NOVARA
CONSULTING GROUP

本文档由机器从英文翻译而成，可能存在错误。英文原文见 https://www.novaracg.com/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/。

本应存在却仍未出现的证据

Heather M. Grizzle · 2026年8月4日

一项调查，追踪2026年最重大人工智能手语宣称背后所缺失的独立证据，涵盖从*SignGemma*到各类商业翻译工具。

第五参数

Novara Consulting Group · Novara Press · 无障碍人工智能治理与证据期刊

《第五参数》第1卷·第13期

作者 Heather M. Grizzle

ORCID 0009-0005-2605-8256

证据日期 截至2026年8月

关键词 手语人工智能；独立评估；证据分类；SignGemma；透明度；信任；聋人无障碍

关于2026年人工智能与手语最重要的事实，并不在于发生了什么。

而在于什么没有发生。

从1月到8月，各公司发布产品、展示系统、筹集资金，并谈论沟通的未来。新工具被宣传为更快、更强大、更接近日常使用。有些吸引了热烈的头条报道。另一些则伴随着关于训练数据、响应时间或可识别手语数量的惊人宣称登场。

没有出现的，是独立证据来证明这些系统确实如宣传所言那样运作。没有任何第三方公开发布过针对该指数所审视的任何主要系统的准确率、错误率或用户理解度的测量结果。对所需资料来源的检索未能找到Signapse、Sorenson AST、Rylo Sign、Sign-Speak、ChatSign、SignVrse、Talksign或谷歌SignGemma的任何独立基准测试。

该发现的正式措辞十分谨慎：截至2026年8月“未能找到”任何A类证据。这为某些证据可能存在于公开记录之外留下了可能性。它或许存放在某家公司内部，掌握在某位客户手中，或包含在尚未发表的研究中。

但无法被审查的证据无法支撑公开的宣称。这一区别之所以重要，是因为该行业并不沉默。它持续不断地发布公告、宣布合作、举办演示、讲述融资故事。给人的印象是进展迅速。然而，用来判断这种进展的公开证据却未能跟上步伐。

一个领域可以忙碌，却未必变得可问责。它可以制造新闻，却未必产生知识。

三类证据

该指数将证据分为三大类。

- A类是独立证据。它来自第三方，包括可供审查的测量结果：准确率结果、错误率、理解度研究、与其他系统的对比，或在已发表基准测试上的表现。
- B类是由负责该系统的公司或组织自行提供的证据。它可能包含有用的技术细节，但提出宣称的组织同时也是被评估的对象。
- C类由断言构成：即公告、宣传材料、访谈、演示或媒体报道中的陈述，这些陈述无法被独立核实。

这并不意味着每一项B类或C类宣称都是虚假的。一家公司可能诚实地描述其产品。一个工程团队可能报告了真实的结果。一次演示可能展示了真正的技术进展。问题在于，这些都无法回答核心问题：当没有利益相关的人来测量该系统时，它的表现究竟如何？

2026年，几乎所有关于人工智能与手语的公开判断，仍然依赖于B类或C类材料。各公司提供了宣称、语言，往往还提供了用来理解这些宣称的衡量标准。独立评估基本上是缺席的。

SignGemma问题

谷歌的SignGemma提供了最清晰的例证。当它被宣布时，该模型吸引了关注，因为它似乎将一家大型科技公司的资源投入到了手语人工智能之中。与之相关的宣称包括基于10,000小时数据进行训练，以及低于200毫秒的延迟。这些数字之所以广为流传，是因为它们同时暗示了规模和速度。

十五个月后，该模型仍未正式发布。谷歌员工在公司的人工智能开发者论坛上证实了这一点。讨论一直持续到2026年6月。来自巴基斯坦、埃及及其他地区的研究人员曾申请访问权限，却未能获得。

这创造了一个不寻常的公共对象：一个可以被讨论却无法被测试、可以被引用却无法被检视、可以被期待却无法被使用的模型。由于无法访问该模型，外部人员无法复现其所

报告的结果。他们无法审查其训练数据是如何汇集的，无法测试其在不同手语间的表现，也无法发现它在受控示例之外的行为。因此，关于训练规模和延迟的宣称仍然只是C类证据。

谷歌的声誉并不能改变这些宣称的性质。权威可能使一项宣称更具说服力，但这并不能使其变得独立。重复并不能将断言变成测量。关键并不在于SignGemma一定失败了，而在于公众没有获得任何可靠的方式来知晓它是否成功。

准确率并非单一数字

独立测试在手语技术中尤为重要，因为“准确率”这个词所掩盖的可能与它所揭示的一样多。一个系统可能在稳定光线下录制的有限词汇量上表现良好，却在自然对话中步履维艰。它可能能识别外貌与其训练数据相似的人所打的手语，却对其他人表现不佳。它可能正确识别单个手语，却在整句中丢失意义。它可能只适用于某一种手语，却被宣传得仿佛适用于所有手语。

手语并不是口语的编码版本。它们拥有各自的语法、地区差异和文化背景。含义可能取决于动作、位置、时机、面部表情以及手语之间的关系。一个能识别手部动作却忽略面部的系统，可能产出单词却丢失了整句的意义。即便一个引人注目的整体准确率也说明不了什么，除非我们知道测试的对象是什么、参与者是谁，以及什么被视为正确结果。

该系统是在孤立的手语符号上测试的，还是在连续对话中测试的？参与者对该模型是否熟悉？聋人手语使用者是否评判过输出是否合理？错误是仅仅被计数，还是被检视以判断其潜在危害？测试是否涵盖了不同年龄、肤色、打手语风格和地区差异？这些都不是次要细节，它们决定了这个数字的真正含义。

这正是理解度研究与技术基准同样重要的原因所在。一个系统可能获得令人印象深刻的分数，却仍然产出令用户感到困惑、不自然或具有误导性的结果。最终的检验标准不在于模型是否能检测动作，而在于人们是否能够理解并信赖它所传达的内容。

期待的代价

关于一个有效系统即将问世的反复承诺是有后果的。一个尚未发布的模型可能占据本应由现行服务填补的空间。它可能影响资助决策，左右政策讨论，并促使各机构推迟对人力供给的投入。它可能使当前的不足显得只是暂时的，即便没有任何时间表或公开证据支持这种看法。

这就是期待的政治学。帮助总是即将到来，因此当下缺乏帮助被视为不那么紧迫。对聋人用户而言，代价是实际存在的。一家医院、大学、雇主或公共机构可能会指向新兴技术，将其作为无障碍状况正在改善的证据。但一项充满前景的公告无法为一次医疗问诊提供口译。一次演示无法保证进入课堂的机会。一个研究人员都无法获取的模型，同样也无法惠及那些其生活被用来证明其重要性的人们。

还有另一重代价。当公众的注意力聚集在未经独立测试的系统上时，那些提供不那么引人瞩目、却更为可靠的支持形式的小型组织，可能从视野中消失。人工口译员、社区主导的服务以及既有的无障碍实践，很少能获得像新型人工智能模型那样的兴奋关注。它们被当作当下存在的开支来评判，而技术却被当作未来的可能性来估值。这种比较是不对等的。现有的服务必须为其当前的局限性负责，而被承诺的系统却因尚未到来而被允许保持完美。

透明度所需要的

独立证据的缺失，原则上并不难以纠正。开发者可以在做出广泛的性能宣称之前，先向合格的研究人员提供访问权限。评估工作可以与聋人手语使用者共同设计，而不仅仅是以他们为测试对象来进行。测试集、方法和定义都可以公开发布。结果可以按语言、情境和用户群体进行细分，而不是压缩成一个单一的头条数字。

研究人员也应当能够在无需依赖被评估公司许可的情况下报告失败情况。这一切都无法消除不确定性。独立研究可能设计不当。基准测试可能只奖励狭隘形式的表现。已发表的结果可能随着系统的变化而过时——然而，不完美的审查也胜过被管控的可见度。一个公开的基准测试为他人提供了可以质疑、重复或改进的东西。一个未经支持的宣称，则只能提供供人传播的东西。

透明度也意味着要清楚说明一个系统做不到什么。为受限场景设计的产品，不应被描述为一种通用解决方案。仅针对一种手语测试的模型，不应被允许借用“手语”这个词所表面暗示的普适性。原型应当被标明为原型。未发布的模型不应被当作公共基础设施对待。

第五参数

技术系统通常依据熟悉的指标来评判：速度、规模、成本和准确率。缺失的那个指标是信任。信任不是又一项性能宣称，它是使宣称可被检验之后的结果。当外部研究人员能够审查一个系统时，当用户能够描述其失败之处时，当证据在公告周期过去之后依然可供查阅时，信任便随之增长。这就是**第五参数**。

以此衡量，人工智能手语领域在2026年8月步入了一个严重的赤字状态。它拥有产品、承诺和宣传，却唯独缺乏针对那些最受关注系统的公开可得的独立证据。

这种缺失不应被误认为只是文书工作中的暂时空白，它是这项技术当前状况的一部分。在独立评估者能够对这些系统进行测量之前，对许多关于其性能的问题，诚实的答案既不是“它们有效”，也不是“它们失败了”。

而是公众尚未获得足够的证据来知晓答案。

利益声明 作者为Novara Consulting Group LLC创始人，该公司构建并授权本文所引用的Sign Language Access Trust Index，并主编《第五参数》。

资助 无。由Novara Consulting Group内部制作。

数据与材料 本文依据所列系统的公开记录，以及谷歌人工智能开发者论坛中有关SignGemma的讨论帖撰写而成。

人工智能使用披露 本文文字均为作者本人撰写。人工智能辅助仅用于将文本排版至品牌母版；文中论点、图表、引文或引用均非模型生成。

建议引用格式 Grizzle, H. M. (2026). The evidence that wasn't there. The Fifth Parameter. Vol 1 Issue 13. Novara Press.

> **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, August 4). 本应存在却仍未出现的证据. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003816