生成式人工智能 APP 个人信息安全保护与规制

——与社会学实证研究相结合

王夏逸轩

（山东科技大学，山东 青岛 266590）

摘 要：生成式人工智能技术的迅猛发展与普及给人们带来极大便利，但同时也存在软件隐私政策文本缺乏可读性与重点提示、软件个人信息处理各环节合规性有所欠缺、数据过度与非法采集风险、个人信息存在被滥用风险、个人信息存储泄露风险、数据删除存在安全风险等多重问题，相关领域的个人信息保护已成为亟待解决的问题。本文通过隐私政策分析、社会学调查等实证研究，深度剖析生成式人工智能个人信息保护相关问题及对策，推动构建适配生成式人工智能发展的个人信息保护体系。

关键词：生成式人工智能 个人信息保护 个人信息处理 全方位监管机制 数据治理

一、引言

随着生成式人工智能软件的快速发展与普及，相关软件在提供智能服务的同时，公民个人信息保护问题也日益突出。此类软件深度渗透于日常生活，涉及大量个人信息的收集与处理，但用户普遍存在对隐私政策“不阅读、难理解”的现象，对信息收集范围、使用规则等认知模糊。而部分软件的隐私政策本身也存在表述晦涩、合规性不足等问题，导致个人信息泄露风险加剧，引发信息泄露、财产受侵犯等一系列问题，相关领域的个人信息保护已成为亟待解决的重要议题。

本文首先依据个人信息保护相关规范，对现有隐私政策进行合规性评估，分析存在的问题及潜在风险；通过问卷调查、访谈、案例分析等实证研究，分析个人信息处理各环节存在的问题；最终总结各种问题，并针对性探索科学有效的解决方案，回应数字时代对个人信息法治保障的迫切需求，促进生成式人工智能技术在法律规范框架内的可持续发展。

二、APP 隐私政策多维分析

本文选取六款市面上较主流的生成式人工智能 APP，并从隐私政策条文、个人信息收集、个人信息存储、个人信息共享转让与披露、信息主体权利[①]等多维度进行合规性分析。具体结果如下：

作者简介：王夏逸轩（2006—），男，本科生，研究方向为法学（刑法、数字法治）。

① 全国信息安全标准化技术委员会：GB/T 35273—2020《信息安全技术个人信息安全规范》,2020 年 3 月.

（一）隐私政策文本分析

应用名称	文本字数	阅读时长	是否存在格式条款	重大问题是否提示	重大问题是否明显提示
Deepseek	9345	31—37min	√	√	√
豆包	13840	46—55min	√	√	√
文心一言	11458	38—45min	√	√	×
Kimi	8603	28—34min	√	√	√
海螺 AI	8993	29—35min	√	√	×
腾讯元宝	9135	30—36min	√	√	×

（二） “个人信息收集”合规性分析

应用名称	目的	方式	范围	敏感信息告知同意	争取授权同意例外
Deepseek	√	√	√	√	√
豆包	√	√	√	√	√
文心一言	√	√	√	√	√
KiMi	√	√	√	√	×
海螺ＡＩ	√	√	√	√	√
腾讯元宝	√	√	√	√	√

（三） “个人信息存储”合规性分析

应用名称	最短时限	超时限处理	去标识化处理	数据加密	生物识别分离存储
Deepseek	×	√	√	√	×
豆包	√	√	√	√	×
文心一言	√	√	√	√	×
Kimi	×	×	√	√	×
海螺 AI	√	√	√	√	×
腾讯元宝	√	×	√	√	×

（四）个人信息共享、转让与披露合规性分析：

应用名称	共享、转让与披露目的	数据接收方信息	接收方行为监管	跨境运输与使用	再次确认流程
Deepseek	√	√	√	√	√
豆包	√	√	√	√	√
文心一言	√	√	×	√	×
Kimi	√	√	×	×	×
海螺 AI	√	√	√	√	√
腾讯元宝	√	√	√	√	×

（五）信息主体权利合规性分析：

应用名称	个人信息查询	个人信息更正	个人信息删除	信息主体权利撤回授权同意	主体注销账户	提前获悉服务停运	获取副本
Deepseek	√	√	√	√	√	√	×
豆包	√	√	√	√	√	√	×
文心一言	√	√	√	√	√	√	×
Kimi	√	√	√	×	√	×	×
海螺 AI	√	√	√	√	√	√	×
腾讯元宝	√	√	√	√	√	√	√

三、问卷调查与深度访谈

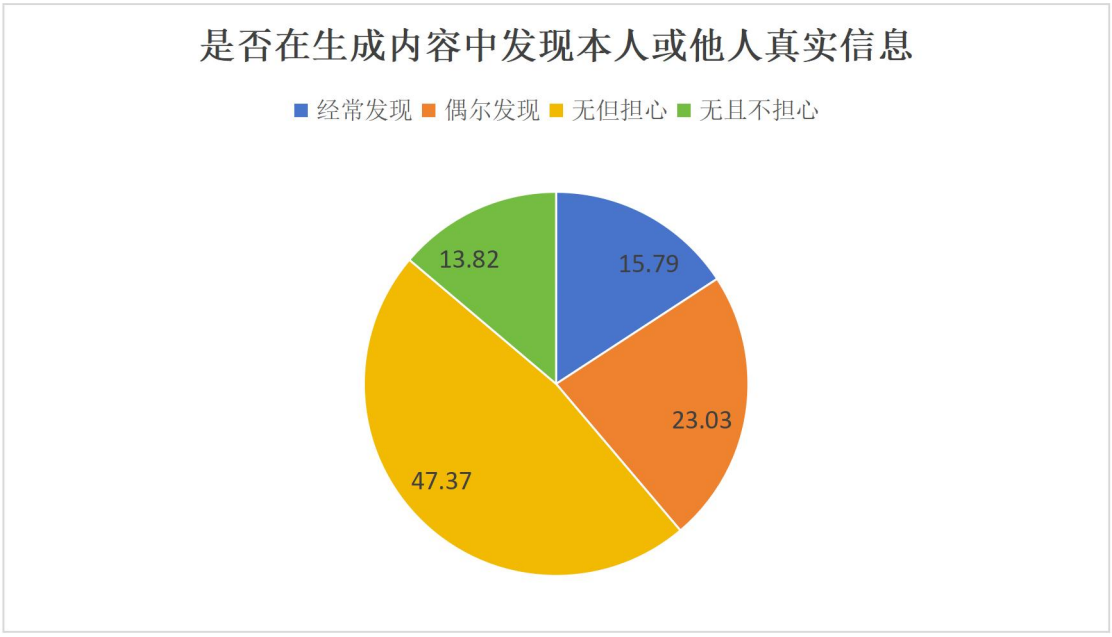
（一）基本概况

从“公众对生成式人工智能软件的认知、需求及面临问题”角度出发，制作调查问卷，同时进行深度访谈。最终收集问卷 152 份并经过信度检验，对三十余人次进行了深度访谈。

（二）数据分析

（1）“主要在哪些场景使用生成式人工智能 APP”问题中，65.79%的受访者主要用于工作，78.29%的受访者主要用于学习，44.74%的受访者主要用于娱乐，50.66%则主要用于生活辅助；在“使用软件时提供过什么信息”问题中，超过 20%的人提供过真实姓名、手机号、地理位置、健康数据等个人敏感信息。可见生成式人工智能已广泛运用于用户生活方方面面，加强用户个人信息安全保护迫在眉睫。

（2）“是否在生成内容中发现本人或他人的真实信息”问题中，发现过相关信息侵权的用户占比高达 38.82%，具体见下图：



且 29.61%的受访者表示收到过与 AI 相关的骚扰信息，37.5%的受访者表示发现 AI 推送过于精准……可见，部分用户个人信息安全遭受侵害，个人信息保护面临多重挑战。

（3）在“现有 AI 软件收集个人信息时是否做到充分告知”问题中，43.42%的受访者认为“有说明但过于复杂难懂，难以真正了解”，26.32%的受访者表示“只是形式展示，难以获取有效信息”；“是否了解软件会将相关个人信息用于模型训练”问题中，不知悉相关问题的用户高达 46.71%。这些数据表明生成式人工智能软件相关隐私政策及“告知同意”规则有待优化。

四、问题总结

（一）软件隐私政策文本缺乏可读性与重点提示

通过对六款主流生成式人工智能软件隐私政策的分析研究，可以发现隐私政策文本平均字数为 10229 字，平均用时 37.5 分钟，且部分软件隐私政策结构分层不合理或不明显，整体缺乏可读性。

占比 50%的软件隐私政策缺乏对格式条款、个人信息处理重大问题的明显说明，导致用户忽略与个人权益息息相关的个人信息授权条款，同时“全有或全无”的隐私政策授权方式也一定程度剥夺了用户的选择权。

（二）软件个人信息处理各环节合规性有所欠缺

通过对六款主流生成式人工智能软件隐私政策的分析研究发现：个人信息存储层面，33.3%的软件在“最短时限”“超时限处理”维度缺乏合规性，所有软件在“生物识别分离存储”维度严重缺乏合规性；个人信息共享、转让与披露层面，33.3%的软件在“接收方行为监管”维度缺乏合规性，16.7%的软件在“跨境运输与使用”维度缺乏合规性，而在“再次确认流程”方面，合规性有所欠缺的软件占比高达 50%；

信息主体权利层面，16.7%的软件在“信息主体权利撤回授权同意”“提前获悉服务停运”维度缺乏合规性，而在“获取副本”维度缺乏合规性的软件占比高达 83.3%。

（三）数据过度与非法采集风险

生成式 AI 软件的运行需要以海量数据为依托，因此需要收集大量用户信息数据，而因规则模糊或未得到践行，在数据收集过程中，存在忽视甚至突破收集规则收集用户的敏感个人信息的情况，同时软件通过技术手段公开获取数据时，公民个人隐私数据容易受到侵犯，甚至会被用于犯罪，例如采集的人脸信息用于诈骗等非法活动。

（四）个人信息存在被滥用风险

生成式人工智能模型底层原理是通过对现有数据语料库的深入学习、分析以回答与解决问题，其通过对海量信息

数据的处理分析，以应对用户问题并输出答案，但其本身并不具备对答案的准确性判断能力，在回答问题、提供数据时，经常会包含其他用户的隐私数据。另外，AI 大模型的深度学习架构、算法不可解释性以及再输出高可识别度的“个人信息画像”等技术特性加剧了个人信息滥用风险[②]。

（五）个人信息存储泄露风险

生成式人工智能 APP 在提供对话、绘图、音视频合成等服务时，会采集并长期存储海量用户信息，如输入的姓名、位置轨迹、声纹人脸、聊天记录、创作草稿等，同时将这些海量信息数据投入到模型训练数据库中，进行模型优化、定向广告推送及衍生内容生成等训练。

② 范懿华.AI 大模型数据利用与个人信息保护的冲突检视及分层治理路径.[J/OL].四川行政学院学报. https://link.cnki.net/urlid/51.1537.D.20250728.1446.006.第 3 页.

但是行业普遍采用的“云端集中+明文存储”架构，缺乏分级加密、令牌化、差分隐私等防护措施，一旦服务器配置错误、API 鉴权缺陷、第三方插件供应链攻击或内部运维人员越权访问，就会导致存储的大量个人信息泄露。同时个人信息可能会被非法转让、分享或被境外相关软件劫取，造成个人信息外部泄露。许多软件的数据处理缺乏透明度或相关通知缺乏时效性，用户对个人信息处理的知情权未得到充分保障，这也加剧了软件存储的个人信息泄露风险。

当前相关部门对“训练数据二次利用”“跨境混合云备份”“边缘节点日志留存期限”等关键环节缺乏细粒度的指引与监管。部分企业以“模型可解释性”为由拒绝披露数据流向，或以“去标识化”之名行“再识别”之实，进一步放大了数据泄露面。

（六）数据删除存在安全风险

我国《个人信息保护法》第 47 条明确了个人信息的主动删除和被动删除义务，对相关个人信息应该及时删除，如果技术上难以删除的，应进行最小化储存且禁止用于其他数据处理活动。

但在具体实践中，生成式人工智能 APP 数据删除环节存在结构性缺陷，用户“撤回”或“注销”等相关操作往往只存在形式上的效果。一方面，预训练阶段写入模型权重的个人信息已与数十亿参数耦合，现有差分隐私、机器遗忘技术仅能降低重现概率，无法定向精准擦除。另一方面，用户只能删除前端聊天记录，而不能删除深层次个人信息数据，且频繁迭代的模型版本，即使删除相关旧数据，仍可通过回滚模型或权重攻击被还原。

通过模型解释性工具也可以发现，软件在微调过程中残留的隐私数据总会形成一定的幽灵特征，即便删除原始数据，这些特征仍会影响后续输出[③]。这些都将导致用户面临“数据一经输入，永久失控”的不可逆风险。

五、对策

（一）优化隐私政策可读性与重大问题提示

要简化语言表达，改善文本结构，建立更为科学的目录导航机制。隐私政策条款措辞应尽量通俗易读，避免使用过多复杂的专业术语及长句，对必要的少量法律、科技术语，应配备相应注释以帮助用户理解其内涵。

要设置清晰的目录结构和跳转链接，采用模块化布局，同时可通过各种图表进行引导与分层，将信息收集、信息存储、信息共享、用户权利等内容分层次呈现，提升用户阅读与理解效率。

对于相关格式条款以及与用户权利相关的重大问题，应通过标注波浪线与下划线、字体加粗、多颜色分层、字体加大或斜体等多种方式组合提示，让用户明晰相关授权条款，同时应优化知情同意规则，取消隐私授权“全有或全无”结构，建立分层、弹性同意授权机制。

要优化分层的差异化同意机制，且敏感信息需加密存储并设置独立访问权限，共享敏感信息需重新获取单独同意；另一方面，要优化动态授权设计，允许用户随时撤回同意，在系统中设置“一键撤回”入口，若信息使用目的变更则需重新获取单独同意，且提供个人信息保障的人工复核渠道[④]，以充分保障用户的知情权与选择权。

（二）完善隐私政策合规性监管机制

企业应首先提升自我监督能力，建立健全规范、有效的合规管理体系。针对高风险的数

③ 宗绍昊，罗世龙．DeepSeek 类生成式人工智能的新型数据安全风险治理[J/OL]．科学学研究．https://doi.org/10.16192/j.cnki.1003-2053.20250729.001.第 14 页.
④ 李金玉：《我国个人信息利用中利益平衡保护的优化路径》，载《陕西行政学院学报》,2025 年第 3 期.

据处理场景，如敏感信息搜集、跨境传输、信息第三方共享等，应设置专门的合规审查流程与机制。同时应完善对数据接收方和外包方的合规管理要求，必要时可引入有资质的第三方评估机构进行安全测评。

政府相关部门应落实对软件开发与运行企业、数据接收外包第三方等监管责任，将企业作为责任承担主体，倒逼企业进行技术革新与自我监督。

（三）建立健全全方位全过程个人信息保护监管机制

要加快出台统一有效的相关法律法规，以填补相关法律空白。政府相关部门要落实监管主体责任，建立“定期评估+不定期抽测”的动态审查与监管机制。分别在生成式人工智能软件的用户信息收集、个人信息存储、个人信息共享、个人数据删除等个人信息处理各环节进行定期评估与不定期抽测，必要时可委托有资质的第三方科技机构进行辅助测评，对辅助测评结果要随机复核，若第三方测评机构与软件开发运行企业恶意串通或评估结果严重偏离客观情况，情节严重的，可以考虑以提供虚假证明文件罪、出具证明文件重大失实罪等追究测评机构的刑事责任。

要提升个人信息共享、转让与披露的处理透明度，企业应细化关键信息的披露范围与表达标准，明确个人信息的具体处理目的、使用场景、数据类型、保存期限等内容，确保用户全面、真实、明确地掌握其个人信息被如何收集、存储与共享。另外，要加强数据共享第三方的信息处理披露义务，第三方应公开数据处理流程与安全保障措施，并及时向软件研发企业与用户进行用途通告。政府监管部门对欠缺透明度或未履行相关义务的企业、第三方应及时责令整改并进行行政处罚。同时政府部门将查获的相关问题应及时向社会、用户通报，以便于用户追究相关企业的民事侵权责任。

（四）加强个人信息保护的刑法规制

对于生成式人工智能软件技术人员未经授权或突破授权恶意收集、泄露用户个人敏感信息，情节严重的，应以侵犯公民个人信息罪追究其刑事责任。

同时生成式人工智能软件技术提供者对下游集体法益犯罪持故意态度时，应视具体情况以帮助信息网络犯罪活动罪或下游犯罪的共犯论处；当技术提供者对下游集体法益犯罪持过失态度时，且未尽相关自我监督管理义务，经政府监管部门责令整改后仍不整改，则应以拒不履行信息网络安全管理义务罪追究刑事责任。

六、结语

目前我国生成式人工智能 APP 相关领域存在软件隐私政策文本缺乏可读性与重点提示、软件个人信息处理各环节合规性有所欠缺、数据过度与非法采集风险、个人信息存在被滥用风险、个人信息存储泄露风险、数据删除存在安全风险等多重问题。

要通过优化隐私政策可读性与重大问题提示、完善隐私政策合规性监管机制、建立健全全方位全过程个人信息保护监管机制、加强个人信息保护的刑法规制等多举措进行个人信息安全保护问题规制。

要将软件开发运行企业作为责任承担主体，倒逼企业进行技术革新、自我监督并履行相关法律义务，通过健全全方位全过程监管机制，以行政责任为主线，鼓励支持用户追究侵权者民事责任，同时以刑事责任为补充与后盾，多措并举以最大限度保护生成式人工智能 APP 中个人信息安全。

参考文献：
[1] 全国信息安全标准化技术委员会：GB/T 35273—2020《信息安全技术个人信息安全规范》,2020 年 3 月.
[2] 范懿华.AI 大模型数据利用与个人信息保护的冲突检视及分层治理路径.[J/OL].四川行政学院学报. https://link.cnki.net/urlid/51.1537.D.20250728.1446.006.第 3 页.
[3] 宗绍昊，罗世龙. DeepSeek 类生成式人工智能的新型数据安全风险治理[J/OL]. 科学学研究. https://doi.org/10.16192/j.cnki.1003-2053.20250729.001.第 14 页.
[4] 李金玉：《我国个人信息利用中利益平衡保护的优化路径》,载《陕西行政学院学报》，2025 年第 3 期.

Personal Information Security Protection and Regulation for Generative AI Applications——Integrating Sociological Empirical Research

WANG Xiayixuan

(Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: The rapid development and widespread adoption of generative artificial intelligence (AI) technologies have delivered substantial convenience while simultaneously revealing multiple risks: privacy policies often lack readability and salient disclosures; compliance gaps persist across the entire life cycle of personal information processing; there are risks of excessive and unlawful data collection; personal information is vulnerable to misuse; storage practices may result in data leakage; and data deletion may entail security risks. Personal information protection in this domain has therefore become an urgent challenge. Using privacy-policy analysis and sociological surveys as empirical methods, this article conducts an in-depth examination of these issues and proposes corresponding countermeasures, with a view to building a personal information protection system calibrated to the developmental characteristics of generative AI.

Keywords: generative AI; personal information protection; personal information processing; comprehensive regulatory mechanisms; data governance