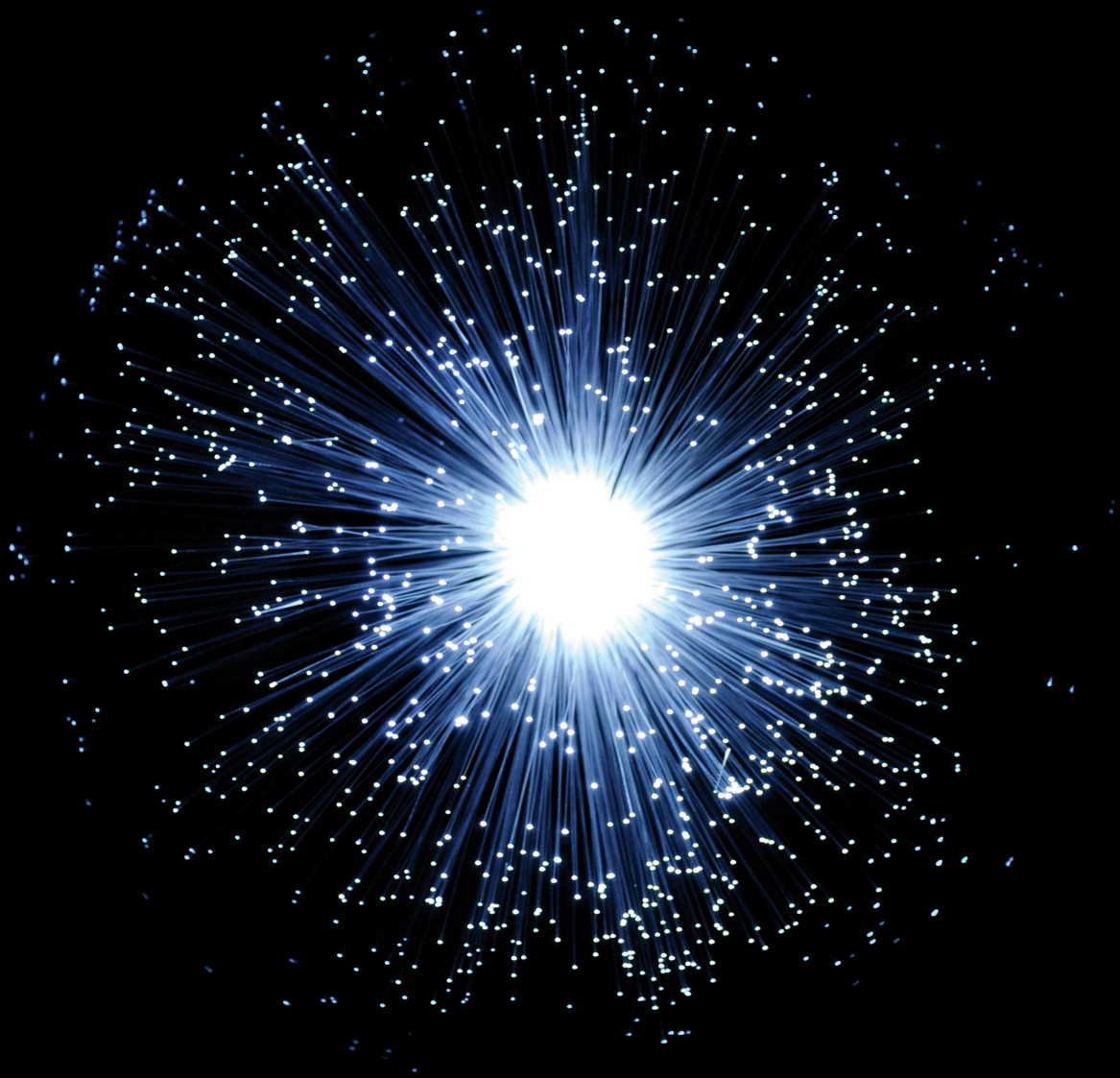
Deloitte.



人工智能与风险管理报告

创新与信心偕行

CENTRE *for*
REGULATORY
STRATEGY
EMEA

目录

1. 概要

说起人工智能（AI），虽然它并不是一个全新的概念，但在近些年金融服务企业开始逐渐意识到它的巨大潜力。

AI可以提高运营效率、降低成本，同时还有助于企业实现战略转型，更多、更好地融入用户参与。。然而，一些客观限制条件，都在或多或少地阻碍着金融服务企业大范围的推广使用AI技术，包括数据量和数据质量上的局限、对于AI潜在风险的认知不足以及公司文化和现行规章制度的限制。

欧盟以及其他国际组织同样对于AI技术有着极大的兴趣。尽管他们认识到AI可以为金融市场、客户以及他们的内部工作带来诸多益处，但也认识到那些受监管公司当采用AI时可能会产生的潜在风险以及意料之外的后果。

近些年来，值得注意的是，金融服务业因对客户和市场的失当行为，从而受到众多经济上及其他形式上的制裁。由此产生的对于公平对待客户和保证市场诚信问题，以及AI技术在监管领域中未经测试与未经检验的特性的关注，都意味着金融服务企业对于采用AI解决方案应该时刻保持谨慎。

“有效的风险管理已经不再是抑制创新的因素，而是公司成功使用AI技术的关键。”

为了能解决这些困难，从而实现充分利用AI技术的优势，并且避免未来可能出现的问题，董事会和高级管理人员必须充分理解这项技术，包括它在企业中现有或潜在的用途，并从风险的角度出发来仔细考虑AI的影响。在这种环境下，有效的风险管理已经不再是抑制创新的因素，而是公司成功使用AI技术的关键。

“我们认为企业面临的重大挑战不是在处理全新类型的风险，而更多应该是关注那些难以用有效且及时的方式去分辨的风险，或它们已经以不同以往的出现方式而显现的风险。”

关于刚刚提到的后一点，这将是本文之后所讨论的重点。我们认为企业面临的重大挑战不是在处理全新类型的风险，而更多应该是关注那些难以用有效且及时的方式去分辨的风险，或它们已经以不同以往的出现方式而显现的风险。在这篇文章中，我们将讨论企业应该如何重新审核并调整现有的风险管理框架，以反映在当部署复杂AI应用时企业需要关注的一些重要差异。

除《企业家第一课》、《企业家功成堂》外，其他公众号分享本期资料的，均属于抄袭！

邀请各位读者朋友尊重劳动成果，关注搜索正版号：《企业家第一课》、《企业家功成堂》

谢谢观看！

企业家第一课，专注做最纯粹的知识共享平台

关注官方微信
获取更多干货

加入知识共享平台
一次付费 一年干货

比如说，AI可以不断的从新数据中学习，并通过建立复杂统计模型，而后得出结论。但这种结果并不是基于那些明确且预先定义好的规则，这就让企业难以理解这其中支撑最终结论的决策机制。在许多方面上，这种挑战与面对管理人力资源时所面临的挑战非常相似。然而，不断发展的AI技术使其可审计性以及可溯源性变得异常困难，并且这种技术发展的速度可能会导致在极短的时间内产生大规模的错误。

企业需要重新审核并且更新他们的风险管理方式，从而在不同的风险管理框架全生命周期（识别，评估，管控，监督）阶段中管理风险。不断发展的AI技术要求这些风险管理方式必须在更短更频繁的时间间隔内得以应用。现有的风险偏好声明也同样需要重新审核，并且需要添加一系列新的内容，比如说公平性原则等，以便为风险管理框架的各个阶段提供信息支持。

“不断发展的AI技术使其可审计性以及可溯源性变得异常困难，并且这种技术发展的速度可能会导致在极短的时间内产生大规模的错误。”

在本篇文章中，我们将会用简单的理论性的风险管理框架来描绘出一些AI所带来的挑战。例如，一个保险公司如何在产品定价过程中，使用AI来做到风险模型管理。最后，我们将总结当监管者监管AI时会遇到的挑战与选择。

本篇文章旨在成为一个理解AI对于现有风险管理方式和更为广泛的监管环境的影响的起步点。通过强调这些需要关注的领域，我们希望帮助企业能够在制定AI政策时，或者更具体来讲，在制定AI风险管理框架的时候，为它们提供更加高效的解决方案和监管机制。

2. AI的简要概述

人工智能的概念最早可以追溯到20世纪50年代，那时候的研究人员开始考虑使用机器模拟人类智能的可能性。然而，AI技术却是在20世纪后期才得以真正蓬勃发展。当时有几个技术因素发展达到顶峰：强大且价格低廉的计算资源、数据总量和种类的增加、访问数据速度的提高以及能够凭借最新且先进的算法来以更加“智能”的方式分析数据。

对于AI技术其实并没有一个单一解释，但从广义的层面来讲，AI是一种计算机系统的理论与发展方式，它能够执行通常需要人类智能才能操作的任务。这一类的任务包括有视觉感知，语音识别，以及在不确定性下的决策和学习。

关于缺乏对于AI定义的共识的原因，可能是因为AI并不是技术，而是集合了模仿人类行为的各种技术这一事实。其中，一些与金融服务企业有关并且将会在本文中提到的关键技术有：

机器学习
Machine Learning

只需直接提供数据而无须遵循明确的程序指令，提高了计算机系统运行的性能。机器学习的核心就是自动发现数据中隐藏的模式并使用它们来进行预测的这样一个过程。

深度学习
Deep Learning

深度学习算法是一系列的机器学习算法。由于它们在与语音和计算机视觉相关的任务中表现的十分有效而变得越来越受欢迎。但这是一种很复杂的技术，大家难以准确地解释每个输入在最后是如何驱动模型结果的，所以通常导致它们被定性为“黑匣子”。

语音识别和自然语言处理
Speech Recognition and Natural Language Processing

拥有能够以人类的方式来理解并自动生成出人类语言的能力。比如说从文本中提取语义信息，或者生成出语义自然、语法正确的可读文本。

视觉识别
Visual Recognition

拥有能够识别图像中的对象，场景和活动的能力。计算机视觉技术使用成像处理中操作和技术的序列来将分析图像的任务拆解成许多可管理的部分。

提高金融服务企业的客户体验

“行为情感分析工具（BEAT）”是德勤的语音分析平台，它使用了深度学习技术和各种机器学习算法来监测并分析语音交互。这其中有三个最核心的功能：

1. 监测客户语音交互
2. 通过自然语言处理 (NLP) 来识别高风险交互
3. 将语音交互映射到潜在的负面结果上（比如投诉或操作行为上的问题），并且提供出现这种问题的可能原因。

BEAT会去分析顾客所说的词语和他们所用的语气，然后使用机器学习技术来不间断的开发并且加强这些分析语音交互的分析能力。而且当可学习的素材越多的时候，最终的风险评估准确度就越高。BEAT的使用，相较于传统解决方案，可以让企业从评估准确率的角度看到显著的提升。

BEAT已经发展到可以分析超过30种不同语言以及30种不同的行为表征。它可以为特定的风险要求和用户需求来进行定制服务。

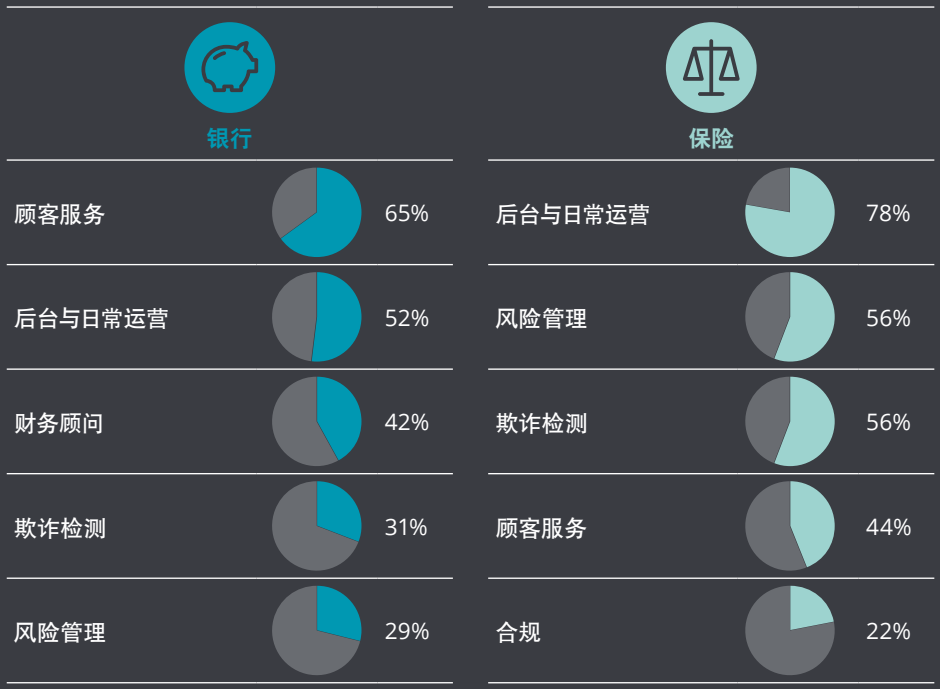
3. 当金融服务企业大规模采用AI技术时可能会遇到的挑战

自从2008年的金融海啸之后，金融服务企业一直致力于提高自身成本效益，并在利润压力下保持竞争力。为实现这一目标，他们所关注的其中一个领域就是技术，并在过去几年中开始更多地去使用AI技术。然而，采用AI技术的方式并不是唯一的，有很多原因导致出现了这种情况。

关于AI应该应用于何处的不同观点

德勤最近对与欧洲财务管理协会（EFMA）合作的3,000多名企业高管调查后显示，AI能对其公司产生最大影响的活动和功能因行业而异。

图1：您认为您公司所开发的人工智能技术在哪一部分的价值链上产生了最大的影响？



总的来说，从这份调查我们可以看到，在金融服务企业中采用AI技术还仍处于起步阶段。在接受调查的公司中，有40%的企业还仍在学习如何在他们的公司中部署AI，11%的企业还没有开展任何这方面的活动。只有32%的企业已经开始参与开发AI解决方案。

数据数量与质量

AI与传统技术解决方案的一个最重要的区别是，后者常常在那些需要用预先设定清楚的规则框架完成任务。然而，AI应用可以自行分析数据来进行模式识别，并以此为基础来制定决策。此外，AI还可根据给出的数据进行一次性或不间断性的学习来完善决策制定机制。

这意味着每一条AI系统所给出的建议都对于所处理的数据的数量和质量十分依赖。总的来说，AI解决方案所遇到的最棘手的困难就是缺少大量高质量数据。对于金融服务企业，因为受到普遍使用的传统系统和组织架构的限制，阻碍了数据之间的无缝流动，并在某种程度上影响了数据质量，导致问题愈发加剧。

透明性，可审计性与合规性

一些AI解决方案拥有许多有关决策制定的隐藏层，都影响着最后生成的结果。在一些复杂AI应用案例中，就比如那些使用了深度学习技术的AI应用，可能在：维护系统，论证其中需要理解的部分，并且对于那些以AI为基础生成的决策拥有控制权等方面上阻碍了金融服务企业的使用。具体来讲，应判断其决策是否包含有适用性、公允性，是否与公司价值观和风险偏好相符合等。

> “那些不断学习，不断进化，并且拥有许多有关决策制定的隐藏层的AI解决方案会使可审计性和可溯源性变得异常困难”

这与公司在面临人力资源上的挑战无异。然而，可以说，那些不断学习，不断进化，并且拥有许多有关决策制定的隐藏层的AI解决方案会使可审计性和可溯源性变得更加困难。此外，AI自我学习及进化的速度可能还会导致在极短的时间内产生大规模的错误。

一些AI解决方案的不透明性还会与现有的法律法规产生冲突，就比如说欧盟的“通用数据保护法案”（GDPR），它规定了在特定情形下企业需要能够向客户解释他们的个人数据是如何被使用的，并且能够给出那些假设性结论的合理解释，和那些会对顾客会产生重大影响且完全自动生成的结论的内部机制。

理解AI及其隐藏含义

AI是一个既复杂但又快速发展的领域，在一些非专业人士的眼中，它被认为是一项难以控制的技术。此外，AI的使用会加剧企业的现有风险，改变风险出现的方式，甚至会为公司带来新的风险。

金融服务企业属于严格管控产业，这其中包含了复杂多样的经营范围和产品，并且当公司处理业务时必须使用严格的审查原则。可以说历史上对于金融服务企业因违规而产生的行政处罚，造成了这些企业在采用相对不了解的技术以用于受监管的业务时非常保守，这也就为创新增加了另一重阻碍。

因为对于这项技术和其所带来的风险，相对而言，不熟悉也不了解，这就造成了过度谨慎的现象。那些重要的利益相关方，比如风险部门、合规部门，以及各业务部门主管，董事会成员和公司高层，除非他们对于这项技术充分了解，否则就可能会在批准使用AI上迟疑，并且对企业内受监管业务AI技术的使用保留解释权。要理解这项技术，就要不止了解它会带来的风险，同时也要知道这些风险可以如何降低，管理并监控。

如何让利益相关方有一个对于AI独立且整体的认知是对公司而言非常困难的。可以使用一些实际的用户案例，使利益相关方知道相关的客户经历可以帮助了解AI能提供的潜在益处，同时也可以认识到哪些方面会出错，以及如何有效的规避或管理风险。

对于人才的影响

对于采用AI的公司，尤其是那些大规模的企业，必须充分了解这种转变对其企业文化和人才战略的影响，并采取必要措施来应对任何不利影响。

企业很可能需要额外的良好技术资源来帮助设计、测试和管理AI应用程序。目前这方面人才的稀缺，以及众多金融服务企业在创新上所面临的困境都使得AI的应用充满挑战性。因此金融服务企业需要升级它们的招聘方式和渠道，技术人员需要选择有利于职业发展的职业道路，同时需要在发展过程中制定继续留任、融合新技能或者是改变职业路径的战略。

AI对现有企业员工工作模式影响可能更为深远。随着AI的技术发展和应用的深入，一些原有人工处理的任务可以借助模式识别等技术自动化地完成，因此相关的劳动力总需求将不断减少。同时，就业方式也会有重大变化，例如减少人员配备需求，或将现有员工重新分配到不同的工作中（这其中可能会存在有相关工作的再培训的问题）。这些改变都可能会影响员工的工作动力，如果不及时解决，可能会导致不必要的员工流失。

如果AI应用程序实施失败或必须在短时间内关停，但这时已经发生的那些过度人员流失，可能会使公司无法保留住那些必备的技术能力和那些能够人工执行流程的技术人员，甚至对公司未来业务发展产生较大影响。

4. 在风险管理体系中嵌入AI

AI的应用和一般的创新过程一样，都要求企业要经历一次不断学习的过程。然而，这样的过程并不是要规避所有与AI相关的风险，而是要去开发工作流程和处理工具，从而让企业相信这些风险，可以在整体的公司的风险观和偏好框架所规定的范围内有效地被识别和管理。因此，尽管存在着一些常见的误解，但一个行之有效的风险管理体系在企业的创新能力发展中起着关键性作用。

AI应用程序的固有风险性质

我们认为，管理AI所产生的挑战并不在于处理这种全新的风险类型，而是要考虑到当我们把AI解决方案的复杂性和发展速度纳入考量，或者当以我们不熟悉的方式出现情况时，我们就难以采用有效和及时的方式来识别这些风险。因此，企业不需要全新的流程来处理AI，但他们需要改进现有流程，把AI因素纳入考量，并填补一些必要的管理空白。同时还需要去解决对所需资源水平以及岗位角色和责任可能造成的影响。

“企业不需要全新的流程来处理AI，但他们需要改进现有流程，把AI因素纳入评估，并填补一些必要的管理空白。”

德勤AI风险管理框架提供了一种识别并管理AI相关风险和管控的机制。在下一页和以下各节中提供的表格中，我们从涵盖了60多个AI风险的总表中列出了一些关键性的评估因素。这些评估因素都会用一般术语来表达，但实际上，风险等级和那些必要控制因素在不同案例下不同的组织下会出现很大差异。

科学的思维方式

采用和推进人工智能需要一个组织和在其中工作的人接纳更科学的思维方式。这意味着需要能够接受最终产品的试错过程意，接受风险和一些最终结果证明是失败的测试工作，并通过引入外部冲击或数据来观察结果，以不断测试产品的可行性。从本质上讲，它意味着在整个组织中创建一个“沙箱”（代表在商业情况下的受控隔离环境）。这种心理上的转变不仅仅适用于业务负责人或部门，而且与组织中的所有领域都相关，包括董事会和其他职能部门，如风险部、合规部、人力资源部和IT部门等。

这其中，需要所有三道防线（业务线，风险/合规和内部审计）都参与进来的这一点尤为重要。作为合规和监控的守护者，充分参与沙箱可以使他们能够了解一些关键的技术信息，并从一开始就帮助形成适合的AI的风险管理政策。

企业风险类别	子类别示例	AI解决方案独特的关键风险因素示例
模型	算法风险—偏差性	• 因为依赖于不断发展的数据集来驱动AI产生决策，这使得识别模型中的固有偏差变得更加困难。 • 输入数据中的固有偏差可能导致运行效率低下或不公允的结果出现。 • 数据科学家缺乏对于偏见性的考虑，使得偏差风险从一开始就注定无法得到充分解决。
	算法风险—不准确性	• 算法类型选择不正确、数据质量不佳或算法参数选用不合理。
	算法风险—反馈	• 未检测到不当反馈的风险增加（尤其在那些允许持续反馈和学习的AI解决方案中），这可能会影响解决方案产生准确结果的能力。
	算法风险—滥用性	• 商业用户可能缺乏对复杂AI模型的充分理解，或错误地解释AI输出结果从而导致出现错误结果的可能性增加。
技术	信息与网络安全	• 当开发者不再支持、更新或免费提供开源组件（软件包，编程语言，API等），企业对其组件的依赖性可能会引入安全漏洞。 • 复杂算法使得人们更难理解AI解决方案是如何做出的决策，从而这可能会受到人类或其他机器的恶意操纵。
	管理层更迭	• 难以识别那些为AI解决方案提供信息的上游系统发生变化的影响，这可能会导致在AI与其外部环境交互时产生无法预料的后果。
	IT运营	• 在某些情况下，AI应用程序对大数据的显著依赖性增加了现有IT基础架构所带来的风险，因为后者可能与AI应用不兼容（例如，现有系统无法处理大数据）。
合规	数据保护	• 由于AI解决方案的不断进步和不透明的特质，这可能会与数据保护法案（例如GDPR）相关的合规风险增加，其中包括，在自动决策生成领域中的数据主体权利。
	合规性	• 管理层很难理解并向监管机构证明复杂的AI应用程序是如何做出这项决策的，例如那些采用神经网络的应用程序，其中包含了许多类似黑匣子的隐藏决策层。
行为	文化	• 由于考虑到实际或可能来自的监管方面和道德方面的问题，大规模的采用AI技术可能会出现文化挑战。 • 担心组织内职位变化从而产生负面影响。
	产品创新	• 已开发的产品不能满足客户需求的风险（即为了使用AI而使用AI），以及可能的大规模不当销售而产生的风险。
人才	岗位与职责	• 在AI全生命周期中，可能无法明确定义职位，职责和责任，同时，利益相关者（合规部，业务部门，IT部及编程人员等）缺乏持续参与和监控可能会增加出错的风险。
	招聘与技术	• 缺乏对正在采用的AI解决方案的理解、使用的经验或适当的监控技能都会增加风险。 • 由于组织内缺乏对于精通AI的人力资源的整合而产生的新风险。 • 过度依赖少数具备AI知识的人才和专家。
市场		• 对于少量的大型第三方AI供应商的过度依赖，会增加过度集中风险，并且，如果当其中一个实体破产或遭受重大运营损失的时候，可能会产生连锁反应。 • 如果算法对某些变量（例如股票市场价格）过于敏感，则由于羊群效应而导致的系统性风险增加（即众多组织与其他市场参与者行为相同）。
供应方		• 黑箱算法的使用可能导致供应商，运营商和AI用户在发生问题时责任分配不明确。 • 同时，黑箱算法会增加第三方AI供应商失败的风险，特别是在那些较小的新公司可能没有足够的管理措施和内部控制的经验下。

风险偏好

公司的风险偏好是公司在任何时候为实现其目标而准备接受的风险量。为了建立有效的风险管理流程和控制系统，任何采用AI的策略都需要从一开始就与整体风险偏好保持一致。

同样，我们也需要重新审视公司的风险偏好，并纳入AI的特定考虑因素。尽管AI的引入不会改变公司整体的风险偏好，但是会影响判别风险的因素，以及会影响衡量和管理风险的工具。

AI解决方案本身就可以增加或减少某些特定类型的风险（例如模型风险），并改变公司当前和未来的风险概况。这就意味着需要在每种风险类型的层面都需要重新考量风险偏好。这其中不仅包含目标风险级别，还有那些能有效支持、管理和监控该风险的政策和管理信息。

当公司需要评估AI使用对其风险偏好的影响时，他们首先应制定一套清晰一致的评估标准。例如："这个AI解决方案是否将面向外部？"，回答这个问题有助于确定AI使用案例中可能会涉及到的风险类型。制定这一套标准问题可以帮助企业了解，无论是单从AI使用的层面还是从整体而言，哪些风险领域需要或多或少的关注。

风险管理框架（RMF）的全生命周期

尽管具体细节和所用术语可能因公司而异，但从概念上讲，风险管理框架全生命周期包含四个关键阶段：

1. **识别**
   通过确定哪些风险可能对公司的业务战略或运营会产生重大不利影响，来了解风险环境。此阶段还涉及监控内外部运营和监管环境，从而确定固有风险格局的变化，并确保框架持续适用于所需的目的。

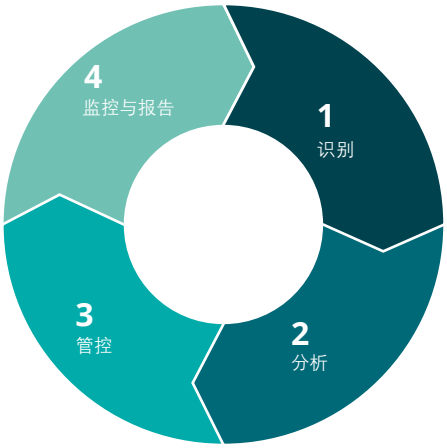
2. **分析**
   定义并建立算法风险评估流程，以评估风险暴露程度。

3. **管控**
   搭建算法风险管控框架，通过控制降低固有风险，使其与风险偏好水平保持一致。

4. **监控与报告**
   设计一个可以评估管控效果的有效方法，其中包括测量其有效性，容差阈值和控制检测等相关指标。

   向相关管理层报告剩余风险概况，控制环境和补救计划的状态。

在下面的章节中，我们将会为风险管理框架的每个阶段提出一些关键的AI考虑因素，并通过实例来说明公司应该如何管理因使用AI解决方案。



1. 识别

金融服务企业中AI的复杂性和相对不成熟性意味着某些风险表现出来的方式及其程度可能会随着时间的推移而发展，在某些情况下可能还会非常迅速。从行为和稳定度的角度来看，这可能对公司产生重大影响（例如大规模的不当销售）。

因此，企业需要定期进行重新评估，以确定AI应用的风险情况自引入以来是否已经发生变化，因为该模型已经学习了新数据并已经进化。

同样，作为概念证明或仅供内部使用而开发的AI解决方案，如果其使用范围扩大，则也需要重新评估。例如，如果公司计划扩展最初开发的AI解决方案的适用范围从而为外部客户提供建议，但此方案最开始时所设计的唯一目的仅为提供内部建议时，则需要了解这些新客户在使用过程中可能产生的风险。

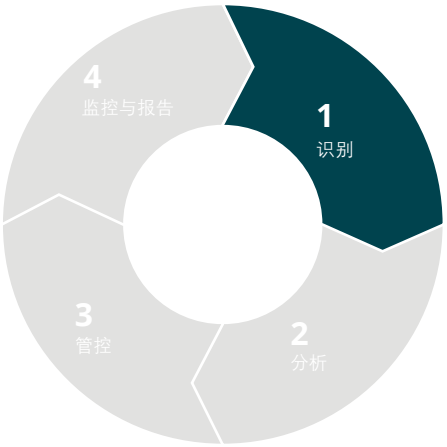
“金融服务企业中AI的复杂性和相对不成熟性意味着某些风险表现出来的方式及其程度可能会随着时间的推移而发展，在某些情况下可能还会非常迅速。”

值得注意的是，AI所对应的自身定义及其风险也将发生变化。例如，随着时间的推移，移动电话的定义及其功能的扩展，使得与移动电话有关的风险也发生了巨大变化。

企业需要确定如何将AI风险考虑因素整合到现有的风险管理框架中，以及需要其改变的程度。这其中的考虑包括了监管与伦理方面的影响，例如算法偏差，以及AI模型在不建立因果关系的情况下从数据集推断的能力。我们在之后实例中说明了这一点。

但是，一般而言，对于复杂和不断发展的AI应用，企业需要审查其自身的管理方法论，采用一个全面且可持续的方法来定义并识别风险。当识别AI风险时，其中应包括与AI应用有关的特定风险（例如风险分析应用程序）以及因在整个公司内广泛采用AI而引入的风险（例如对员工关系和企业文化的影响。

为了识别AI解决方案产生的风险，同样重要的是要考虑那些，更加广泛意义上的公司层面的影响，及其在短期与长期层面上对公司的人力资本而产生的影响。



1. 识别—实例

- 如上所述，风险分析AI模型产生的主要风险之一是算法偏差和AI模型在不建立因果关系的情况下从数据集推断的能力。
- 例如，AI财产保险定价模型可能会使用各种非结构化数据来评估财产。这些数据可能根据那些仅发生一次的本地事件（例如狂欢节或示威游行）从而将其捕捉到该区域的风险概况中。然而这就带来了许多风险，其中最主要的风险来源于不确定性较强的一些特征，例如，那些用于定价的决策驱动因子。次要风险是，在该地理位置发生的任何一次性事件都有可能会被定性为该位置的永久型风险。
- 此外，相同的数据可能会在未来用于不同的AI模型中，并且在无意中会告诉我们其他人的风险状况。例如，不同的AI模型可能使用相同的评估数据用于评估风险状况，通过标记上述事件中的参与者或旁观者的照片并用他们的社交媒体所呈现的内容来对他们进行评估，从而为他们设计个人汽车或假日保险的产品，而无需他们的授权。
- 在这个例子中所产生的风险是多种多样的，包括数据保护，客户授权和错误定价，更不用说道德考量了。尽管偏差性、模型、声誉和监管风险都不是新型企业风险，但在AI使用案例中，它们可能以全新的或不熟悉的方式表现出来，使它们更加难以被识别。

译者注：

- 又例如，在债券风险评估、银行信贷检查中，对于发债企业、贷款客户信用风险的分析不仅基于企业的基本面信息（如财务状况），企业的舆情信息也是AI模型分析的重点。而舆情分析涉及到金融文本这一类非结构化数据，其来源为众多新闻网站。随着有价值的新闻网站被不断发掘，舆情监控范围也在随之扩大，AI模型在最初训练时的数据源和后续的数据来源将有所不同。而数据源的变化将很可能导致数据特征的迁移，例如不同新闻网站的编辑具有不同写作风格和词语表达，最终可能导致最初的模型逐渐失效。一个比较好做法是对AI模型根据每天或一个时间窗口内的新增数据，进行模型全量或增量优化训练，确保AI模型适用于最新的数据源。
- 部分新闻网站对于通过网络爬虫的方式获取其舆情信息，可能会存在限制的情况。严格遵守各网站的机器人协议进行数据爬取，或进行数据方面的合作才是正确的做法，否则若处理不当可能会引发声誉和法律风险。
- 在这个例子中所产生的风险是多种多样的，包括数据保护和授权、数据特征迁移，其中也涉及到道德风险。尽管算法偏差、声誉和监管风险都不是新型企业风险，但在AI模型的使用情景中，它们可能以全新的或人们所不熟悉的方式表现出来，使它们更加难以被识别。

2. 分析

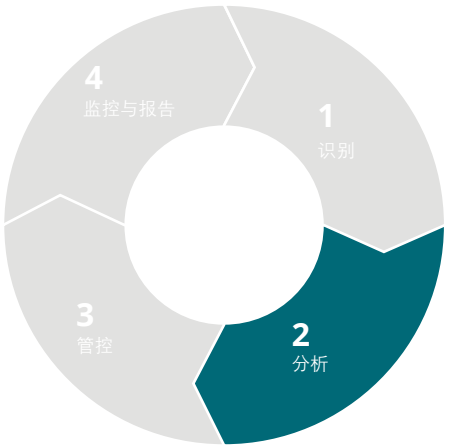
在每个AI应用开发之前，都应该去设计其风险评估过程，并由公司管理层同意。该过程应仔细考虑那些可能会使某一特定应用的风险更高或更低的关键性因素（例如监管、客户、财务或声誉的影响）。例如，那些为客户提供财务建议的AI解决方案，其固有风险水平以及审查过程的程度，会与那些为内部员工提供IT故障排除支持的解决方案不同。

现有风险偏好和评估框架可能不够全面，无法涵盖AI解决方案中的一些特定考虑因素。例如，为了评估AI模型中的偏差，企业首先需要定义诸如“公允性”等概念，以及如何衡量它们。就如同“公平性”等的这种企业价值观，在评估某些风险特性的方面发挥着重要作用，特别是当我们从行为和声誉的角度来看。

此外，由于AI模型可以随着时间的推移而不断进化，企业可能会发现以前的一些定义和评价指标可能无法充分解释或评估模型中的决策驱动因素。因此，评估工作的开展需要变得更加的频繁和动态，同时，我们也需要以“自下而上”（针对每个单独的应用），和“自上而下”（整体风险偏好）的方式来进行不断修正评估流程。

同时，评估工作还需要有更高的参与度，以及更加广泛的利益相关者的赞成，这其中包括AI方面的专家，风控部门（如技术风险和合规性），以及那些业务代表们。

还需要注意的是，AI应用一般使用敏捷开发方法，相比之下，许多技术风险管理框架通常采用传统的瀑布模型。因此，那些为了评估传统技术开发而设立的流程，政策及管理方式都需要改变，变得更加动态。在实际问题中，这可能意味着，至少对于那些高风险应用，在整个开发阶段，风险方面的考量可能会更加频繁出现在日常工作之中。但这，可能会给企业所拥有的现有资源带来压力。



2. 分析—实例

- 在保险产品定价实例中，它的AI解决方案使用了大量结构化和非结构化的数据源来为定价结果提供风险权重。对于我们来说十分重要的工作就是，评估其结果是否与那些只使用了静态和一些可识别的决策驱动因素的非AI系统所产生的结果一致，并了解其中产生任何偏差的基本原理。例如，对于商业房地产定价，一个非AI模型可能仅考虑房产的物理特征及其周围环境，而AI模型可能包含了范围更庞大的影响因素集。
- 同样，如果是通过一系列离散的AI解决方案以模块化的方式来完成定价，即将一个AI系统的结果输入另一个中，每个模块的结果都应由利益相关者进行评估，从而审核并且检测这些决策驱动因素的有效性，特别是当它们与生成定价结果的风险权重之间是没有因果关系的时候。
- 评估时应该包含模型的技术参数（例如偏差度和分类失误率），还应该包括商业数据（例如客户部门出保量）和操作参数（例如保单从开始到生效的速度）。

译者注：

- 在债券风险预警的例子中，AI模型使用了大量动态的结构化和非结构化数据源来为发债企业预警结果提供风险权重，包括发债企业财务数据、经营数据、行业和地区数据、舆情数据、债券交易数据等。对于我们来说十分重要的工作就是，评估其结果是否与那些只使用了静态和一些可识别的决策驱动因素的非AI系统所产生的结果一致，并了解其中产生偏差的基本原理。例如，对于投资人对发债企业的情绪，一个非AI模型可能仅考虑该发行人所发行债券的成交价格，而AI模型可能包含了范围更庞大的影响因素集，例如舆情因素。
- 对于应用AI模型的不同场景，相关业务人员对模型犯错的容忍度是不同的。一些和投资以及风险管理联系紧密的场景中（如智能投顾），人们对于模型的犯错几乎是零容忍的，所以在此类场景中，AI模型的输出结果往往需要经过人工的确认或者该输出结果仅仅是作为人工决策的辅助参考或验证。在另外一些场景中，模型的输出结果可以不经人工确认而直接使用。在评估AI模型前，对于模型适用环境的确认是非常重要的。
- AI模型评估时应该包含对模型自身技术参数（例如预警准确率、预警提前量）的分析，还应该包含对输入数据质量的分析以及模型运行维护环境的分析，其中模型使用和运维人员是否具备足够的专业性也是特别需要关注的问题。

3. 管控

管控和测试过程也需要更加动态。在实际工作中，可能需要定期且频繁地测试并监控这些AI解决方案，可以说，这部分的工作量会远远超出AI方案的最初开发阶段和初始数据集的训练过程。

与传统技术的解决方案相比，这可能会增加所需的测试量。我们应使用基于风险的解决方案来确定每个实际应用其最适合的管控等级，而且也要与公司的整体风险评估框架成比例相关，并与之保持一致。

此外，由于AI的采用将对整个公司产生广泛性影响，与之相关的管控工作可能会横跨多个领域（例如人力资源，技术，运营等）。这进一步强调了在整个风险管理的生命周期中需要广泛的利益相关者来参与其中。

企业可能还需要重新定义业务连续性计划，以便在当系统不可用的情形下或在当AI使用过程中出现管控失败的情况（例如部署“紧急开关”）时，使公司能够重新回跳到当前的已有进程。该算法还应该定期进行压力测试，以分析它在响应严重问题时的反馈，和在遇到非典型性行为时是否会产生正确的处理方式。

管控过程应当考虑AI会如何与利益相关者（客户，承销商）进行交互，以及它们之间的关联点是什么。对于公司来说，当测试从最初参与到AI解决方案产生的结果时，这其中的用户体验尤为重要。公司需要时常进行测试，并在必要时，于早期阶段就纠正那些出现的反常现象和异常值。

同样，公司还应该有一个完善的“交还人工”的流程，这说的就是当AI解决方案负责的工作需要交还给人工来完成的情况，就比如当算法无法在预先定义好的风险容差框架内产生输出的时候。比如，如果该算法无法十分确定产品价格的时候，就应该将该工作交给人类来完成）。

> “实际工作中，可能需要定期且频繁地测试并监控这些AI解决方案，可以说，这部分的工作量会远远超出AI方案的最初开发阶段和初始数据集的训练过程。”

我们还需要使用测试外样本来设计一些关键性能指标，也就是使用全新数据来运行一些测试人员已经知道正确的结果的AI模型。同时应该对算法（包括其中的模型驱动程序）进行频繁且连续的测试和统计分析，以确保当用于生产环境下的AI解决方案使用全新或被更新的数据集时，它的性能可以符合预期和公司的风险偏好。

最后还应该注意到，管理整体模型风险和提高算法透明度的其中一种方法也可以去构建模块化的解决方案，在这其中我们将使用一系列处理能力有限的微型算法来确定最终输出，而不是单一且复杂的模型。这将使得我们更容易理解和控制算法的推理过程和这些决策驱动因素。

3. 管控—实例

- 针对不同AI模型不同的应用场景，管控的措施也有所不同。以债券预警中的舆情分析为例，在该场景中对负面舆情识别的漏报比误报要严重很多。因此在模型效果评估中，应赋予漏报率相较于误报率更高的权重，同时包含模型的技术参数（例如偏差度和分类失误率），还应该包括商业数据（例如客户部门出保量）和操作参数（例如保单从开始到生效的速度）。
- 应该不断训练算法以理解决策驱动因素所产生的不同结果。例如，财产保险定价算法可以通过收集卫星图片所反馈的建筑物裂缝的测量数据，并收集那些有裂缝和无裂缝的建筑物图片来不断训练。
- 一旦在评估过程中发现其结果与非AI定价系统所返回的结果出现任何差异（无论正面还是负面），公司都应该进行人工审查或使用其他模型来进行分析其结果。
- 对于金融机构应用AI模型的场景，若与投资和风险管理决策相关程度较大，则人们往往对模型输出结果的可解释性有更高要求，而这与一些AI模型，如神经网络的“黑箱”这一不可解释性的特征相矛盾。（目前一些深度学习算法通过引入查询和注意力机制，将业务专家在做人工判断时的关注点输入给AI模型学习，一定程度上可实现AI模型输出结果的同时也能有效输出判断依据。）
- 对保险产品定价模型的控制应涵盖算法的有效性，相关性和准确性以及数据等方面：
  - 算法的准确性：如上所述，AI算法结果应与非AI定价系统产生的结果比对从而检查模型性能的准确性。 此外，还应该基于不同的数据源上来测试算法，分析这其中为定价而生成的这些风险权重是否具有一致性。
  - 应该在不同的数据集上训练和测试算法，以确保在模型面对新数据时输出的结果是否还保持有效。可以说有众多不同的方法来实现这个目的，其中一种方式就是划分可用数据集（例如，过去的保险产品价格数据），例如仅在80%的数据上来训练算法，然后用剩余的20%的数据来测试其生成的结果，并确认结果的准确性和公允性。
  - 管控措施应确保算法所使用的训练数据具有良好的准确度，并且保证如果一旦开始不断输入实时数据了，还能够保持其相对稳定的准确度。也就是说，AI算法可以时刻根据其已定义好的预警给出最佳的预警结果。
  - 偏差数据的使用：需要将带有偏差的数据输入算法，从而查看其返回值是否反映出了偏差度。

4. 监控与报告

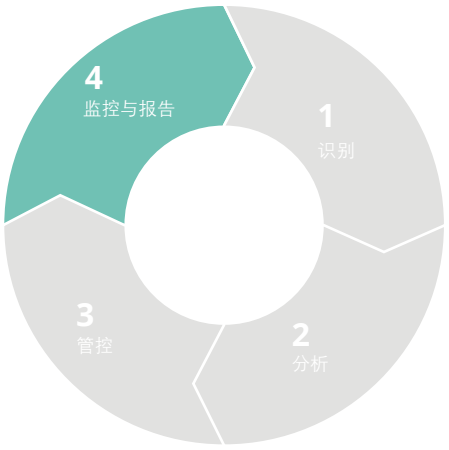
由于算法在不断发展和进化，公司需要采用更加动态的监控方法来确保模型仍然按照其为了特定应用而设计的预期目标来执行。

此外，必须定期监控那些与AI解决方案有关的限制因素和预期目标（例如关键绩效指标—KPI）的适当性，相关性和准确性。

监控和报告工作都应涵盖模型的技术性能指标，以及因该模型而实现的业务和运营成果。

“由于算法在不断发展和进化，公司需要采用更加动态的监控方法来确保模型仍然按照其为了特定应用而设计的预期目标来执行。”

监控工作还应包括关注所有可能涉及到需要更改模型架构的法律和监管措施的更新，同时也要关注可能会影响模型数据、模型结果的外部事件。那些静态技术解决方案也会受到这些改变的影响，只不过，在这些方案中可以相对容易地识别它们对决策驱动因素和结果的影响。在AI解决方案中这些不断变化的决策驱动因素使得我们难以隔离分析，评估并监控这些新的外部事件对决策驱动因素的影响程度。



4. 监控与报告—实例

- 公司应该设定明确且精确的关键绩效指标 (KPI) 来监控其算法。该指标中应当包含有公司的公平和反歧视原则。例如，客户在样本外测试中请求被拒绝的次数就可以作为一个指标，即如果某一特定类型的人群一直被拒绝，则说明算法可能存在某种程度上的不公平性。
- 应根据预定义的绩效指标来评估算法。公司应当评估所使用的算法是否产生了歧视性的结果，以及是否已采取措施来抵消任何歧视性因素的影响。
- 相关工作人员应监控那些可能对保险产品定价模型的架构产生影响的潜在市场或监管方式的变化。例如，在任何保护弱势客户的公平待遇的法规的出台或修订时都需要修改算法设计，以确保不会导致带有歧视性的输出结果。
- 那些认为自己受到歧视性待遇的客户的投诉，应当包含在算法的审核过程之中，如果有必要的话，需要相应地对算法进行更改。
- 模型性能的持续分析应由人来完成：
  - 边界情况分析 (例如，对于那些投保了每次都会拒保的个人和那些投保了每次都会被批准的个人之间的比较分析)。
  - 并且在人工验证后对模型进行反馈修正。
- 应分析那些导入数据的分布状况，以确保为模型提供的数据集本质不会发生一些浅层变化。
- 业务KPI值应包含许多指标，例如AI模型所生成的保费、亏损率、销售成本和总体利润与非AI定价模型所产生的这些项目之间的比较。
- 应适当监控利润和投资组合的状态，以确保某些客户群数量没有由于价格增加或受到不公平对待而明显下降，同时也需要保证来源于某部分客户群的利润也没有大幅增加。
- 运营监督中还应包括一些捕获性和比较性的指标，例如被AI系统推送给人处理的产品交易量，以及相较于非AI系统，当部署AI解决方案时完成投保的速度。

译者注：

- 对于AI模型的表现，应根据预定义的成功指标来评估算法，相关工作人员应当评估所使用的AI模型是否产生了结果偏差，以及是否已采取措施来抵消任何导致偏差的因素影响。
- 在债券预警案例中，相关工作人员应监控那些可能对债券预警模型的架构产生影响的潜在市场或监管方式的变化。例如，在任何保护性或限制性法规 (如支持绿色债券发展) 的出台或修订时都可能需要修改算法设计，以确保不会导致带有偏差的输出结果。在银行信贷模型案例中，也需要重点关注人行、监管机构对相关政策对模型参数的影响。
- 模型性能的持续分析应由人来完成：
  - 边界情况分析 (例如，对于那些发行人资质较好但风险预警级别较高和发行人资质一般但风险预警级别较低的债券之间的比较分析)。
  - 并且在人工验证后对模型进行反馈修正。
- 应分析那些导入数据的分布状况，以确保为模型提供的数据集不会发生本质变化。以发债企业舆情为例，随着时间的迁移，舆情中可能会出现全新的词汇，如央行新推出的公开市场操作；也可能会出现全新的事件类型，如外评AAA的房地产企业发债失败等，这些从未出现的词语以及事件将对现有的模型产生影响。因此在该场景中，需要每日对舆情数据和模型输出进行人工检查和模型更新，确保模型适用于对最新的舆情。

5. 监管者所看重的具体是什么呢?

从国际上、欧盟和英国当局所发表的众多声明和文件上来看，充分了解应用人工智能对这些受监管公司可能造成的影响，已经成为监管机构关注的重点。

一般而言，这些准备采用或已经使用AI技术的公司可以预见到，在未来，监管机构对它们的审查水平只会不断增加。

尽管没有十分完善的AI法律法规，但已经有相关规定，包括算法交易的现行规定和监察描述，以及英国的高级管理人员及认证制度（SM&CR）的规定，还有系统控制方面的众多要求。以上提到的对于AI治理和风险管理的所有规定，可以很好的给出监管机构指导意见。

根据这些文件，以及我们服务客户的经验，可以总结出当公司在采用AI时应该考虑与监管相关的重要原则和措施。这些原则在很大程度上都源于有关算法交易的一些完善的应用。在多大程度上这些考虑因素可以适用于其他AI应用，这将取决于它们特性和复杂度。

“这些准备采用或已经使用AI技术的公司可以预见到，在未来，其监管人员对它们的审查水平只会不断增加。”

治理、监控以及问责制

- 监管机构希望公司能够实行强而有效的治理方式，其中就包括建立风险管理框架，从而可以去识别，减少并控制整个公司内每一个AI应用程序在开发和持续使用阶段所产生的风险。除了风险管理框架应该得到董事会的批准以外，公司还应该能够向他们的监管机构解释每个AI应用程序的工作原理，以及它是如何与有关的监管条例和公司的风险偏好相符合的。

- 由于AI的快速发展，以及公司内采用AI解决方案的比例不断提高，我们应定期审查风险敞口和相关控制措施，以确保它们与公司的风险偏好时刻保持一致。这其中需要考虑诸如公司内AI的使用程度，内部AI的处理能力以及从外部而来的威胁和事件等因素。

- 根据问责制度，其中特别是英国的高级管理人员及认证制度监管机构希望公司有明确的责任制和问责制，其中包括每个AI应用程序都有明确的负责人。根据预先定义完善的测试和批准流程，负责人将全权负责审核和批准AI算法。只要存在任何可能影响其准确性、公平性或合规性的可能因素（例如市场或监管方式的变化），负责人都应负责开始对AI应用程序的重新审核和更新工作。

- AI监管管理委员会中的成员，应该除了接受培训来了解与AI应用程序相关的风险以外，还应该建立测试和批准流程，这其中包括质量保证标准，还同时需要定期审查AI应用程序的性能，以确定是否有任何新问题出现。

- 为了能够反映出AI在整个实体中所具有的深远影响，一个有效的AI治理过程应当更加广泛地涉及来自整个公司的所有利益相关者。尤其在那些关键的开发和测试阶段，除了应该包括AI技术专家以外，还应该有来自第一，第二和第三道防线的相关代表们。

“为了能够反映出AI在整个实体中所具有的深远影响，一个有效的AI治理过程应当更加广泛地包含来自整个公司的所有利益相关者。”

- 公司还应起草与人工操作“紧急开关”或“退出滑槽”有关的步骤和管控细则，以便一旦检测到错误或异常行为时，就立即停止算法运行。企业应该围绕这些管控细则的实施来制定管理流程，其中应当包括有业务连续性计划和补救预案。

- 所有AI算法都应定期重新验证。这些重新验证工作的频率将根据公司、客户或其他市场参与者在算法出现故障时可能面临的风险程度不同而有所差异。这一频率还应考虑到算法随时间而进化的程度，以及决策的关键驱动因素的波动性，例如宏观经济指标。

管控工作中应具备的能力和参与度

- 公司需要确保风险、合规和内部审计团队中的员工具备足够的专业知识，以正确理解所采用的每个AI解决方案的风险。此外，他们应该有足够的权力来与业务主管人员相抗衡，并在必要时施加额外的控制措施，以确保风险管理有效进行。

- 特别是风险和合规方面的工作应充分地融入到新的AI应用程序开发和实施过程中的每个关键阶段，以便能够为建立适合的风控系统提供帮助，确定其是否符合风险偏好，并对任何潜在的行为和监管风险进行独立调查。

- 内部审计工作应确保对AI应用程序和模型的审查是其工作计划中的一部分，并且还应该考虑是否需要持续不断的监控。

文档和审计跟踪

- 公司应清楚全面地了解其公司部署的所有AI应用程序，相关负责人，以及现有的合规性和风险控制工作。

- 应记录测试和批准的过程，其中包括明确说明AI模型在真正实施之前需要满足的条件。

- 同样，监管机构也希望公司能有一个符合审计标准的工作流程来跟踪和管理任何已识别出的问题。

- 最后，还应清楚地记录现有算法的任何变化。企业应该定义什么属于“重大变化”，并确保这些标准在整个企业中能得到一致地应用。任何重大变化都应经过严格且记录在案的测试工作，其程度应与这个变化可能带给公司带来的风险相称。

“受监管公司在任何情况下都不能将应该履行监管义务的责任外包给第三方。”

第三方风险和外包

- 受监管公司在任何情况下都不能将应该履行监管义务的责任外包给第三方。同理，任何由外部供应商设计和提供的AI模型及其相关的风险控制，都应接受那些在部署之前在公司内部就形成的同等严格的测试和监控流程所约束。

- 为了应对第三方供应商开发的AI解决方案停止工作的情况，或者供应商无法提供服务（例如，供应商受到网络攻击时）的情况，公司应设计一套行之有效的业务连续性计划以维持日常运营。因为目前市场上企业AI第三方供应商的数量相对较少，其中普遍还大多是小型的初创企业，所以这一点显得尤为重要。

AI与GDPR

现在企业开始越来越多地使用AI解决方案来设计更适合客户需求的一些定制服务及产品，并且更有效地来确定客户的个人风险状况。

我们能够有效利用这些技术的大前提是我们有大量可用的相关客户数据。随着GDPR的问世，这将会检测公司是否能够在符合数据保护法规的同时，有效地使用客户数据。

GDPR保证了消费者有权理解并且管控公司是如何使用他们的个人数据。只要公司的商业模式是依赖于批量处理客户个人数据这一基础，无论他们是否已经开始使用AI解决方案，都需要在2018年5月之前做好各项准备工作。这意味着，一旦监管程序正式启动，公司就应该满足监管人员的各种规定，更重要的是，公司需要以有意义，透明且易懂的方式来回应客户的质询。

"[……]如果机器产生的决定对个人会产生重大影响，则GDPR允许他们有权质疑该决定并要求公司向他们解释。[…]"

英国信息专员Elizabeth Denham在下议院科学和技术委员会的口头证据，2018年1月

为了应对2018年5月这一GDPR实施截止日期，公司需要制定完善计划，为正在那些处理不断发展变化的客户数据的AI应用程序完成数据隐私影响评估，并在必要时制定补救计划以确保其持续合规。

一般来讲，公司应该采用算法问责制和可审计性原则。这就要求公司能够证明它们拥有企业性的和技术领域上的明确工作流程，同时第三方也能够检查和审核算法是否符合数据保护要求。同样也非常重要的是，公司需要确保用于处理的数据可以合法地使用并且没有任何偏差性。

"[…]作为一个监管机构，我们可能需要在幕后来观察这些公司使用了哪些数据，使用了哪些训练集，哪些因素在系统中被考虑，以及AI系统本身是被训练去回答哪些问题的。"

英国信息专员Elizabeth Denham在下议院科学和技术委员会的口头证据，2018年1月

GDPR需要企业在公司和行业两个层面上与数据保护监管机构之间的关系发生一些变化。这意味着公司需要建立一支更有组织的且资金充足的监管事务团队，与数据保护监管机构定期进行汇报。这其中需要讨论一下它们的数据隐私战略和一些准备实施的任何具有高风险的数据自动处理计划。

6. 监管AI中的一些反思

我们需要认识到不断增加使用AI的影响以及风险不仅是针对金融服务企业的挑战，也是对其监管机构的挑战。后者知道AI技术可以凭借提供更优质的服务和量身定制的产品的这一形式为金融市场提升效率并为消费者带来利益。事实上，监管机构本身其实一直在探索如何在自己的日常工作中应用到AI。

然而，正如我们之前提到的，监管机构也越来越关注受监管公司因使用AI而可能带来的潜在风险和意外后果。从金融稳定性的角度来看，潜在的网络和羊群效应以及网络安全问题是一些需要重点关注的领域。从行为的角度来看，监管机构注意到因为一些不准确的AI模型可能造成无法公平对待每一位客户（即大数据“杀熟”），或出现不当销售的情况，客户缺乏对数据处理方式的理解，金融排斥现象（译者注：指社会中的某些群体没有能力进入金融体系，没有能力以恰当的形式获得必要的金融服务。）的增加，以及产生对弱势消费者的负面结果等问题。

与公司一样，大多数风险对监管机构来说并不陌生。监管机构在AI方面面临的挑战，或者更普遍来讲，在创新技术方面面临的挑战，需要在支持良性创新和竞争、保护客户和市场诚信，以及稳定财务之间找到一个适当的平衡点。

由于新技术发展和被采用的速度与新法规的制定和实施的速度之间的不匹配性，使得找到这种平衡变得特别困难。例如，当时二号欧盟金融工具市场指导（MiFID II）于2011年首次被提出，旨在解决金融市场中算法交易日益增多的问题。但是，二号欧盟金融工具市场指导是在七年后，即2018年1月才实施。

监管机构已经意识到这种滞后性，并且在历史上已经通过采用“技术中立”这一原则解决了这个问题，也即，无论他们那些受监管活动所使用的技术具体是哪一类，可以使用类似的监管原则适用于其上。技术中立这一监管措施确实有助于降低法规迅速过时的这一风险，但也可能阻碍监管机构应对个别特殊技术和应用产生的风险的能力。

然而，我们可以看到一些迹象表明，如果某些特定技术的使用对于全系统会成为或有可能在未来成为十分重要的一部分，监管机构就会开始准备着手摆脱其技术中立的立场。有关算法交易的二号欧盟金融工具市场指导就是一个很好的例子。

我们也越来越多地看到监管机构开始发布一些详细的技术指导，使他们在许多领域上对公司的一些要求变得更加清晰，其中包括智能投顾、外包云平台，以及最近的算法交易。

“技术中立这一监管措施确实有助于降低法规迅速过时的这一风险，但也可能阻碍监管机构应对个别特殊技术和应用产生的风险的能力。”

就AI而言，我们认为监管指导是一种强有力的工具，可以帮助企业了解监管机构对其风险管理方法的一些要求。这同时也使得管理主体和高级管理层不光可以更有信心地推进其创新计划，也能帮助他们找出一些亟需解决的问题。更多针对AI的指导方案也将有助于监管者自身的日常工作，因为这些方案使监控工作更加具有统一性，也同时提高他们发现行业内合规缺陷和剩余风险上的能力。

但我们面临的挑战就是，要真正有效地向企业提供有关监管机构是如何希望企业来遵守现有监管制度这一问题上充足的信息。同时，挑战也包括，任何AI的指导方案都需要具体问题具体分析，而非套用普遍的情况。尽管最近审慎监管局(Prudential Regulation Authority)所给出的一些关于算法交易的草案声明中提到的原则与AI应用程序更相关，但它们的真正目的是在于它们对于处理不同算法交易活动上所表现出的特殊性。鉴于AI使用案例的广度及其复杂性，监管机构需要使用基于风险程度的方法，来仔细选出在哪一具体方面上来集中他们有限的资源去监管。监管领域的沙箱（TechSprints）和行业圆桌会议都将继续成为监管机构能够有效做到这些目标的重要方式。

监管者的另一个措施就是去定义那些要解决的问题，随后要求行业内部自行来制定与之相符的AI标准和行为准则。这类似于英国的市场竞争管理局在调查零售银行后所做的事情。当时它要求英国九大银行自行制定开放应用程序接口的标准细则。这种方法还得到了来自英国信息专员办公室（ICO）的信息专员（UK Information Commissioner）的支持。他最近解释说，在AI和数据保护这一领域下，行业自行开发特定的行为准则，而随后经过相关监管机构的认证这种方式是值得推行的。

当然，对于AI的监管不仅仅是对金融服务领域的挑战，同时也不能被地理边界所框限。监管机构需要不断克服不同国家和部门之间所存在的界限约束，并广泛地与对手方进行合作，除了不仅要制定有效应对新型风险的政策，还要更广泛地着手解决公共政策和道德领域上出现的问题。

“监管者的另一个措施就是去定义那些要解决的问题，随后要求行业内部自行来制定与之相符的AI标准和行为准则。”

7. 结语

AI会越来越多地成为众多金融服务公司企业战略的核心组成部分，从而提供更好的客户服务，提高运营效率，并赢得竞争优势。

然而总的来说，在金融服务公司中采用AI这一技术尚属初期。这些企业仍在不断摸索这一技术，并发掘，当基于它们各自的商业模式，产品和服务时，哪些可应用AI的工作可以为他们带来最大的价值。

这整个学习过程中一个最重要部分就是要从风险的角度出发来理解AI所蕴含的含义。对于金融服务行业的严格监管，这不仅是业务需要，也是监管机构希望看到的。

对于公司而言，重要的是他们要认识到这是一个双向学习的过程。董事会，高级经理队伍以及业务和管控方面的职能部门需要增加对AI的理解，而AI专家们如果还没有对于风险和监管方面的理解，那他们需要增强对这方面的认识，并一定会从中有所收获。能够认识到这种跨领域合作的重要性并激励员工以这种方式进行协作的金融服务公司将能够最大限度地利用AI所带来的优势。

这种“合作关系”会使企业认识到，AI所带来的一些已经非常普遍的风险类别（例如偏差性）可能表现出的方式与以往不同，或这些风险出现的速度以及强度方面也会出现与往常不同的情况。这意味着，在采用AI技术时，企业需要仔细考虑如何将那些与AI相关的特定考虑因素整合到现有的风险管理框架中，以确保这些框架时刻与最初设定的目的相吻合，并且能够让公司相信在企业文化和风险偏好的框架内，AI相关的风险能够被有效地识别和管理。

监管机构也越来越关注在金融服务企业中采用AI技术的潜在风险和可能出现的意外后果，以及如何在支持良性创新和竞争以及保护客户，市场诚信和财务稳定性之间找到一个平衡点。其中可能采用的应对措施可能包括在某些情况下脱离其当前所一直保持的技术中立地位，或要求全行业与监管机构进行合作，从而制定针对特定应用情况的AI技术标准和行为准则。

我们还应该认识到，管理AI技术这一任务并非只是应该落在金融服务企业这一方身上，而是，全行业连同监管机构都必须共同努力，参与到那些有关于AI技术产生的大范围且长期社会和道德影响的讨论之中，同时研究应对这些情况可以采用的政策应该有哪些。

“对于公司而言，重要的是他们要认识到这是一个双向学习的过程。董事会，高级经理队伍以及业务和管控方面的职能部门需要增加对AI的理解，而AI专家们如果还没有对于风险和监管方面的理解，那他们需要增强对这方面的认识，并一定会从中有所收获。”

后记：中国AI风险管理和监管现状

从2017年国务院发布《关于印发新一代人工智能发展规划的通知》对中国人工智能发展给予指导性规划以来，发展人工智能已经成为国家重要战略，推动互联网、大数据、人工智能和实体经济深度融合被写入十九大报告，人工智能应用也已遍及智能制造、智能医疗、智能金融、智能电网、智能安防和智能家居等各个领域。

在应用人工智能技术驱动行业创新、创造商业价值的过程中，国内企业对于人工智能相关风险管理认识普遍不足、重视程度亟待提高。人工智能风险控制不到位造成的企业声誉影响、经济损失、监管问责等事件屡见不鲜：如国内部分互联网新媒体由于人工算法的信息推送导向不正、格调低俗等问题被监管机构行政处罚甚至整改关停；国内多家电子商务企业由于疑似利用人工智能技术开展大数据杀熟引发大量用户投诉和舆论批评；国内某金融结构由于自动化交易程序漏洞造成乌龙指发生引发巨大损失等，这些事件都对企业人工智能风险管理敲响了警钟。

从监管环境来看，在新出台的《网络安全法》、《电子商务法》、《资管新规》等法律法规中都包含了人工智能应用有关的法规要求，国内金融行业监管机构也正在抓紧制定针对人工智能、金融科技等颠覆性技术所带来的新型风险的应对指导意见和监管要求。

因此，对于已经在或者计划在业务中应用人工智能技术的各类企业来说，都应当在企业风险管理过程中考虑人工智能风险，充分评估人工智能技术蕴含的复杂性、不确定性、难解释性以及发展速度等风险因素，开展人工智能风险管理工作已经刻不容缓。

部分国内人工智能法律法规要求

- **《中华人民共和国网络安全法》—规范人工智能个人信息保护要求**

  第四十一条网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意。

  网络运营者不得收集与其提供的服务无关的个人信息，不得违反法律、行政法规的规定和双方的约定收集、使用个人信息，并应当依照法律、行政法规的规定和与用户的约定，处理其保存的个人信息。

- **《中华人民共和国电子商务法》—约束大数据杀熟等违法行为**

  第十八条电子商务经营者根据消费者的兴趣爱好、消费习惯等特征向其提供商品或者服务的搜索结果的，应当同时向该消费者提供不针对其个人特征的选项，尊重和平等保护消费者合法权益。

- **《关于规范金融机构资产管理业务的指导意见》（2018资管新规）—明确金融机构资管业务人工智能应用责任主体和风险应对要求**

  第二十三条：运用人工智能技术开展投资顾问业务应当取得投资顾问资质，非金融机构不得借助智能投资顾问超范围经营或者变相开展资产管理业务。

  金融机构运用人工智能技术开展资产管理业务应当严格遵守本意见有关投资者适当性、投资范围、信息披露、风险隔离等一般性规定，不得借助人工智能业务夸大宣传资产管理产品或者误导投资者。金融机构应当向金融监督管理部门报备人工智能模型的主要参数以及资产配置的主要逻辑，为投资者单独设立智能管理账户，充分提示人工智能算法的固有缺陷和使用风险，明晰交易流程，强化留痕管理，严格监控智能管理账户的交易头寸、风险限额、交易种类、价格权限等。金融机构因违法违规或者管理不当造成投资者损失的，应当依法承担损害赔偿责任。

  金融机构应当根据不同产品投资策略研发对应的人工智能算法或者程序化交易，避免算法同质化加剧投资行为的顺周期性，并针对由此可能引发的市场波动风险制定应对预案。因算法同质化、编程设计错误、对数据利用深度不够等人工智能算法模型缺陷或者系统异常，导致羊群效应、影响金融市场稳定运行的，金融机构应当及时采取人工干预措施，强制调整或者终止人工智能业务。

译者

朱磊
德勤中国风险咨询合伙人
jaczhu@deloitte.com.cn

何铮
德勤中国风险咨询合伙人
zhhe@deloitte.com.cn

俞宁子
德勤中国风险咨询合伙人
jerryu@deloitte.com.cn

作者

Tom Bigham
德勤英国风险咨询总监
tbigham@deloitte.co.uk

Valeria Gallo
德勤英国风险咨询经理
vgallo@deloitte.co.uk

Suchitra Nair
德勤英国风险咨询总监
snair@deloitte.co.uk

Michelle Lee
德勤英国风险咨询经理
michellealee@deloitte.co.uk

Sulabh Soral
德勤英国咨询总监
ssoral@deloitte.co.uk

Tom Mews
德勤英国风险咨询经理
michellealee@deloitte.co.uk

Alan Tua
德勤英国风险咨询总监
altua@deloitte.co.uk

Morgane Fouché
德勤英国风险咨询高级顾问
mfouche@deloitte.co.uk