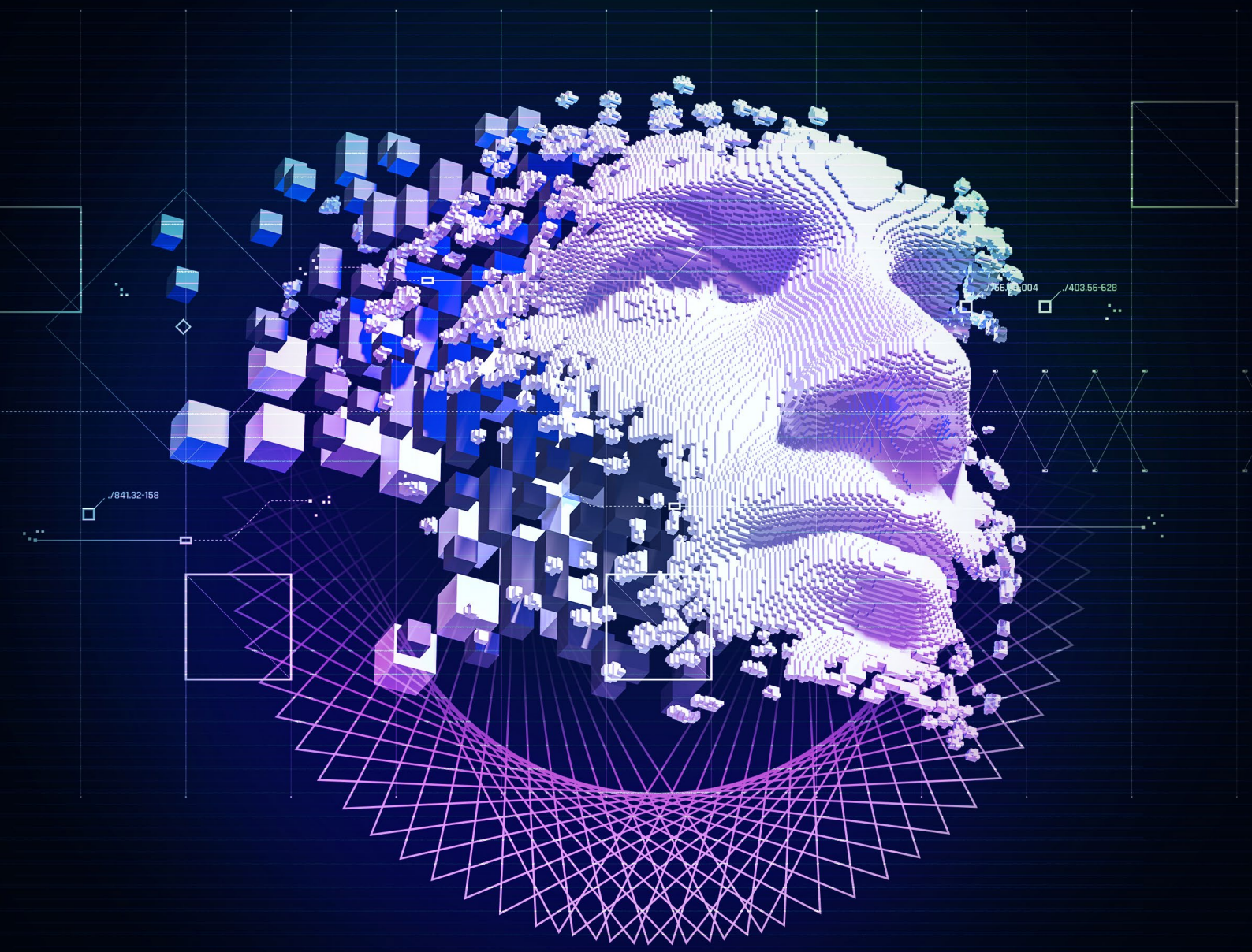


Centre for
Regulatory
Strategy

亚太地区生成式人工智能应用与监管

引言

过去一年，大语言模型 (LLM) 和自然语言处理模型等人工智能 (AI) 技术的发展取得了重大突破。这些技术已经通过OpenAI的ChatGPT、微软Bing AI Chat和谷歌Bard AI等工具得到广泛传播，并引起全球消费者的热议、追捧和警惕。

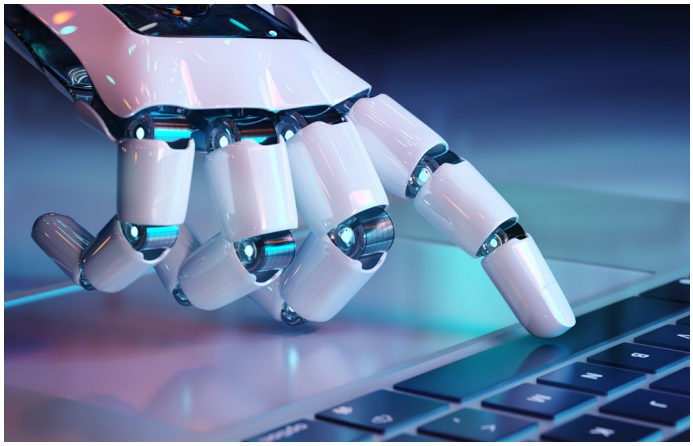
AI平台对广大用户的可触达性突显了AI技术在各行各业（包括金融服务业）的应用潜力。许多企业开始利用AI技术提高自身竞争优势。监管机构和立法机构需更加迅速、敏捷、主动地应对AI应用所带来的相关风险。

在德勤2022年发布《人工智能在金融服务业的可靠应用》报告时，亚太地区许多监管机构仍处于商讨和/或实施AI原则的起步阶段。随着AI工具在金融服务业得到应用和普及，部分立法和监管机构已经开始研究AI应用的相关风险，以保障消费者权益。在本篇后续内容中，我们将进一步探讨金融服务业使用AI的相关风险、亚太地区监管现状以及金融机构在准备应对即将出台的相关法律法规时的考量因素。

第一部分：传统人工智能与生成式人工智能

知识更新：了解传统AI与生成式AI

传统AI是指可以自动处理预定义输入的系统。此类AI系统能够从训练数据中获取知识，并利用这些知识做出决策或预测。例如，许多企业利用AI聊天机器人提供精简高效的客户支持。传统AI聊天机器人在处理常见问题方面尤其有效。凭借内部搭建的知识库，其可针对常见问题提供准确一致的回复并进行用户意图预测。



生成式AI可以编写文本、生成代码、制作音频和图像，其水平与人类不相上下，甚至超越人类。例如，生成式AI工具包括可用于生成书面文本（如营销文案、软件代码等）和图像等内容的LLM。生成式AI模型具有生成连贯文本和超逼真图像的能力，其可采用以前只能通过人类的思维、努力和创造力才能实现的方式生成数据。

传统AI和生成式AI的不同功能驱动了不同用例。就金融服务业而言，传统AI可以用于开展预测分析或检测可疑交易，而生成式AI可以加速完成从交易和研究到通过生成相关报告为合规职能提供关键支持等任务，本报告将对此作进一步阐述。

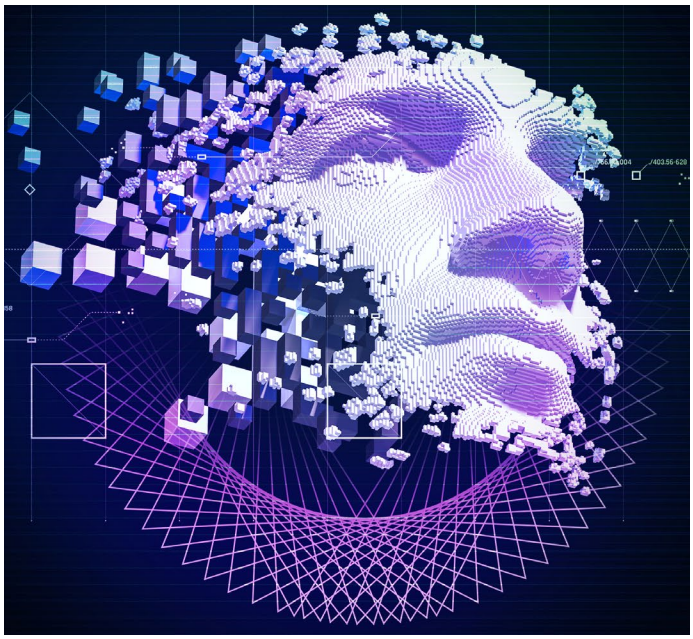


图1：传统AI与生成式AI

传统AI是指根据预定义指令或策略执行特定任务的系统。

- 预设规则
- 训练算法
- 预测和模型分析

特点

生成式AI是一种能够根据用户提示创建新内容的人工智能。

- 文本生成
- 图像生成
- 计算机代码生成

示例

第二部分：生成式人工智能相关风险

在2022年发布的《人工智能在金融服务业的可靠应用》报告中，探讨了亚太地区监管机构希望通过AI监管原则解决的常见风险要素：透明度、问责制、公平性、稳健性、隐私和数据安全。目前此类风险和担忧依然存在，而生成式AI的兴起又给市场带来了新的风险：

- **缺乏透明度**：考虑到生成式AI模型的复杂性及其所涉信息的专有性，人们普遍认为生成式AI缺乏透明度。此外，在衡量或评估生成式AI模型的透明度方面缺乏标准化的工具和方法，这可能导致在比较不同模型和追踪长期进展时变得困难。
- **歧视和偏见**：生成式AI可能会将一些偏见与训练数据中的模式形成关联，从而生成歧视性或误导性内容。
- **缺乏准确性和产生错误观念**：生成式AI可能会利用不完整、不准确或有偏见的数据生成不准确或有误导性内容，或者干脆生成虚构事实。生成式AI模型没有固有的“客观真理(objective truth)”，可能会生成错误甚至有害的内容和观点。
- **知识产权和版权问题**：生成式AI模型可能会以受版权保护的材料为基础进行训练，从而生成与受版权保护的材料非常相似的内容。生成式AI模型还可能用于制造假冒或盗版商品，侵犯知识产权。
- **欺诈**：生成式AI可能生成深度伪造和合成数据，这些数据可以用于实施欺诈、传播错误信息或造成系统漏洞。

第三部分：亚太地区人工智能监管措施

生成式AI的出现迫使亚太地区政策制定机构和监管机构重新评估之前实施的AI框架是否同样适用于降低新兴技术风险。某些监管机构已经实施AI指引和计划，为企业和行业提供最佳实践建议。下表 (图2) 列举了亚太司法管辖区在开展AI监管或为AI风险管理提供建议方面所采取的措施，包括制定AI原则、提供指导和工具、出台立法以及将AI应用纳入国家战略：

- **AI原则：**AI原则为有效管理与各行业使用AI相关风险提供了指引。例如，欧盟以AI原则为入手点开展AI监管以及出台立法。值得注意的是，某些选择针对AI风险出台立法或开展监管的司法管辖区也推出了AI原则。举例而言，中国大陆在对AI应用进行立法的同时，国家新一代人工智能治理专业委员会发布了《新一代人工智能治理原则——发展负责任的人工智能》。
- **指导和工具：**指导和工具通常用于支持AI原则的实施。以新加坡为例，由新加坡金融管理局领导的Veritas联盟发布了五份白皮书，阐述了公平、道德、负责和透明 (FEAT) 原则的评估方法。为推动金融机构加快采用FEAT方法和原则，联盟开发了Veritas Toolkit 2.0版。与1.0版相比，2.0版改进了公平原则评估方法，并纳入了道德、负责和透明原则评估方法。2022年5月，资讯通信媒体发展局和个人数据保护委员会推出全球首个AI治理测试框架和工具包——A.I. Verify，适用于旨在以客观和可验证的方式展示负责任的AI的企业。
- **立法：**韩国、中国大陆、菲律宾和越南等司法管辖区采取了针对保险业出台AI专项立法的措施，其中中国大陆和越南已通过AI专项立法。
- **国家战略：**泰国、印度尼西亚、日本、中国大陆和马来西亚等许多亚太司法管辖区已将AI确定为战略重点，并制定了促进可信AI应用的国家战略，但是某些司法管辖区尚未在实施战略或向业界提供结构化框架方面取得进展。

图2：监管机构和立法机构为应对AI相关风险而采取的措施

澳大利亚——澳大利亚的《人工智能伦理框架》(Artificial Intelligence Ethics Framework)[1]是指导负责任的AI解决方案的自愿原则。该等原则旨在确保AI对人类有益且用于预期目的，以及AI系统负责人对系统的影响负责。

中国香港——2021年，个人资料私隐专员公署 (PCPD) 发布了《开发及使用人工智能道德标准指引》[2]，旨在帮助组织理解和遵守《个人资料 (私隐) 条例》(第486章) (《私例》) 下相关的个人资料隐私保护要求，以便在开发和使用人工智能时予以遵循。该指引的内容包括人工智能的数据管理价值和伦理原则，并提供人工智能战略治理实践指导，以帮助组织制定适当的人工智能战略和管理模式，进行风险评估，并制定相关的监督安排等。

日本——经济产业省的《实施人工智能原则的治理指南》(Governance Guidelines for AI Principles) 对AI应用的潜在影响进行了考量，并就最大程度减少负面影响提供了指导。[3]该文件是2019年发布的《以人为中心的人工智能社会原则》(Social Principles of Human Centric AI) 的延伸。相关原则包括以人为中心，教育应用，隐私保护，安全保障，公平竞争，公正、问责和透明以及创新。

中国台湾——2023年8月，台湾金融监督管理委员会基于台湾AI行动计划2.0和全球AI指引起草《金融业运用AI核心原则与相关推动政策》草案，以指导金融机构使用AI。该草案详细阐述了六项原则，包括治理、以人为本的价值观、隐私保护、系统安全、透明度和可持续发展。[4]

新加坡——由新加坡金融管理局领导的Veritas联盟发布了五份白皮书，阐述了公平、道德、负责和透明 (FEAT) 原则的评估方法。[5]为推动金融机构加快采用FEAT方法和原则，联盟开发了Veritas Toolkit 2.0版。与1.0版相比，2.0版改进了公平原则评估方法，并纳入了道德、负责和透明原则评估方法。

中国大陆——中国国家互联网信息办公室于2023年发布了《生成式人工智能服务管理暂行办法》[6]，旨在促进生成式人工智能的健康发展和规范应用，保障国家安全和社会公共利益，同时保护公民、法人及其他组织的合法权益。

菲律宾——菲律宾正在寻求通过立法[7]，成立“人工智能发展管理局”，负责制定国家AI战略和框架，指导企业在菲律宾开发和部署AI技术。

韩国——2023年2月，韩国国会通过《促进人工智能产业和建立可信人工智能框架法案》(Act on Fostering the AI Industry and Establishing a Foundation for Trustworthy AI)，这是韩国首部全面规范AI应用的立法，[8]主要内容包括禁止任何人在未经政府批准的情况下开发AI以及对被认为足以影响人类生活的“高风险”AI进行分类。

越南——2022年6月，新《保险业务法》(Law on Insurance Business)[9]获得通过，允许在保险业务活动中使用技术。政府鼓励保险公司使用包括AI在内的技术来销售创新型保险产品和服务。

印度尼西亚——《国家人工智能战略》(National Strategy for AI)[10]以政府支持的“印尼制造4.0”(Making Indonesia 4.0) 计划为基础，该计划旨在推动印尼社会各部门实现自动化，其中提出针对AI、机器人技术和印尼科技企业进行投资，同时吸引来自日本、中国和韩国的领先科技企业进行投资。

马来西亚——科技创新部发布《2021-2025年国家人工智能路线图》(2021-2025 National AI Roadmap)[11]，阐明国家促进AI发展的六项战略以及负责任AI的七项原则。

泰国——泰国目前暂无AI和机器学习专项法律。[12]不过，泰国政府正在积极制定国家AI战略，其中或将包含更详细的AI监管规定。

就其他地区而言，欧盟和美国等司法管辖区也已着手采取措施来应对生成式AI的快速发展。欧盟《人工智能法案》(Artificial Intelligence Act) 是欧盟委员会为规范欧盟AI系统而提出的立法，并已纳入欧盟确保以负责任的方式开发和使用该技术的整体战略。《人工智能法案》旨在建立基于风险的框架，以应对AI相关风险，同时促进创新和提高竞争力。相比之下，美国采取的AI监管措施较为分散。美国的法律和监管结构以州为基础，在联邦层面尚未颁布或提出规范生成式AI的法律。某些州 (包括加利福尼亚州和科罗拉多州) 已着手推进AI立法，而某些州则在监测不断变化的风险。

图3：其他监管辖区为应对AI相关风险而采取的立法和监管措施

欧盟

- *《人工智能法案》*已于2023年底通过成为法律。[13]
- 《人工智能法案》旨在整合欧盟所有国家的AI相关法律，制定AI整体立法，并将其框架推广为全球基准。
- AI风险评级分为四个等级：
  - **禁用AI系统**：因“侵犯基本权利”而被明确禁用的系统，例如公共场所的实时生物识别系统。此外，《人工智能法案》严禁任何蓄意利用成人或儿童的弱点造成伤害的系统。
  - **高风险AI系统**：用于或属于受欧盟产品安全立法约束的产品或《人工智能法案》附件III所列产品的系统。高风险AI系统受《人工智能法案》全面监管并须满足多项要求。
  - **有限风险AI系统**：可与人类直接交互的系统。此类系统必须满足透明度要求并告知用户他们正在使用AI。
  - **低风险或最小风险AI系统**：简单的AI系统，例如垃圾邮件过滤器或AI视频游戏。此类系统不受任何限制。

美国

- 美国联邦、各州和各行业的AI监管法规各不相同。
- 科罗拉多州等州针对AI在保险业的应用制定的专项法律已于2023年生效。
- 目前仍有20多个州未制定AI相关法律法规。
- 2022年10月，白宫科技政策办公室发布《人工智能权利法案蓝图》（Blueprint for an AI Bill of Rights）[14]，其中提出五项原则：
  - 安全有效的系统
  - 算法歧视保护
  - 数据隐私
  - 通知和解释
  - 替代方案

该蓝图并非法规，无法强制执行，其目的是对白宫（科技政策办公室）认为存在问题的AI用途提供指导。

挑战和考虑因素

生成式AI的快速发展给亚太地区监管机构带来了新的挑战，具体法规的制定和实施往往被认为没有效果，且可能很快过时。监管机构在应对生成式AI给金融服务业带来的新挑战和新风险方面也面临困难。

由于人才短缺以及公共和私营部门为吸引具备适当AI技术能力的人才而展开的激烈竞争，导致某些监管机构无法灵活应对AI技术带来的新兴和不断变化的风险和发展趋势。立法机构和监管机构在监督和执行相关指令方面也举步维艰。例如，就AI定义达成共识是实施AI法律法规的关键问题。

虽然AI应用可能因利益相关方而异，但就AI原则达成共识对于利用AI改善金融服务且不影响安全性、公平性或消费者保护至关重要。归根结底，有效的监管应当推动创新，同时保障金融生态系统中所有相关方的利益。

行业和监管机构可以加强区域合作，建立跨境治理框架，开展联合研究，确定最佳实践，以实现监管标准化。此类联盟有助于确保监管措施始终依据实际见解应对全球挑战及平衡行业增长和社会保障。

第四部分：高可信人工智能框架的应用

相比传统AI，生成式AI可能会对使用AI应用的金融机构提出更具挑战性的风险管理要求。由于大多数司法管辖区仍处于制定或实施AI法律法规的起步阶段，金融机构必须尽早建立自有AI治理框架并将全球/区域AI原则纳入其中。

金融机构应当利用该框架系统管理与使用生成式AI相关的风险。这对确保AI监管合规、加强用户保护以及进一步推动AI应用的成功实施至关重要。《人工智能在金融服务业的可靠应用》报告中，我们简要介绍了德勤高可信人工智能框架。在本节中，我们将探讨在不同用例中如何使用德勤高可信人工智能框架管理潜在的AI相关风险。

图4：德勤高可信人工智能框架[16]