



在银行业管理人工智能： 我们准备好合作了吗？

帕布罗·埃尔南德斯·德·科斯

1986年，历史学家梅尔文·克兰兹伯格发表了《技术六定律》，其将“技术既不好也不坏；它也不是中性的”列为首位定律（Kranzberg, 1986）。这一观点可以无缝地适用于当下对银行业应用人工智能和机器学习的讨论中。

关于在银行业中使用生成式人工智能（Generative AI）和大型语言模型（Large Language Models）的前景和陷阱的讨论正变得越来越普遍。随着银行开始使用、投资人工智能和机器学习，讨论也在转换为行动。更大胆的是，一些人已经在谈论尚未建立的人工通用智能（Artificial General Intelligence）对银行业的潜在影响（Silva, 2024）。

现在有一种新兴的说法，称赞银行业应用人工智能的所谓好处，例如在运营效率和风险管理方面的改善，也警告所带来的数据隐私、模型幻觉和声誉风险等挑战。

然而，我们仍有些悬而未决的问题：在银行业中使用人工智能和机器学习对全球金融稳定，甚至对整个社会来说，是正面还是负面的？我们是否考虑过人工智能和机器学习在银行业发挥重要作用的世界里所有的潜在场景？如果我们未能退后一

步，考虑更大的图景，我们是否面临着关于人工智能和机器学习的“共识幻觉”风险？

本文将遵循克兰兹伯格的框架，并结合宏观审慎及金融稳定等考量因素。本文的主要观点是，人工智能在银行业的应用，对审慎和金融稳定提出了重要挑战。我们还没有看到人工智能和机器学习在整个金融周期中的表现。如果不加以控制，这些模型可能会放大未来的银行业危机。但如果中央银行和监管机构能够适应这一新现实并有效合作，这些挑战和限制并非不可克服。

一、银行业应用人工智能和机器学习的潜在好处

一般而言，人工智能和机器学习可以帮助提高银行的运营效率、风险管理能力和产品供应。具体来讲，与传统方法相比，人工智能和机器学习的应用可能包括：①更优越的模式识别能力和预测能力，例如改善投资表现或扩大信贷渠道；②成本效率，即人工智能和机器学习方法可能以更低的成本获得同样的表现结果，例如能够开发多渠道客户访问和增加客户自助服

帕布罗·埃尔南德斯·德·科斯（Pablo Hernández de Cos），巴塞尔银行监管委员会主席、西班牙中央银行行长。本文根据帕布罗·埃尔南德斯·德·科斯2024年4月17日在国际金融协会（Institute of International Finance）全球展望论坛上的主旨演讲整理。原文出自 <https://www.bis.org/speeches/sp240417.pdf>。喻雪莹编译。



务；③与需要更多人力投入和更高“操作错误”的方法相比，处理的准确性和一致性更高，例如反洗钱监测中异常检测业务；④改善客户体验，即通过消除手动步骤来帮助简化客户交互，例如贷款申请业务；⑤对大数据、非结构化数据的适应及处理能力更强更有效，例如能够更深入地了解客户需求，提供定制化程度更高的服务。

实践中已涌现出许多这些技术的应用案例。一个值得注意的领域是在“科学的”应用中使用人工智能和机器学习，例如利用图神经网络（Graph Neural Networks）监测洗钱活动。一个例子是由国际清算银行创新中心的 Aurora 项目，通过人工智能来识别传统方法无法检测到的交易数据中的异常情况，从而加强反洗钱交易监控（BIS Innovation Hub, 2023）。在这种情况下，人工智能可以帮助降低银行和银行体系的风险。

我们还看到，一些银行正在尝试将自然语言处理或机器学习方法用于数字助理（机器人顾问）、总结文档、代码生成和测试、欺诈和洗钱检测，以及算法交易策略。

同样，监管机构也一直在探索利用人工智能和机器学习等高科技提高履职能力。例如国际清算银行创新中心已经开发了各种概念证明，依靠人工智能和机器学习来提升监管报告和数据分析（项目 Ellipse, BIS Innovation Hub, 2022），并促进气候风险分析（项目 Gaia, BIS Innovation Hub, 2024）。

人工智能和机器学习改变银行系统格局的例子不胜枚举。因此，我们更应确保仔细评估相关的限制和风险。

二、银行业应用人工智能和机器学习的潜在风险

在某些方面，银行业中的人工智能和机器学习模型本质上是现有建模技术的更复杂、更强大的版本。但它们潜在的规模和范围带来了重要的审慎风险和限制。

首先是关于这些模型中使用的输入数据。更大的数据量、更高的速度和不同的种类，可能会在确保数据质量、相关性、安全性和保密性方面带来更大的数据治理挑战。用户可能不知道生成产出的来源或这些来源如何加权，金融机构可能无法完全理解或控制所使用的数据集。在广泛依赖这些模型之前，我们需要回答一些关于数据治理和伦理问题的基本问题。另外，人工智能和机器学习模型的可改进范围也是一个问题。迄今为止，这些模型依赖于越来越大的数据集来提高准确性。然而，还有多少未使用的数据可供这些模型的下一代使用？一项估计表明，现有的高质量语言数据最早可能在 2025 年用尽（Villalobos et al., 2022 ; The Economist, 2023）。因此，未来可能更需要依靠其他技术来增强这些模型及其在银行业中的实用性。

其次，与传统模型一样，人工智能和机器学习模型会反映训练数据中的偏差和不准确性。幻觉^①产生错误和有缺陷的输出，有时甚至以令人信服的方式。需要说明的是，人类在决策时也容易产生偏见。而且在许多情况下，这些模型无法被“解释”——换句话说，我们并不总是理解一个模型为什么及如何产生特定的输出。现有模型无法被解释的部分主要在于统计误

① 这个术语将模型拟人化，不但毫无帮助，反而淡化了不准确的严重性。



差和残差，而人工智能和机器学习模式相较于现有模型又增加了一层关于模型可解释性的问题。因此，银行和监管机构必须对它们在关键银行服务中的潜在应用的稳健性感到足够放心和自信。

对于这样的模型，我们的风险承受水平应该是多少？一个模型给出的1万个“正确”答案，值得付出一次“幻觉”的代价吗，尤其是对于具有系统重要性的银行活动而言？我们目前是否有足够有力的缓解措施来应对风险模型产生的不正确、有偏见或不道德的输出？如果我们不知道为什么会得到这样的结果，这是否要紧？这些都是银行和监管机构应该考虑的重要问题。

最后，当前的人工智能和机器学习格局可能会提高银行对第三方的依赖，加剧集中度风险。迄今为止，大多数基础模型是由少数提供者开发的。这导致银行在使用这些供应商开发的现成或改编模型时行为趋同、路径依赖。人工智能和机器学习的部署通常还涉及云技术的使用，进一步巩固了这种互联网络，这可以测试银行的运营弹性。网络效应和规模经济会强化这些服务提供商的作用，并进一步加剧银行的依赖风险。监管此类第三方服务提供商时，需要一个简单的原则：对其治理、风险管理和运营弹性所要求的实践水平和稳健程度应等同于对银行机构的要求。

三、银行业应用人工智能和机器学习并非中立

除了这些主要是微观层面的考虑之外，在银行业中使用人工智能和机器学习时，可能还有一个更根本的金融稳定挑战。人

工智能和机器学习模型本质上是“中立的”，还是通过设计渲染和夸大了人类行为？

人工智能和机器学习模型从现有模型中脱颖而出的是以下特质：①自动化，即它们在没有或很少人为参与情况下运行的能力；②速度，因为它们的处理能力和对大数据的依赖；③可能无处不在，因为广泛的、跨部门的采用和整合（即“网络效应”）。

借用蠢朋克乐队的说法，人工智能和机器学习模型可能比现有模型更硬、更快、更强大^①（Daft Punk, 2001），但它们一定更有利于金融稳定吗？

从整个系统的角度来看，银行对人工智能和机器学习模型的依赖会加剧金融周期自我强化的特征，包括明显的顺周期演变和总体风险的放大（Hernández de Cos, 2023）。这是因为随着时间的推移，人工智能和机器学习模型可能对内生风险的识别不那么可靠稳健（Danielsson & Uthemann, 2024；Danielsson & Shin, 2002；Danielsson et al., 2012）。这些模型严重依赖于历史模式，未能捕捉到人类在压力时期的行为动态。

在金融大危机期间，我们从现有的银行内部模型中学到了这一痛苦的教训。当这些模型在压力事件中崩溃时，内部模型“贯穿周期”的稳健性显然受到了质疑。回想起2007年8月，一家具有全球系统重要性银行的首席财务官观察到，他们“连续几天看到了25个标准差的波动”（Larsen, 2007），这真令人难以置信。事实上，《巴塞尔协议III》的一些改革旨在减少这些模型的作用，增强它们的稳健性（BCBS, 2017）。

① 引用了Daft Punk 2001年10月发行的Discovery专辑中的歌曲名“Harder Better Faster Stronger”。



在这个阶段，没有任何迹象表明人工智能和机器学习模型会有更好的表现。如果有什么不同的话，那就是它们的自动性、速度和普遍性，可能会加剧甩卖、挤兑和去杠杆化等放大机制，从而加剧未来的金融压力事件。在著名的人工智能系统中，参数的数量远远超过了100亿，有些甚至超过了1万亿。因此，模型错误设定(Model Misspecification)的风险高得令人难以置信(Giattino et al., 2023)。

人工智能和机器学习模型还会给银行带来间接的结构性风险。最近的研究表明，人工智能和机器学习有可能重塑劳动力市场(Cazzaniga et al., 2024)。在某些场景下，有些劳动力部门或行业的收入不平等程度和失业率可能会提升。这可能会令银行对此类家庭或企业的敞口产生负面的第二轮效应——可能在很短的时间内。因此，作为风险管理的一部分，银行必须对此类潜在情况保持警惕。

我们也知道现有的大型语言模型面临着重要的认知限制(Perez-Cruz & Shin, 2024)。当今一代的模型无法理解井字游戏^①或篮球等游戏的规则，即使它们可以在国际象棋或围棋等游戏中击败人类。因此，我们又如何能明智地依靠它们来理解不断演变的复杂金融游戏？换句话说，虽然人工智能和机器学习模型可能擅长在现有的干草堆中找到一根针，但当干草堆本身随着时间的推移而演变时，它们不太可能表现得那么好。

这就是为什么人类的判断必须继续成为银行业的核心部分，而银行自身的风险管理和治理安排则是其核心(Hernández de Cos, 2022)。人工智能中的“A”将永远

是人工的，不能取代人类的关键作用，这既体现在银行业中涉及关系的部分，也体现在挑战和推翻模型输出结果的能力。没有银行、银行家或监管者的银行业将是一个“美丽新世界”。

事实上，克兰茨伯格也注意到了这一点。他的另一条技术定律指出，“尽管技术可能是许多公共问题的主要因素，但非技术因素在技术政策决定中占据优先地位”(Kranzberg, 1986)。在讨论人工智能/机器学习在银行业中的作用时，我们不能忽视这些更大动态变量。

四、结论

数字创新将进一步推动跨境和跨部门金融互联互通。因此，维护全球金融稳定需要持续的合作，以确保我们实现适当的监管基线和监督方法，以监督人工智能和机器学习在银行业及其他领域的使用。鉴于非银行机构的持续增长和监管界限的模糊，这种合作可能会扩展到更广泛的行政管理机构，而不仅仅是中央银行和银行监管机构。有些人甚至呼吁对人工智能治理采取“技术审慎”的方法，在宏观审慎框架的基础上监督全系统的金融风险(Bremmer & Suleyman, 2023)。

就银行业而言，银行预测、监控人工智能和机器学习带来的风险和挑战——无论是微观还是宏观层面——并将其纳入日常风险管理和治理安排中，这一点至关重要。此外，人类的判断作为这个框架的一部分发挥着核心作用，这一点也至关重要。

中央银行和监管机构还必须预测并减轻银行体系的潜在风险和脆弱性。巴塞尔

① 井字棋(Tic Tac Toe)，是一种在3×3格子上的连珠游戏。



委员会已经发布了关于人工智能、机器学习及使用第三方和第四方服务提供商的监督通讯 (BCBS, 2022)。此外, 它将很快发布一份关于金融数字化及其对监管的影响的更全面的报告。

最后, 回到克兰茨伯格关于技术的最

后一条定律, “技术是一种非常人性化的活动” (Kranzberg, 1986)。尽管有人试图为银行业描绘一个机器人的未来, 但事实是, 人类——尤其是人类的判断力——能够而且必须继续发挥不可或缺的作用。^[N]

学术编辑: 韦燕春

参考文献

- [1] BCBS.High-level summary of Basel III reforms[R].The Basel Committee on Banking Supervision,2017.
- [2] BCBS.Newsletter on artificial intelligence and machine learning[R].The Basel Committee on Banking Supervision,2022a.
- [3] BCBS.Newsletter on third- and fourth-party risk management and concentration risk[R].The Basel Committee on Banking Supervision,2022b.
- [4] BIS Innovation Hub.Project Aurora: the power of data, technology and collaboration to combat money laundering across institutions and borders[R].Bank for International Settlements,2023.
- [5] BIS Innovation Hub.Project Ellipse: an integrated regulatory data and analytics platform[R].Bank for International Settlements,2022.
- [6] BIS Innovation Hub.Project Gaia: enabling climate risk analysis[R].Bank for International Settlements, 2024.
- [7] Bremmer I,Suleyman M.The AI power paradox[J].Foreign Affairs,2023(08).
- [8] Cazzaniga M,Jaumotte F,Li L,et al.Gen-AI: artifical intelligence and the future of work[R].IMF Staff Discussion Note,2024(01).
- [9] Danielsson J,Shin H S,Zigrand J-P.Procyclical leverage and endogenous risk[R].Risk Research Working Paper,2012.
- [10] Danielsson J,Shin H S.Endogenous risk[M]//Field P.Modern Risk Management,London: Risk Books,2002.
- [11] Danielsson J,Uthemann A.On the use of artificial intelligence in financial regulations and the impact on financial stability[J]arXiv:2310.11293,2024.
- [12] Giattino C,Mathieu E,Samborska V,et al.Data page: parameters in notable artificial intelligence systems [DB].Our World in Data,2023.
- [13] Hernández de Cos P.Back to the (macroprudential) future: reflections and questions on macroprudential policy[Z].speech at the HKMA-BIS joint conference "Future-proof supervision for an innovative banking world", Hong Kong SAR,24 March,2023.
- [14] Hernández de Cos P.International cooperation in a world of digitalisation[Z].speech at the 22nd International Conference of Banking Supervisors (virtual),29 November,2022b.
- [15] Hernández de Cos P.Trust, digitalisation and banking: from my word is my bond to my code is my bond?[Z].speech at the Eurofi 2022 Financial Forum,Prague,9 September,2022a.
- [16] Kranzberg M.Technology and history: Kranzberg's Laws[J].Technology and Culture,1986,27(3).
- [17] Larsen T P.Goldman pays the price of being big[N].Financial Times,2007-8-13.
- [18] Perez-Cruz F,Shin H S.Testing the cognitive limits of large language models[EB].BIS Bulletin,No. 83, 2024.
- [19] Sevilla V P J,Heim L,Besiroglu T,et al.Will we run out of data? An analysis of the limits of scaling datasets in machine learning[J].arXiv preprint,2306.155456,2022.
- [20] Silva L.How artificial general intelligence will drive an inclusive financial sector in Latin America[R].World Economic Forum Annual Meeting,2024.
- [21] The Economist.AI is setting off a great scramble for data[J].The Economist,2023-8-13.