引用格式：张夏恒，刘彩霞. 词元数据跨境流动模式、风险与治理研究[J]. 技术经济，2026，45(9)：29-38.

Zhang Xiaheng，Liu Caixia. The patterns，risks，and governance of cross-border token data flow[J]. Journal of Technology Economics，2026，45(9)：29-38.

词元数据跨境流动模式、风险与治理研究

张夏恒[1]，刘彩霞[2]

（1. 浙江大学中国数字贸易研究院，杭州 310058；2. 中共宁夏区委党校经济学教研部，银川 750021）

摘 要：随着人工智能技术的迭代升级与数字经济全球化的深度推进，词元作为大模型信息处理的最小计算单元与核心价值载体，已成为人工智能时代新型跨境数据形态。词元数据具有算力依附性、衍生动态性、批量规模化、价值复合性等特点，其跨境流动打破了传统数据治理的规则体系和监管框架。本文系统界定词元数据的核心内涵、数据属性与跨境流动本质，梳理当前中国词元数据跨境流动的主流模式与实践形态，深入剖析其在国家安全、数据安全、产业安全、合规体系、技术伦理等维度的新型风险，厘清当前词元数据跨境流动治理存在的制度短板、监管盲区、技术短板与国际规则适配难题。基于统筹发展与安全的核心原则，从制度体系、监管机制、技术赋能、产业协同、国际治理五个维度构建适配中国人工智能产业发展的词元数据跨境流动治理体系，为规范词元数据跨境流转、防范新型数据风险、推动中国人工智能算力合规出海、参与全球数字治理规则构建提供理论支撑与实践路径。

关键词：词元数据；数据跨境流动；数据安全；数字治理；人工智能；安全风险

中图分类号：G203 **文献标志码**：A **文章编号**：1002-980X(2026)09-0029-10

DOI：10. 12404/j. issn. 1002-980X. J26062003

一、引言

全球数字经济进入人工智能(AI)深度赋能的全新阶段，各类大模型产业化推动数据要素形态、流转模式与价值实现方式发生颠覆性变革。传统数据要素以结构化业务数据、个人信息、文本资料为核心，而 AI 算力服务普及催生了新型基础数据单元，即词元(Token)。词元不仅是智能时代的价值锚点，更是连接技术供给与商业需求的结算单位，为商业模式的落地提供了可量化的可能①。词元作为大模型文本拆分、信息计算、内容生成的最小单元，是衡量 AI 算力消耗、量化智能服务价值的核心计量指标，成为数字经济全新的价值载体与贸易标的。在全球 AI 产业竞争日益激烈趋势下，算力资源、电力资源与数据资源的区域重构推动词元数据跨境流动成为数字贸易新赛道。中国凭借稳定的绿电供给、完善的算力基础设施、规模化的大模型产业体系，形成了全球领先的词元生产与输出能力。中国词元日均调用量从 2024 年初的 1000 亿跃升至 2026 年 3 月的 140 万亿，两年间增长超千倍。汕头等试点区域已实现日均百亿级词元跨境调用，词元出海进入规模化落地阶段②。不同于传统数据跨境传输[1-2]，词元跨境流动本质是 AI 算力服务的跨境输出，突破了传统能源、算力资源无法直接跨境贸易的局限，成为中国数字经济对外开放的新形态。与此同时，词元数据的新型属性对现有数据跨境治理体系形成强烈冲击。《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》和《数据出境安全评估办法》主要针对传统结构化数据、个人敏感信息、政务数据等进行规

收稿日期：2026-06-20

基金项目：国家社会科学基金一般项目“跨境电商推进我国数字贸易强国建设机制与路径研究”(22BJY014)

作者简介：张夏恒(1982—)，博士，浙江大学中国数字贸易研究院教授，研究方向：数字经济；刘彩霞(1980—)，博士，中共宁夏区委党校副教授，研究方向：数字经济。

① 人民日报，《“Token”中文名定了：词元》，2026-03-25，https://www. nda. gov. cn/sjj/swdt/mtsy/0325/20260325113219152129329_pc. html。

②人民日报，《我国日均词元调用量突破 140 万亿》，2026-03-24，https://paper. people. com. cn/rmrb/pc/content/202603/24/content_30147015. html。

制[3-4]，尚未针对词元数据的算力依附性、衍生不确定性、批量流转性等特征建立专项治理规则。全球范围内，美国自由流动导向的跨境数据规制模式与欧盟严格保护导向的跨境数据规制模式形成了二元对立格局[5-6]。一方面，词元跨境流动能够释放中国算力产业优势、拓展全球数字贸易市场；另一方面，无规制的词元跨境流转可能引发数据泄露、信息滥用、影响国家安全、产业规则失衡等多重风险，新业态的快速发展与治理体系的滞后性矛盾日益凸显。在此背景下，系统研究词元数据跨境流动的模式、风险与治理路径，成为统筹中国AI产业全球化发展与数据安全、国家安全的重要课题。

关于数据跨境流动的研究，一方面，关注到数据跨境流动的影响。例如，数据跨境流动对经济增长的影响[7]、劳动生产率的影响[8]、抑制国际贸易逃税的影响[9]、数字服务出口的影响[10]等。另一方面，关注到数据跨境流动的规制。例如，中国数据跨境流动监管规则[11]、数据跨境流动国际规则[12]、中欧数据跨境流动合作法律机制[13]、数据跨境流动的治理困境[14]等。然而，学界罕有关注词元数据跨境流动议题，较少探讨如何对其进行规制。关于词元的研究，学界开始对其关注，更多聚焦词元经济的研究，包括词元经济的内涵[15-16]、产业逻辑[17]、生成基础[18]、基本规律[19]、价值机制[20]及词元交易市场[21]等。整体上看，词元逐渐成为学界关注热点，但对词元数据乃至词元数据跨境流动仍存在研究空间。这正是本文研究的突破点与探讨的核心议题。

二、词元数据跨境流动的内涵诠释

数字技术迭代推动数据要素跨境流通常态化，词元数据作为新型基础数据形态，跨境流动场景愈发广泛。无论词元数据还是词元数据跨境流动，都属于新生事物，需要对其内涵进行诠释，尤其需要厘清词元数据跨境流动与传统跨境数据流动的差异。

（一）词元数据的概念界定

词元是AI大模型进行文本解析、语义理解、逻辑运算、内容生成的最小基础计算单元，是大模型算力消耗、任务执行、价值量化的核心载体。从数据形态来看，词元数据并非独立的原始数据，而是原始文本、语音、图像等输入数据经大模型分词、编码、量化处理后形成的标准化计算数据，是原始数据的衍生数据与加工数据。从价值属性来看，词元数据的生成、运算、输出全程依附于AI算力，其价值直接对应算力消耗、电力资源消耗与模型算法成本，是AI算力贸易的核心计量标的，具备可量化、可交易、可流转的商品属性。区别于传统数据，词元数据呈现出显著的特征：一是算力依附性，词元数据无法脱离大模型算力和电力资源而独立存在，其生成与流转全程依托算力基础设施；二是衍生动态性，词元由原始数据实时加工生成，无固定形态，输出内容随算法、场景、输入数据动态变化，具备极强的不确定性；三是批量规模化，词元跨境流转以百万、亿级为单位，呈现批量、高频、实时流转特征，远超传统数据流转规模；四是价值复合性，词元同时承载数据价值、算力价值、能源价值与技术价值，是多重价值的复合载体。

（二）词元数据跨境流动的内涵界定

词元数据跨境流动指境内算力基础设施依托各类大模型，对境内外用户输入的原始数据进行加工处理生成词元数据，并通过合规网络通道完成词元运算结果的跨境传输，为境外用户提供AI服务的全部流程。词元数据跨境流动本质并非传统意义上的原始数据跨境传输，而是境内算力与AI服务的跨境输出，实现了境内算力生产、境外场景消费的新型数字贸易模式。

词元数据跨境流动是数字经济与AI技术深度融合下诞生的新型数据跨境形态，区别于传统数据跨境传输模式，具备独特的技术属性、流通逻辑与贸易特征。词元数据跨境流动是指以境内合规算力基础设施作为核心支撑载体，依托各类通用、垂直领域大语言模型及AI算法系统，针对境内外用户在交互过程中输入的文本、指令、信息片段等各类原始数据，进行清洗、拆分、编码、语义解析、重组运算等系列智能化加工处理，最终生成标准化、结构化、可运算的词元数据，并依托国内合规网络通道、跨境数据传输专线等规范路径，将词元数据的运算结果、智能应答内容、分析成果等进行跨境传输，以此为境外企业、非盈利机构、个人等各类主体提供智能问答、内容生成、数据分析、场景赋能等AI服务的完整业务内容。从核心本质来看，词元数据跨境流动彻底突破了传统数据跨境的核心逻辑，并非简单的原始用户数据、底层信息资料的跨境转移与流

通。与此同时，词元数据跨境流动的本质是境内算力资源与 AI 服务的一体化跨境输出，重构了传统数字贸易的发展模式。在这一流动模式中，数据加工、智能运算、模型推演等核心生产环节全部依托境内算力资源完成，充分发挥了境内 AI 技术、算力基础设施的产业优势；最终的服务成果、运算结果则面向境外场景落地应用，形成了境内算力生产、境外场景消费的新型数字贸易格局。这种新型数据流动形态，既实现了中国算力资源、AI 技术能力的商业化输出，也通过合规可控的数据流转方式，平衡了数字贸易发展与数据安全、个人信息保护的关系，是当下 AI 跨境服务、数字贸易全球化发展的核心形态之一。

（三）词元数据跨境流动与传统跨境数据流动的对比

词元数据跨境流动与传统跨境数据流动存在显著差异（表 1）。传统数据跨境流动以原始数据为流转核心，存在数据泄露、信息滥用、隐私侵权等安全风险，监管重点聚焦原始数据的出境合规与安全管控。词元数据跨境流动的核心载体是经过 AI 模型深度加工后的衍生数据，原始输入数据仅作为运算素材留存于境内算力系统，并未发生跨境流动，有效规避了原始数据出境的各类安全隐患。

从数据形态维度来看，传统跨境数据流动的客体多为静态结构化原始数据，涵盖用户信息、交易数据、政务数据、企业经营数据等常规数据类型。这类数据经过人工规整的标准化录入，数据结构固定、格式统一，内容不会随时间任意变动，具备极强的稳定性，且每一条数据的来源、生成时间、流转路径均可精准追溯，溯源取证难度较低。词元数据是 AI 模型对原始文本、语音、图像等多源数据进行拆分、解析、编码后形成的动态衍生加工数据，并非原生原始数据。其形态具备极强的可变性，可以根据模型训练需求、算力调度场景、指令参数进行实时重组、迭代与生成，无需依托固定数据模板，呈现出即时生成、动态更新、非结构化的典型特征，数据形态无固定标准，溯源难度大幅提升。

从数据流转核心维度来看，传统跨境数据流动的核心逻辑是原始数据本身的跨境传输与存储，核心客体是数据内容，本质是数据资源的跨境流转，整个流程围绕数据的迁移、存储、使用展开。词元数据跨境流动的核心发生了本质转变，不再以原始数据跨境转移为核心，而是以境内智能算力、算法服务的跨境输出为核心，词元数据仅作为算力服务落地的载体存在。在大模型跨境服务、云端算力共享等场景中，境内依托本土数据训练的模型算力向外输出，词元数据随算力调度同步流转，实现了从数据出海到算力出海的模式升级。

从流转规模维度来看，传统跨境数据流动的流转节奏平缓、体量可控，多为单次的、定向的、小批量的数据传输，日常流转频次较低，多在业务对接、数据同步等特定场景产生流转行为，整体流转量级可精准统计、风险可控。与之相对，词元数据依托 AI 持续训练、实时交互、云端推理等场景，实现了日均百亿级的超高流转量级，且全程保持高频、实时、批量的传输状态。这种全天候、自动化、规模化的流转模式，彻底颠覆了传统数据跨境流动的低频可控特征，海量数据瞬时流转，难以实现逐批次精准监管。

从风险特征维度来看，二者的风险特征差异显著。传统跨境数据流动的风险集中在数据本身，以个人隐私泄露、原始数据被非法篡改、无资质违规传输等基础数据安全风险为主，风险影响范围局限于数据本身及对应主体。词元数据跨境流动的风险更具系统性、产业性和战略性，主要体现为算法滥用导致的决策偏差、跨境算力输出催生的算力霸权、动态衍生数据无节制迭代引发的数据失控，以及国内外算力、数据规则不对等造成的数字产业规则失衡，风险辐射范围覆盖个人、企业乃至整个数字产业与国家数字主权。

从合规体系维度来看，传统跨境数据流动经过长期发展，已构建起成熟完备的合规体系，涵盖数据出境安全评估、跨境数据合同备案、数据合规认证等标准化规制规则，形成了权责清晰、流程规范的监管框架，可以

表 1　词元数据跨境流动与传统跨境数据流动的对比

对比维度	传统跨境数据流动	词元数据跨境流动
数据形态	静态结构化原始数据，形态固定、可追溯	动态衍生加工数据，形态可变、实时生成
流转核心	原始数据的跨境传输与存储	境内算力服务的跨境输出，数据为算力载体
流转规模	单次流转量级可控、频次较低	百亿级日均流转、高频实时批量传输
风险特征	隐私泄露、数据篡改、违规传输	算法滥用、算力霸权、衍生数据失控、产业规则失衡
合规体系	具备完善的出境评估、合同、认证规则	无专项规制规则，适配传统合规体系难度大

有效适配传统数据流转场景。词元数据跨境流动属于新型数字业态,目前国内尚无针对性的专项规制规则。现有合规制度基于传统静态数据流转制定,无法适配其动态、算力载体化、超大规模的流转特征,传统合规体系的适配性、有效性大幅下降,存在明显的监管空白与合规难题。

三、词元数据跨境流动的主要模式

依托中国算力基础设施、绿电资源优势与试点政策支撑,当前国内词元数据跨境流动已形成多种实践模式。结合目前试点区域落地实践与词元行业特征,根据流转场景、运作机制、合规路径的差异,将中国词元数据跨境流动划分为终端产品搭载出海模式、离岸算力服务跨境模式、应用编程接口(API)接口中转模式、定制化AI服务跨境模式四种类型。各类型具备差异化的运行逻辑、适用场景与实践特征。

(一)终端产品搭载出海模式

终端产品搭载出海模式是目前国内落地最成熟、合规体系最完善、市场化可行性最高的词元出海路径,有效破解了传统AI算力出海的数据合规难题与市场推广壁垒,其中汕头华侨试验区的先行试点实践,为全国该模式的规模化推广提供了可复制、可落地的标杆范例。随着全球消费级智能硬件市场持续扩容,国产AI终端出海需求激增,该模式凭借轻量化、高合规、快落地的核心特质,成为国产大模型技术出海、抢占全球消费级AI市场的核心渠道。

终端产品搭载出海模式拥有清晰完整的闭环运行逻辑,打通了国内算力生产、境外场景消费的跨境AI服务链路。国内深耕AI智能硬件领域的企业,依托成熟的跨境电商产业体系,研发生产AI玩具、智能服务机器人、家用智能交互终端等各类消费级跨境产品,在硬件出厂前统一搭载自主可控的国产大模型交互系统,摒弃海外第三方模型接口,实现核心技术自主化。境外消费者通过跨境电商平台购买终端产品后,可在本地使用设备发起语音交互、智能问答、内容生成等各类AI指令。所有用户交互指令数据通过该模式经营主体搭建的专属合规跨境数据通道,安全回传至境内合规算力机房,由境内图形处理器(GPU)算力集群完成词元数据解析、运算推演、智能内容生成等全流程工作,最终将处理完成的交互结果实时反馈至境外终端设备,全程实现核心词元算力、数据生产环节留存境内,消费应用场景落地海外。在商业模式层面,终端产品搭载出海模式优化了定价与盈利逻辑,极大地降低了国产AI产品海外拓展的客户门槛。终端产品搭载出海模式将大模型词元服务成本、算力运维费用整体打包融入终端产品售价,境外用户购买设备后即可免费享受全程AI交互服务,无需单独支付词元调用、流量续费等额外费用。这种一体化定价模式简化了海外用户的消费流程,降低了市场教育成本,大幅提升了国产AI智能硬件的性价比与市场竞争力,助力国内硬件企业依托产品出海,带动国产大模型技术同步输出,形成“硬件+算力服务”的一体化出海业态。

从业态发展来看,终端产品搭载出海模式优势与短板特征十分鲜明。其核心优势在于业态轻量化、落地周期短、合规链路清晰,无需搭建复杂的海外算力布局,适用于AI终端、智能硬件、轻量化消费级AI服务等大众场景,能够快速实现市场化落地。目前该模式主要辐射东南亚、中东等新兴海外市场,适合当地宽松的监管环境与消费级AI需求。同时,该模式短板也较为突出,整体业态相对单一,仅聚焦消费级轻量化场景,算力服务附加值偏低,无法适配大型企业级定制、专业级高精度算力服务等高端场景。此外,面对美国、欧盟等国家和地区严苛的数据跨境监管体系,该模式的适配性较弱,难以快速打开美国、欧盟等国家和地区高端市场,后续仍需依托政策迭代与技术升级,持续拓宽业态边界与市场覆盖范围。

(二)离岸算力服务跨境模式

离岸算力服务跨境模式是中国算力产业出海的全新规模化路径,也是适配数字贸易合规发展的企业级算力服务新模式。该模式聚焦企业级市场,以专业化、规模化的词元出海为核心载体,依托国内建成的离岸算力中心、官方认证的跨境算力试点平台搭建合规服务体系,主要面向境外企业、科研机构、跨境服务商等主体,提供大模型推理、弹性算力租赁、批量词元调用等定制化、一体化AI算力服务,重构了传统国产大模型出海的商业模式与收益格局。

离岸算力服务跨境模式拥有成熟且合规的闭环运行逻辑,形成了境内整合、跨境服务、市场化结算的完整链路。在资源整合端,境内合规离岸算力平台集中对接阿里千问、DeepSeek、豆包等国内主流成熟国产大

模型，深度整合国内风电、光伏等绿色电力算力资源与优质算力节点，依托国内完善的算力基础设施网络，搭建高稳定、高算力储备的服务底座。在服务输出端，平台精准对接境外科技企业、跨境服务机构、海外内容创作团队、跨境电商企业等多元客户群体，提供标准化、批量式的词元调用核心服务。境外客户无需搭建本地算力与模型体系，主要通过专属跨境合规通道安全提交算力调用、模型推理等需求，全程规避跨境数据传输风险。境内平台接收需求后，依托本土算力资源完成词元生成、数据运算、模型推理等全流程处理，最终向境外输出标准化、可落地的 AI 服务成果，并根据客户词元调用量级、服务类型实行市场化阶梯结算。离岸算力服务跨境模式实现了境内电力、境内算力、境内模型的全链条本土化落地，词元产品从算力供给、模型运算到服务产出全部在国内完成，真正实现词元产品的本土生产、全球出口，充分释放中国绿电资源充足、算力成本低廉、大模型技术成熟的核心竞争优势，让算力产业的核心收益全部留存国内，大幅提升中国数字产业的对外价值转化率。离岸算力服务跨境模式应用场景比较广泛，全面覆盖各类企业级跨境 AI 服务场景，具备极强的市场适配性。在 AI 内容生成领域，能够助力境外机构完成多语种文本创作、图文生成、视频脚本制作等；在跨境智能客服领域，能够为跨境企业提供 7×24 小时多语种智能应答服务。

从产业特性来看，离岸算力服务跨境模式具备规模化程度高、产业附加值高、盈利模式清晰的突出优势，能够快速形成产业化、规模化的出海效应，助力中国算力产业打造全新外贸增长点。但同时该模式也存在明显发展短板，对核心配套条件要求严苛，不仅需要顶尖的算力基础设施、稳定合规的跨境传输通道、低延迟的跨境网络体系，还需要完善的跨境数字合规体系支撑。此外，行业发展面临诸多外部壁垒，境外各国的算力准入规则、数据属地化监管政策存在较大差异，严格的跨境数据监管、市场准入限制，成为制约离岸算力服务规模化扩张的核心阻碍，也是未来行业规范化、全球化发展需要重点突破的关键难题。

（三）API 接口中转模式

API 接口中转模式是当前词元跨境流动领域应用最普及、市场渗透率最高且合规争议最为突出的主流业务模式，广泛应用于国内 AI 算力出海、境外轻量化 AI 服务接入等跨境业务场景。国内大模型算力资源成熟、性价比突出，依托 API 接口中转实现词元数据跨境流转的服务模式应运而生，并迅速成为中小 AI 出海企业的核心业务载体。该模式主要依托专业的 API 中转站、接口聚合服务平台搭建跨境传输通道，打通境内合规大模型算力接口与境外多元化 AI 应用终端，实现词元数据的双向跨境流转与 AI 算力服务的全球化输出。

从运行逻辑来看，API 接口中转模式具备清晰的轻量化运营逻辑。跨境服务商通过搭建中立化、集成化的 API 中转平台，一边对接国内各类通用、垂直领域大模型的算力服务端口，整合文本生成、图像创作、语言翻译、智能问答等多元算力资源；另一边面向海外开发者、中小企业、互联网应用终端开放接口服务。平台通过接口代理、数据中转、请求分发、算力调度等核心技术手段，精准匹配境外用户的 AI 服务需求与境内闲置、富余的大模型算力资源，全程完成高频次、碎片化、小体量的词元跨境调用交互。相较于服务器出海、专线直连等传统跨境算力模式，该模式无需复杂的硬件部署、网络适配和场地运维，接入门槛极低，能够快速适配全球各地的轻量化 AI 应用场景。API 接口中转模式具备鲜明的技术中立性、跨境匿名性、运营隐蔽性与服务去中心化等特征。在技术层面，平台仅承担数据传输与算力调度的中介角色，不直接生产 AI 内容，呈现出中立技术服务属性。在跨境传输层面，平台通过接口封装、链路加密等技术处理，隐藏数据源头与终端用户信息，具备较强的匿名特征。在运营层面，多数中转平台无固定服务场景、无显性业务标识，线上化、轻量化的运营模式具备极强的隐蔽性；同时采用去中心化的服务架构，打破了传统跨境算力服务的地域与端口限制，适配全球海量分散的终端需求。凭借这些优势，API 接口中转模式流转效率极高、场景适配性极强，成为中小 AI 企业出海的最优选择。

API 接口中转模式深度渗透海外大众 AI 服务赛道，覆盖海外短视频智能文案生成、多语种 AI 实时翻译、轻量化智能绘图、日常场景智能问答、海外社交媒体内容优化等海量碎片化场景，支撑了海量中小海外开发者与小微企业的 AI 服务需求，降低了全球轻量化 AI 应用的落地成本。但该模式的合规漏洞与安全风险日益凸显，成为当前中国词元跨境流动风险的核心高发领域。现阶段中国数据跨境、AI 服务监管的现行法律法规与行业规范，尚未针对 API 中转平台的法律定位、业务经营范围、合规审核标准、风险管控要求及

政企多方责任划分作出明确、细化的规定，行业长期处于监管模糊地带。由于合规体系缺失，行业整体呈现无序发展态势，多数中小型API中转平台缺乏完善的数据审核、内容风控、链路溯源机制，大量未经安全审核、合规校验的用户原始指令、敏感词元数据、场景化交互数据通过中转通道无序跨境流转。

（四）定制化AI服务跨境模式

定制化AI服务跨境模式是数字经济全球化背景下，中国AI算力服务出海的高端细分形态，是聚焦高端专业市场的精细化、定制化算力服务出海路径。该模式精准瞄准境外优质高端客户群体，核心服务对象涵盖境外大型跨国企业、顶尖科研机构、跨境集团及高端涉外政务机构，主要承接各类场景复杂、精度要求高、合规性严苛的专业化AI算力需求，是中国AI产业从基础服务出海向高端技术服务出海升级的重要体现。

定制化AI服务跨境模式具备清晰且成熟的闭环运行逻辑，全程围绕境外客户个性化需求落地专业化服务。在服务前期，境内头部AI科技企业、大型算力平台会深度对接境外客户，结合其所属行业属性、具体业务场景、核心算力痛点与长期发展规划，开展全方位需求调研与方案拆解。基于精准的需求分析，团队针对性定制专属大模型算力解决方案，摒弃通用化模板服务，实现一场景一方案、一客户一适配。方案落地阶段，依托国内成熟的AI算力集群与技术体系，为客户提供定制化词元运算、专属模型推理、多维数据智能分析、行业智能算法适配等高端核心服务。目前，定制化AI服务跨境模式广泛落地于工业智能智造、跨境联合科研、国际金融风控、跨境政务辅助、智慧跨境物流等多个高精尖专业领域，适合各类高端跨境业务场景的智能化升级需求。定制化AI服务跨境模式拥有差异化的核心特征，集中体现为定制化、高附加值、低频次、高精度。其词元生成、算力运算、模型推演等功能均具备极强的场景针对性，契合客户专属业务逻辑。从服务属性来看，此类高端定制服务无需大规模批量输出，单次服务的技术价值、产业价值远高于基础算力服务，形成了小体量、高价值、高精度的独特服务形态。定制化AI服务跨境模式普遍采用一事一议的专项合规审核机制，针对每一项跨境服务、每一次数据流转、每一组模型输出，开展独立合规核查。服务全程依托企业跨境数据合规资质认证、国家级数据安全评估、跨境数据流动备案等正规流程完成合规流转，严格规避跨境数据泄露、合规违规、知识产权纠纷等风险，整体合规管控标准远高于普通出海服务，保障高端跨境AI服务的安全性与规范性。

从产业发展维度来看，该模式具备突出的发展优势与亟待突破的短板。其核心优势在于产业附加值极高，能够打破中国传统算力出海的低端同质化困境，实现AI服务的高端化、专业化输出，同时依托严格的合规体系实现风险可控，精准适配全球高端算力市场需求，持续推动中国AI算力产业向价值链上游升级。但该模式存在明显发展短板，高端定制属性导致服务准入门槛高、标准化程度低，难以快速实现规模化复制与批量推广，前期技术研发、合规审核、方案定制的成本居高不下。此外，美国、欧盟等国家和地区高端市场的技术垄断、严格的数据准入壁垒、严苛的知识产权审查机制，也为该模式的国际化拓展带来了多重挑战，是未来高端AI服务出海需要重点突破的核心难题。

四、词元数据跨境流动的多维风险识别

词元数据的算力依附性、动态衍生性、批量流转性等独特特征，使其跨境流动风险区别于传统数据跨境风险，呈现出隐蔽性、复合型、扩散性、长期性的特征。结合四类模式的实践特征，从国家安全风险、数据安全风险、产业安全风险、合规法律风险、技术伦理风险五个维度，系统识别词元数据跨境流动的新型风险。

（一）国家安全风险：数据主权与意识形态安全隐患

词元数据跨境流动衍生新型国家安全风险，主要体现在数据主权弱化与意识形态安全两大层面。一是数据主权弱化风险。词元数据由境内算力、国产算法、本土能源资源生成，承载着中国算力主权与数字要素主权。无规制的跨境批量流转可能导致中国核心数字生产要素零成本外流，弱化中国在全球AI算力竞争中的主权优势。同时，部分API中转模式存在匿名化、隐蔽化流转特征，监管部门无法精准追溯词元流转去向、使用场景与用途，存在核心算力资源被境外非法利用、窃取的风险，导致数据主权管控失效。二是意识形态安全风险。词元数据承载语义信息、价值导向与内容逻辑，境外各类社会舆论、政治观点、意识形态相关的指令数据输入境内大模型，经词元运算生成跨境输出内容，可能出现数据筛选操纵、算法偏差、内容误

导等问题。境外势力会通过隐蔽化词元调用渠道，利用中国大模型算力生成不实信息、负面舆论、意识形态渗透内容[22]，反向回流到境内进行传播，对中国网络舆论生态、意识形态安全造成冲击，甚至引发跨境舆情风险与政治风险。

（二）数据安全风险：衍生数据失控与隐私泄露风险

词元数据的动态衍生性与批量流转性，催生新型数据安全风险，突破了传统隐私泄露、数据篡改的风险边界。一是原始数据存在泄露风险。境外用户输入的指令可能包含个人敏感信息、商业秘密、行业隐私数据。这类原始数据传入境内经词元化处理后，虽形态发生改变，但部分敏感信息仍会通过算法还原、词元拆解实现逆向解析，导致跨境敏感数据泄露。同时，在批量高频的词元流转过程中，海量衍生数据在境内算力平台集聚，若缺乏完善的分级管控、脱敏处理机制，容易形成数据黑洞，引发大规模数据泄漏风险。二是数据流转溯源存在失控风险。传统数据跨境流转通过数据备案、日志留存等实现全程溯源，而 API 接口中转模式下的词元流转具备匿名化、去中心化特征，流转链路隐蔽、调用主体模糊、场景无法界定，一旦发生数据滥用、泄露、篡改等问题，无法精准定位责任主体、追溯风险源头，形成监管溯源盲区。最后，数据安全防护面临冲击风险。词元数据具有动态生成、实时流转的特征，导致传统静态数据加密、脱敏、存储安全防护技术难以适配，进而会加剧数据安全防护体系受冲击的风险。

（三）产业安全风险：全球算力竞争与规则失衡风险

词元数据跨境流动作为 AI 算力贸易的核心业态，直接关联中国数字产业安全与全球 AI 产业竞争格局，会衍生多重产业安全风险。一是产业低端锁定风险。当前中国词元出海多以基础算力输出、轻量化服务为主，处于全球 AI 算力贸易的中低端环节，高端定制化服务、算法解决方案输出占比较低。长期大规模出口基础词元算力，可能导致中国 AI 产业陷入“算力代工”的低端锁定困境，核心算法、高端技术优势难以释放，产业附加值持续偏低。二是国际产业壁垒挤压风险。美国、欧盟等发达经济体出于产业保护与安全管控诉求，构建了严格的 AI 数据、算力跨境准入壁垒，推行数据本地化、算力属地化政策，禁止境外算力跨境处理本土数据，对中国词元出海设置严苛门槛。美国依托技术优势掌控全球大模型核心算法与算力标准，欧盟通过高标准规制体系输出区域规则，但中国在全球词元数据治理、算力贸易规则制定中话语权相对不足，面临规则被动适配、产业发展被动受限的风险。三是产业无序竞争风险。当前词元出海行业准入门槛较低，大量中小企业无序入局，部分企业为抢占市场，规避合规成本，采用无资质 API 接口中转等模式开展业务，引发低价竞争、无序竞争乱象，扰乱国内词元算力产业市场秩序，损害中国词元出海的整体品牌形象与产业竞争力。

（四）合规法律风险：制度适配不足与违规运营风险

当前中国数据跨境合规体系基于传统静态数据构建，与词元数据新型业态严重不适配，引发全方位合规法律风险。一是制度建设不足导致的合规失范风险。中国现行法律法规未明确词元数据的法律属性、数据分类、跨境流转等合规标准，未针对词元批量、实时、动态流转特征制定专项合规规则，传统的数据出境评估、标准合同、认证机制难以适配词元出海场景。大量词元跨境流转行为处于法律真空地带，企业无法可依、监管部门无规可管，合规边界模糊。二是跨境规制冲突引发的合规风险。全球词元数据规制规则碎片化特征显著，美国主张数据自由流动、弱化政府监管，欧盟实行严格的数据保护与准入管控，各国数据属地化、隐私保护、算力准入规则差异巨大。中国企业开展词元跨境业务时，需要同时适配国内安全规制与境外属地监管规则，双重合规成本高、规则冲突问题突出，容易触发境外合规处罚、业务封禁风险。三是主体权责不清引发的法律纠纷风险。词元跨境流动涉及算力平台、API 中转服务商、终端企业、境外用户等多方主体，现行规则未明确各方主体的权责边界、合规义务、责任承担方式。一旦发生数据泄露、内容违规、知识产权纠纷、跨境侵权等问题，容易出现权责推诿、追责困难的局面，引发跨境法律纠纷。

（五）技术伦理风险：算法滥用与内容失范风险

词元数据依托大模型算法生成，其跨境流转与应用可能衍生多重技术伦理风险。一是算法偏见与内容失范风险。大模型算法存在固有训练偏差、语义识别误差[23]，不同国家、区域的文化习俗、价值观念、语义体系差异显著，境外多元指令输入境内模型后，经词元运算可能生成违背公序良俗、文化伦理、社会价值观的

跨境内容,引发文化冲突和伦理争议。同时,境外恶意主体通过精准指令诱导,操控大模型生成虚假信息、低俗内容、偏见内容,造成跨境内容失范传播。二是算力滥用与技术滥用风险。无规制的词元跨境调用,可能被境外恶意主体利用,开展大规模网络爬虫、算力破解、恶意攻击、虚假信息自动生成等违规活动,境内算力资源被滥用为境外违规行为的技术支撑,引发网络安全、技术滥用伦理风险。此外,批量词元跨境流转可能导致算法技术、模型参数隐性外泄,存在核心AI技术被境外窃取、复刻的风险,威胁中国AI技术安全。

五、词元数据跨境流动的治理路径

针对当前中国词元数据跨境流动的实践模式、多维风险,需要立足统筹发展和安全、兼顾开放与管控、平衡产业与安全的原则,构建系统化、精准化、可落地的词元数据跨境流动治理体系,推动中国词元产业规范化和规模化发展,同时筑牢数据安全、国家安全与产业安全底线。

(一)完善顶层制度体系,构建专项规制规则

针对当前立法滞后、适配性不足等问题,加快构建适配词元数据业态特征的专项制度体系,夯实治理制度基础。一是明确核心法律条文界定。通过修订《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》配套规章,出台词元数据跨境流动专项规制文件,明确词元数据的法律属性、数据分类、权利边界与监管定位,区分词元衍生数据与传统原始数据的规制差异,丰富新型数据业态立法内容。二是优化差异化合规机制。结合四类流动模式的特征,构建分类合规体系:对终端产品搭载出海模式,简化合规流程,推行批量备案机制;对离岸算力服务跨境模式,建立专项安全评估制度;对API接口中转模式,明确准入资质与合规标准,严禁无资质违规流转;对定制化AI服务跨境模式,推行一事一议专项审核机制,实现精准合规管控。三是推广试点经验标准化。系统总结汕头等试点的合规闭环经验,提炼可复制、可推广的全链路合规流程、安全标准与管控规范,在全国算力枢纽、数字经济试点区域全面推广,构建全国统一的词元跨境合规标准体系。四是完善配套制度规则。建立词元数据分级分类管理制度,根据流转场景、敏感程度、风险等级划分管控层级;出台词元算力计量标准、脱敏技术标准、跨境传输安全标准,统一行业技术规范;完善词元知识产权保护、跨境交易结算、权益分配相关制度,规范产业发展秩序。

(二)优化协同监管机制,实现全流程精准管控

打破传统静态监管模式,构建事前预防、事中管控、事后溯源、跨域协同的全流程动态监管体系,破解监管滞后、权责模糊等难题。一是明确多元监管权责边界。厘清网信、工信、公安、商务、海关等部门的监管职责,建立网信统筹、多部门协同、属地负责的监管机制,杜绝多头监管与监管盲区。设立词元数据跨境治理专项工作体系,统筹协调产业发展、合规监管、风险防控各项工作。二是构建动态监管模式。转变事后处罚的被动监管模式,建立事前风险评估、事中实时监测、事后溯源追责的全流程动态监管体系,针对高频批量流转特征,实现风险实时预警、动态管控。三是建立行业标准化监管体系。统一词元跨境调用计量、安全审核、数据脱敏、溯源留存的行业标准,规范企业合规操作与监管执法流程,解决监管标准不统一、执法随意性问题。四是健全跨境协同监管机制。积极与重点出海市场国家建立数据治理协同合作机制,搭建跨境风险信息共享、违规行为联动处置、争议纠纷协同解决平台,破解跨境监管壁垒。

(三)强化技术赋能治理,筑牢安全技术屏障

聚焦技术治理短板,加大核心技术研发投入,构建适合词元数据动态流转特征的技术防护与监管体系,以技术创新赋能风险治理。一是研发专项安全防护技术。突破词元实时脱敏、动态加密、敏感语义拦截等核心技术,针对词元动态衍生、批量流转等特征,搭建智能防护系统,实现原始敏感数据实时过滤、词元衍生内容安全审核、跨境传输全程加密,从源头防范数据泄露与内容失范风险。二是搭建智能化监管溯源平台。依托大数据、AI、区块链等技术,构建全国统一的词元跨境流动监管溯源平台,实现对词元调用主体、流转链路、应用场景、消耗量级的全程可追溯、可监测、可预警,精准识别匿名流转、违规调用、算力滥用等风险。三是统一行业技术标准。牵头制定国内词元编码、计量、脱敏、溯源、安全防护的统一技术标准,推动各大模型平台、算力企业技术标准对接兼容,实现行业技术规范化发展。第四,构建技术风控体系。建立词元跨境风险智能化研判模型,基于海量流转数据实现风险自动识别、分级预警、智能处置,提升监管精准度与效率;鼓

励企业搭建内部合规风控技术系统，实现企业自主合规、智能风控，形成政企协同的技术治理格局。

（四）规范产业发展秩序，构建多元共治体系

立足产业高质量发展要求，规范行业发展秩序，破解业态乱象、结构失衡、自律缺失等问题，推动词元出海产业提质增效。一是优化产业结构布局。依托中国算力与绿电优势，推动词元出海从低端轻量化服务向高端定制化、企业级算力服务升级，重点培育工业智能、跨境科研、高端数字服务等高附加值业态，破解低端锁定风险，提升产业核心竞争力。二是完善词元产业生态。统筹全国算力资源，打造规模化词元出海产业集群，培育头部龙头企业，完善上下游产业链配套，构建完整的词元算力出海产业生态。三是规范行业市场秩序。严格落实行业准入机制，取缔无资质、违规运营的 API 中转平台与跨境服务主体，严厉打击低价竞争、违规流转、算力滥用等乱象，净化行业发展环境。四是强化企业合规能力建设。开展企业合规专项培训，出台企业词元跨境合规操作指南，引导企业建立完善的内部合规体系、风险防控制度与溯源管理制度，提升企业自主合规能力。第五，构建行业自律机制。牵头成立全国词元算力产业自律协会，制定行业自律公约、经营规范与信用评价体系，推动行业自我约束、自我规范、协同共治，形成政府监管、行业自律、企业自治的多元产业治理格局。

（五）参与全球规则构建，提升国际治理话语权

立足全球数字治理碎片化格局，主动对接高标准国际经贸规则，积极参与全球词元与 AI 数据跨境治理体系构建，破解国际壁垒、提升中国话语权。一是对接国际高标准规则体系。主动对标全面与进步跨太平洋伙伴关系协定（CPTPP）、数字经济伙伴关系协定（DEPA）等高标准数字贸易协定中的数据跨境流动规则，适度优化中国词元跨境开放政策，在守住安全底线的前提下，放宽合规企业、低风险场景的词元跨境流动限制，提升国际规则适配性。二是深化区域数字治理合作。依托区域全面经济伙伴关系协定（RCEP）框架，加强与东盟、日韩等区域国家的词元算力贸易与数据治理合作，构建区域统一的词元跨境合规标准、流动规则与协同治理机制，打造区域数字治理合作样板，突破美国、欧盟规则壁垒。三是主动参与全球规则制定。依托世界贸易组织、联合国等多边平台，积极参与全球 AI 数据、词元算力跨境流动、数字贸易等相关规则的制定与修订，输出中国治理经验与标准体系，打破美欧规则垄断，提升中国在新型数字要素治理领域的国际话语权。四是构建中国特色治理方案。立足中国产业优势与安全诉求，平衡自由流动与安全保护，构建安全可控、有序开放、分类管控、互利共赢的词元数据跨境流动中国方案，为全球新型数字要素治理提供新路径，提升中国数字治理国际影响力。

参考文献

[1] 杨松，汪宓. 数据跨境流动国际规则的困境及应对[J]. 社会科学辑刊，2025(3)：225-236，239，241.

[2] 梅傲. 数据跨境传输规则的新发展与中国因应[J]. 法商研究，2023，40(4)：58-71.

[3] 丁晓东. 数据跨境流动的法理反思与制度重构——兼评《数据出境安全评估办法》[J]. 行政法学研究，2023(1)：62-77.

[4] 郑曦.《刑事诉讼法》与《数据安全法》《个人信息保护法》衔接问题研究[J]. 中国刑事法杂志，2024(2)：107-123.

[5] 陈元馨，康宁. 美国数据跨境流动规制的国内研究观点分析[J]. 情报杂志，2023，42(4)：132-139.

[6] 金晶. 欧盟的规则，全球的标准？数据跨境流动监管的"逐顶竞争"[J]. 中外法学，2023，35(1)：46-65.

[7] 王立勇，常清，黄京磊，等. 数据跨境流动与经济增长：理论框架与经验证据[J]. 经济研究，2026，61(2)：119-141.

[8] 王恕立，朱梦杰. 数据跨境流动对劳动生产率的影响研究[J]. 经济体制改革，2026(2)：49-57.

[9] 房超，沈雨婷. 数据跨境流动能否抑制国际贸易逃税——来自海关信息交换的证据[J]. 数量经济技术经济研究，2024，41(3)：89-111.

[10] 韦倩青，刘敏. 数据跨境流动限制措施如何影响数字服务出口[J]. 技术经济，2025，44(5)：1-13.

[11] 马其家，李晓楠. 论我国数据跨境流动监管规则的构建[J]. 法治研究，2021(1)：91-101.

[12] 何波. 中国参与数据跨境流动国际规则的挑战与因应[J]. 行政法学研究，2022(4)：89-103.

[13] 舒静怡. 论中欧数据跨境流动合作法律机制的构建[J]. 社会科学战线，2026(3)：222-235.

[14] 陈兵，马贤茹. 系统观念下数据跨境流动的治理困境与法治应对[J]. 安徽大学学报(哲学社会科学版)，2023，47(2)：58-66.

[15] 潘宏亮，杨文光，叶璐. 我国词元经济发展的条件、约束与推进路径[J]. 中国流通经济，2026，40(6)：3-14.

[16] 夏璋煦，徐政. "十五五"时期智能经济新形态下"词元经济"的内涵特征、生成逻辑与推进路径[J/OL]. 重庆大学学报(社会科学版)，1-16[2026-06-26]. https://link.cnki.net/urlid/50.1023.C.20260518.1128.006.

[17] 李国庆. "人工智能+"背景下词元经济的技术内涵、产业逻辑与治理体系[J]. 统一战线学研究，2026，10(4)：99-110.

[18] 蒙慧, 肖丙浩. “十五五”时期词元经济的生成基础、价值机制与治理路径——基于马克思主义政治经济学的分析[J]. 统一战线学研究, 2026, 10(4): 86-98.
[19] 欧阳日辉, 黄芝萱. 词元(Token)经济的内涵特征、基本规律与发展策略[J]. 河海大学学报(哲学社会科学版), 2026, 28(3): 127-142.
[20] 任保平, 冯星源. 词元经济价值实现的逻辑与方略[J]. 改革, 2026(6): 63-74.
[21] 杨小强, 张江朋. 词元(Token)交易市场的微观机制与制度设计——基于中国多地实践的研究[J]. 当代经济管理, 2026, 48(9): 20-28.
[22] 谢俊, 吴阳琴. 智能算法推荐对主流意识形态话语权构建的挑战及其应对[J]. 西南大学学报(社会科学版), 2024, 50(6): 144-154.
[23] 张夏恒. ChatGPT 的政治社会动能、风险及防范[J]. 深圳大学学报(人文社会科学版), 2023, 40(3): 5-12.

The Patterns, Risks, and Governance of Cross-border Token Data Flow

Zhang Xiaheng[1], Liu Caixia[2]
(1. China Digital Trade Research Institute, Zhejiang University, Hangzhou 310058, China;
2. Economics Teaching and Research Department, Party School of the CPC Ningxia Committee, Yinchuan 750021, China)

Abstract: With the iterative upgrading of artificial intelligence (AI) technology and the deepening globalization of the digital economy, Token, as the smallest computational unit and core value carrier for large model information processing, have emerged as a new type of cross-border data form in the AI era. Token data possesses characteristics such as computational dependency, derivative dynamism, batch scalability, and value complexity. Its cross-border flow has disrupted the traditional data governance rules and regulatory frameworks. The core connotation, data attributes, and essence of cross-border Token data flow was systematically defined. The current mainstream models and practical forms of cross-border Token data flow in China was reviewed. Its new risks in dimensions such as national security, data security, industrial security, compliance systems, and technological ethics, and clarifies the existing institutional shortcomings, regulatory blind spots, technological weaknesses, and challenges in adapting to international rules in the governance of cross-border Token data flow were reviewed. Based on the core principle of balancing development and security, this paper constructs a governance system for cross-border Token data flow that is suitable for the development of China's AI industry from five dimensions: institutional system, regulatory mechanism, technological empowerment, industrial collaboration, and international governance. It provides theoretical support and practical paths for regulating lexeme data cross-border circulation, preventing new data risks, promoting China's AI computing power to go overseas in compliance, and participating in the construction of global digital governance rules.

Keywords: Token data; cross-border data flow; data security; digital governance; artificial intelligence; security risks