Using the Digital Trust Ecosystem Framework to Achieve Trustworthy AI

Governance

CONTENTS

ABSTRACT

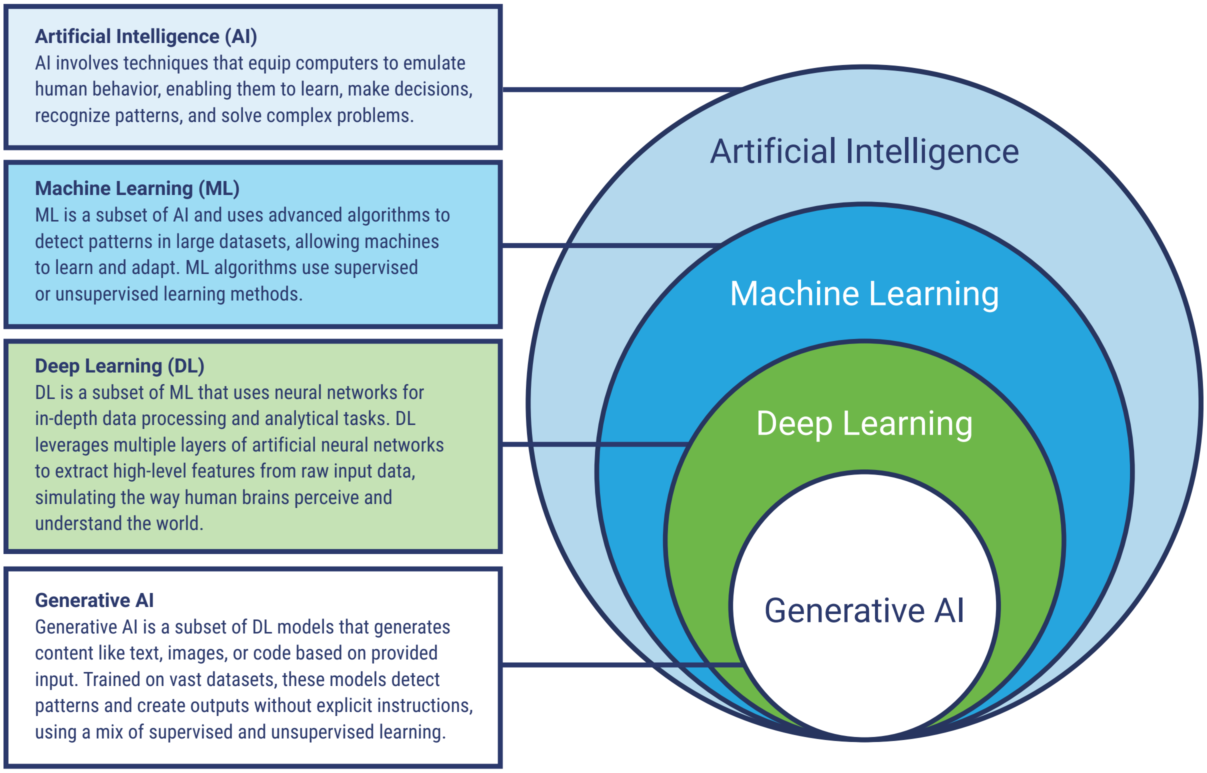
This white paper explores the benefits of using ISACA's *Digital Trust Ecosystem Framework* (DTEF) for enterprises adopting artificial intelligence (AI)-enabled technologies and services. It provides an understanding of how the DTEF supports the evaluation of emerging technology risk and provides guidance on building the governance structure to benefit organizations throughout an AI life cycle. The white paper highlights key components of digital trust considered foundational to successful AI technology integration and service delivery using example use case scenarios organizations typically face.

Introduction

Instances of artificial intelligence (AI) are ubiquitous in society; some examples include chatbots, financial fraud detection, and navigation software. The term AI encapsulates machine learning (ML), deep learning, and generative AI (**figure 1**); and, collectively, it continues to revolutionize all industries, offering efficiency and benefits with speed and scale. For example, within healthcare, AI enables personalized treatment plans and manages predictive diagnostics; in financial services, AI enhances fraud detection and risk management with data-driven insights; while within the energy sector, AI optimizes grid management and predictive maintenance, unleashing real-world benefits of efficiency and sustainability.

The widespread influence of AI extends beyond explicit enterprise adoption, as evidenced by its integration into various third-party applications such as popular office productivity software and everyday tasks. Further, employees in certain departments (e.g., HR, marketing) are likely already using web-based software to screen job applicants or write marketing content. In short, employees are likely already using AI even if they are unaware of it. Ultimately, the growing catalog of generative AI products not explicitly permitted for business use represents an evolution of shadow IT.

FIGURE 1: A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI



Artificial Intelligence (AI)
AI involves techniques that equip computers to emulate human behavior, enabling them to learn, make decisions, recognize patterns, and solve complex problems.

Machine Learning (ML)
ML is a subset of AI and uses advanced algorithms to detect patterns in large datasets, allowing machines to learn and adapt. ML algorithms use supervised or unsupervised learning methods.

Deep Learning (DL)
DL is a subset of ML that uses neural networks for in-depth data processing and analytical tasks. DL leverages multiple layers of artificial neural networks to extract high-level features from raw input data, simulating the way human brains perceive and understand the world.

Generative AI
Generative AI is a subset of DL models that generates content like text, images, or code based on provided input. Trained on vast datasets, these models detect patterns and create outputs without explicit instructions, using a mix of supervised and unsupervised learning.

Source: Unraveling AI Complexity - A Comparative View of AI, Machine Learning, Deep Learning, and Generative AI, https://commons.wikimedia.org/wiki/File:Unraveling_AI_Complexity_-_A_Comparative_View_of_AI,_Machine_Learning,_Deep_Learning,_and_Generative_AI.jpg. This figure is available under the *Creative Commons Attribution-ShareAlike 4.0 International* license

Regardless of AI's business enhancements and optimization, it can in many cases increase risk. As with any technology, it was not long before bad actors began leveraging AI for nefarious purposes. To date, generative AI has improved the credibility of emails associated with business email compromise[1] (BEC) while also decreasing the technical capabilities necessary to implement BEC attacks.[2] Additionally, AI has been weaponized for geopolitical reporting,[3] image-based abuse,[4] and political campaigns[5] and is credited with a successful multimillion-dollar scam.[6] In total, not only does AI mean different things to different people, but the associated risk for each type of AI is also highly variable.

AI's pervasiveness cannot be overstated. According to Deloitte, the world is experiencing the first-of- its-kind, most readily available technological revolution of epic proportion in generative AI.[7] Despite AI being first conceptualized in 1950,[8] the magnitude and velocity of advancements and associated impacts are far from being realized. Similar to past transformative changes in technology, for many, any mention of AI invokes fear, skepticism, and distrust. However, it is noteworthy that generative AI is envisioned to unlock, not replace, human potential, so the immediate and perhaps enduring challenge entails the distribution and management of tasks that machines and humans do well. The use of AI necessitates cross-functional leadership and careful consideration of business and societal implications—in addition to the traditional elements of people, processes, and technology. Enter ISACA's Digital Trust Ecosystem Framework (DTEF).[9]

AI's pervasiveness cannot be overstated. According to Deloitte, the world is experiencing the first-of- its-kind, most readily available technological revolution of epic proportion in generative AI.

The DTEF supports the establishment and maintenance of digital trust from all organizational stakeholder perspectives. Digital trust is broader than technology; it applies to the entire organization and all its external stakeholders.

Selecting, establishing, and maintaining digital relationships requires confidence and transparency from all parties involved. The needs, principles, values, and objectives of providers and consumers influence the levels of trust necessary. The DTEF provides an innovative digital trust and transformation approach for enterprises to prioritize digital trust—a concept associated with any mention of AI.

This white paper explores how organizations can achieve the digital trust worthiness of AI-enabled technologies and service solutions using ISACA's DTEF.

[1] ISACA, "ISACA Glossary," https://www.isaca.org/resources/glossary

[2] Kelley, D.; "Wormgpt - the Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks," SlashNext, 13 July 2023, https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/; Shiebler, D.; "Generative AI Enables Threat Actors to Create More (and More Sophisticated) Email Attacks," AbnormalSecurity, 14 June 2023, https://abnormalsecurity.com/blog/generative-ai-chatgpt-enables-threat-actors-more-attacks

[3] Klepper, D.; "Fake Babies, Real Horror: Deepfakes From the Gaza War Increase Fears About AI's Power to Mislead," APNews, 28 November 2023, https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47

[4] Saner, E.; "Inside the Taylor Swift Deepfake Scandal: 'It's Men Telling a Powerful Woman to get Back in her Box'," The Guardian, 31 January 2024, https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box

[5] Hickey, M.; "Vallas Campaign Condemns Deepfake Video Posted to Twitter," CBS News, 27 February 2023, https://www.cbsnews.com/chicago/news/vallas-campaign-deepfake-video/; Harper, A.; Gehlen, B.; et al.; "AI Use in Political Campaigns Raising Red Flags into 2024 Election," ABC News, 8 November 2023, https://abcnews.go.com/Politics/ai-political-campaigns-raising-red-flags-2024-election/story?id=102480464; Ulmer, A.; Tong, A.; "Deepfaking it: America's 2024 Election Collides with AI Boom," Reuters, 30 May 2023, https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/

[6] Edwards, B.; "Deepfake Scammer Walks off With $25 Million in First-of-its-kind AI Heist," ARS Technica, 5 February 2024, https://arstechnica.com/information-technology/2024/02/deepfake-scammer-walks-off-with-25-million-in-first-of-its-kind-ai-heist/

[7] Deloitte, "Generative AI and the Future of Work," https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-ai-institute-generative-ai-and-the-future-of-work.pdf

[8] Turing, A.M.; *Computing Machinery and Intelligence*, Mind 49, 1950, https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf

[9] ISACA, *Digital Trust Ecosystem Framework*, USA, 2022, www.isaca.org/dtef-ebook

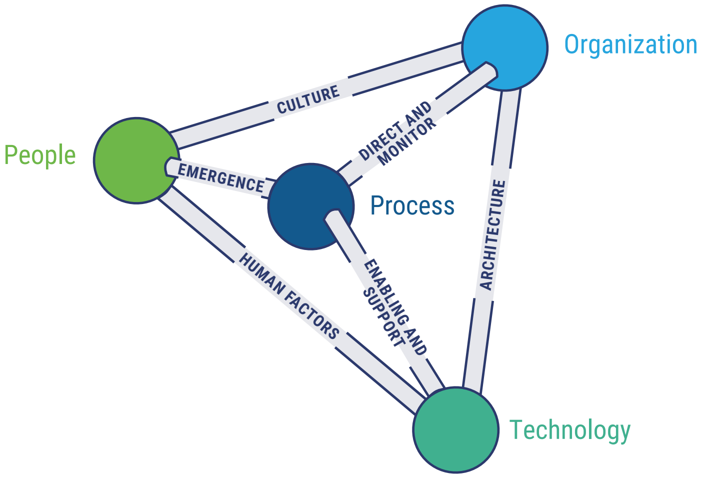
DTEF Overview

The DTEF defines the core ingredients to create a digitally trusted ecosystem that considers all stakeholders to ensure that all digital interactions and transactions are legitimate, trusted and consider the elements of integrity, security, privacy, resilience, quality, reliability, and confidence.[10] Trust is an important principle that must be incorporated into the AI adoption and deployment strategy of any enterprise. For example, the ability to trust AI outputs will depend on factors such as the level of data quality and data protection. The AI model must be developed based on trusted code and transparency to include operation decision making. The end user or consumer of AI-enabled services trusts the AI algorithms driving decisions are proven to be accurate, protects the privacy of the end user, is secure, and is free from bias. Finally, proper AI governance will help senior management and key stakeholders trust AI decisions are explainable, fairness and ethical practices are followed, and regulatory and legal compliance can be demonstrated.

The DTEF creates a body of knowledge and helps to address a dynamic changing legal, regulatory, and technology landscape; modern day business requirements; external/internal influence and emergence; and the risk and controls required to create and operate in a digitally trusted environment.

The DTEF is a three-dimensional model based on the theory that there are multiple dependencies between four primary elementary nodes: people, process, technology, and organization. Nodes interact with each other through multiple dynamic interactions, and in the model, these nodes are the highest level. The four nodes are interconnected by six domains—Culture, Emergence, Enabling and Support, Human Factors, Direct and Monitor, and Architecture. **Figure 2** illustrates the relationships between nodes and domains.

FIGURE 2: DTEF Model



Source: ISACA, *Digital Trust Ecosystem Framework (DTEF)*, 2022

[10] *Ibid.*

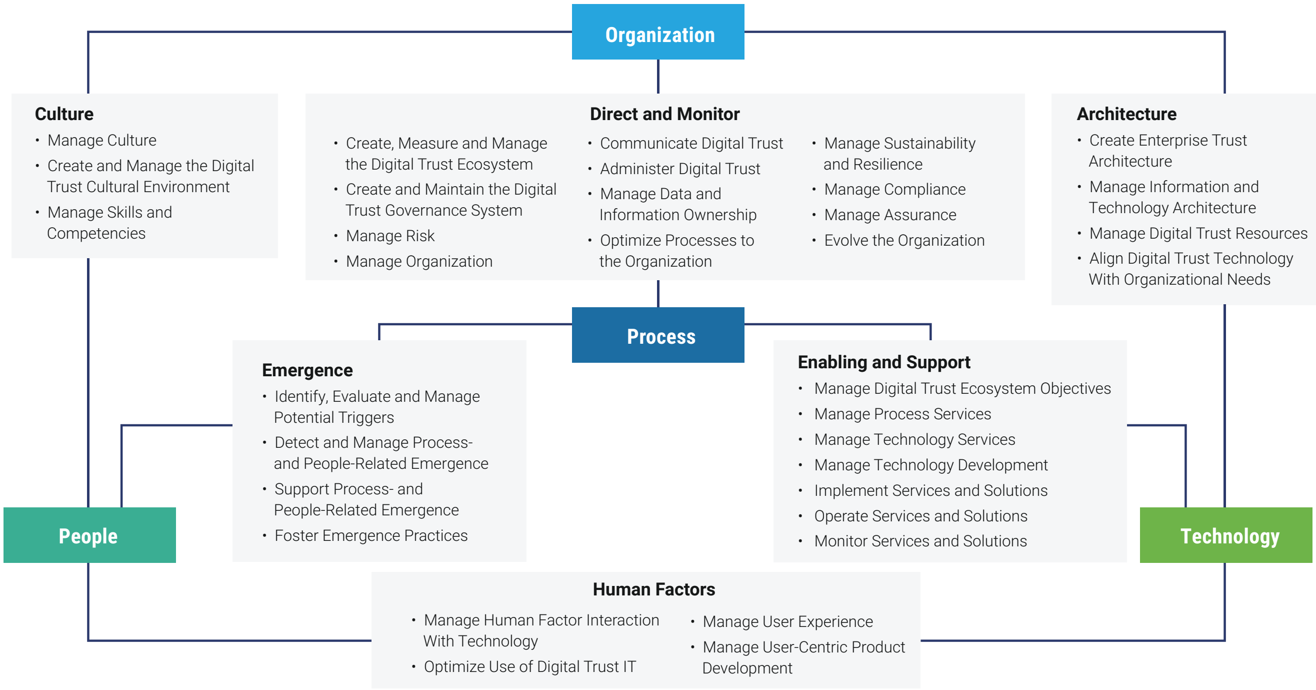
Domains influence one or more nodes. For example, a change in the Architecture domain will inevitably influence the Organization node and/or the Technology node. The domains also interact with each other in a systematic sense. The domains play a vital role in managing the interconnections and complexities that exist in an organization, as they work in tandem with changing regulations, emerging technologies, new threats, procedural changes, etc. Domains are comprised of a set of constituting and structuring elements.

The DTEF uses trust factors to establish a foundation of content within each domain. For example, the Architecture domain is categorized into the following four trust factors:

1. Create Enterprise Trust Architecture
2. Manage Information and Technology Architecture
3. Manage Digital Trust Resources
4. Align Digital Trust Technology with Organizational Needs

Trust factors describe the overarching actions required to maintain digital trust and help to avoid or reduce bias. The components of the framework (refer to **figure 3**) can be used to ensure digital trust principles are demonstrated with any application of AI.

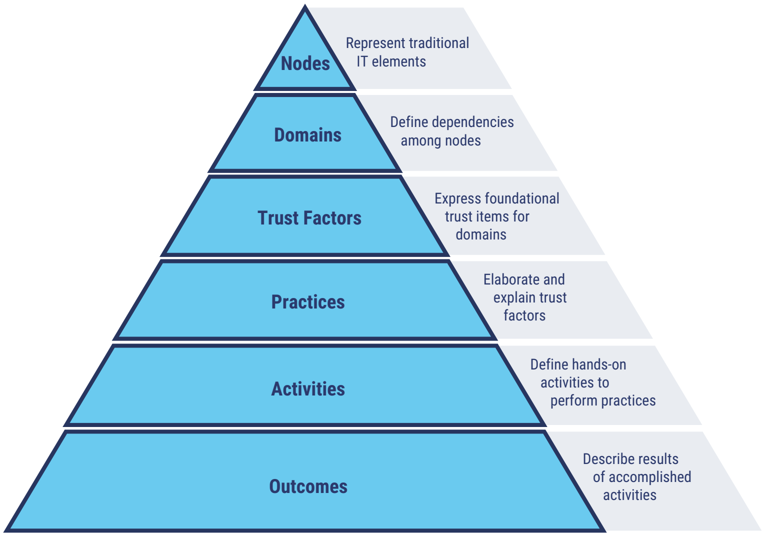
FIGURE 3: DTEF Components


Organization
Culture
• Manage Culture
• Create and Manage the Digital Trust Cultural Environment
• Manage Skills and Competencies
Direct and Monitor
• Create, Measure and Manage the Digital Trust Ecosystem
• Create and Maintain the Digital Trust Governance System
• Manage Risk
• Manage Organization
• Communicate Digital Trust
• Administer Digital Trust
• Manage Data and Information Ownership
• Optimize Processes to the Organization
• Manage Sustainability and Resilience
• Manage Compliance
• Manage Assurance
• Evolve the Organization
Architecture
• Create Enterprise Trust Architecture
• Manage Information and Technology Architecture
• Manage Digital Trust Resources
• Align Digital Trust Technology With Organizational Needs
Process
Emergence
• Identify, Evaluate and Manage Potential Triggers
• Detect and Manage Process- and People-Related Emergence
• Support Process- and People-Related Emergence
• Foster Emergence Practices
Enabling and Support
• Manage Digital Trust Ecosystem Objectives
• Manage Process Services
• Manage Technology Services
• Manage Technology Development
• Implement Services and Solutions
• Operate Services and Solutions
• Monitor Services and Solutions
People
Technology
Human Factors
• Manage Human Factor Interaction With Technology
• Optimize Use of Digital Trust IT
• Manage User Experience
• Manage User-Centric Product Development

Source: ISACA, *Digital Trust Ecosystem Framework*, USA, 2022

The DTEF provides a structure to support digital trust throughout an entire ecosystem, considering how relationships influence the level of engagement with consumers, customers, and users. **Figure 4** illustrates key structural elements of the DTEF.

FIGURE 4: DTEF Hierarchy


Nodes
Represent traditional IT elements
Domains
Define dependencies among nodes
Trust Factors
Express foundational trust items for domains
Practices
Elaborate and explain trust factors
Activities
Define hands-on activities to perform practices
Outcomes
Describe results of accomplished activities

Digital trust[11] is not just about digital information and technology. Digital trust impacts the entirety of businesses; therefore, enterprises that can demonstrate digital trustworthiness gain considerable competitive advantage and build better relationships with consumers.[12]

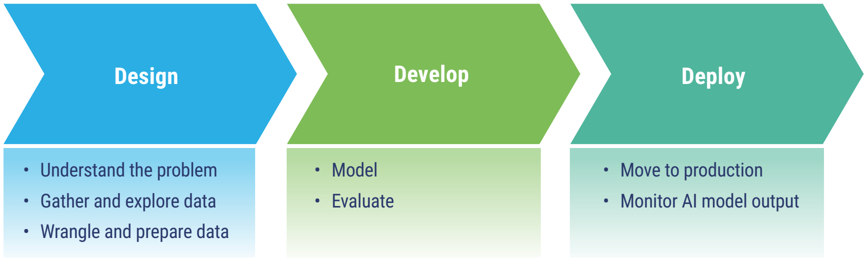
AI Life Cycle

Uncontrolled use of AI products can introduce considerable risk. Due to its potential to adversely impact the entire business (e.g., intellectual property [IP] loss, brand damage, litigation), enterprises must understand not only what may already be in use, but also what business problems employees are attempting to solve by using AI. In many ways, AI is just another technology, but its nuances are what necessitates deliberate governance and risk management within all aspects of its life cycle. An AI life cycle is an iterative process of moving from a business problem to an AI-based solution to solve that problem.[13]

Each of the steps depicted in **figure 5** continuously iterates throughout design, development, and deployment phases. Readers may be familiar with other AI life cycles, which may include additional detail. Regardless of their differences, AI life cycles all generally address four components: business requirements and goals, data collection and preparation, model development and evaluation, and operational deployment and monitoring.

[11] ISACA defines digital trust as the confidence in the integrity of the relationships, interactions, and transactions among providers and consumers within an associated digital ecosystem. This includes the ability of people, organizations, processes, information, and technology to create and maintain a trustworthy digital world.
[12] Tower-Pierce, J.; "Wake up America: Digital Trust can Positively Impact Revenue," InCyber, 10 July 2023, https://incyber.org/en/article/wake-up-america-digital-trust-can-positively-impact-revenue/
[13] IT Modernization Centers of Excellence, "AI Guide for Government: A Living and Evolving Guide to the Application of Artificial Intelligence for the U.S. Federal Government," 26 March 2024, https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html

FIGURE 5: AI Life Cycle



Design

Understanding the Problem

AI process and system owners and other relevant stakeholders are responsible for identifying the key project objectives and requirements to effectively define the desired business outcome. The organization must clearly define what business problem they want AI to solve. No AI solution will succeed without clearly and precisely understanding the business challenge being addressed and the desired outcome.

Data Gathering and Exploration

Data is the foundation for any AI solution. In this step, data is gathered and evaluated for its appropriateness for use in the proposed AI application. It requires discovering available data sets, identifying data quality problems, and deriving initial insights into the data and perspectives on a data plan. The AI model should only use the data with a clear understanding of the data required and the makeup of that data.

Data Wrangling and Preparation

This step comprises all activities used to construct the working data set from the initial raw data into a format that the AI model can use. This step can be time consuming and tedious, but it is critically important to develop an AI model that achieves the goal of solving the problem(s) identified in the first step.

Develop

Modeling

This step focuses on experimenting with data to determine the right AI model. Often during this phase, the team trains, tests, evaluates, and retrains many different AI models to determine the best AI model and settings to achieve the desired outcome. The AI model training and selection process is interactive. No AI model achieves the best performance the first time it is trained. It is only through iterative fine-tuning that the model is honed to produce the desired outcome. Depending on the amount and type of data being used, this training process may be very computationally expensive; it may require special equipment to provide enough computing power, for it cannot be performed on a typical laptop.

Evaluation

Once one or more AI models have been built that appear to perform well based on relevant evaluation metrics, the AI models are tested on new data to ensure they generalize well and meet the business goals.

Deploy

Move to Production

Once an AI model has been developed to meet the expected outcome and performs at a level determined ready for use on live data, it is deployed into a production environment. In this case, the AI model will take in new data that was not a part of the training cycle.

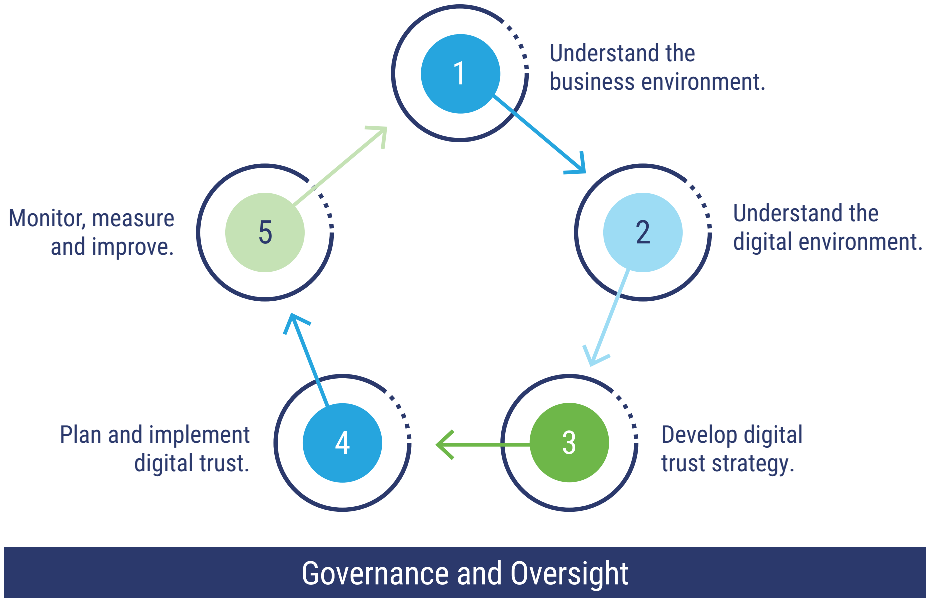
Monitor AI Model Output

Once deployed, the AI production output must be monitored to ensure that it is adequately able to produce the intended outcome—a process known as generalization, or the AI model's ability to adapt properly to new, previously unseen data. In production, AI models can "drift," meaning that the performance will change over time. Careful monitoring of drift is important and may require the continuous updating of the AI model. AI systems must undergo rigorous and continuous monitoring and maintenance to ensure they continue to perform as trained, meet the desired outcome, and solve the business challenges.

AI-focused DTEF Implementation

The DTEF implementation model can be very helpful when thinking about AI projects holistically. There are five phases of the DTEF implementation model with a sixth phase, Governance and Oversight, that underpins the other five phases (refer to **figure 6**).

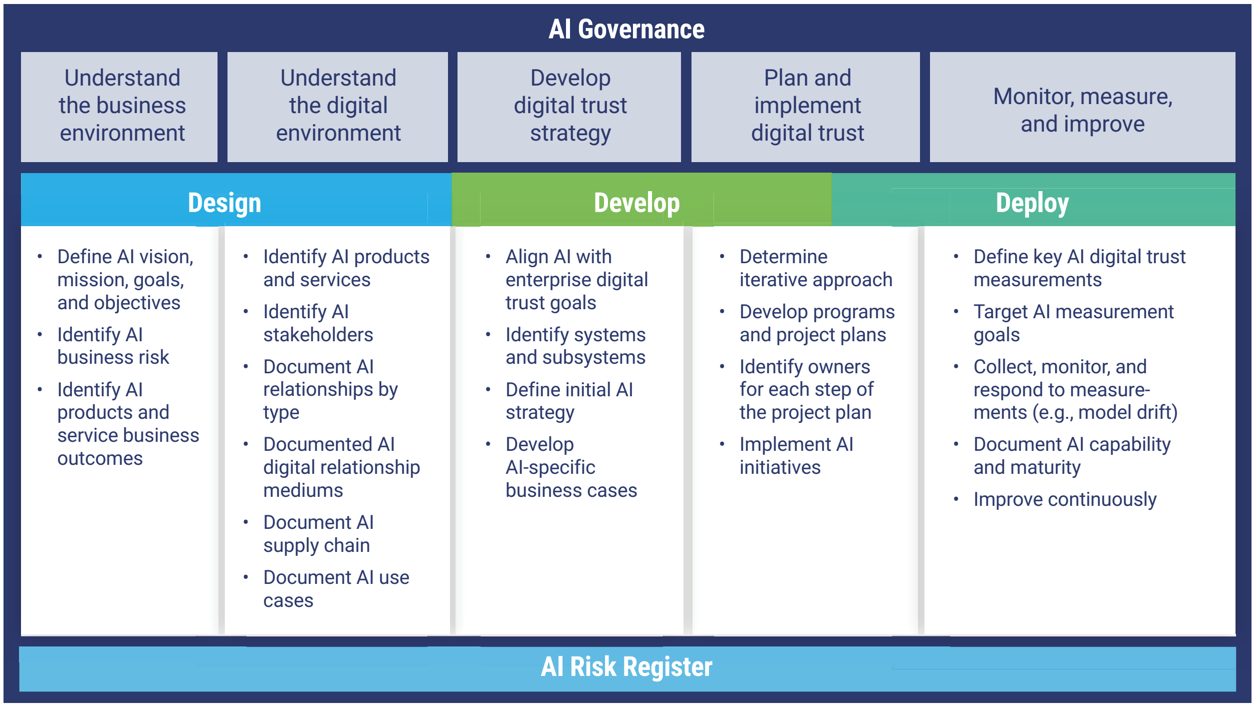
FIGURE 6: DTEF Implementation Model


1
Understand the business environment.
2
Understand the digital environment.
3
Develop digital trust strategy.
4
Plan and implement digital trust.
5
Monitor, measure and improve.
Governance and Oversight

Source: ISACA, *Digital Trust Ecosystem Framework Implementation Guide*, 2024

Figure 7 illustrates the relationship between the DTEF implementation model, AI life cycle, and key AI-specific activities.

FIGURE 7: Mapping of AI Life Cycle to DTEF Implementation Model



AI Governance

Understand the business environment	Understand the digital environment	Develop digital trust strategy	Plan and implement digital trust	Monitor, measure, and improve
Design		**Develop**		**Deploy**
• Define AI vision, mission, goals, and objectives • Identify AI business risk • Identify AI products and service business outcomes	• Identify AI products and services • Identify AI stakeholders • Document AI relationships by type • Documented AI digital relationship mediums • Document AI supply chain • Document AI use cases	• Align AI with enterprise digital trust goals • Identify systems and subsystems • Define initial AI strategy • Develop AI-specific business cases	• Determine iterative approach • Develop programs and project plans • Identify owners for each step of the project plan • Implement AI initiatives	• Define key AI digital trust measurements • Target AI measurement goals • Collect, monitor, and respond to measurements (e.g., model drift) • Document AI capability and maturity • Improve continuously

AI Risk Register

Understand the Business Environment

Any business initiative necessitates a clear understanding of the current business environment inclusive of vision, mission, goals, and objectives; for without these, there may be misalignment. Early stakeholder engagement is especially important when it comes to AI given its current popularity and oftentimes lofty expectations. It is important to identify enterprise risk and tolerance, business pain points, and desired outcomes, because not everything should be automated. This work serves to identify scope, assumptions, and constraints. The tasks as they relate to AI in this phase are:

1. Formulate the enterprise's AI vision, mission, goals, and objectives.
2. Understand the business risk related to AI.
3. Identify the products and services that use AI.

Understand the Digital Environment

Similar to the preceding phase, this phase involves data gathering, including documenting the digital aspects of the organization's products and services, and identifying various relationships and use cases for the digital interactions. The tasks related to AI in this phase are:

1. Identify AI-enabled products and services currently used in the enterprise.
2. Identify a diverse slate of AI stakeholders.
3. Define AI digital relationship types (e.g., business to consumer, business to employee, government to constituent).
4. Define AI digital relationship mediums.

5. Understand the AI digital supply chain.
6. Create AI digital interaction use cases.

Develop the Digital Trust Strategy

This phase requires an enterprise to develop a digital trust strategy based on the enterprise's business and digital environments. The tasks as they relate to AI in this phase are:

1. Document strategic AI digital trust goals.
2. Map AI systems and subsystems.
3. Develop initial AI strategy.
4. Create an AI business case.

Plan and Implement Digital Trust

This phase is where the work of AI implementation happens as it pertains to execution. To be successful, an organization must, based on the first three phases, plan the AI digital trust initiative and then implement it. The tasks in this phase are:

1. Create an iterative approach.
2. Develop AI program/project plan.
3. Assign AI owners to each step of the project plan.
4. Implement AI.

Monitor, Measure, and Improve

Finally, the fifth phase is the vital link to creating a continuous model. This includes continuous monitoring, measurement, and improvements and is the link back to phase one, which triggers the continuous model. Continuous monitoring in certain AI applications such as ML is key, because the AI output is constantly changing based on new data sets being added to the model. The tasks as they relate to AI in this phase are:

1. Identify key AI digital trust measurements.
2. Set target AI measurement goals.
3. Collect, monitor, and respond to measurements.
4. Assess AI capability and maturity.
5. Improve continuously the AI digital trust environment.

Governance and Oversight

Underlying all these phases is governance and oversight, which ensures that digital trust initiatives, including AI, are properly scoped, executed, and improved over time. The tasks in this phase are:

1. Identify and adopt AI governance.
2. Create and manage the AI risk register.
3. Integrate AI digital trust in the enterprise governance, risk, and compliance (GRC) structure.

Applying the DTEF

Unlike traditional software systems, AI systems pose unique challenges that necessitate a specialized governance framework. As AI systems continue to evolve at an unprecedented pace, their impact on businesses becomes increasingly more significant. The rapid increase in AI usage requires oversight and a structured approach to reduce harm. The magnitude of variability in enterprise purpose and infrastructure coupled with each unique use case make it impractical to cover every possible way the DTEF can help enterprises maximize organizational value while also minimizing AI risk.

The rapid increase in AI usage requires oversight and a structured approach to reduce harm.

Today, a great deal of AI is adopted as either an integration to existing software or via an application programming interface (API). Ideally, enterprise software updates that include AI functionality should initiate a review using established evaluation and approval processes to enable a business review of accompanying terms of use and business risk vs. reward determination. Of note, the ease of disabling AI functionality varies by product and is not always intuitive.[14]

Siloed vendor management and procurement processes are insufficient for the complexity associated with today's enterprise digital ecosystem. Industry data paints an ominous picture of companies inadequately managing third-party risk[15] and rise of API attacks[16] when more than 40% of companies lack awareness of all APIs in use.[17] Recognizing that more than half of those affected by supply chain breaches are attributed to supply chain failures,[18] enterprises must place more emphasis on API security to address common weaknesses.[19] The remainder of this paper will explore a common use case and a representative number of DTEF components; none of which are meant to be exhaustive.

Use Case: Customer Service Chatbots

The rise of generative AI is attributed to 70% of enterprise customer experience (CX) leaders re-evaluating the CX according to a Zendesk survey.[20] That same survey revealed ten trends across three themes, all of which involve AI and necessitate awareness, engagement, and oversight by digital trust practitioners of any primary discipline. Those themes are depicted in **figure 8**.

[14] Kaelin, M.; "How to Disable Windows 11 Copilot Through Registry File or Group Policy Editor," TechRepublic, 20 October 2023, https://www.techrepublic.com/article/how-to-disable-copilot-windows-11/; Salesforce, "Enable and Disable Data Cloud AI and Beta Features with Feature Manager," https://help.salesforce.com/s/articleView?id=release-notes.cdp_rn_2024_winter_feature_manager.htm&release=246&type=5

[15] Bolton, R.; "Third-Party Cybersecurity Risk Management — Updates for a Changing Risk Environment," Community Banking Connections, https://www.communitybankingconnections.org/Articles/2023/I2-I3/third-party-cybersecurity

[16] SALT Labs, "State of API Security Q1 2023," SALT, https://content.salt.security/rs/352-UXR-417/images/SaltSecurity-Report-State_of_API_Security.pdf; Matson, K.; "The Next big API Security Breach Looms: Here's how to Prepare," SC Media, 19 October 2023, https://www.scmagazine.com/perspective/the-next-big-api-security-breach-looms-heres-how-to-prepare

[17] Nagaraj, S.; "The State of API Security in 2023," InfoWorld, 2 November 2023, https://www.infoworld.com/article/3709450/the-state-of-api-security-in-2023.html

[18] *Op cit* Bolton, https://www.communitybankingconnections.org/Articles/2023/I2-I3/third-party-cybersecurity

[19] OWASP, "OWASP API Security Project," https://owasp.org/www-project-api-security/

[20] ZenDesk, "CX Trends 2024," https://cxtrends.zendesk.com/reports/cx-trends-report

FIGURE 8: Customer Experience Trends 2024

	Theme		
	AI and Intelligent Experiences	Data and Trustworthy Experiences	Next Gen and Immersive Experiences
Trends	1. Generative AI accelerates delivery of a more humanized journey. 2. Increased capability of chatbots. 3. Growing disconnect between CX leaders on AI strategy, tools, and role impact. 4. AI transparency and decision-making shifts from exception to norm.	1. Business focus on data-driven dynamic user experience. 2. CX leaders become prominent stakeholders for data privacy. 3. Security-by-design is normal-ized.	1. Live experience influences future of online shopping. 2. Voice focus on handling complex and/or issue escalation. 3. Predictive agent management tools overtake traditional methods.

Source: Adapted from Zendesk, "CX Trends 2024," https://cxtrends.zendesk.com/

Expectations for generative AI are high in the customer service area, so it should come as no surprise that more than half of CX leaders are exploring AI vendors. Those involved in CX work are fueled with lofty expectations that chatbots will continue to transform into digital agents with far greater capabilities. In the interim, enterprises are wise to form cross-functional enterprise teams to manage compliance issues and minimize the risk associated with integrations to their existing digital ecosystem.

To transition from these expectations to the practical application of chatbots, it is important to understand their categorization. Chatbots are commonly categorized as simple, smart, or hybrid. Chatbots have until recently been primarily rule-based, which offers a consistent, reliable experience but is increasingly shifting to AI by leveraging natural language processing (NLP). Besides being rule-based, bots may be keyword-based, menu-based, intelligent (contextual), hybrid, or voice-enabled.[21] Collectively, chatbots enhance customer service, overcome language barriers, and attempt to eliminate any frustration associated with lengthy call hold times. The DTEF is poised for the challenges involved in strategy, vendor selection and management, implementation, and continuous improvement.

How can DTEF Help?

Regardless of the type of AI currently in use or being considered for use, it is imperative that enterprises have an AI strategy aligned to corporate strategy and outcomes. The specific strategy and development cycle will vary widely depending on the type of AI being considered.

The following include some, but not all, of applicable trust factors, practices, and activities within specific domains that provide useful guidance on the AI chatbot example.

[21] Shenoy, A.; "6 Types of Chatbots – How to Choose the Best for Your Business?," Yellow.ai, 9 January 2024, https://yellow.ai/blog/types-of-chatbots/

Culture

AI is subject to fear, uncertainty, and doubt, especially regarding the extent that AI will affect an individual's or occupation's work. Enterprises are encouraged to develop an organizational strategy that accounts for shifts in which types of work are performed by whom.

Open dialogue between those responsible for organizational development and hiring managers can foster buy-in with employees about the risk and opportunities provided by AI. Conversely, clarity between an enterprise and external stakeholders regarding the purpose, use, and governance of AI would be beneficial. An example of this is frequently asked questions (FAQ) pages.

AI-related considerations within the Culture domain include:

- Defining the target culture for the deployment and use of AI: Why is it being used, how is it being used, and what is the cultural 'footprint' vis-a-vis internal and external stakeholders
- Addressing any cultural inhibitors to successful AI development and integration
- Managing the target culture in terms of the wider environment, e.g. levels of acceptance, adoption rate, general knowledge and skills, common usage patterns, etc.

All three trust factors within the Culture domain are especially relevant whenever an enterprise is considering the adoption of AI:

1. *Manage Culture (CU.01)*: Evaluate, adjust, and promote an organizational and human culture that fosters digital trust.
   Practice number CU.01.02: Modify Culture, includes activities that communicate the company's strategy and values and incorporates lessons learned and factors in external requirements and expectations. This practice underscores the need for demonstrative top-down ethical leadership.
   Practice number CU.01.03: Promote Culture, includes training and awareness for employees to understand and uphold the digital trust strategy.
2. *Create and Manage the Digital Trust Cultural Environment (CU.02)*: Define and establish the management system for digital trust across the ecosystem.
   Within the practice titled, Modify Culture, is activity CU.02.02.6: Communicate the information at the appropriate level of detail for respective external stakeholders.
3. *Manage Skills and Competencies (CU.03)*: This trust factor focuses on the identification and maintenance of optimal skills, competencies, and capacities of required human resources. This information will be essential for managing AI-related risk, which could result when key stakeholders fail to understand their roles and responsibilities relative to ensuring and maintaining digital trust in emerging technology environments.

Human Factors

The Human Factors domain can help an enterprise anticipate staffing needs. It can also address areas of CX that require consideration and close monitoring. While technology-related knowledge, skills, and competency gaps are not unique to AI, AI is the newest and most dynamic, and therefore will—at least for the foreseeable future—require deliberate talent management strategies to identify and manage up/reskill needs for overall AI program success and individual projects. Further, AI requires continuous monitoring to confirm the model is performing as specified.

AI-related considerations in Human Factors include:

- Making the AI implementation understandable and accessible, using multiple communication vectors and multiple presentation layouts for AI-driven results

- Defining controls supporting the user experience and verification of AI-driven results (discriminator, professional skepticism, etc.)
- Defining and deploying continual improvement process, introducing feedback and feedforward loops for human interaction with the AI actor
- Including rating and scoring mechanisms to enable human actors to attribute levels of trust to the AI implementation

In the chatbot example, practitioners may want to look at:

- Practice HF.01.04 and all its associated activities: Evaluate and Manage Technology Human Factor Capacity.
    - Activity HF.01.05.1: Identify and remediate sources of errors and interruptions related to digital trust.

Architecture

A successful AI integration harmonizes data, people, processes, and technology.[22] To accomplish this, enterprises need a strategy—a conscious, deliberate process involving choice, compromise, and saying no when necessary to satisfy business objectives. According to the Open Group, enterprise architecture (EA) is a strategy tool to identify and close gaps between the current and future state. Good EA allows enterprises to translate strategy into execution.[23]

A successful AI integration harmonizes data, people, processes, and technology.

The DTEF Architecture domain covers topics that define, develop, and manage the overall enterprise architecture. It includes areas such as plans, policies, and standards for business, data, application, and technology layers of the EA model.

AI-related considerations in Architecture include:

- Mapping the AI Universe to the overall informational communications technology (ICT) environment and business-facing use cases
- Embedding AI actors in the overarching architecture, including third parties and longer supply chains
- Defining and securing the resource requirement to run, control, and manage AI actors
- Considering the domain Culture and Emergence, continuously adopting new or alternative use cases for AI actors to enhance the architecture
- Managing and regularly reviewing the business and financial case for AI use

Relevant to the chatbot use case, the following trust factors, practices, and activities are useful:

1. *Manage Digital Trust Resources (AR.03)*: This trust factor entails the identification, management, and controlling of resources required to ensure digital trust. This is necessary to manage all elements and components of the infrastructure and the following practices and activities would be relevant:
    - Practice AR.03.02: Manage Digital Trust Applications.
        - Activity AR.03.02.1: Manage customer-facing applications (e.g., front end) for digital trust.
    - Practice AR.03.05: Manage Digital Trust Operations: Manage and control all operational and production aspects of digital trust and the related architecture.
        - Activity AR.03.05.1: Estimate the size, effort, duration, and cost of the work and resources needed to develop, acquire, or deliver an operations plan related to digital trust.

[22] Strickrodt, D.; "The Future of Enterprise Architecture and AI Integration. Embrace the Synergy.," Bizzdesign, 27 October 2023, https://bizzdesign.com/blog/the-future-of-enterprise-architecture-and-ai-integration/

[23] The Open Group, "A Practitioners' Approach to Developing Enterprise Architecture Following the TOGAF® ADM," https://pubs.opengroup.org/togaf-standard/adm-practitioners/adm-practitioners_3.html

2. *Align Digital Trust Technology With Organizational Needs (AR.04)*: This trust factor pertains to the identification of organizational needs and aligning technology to those needs.
    - Practice AR.04.01: Align Technology to Business Needs.
        - Activity AR.04.01.1: Identify relevant business objectives.

Direct and Monitor

The potential risk associated with AI[24] has been predominantly characterized as a future concern. However, proactively addressing AI risk can help to ensure the safe and responsible deployment of AI technology within the business. Despite the importance of recognizing future risk, the present state of AI technology and current utilization patterns present immediate risk that must be addressed. Plainly stated, AI can exacerbate any existing issues, such as a lack of quality control or poor data integrity, leaving systems vulnerable to cyberattacks and possibly introducing new ones.

Despite the importance of recognizing future risk, the present state of AI technology and current utilization patterns present immediate risk that must be addressed. Plainly stated, AI can exacerbate any existing issues, such as a lack of quality control or poor data integrity, leaving systems vulnerable to cyberattacks and possibly introducing new ones.

The Direct and Monitor domain contains topics related to creating, measuring, managing, and governing the digital trust ecosystem to include risk, communications, information, sustainability and resilience, compliance, assurance, and overall evolution of the enterprise.

AI-related considerations in Direct and Monitor include:

- Embedding the AI universe as part of a larger GRC framework in IT
- Recursively applying elements of governance, risk management, compliance, and assurance to the AI universe within (and without) the organization
- Organizing human and organizational structures to steer and control AI actors (including their life cycle from onboarding to decommissioning)
- Embedding AI in the three lines of defense

While there are many pertinent trust factors, practices, and activities that have application in the chatbot example, the following come to mind:

1. *Manage Risk (DM.03)*: This trust factor involves continually identifying, assessing, and reducing digital ecosystem-related risk within tolerance levels set by enterprise executive management.
    - Practice DM.03.01: Direct and Monitor Risk Management.
        - Activity DM.03.01.1: Identify roles and responsibilities.
        - Activity DM.03.01.2: Establish risk appetite and risk-tolerance levels.
    - Practice DM.03.02: Identify Digital Ecosystem Risk.
        - Activity DM.03.02.2: Identify risk owners.
        - Activity DM.03.02.3: Identify current risk controls/control environment.
        - Activity DM.03.02.6: Integrate digital ecosystem risk into the larger enterprise risk. management (ERM)

[24] ISACA, "The Promise and Peril of the AI Revolution," 12 September 2023, https://www.isaca.org/resources/white-papers/2023/the-promise-and-peril-of-the-ai-revolution

2. *Manage Organization (DM.04)*: This trust factor requires an enterprise to identify and organize structures to support the digital trust ecosystem.
    - Practice DM.04.01: Manage Organizational Structure.
        - Activity DM.04.01.4: Establish roles and responsibilities, including, if appropriate, establishing a trust steering committee (or equivalent) composed of executive, business, and information and technology (I&T) management to track the status of projects, resolve resource conflicts, and monitor service levels and service improvements.
3. *Administer Digital Trust (DM.06)*: This trust factor calls for an organization to maintain appropriate digital trust practices as applied to information.
    - Practice DM.06.01: Inventory Information Assets.
        - Activity DM.06.01.3: Discover and inventory contracts and other documents governing external relationships.

Emergence

In today's fast-paced and digitized environment, organizations need to be aware of external environmental factors that may influence future success. Organizations that consider these factors can increase their agility and resilience. From an agility perspective, they can move and adapt more quickly, flexibly, and decisively. From a resilience perspective, they can anticipate, respond, and adapt to changes or disruptions.

In today's fast-paced and digitized environment, organizations need to be aware of external environmental factors that may influence future success.

The Emergence domain focuses on events and activities that could trigger opportunities at process and people levels, including internal changes, external influences, and people-driven deviations.

AI-related considerations in Emergence include:

- Anticipating emergence through generative and nongenerative AI (e.g., defining the expected boundaries of AI actors)
- Analyzing and verifying process and outcomes models (e.g., what is the AI actor supposed to achieve, and what are possible deviations [be mindful of reasonably foreseeable misuse])
- Controlling input variables, training, and evolution of the AI actor
- Monitoring the emergent behavior of the AI actor and making adjustments where necessary to maintain an acceptable level of trust

Relevant to the chatbot use case, the following trust factors and practices are useful:

1. *Identify, Evaluate, and Manage Potential Triggers (EM.01)*: This trust factor entails identifying, evaluating, and managing potential triggers.
    - Practice EM.01.01: Identify and Manage Internal Signals and all its associated activities.
2. *Detect and Manage Process- and People-Related Emergence (EM.02)*: This trust factor requires enterprises to identify imminent or completed change through emergence and manage the results.
    - Practice EM.02.01: Detect Internal Emergence and all its associated activities.

Enabling and Support

The Enabling and Support domain is the dynamic interconnection through which technology enables processes, which in turn supports the deployment and operation of technology. Adoption of practices within the Enabling and Support domain can assist in identifying problems before chatbots are deployed to production.

AI-related considerations in Enabling and Support include:

- Embedding the AI Universe in the service value chain and service management
- Defining and describing the AI actors as part of processes, services, and the overall service portfolio
- Steering and controlling further AI development (considering the provider/user transition)
- Continuously monitoring AI operation, with interfaces to Emergence and Human Factors

Relevant to the chatbot use case, the following trust factors, practices, and activities are useful:

1. *Manage Digital Trust Ecosystem Objectives (ES.01)*: This trust factor is about determining process and technology objectives as well as enabling expectations, including quality considerations.
    - Practice ES.01.01: Determine Process Objectives and all its associated activities.
    - Practice ES.01.03: Determine Process Specifications and all its associated activities.
2. *Implement Services and Solutions (ES.05)*: This trust factor requires an organization to plan, coordinate, and implement services and solutions.
    - Practice ES.05.01: Plan Implementation and all its associated activities.
3. *Monitor Services and Solutions (ES.07)*: This trust factor contemplates continuously monitoring the operation of services, solutions, and related technology for digital trust.
    - Practice ES.07.01: Monitor Process Operation.
        - Activity ES.07.01.1: Manage changes to process operations.
        - Activity ES.07.01.2: Monitor operational metrics and controls.
        - Activity ES.07.01.4: Monitor process alignment.
        - Activity ES.07.01.7: Identify improvement items and feed into quality management process.

Conclusion

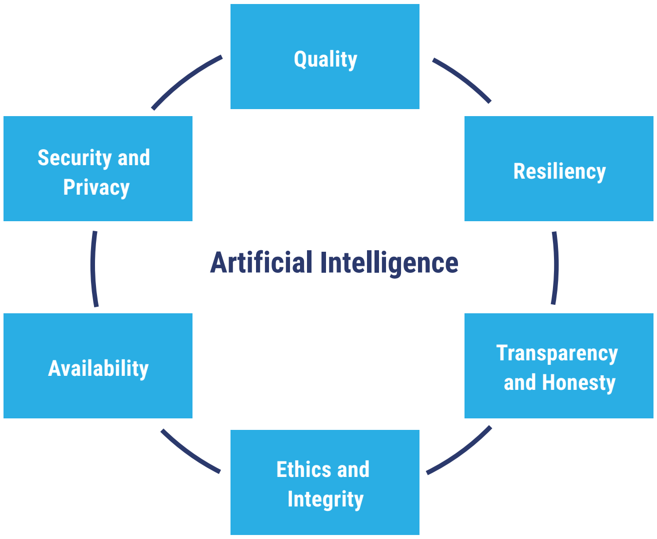
AI is not new, but recent developments—namely the rise of generative AI—thrust the market into overdrive. Enterprises are wise to embrace the reality that at least some employees have used some sort of AI and possibly put intellectual property at risk. Meanwhile, software of all types is being adopted that has AI incorporated into it, which requires enterprises to perform due diligence prior to deploying updates. Further, the vendor market is heating up as solution providers race to remain competitive.

The reality is AI is here and despite any risk, it is unrealistic that it be halted, given its many benefits, regardless of industry or business function. Legislation and standards will be the primary driver in establishing guardrails to attain the ethical, responsible use of technology, and the former is likely to follow the same course and trajectory that privacy regulations have—a complex web of nonuniform laws certain to create headaches for GRC professionals.

The question is no longer whether an enterprise will adopt AI-based technology, but how much. Even if enterprises do not develop their own private models, AI is pervasive, and the nonexplicitly permitted use of popular generative AI products represents an evolution of shadow IT. Enterprises need an AI strategy aligned to business objectives and must validate that AI instances are used to solve business problems and are within accepted risk tolerance levels.

To accomplish this, enterprises need to have a robust framework to help ensure compliance with AI-related legal and regulatory requirements, international standards guidance, and customer contractual obligations. Although there are many frameworks available in the industry for different practice areas, the DTEF framework helps enterprises achieve quality, resiliency, transparency and honesty, ethics and integrity, availability, security, and privacy in their AI-enabled processes and systems (**figure 9**).

FIGURE 9: Elements of Trustworthy AI



DTEF is also flexible enough to be leveraged with other frameworks. Adopting the DTEF to implement AI-based technology means all these aspects would be considered throughout the life cycle, and it also helps to break organizational silos, which can surface between processes, people, and technology. While AI is designed, developed, and deployed in organizations, the benefit DTEF offers is that, if implemented effectively, it should help the organization demonstrate compliance with regulations, customer requirements, international standards, organizational policies and procedures, but most importantly, that solutions achieve business objectives in an ethical, responsible manner.

Acknowledgments

ISACA would like to acknowledge:

Lead Developer

Chetan Anand

CDPSE, CPISI, ICBIS, ICCP

AVP, Information Security and CISO, Profinch Solutions Pvt Ltd., India

Expert Reviewers

Joyce Chua

CISA, CISM, CDPSE, CAEG (Professional), (C)CISO, CFE, CIA, CIMP, CIPM, CIPP(A), CIPP(E), FIP, ITIL, MCP, PMP

United Overseas Bank, Singapore

J. Winston Hayden

CISA, CISM, CGEIT, CRISC, CDPSE

South Africa

Ed Moyle

CCSK, CISSP

Chief Information Security Officer for Drake Software, USA

Geetha Murugesan

CISA, CGEIT, CRISC, CDPSE, ISO 22301:2019, ISO 27001:2013, ISO 31000:2018, ISO 9000:2015

India

Rolf von Roessing

CISA, CISM, CGEIT, CDPSE, CISSP, FBCI

Partner, FORFA Consulting AG, Switzerland

Board of Directors

John De Santis, Chair

Former Chairman and Chief Executive Officer, HyTrust, Inc., USA

Brennan P. Baybeck, Vice-Chair

CISA, CISM, CRISC, CISSP

Senior Vice President and Chief Information Security Officer for Customer Services, Oracle Corporation, USA

Stephen Gilfus

Managing Director, Oversight Ventures LLC, Chairman, Gilfus Education Group and Founder, Blackboard Inc., USA

Niel Harper

CISA, CRISC, CDPSE, CISSP, NACD.DC

Former Chief Information Security Officer, United Nations Office for Project Services (UNOPS), USA

Gabriela Hernandez-Cardoso

NACD.DC

Independent Board Member, Mexico

Jason Lau

CISA, CISM, CGEIT, CRISC, CDPSE, CIPM, CIPP/E, CIPT, CISSP, FIP, HCISPP

Chief Information Security Officer, Crypto.com, Singapore

Massimo Migliuolo

Independent Director, Former Chief Executive Officer and Executive Director, VADS Berhad Telekom, Malaysia

Maureen O'Connell

NACD.DC

Board Chair, Acacia Research (NASDAQ), Former Chief Financial Officer and Chief Administration Officer, Scholastic, Inc., USA

Erik Prusch

Chief Executive Officer, ISACA, USA

Asaf Weisberg

CISA, CISM, CGEIT, CRISC, CDPSE, CSX-P

Chief Executive Officer, introSight Ltd., Israel

Pamela Nigro

ISACA Board Chair 2022-2023

CISA, CGEIT, CRISC, CDPSE, CRMA

Vice President, Security, Medecision, USA

Gregory Touhill

ISACA Board Chair 2021-2022

CISM, CISSP

Director of the CERT Division at Carnegie Mellon University's Software Engineering Institute, USA

Tracey Dedrick

ISACA Board Chair, 2020-2021

Former Chief Risk Officer, Hudson City Bancorp, USA

About ISACA

ISACA® (www.isaca.org) is a global community advancing individuals and organizations in their pursuit of digital trust. For more than 50 years, ISACA has equipped individuals and enterprises with the knowledge, credentials, education, training and community to progress their careers, transform their organizations, and build a more trusted and ethical digital world. ISACA is a global professional association and learning organization that leverages the expertise of its 180,000+ members who work in digital trust fields such as information security, governance, assurance, risk, privacy and quality. It has a presence in 188 countries, including 225 chapters worldwide. Through the ISACA Foundation, ISACA supports IT education and career pathways for underresourced and underrepresented populations.

DISCLAIMER

ISACA has designed and created *Using the Digital Trust Ecosystem Framework to Achieve Trustworthy AI* (the "Work") primarily as an educational resource for professionals. ISACA makes no claim that use of any of the Work will assure a successful outcome. The Work should not be considered inclusive of all proper information, procedures and tests or exclusive of other information, procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific information, procedure or test, professionals should apply their own professional judgment to the specific circumstances presented by the particular systems or information technology environment.

RESERVATION OF RIGHTS

ISACA
1700 E. Golf Road, Suite 400
Schaumburg, IL 60173, USA

Phone: +1.847.660.5505

Fax: +1.847.253.1755

Support: support.isaca.org

Website: www.isaca.org

Participate in the ISACA Online Forums:
https://engage.isaca.org/onlineforums

X: www.x.com/ISACANews

LinkedIn:
www.linkedin.com/company/isaca

Facebook:
www.facebook.com/ISACAGlobal

Instagram:
www.instagram.com/isacanews/