

汉坤法律评述

2026 年 9 月 14 日

北京 | 上海 | 深圳 | 杭州 | 武汉 | 海口 | 香港 | 新加坡 | 纽约 | 硅谷 | 伦敦

加州集中出台未成年人在线安全及 AI 监管新法：“州级先行”科技立法下的企业合规观察

作者：段志超 | 刘倩倩 | 刘若愚

2026 年 9 月 9 日至 10 日，加州州长加文·纽森（Gavin Newsom）接连签署多项与人工智能（Artificial Intelligence, AI）及未成年人在线安全相关的法案¹。其中，AB 1709 和 SB 1119 两项法案分别针对社交媒体平台的“成瘾性功能”（Addictive feature）以及 AI 陪伴型聊天机器人（Companion chatbot）建立新的未成年人保护制度；AB 1405 和 SB 813 两项法案则基于第三方审计机构和独立验证组织两个维度，建立 AI 风险评估及监督机制。同期签署的还包括提高大型社交媒体平台伤害儿童民事责任的 AB 2、涉及应用年龄验证信号的 AB 1856、要求建立儿童性虐待材料举报机制的 AB 1946，以及针对儿童 AI 陪伴型玩具的 SB 867 等法案。

本次立法发生在美国未成年人在线安全及 AI 安全监管持续受到关注的背景下，在既有内容安全、算法推荐及未成年人保护规则的基础上，上述法案进一步细化对产品功能和默认设置的限制，并加强风险评估及第三方审计、验证机制。同时，在联邦层面尚未形成统一的综合性 AI 监管框架的背景下，加州本轮集中立法延续了“州级先行”的探索路径。州长纽森称此次立法将进一步巩固加州在 AI 监管领域的领先地位，并呼吁联邦政府建立相应的全国性监管规则。

本文选取其中与平台产品设计和 AI 治理架构关系最为直接的四项法案进行介绍，并结合立法背景与各方立场，对相关企业可能面临的合规影响进行简要分析。

一、未成年人保护相关立法：从限制社交媒体“成瘾性设计”到建立 AI 聊天机器人事前安全评估机制

（一）本轮未成年人在线安全立法的背景

围绕社交媒体及 AI 聊天机器人的未成年人安全诉讼，是理解加州本轮立法的重要背景。2026 年 8 月，Meta 就未成年人社交媒体成瘾相关诉讼与美国绝大多数州的总检察长达成和解（尚待法院批准），同意在十年内支付最高 180 亿美元，并对 Instagram 和 Facebook 采取包括未成年人每日使用时长默认

¹ <https://www.gov.ca.gov/2026/09/09/governor-newsom-signs-first-in-the-nation-ai-safeguards-to-protect-californians-calls-on-the-federal-government-to-do-its-part/>; <https://www.gov.ca.gov/2026/09/10/governor-newsom-signs-the-strongest-child-safety-chatbot-and-social-media-laws-in-the-nation/>。

上限、夜间使用限制及非个性化信息流选项在内的新的青少年保护措施²。相关诉讼的核心指控之一，是 Meta 通过产品设计和用户参与机制，诱导未成年人持续使用平台。SB 1119 法案又称为 Adam's Law，以 2025 年因与 ChatGPT 互动后自杀身亡的加州青少年 Adam Raine 命名。其父母于同年对 OpenAI 等被告提起民事诉讼，指控其产品设计及安全保护缺陷，在一定程度上导致了 Raine 的悲剧³。上述案件反映出美国社会目前对未成年人在线安全问题的高度关注，Meta 和解协议中的时长上限、夜间限制等承诺，也与下文 AB 1709 和 SB 1119 的法定要求高度呼应。同时，加州并非在此领域单独行动：纽约州的陪伴型聊天机器人法已于 2025 年 11 月生效，爱达荷、俄勒冈、华盛顿等州亦于 2026 年通过了类似规则，各州立法正呈现趋同态势⁴。

（二）AB 1709：禁止向 16 岁以下用户提供特定“成瘾性功能”⁵

AB 1709 法案的监管重点在于平台的成瘾性功能及相关产品设计。法案禁止受监管平台向 16 岁以下的加州用户提供“成瘾性功能”，包括“成瘾性信息流”（Addictive feed）、自动播放及后续实施规则指定的其他功能。法案将“成瘾性功能”定义为任何通过利用用户心理弱点，旨在最大限度提高用户参与度，并且可以预见会导致用户持续使用的平台功能。受监管平台不限于社交媒体，但相关功能须构成平台服务的重要组成部分，仅涉及商业交易、消费者评价的特定平台，以及主要用于云存储的信息流等，存在法定排除。“成瘾性信息流”主要涉及依据用户或其设备相关信息推荐、筛选或优先展示用户生成或分享的内容，其定义同时设有具体例外，因此并非所有信息流或推荐功能均受同等限制。

同时，AB 1709 法案亦在年龄识别方面，与加州 2027 年 1 月即将生效的《数字年龄保障法案》（Digital Age Assurance Act）衔接⁶，要求受监管平台在提供“成瘾性功能”前通过从操作系统及应用商店等获取用户年龄区间数据，或自行核验用户年龄，从而确保 16 岁以下用户不会被提供相关功能。在账户管理方面，平台可以允许 16 岁以下用户创建或保留不提供“成瘾性功能”的账户；不采用该安排的，应依照法案删除相关账户及关联个人信息。因此，企业除调整具体功能外，还需评估未成年人账户的保留、功能隔离及删除机制。

在执行层面，AB 1709 法案授权加州总检察长制定进一步的实施规则（包括在必要时调整“受监管平台”的范围），并规定仅由总检察长或地方检察机关通过民事诉讼实施执法与请求民事罚款：明知违法的，每名受影响未成年人最高 5 万美元；过失违法的，每名最高 2.5 万美元；法院量罚时须考虑平台规模、违法的严重程度与持续时间以及平台的善意合规努力。法案本身未设私人诉权。法案亦未另行规定生效日期，按加州一般规则将于 2027 年 1 月 1 日生效。法案同时在加州司法部内部设立电子安全咨询委员会（e-Safety Advisory Commission），就在线安全问题向州政府提供建议，并定期向州议会和州长报告。

AB 1709 法案延续并强化了 2024 年通过的 SB 976 法案对成瘾性信息流及相关产品机制的监管⁷。现行的 SB 976 法案主要针对“成瘾性信息流”作出限制，而 AB 1709 则进一步将监管对象扩展至“成瘾性功能”。但其变化不仅在于将监管对象扩展至自动播放等“成瘾性功能”，还在于对 16 岁以下用

² <https://www.theguardian.com/technology/2026/aug/26/meta-social-media-addiction-trial-settlement>; 另见 <https://abcnews.com/Business/meta-settles-states-landmark-social-media-addiction-trial/story?id=135967095>。

³ <https://www.bbc.co.uk/news/articles/cgerwp7rdlvo>。

⁴ <https://www.kqed.org/news/12099017/gov-newsom-signs-sweeping-new-rules-for-online-child-safety>。

⁵ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1709。

⁶ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1043。

⁷ https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB976。

户实施更严格的限制：与 SB 976 的相关安排不同，AB 1709 未设置凭家长同意即可向该年龄段用户提供“成瘾性功能”的一般例外。因此，AB 1709 法案的通过意味着未来的合规审查需要进一步覆盖信息流、推荐机制、自动播放等具有较强用户留存导向的产品设计与功能，并结合年龄识别机制，评估针对未成年用户的功能限制是否得到有效落实。

（三）SB 1119：对陪伴型 AI 聊天机器人进行未成年人安全风险评估⁸

SB 1119 法案针对在加州向用户提供服务的陪伴型 AI 聊天机器人（Companion chatbots），该法案并不当然适用于所有生成式 AI 企业或聊天机器人。其风险评估等多数义务主要适用于依法确定用户年龄后仍允许未成年人使用的运营者；仅用于客服、特定业务运营、内部研究或技术支持等用途的机器人，可能不属于法定陪伴型聊天机器人的范围。法案规定，自 2027 年 7 月 1 日起，运营者在向加州用户提供新的陪伴型 AI 聊天机器人，或对现有产品进行重大修改前，应开展并记录全面的未成年人安全风险评估。评估应覆盖产品的设计、配置和运营，识别法定损害风险，包括可合理预见的身体或财产损害、严重且可合理预见的心理或情感损害、对受保护隐私权的高度冒犯性骚扰及违法歧视，并采取和记录合理的风险缓解措施。评估须载明对各项法定损害的评估结果摘要及评估方法（包括所参考的公开基准和研究，以及未依赖公开基准的评估）。一组可比较的聊天机器人可合并进行一次评估；为遵守其他法律所做的、范围与效果实质相似的评估亦可用于满足本法要求。此外，高等教育机构仅在教育场景提供、以及仅向员工或承包商在工作场景提供的聊天机器人，不属于法案所称“运营者”。

对于允许未成年人使用的陪伴型 AI 聊天机器人，法案要求建立心理危机响应，针对未成年人用户的适龄披露，特定功能的默认设置与家长控制等机制。其中，默认设置仅可由家长更改，包括关闭持久性对话记忆、关闭推送通知、单次连续使用不超过 1 小时、每日累计不超过 2 小时；家长控制须包含禁止 16 岁以下儿童使用的选项；未关联家长账户的，默认设置不得更改。法案还列举了运营者须采取合理措施防止聊天机器人实施的行为，包括鼓励自残或饮食失调、未经授权诊断或治疗未成年人的身心健康问题、声称自身具有意识或情感、向未成年人表达或模拟恋爱兴趣、鼓励对聊天机器人形成情感依赖、使用与语境不成比例的过度赞美、诱导付费以维持“关系”、鼓励规避家长控制等。此外，运营者不得向儿童展示跨场景行为广告或在对话中利用儿童个人信息定向投放广告，不得出售儿童个人信息，并须对涉及严重自残风险的对话记录保存至少三年。SB 1119 法案同时要求运营者基于《数字年龄保障法案》对用户年龄进行验证，并在无法验证时向所有用户统一适用法案指定的儿童保护措施。

SB 1119 法案还确立了独立未成年人安全审计制度。首次审计应于 2029 年 1 月 1 日或运营者首次公开提供陪伴型聊天机器人之前完成，以较晚者为准，此后原则上每两年开展一次；对于风险评估显示将增加儿童安全风险的重大修改，还须在修改后的产品发布前开展审计。2032 年 1 月 1 日前，上一年度总收入不足 5 亿美元的运营者暂不适用该审计条款。适用运营者须在收到审计报告后 30 个工作日内向总检察长提交报告摘要，并在 90 日内公开概要；总检察长亦可在具备正当理由时要求提供完整报告。相关审计服务还需与下文 AB 1405 的审计师注册及行为规范相衔接。在执行层面，法案不仅允许总检察长与其他检察机关进行执法（过失违法每名受影响儿童最高 5,000 美元、故意违法最高 15,000 美元），在未成年人因运营者违反法案所列产品保护义务（危机响应、儿童安全保障、默认设置、AI 身份提示及禁止性行为）而遭受实际损害时，该未成年人及其父母或监护人也可提起民事诉讼；但财产损失须超过 1,000 美元，情感损害须构成严重情感困扰，且不得依据加州《不正当竞争法》另行提起私人诉讼。

SB 1119 法案是本轮立法中与生成式 AI 企业关系最为直接的法案之一，其重要变化在于将安全义

⁸ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB1119。

务前移至产品上线之前。这意味着，相关企业需在产品开发与设计阶段识别和评估未成年人安全风险，采取相应缓解措施，并留存可供核查的记录。未履行相关义务可能引发监管执法，违反法案指定的产品保护义务并造成实际损害的，还可能引发法定私人诉讼及相应民事责任。

（四）SB 1119 与现行 SB 243 的关系

从立法沿革看，SB 1119 并非加州首部陪伴型聊天机器人立法。2025 年通过的 SB 243 已要求运营者在用户可能误以为对方是人类时作出 AI 身份提示、维持防止生成自杀及自残内容的应对协议，并对已知的未成年用户课以额外的提示义务⁹。SB 1119 在保留 SB 243 上述一般性义务的同时，删除了其中针对“已知未成年用户”的条款，并以更为完整的“Adam’s Law”专章取而代之——即将原来以运营者“实际知悉”为前提的被动义务，转变为以年龄识别为前提、覆盖上线前评估、默认设置、家长控制、独立审计与私人诉权的主动治理框架。

（五）两项未成年人保护法案的关联性

AB 1709 和 SB 1119 在监管对象和具体义务上有所不同，但均进一步强化了产品设计阶段的未成年人保护要求。前者重点限制向特定年龄用户提供“成瘾性功能”，后者则将风险评估、默认设置、危机响应及独立审计等要求纳入陪伴型聊天机器人的治理流程。对于相关平台而言，未成年人保护领域的合规审查可能愈发集中于从设计阶段、默认设置和算法机制等层面识别并降低可预见的风险，并确保年龄识别结果能够转化为实际生效的功能限制和保护措施。

二、AI 监管相关立法：建立第三方审计和独立验证制度

（一）AB 1405：建立 AI 审计师注册制度¹⁰

AB 1405 法案将在加州建立 AI 审计师注册制度。根据该法案，加州政府运营局（Government Operations Agency）需要最迟于 2029 年 1 月 1 日建立公开的 AI 审计师注册名录。该注册制度并不覆盖所有针对 AI 模型或企业的内外部审计，而仅针对“对 AI 系统或模型的内控、流程及相关系统进行的合规审计”。审计师需要在注册时，向政府披露包括适用的法律法规依据，以及审计标准运作流程（SOP）等信息。对于具备公共会计师执业资格的机构或个人，法案则设置了相应的衔接安排。未经注册的机构或个人则不得提供相应的合规审计业务。

AB 1405 法案规定在向企业提供审计服务时，注册 AI 审计师应当基于行业标准，对接受审计的企业出具相关审计报告，并将审计有关材料保存 10 年。报告内容则应当包括审计发现的各项缺陷及治理措施、被审计企业的内部安全标准和协议、以及审计本身的局限性与缺口等内容。

AB 1405 法案尤其强调审计独立性、客观性和诚信。例如，审计师在参与审计期间不得向被审计对象寻求、协商或接受雇佣关系；审计师不得审计自己此前重大设计、开发、实施或者运营过的系统、流程、控制措施或评估工作；若审计师与被审计对象之间存在可能合理影响独立性或客观性的财务、商业、雇佣或其他关系，也不得开展相关审计。此外，法案同时建立了针对违规行为的举报、调查和移除机制，并对举报人提供一定程度的保护。

⁹ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243。

¹⁰ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1405。

（二）SB 813：建立独立的 AI 风险评估框架¹¹

SB 813 法案则将在 AI 审计师制度的基础上，建立一套独立的 AI 风险评估框架。这一框架的核心，是所谓“独立验证组织”（Independent Verification Organizations, IVOs）。根据 SB 813 法案的定义，IVO 应当是具有评估 AI 系统或模型风险并具备确定相关风险评估所采用指标和方法的专业能力的 AI 审计师。IVO 应保持与被评估方在财务、运营和管理层面的独立，被指定的 IVO 须每年向政府及州议会报告其评估方法、内部治理及资金来源变化等信息。

根据 SB 813 法案，加州政府运营局应于 2028 年 1 月 1 日前采取措施，建立对独立验证组织的选择和监管框架。法案规定，该框架应当包括 IVO 申请、指定、资格审查、利益冲突、暂停及终止等一系列机制。在指定和修订标准时，政府运营局应当组织 IVO、AI 企业工程师、AI 安全专家参与并征求意见。

SB 813 法案在其最终版本中未对 AI 企业做出任何强制性的审计或评估要求，而仅说明如果某个 AI 系统已经接受符合该法要求的独立评估，该评估事实在未来涉及损害的诉讼中可以作为相关证据予以考虑，但不具有决定性效力。

（三）两项 AI 监管法案的关联性

与 AB 1405 相比，SB 813 试图在 AI 审计师制度的基础上，建立一个具备更强专业能力、方法论和独立性的 IVO 体系。值得注意的是，AB 1405 和 SB 813 法案本身并不直接要求企业进行合规审计或评估，但通过进一步规范审计或评估机构的资格、审计独立性以及相关服务的监管机制，加州政府及州议会无疑希望通过两项法案的结合，先行建立一套从“企业自我风险管理+第三方审计/评估+政府监督”的多层 AI 治理框架。

三、科技巨头的监管立场变化及其政策意义

OpenAI、Anthropic 和 Meta 这三家科技巨头对近期加州 AI 立法的态度尤其值得关注。

2026 年 8 月 31 日及 9 月 9 日¹²，OpenAI 分别公开支持包括 SB 1119、SB 813 和 AB 1405 在内的加州相关 AI 监管立法，并同时呼吁美国在联邦层面建立具有约束力的 AI 安全监管规则。据 KQED 报道，Anthropic 已于 2026 年 8 月公开支持 SB 813 和 AB 1405 两项法案，其政府事务发言人特别肯定了两项法案对“审计”（评估合规所需的内控）与“评估”（评估 AI 系统本身的风险）所作的区分¹³。Meta 则为算法个性化推荐在提供适龄内容方面的作用进行辩护；对于 SB 1119，其表示仍在评估该法，同时支持赋予家长决策权并保留青少年教育工具使用机会的聊天机器人监管方向¹⁴。

从政策角度看，上述立场体现出当前头部科技企业监管立场的变化趋势：企业并非一味反对具有实质约束力的安全要求，但更倾向于由联邦政府建立统一、可预期的监管标准。这一方面有助于降低企业应对不同州级监管时的合规成本，另一方面，大型科技企业也能将其积累的“安全合规能力”转化为竞争优势，将监管要求转化为自身的竞争壁垒。

¹¹ https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB813。

¹² <https://openai.com/index/supporting-california-bill-advance-ai-youth-safety/>; <https://openai.com/index/ai-policy-window/>

¹³ <https://www.kqed.org/news/12099017/gov-newsom-signs-sweeping-new-rules-for-online-child-safety>; 另见 <https://iapp.org/news/a/a-view-from-dc-what-will-all-these-ai-auditors-be-auditing>。

¹⁴ <https://www.kqed.org/news/12099017/gov-newsom-signs-sweeping-new-rules-for-online-child-safety>。

OpenAI 在声明中强调，其希望通过“反向联邦主义”（reverse federalism），通过趋同的州级监管规则逐步形成统一的安全标准，最终由国会将这些标准制度化为联邦层面的规则¹⁵。这一立场与加州政府及州议会，在联邦层面尚未形成统一监管框架之前，率先推动科技监管的政策取向具有较高一致性。可以说，加州本轮立法也因此具有明显的“监管试验场”特征。其未来与联邦政府及其他州监管机构之间的制度协调与博弈，也值得企业持续关注。

也应看到，美国国内对本轮立法存在不同评价。有评论指出，这批法案是在主要 AI 企业深度参与下形成的“行业可接受”版本：AB 1405 与 SB 813 本身并不强制任何企业接受审计或评估，审计师注册限制要到 2029 年才实施，在此之前企业仍可继续自我认证；而州长纽森在 2024 年否决了更为严格的 SB 1047 之后，一贯倾向于签署经过行业协商的较温和方案¹⁶。从这一角度看，加州本轮 AI 立法更多是在搭建第三方审计与评估的市场基础设施，其实际约束力将取决于后续实施规则，以及其他法律（如 SB 1119）是否引入强制审计要求。

四、对中国出海企业的启示

加州上述法案均以“向加州用户提供服务”为适用标准，不以运营者所在地为限。面向美国市场的中国企业，无论是带有推荐信息流或自动播放功能的社交、短视频及内容社区平台，还是提供 AI 陪伴、角色扮演类聊天产品的开发者，抑或涉足 AI 玩具、儿童智能硬件的厂商，只要其产品可被加州 16 岁以下（AB 1709）或 18 岁以下（SB 1119）用户使用，即可能落入监管范围。特别是 AB 1709 的“受监管平台”不限于社交媒体，且总检察长有权通过规则调整其范围，企业不宜仅凭产品类别自行排除适用。

就具体准备而言，第一，年龄识别是两部法案共同的前提。《数字年龄保障法案》自 2027 年 1 月 1 日起要求应用在下载及启动时通过操作系统或应用商店的接口获取用户年龄区间信号，企业应尽早评估现有账户体系及产品架构能否接收并响应该信号，并确保识别结果能够转化为功能限制或默认设置。第二，将未成年人安全风险评估嵌入产品发布与重大更新流程，并形成可供审计核查的文档；SB 1119 允许对可比较的一组产品合并评估，并认可为其他法律所做的实质相似评估，企业可在全球未成年人保护合规框架下统筹设计，避免重复工作。第三，关注私人诉权带来的诉讼风险。SB 1119 为受实际损害的未成年人及其家长提供了诉权，结合近年美国针对社交媒体和聊天机器人的系列诉讼及 Meta 和解案，产品设计文档与风险评估记录在未来诉讼中的证据意义不容忽视。

与国内规则对照，中国的《未成年人网络保护条例》、《移动互联网未成年人模式建设指南》以及生成式 AI 服务相关规定，同样强调防沉迷、时长管理、内容过滤及家长监护；但加州新法在“默认关闭对话记忆”“禁止表达恋爱兴趣或鼓励情感依赖”“禁止过度赞美”等针对陪伴型 AI 的具体行为规范，以及事前风险评估与独立审计的制度化方面走得更远。已在国内建立未成年人模式的企业具备一定基础，但仍需针对加州要求逐项映射，而非直接照搬国内方案。

五、结论

加州本轮立法表明，其科技监管模式正在从一般性的内容安全和企业自我合规管理，逐步转向产品设计监管、事前风险评估以及第三方独立验证等“事前监管”模式。从未成年人保护角度看，AB 1709 和 SB 1119 分别将传统社交媒体的“成瘾性功能”问题与 AI 陪伴型聊天机器人的相关风险纳入更加具体的产品监管框

¹⁵ <https://openai.com/index/advancing-ai-safety-through-state-and-federal-action/>。

¹⁶ <https://startupfortune.com/newsom-signs-ab-1405-creating-californias-first-ai-auditor-registry/>。

架；从 AI 监管角度看，AB 1405 和 SB 813 则进一步建立对 AI 审计师和独立验证组织的监管机制。上述法案的相互配合，代表着加州在 AI 监管中形成了更为完整的“企业治理+第三方审计/评估+政府监督”体系。

需要注意的是，上述法律的具体义务分阶段实施。AB 1709 原则上于 2027 年 1 月 1 日生效；SB 1119 的风险评估及主要产品保护义务自 2027 年 7 月 1 日起实施，独立审计另设实施期限及过渡安排；SB 813 要求主管机关于 2028 年 1 月 1 日前建立相关制度，AB 1405 的审计师注册限制则自 2029 年 1 月 1 日起实施。企业应区分法律生效时间与各项义务的具体实施时间。从企业合规角度而言，社交媒体及内容平台应先判断是否属于受监管平台，再审查“成瘾性信息流”、自动播放及相关功能，并将年龄识别与功能限制、账户管理相衔接。允许未成年人使用的陪伴型 AI 聊天机器人运营者，应将安全风险评估纳入产品发布及重大更新流程，落实默认设置、家长控制和危机响应机制，并根据收入规模及法定期限安排独立审计。其他 AI 企业则应结合自身适用的具体法律判断审计或评估需求，并关注 AB 1405 和 SB 813 后续实施规则及相关标准。

特别声明

汉坤律师事务所编写《汉坤法律评述》的目的仅为帮助客户及时了解中国或其他相关司法管辖区法律及实务的最新动态和发展，仅供参考，不应被视为任何意义上的法律意见或法律依据。

如您对本期《汉坤法律评述》内容有任何问题或建议，请与汉坤律师事务所以下人员联系：

段志超

电话： +86 10 8516 4123

Email: kevin.duan@hankunlaw.com