ISACA

利用数字信任生态系统框架实现可信的人工智能

治理

目录

摘要

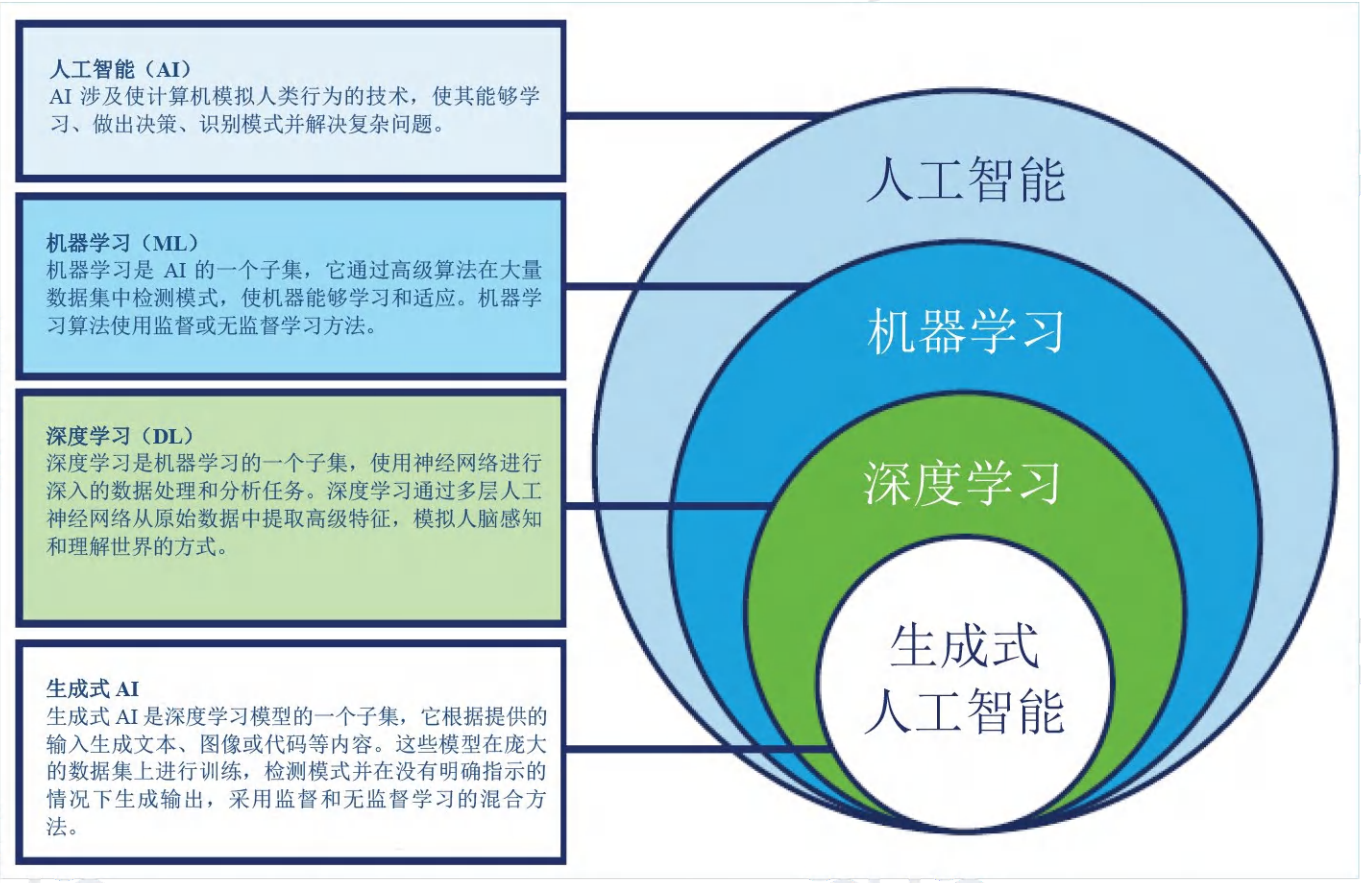
本白皮书探讨了企业在采用人工智能（AI）技术和服务时使用 ISACA 数字信任生态系统框架（DTEF）的优势。它帮助理解 DTEF 如何支持新兴技术风险评估，并提供建立治理结构的指导，以在整个 AI 生命周期内为组织带来帮助。白皮书重点介绍了数字信任关键要素，它们被视为成功整合 AI 技术和服务交付的基础，并通过典型用例场景展示组织通常遇到的情况。

导言

人工智能（AI）无处不在，具体实例包括聊天机器人、金融欺诈检测和导航软件。AI 涵盖机器学习（ML）、深度学习和生成式 AI（图 1）。总体而言，它继续革新所有行业，以速度和规模提供效率和利益。例如，在医疗保健领域，AI 支持个性化治疗方案和预测诊断管理；在金融服务领域，AI 通过数据驱动的洞察力增强了欺诈检测和风险管理；在能源部门，AI 优化电网管理和预测性维护，释放了效率和可持续性的现实效益。

人工智能的广泛影响不仅体现在企业的明确采用，还体现在它与各种第三方应用程序（如主流的办公生产力软件和日常任务）的集成。此外，某些部门（如人力资源、市场营销）的员工可能已经在使用基于网络的软件来筛选求职者或编写营销内容。简而言之，即使员工没有意识到，他们可能已经在使用人工智能了。最终，未被明确允许用于商业用途的生成式 AI 产品数量的增加，代表了一种影子 IT 的演变。

图 1： AI、机器学习、深度学习和生成式 AI 的比较视图



资料来源：揭开 AI 复杂性的面纱 - AI、机器学习、深度学习和生成式 AI 的比较视图，https://commons.wikimedia.org/wiki/File:Unraveling_AI_Complexity_-_A_Comparative_View_of_AI,_Machine_Learning,_Deep_Learning,_and_Generative_AI.jpg。本图由 *Creative Commons Attribution-ShareAlike 4.0 International* 许可证提供。

尽管 AI 能够提升和优化业务，但在许多情况下，它也可能增加风险。和任何技术一样，不法分子很快便开始利用 AI 实现不良目的。目前，生成式 AI 已经提高了与商业电子邮件诈骗（BEC）[1] 相关的电子邮件的可信度，同时降低了实施 BEC 攻击所需的技术能力。[2]此外，AI 还被用于地缘政治报道[3]、基于图像的滥用[4]和政治活动[5]，并且成功实施了数百万美元的诈骗。[6] 总的来说，AI 对不同的人意味着不同的东西，并且每种 AI 所关联的风险也是高度变化的。

AI 的普及程度不可低估。根据德勤的报告，全球正经历一场前所未有、最易获取的技术革命，这场革命以生成式 AI 的形式出现，其规模堪称史诗级。[7]尽管 AI 的概念早在 1950 年就已提出，[8] 但其进展的规模和速度以及相关影响还远未完全显现。像过去的技术变革一样，许多人一听到 AI 就会产生恐惧、怀疑和不信任。然而，值得注意的是，人们认为，生成式 AI 将释放人类潜力，而非取代人类。因此，当前以及未来的挑战在于合理分配和管理机器和人类各自擅长的任务。使用 AI 不仅需要关注传统的人员、流程和技术因素，还需要跨职能部门的领导力以及仔细考量其对业务和社会的影响。在此结合 ISACA 的数字信任生态系统框架（DTEF）。[9]

AI 的普及程度不可低估。根据德勤的报告，全球正经历一场前所未有、最易获取的技术革命，这场革命以生成式 AI 的形式出现，其规模堪称史诗级。

DTEF 支持从所有组织利益相关者的角度建立和维护数字信任。数字信任不仅仅是技术问题；它适用于整个组织及其所有外部利益相关者。

选择、建立和维护数字关系需要各方的信心和透明度。供应商和消费者的需求、原则、价值观和目标影响所需的信任水平。DTEF 为企业提供了一种创新的数字信任和转型方法，以优先考虑数字信任——这是一个与所有 AI 种类 相关的概念。

本白皮书探讨了组织如何利用 ISACA 的 DTEF 框架实现由 AI 赋能的技术和服务解决方案的数字信任可靠性。

---

1 ISACA，《ISACA 词汇表》，https://www.isaca.org/resources/glossary

2 Kelley, D.;“WormGPT——网络犯罪分子用于发起商业电子邮件诈骗攻击的生成式 AI 工具”，SlashNext，2023 年 7 月 13 日，https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/; Shiebler, D.;“生成式 AI 使威胁行为者能够制造更多（更复杂）电子邮件攻击”，Abnormal Security，2023 年 6 月 14 日，https://abnormalsecurity.com/blog/generative-ai-chatgpt-enables-threat-actors-more-attacks

3 Klepper, D.;“虚假婴儿，真实恐怖：加沙战争中的深度伪造技术加剧了对 AI 误导能力的担忧”，APNews，2023 年 11 月 28 日，https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47

4 Saner, E.;“泰勒·斯威夫特深度伪造丑闻内幕：这是男人在告诉一位有力量的女性回到她的框框里”，The Guardian，2024 年 1 月 31 日，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.;“瓦拉斯竞选团队谴责发布到推特上的深度伪造视频”，CBS News，2023 年 2 月 27 日，https://www.cbsnews.com/chicago/news/vallas-campaign-deepfake-video/; Harper, A.; Gehlen, B.; et al.;“在政治竞选中使用 AI 引发对 2024 年选举的警示”，ABC News，2023 年 11 月 8 日，https://abcnews.go.com/Politics/ai-political-campaigns-raising-red-flags-2024-election/story?id=102480464; mer, A.; Tong, A.;“深度伪造：美国 2024 年选举与 AI 热潮碰撞”，路透社，2023 年 5 月 30 日，https://www.reuters.com/world/us/deepfaking-it- americas-2024-election-collides-with-ai-boom-2023-05-30/

6 Edwards, B.;“深度伪造诈骗者通过首例 AI 盗窃案骗取 2500 万美元”，ARS Technica，2024 年 2 月 5 日，https://arstechnica.com/ information-technology/2024/02/deepfake-scammer-walks-off-with-25-million-in-first-of-its-kind-ai-heist/

7 德勤，“生成式 AI 与未来的工作”https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-ai-institute- generative-ai-and-the-future-of-work.pdf

8 图灵，A.M.；“计算机器与智能”，Mind 49，1950 年，https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf

9 ISACA，“数字信任生态系统框架”，美国，2022 年，www.isaca.org/dtef-ebook

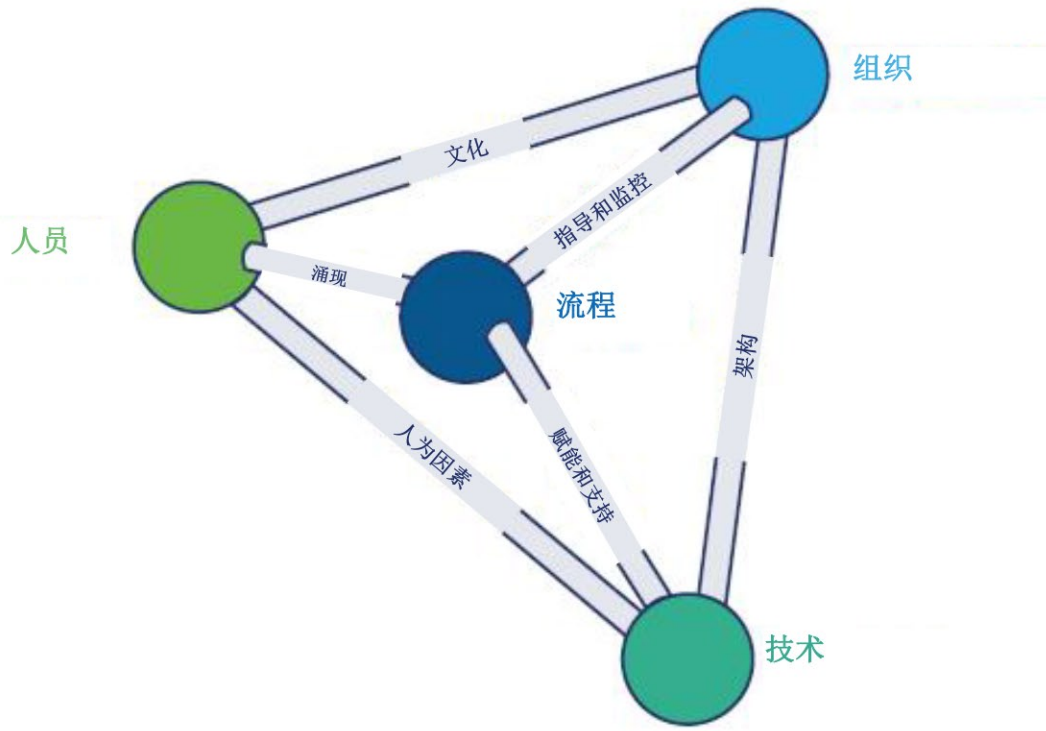
DTEF 概览

DTEF 定义了创建数字信任生态系统的核心要素，该框架考虑到所有利益相关者，以确保所有数字互动和交易的合法性、可信度，并涵盖完整性、安全性、隐私、韧性、质量、可靠性和信心等要素。信任是任何企业在采用和部署 AI 策略时必须纳入的重要原则。例如，信任 AI 输出的能力将取决于数据质量和数据保护的水平。AI 模型必须基于可信代码和透明性开发，包括操作决策。人工智能服务的最终用户或消费者相信，驱动决策的人工智能算法经证明是准确的，保护了最终用户的隐私，是安全的，并且没有偏见。最后，适当的人工智能治理将有助于高级管理层和主要利益相关者相信人工智能决策是可以解释的，公平和道德规范得到遵守，并且能够证明符合监管和法律规定。

DTEF 创建了一个知识体系，有助于应对动态变化的法律、监管和技术环境、现代商业要求、外部/内部影响和新兴因素、以及创建和操作数字信任环境需要面对的风险和需要采取的控制措施。

DTEF 是一个三维模型，其理论基础是人员、流程、技术和组织这四个主要基本节点之间存在多重依赖关系。节点之间通过多种动态互动进行交互，在该模型中，这些节点是最高级别的。这四个节点通过六个域相互关联--文化、涌现、赋能和支持、人为因素、指导与监控以及架构。**图 2** 展示了节点与域之间的关系。

图 2：DTEF 模型



资料来源：ISACA，数字信任生态系统框架，美国，2022 年

域会影响一个或多个节点。例如，架构域的变化必然会影响组织节点和/或技术节点。域之间也以系统的方式相互作用。域在管理组织内部存在的相互联系和复杂性方面发挥着至关重要的作用，因为它们与不断变化的法规、新兴技术、新威胁、程序变化等协同工作。域由一系列构成要素和结构要素组成。

DTEF 使用信任因素在每个域内建立内容基础。例如，架构域分为以下四个信任因素：

1. 创建企业信任架构
2. 管理信息和技术架构
3. 管理数字信任资源
4. 根据组织需求调整数字信任技术

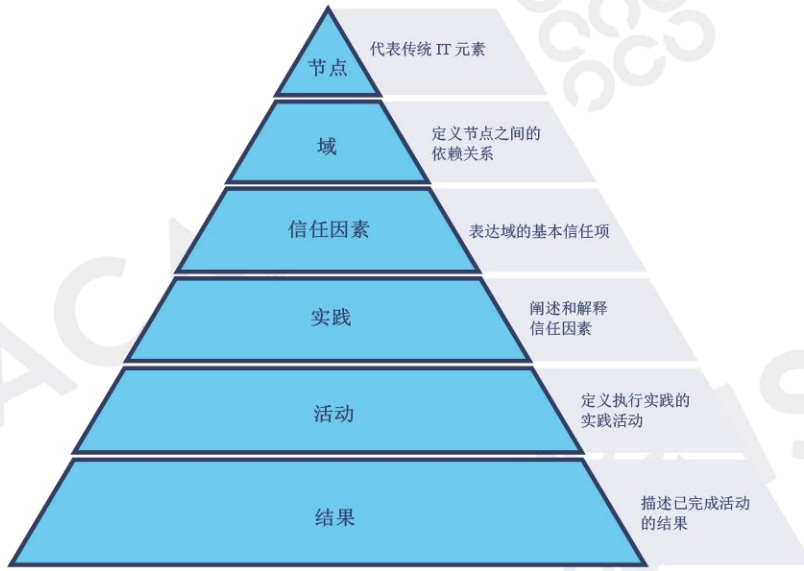
信任因素描述了维护数字信任所需的总体行动，有助于避免或减少偏见。该框架的组成部分（见**图 3**）可用于确保在任何 AI 应用中展示数字信任原则。

图 3：DTEF 组成

资料来源：ISACA，数字信任生态系统框架，美国，2022 年

DTEF 提供了一个支持整个生态系统中数字信任的结构，考虑了各种关系如何影响与消费者、客户和用户的互动水平。**图 4** 展示了 DTEF 的关键结构要素。

图 4：DTEF 层次结构



数字信任[11]不仅仅涉及数字信息和技术。数字信任影响着整个企业；因此，能够展示数字可信度的企业可以获得相当大的竞争优势，并与消费者建立更好的关系。[12]

AI 生命周期

无节制地使用 AI 产品会带来相当大的风险。由于 AI 可能会对整个企业造成不利影响（如知识产权损失、品牌损害、诉讼），企业不仅必须了解可能已经在使用的 AI 产品，还必须了解员工试图通过使用 AI 解决哪些业务问题。在许多方面， AI 只是另一种技术，但其细微差别却要求在其生命周期的所有方面进行深思熟虑的治理和风险管理。AI 生命周期是一个从业务问题到基于 AI 的解决方案的迭代过程。[13]

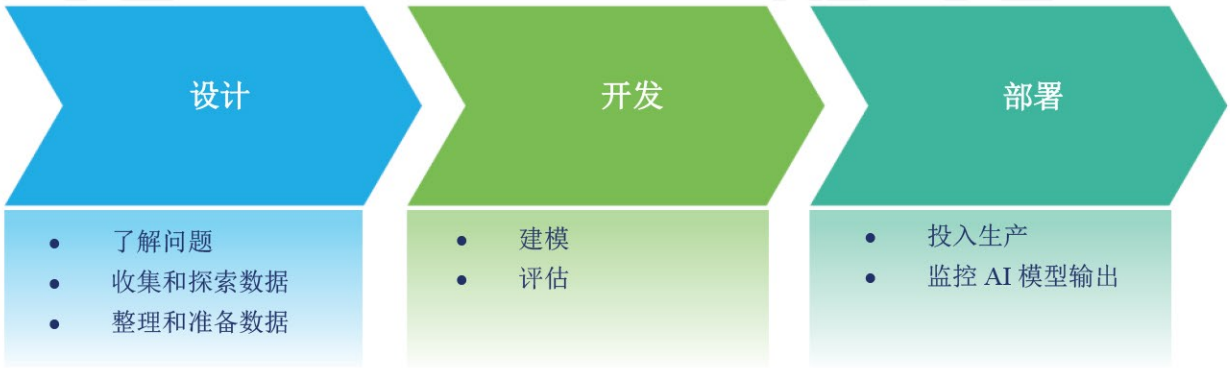
图 5 中显示的每个步骤在设计、开发和部署阶段不断迭代。读者可能熟悉其他 AI 生命周期，这些生命周期可能包含更多的细节。无论有什么不同，AI 生命周期一般都包括四个组成部分：业务需求和目标、数据收集和准备、模型开发和评估、以及运营部署和监控。

11 ISACA 将数字信任定义为对相关数字生态系统中供应商和消费者之间关系、互动和交易完整性的信心。这包括人员、组织、流程、信息和技术创建和维护可信数字世界的能力。

12 Tower-Pierce, J.; “唤醒美国：数字信任可对收入产生积极影响”，InCyber，2023 年 7 月 10 日，https://incyber.org/en/article/wake-up-america-digital-trust-can-positively-impact-revenue/

13 IT 现代化卓越中心，“政府 AI 指南：适用于美国联邦政府的人工智能应用的动态发展指南”，2024 年 3 月 26 日，https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html

图 5：AI 生命周期



设计

了解问题

AI 流程和系统负责人及其他相关利益者负责确定关键项目目标和要求，以有效定义预期的业务成果。组织必须明确定义他们希望 AI 解决什么业务问题。如果不能清楚准确地了解所要解决的业务挑战和预期的结果，任何 AI 解决方案都不会成功。

数据收集和探索

数据是任何 AI 解决方案的基础。在此步骤中，要收集数据并评估其是否适合用于拟议的 AI 应用。这需要发现可用的数据集，找出数据质量问题，并对数据和数据计划的观点有初步的了解。AI 模型只能在清楚了解所需数据和数据构成的情况下使用数据。

数据整理和准备

此步骤包括将工作数据集从最初的原始数据转换成 AI 模型可以使用的格式的所有活动。这一步可能既耗时又乏味，但对于开发 AI 模型以实现解决第一步中确定的问题这一目标来说，却是至关重要的。

开发

建模

此步骤专注于对数据进行实验，以确定正确的 AI 模型。在这一阶段，团队通常会对许多不同的 AI 模型进行训练、测试、评估和再训练，以确定实现预期结果的最佳 AI 模型和设置。AI 模型的训练和选择过程是互动的。没有哪个 AI 模型在第一次训练时就能达到最佳性能。只有通过反复的微调，才能使模型达到预期效果。根据所使用数据的数量和类型，这一训练过程的计算成本可能会非常高；可能需要特殊的设备来提供足够的算力，因为它不可能在普通的笔记本电脑上进行。

评估

一旦构建了一个或多个基于相关评估指标表现良好的 AI 模型，就会在新数据上对 AI 模型进行测试，以确保它们能很好地泛化并达到业务目标。

部署

投入生产

一旦开发的 AI 模型达到了预期的结果，并在实时数据上达到了可以使用的水平，就可以将其部署到生产环境中。在这种情况下，AI 模型将采用不属于训练周期的新数据。

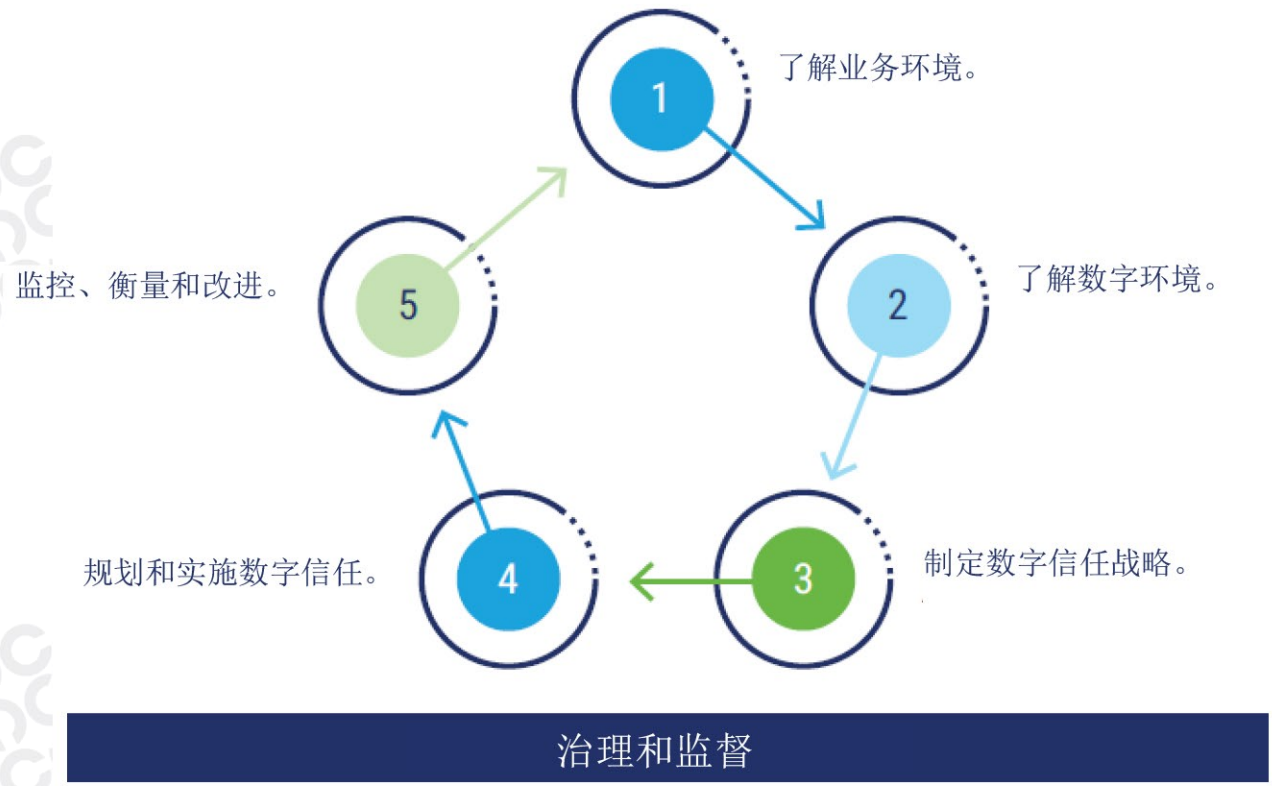
监控 AI 模型输出

一旦部署完毕，就必须对 AI 的生产输出进行监控，以确保其能够充分产生预期结果--这一过程被称为泛化，即 AI 模型对新数据的适应能力。在生产过程中，AI 模型可能会“漂移”，这意味着其性能会随着时间的推移而发生变化。仔细监控漂移非常重要，AI 模型可能需要根据监控的结果不断更新。AI 系统必须经过严格、持续的监控和维护，以确保它们能继续保持训练时的表现，达到预期结果，并解决业务挑战。

专注 AI 的 DTEF 实施

从整体上考虑 AI 项目时，DTEF 实施模型将会变得非常有帮助。DTEF 实施模型分为五个阶段，再加上第六个阶段也就是治理和监督，它是其他五个阶段的基础（参见**图 6**）。

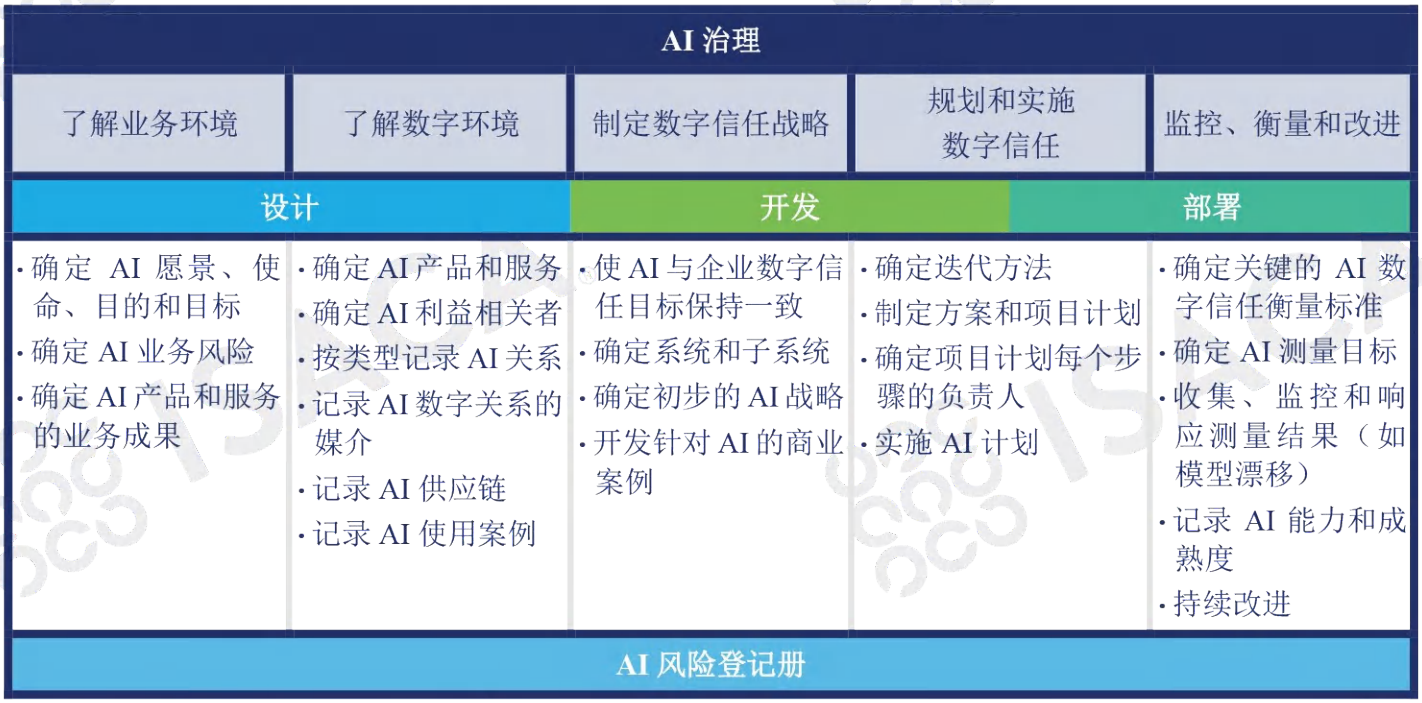
图 6：DTEF 实施模型



资料来源：ISACA，数字信任生态系统框架，美国，2024 年

图 7 展示了 DTEF 实施模式、AI 生命周期和关键 AI 特定活动之间的关系。

图 7：AI 生命周期与 DTEF 实施模型之间的映射



了解业务环境

任何业务计划都必须清楚地了解当前的业务环境，包括愿景、使命、目的和目标，否则就可能出现偏差。考虑到 AI 目前非常流行且人们对它的期望值通常很高，利益相关者的早期参与尤为重要。确定企业风险和容忍度、业务痛点和预期结果非常重要，因为并不是所有事情都应该自动化。这项工作有助于确定范围、假设和限制因素。本阶段与 AI 相关的任务包括：

1. 制定企业的 AI 愿景、使命、目的和目标。
2. 了解与 AI 相关的业务风险。
3. 确定使用 AI 的产品和服务。

了解数字环境

与前一阶段类似，本阶段涉及数据收集，包括记录组织产品和服务的数字化内容，并确定数字交互的各种关系和用例。本阶段与 AI 相关的任务有：

1. 确定企业当前使用的 AI 产品和服务。
2. 确定不同的 AI 利益相关者。
3. 定义 AI 的数字关系类型（如企业对消费者、企业对员工、政府对选民）。
4. 定义 AI 的数字关系媒介。
5. 了解 AI 的数字供应链。
6. 创建 AI 的数字互动用例。

制定数字信任战略

本阶段要求企业根据其业务和数字环境制定数字信任战略。本阶段与 AI 相关的任务有：

1. 记录战略性 AI 数字信任目标。
2. 规划 AI 系统和子系统。
3. 制定初步的人工智能战略。
4. 创建 AI 商业案例。

规划和实施数字信任

这一阶段是人工智能实施工作的执行阶段。要想取得成功，企业必须在前三个阶段的基础上，规划 AI 数字信任计划，然后加以实施。这一阶段的任务包括：

1. 创建迭代方法。
2. 制定人工智能方案/项目计划。
3. 为项目计划的每个步骤指定 AI 负责人。
4. 实施 AI。

监控、衡量和改进

最后，第五阶段是创建持续模型的重要环节。本阶段包括持续监控、衡量和改进，是回归到第一阶段的纽带，触发持续模型。在某些 AI 应用（如机器学习）中，持续监控是关键，因为 AI 输出会根据模型中添加的新数据集不断变化。本阶段与 AI 相关的任务有：

1. 确定关键 AI 数字信任的衡量标准。
2. 设定 AI 衡量目标。
3. 收集、监控和响应衡量结果。
4. 评估 AI 的能力和成熟度。
5. 不断改善 AI 的数字信任环境。

治理和监督

所有这些阶段的基础都是治理和监督，它可确保包括 AI 在内的数字信任计划得到适当的范围界定、执行和长期改进。这一阶段的任务包括：

1. 确定并采用 AI 治理。
2. 创建并管理 AI 风险登记册。
3. 将 AI 数字信任纳入企业治理、风险与合规（GRC）结构。

DTEF 的应用

与传统软件系统不同，AI 系统带来了独特的挑战，因此需要专门的治理框架。随着 AI 系统以前所未有的速度不断发展，其对企业的影响也变得越来越大。AI 使用的快速增长需要监督和结构化的方法来减少危害。企业宗旨和基础设施的差异巨大，再加上每个应用场景的独特性，要在 DTEF 中涵盖可以帮助企业实现组织价值最大化同时最大限度降低 AI 风险的所有可能方式是不切实际的。

AI 使用的快速增长需要监督和结构化的方法来减少危害。

如今，大量 AI 都是作为现有软件的集成或通过应用编程接口 (API) 采用的。理想情况下，包含 AI 功能的企业软件更新应使用既定的评估和审批流程启动审查，以便对附带的使用条款进行业务审查，并确定业务风险与回报。值得注意的是，禁用 AI 功能的难易程度因产品而异，而且并不总是很直观。[14]

孤立的供应商管理和采购流程无法应对当今企业数字生态系统的复杂性。行业数据描绘了一幅不祥的图景：公司对第三方风险的管理不足[15]，API 攻击事件增多[16]，超过 40% 的公司对使用中的所有 API 缺乏了解。[17] 认识到受供应链漏洞影响的人中有一半以上归因于供应链故障，[18]企业必须更加重视 API 安全，以解决普遍的弱点。[19] 本文的其余部分将探讨一个常见的使用案例和代表性数量的 DTEF 组件，但这些并不能涵盖所有的可能性。

用例：客户服务聊天机器人

根据 Zendesk 的一项调查，生成式 AI 的兴起归因于 70% 的企业客户体验领导者重新评估客户体验。[20] 同一调查还揭示了三个主题的十大趋势，所有这些趋势都涉及 AI，需要任何主要学科的数字信任从业者提高认识、参与和监督。这些主题如**图 8** 所示。

14 Kaelin, M.; “如何通过注册表文件或组策略编辑器禁用 Windows 11 Copilot ”，TechRepublic，2023 年 10 月 20 日，https:// www. techrepublic. com/article/how-to-disable-copilot- windows -11/; Salesforce，“用功能管理器启用和禁用数据云 AI 和测试版功能”，https: //help.salesforce.com/s/articleView?id=release-notes.cdp_rn_2024_winter_feature_manager.htm&release=246&type=5

15 Bolton, R.; “第三方网络安全风险管理--不断变化的风险环境中的更新”，Community Banking Connections, https:// www.communitybankingconnections. org/ Articles/ 2023/I2-I3/ third -party- cybersecurity

16 SALT Labs, “2023 年第一季度 API 安全状态”，SALT, https://content.salt. security/rs/352-UXR-417 /images/ SaltSecurity-Report-State_of_API_Security.pdf; Matson, K.; “下一次大型 API 安全漏洞即将来临：如何准备”，SC Media，2023 年 10 月 19 日，https://www.scmagazine.com/perspective/ the-next-big-api-security-breach-looms-heres -how-to- prepare

17 Nagaraj, S.; “2023 年 API 安全状态”，InfoWorld，2023 年 11 月 2 日，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 安全项目”，https://owasp.org/www-project-api-security/

20 ZenDesk, “2024 年客户体验趋势”，https://cxtrends.zendesk.com/reports/cx-trends-report

图 8：2024 年客户体验趋势

	主题		
	AI 和智能体验	数据和可信体验	下一代和沉浸式体验
趋势	1. 生成式人工智能加速提升人工智能的人性化。 2. 聊天机器人的能力增强。 3. 客户体验领导者在 AI 战略、工具和角色影响方面的脱节日益加剧。 4. 人工智能的透明度和决策从例外变为常态。	1. 企业关注数据驱动的动态用户体验。 2. 客户体验的领导者成为数据隐私的主要利益相关者。 3. 安全设计常态化。	1. 现场体验影响未来的在线购物。 2. 语音关注处理复杂和/或升级的问题。 3. 预测性代理管理工具超越传统方法。

资料来源：改编自 Zendesk，“2024 年客户体验趋势”，https://cxtrends.zendesk.com/

客户体验领域对生成式人工智能的预期很高，因此半数以上的客户体验领导者正在探索人工智能供应商也就不足为奇了。客户体验领域的从业者对聊天机器人将继续转变为具有更强大功能的数字客服寄予厚望。在此期间，企业最好组建跨职能部门的企业团队来管理合规问题，并将与现有数字生态系统集成相关的风险降至最低。

要从这些预期过渡到聊天机器人的实际应用，了解其分类非常重要。聊天机器人通常分为简单型、智能型和混合型。直到最近，聊天机器人还主要是基于规则的，它能提供一致、可靠的体验，但现在聊天机器人正通过越来越多地利用自然语言处理技术（NLP）转向人工智能。除了基于规则，机器人还可以是基于关键字的、基于菜单的、智能的（上下文）、混合的或支持语音的。[21]总之，聊天机器人可以提高客户服务水平，克服语言障碍，并尝试消除因通话时间过长而产生的任何挫败感。DTEF 已经为应对战略、供应商选择和管理、实施和持续改进方面的挑战做好了准备。

DTEF 如何提供帮助？

无论目前正在使用或考虑使用哪种类型的人工智能，企业都必须制定与企业战略和成果相一致的人工智能战略。具体战略和开发周期会因考虑使用的人工智能类型不同而有很大差异。

以下内容包括一些（但并非全部）特定领域内适用的信任因素、实践和活动，可为人工智能聊天机器人示例提供有用的指导。

[21] Shenoy, A.; “六种类型的聊天机器人 – 如何为你的企业挑选最佳的聊天机器人？”，Yellow.ai，2024 年 1 月 9 日，https://yellow.ai/blog/types-of-chatbots/

文化

人工智能会给人带来恐惧、不确定性和疑虑，尤其是关于人工智能将在多大程度上影响个人或职业这一点上。我们鼓励企业制定组织战略，该战略应考虑到工作类型的变化，以及这些工作将由谁来完成。

负责组织发展和招聘的管理人员之间的开放对话可以促使员工了解 AI 带来的风险和机会。相反，企业与外部利益相关者之间就 AI 的目的、使用和治理达成明确的共识是有益的。常见问题（FAQ）页面就是一个很好的例子。

文化领域中与人工智能相关的考虑因素包括：

- 确定部署和使用人工智能的目标文化：为什么使用人工智能，如何使用人工智能，以及相对于内部和外部利益相关者的文化“足迹”是什么。
- 消除阻碍人工智能成功开发和整合的文化因素。
- 从更广泛的角度管理目标文化，例如接受程度、采用率、一般知识和技能、常见使用模式等。

在企业考虑采用人工智能时，文化领域的所有三个信任因素都尤为重要：

1. *管理文化（CU.01）*：评估、调整和推广能够促进数字信任的组织和人类文化。

   实践编号 CU.01.02： 修改文化，包括传达公司战略和价值观的活动，并将吸取的经验教训和外部要求及预期纳入其中。这一做法强调了自上而下的道德领导力的必要性。

   实践编号 CU.01.03：促进文化，包括对员工进行培训，提高他们的认识，使其了解并坚持数字信任战略。

2. *创建和管理数字信任文化环境（CU.02）*： 定义并建立整个生态系统的数字信任管理系统。

   实践修改文化（CU.02.02.6）：以适当的详细程度向各外部利益相关者传达信息。

3. *管理技能和能力（CU.03）*： 这一信任因素的重点是识别和维护所需人力资源的最佳技能、能力和才干。这些信息对于管理与人工智能相关的风险至关重要，如果关键利益相关者不了解他们在新兴技术环境中确保和维护数字信任方面的作用和责任，就可能导致风险。

人为因素

人为因素领域可以帮助企业预测人员需求。它还可以解决需要考虑和密切监控的客户体验领域。虽然与技术相关的知识、技能和能力差距并非人工智能所独有，但人工智能是最新、最具活力的技术，因此，至少在可预见的未来，需要深思熟虑的人才管理策略来识别和管理提升/技能需求，以确保整体人工智能项目和单个项目的成功。此外，人工智能需要持续监控，以确认模型是否按规定运行。

人为因素中与人工智能相关的考虑因素包括：

- 使人工智能的实施易于理解和使用，利用多种交流载体和多种展示布局来实现人工智能驱动的结果
- 定义支持用户体验和人工智能驱动结果验证的控制措施（鉴别器、专业怀疑等）
- 定义和部署持续改进流程，引入反馈和前馈回路，促进人类与人工智能行为者的互动
- 包括评级和评分机制，使人类行为者能够对人工智能实施赋予信任度

在聊天机器人示例中，实践者可能需要了解：

- 实践 HF.01.04 及其所有相关活动： 评估和管理技术的人为因素能力。
  - 活动 HF.01.05.1：识别并修复与数字信任相关的错误来源和中断问题

架构

成功的人工智能集成可以协调数据、人员、流程和技术。[22]要做到这一点，企业需要一个战略——一个有意识的、深思熟虑的流程，其中涉及选择、妥协以及在必要时说不，以满足业务目标。根据 Open Group 的观点，企业架构（EA）是一种战略工具，用于识别和缩小当前与未来状态之间的差距。良好的企业架构可以让企业将战略转化为执行力。[23]

成功的人工智能整合可以协调数据、人员、流程和技术。

DTEF 架构领域涵盖定义、开发和管理整体企业架构的主题。它包括企业架构模型的业务、数据、应用和技术层的计划、政策和标准等领域。

架构中与人工智能相关的考虑因素包括：

- 将人工智能世界映射到整体信息通信技术（ICT）环境和面向业务的用例中
- 在总体架构中嵌入人工智能行为者，包括第三方和更长的供应链
- 定义并确保运行、控制和管理人工智能行为者所需的资源
- 考虑域文化和涌现，不断为人工智能行为者采用新的或替代用例，以加强架构
- 管理并定期审查使用人工智能的业务和财务案例

在聊天机器人用例中，以下信任因素、实践和活动非常有用：

1. *管理数字信任资源（AR.03）*： 该信任要素要求识别、管理和控制确保数字信任所需的资源。这对于管理基础架构的所有元素和组件是必要的，以下实践和活动也与此相关：
    - 实践 AR.03.02： 管理数字信任应用程序。
        - 活动 AR.03.02.1：管理面向客户的数字信任应用程序（如前端）。
    - 实践 AR.03.05：管理数字信任操作： 管理和控制数字信任及相关架构的所有运营和生产环节。
        - 活动 AR.03.05.1： 估算开发、获取或交付与数字信任相关的运营计划所需的工作和资源的规模、工作量、持续时间和成本
2. *根据组织需求调整数字信任技术（AR.04）*：该信任要素涉及组织需求的识别和根据这些需求调整技术。
    - 实践 AR.04.01：根据业务需求调整技术。
        - 活动 AR.04.01.1： 确定相关业务目标。

指导和监控

与人工智能相关的潜在风险[24]主要被描绘成一个未来的担忧。但是，积极应对人工智能风险有助于确保在企业内部安全、负责任地部署人工智能技术。

---

22 Strickrodt, D.;“企业架构和 AI 集成的未来。拥抱协同。”，Bizzdesign，2023 年 10 月 27 日，https://bizzdesign.com/ blog/the-future-of-enterprise- architecture-and- ai- integration/

23 The Open Group,“使用 TOGAF ADM 开发企业架构的实践者方法”，https://pubs.opengroup.org/togaf- standard/adm-practitioners/ adm-practitioners_3. html

24 ISACA，“AI 革命的承诺与风险”，2023 年 9 月 12 日，https://www.isaca.org/resources/white-papers/2023/the-promise-and-peril- of-the-ai-revolution

尽管认识到未来风险的重要性，但人工智能技术的现状和当前的使用模式也带来了必须解决的直接风险。简而言之，人工智能可能会加剧任何现有问题，如缺乏质量控制或数据完整性差，使系统容易受到网络攻击，并可能引入新的问题。

尽管认识到未来风险的重要性，但人工智能技术的现状和当前的使用模式也带来了必须解决的直接风险。简而言之，人工智能可能会加剧任何现有问题，如缺乏质量控制或数据完整性差，使系统容易受到网络攻击，并可能引入新的问题。

指导和监控域包含与创建、衡量、管理和治理数字信任生态系统有关的主题，其中包括风险、通信、信息、可持续性和韧性、合规、鉴证和企业的整体发展。

指导和监控中与人工智能相关的考虑因素包括：

- 将人工智能嵌入到 IT 中更大的 GRC 框架中
- 将治理、风险管理、合规和保证等要素递归应用到组织内部（和外部）的人工智能环境中
- 组织人力和组织结构，以引导和控制人工智能行为者（包括其从入驻到退役的生命周期）
- 将人工智能嵌入三道防线

虽然有许多相关的信任因素、实践和活动都适用于聊天机器人示例，但我们想到了以下几点：

1. 管理风险（DM.03）： 该信任因素涉及持续识别、评估和降低与数字生态系统相关的风险，并将其控制在企业执行管理层设定的可承受范围内。
    - 实践 DM.03.01：指导和监控风险管理。
        - 活动 DM.03.01.1： 确定角色和职责。
        - 活动 DM.03.01.2：确定风险偏好和风险容忍度。
    - 实践 DM.03.02：确定数字生态系统风险。
        - 活动 DM.03.02.2：确定风险所有者。
        - 活动 DM.03.02.3：识别当前的风险控制/控制环境。
        - 活动 DM.03.02.6：将数字生态系统风险纳入更大的企业风险管理（ERM）。
2. *管理组织（DM.04）*：这一信任要素要求企业确定和组织支持数字信任生态系统的结构。
    - 实践 DM.04.01：管理组织结构。
        - 活动 DM.04.01.4：确立角色和职责，包括在适当情况下，建立一个由行政、业务和信息与技术管理人员组成的信任指导委员会（或同等机构），以跟踪项目状态、解决资源冲突、监控服务水平和服务改进。
3. *管理数字信任（DM.06）*：这一信任要素要求组织对信息保持适当的数字信任做法。
    - 实践 DM.06.01：清点信息资产。
        - 活动 DM.06.01.3：发现并清点管理外部关系的合同和其他文件。

涌现

在当今快节奏和数字化的环境中，组织需要了解可能影响未来成功的外部环境因素。考虑到这些因素的组织可以提高其敏捷性和应变能力。从敏捷性的角度来看，它们可以更快速、灵活和果断地行动和适应。从应变能力的角度来看，组织可以预测、应对和适应变化或干扰。

在当今快节奏和数字化的环境中，组织需要了解可能影响未来成功的外部环境因素。

涌现域侧重于可能在流程和人员层面引发机遇的事件和活动，包括内部变化、外部影响和人员驱动的偏差。

与人工智能相关的考虑因素包括：

- 通过生成式和非生成式人工智能预测新出现的情况（例如，定义人工智能行为者的预期边界）
- 分析并验证过程和结果模型（例如，人工智能行为者应该实现什么目标，可能出现哪些偏差[注意可合理预见的滥用]）
- 控制人工智能行为者的输入变量、训练和演化
- 监控人工智能行为者的突发行为，并在必要时进行调整，以保持可接受的信任度

在聊天机器人用例中，以下信任因素和做法非常有用：

1. *识别、评估和管理潜在触发因素（EM.01）*：该信任因素要求识别、评估和管理潜在触发因素。
    - 实践 EM.01.01：识别和管理内部信号及其所有相关活动。
2. *检测和管理与流程和人员有关的涌现（EM.02）*： 这一信任要素要求企业通过识别即将发生或已经完成的变革，并对结果进行管理。
    - 实践 EM.02.01： 检测内部变革及其所有相关活动。

赋能和支持

赋能和支持域是动态的相互联系，技术通过这种联系赋能流程，而流程又反过来支持技术的部署和运行。采用“赋能和支持”域中的实践有助于在聊天机器人部署到生产环境之前发现问题。

赋能和支持中与人工智能相关的考虑因素包括：

- 将人工智能嵌入服务价值链和服务管理中
- 定义和描述作为流程、服务和整体服务组合一部分的人工智能角色
- 指导和控制人工智能的进一步发展（考虑提供商/用户的过渡）
- 持续监控人工智能的运行，并与“涌现”和“人为因素”建立联系

在聊天机器人用例中，以下信任因素、实践和活动非常有用：

1. *管理数字信任生态系统目标（ES.01）*：该信任因素涉及确定流程和技术目标以及实现预期，包括质量考虑因素。
    - 实践 ES.01.01： 确定流程目标及其所有相关活动。
    - 实践 ES.01.03：确定流程规范及其所有相关活动。
2. *实施服务和解决方案（ES.05）*：该信任要素要求组织计划、协调和实施服务与解决方案。
    - 实践 ES.05.01：计划实施及其所有相关活动。
3. *监控服务和解决方案（ES.07）*： 这一信任要素要求持续监控服务、解决方案和相关技术的运行，以确保数字信任。
    - 实践 ES.07.01： 监控流程运行。
        - 活动 ES.07.01.1： 管理流程操作的变更。
        - 活动 ES.07.01.2：监控运营指标和控制。

- 活动 ES.07.01.4：监控流程调整。
- 活动 ES.07.01.7：确定改进项目并将其纳入质量管理过程。

结论

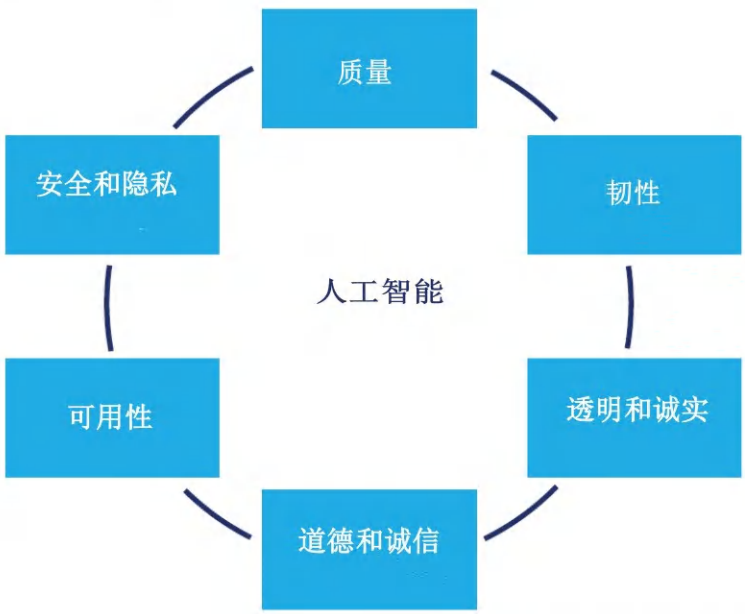
人工智能并非新生事物，但最近的发展--尤其是生成式人工智能的兴起--将市场推向了超速运转的状态。至少有些员工已经使用了某种人工智能，并可能将知识产权置于风险之中，企业接受这一现实是明智之举。与此同时，各种类型的软件都在采用人工智能，这就要求企业在部署更新之前进行尽职调查。此外，随着解决方案提供商竞相保持竞争力，供应商市场也在不断升温。

现实情况是，人工智能已经到来，尽管存在各种风险，但考虑到其诸多优势，无论行业或业务功能如何，阻止人工智能的发展是不现实的。立法和标准将成为建立护栏的主要驱动力，以实现合乎道德、负责任地使用技术，而前者可能会遵循与隐私法规相同的路线和轨迹--复杂的非统一法律网络肯定会让 GRC 专业人员头疼不已。

问题不再是企业是否会采用人工智能技术，而是采用多少。即使企业不开发自己的私有模型，人工智能也是无处不在的，未明确许可的流行生成式人工智能产品的使用代表了影子 IT 现象的演变。企业需要与业务目标相一致的人工智能战略，并且必须验证人工智能实例是否用于解决业务问题，是否在可接受的风险承受范围内。

要实现这一目标，企业需要有一个强大的框架，以帮助确保符合人工智能相关的法律和监管要求、国际标准指南以及客户合同义务。虽然业界有许多针对不同实践领域的框架，但 DTEF 框架可帮助企业在其人工智能流程和系统中实现质量、弹性、透明和诚实、道德和诚信、可用性、安全和隐私（**图 9**）。

图 9：可信人工智能的要素



DTEF 还具有足够的灵活性，可与其他框架结合使用。采用 DTEF 来实施基于人工智能的技术，意味着所有这些方面都将在整个生命周期中得到考虑，而且还有助于打破流程、人员和技术之间可能出现的组织孤岛。虽然人工智能是在组织中设计、开发和部署的，但 DTEF 提供的好处是，如果能有效实施，它应能帮助组织证明其符合法规、客户要求、国际标准、组织政策和程序，但最重要的是，解决方案能以合乎道德、负责任的方式实现业务目标。

致谢

ISACA 谨向以下人员表示诚挚谢意：

首席开发人员

Chetan Anand

CDPSE. CPISI. ICBIS, ICCP

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

专家评审

Joyce Chua

CISA. CISM. CDPSE. CAEG (Professional), (C)CISO, CFE. CIA. CIMR CIPM. CIPP(A), CIPP(E), FIR ITIL. MCR 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. CISSR FBCI

Partner, FORFA Consulting AG, Switzerland

董事会成员

John De Santis, Chair

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

Brennan R 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, CISSR FIR 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

关于 ISACA

ISACA® （www.isaca.org）是一个推动个人和组织追求数字信任的全球社区。50 多年来，ISACA 为个人和企业提供了知识、证书、教育、培训和社区，推动职业进阶，为组织带来改变，并建立一个更可信、更道德的数字世界。ISACA 是一个全球性的专业协会和学习组织，拥有来自信息安全、治理、鉴证、风险、隐私和质量等数字信任领域工作的超过 18 万名成员，在 188 个国家设有分支机构，包括全球 225 个分会。 ISACA 通过基金会支持资源不足和代表性不足的人群获得 IT 教育、职业进阶。

ISACA 中国办公室成立于 2017 年，是 ISACA 在美国以外建立的第一个直属机构，旨在服务 ISACA 在中国大陆的持证人员以及 IT 和安全行业的专业人士，引进 ISACA 全球先进的标准、框架体系和知识，并向全球同行输出中国业界的最佳实践。

免责声明

ISACA 设计并开展了《利用数字信任生态系统框架实现可信的人工智能》调查（以下简称“作品”），主要用作专业人员的学习资料。ISACA 无法保证使用本作品就一定能够实现成功的结果。本作品不应被视为包含所有适用的信息、程序和测试，不排除在其它信息、程序和测试的合理指导下获得同样结果的可能。在确定任何具体信息、程序或测试的适宜性时，专业人员应就具体的情况（特定的系统或信息技术环境）做出自己专业性的判断。

ISACA 中国办公室

地址：北京市东城区隆福寺街 95 号隆福文创园 1 号楼 WeWork 3 层

电邮：ISACAChina@isaca.org

客服：support.isaca.org　18515911939 （同微信）

保留权利

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

电话：+1.847.660.5505

传真：+1.847.253.1755

支持：support.isaca.org

网站：www.isaca.org

推特：

www.x.com/ISACANews

领英：

www.linkedin.com/company/isaca

脸书：

www.facebook.com/ISACAGIobal

Instagram:

www.instagram.com/isacanews/