NOVARA
CONSULTING GROUP

本文件由机器从英文翻译而来,可能包含错误。英文原文见 https://www.novaracg.com/2026/07/27/the-consent-gap-undisclosed-data-collection-in-sign-language-ai-interpreting-systems/。

同意缺口:手语人工智能翻译系统中未披露的数据收集

Heather Grizzle · 2026年7月27日

Sign language AI depends on video of Deaf users. What vendors disclose about collection, retention, training use and sharing is inconsistent or absent.

目录

1. 执行摘要
2. 政策问题
3. 机构分析
4. 治理影响
5. NCG政策立场
6. 实施考量
7. 参考文献

执行摘要

无论是用于识别、翻译还是虚拟化身生成,手语人工智能在结构上都依赖于对Deaf及听障用户的视频或生物特征采集。这种依赖性造成了一个数据治理问题,而该问题所受到的制度审查,远少于本系列先前几期所记录的语言学与理解层面的失败。在NCG依据SLAT(Sign Language Access Trust)框架审查过的供应商范围内,关于收集了什么数据、保留多久、是否用于训练未来的模型、以及是否与第三方共享等信息的披露,做得好的情况参差不齐,做得差的情况则完全缺失。这并不是在声称任何特定供应商存在恶意行为。目前公开证据尚不足以将意图归咎于任何具名公司,本期也不做此类归咎。这里所主张的是:市场尚未形成与所涉数据敏感性相称的披露标准,而这一缺口本身就是一种治

理失败,与任何单个供应商的行为无关。下文的政策立场主张,同意与披露的充分性应被视为采购中的门槛性标准,而非在功能评估之后才考虑的次要事项。

政策问题

手语采集即生物特征采集。对签手手部、面部语法及身体姿态的视频记录,并非可与提交给文本系统的一句打字输入相提并论的附带性输入。它具有唯一识别性,依其定义即编码了残障状态信息,且在许多司法辖区,它符合受到更高保护的生物特征数据或特殊类别数据的法定定义。例如,伊利诺伊州的《生物特征信息隐私法》(BIPA)要求在采集生物特征标识符之前提供书面通知并取得同意,并规定即使无法证明实际损害也可追究责任。欧盟的《通用数据保护条例》(GDPR)将揭示与残障相关信息的数据,以及用于身份识别的某些生物特征数据,归为第9条下的特殊类别数据,要求在普通同意之外具备明确的法律依据。加州的《消费者隐私法》(CCPA)与《隐私权法》(CPRA)也对生物特征信息施加了各自的披露与选择退出义务。

尽管存在这样的监管环境,面向消费者的手语人工智能产品却常常依赖通用服务条款,而这些条款在拟定时并未考虑Deaf用户或生物特征采集的问题。NCG所做的供应商快照调查中,反复出现三类缺陷:第一,没有单独或显著的披露说明视频输入构成受特定法定保护的生物特征数据;第二,没有明确说明所采集的视频或衍生特征是否用于模型训练,若用于训练,保留多久、享有何种删除权利;第三,没有无障碍格式的披露,也就是说,同意条款本身往往未被译成ASL,也未以其数据被收集的群体可使用的格式呈现。最后这一点并非技术性细节。若一套披露制度要求具备第二语言的书面阅读能力才能理解自身生物特征数据将被如何处理,那么对于英语读写能力参差不齐、以ASL为主要理解语言的群体而言,这套制度并不能算作真正发挥作用的知情同意。

机构分析

> 手语采集本质上就是生物特征采集。

造成这一缺口的机构性动因是结构性的,而非阴谋性的。手语人工智能仍是一个年轻的产品类别,大多数供应商是小型公司,使用的是整个软件行业通用的隐私政策模板,即便有所调整也十分有限。与此同时,机构层面的采购流程历来是依据功能性标准来评估这些产品:虚拟人形象呈现是否清晰可辨,识别系统是否达到可接受的准确度,供应商在开发过程中是否有Deaf员工参与。数据治理则被当作法律或IT合规问题,在实质性采购决策之后才加以处理,如果有处理的话。这种先后顺序本身

就是真正的治理失败。等到数据治理方面的缺陷显现出来——无论是通过数据泄露、监管调查,还是来自数据被采集的Deaf社群的公开批评——机构关系与财务承诺早已确立,此时要求补救的筹码相应也就更弱。

此外还存在一个与同意问题相邻、但在分析上有所区别的训练数据来源问题。若干手语人工智能系统部分是基于从公开发布内容中提取的视频语料库构建的,其中包括Deaf ASL使用者及教育工作者在社交平台上创作的内容。这种使用是否属于平台自身服务条款所允许的范围,是否构成版权法意义上的合理使用,以及那些手语出现在这些影像中的个人是否被有意义地告知其内容将被用于训练商业人工智能系统,这些都是悬而未决的问题,NCG尚未就任何特定供应商独立核实,本文在此也不主张其为既定事实。就治理层面而言,关键在于:评估这些系统的采购官员几乎完全无法了解训练数据的来源,而当前行业的披露规范也并不要求供应商提供这类信息。

治理影响

采购手语人工智能的机构,不仅仅是一件翻译工具的购买者。在该系统对学生、患者、员工或公众的实时视频进行处理的情况下,该机构在大多数适用的隐私框架下同时也是数据控制者或处理者,并且无论其在采购时是否被告知,该机构都会承继供应商数据实践所带来的治理风险。若一家机构在部署前未核实供应商的数据收集、保留及训练使用做法,便无法向自身的利益相关方、监管机构,或数据被采集的Deaf个人,声称自己已履行了数据保护义务。无论该供应商的实际行为如何,这一点都成立,因为未经核实的披露缺失本身就是治理缺陷,与任何特定供应商是否最终被证明负责任地处理数据无关。

NCG政策立场

NCG的立场是,数据治理与隐私披露的充分性应作为手语人工智能采购中的门槛性标准,在功能性能评估之前而非之后进行评估。若供应商无法提供清晰、可通过ASL获取、且符合司法辖区要求的披露,说明所收集的生物特征数据是什么、保留多久、是否以及如何用于模型训练、存在何种删除权利,则无论其在评估的其他方面表现出怎样的语言或技术水平,都应被视为不具备机构部署资格。这一立场已在SLAT框架的“数据治理与隐私”及“透明度与披露”领域中得到落实,本期主张,鉴于机构所面临的法定风险,以及不具无障碍性的同意语言给受影响群体带来的理解障碍,这些领域相对于该框架目前的权重设定,理应获得更高的权重。

实施考量

评估手语人工智能的采购官员应要求供应商作为响应条件,提供涵盖上述四项要素的书面数据治理披露文件,并要求该披露文件除书面英文版本外还提供ASL视频格式版本,并将供应商拒绝或无法提供此类披露视为不利认定,而不论该系统在功能层面表现如何。已有部署的机构应主动进行回溯性审查,而非等到合同续签之时,因为治理风险的累积始于部署之日,而非发现之日。若供应商的训练数据来源无法核实,机构应要求就数据来源合法性提供合同保证,而不应将无从回答视为一种中性的未知状况。

> **四项披露要素**
>
> 收集了哪些生物特征数据、保留多久、是否以及如何用于模型训练,以及存在何种删除权利。

参考文献

- 《生物特征信息隐私法》,740 ILCS 14(伊利诺伊州)。
- 经《加州隐私权法》修订的2018年《加州消费者隐私法》,Cal. Civ. Code § 1798.100 et seq.。
- 欧盟第2016/679号条例(《通用数据保护条例》),第9条。
- 《联邦贸易委员会法》,15 U.S.C. § 45(不公平或欺骗性行为或做法)。
- Novara Consulting Group,《手语无障碍信任*(SLAT)*框架》,2026年。

> **TIPS**
>
> Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

引用方式

Grizzle, H. M. (2026, July 27). 同意缺口:手语人工智能翻译系统中未披露的数据收集. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003818