such as, a computer-vision model biased towards texture of the background, while detecting objects in the foreground, or a face detection model biased towards gender of the person or the color of one's skin. These are latent since there are no explicit features being provided to the model while training. One of the prominent ways to avoid bias is to maintain diversity at every step of data handling, aiming to cover multiple features. It is important to maintain similar diversity during data annotation, aiming to include people from all backgrounds, age-groups, and ethnicities.



Human-centric

AI systems need to be human-centric, socially beneficial, and safe. Understanding the needs of the people is imperative to strategize AI engagements and design use cases. Human-centered designs recommend that the business use cases should cater to and address these human needs. While businesses use automation to boost productivity, it is important to ensure that the most impactful decisions are taken by people using human-in-the-loop design principles, thus making the implementation process human-centric. The AI systems that are developed should have a net positive impact on society.



Transparency and Explainability

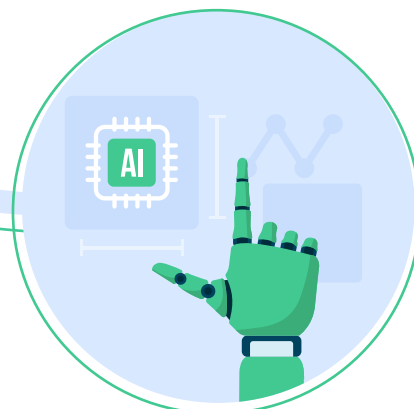
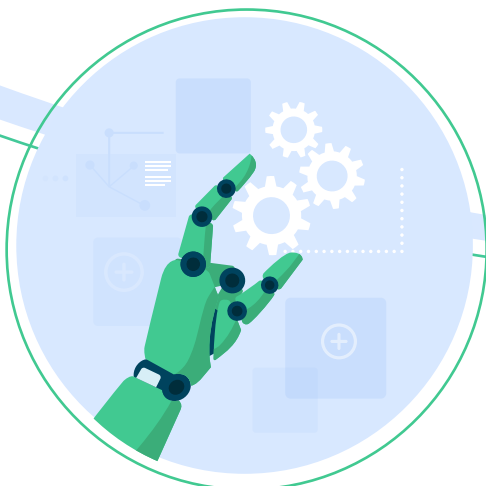
Transparency and explainability advocate maintaining complete transparency of data handling as well as model techniques. Transparency in modeling techniques includes the availability of details like training data used, proportions of data used, and validation strategies regarding the data processing. All the algorithmic choices and the reasoning behind their use should be well documented and supported.

Explainability strives to achieve an AI system that is self-explanatory. Best practices strive to add mechanisms that help the users in understanding the algorithms and their working principles. Transparency and explainability play a crucial role in improving the trust factor in AI, which leads to an increased adoption of AI by organizations and business units.



Safety, Robustness and Reliability

The AI systems' performance, speed, scalability and safety should be robust. Modern AI systems need to be evaluated for multiple metrics, and should aim at achieving high performance for those metrics. Rigorous testing for scalability and safety needs to be performed in complex simulations that depict real-life environments.



Socially Beneficial

AI and ML should be used to build solutions and tools that deliver a net positive impact on society by addressing the most pressing challenges. AI systems can be used to safeguard businesses and communities against potential threats.

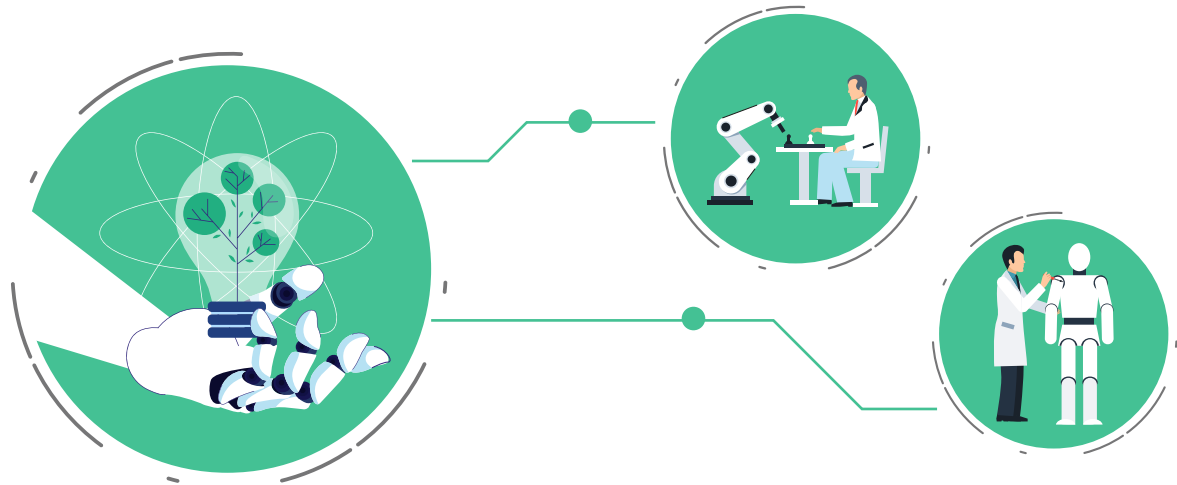


Scientific Rigor

AI systems should follow high standards of science-based approaches with critical evaluation of design, methods, results and conclusions. Organizations should collaborate with the scientific community to advance state-of-the-art techniques and push the boundaries further, and to validate the truthfulness and generalization of the AI systems being built at-scale. Also, interesting studies and insights regarding the AI-systems should be published for the collective advancement of science and the community.

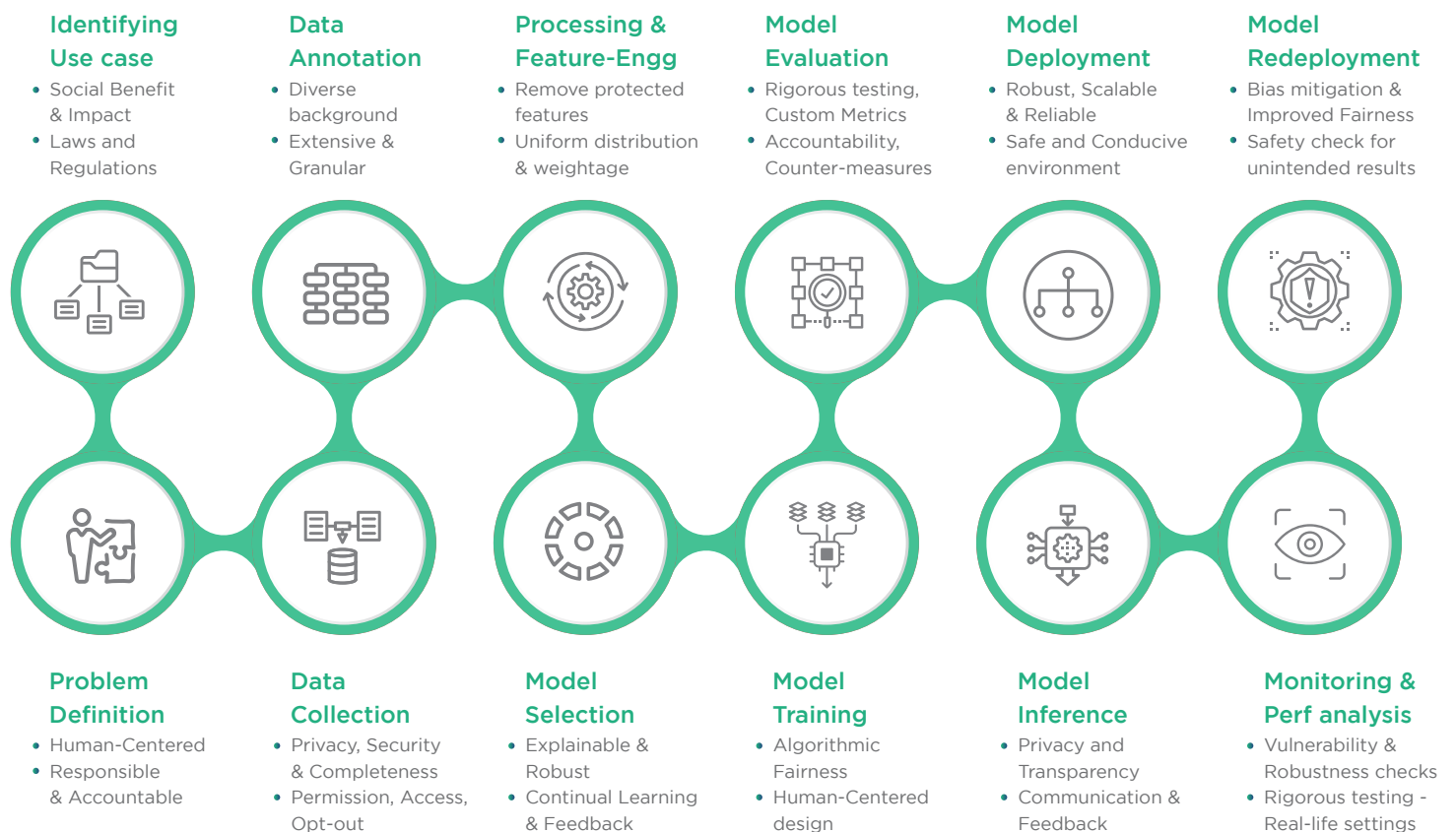


The Quantiphi Difference



While most ethical AI is in the theoretical stage, Quantiphi, a leader in AI, has leveraged its expertise and experience to create a responsible AI framework where we integrate the Responsible and Ethical AI principles, transform them from theoretical principles to actionable tasks in our end to end AI lifecycle.

Integrate Responsible AI into ML Lifecycle



Quantiphi redefines your organization's approach by ensuring that AI systems built for customers and partners are ethically compliant all the time. Being an AI-first digital engineering company driven by the desire to solve transformation problems at the heart of business, Quantiphi provides responsible AI solutions that will curate the future. Quantiphi can help train specific teams in the organizations in these areas to sensitize them on:



AI-Principles

Understanding the ethical principles, dos and don'ts and setting the boundaries of usage of AI systems.



Governing Processes

Potential risks and pitfalls of lack of Responsible-AI, as well as benefits and requirements of Responsible-AI. Guidelines for review of processes and best-practices being followed.



Technical Knowledge

Gaining theoretical and practical knowledge of available tools and technologies to implement Responsible-AI at scale, at various stages of AI-lifecycle.

Ethical ML practices make AI-systems more trustable by:



Accelerating Innovation

Trustworthy AI enables organizations to re-image products with a digital first mindset. Quantiphi allows this by offering domain adapted speech transcription, visual safety and surveillance, and medical imaging solutions. Using innovative techniques in responsible AI helps detect bias and alleviate opacity.



Enhancing Efficiency

Ethical AI practices improve operational efficiency and remove redundancies. Quantiphi helps healthcare organizations to achieve this by offering federated learning solutions, promoting safety and ensuring transparent AI/ML systems.

The Quantiphi Difference

AI is a powerful technology that has already demonstrated improvements in our everyday lives. However, to fully understand its potential benefits, manufacturers and developers must be committed to building a system of trust. Issues of bias, explainability, data handling, data selection, transparency on data policies, and design choices must be addressed in a responsible and open way. By infusing AI products with robust and proven bias detection, mitigation capabilities, and the ability to explain how their decisions are made, AI developers can bridge the trust gap and create an effective path for economic growth and societal benefit.

It's best for healthcare organizations to embrace AI early and leverage it to deliver better care and services to the customers. In the future, customers, companies, and regulators will target unethical AI and train employees on ethical best practices and conduct robust testing and monitoring throughout the lifecycle. Companies looking to build trust with both employees and consumers need to ensure that their AI systems are responsible and ethical.

Such an approach is adopted by Quantiphi to ensure issues are identified, discussed, disclosed, and resolved in a cooperative environment. It is this kind of interdisciplinary and cooperative approach that produces the best solutions and creates a comprehensive environment for trustworthy AI.



Build a better future for healthcare

Let's chat

Get in touch with our experts. Write to us at appliedai@quantiphi.com

Featured Resources

Interested in integrating electronic health systems, improving your mission-critical workflows, and delivering services more efficiently.? Check out the resources below.

Transforming Healthcare with AI-assisted Computer Medical Imaging

Revolutionize the Future of Healthcare with Intelligent Virtual Assistants

Drive Efficient Drug Discovery with AI-powered Virus Mutation Prediction
Leveraging Sequencing Data

Total Experience Automation: The Future Of Healthcare Services

Explore solutions and demos

Learn more about our suite of Healthcare and Life Science solutions

Quantiphi is an award-winning AI-first digital engineering company driven by the desire to reimagine and realize transformational opportunities at the heart of business. We are passionate about our customers and obsessed with problem-solving to make products smarter, customer experiences frictionless, processes autonomous and businesses safer.

Amsterdam | Boston | Bangalore | London | Mumbai | Princeton | San Jose | Toronto | Trivandrum

 **quantiphi**
Solving What Matters

Follow us on:



appliedai@quantiphi.com