日均Token消耗升至百亿量级 银行业AI应用进入下半场

中经记者 秦玉芳 广州报道

“日均Token消耗量”正在成为衡量银行数智化转型的关键标尺。

2026年上半年，不少银行密集披露了日均Token消耗数据，从数十亿到数百亿不等，较去年普遍数倍乃至数十倍增长。

Token消耗量激增背后，是银行AI应用从“技术验证期”全面迈入“规模化应用期”的深层信号。

面对Token账单的快速攀升，银行业正从“拼规模”转向“拼效率”，通过优化模型架构、建设混合算力平台、建立ROI考核机制等手段，力求让每一笔Token支出都能转化为可见的业务价值。

银行AI应用从“技术验证期”全面迈入“规模化应用期”。 视觉中国/图

从“拼规模”转向“拼效率”

过去Token只是模型调用的计量单位，如今它直接关联业务效率、客户体验和收入增长。

招商银行（600036.SH）首席信息官周天虹近日透露，截至5月底，该行仅业务单位在经营管理服务中应用大模型的日均Token消耗已达330亿。

形象点说，若把330亿个Token打印成标准的A4纸（每张纸约印1000个Token），那将得到3300万张纸。把这些纸首尾相连铺成一条纸带，长度超过9800公里。

不仅招商银行，越来越多银行日均Token消耗量也在快速突破百亿规模大关。

邮储银行（601658.SH）副行长、CIO牛新庄日前也披露，至2026年年底，预计该行日均大模型调用次数将超千万次，Token吞吐量超300亿。这意味着，该行的大模型调用次数、Token消耗量将翻倍甚至还不止。

微众银行数据也显示，截至2026年7月，该行日均Token消耗量已超过200亿，每天人均Token消耗量520万，且仍保持快速上升趋势。

业内人士普遍认为，今年，银行业Token消耗量呈现爆发式增长。

从结构差异来看，零壹财经方面认为，民营银行增速最快，股份行紧随其后，国有大行绝对量大但增速相对稳健，中小银行仍在起步阶段。这种分化背后是战略路径的分野：增速快的银行将AI视为“重构业务模式”的核心引擎，愿意承担短期高成本换取长期结构性优势；增速慢的更多将AI作为效率提升的补充手段，注重投入产出比。

从业务层面来看，北京大学光华管理学院院长、北京大学博雅特聘教授田轩进一步表示，股份行Token消耗主要由业务部门产生，追求业务场景的深度渗透与效率提升；国有大行凭借庞大的零售客群和全牌照业务场景，Token消耗绝对量领先，以“场景覆盖”为驱动，聚焦高频、高价值场景的规模化落地；部分中型银行则处于“考核驱动型”，更注重短期业绩考核与AI应用指标的达标，Token消耗增长迅猛但场景深度有待加强。

与外资银行比，科技部国家科技专家库专家周迪指出，国内银行增长呈现两个明显特点：一是增长速度更快，很多银行半年就实现了外资银行一两年的增长幅度；二是应用更偏向零售和普惠金融，比如客服、信贷审批这些高频场景，而外资银行更多集中在投行、风控等高端复杂场景。“这种差异主要是因为国内银行正处于AI规模化应用的爆发期，而外资银行相对更成熟，增长更平稳。”

整体来看，零壹财经方面认为，银行业Token消耗将从“野蛮生长”转向“结构性增长”。短期看增速依然强劲，特别是中小银行将进入快速追赶期；中期看增速趋于理性，增长驱动力从“场景数量”转向“场景深度”。

为何银行越来越关注Token的消耗量？

零壹财经方面指出，Token原本是大模型处理文本的基本计算单元，是纯技术术语。2025年招商银行等开始首次披露Tokens吞吐量，到2026年上半年多家银行在股东会上密集披露Token数据，该指标迅速成为衡量AI应用深度的“新标尺”。对银行而言，AI定位正在从“技术投入”升级为“生产要素”，传统指标衡量“投入端”，Token消耗量衡量“使用端”，即AI实际被调用的频率和深度，Token正成为智能时代的“水电气”。

在田轩看来，银行密集披露Token消耗量，标志着AI已从“技术验证期”迈入“规模化应用期”，Token消耗不再是技术投入的象征，而是业务产出的直接映射。

田轩进一步指出，过去Token只是模型调用的计量单位，如今它直接关联业务效率、客户体验和收入增长，成为衡量银行数字化成熟度的关键标尺。这反映出银行AI战略的底层逻辑已从“工具赋能”向“价值创造”转变，即AI不再仅仅是后台的辅助工具，而是直接嵌入核心业务流程，成为驱动业务增长的关键引擎。

周迪补充道，这背后也是银行AI战略的底层逻辑大转弯：从“为了AI而AI”的技术驱动，转向“用AI赚效益”的业务驱动。不再只看用了多少AI，更要看AI用在什么地方、带来多少实际价值，高层也开始追问“这些Token到底烧出了什么”，要求建立投入产出比评估机制。

从“野蛮生长”到“精打细算”

用技术手段监控和优化消耗，最终让AI从“成本中心”变成“利润中心”。

随着Token消耗量的持续增长，优化结构、压降成本，也正在成为银行AI布局的关注重点。

田轩认为，当前行业正从“拼规模”向“拼效率”切换，日均Token消耗量反映的是AI应用的广度，而单位Token产出价值才是衡量AI战略质量的终极标尺，反映每消耗百万Token带来的收入增量、成本节约或客户体验提升。

从成本来看，日均Token消耗量的成本支出也在快速增长。每日上百亿的Token消耗量是何概念？若按国产大模型平均8元/百万Token计算，日均100亿的Token消耗量，意味着每天至少要支出8万元，且Token单位成本还会随着消耗量级增加而递增。

田轩认为，在净息差持续承压的背景下，Token支出快速增长给银行带来的成本压力不容忽视。他指出，Token支出激增虽绝对值占营收比重尚低（通常不足1%），但其指数级增长趋势已引发管理层对“算力通胀”的警惕。若缺乏管控，AI运营成本可能侵蚀数字化转型的红利。

因此，周迪认为，下一步银行会更注重“精细算钱”，比如给不同业务场景设定Token预算，用技术手段监控和优化消耗，最终让AI从“成本中心”变成“利润中心”。

从成本优化角度来看，银行也在通过基础建设与架构优化，降低成本消耗。毕马威在《2026年中国银行业调查报告》中明确，在宏观演进的背景下，中国银行业在基础设施与架构选型上全面摒弃了对单一全能基础大模型的依赖，确立了大规模参数模型与轻量化小模型双重支撑的混合协同架构。这种架构在保障业务输出质量的同时，实现了算力资源消耗的最优化配置，解决了商业银行实际部署中的成本制约。

微众银行方面也表示，AI基础设施建设是AI走向规模化应用的关键。围绕算力、模型和工具能力建设，微众银行打造统一的AI工程化平台，通过经济高弹性的混合算力架构、灵活可插拔的大模型适配能力、各类可复用的低代码工具及Skills，推动AI从少数技术团队的试验场，转变为全行可调用、可组合、可复用的基础能力。“随着AI应用逐步进入真实业务流程，微众银行也在持续探索AI原生组织转型。”

田轩强调，随着模型架构持续优化、推理芯片快速迭代以及开源生态日益成熟，Token的边际成本将加速下降。

周迪也认为，未来一两年Token单价下降是大概率事件，一方面是大模型技术成熟带来的成本下降，另一方面是银行自建算力、优化模型带来的边际成本降低。

在此行业背景下，田轩认为，随着AI应用从试点走向规模化，预计Token成本管控将呈现三大趋势：一是从“粗放式总量控制”转向“精细化单元经济模型”，重点监控单客户、单交易的Token边际收益，实现每笔业务消耗的Token成本与收益的实时匹配；二是AI运营中台将逐步普及，通过统一监控、调度和优化，降低整体Token浪费；三是银行与模型厂商的合作模式将从“按量付费”向“效果付费”演进，进一步推动成本与价值对齐。

周迪认为，银行该做的规划很明确：先把资源集中到高价值场景，比如智能风控、反欺诈、精准营销这些能直接赚钱的领域；再优化模型和流程，减少闲聊、重复查询这些无效消耗；最后可以考虑自建或联合共建算力，降低长期成本，别再盲目追求Token数量的“面子工程”。

在周迪看来，行业竞争核心指标正在从“比谁用得多”向“比谁用得值”切换，单位Token产出价值正成为新的“分水岭”。

筑牢“三道防线” 监管完善银行保险网络安全管理

中经记者 慈玉鹏 北京报道

为加强银行业金融机构、保险业金融机构和金融控股公司网络安全监督管理，规范网络安全工作，金融监管总局起草了《银行业保险业网络安全管理办法（征求意见稿）》（以下简称《办法》），近日面向社会公开征求意见。

《中国经营报》记者采访了解到，《办法》出台标志着金融业网络安全治理从“合规驱动”正式迈入“责任驱动”与“实战驱动”并重的新阶段，其突破性在于构建了网络安全主管部门、网络安全风险管理部门、审计部门“三道防线”，明确了权责清晰、覆盖全面、分级管理的现代化治理体系。

明确多层责任人

《办法》总体要求指出：“金融机构应当依法履行网络安全保护义务，建立网络安全治理架构，构建网络安全管理体系，开展网络安全风险管理，与业务连续性管理、外包风险管理、数据安全管理等体系机制有效衔接，将网络安全风险纳入全面风险管理，确保网络安全管理能力有效支撑业务安全稳健运行。金融机构应当将境内外分支附属机构网络安全工作纳入整体网络安全管理体系。金融机构应积极配合公安机关、国家安全机关依法开展网络安全保卫、防范违法犯罪活动和监督检查，并提供必要技术支持和协助。”

值得注意的是，《办法》要求金融机构建立网络安全责任制，党委（党组）、董（理）事会对网络安全管理工作负主体责任。金融机构主要负责人为网络安全第一责任人，分管网络安全的领导班子成员（高级管理人员）为直接责任人。金融机构应当明确各层级网络安全责任，明确违规情形和责任追究事项，落实问责处置机制。

长三角科技产业金融研究联盟秘书长陆岷峰告诉记者，《办法》最大的亮点之一是确立了“党委（党组）领导下的网络安全负责制”，并且将主要负责人的责任法定化，这就从根本上解决了长期以来网络安全在董事会层面“听得多、议得少、责不明”的治理上的难题，迫使管理层将网络安全视为与信贷风险同等重要的生存底线。

中国银行（601988.SH）深圳市分行大湾区金融研究院高级研究员曾圣钧告诉记者，网络安全的责任体系从“技术层面责任”升级为“治理层面工程”，首次以规章形式明确“金融机构主要负责人为网络安全第一责任人”。过往监管文件多将网络安全归口于信息科技部门，责任停留在技术层面。《办法》第七条明确规定：党委（党组）、董（理）事会负主体责任，主要负责人是第一责任人，分管领导是直接责任人。这一规定将网络安全从科技条线工作提升至公司治理层面，相较以往政策，适用范围也从数据安全扩展至整个网络安全领域，力度更大。

同时，《办法》正式确立“三道防线”协同共治机制，以法定形式明确三道防线的独立职责与协同关系。曾圣钧告诉记者，网络安全主管部门（第八条）作为第一道防线，负责归口管理、技术保障、应急协调；网络安全风险管理部门（第九条）作为第二道防线，明确要求“保持独立性”，负责风险识别、评估、计量、监测；审计部门（第十条）作为第三道防线负责网络安全审计，允许委托第三方机构。《办法》从制度层面强制风险管理与审计职能独立，构建“决策—管理—执行—监督”的完整闭环，体现“大网络安全”治理理念。

曾圣钧还指出，《办法》实现全集团统一管理与穿透式监管，境内外分支附属机构全部纳入统一管理体系。《办法》第四条明确要求，将境内外分支附属机构网络安全工作纳入整体网络安全管理体系，风险评估、渗透测试范围必须覆盖境内外分支机构和附属机构（第四十一条）。这针对了部分金融机构以往“总部严、子公司松”“境内严、境外松”的管理断层问题。

值得注意的是，《办法》针对关键信息基础设施（CII），设置了比一般机构严的保护标准。陆岷峰指出，《办法》不仅强制要求境内运维、7×24小时监测及年度攻防演练，更明确了“1小时上报”的应急响应时限，这种“抓关键少数”的监管思路，有效应对了金融基础设施故障可能引发的社会性“多米诺骨牌效应”。因此，《办法》目前已经将网络安全工作，从一个单纯的技术运维议题，已经上升到法人治理的核心议题，从而为金融数字化转型筑牢了制度底座。

或可实施差异化分类指导

近年来，金融网络安全建设愈发重要。

从风险特点上看，陆岷峰指出，供应链“投毒”风险日益凸显，金融机构对第三方服务商及开源组件的依赖度极高。Log4j类漏洞的爆发暴露出，单一组件缺陷足以引发全局性瘫痪，且外包人员权限管理失控已成为数据泄露的高频诱因。同时，API接口安全成为新的短板，在开放银行模式下，账户、支付、风控等核心能力通过API向外输出，“内网”概念被彻底击穿，攻击者往往利用API未授权访问或业务逻辑漏洞，绕过传统防火墙进行批量数据爬取或发起撞库攻击，这类行为常被误判为正常业务流量而难以察觉。

一位地方银行人士表示，勒索软件攻击正从“加密勒索”转向“数据窃取+敲诈勒索”的双重勒索模式，加之APT组织对国家金融基础设施的长期觊觎，使得网络安全威胁不仅关乎数据保密，更直接关联国家经济安全与社会稳定，这对金融机构的动态防御能力提出了一定挑战。

而《办法》的推出，能够助力金融网络安全建设完善。但与之同时，落地或存在一定问题需要破解。

陆岷峰指出，《办法》提供了较为完善的制度框架，但在落地执行层面，中小金融机构的资源禀赋与监管要求之间的错配及跨部门协同的“软性壁垒”，是当前推进工作中的最大难点，需要通过差异化监管与内生机制改革来加以化解。

陆岷峰表示，目前众多城商行、农商行及小型保险机构，面临着“一把手有责无权、科技部门有人无力”的窘境，由于受到预算有限与技术人才匮乏等原因，很难独立搭建符合《办法》要求的威胁情报分析系统与灾备体系，也容易掉入“纸面合规”的形式主义陷阱中。同时，“业务为先”的传统惯性也可能导致安全防线难以前置，业务部门为抢占市场往往要求“特事特办”，在安全审查上寻求豁免，致使第九条要求的风控独立性在执行中被稀释，从而形成“业务跑得快、安全追不上”的内耗局面。除此之外，针对重大事件“1小时上报”的硬性指标，对于机构的监测预警与研判能力也提出了极高要求，中小机构往往因缺乏自动化工具而错过最佳处置窗口。

而在曾圣钧看来，政策实施过程中或面临的难点包括：一是治理架构重塑成本较高。涉及部门重组、权责重新划分、利益格局调整，组织变革会有阻力。二是适配的专业人才短缺，网络安全人才供需缺口巨大，既懂金融业务又懂安全合规管理的复合型人才尤为稀缺。三是存量系统改造任务繁重，大量存量信息系统建设年代久远，安全基线不达标，网络架构不符合安全域划分要求，访问控制策略粗放。全面整改涉及系统重构、网络改造、数据迁移，投入大、周期长，且不能影响正常业务运营，实施难度高。

曾圣钧建议，金融机构突破路径建议顶层设计先行，分步推进架构重塑。根据《办法》要求优化高层议事决策机制，明确建立“党委（党组）—董（理）事会—主要负责人—直接责任人”的责任体系，并与考核评价机制相衔接；同时，强化网络安全主管部门、风险管理部门、审计部门的职能与职责，制定三道防线网络安全责任清单，完善跨部门协调工作机制；分级分类推进技术体系建设，优先保障关键信息基础设施和核心业务系统达标，按系统重要性分批次整改，重构网络安全架构，逐步替代传统边界防护模式。

陆岷峰则建议，监管层应当实施差异化的分类指导，或可允许中小机构通过加入行业联合安全运营中心（SOC）或采用监管认可的云服务来履行部分条款；机构内部则需重塑考核激励机制，将安全合规指标纳入业务部门KPI，确立风险管理部门的“一票否决权”，并通过常态化的“不预告”实战攻防演练，将《办法》书面规定的要求转化为全员的自觉行为，真正实现从“被动合规”向“主动防御”的质的转变。