

私密通讯记录人格蒸馏的法律规制 ——隐私权与个人信息保护的协同适用

李云飞

北京外国语大学法学院, 北京 100080

DOI:10.61369/SE.2026060035

摘 要 : 以通讯记录训练人格代理的“人格蒸馏”实践,使封闭于双边关系场域的私密通讯记录进入生成式人工智能的处理链条。本文提出,此类记录属于关系性私密信息,一方参与通讯的同意仅及于原通讯目的,单方上传用于人格建模构成《民法典》第一千零三十三条第五项所称“处理他人的私密信息”;平台在构成个人信息处理者的前提下,应接受《个人信息保护法》第十三条合法基础检验。二元路径应协同适用而非择一:对上传者,为请求权基础并存与救济聚合,对平台则属于多数人责任的分别判断;救济命令应按数据对象分层配置,输出端拦截不得替代可删除对象的删除义务。本文据此建构场域判断、场景分层与救济分层的裁判规则。

关 键 词 : 人格蒸馏; 私密通讯记录; 关系性私密信息; 隐私权; 个人信息保护; 协同适用

Legal Regulation of Personality Distillation in Private Communication Records — Coordinated Application of Privacy Rights and Personal Information Protection

Li Yunfei

Law School, Beijing Foreign Studies University, Beijing 100080

Abstract : The practice of "personality distillation" that uses communication records to train personality agents enables private communication records that are closed in the bilateral relationship field to enter the processing chain of generative artificial intelligence. This article proposes that such records are relational private information, and the consent of one party to participate in the communication is only for the original purpose of communication. Unilateral uploading for personality modeling constitutes "processing of other people's private information" as referred to in Item 5 of Article 1033 of the Civil Code; on the premise that the platform constitutes a personal information processor, it should accept the legal basis test of Article 13 of the Personal Information Protection Law. Binary paths should be applied collaboratively rather than selectively: for the uploader, it is the coexistence and relief aggregation based on the claim rights, while for the platform, it is the separate judgment of the majority responsibility; the relief order should be configured hierarchically according to the data object, and output-side interception should not replace the deletion obligation of deletable objects. Based on this, this article constructs referee rules for field judgment, scene stratification and relief stratification.

Keywords : personality distillation; private communication records; relational private information; privacy rights; personal information protection; collaborative application

引言

开源社区和商业应用中出现的“dot-skill”等项目,已将聊天记录与邮件等材料转化为可持续调用的人格代理,呈现“以材料蒸馏一个人”的技术想象。^[1]当上传者将其与前任、亲友的大量私密通讯记录交给生成式人工智能,提取对方的表达习惯与决策偏好并封装为可对话的人格代理时,问题已超出一概“聊天记录泄露”:其对象兼具私密信息与个人信息的双重属性,其用途横跨本地自用、平台训练与商业化经营。既有研究已从生成式人工智能侵权责任、处理规则、隐私与个人信息界分等角度提供基础分析,^{[2][3][4]}但对这一“共同生成、单方上传、算法再造”的特殊对象,仍缺少可直接用于裁判的协同适用规则。

本文的核心问题是:当一方未经另一方同意,将双方私密通讯记录用于人工智能人格蒸馏时,隐私权路径与个人信息保护路径如何分别成立,又如何在不同场景中协同适用。本文聚焦民事请求权与平台合规义务,不展开刑事与行政责任。本文的贡献不在于创设新概念,而在于提出关系性私密信息的场域判断规则,并将救济命令按数据对象的技术层级分层,使裁判结论可涵摄、可执行。

一、人格蒸馏的行为结构与问题边界

人格蒸馏包括原料取得、算法处理与封装利用三个环节：上传者导出双边通讯记录，人工智能服务对其提取与清洗并向量化或者微调，生成可被召唤的“人格代理”，其利用横跨本地自用、平台调用与收费经营。主体上应区分上传者、平台服务提供者、基础模型开发者与第三方调用者：上传者直接打破通讯场域边界；平台是否构成处理者，取决于其是否实际决定处理目的与方式；基础模型开发者仅提供通用模型且不接触具体记录的，不当然承担同等责任；第三方调用者明知来源未经授权仍传播与售卖的，进入共同侵权评价。

私密通讯记录由此呈现双重属性。其一，它是私密信息。《民法典》第一千零三十二条以“私人生活安宁”和“不愿为他人知晓”界定隐私，第一千零三十三条第五项禁止处理他人的私密信息，第一千零三十四条第三款规定个人信息中的私密信息优先适用隐私权规则。^[9] 双方选择私信而非公开发布，本身即体现对流通范围的限定。其二，它也是个人信息。记录中的称谓、语气与关系状态，足以单独或结合其他信息识别特定自然人；涉及行踪、健康或者亲密关系的，还可能构成敏感个人信息。还应注意，第一千零三十三条的“处理”服务于人格权侵害判断，与个人信息保护法第四条并非同一语境，涵摄应分层对应：导出上传对应提供与加工，平台留存对应收集与存储，向量化微调对应加工与使用；代理输出相似表达是否构成再次处理，须结合可识别性另判。

人格蒸馏亦不同于以公开人格标识塑造虚拟形象的既有案型。“AI 陪伴”案确认，未经同意利用他人姓名与肖像并通过算法开放给用户，可能侵害人格权；AI 声音案则确认，可识别的合成声音仍受声音权益保护。^{[10][7]} 但人格蒸馏所指向的核心客体，并非姓名、肖像与声音这类单一外显的人格标识，而是从关系性文本中经由算法抽取的整体人格轮廓，其识别依赖于表达风格与关系脉络的综合呈现。故既有肖像与声音规则难以径直套用，评价必须回到二元路径。

二、隐私权路径：关系性私密信息的场域判断

隐私权路径第一步是确认记录的私密性。文义上，封闭通讯中的亲密表达、冲突与关系判断，均属通常人不愿向第三方开放的信息。体系上，第一千零三十三条第五项的“处理”可参照第一千零三十五条的行为列举理解，上传至平台至少构成“提供”与“加工”。目的上，隐私权保护的不仅是秘密不被公开，更是个人对私人关系场域边界的合理期待；将私密通讯用于生成可替代对方发言的人格代理，破坏的正是这种期待。

第二步是说明为何上传者作为对话参与者仍不得单方处分全部记录。传统理论易把信息主体设想为单一个体，误以为“我参与的聊天记录，我就可以自由使用”。本文以“关系性私密信息”回应：双边记录由双方共同生成，涉及各自人格与私密表达的部分，双方各享平行的隐私权益。尼森鲍姆的“情境完整性”理论提示，隐私侵害常发生在信息脱离原有社会情境而被不当流转之时。^[8] 转化为裁判规则即：一方参与通讯所作的同意，作为违法阻却事由仅及于原通讯目的内的阅读、保存与回应；用于人格建模已超出同意范围，阻却效果不复存在。所谓“信息共创者”仅是这一场域判断

的描述性标签：它不创设权源，亦非共有或共同处理，对抗单方处分的效力始终来自对方依第一千零三十二条享有的隐私权。

该规则可用三个边界案型校验。其一，上传者仅以自己一方的消息训练模型，但内容足以折射对方人格与关系细节的，仍落入场域保护，因为场域识别不以逐条消息的形式归属为准。其二，上传者将记录用于诉讼举证或者依法维权的，属原通讯目的的合理延伸或法定事由，不在禁止之列。其三，双方消息高度交织而难以逐条切分归属时，无须逐条认定权属，径以记录整体为判断对象即可确定是否越界。三个案型共同表明，规则所禁止者乃人格建模这一目的变更，而非记录的一切后续使用。

隐私权路径可以直接回应人格尊严与私人生活安宁受损的问题：被蒸馏者纵无财产损失，也可请求停止侵害、删除原始记录、赔礼道歉与精神损害赔偿。然而该路径存在结构性局限。其一，隐私权以人格尊严为内核，长于防御私密性本身，拙于处理平台批量训练与算法留存等技术环节。其二，对已进入模型参数的信息，单纯的删除请求难以转化为可执行的裁判命令。其三，一旦行为扩展为面向公众的商业服务，平台的告知、同意、最小必要与安全保护等义务，已非隐私权规则所能周延涵盖，此即引入个人信息保护路径的规范动因。

三、个人信息保护路径：七项合法基础与第72条边界

个人信息保护路径的前提是确认处理行为与主体。依个人信息保护法第四条与第七十三条，^[9] 人工智能服务提供者只有实际决定处理目的、方式、留存或训练用途时方构成处理者；仅提供底层算力或通用接口的，不当然成立。个人信息权益是工具性权利与目的性法益的集合，其保护重心在于规制处理风险。^[10] 仅在本地为个人情感整理而处理的，可能落入第七十二条个人或者家庭事务的排除；一经上传平台、开放第三人或以此牟利，即超出该边界。

第十三条规定的七项合法处理基础应当逐项检验。其一，同意不成立：第十四条要求充分知情下自愿而明确的同意，第二十九条对敏感个人信息还要求单独同意，被蒸馏者不知晓上传与训练目的，发送私信不能推定为同意蒸馏。其二，合同必要不成立：通讯关系或平台合同的履行，均不需要把对方人格特征交给模型学习。其三，“合理处理已公开个人信息”不能援用：私信发给特定接收人并非向公众公开；即便片段曾被发布，第二十七条仍要求合理范围内处理，重大影响的应取得同意。其余法定职责、紧急保护、新闻报道等四项，与私人蒸馏场景无涉，不赘。

还需澄清：第七十二条排除的仅是本法在私人事务场景的适用，既不排除隐私权规则，也不当然排除平台义务。平台介入后，依《生成式人工智能服务管理暂行办法》第七条与第九条，应确保训练数据来源合法、取得同意或符合法定情形，并承担处理者责任。^[11] 即将施行的《人工智能拟人化互动服务管理暂行办法》第十六条保障用户对交互数据的复制与删除，并要求敏感交互数据用于训练须单独同意。^[12] 但应注意两点边界：其一，两办法均以“向境内公众提供服务”为门槛，私有化部署与本地自用未必落入；其二，拟人化办法以“用户”为中心，第十六条保护的是用户交互数据，被蒸馏者恰非用户。这一“用户中心、第三人缺位”的落差，恰恰说明被蒸馏者的救济仍须回到二元路径。

四、二元路径的协同适用与场景化救济

人格蒸馏同时该当隐私权侵害与个人信息违法处理，二元路径应分三层把握，学理上“平行还是交叉”之争亦应在此安放。^[13]其一，对同一上传者，两项请求权基础并存，权利人可择一或一并主张；第一千零三十四条第三款虽确立私密信息优先适用隐私规则，但“优先”并非排斥，而是以人格尊严为基底、允许个人信息保护规则补充，契合“两头强化、三方平衡”的理念。^[14]其二，救济层面，停止侵害、删除、限制处理与损害赔偿可聚合行使，但同一精神损害不得重复填补。其三，上传者与平台之间并非竞合，而是多数人责任的分配：平台构成帮助侵权的，依第一千一百六十九条与上传者承担连带责任；仅违反自身处理者义务的，按原因力承担分别责任。

救济配置应避免把“删除模型数据”表述为单一命令，而须按数据对象的技术层级分别配置。对于原始记录、上传文件与交互日志，可以径行判令删除；对于向量库与检索索引，可以判令删除并重建；对于角色卡与人格配置文件，可以判令下架禁用；对于外挂的微调组件，可以判令卸载并删除对应训练集。至于已融入基础模型参数的部分，因剔除成本畸高且遗忘效果难以验证，原则上不宜迳行判令删除，而应依第四十七条但书转化为停止继续训练与停止调用相关版本并提交遗忘效果评估。输出端拦截须防范语义绕过与提示词注入，不能止步于关键词屏蔽，而应辅以相似度检测与对抗性测试；且其仅是参数层删除不可行时的替代，不得用以豁免前述可删对象的删除义务，技术不可行亦应由平台负担举证责任。

第一，本地私人使用场景，以隐私权路径为主。记录既未上传至平台，也未向第三人展示的，个人信息保护法可能因第七十二条被排除，裁判重点转向第一千零三十二条与第一千零三十三条，即审查上传者是否超出原通讯目的处理私密信息，以及是否损害私人生活安宁或者人格尊严。即便没有公开传播，也不宜当然免责，因为人格代理能够持续替代对方表达，本身即具持续侵扰风险。

第二，平台接入但不公开共享场景，采隐私权与个人信息保护并行。上传者虽主张个人目的，但平台并非家庭事务主体，其收集、保存与分析输入记录或者默认将输入用于模型优化的，应当履行告知同意、最小必要、安全保护与删除响应义务。被蒸馏者可以同时请求上传者停止侵害，并请求平台删除原始记录与可

定位样本，或者采取前述替代措施。

第三，开放共享或者第三方调用场景，以个人信息保护为主，隐私权补强精神损害救济。此时信息流通从封闭关系转为平台化传播，事前影响评估、权限控制、投诉举报、删除与停止处理更具可执行性；若拟人化代理足以使他人误认正在与自然人互动，还应结合《互联网信息服务深度合成管理规定》与拟人化办法第十八条的标识义务，防止公众混淆。^{[12][15]}

第四，商业化利用场景，应当叠加个人信息保护、隐私权与一般人格权三重路径。经营者将人格代理包装成收费产品，实质上把他人的私密表达与人格特征转化为经营资源，既放大处理风险，也加重人格尊严损害。此时可以参照“AI 陪伴”案关于平台算法推荐与“调教”机制构成实质参与的思路，否定单纯技术中立的抗辩；^[6]平台知道或者应当知道用户上传的记录涉及他人私密通讯，仍然提供训练、共享与售卖功能的，可以结合《民法典》第一千一百九十七条认定其承担相应责任。在个人信息保护维度上，第六十九条对损害赔偿确立过错推定，但以构成处理者为前提：处理者不能自证无过错即须担责；^[9]仅提供接口或者算力的，回归第一千一百九十七条的过错审查。

还应说明不同平台形态的责任差异，以免场景类型化因技术假定单一而出现盲区。开源模型本地部署的情形，因无中心化处理者介入，责任应当收缩至上传者本人，回归第一场景的隐私权路径；应用程序接口套壳与转售的情形，应以何者实际决定处理目的与方式识别处理者，直接面向用户的应用方通常首当其冲；境外平台难以直接执行删除命令的，义务重心应落在境内应用提供者与分发渠道，以免救济因执行障碍落空。

五、结论

私密通讯记录人格蒸馏的法律评价，不能在隐私权与个人信息保护之间简单二选一。隐私权路径负责确认关系性私密信息的场域边界，个人信息保护路径负责审查平台处理的合法基础与数据义务；就上传者而言，二者表现为请求权基础并存与救济聚合，就平台而言，则转入多数人责任的分别判断。全部规则的重心不在概念创设，而在可涵摄与可执行：参与通讯不等于授权建模，删除命令应随数据对象分层，输出拦截不得替代可删对象的删除。唯其如此，既能避免对私人生活的过度规制，也能防止平台以“用户上传”为由转移人格蒸馏的系统性风险。

参考文献

- [1] TITANWINGS.colleague-skill/dot-skill[EB/OL].<https://github.com/titanwings/colleague-skill>.
- [2] 王利明.生成式人工智能侵权的法律应对[J].中国应用法学,2023(5):27-38.
- [3] 程啸.论我国个人信息保护法中的个人信息处理规则[J].清华法学,2021(3):55-73.
- [4] 王利明.和而不同：隐私权与个人信息的规则界分和适用[J].法学评论,2021(2):15-24.
- [5] 中华人民共和国民法典[Z].(2021-01-01施行).
- [6] 最高人民法院.民法典颁布后人格权司法保护典型案例：“AI 陪伴”软件侵害人格权案[EB/OL].(2022-04-11).
- [7] 最高人民法院.殷某某声音权益保护案[EB/OL].(2025-06-11).
- [8] NISSENBAUM H F. Privacy in Context: Technology, Policy, and the Integrity of Social Life[M].Stanford:Stanford University Press,2010.
- [9] 中华人民共和国个人信息保护法[Z].(2021-11-01施行).
- [10] 王锡锌.个人信息权益的三层构造及保护机制[J].现代法学,2021(5):105-123.
- [11] 国家互联网信息办公室等.生成式人工智能服务管理暂行办法[Z].(2023-08-15施行).
- [12] 国家互联网信息办公室等.人工智能拟人化互动服务管理暂行办法[Z].(2026-07-15施行).
- [13] 周汉华.平行还是交叉——个人信息保护与隐私权的关系[J].中外法学,2021(5):1167-1187.
- [14] 张新宝.从隐私到个人信息：利益再衡量的理论与制度安排[J].中国法学,2015(3):38-59.
- [15] 国家互联网信息办公室等.互联网信息服务深度合成管理规定[Z].(2023-01-10施行).