

Deloitte.

德勤



智海领航

洞悉人工智能在金融服务领域的战略及监管风险

目录

引言	3
人工智能的挑战	4
人工智能的可解释性	6
系统性风险和人工智能	8
歧视与公平	11
其他不确定性	13
算法受托	13
算法合谋	13
结论	15
联系人	16

引言

各位同事，

自2015年以来，Deloitte Consulting LLP (以下简称“德勤”) 一直与世界经济论坛 (以下简称“论坛”) 合作，共同探讨金融服务业变革的驱动力量。这一努力最终，或者说不可避免地，会触及[人工智能 \(AI\) 在金融服务领域的应用](#)这一话题。

我们也在思考：金融机构在有效应对人工智能带来的风险的同时，如何从人工智能中获益？这也是论坛于2019年发布的[《智海领航：在金融服务领域使用 AI 进行负责任创新的路线图》](#)报告所探讨的专题之一。

为获得更深刻的洞察，我们开展了超过十个月的调研，访谈了250多位全球各地的行业专家，其中包括许多知名金融机构的高管。此外，我们还在世界各地举办了一系列的线下研讨会。

通过这些调研，我们提炼出三大焦点问题：

- 各利益相关方如何才能了解人工智能系统进行决策的原理？
- 人工智能算法是否会破坏金融系统的稳定性？
- 金融机构如何避免人工智能解决方案的歧视问题？

为深入了解上述及其他相关问题，我们探讨了潜在的风险和不确定性，有效管理应对此类风险和不确定性带来的机遇，以及金融机构应对此类问题的领先实践。我们希望本论坛报告能对金融服务领域合理运用人工智能有所启示，并为企业在未来不确定性中平稳前行保驾护航。

Bob Contri

德勤全球
金融服务业领导合伙人
bcontri@deloitte.com

Rob Galaski

德勤管理咨询
银行业及资本市场
全球咨询领导合伙人
rgalaski@deloitte.ca

人工智能的挑战

伴随人工智能的兴起，金融行业进入了一片未知的“海域”。早期的“航海家”可能在这片海域收获颇丰，但也可能在航程中面对风险和不确定性。

行业能否顺利进入人工智能驱动时代，取决于是否具备以下能力：

合理部署现行人工智能系统。人工智能系统能以人类尚不熟知的方式进行自我学习和“思考”。这种难以理解的特性会导致一系列问题，诸如对客户的歧视、系统不稳定性，尤其在人工智能系统之间以更高的频率和速度进行交互时。

因此，金融行业可能不得不采用新的治理方式，其中可能包括应对人工智能风险事件所采取的例如检测及预防等保障措施，风险发生时的补救机制和提供必要干预手段的流程。此外，针对每个案例，系统自身可能需要具备某种形式的审核机制。

合理扩大人工智能在未来金融系统中的应用。人工智能的应用正推动多个行业重新审视竞争政策、隐私权和运营弹性。这一过程中的共同关注点是什么？答案是数据。金融机构之间的数据交互日益频繁。大型数据存储为亟需数据的人工智能系统提供了“赢家通吃”的环境。与此同时，客户也逐渐意识到个人数据的价值。

上述动态变化将决定金融机构如何部署人工智能，以及采用何种战略。而潜在的影响不仅包括数据的应用，还包括市场参与者、适用的商业模式和法规等方面的变化。行业的领导者应当积极地参与相关决策，以确保相关政策能对金融机构产生正面的影响。

激发具备可靠人工智能技术的金融系统的潜力。审慎应用人工智能不仅可规避风险，还可以推动行业向善。某些机构已使用人工智能打造金融产品，从而更好地服务于客户和社会。（图1）

金融服务严监管的特性对于人工智能而言反而是“一线生机”。消费者愈加关注其数据权利和数字主权，这导致科技巨头的优势被逐步削弱，而金融机构则有机会利用[更高客户信任度这一优势¹](#)，实现客户全新的价值主张。

图1: 如何利用人工智能提高金融系统的道德标准



实现上述目标应该如何行动呢?最重要的一点就是行业领袖需要同监管机构及政策制定者紧密合作,以缓解人们对金融服务中人工智能应用的担忧。人们最担忧的有两个问题,其一是不清楚人工智能是如何产生结果的,其二是人工智能可能破坏金融系统的稳定性。此外,人工智能还可能强化金融决策中的歧视现象。另外还有两个较少讨论的不确定事项也需要注意:算法受托和算法合谋。

上述这些是金融服务业在运用人工智能过程中普遍存在的一些障碍,我们需要更深入地分析,才能找到解决办法。所以让我们来逐一分析这些障碍。

金融领域运用人工智能的五大关注点:



人工智能的可解释性

随着深度学习及其他先进方法的出现，人工智能变得越来越复杂。一定程度上是因为人工智能系统可容许大量数据输入，而那些对输出结果影响最大的输入项却很难被找到。此外，现有系统在数据输入（例如姓名、收入和信用评级）与结果输出（例如提供利息为12%的5,000美元贷款）间通常存在数百个处理步骤，这不利于让人们理解人工智能生成决策的逻辑。

其他挑战则深入到认知技术的本质。由于人工智能会随着时间推移进行自我调整，因此过往的决策未必能预测未来决策。下面这一点更为关键：人工智能的运行方式与人类智能迥异。人类做出的判断可能不会对人工智能的决策过程产生影响，这将让人更加难以理解系统的决策逻辑。

知情信任

上述的各种情况都能将人工智能系统变成一个“黑匣子”，甚至对其开发者也如此。这促使金融机构和监管机构寻求一个中间地带——介于对人工智能盲目信任（将导致极大风险）和对人工智能不屑一顾（将严重阻碍创新）之间。这种平衡关键在于向用户提供充足的人工智能系统的相关信息，使用户可以信任人工智能系统作出的决策。

怎样才能建立这种程度的知情信任呢？这需要视情况而定。某些情况下，如个性化营销推广可能不需任何解释。但其他情况可能就需要：

- **提供足够的透明度** 使用用户理解人工智能系统如何做出决策
- **对应到具体的场景** 使用用户理解人工智能系统为何作出特定选择
- **进行必要的管控** 使用用户相信人工智能生成的决策会符合道德、法律和商业要求

换言之，可能需要向用户解释的原因有很多，并且每个原因相应都需要对底层技术采用相应的措施。（图2）

图2：在人工智能系统中管理可解释性的措施



值得注意的是，提升可解释性这一工作并非一劳永逸。人工智能技术日新月异，金融机构需要随着技术的发展定期重新评估其人工智能战略。我们已经可以想象到未来将会发生的变化。目前，满足监管的要求是推动可解释性提升的主要因素。然而，随着人脸或语音识别等人工智能解决方案愈发依赖敏感数据，可解释性将成为消费者让渡数据主权的前置条件和机构之间差异化竞争力的重要体现。

对高管的重要启示



人工智能系统应可以进行自我解释，并满足在相互关联的机构间进行有效风险管理及问责这一前提条件。



根据场景，仅需提供特定类型的知情信任。



行业内部充分沟通对于制订可解释性程度的行业标准有着重要作用。



通过人工智能的可解释性获得客户或员工的信任可能成为一项战略决策，金融机构需为此加大投资。



可解释性未必能防止现实世界的歧视被编入算法中。（后文将详细介绍）

系统性风险和人工智能

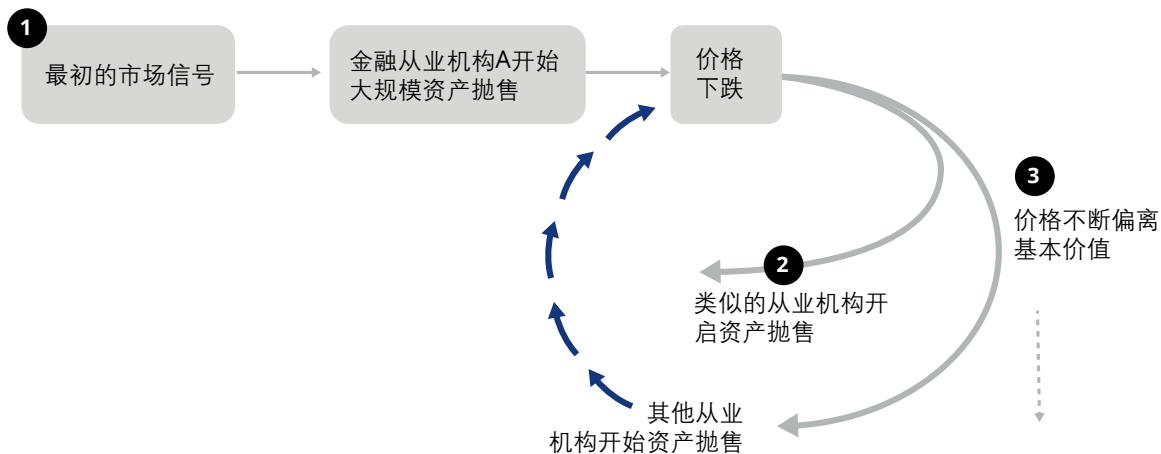
20世纪70年代以前，投资者在金融市场中享有相对公平的竞争环境。但在此之后，基于计算机的交易导致竞争环境发生天翻地覆的变化，为有能力直接开展自动化交易的大型专业机构提供了有利条件。这些企业闪电般的交易速度常常会加剧市场波动，当其他投资者还没来得及作出反应，市场价格就已迅速暴跌。

全新风险来源

如今，人工智能将再次推动金融市场进行转型。它广泛适用于金融服务行业的各个领域，能积极推动人与机器之间更加频繁的互动。同时，人工智能的黑匣子效应致使投资者更难解读不断变化的市场动态，也更难识别或洞悉新兴风险。总之，上述情况可以通过以下方式增加整个系统的风险：

从众效应。金融机构以相似方式解读市场信号时将会出现从众效应。举例来说，这些机构可能通过出售资产来对某个新闻事件作出反应。大规模抛售将导致资产价格下跌，进而造成进一步的抛售。(图3)

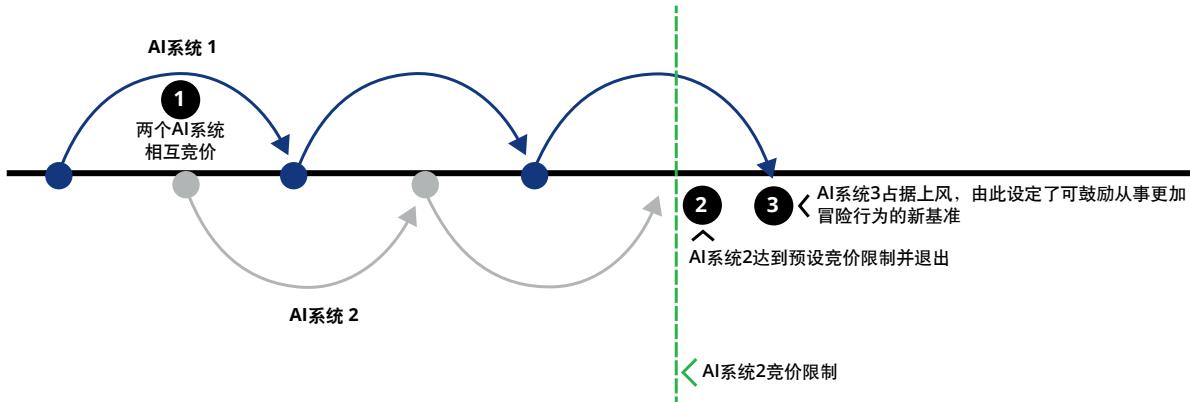
图3：从众行为



通过建立熔断机制等相关合理机制可以缓解从众效应，这样做可以暂停市场参与者的行动并将真正的市场信号从算法噪声 (algorithmic noise) 中分离出来。此外，人工智能设计的多样化亦有助于降低此类效应，尽管这可能需要行业统一制定并执行一套健全的人工智能开发标准。

算法竞争。假设两个人工智能系统持续相互竞价，并促使其不断优化策略以获得最高的市场价格。两者在竞价过程中，市场价格不断上涨，直至其中一个系统因盈利限制因素而被迫退出。这真的会发生么？答案是肯定的。但这可能会鼓励参与者从事风险更大的活动以实现超级盈利或导致其完全退出市场，从而严重损害其财务状况。(图4)

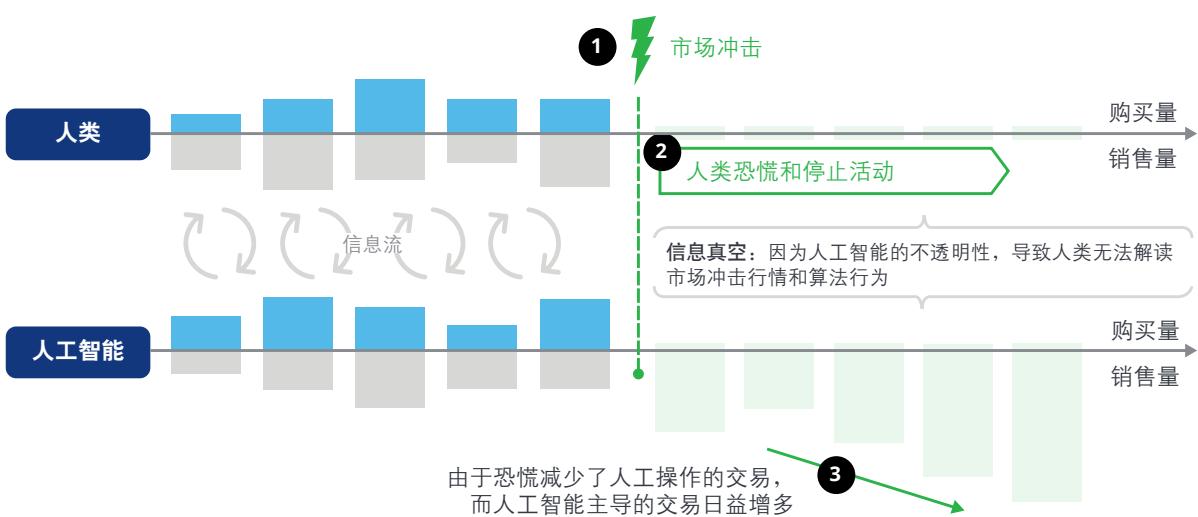
图4: 算法竞争



场境建模有助于识别人工智能不稳定行为可能发生的时机。但管理此类行为可能需要在资源配置方面设定多样化的原则，或限制自主人工智能重新分配资产的速度。

信息真空。一个典型情况是自动化系统触发大规模抛售。随着价格直线下降，在信息严重不对称的情况下，投资者往往无法理解当前的局势，最终只能停止一切交易。而算法会将投资者的这一行为解读为不感兴趣或欲低价清仓。只要信息不对称持续存在，市场就会一直处于失衡状态。(图5)

图5: 信息真空期的恐慌情绪



熔断机制通过暂停所有活动来中断由算法噪声引起的反馈循环，以便参与者有时间验证其信息源。这也说明，短期内最重要的优先任务反而是设计一个用于分析资讯和市场情绪的人工智能系统，以防止市场信息误读或伪造市场信息。

侵蚀现有防御措施

伴随人工智能而来的新的风险来源，可能导致金融稳定性的传统防御措施失效。例如，当前金融系统依赖人们的判断来发现和防范风险，而依赖人工智能后，人们可能逐渐丧失风险防范的相关技能。此外，若人工智能引发更多小规模干扰，人们可能会逐渐习以为常而麻痹大意，从而无法预见或防止破坏稳定性重大事件。人工智能系统有时也可能有意或无意地与监管规则或预测性监管系统进行博弈。

为维持现行的保障措施，金融机构可：

- 为相关人员提供有关不断变化的市场和消费者行为的关键信息
- 采用交互式及可解释型人工智能系统来培养员工相关能力
- 侦测新兴的系统性风险
- 系统性分析市场冲击事件
- 在危机中对基于人工智能博弈的影响进行压力测试
- 基于金融机构已有规则开展自主的“完备性检查”

然而，由于金融系统自身在发生变化，上述策略及其它策略也将随之改变。新兴企业正进军金融服务市场，现有的金融机构也在不断尝试新的商业模式。在这些结构性转变中，系统弹性可能很快会成为危机的焦点。

对高管的重要启示



熔断机制等旧工具将与新工具结合使用，从而有效管理人工智能环境下的由市场恐慌所带来的风险。



从长远来看，可解释型人工智能有助于构建系统性的安全环境。



针对机器集成式金融系统的动态风险，监管机构需承担新的角色并与金融机构建立更加深入的互利共赢伙伴关系。

未来监管机构

在瞬息万变、日趋复杂的金融行业，监管机构需承担如下新的责任：



模拟和分析人工智能的潜在影响



发布针对金融服务领域重要新趋势的见解



宣传人工智能设计中具体应用的最佳实践



发布预防性警报以整合相关机构的数据



在市场崩盘期间从算法噪声中分离出市场真实趋势



为金融行业提供测试新型人工智能解决方案的沙盒

歧视与公平

为了给产品和服务合理定价，金融机构须评估客户的风险。但从历史来看，金融决策依据有时会包含种族和性别等与风险无关的因素。这样做的结果会阻碍客户积累财富，抑制经济获得新的生产力，还可能使金融机构面临更严格的监管审查，并引发公众的强烈不满。

人工智能的角色

就歧视而言，人工智能的潜能错综复杂。一方面，人工智能可创造以下机会：

- 为无银行账户者远程提供高质量低成本的服务
- 利用社交、远程信息及其它可替代性数据，向无银行账户或缺乏银行服务的人提供服务
- 按需提供小额贷款或短期政策性产品等个性化产品

另一方面，人工智能也会：

- 从新数据源中产生无意识的歧视
- 增加识别无意识歧视的难度
- 更快更广泛地在系统中传播歧视，并放大歧视的影响

歧视会在金融系统中以多种方式呈现。

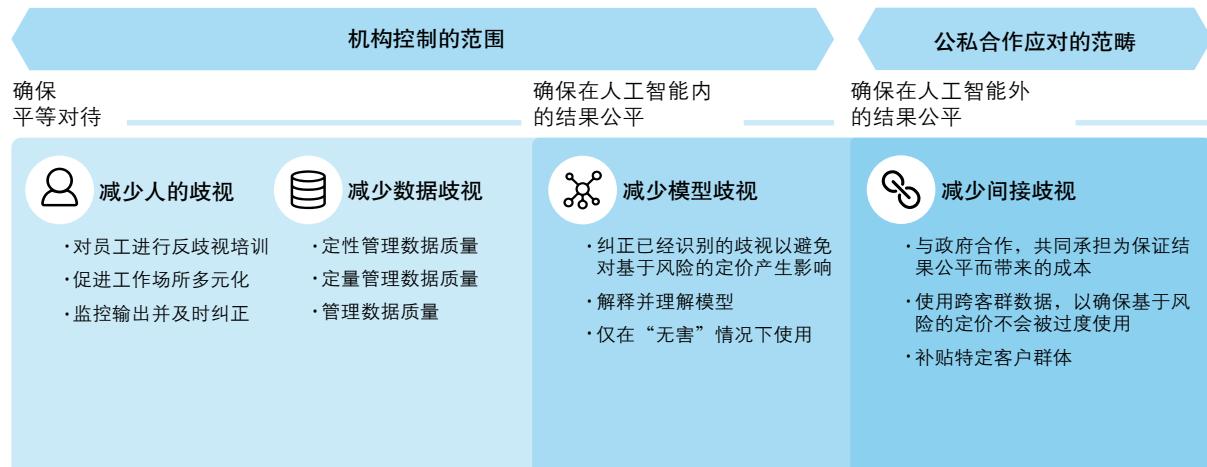
来自人的歧视。人类思维的系统性错误会影响人们的决策和判断。这些歧视可能是有意或无意的。人工智能在系统设计和数据收集阶段可能会引入来自人的歧视。此外，监督式学习和模型应用也会将人类的歧视带入人工智能系统。

数据歧视。若使用不准确、歪曲或其他歧视性数据进行决策系统的开发或训练，可能导致系统对特定人群产生扭曲的结果。

模型歧视。计算能力、系统设计或编程逻辑方面的限制——更不必说用户造成的错误等，均会导致人工智能系统的输出结果存在歧视。模型歧视包括因果关系的混淆和对弱势群体的过度保护。

间接歧视。不幸地是，教育、就业和法律诉讼等生活的方方面面目前均存在不公平的歧视。这显然增加了遭受歧视的个人可能面临的金融风险。与其他歧视不同，构建或使用给定模型的人不能直接控制间接歧视；为减少此类歧视，可能需要机构放弃一些对统计精确度的要求。这些间接歧视可能通过更具广度和深度的数据（包括来自第三方的数据），输入给人工智能。

图5：为最大程度降低人工智能中产生的歧视，需定义“公平”的范畴，以及机构在“间接歧视”的这一问题上的职责边界



对高管的重要启示



人工智能在金融服务领域推广金融包容性提供了新机遇。同时，人工智能可能增加不公平及歧视性金融决策的风险。



尽管金融决策中的歧视并不鲜见，但人工智能可能通过建立新模式将歧视引入决策、加快歧视的传播速度并加大其被侦测的难度，从而引发新的忧虑。



现有工具和系统有助于识别和减少来自人类、数据和算法的歧视。但人工智能会带来复杂的结果，尤其是在间接歧视方面（歧视的根源不在金融机构的直接控制或影响范围内）。



减少歧视需摒弃纯粹基于风险的方法，尤其是对间接歧视而言。金融机构和政策制定者需共同确定对公平的定义来应对此类挑战，这将涉及企业与政府间的成本分担问题。

其他不确定性

就金融领域的人工智能而言，可解释性、系统性风险及歧视与公平是三大重点关注问题。此外，还有两大问题值得提及：一是人工智能系统能否履行受托义务，即始终以客户的最佳利益行事；二是与算法合谋相关，即人工智能系统学会从事反竞争行为。

算法受托

人工智能系统开始承担过去仅由人类负责的金融任务。例如，人工智能可以：

- 从事证券投资活动并对其进行管理以获得高于市场的回报
- 推荐符合客户长期利益目标的资产配置方案
- 提供保险、退休、税务或财务规划等一系列产品和服务相关的全方位的日常金融管理。

这些均存在受托责任相关的重要问题。在美国，人工智能算法已在注册投资顾问中广泛应用，应对这些问题已成为当务之急。值得高兴的是，自动化系统有可能达到与传统公司相同的受托标准。这在一定程度上是由于人工智能可有效地搜索、优化和汇总与每个客户财务情况相关的数据。

然而，也存在先前讨论的可解释性问题。人工智能作出决策及建议背后的依据应告知受托人。但随着人工智能系统为应对日益繁多的产品和服务而整合了更为广泛的信息之后，人们越发难以理解特定决策背后的逻辑。

此外，如果人工智能系统未能履行其受托义务，对此产生的后果也具有不确定性。谁应对此负责：开发人员、机构、监管层还是其他人员？或者，利用对冲性保险来承担后果？无论采取何种方式，监管机构和金融机构可能需要通过其他方法管理人工智能系统，以保证受影响的各方能尽可能快速、无缝地从危机中完全恢复。

算法合谋

人工智能系统会导致市场合谋行为的监督和管理复杂化。原因何在？不妨思考一下现有的合谋，即机构间通过沟通与合作，以获取特定结果或从改变竞争态势中受益。

根据此定义，人工智能系统间的合谋可能难以被确认。譬如，人们可能不清楚决策属于独立有效的分析结果还是合谋行为。另一难题是人工智能可实时整合大量源自外部的结构化和非结构化数据。有鉴于此，其他市场参与者的动机变得更加易于理解，而无需任何实质性的交互。

因此，在现行法规下，人工智能合谋因其似是而非的形式而更加难以被侦测、确认和检举。这增加了界定其非法和合法的不确定性。

即便如此，金融机构也并非完全无力降低此类风险。它们可以对人工智能系统和模型进行限定，使人工智能系统：

- 根据明确正当的商业目的开展交互
- 基于有效、合法的商业理由作出决策
- 通过人工监督以验证人工智能决策是否符合业务实际

除此之外，算法合谋几乎不存在。算法合谋发生的前提是金融产品和服务相对同质化。同时，各金融机构也具有相似的营业利润，且买方权利被大大限制。

对高管的重要启示



解决受托责任和算法合谋相关的问题需要弥合人工智能专家与政策制定者或法律专家在认知上的差距。



通过单一的框架对人类和机器参与者进行管理是否足够，抑或是否需要专门为机器代理设计单独的框架，这一问题仍有待观察。



通过阻止算法合谋及承担受托责任，企业可赢得信任并使其有别于标准较宽松的竞争对手。

结论

人工智能涉及计算的全新形态。金融机构需克服人工智能在可解释性、系统性风险和公平等方面的挑战，进而从中受益。总而言之，金融行业应当：

- **在现行金融生态系统中合理地部署人工智能。**人工智能要求金融行业开发并合理使用全新工具来保护金融体系。
- **合理扩大基于人工智能的金融生态体系的范畴。**随着时代的发展，人工智能将改变金融机构的运营及参与市场的方式。监管机构势必出台新的监管政策，而这些政策将从根本上改变金融机构可能的战略选择。如果金融机构希望对未来的政策变化有所影响，目前就应该积极参与到与监管机构前期的沟通、讨论与合作中来。
- **激发具备可靠人工智能技术的金融生态体系的潜力。**与传统系统相比，人工智能系统能够更迅速地产生更广泛的影响。因此，人工智能系统应具备更高的标准。

首先，重点是认识到人工智能并非孤立存在。人工智能已与云和物联网 (IoT) 等技术紧密结合，量子计算、5G 和其他新兴技术很快也将加入其中。在这一科技飞速发展的环境中，金融行业更加需要提前考虑人工智能的潜在影响。

人工智能的影响到底是什么呢？首先是战略上的转变。人工智能（及其他新兴技术）的普及正推动金融机构与监管层更频繁的对话，重点探讨如何对新技术进行有效的部署和应用。金融机构若想有效制定其未来战略，就应主动与所有利益相关者沟通。同样地，该方法也适用于应对监管政策变化对运营、战略和人力资本的影响。

其次是对优先级事项的重新梳理。对于人工智能的应用，其公司治理是首要工作——无论是为了应对风险预警，还是为了与客户的利益保持一致。重视公司治理不仅为金融机构提供实现差异化的契机，而且可以通过与监管机构和金融服务生态系统中的其他参与者建立协作文化来实现数据驱动的决策。

最后，我们大可不必针对人工智能在金融服务领域的应用而建立一套全新的道德标准。真正需要我们做的，是以开放的态度拥抱适用于人工智能应用的新方法。需谨记，无论这些方法如何创新，我们都应以公正性、稳定性、透明度与可操作性四大标准予以权衡，并决定是否采纳。正是这套标准，使得金融行业能够在过往的技术变革浪潮中脱颖而出。

联系人

全球联系人

Bob Contri

德勤全球
金融服务业领导合伙人
bcontri@deloitte.com

Rob Galaski

德勤管理咨询
银行业及资本市场
全球咨询领导合伙人
rgalaski@deloitte.ca

作者

Ishani Majumdar

德勤加拿大
Omnia人工智能高级咨询顾问
ismajumdar@deloitte.ca

中国联系人

吴卫军

德勤中国
副主席
金融服务业领导合伙人
电话: +86 10 8512 5999
电子邮件: davidwjwu@deloitte.com.cn

张庆杰

德勤中国
管理咨询
分析与数据管理领导合伙人
电话: +86 10 8520 7830
电子邮件: qjzhang@deloitte.com.cn

薛梓源

德勤中国
风险咨询
科技风险及网络安全领导合伙人
电话: +86 10 8520 7315
电子邮件: tonxue@deloitte.com.cn

尤忠彬

德勤中国
管理咨询
分析与数据管理合伙人
电话: +86 21 2316 6172
电子邮件: zhyou@deloitte.com.cn

特此感谢德勤加拿大Courtney Kidd Chubb和Denizhan Uykur对撰写本报告所提供的帮助。

尾注

1. BaFin, Big Data meets artificial intelligence, 16 July 2018, https://www.bafin.de/SharedDocs/Downloads/EN/dl_bdai_studie_en.html;jsessionid=DF0654DBF78C8FA8AE5E4AA58148CB44.1_cid381?nn=9866146, accessed 5 November 2019.
2. "New Scoring Models For The Underserved Gaining International Attention," PYMNTS, 3 July 2018, <https://www.pymnts.com/news/banking/2018/alternative-credit-scoring-underbanked-unbanked-nova/>, accessed 5 November 2019.
3. Sudipto Ghosh, "Temenos Acquires a SaaS-based, Patented, Explainable AI (XAI) Platform," AiThority, 18 July 2019, <https://www.aithority.com/natural-language/information-extraction/temenos-acquires-a-saas-based-patented-explainable-ai-xai-platform/>, accessed 5 November 2019.
4. "Personetics leverages AI to help chip away at student loan crisis," FinTech Futures, 30 October 2017, <https://www.fintechfutures.com/2017/10/personetics-leverages-ai-to-help-chip-away-at-student-loan-crisis/>, accessed 5 November 2019.
5. Financial Stability Board, Artificial intelligence and machine learning in financial services: Market developments and financial stability implications, 1 November 2017, <https://www.fsb.org/wp-content/uploads/P011117.pdf>, accessed 5 November 2019.



关于德勤

Deloitte（“德勤”）泛指一家或多家德勤有限公司，以及其全球成员所网络和它们的关联机构。德勤有限公司（又称“德勤全球”）及其每一家成员所和它们的关联机构均为具有独立法律地位的法律实体。德勤有限公司并不向客户提供服务。请参阅 www.deloitte.com/cn/about 了解更多信息。

德勤亚太有限公司（即一家担保有限公司）是德勤有限公司的成员所。德勤亚太有限公司的每一家成员及其关联机构均为具有独立法律地位的法律实体，在亚太地区超过100座城市提供专业服务，包括奥克兰、曼谷、北京、河内、香港、雅加达、吉隆坡、马尼拉、墨尔本、大阪、上海、新加坡、悉尼、台北和东京。

德勤于1917年在上海设立办事处，德勤品牌由此进入中国。如今，德勤中国为中国本地和在华的跨国及高增长企业客户提供全面的审计及鉴证、管理咨询、财务咨询、风险咨询和税务服务。德勤中国持续致力为中国会计准则、税务制度及专业人才培养作出重要贡献。德勤中国是一家中国本土成立的专业服务机构，由德勤中国的合伙人所拥有。敬请访问 www2.deloitte.com/cn/zh/social-media，通过我们的社交媒体平台，了解德勤在中国市场成就不凡的更多信息。

本通信中所含内容乃一般性信息，任何德勤有限公司、其成员所或它们的关联机构（统称为“德勤网络”）并不因此构成提供任何专业建议或服务。在作出任何可能影响您的财务或业务的决策或采取任何相关行动前，您应咨询合资格的专业顾问。任何德勤网络内的机构均不对任何方因使用本通信而导致的任何损失承担责任。

© 2020。欲了解更多信息，请联系德勤中国。
Designed by CoRe Creative Services. RITM0409216