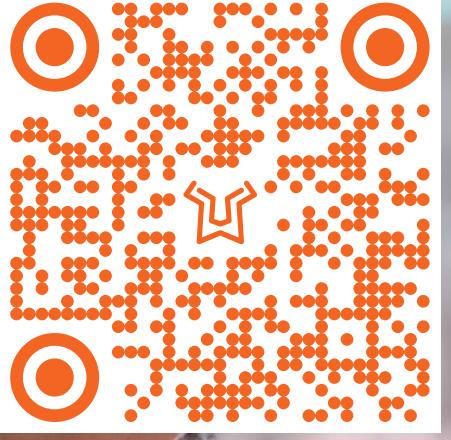


# 创新与行为规范的平衡： 版权所有者对于在知识产权本地化过程中使用 GenAI 的看法



LIONBRIDGE



作者: Melinda Johnson, Megan Duff, Kathryn Nolte, Stephanie Casale

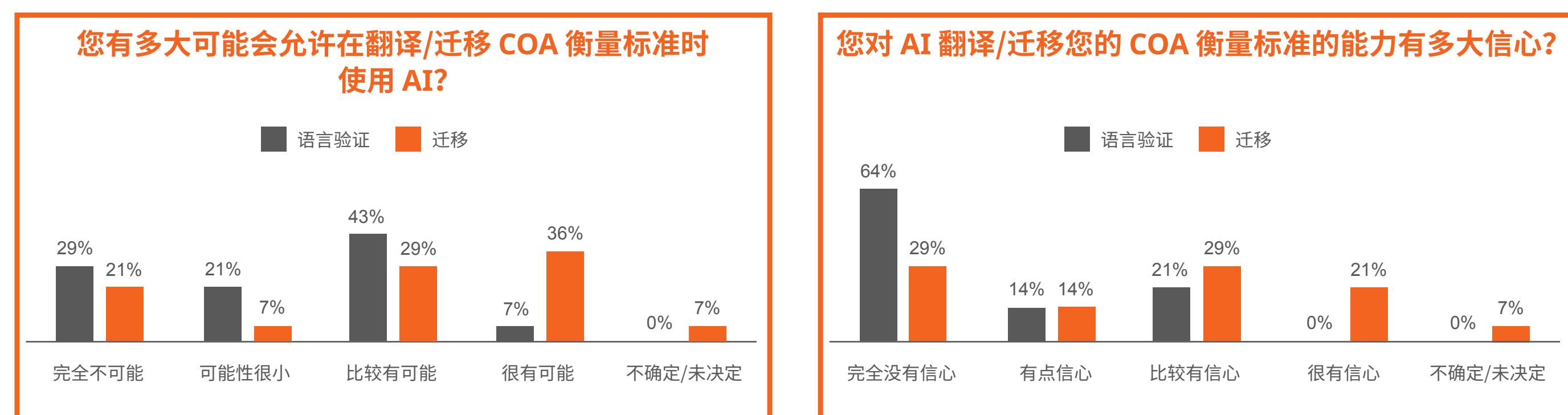
## 目标

由于在将生成式人工智能 (GenAI) 应用于临床结果评估 (COA) 知识产权 (IP) 方面缺乏明确的指导, 将 GenAI 整合到临床试验运营时 — 尤其是在 COA 领域 — 遭遇了阻碍。这种不确定性减缓了 GenAI 在 COA 语言验证和迁移等关键领域的应用, 而这些领域对于倾听患者意见至关重要。本研究旨在收集并传播 COA 版权所有者对于在这些流程中使用 GenAI 的看法, 以指导未来的部署策略。

## 结果与解读

### 1) 对于在语言验证和迁移过程中使用 AI 的允许度及信心

a. 结果: 有 64% (9/14) 的 COA 版权所有者表示对于在语言验证过程中使用 AI 来翻译其 COA 衡量标准的做法“完全没有信心”。然而, 只有 29% (4/14) 的人表示“完全不可能”允许在语言验证过程中使用 AI 来翻译其 COA 衡量标准。关于在 eCOA 迁移过程中使用 AI 的信心与允许度, 反馈则较为一致: 分别有 29% (4/14) 和 21% (3/14) 的受访者表示“完全没有信心”和“完全不可能”。

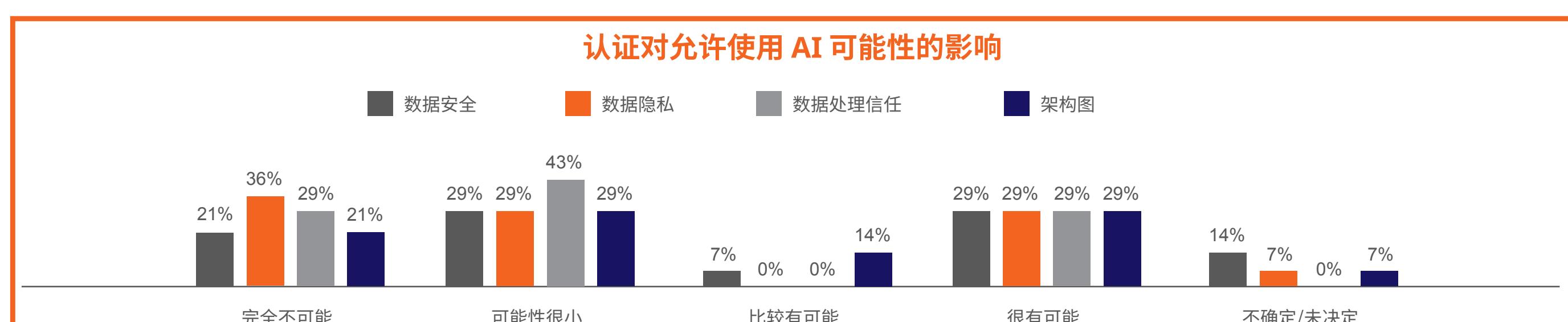


b. 解读: 虽然版权所有者对 AI 能力的信心因使用场景(语言验证和迁移)而有所差异, 但允许在这两种工作流程中使用 AI 的可能性始终呈积极倾向, 只是程度有所不同(“比较有可能”与“很有可能”), 其中允许在语言验证中使用的比例为 50% (7/14), 允许在迁移中使用的比例为 65% (9/14)。根据对于开放式问题的文本回复, 版权所有者对 AI 的整体信心仍然较低。然而, 在通过细致深入的沟通后, 大多数受访者表示愿意允许使用 AI。

### 2) 认证与信任度

a. 结果: 我们询问了版权所有者以下几项认证和/或文件是否会影响其批准使用 AI 的可能性: 1) 数据安全认证(例如 ISO27001:2022), 2) 数据隐私认证(例如 ISO27001:2019), 3) 数据处理信任认证(例如 SOC2 Type 2), 以及 4) AI 架构图。不同认证/文件的影响程度不一, 其中对数据隐私认证和数据处理信任认证持怀疑态度的人最多。但对于所有四类文件, 都有 29% (4/14) 的受访者表示在提供相应文件的情况下, “很有可能”会批准使用 AI。

b. 解读: 尽管有一些受访者对不同类型的文件(认证/架构图)、数据安全认证、数据隐私认证以及 AI 引擎的运行方式持怀疑态度, 但这些文件确实能使至少 29% 的版权所有人更倾向于允许在翻译其 IP 时使用 AI。



## 总结

与其它利益相关方一样, COA 衡量标准的版权所有者目前所处的行业环境仍缺乏关于在本地化其 IP 时使用 AI 的明确指导。然而, 尽管这些利益相关方对在本地化其 IP 时合理使用 AI 的信心有待提升, 但他们似乎愿意就此开展讨论、制定政策、修订许可协议, 以及审查相关文件以验证其 IP 在与 AI 引擎交互时的安全性。最后需要指出, 由于受访者人数有限, 本次研究存在一定的局限性, 需要进一步分析来验证这些结果。

## 方法

为实现此目标, 我们向 100 位 COA 版权所有者发送了一份包含 15 个问题的电子调查问卷, 问卷涵盖以下三个主题。在为期三周的规定时间内, 共有 14 位版权所有者回复了这份匿名调查问卷:

- 1) 对于在 COA 翻译(以下称为“语言验证”)与 eCOA 迁移(以下称为“迁移”)过程中使用人工智能(AI)的允许度及信心
- 2) 认证与信任度
- 3) 当前的 AI 方法、政策和许可

### 3) 当前的 AI 方法

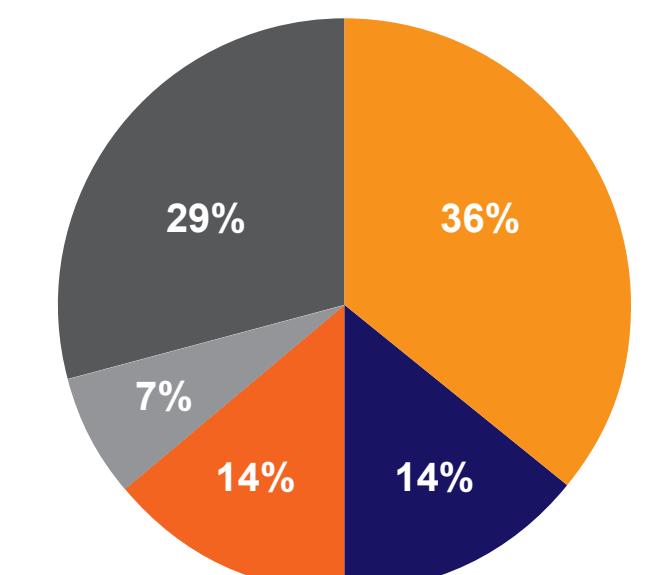
a. 结果: 我们针对此主题提了四个问题。

1) 当被问及在语言验证和迁移中使用 AI 的方法, 与以往使用计算机辅助翻译(CAT)工具解决方案的方法有何不同时, 多数受访者(36%, 5/14)表示“不确定/未决定”。相比之下, 28.5% (4/14) 的受访者表示“有很大不同”。仅有 28.5% (4/14) 的受访者表示与 CAT 技术相比, 这项技术的应用方式“没有不同”或“略有不同”。

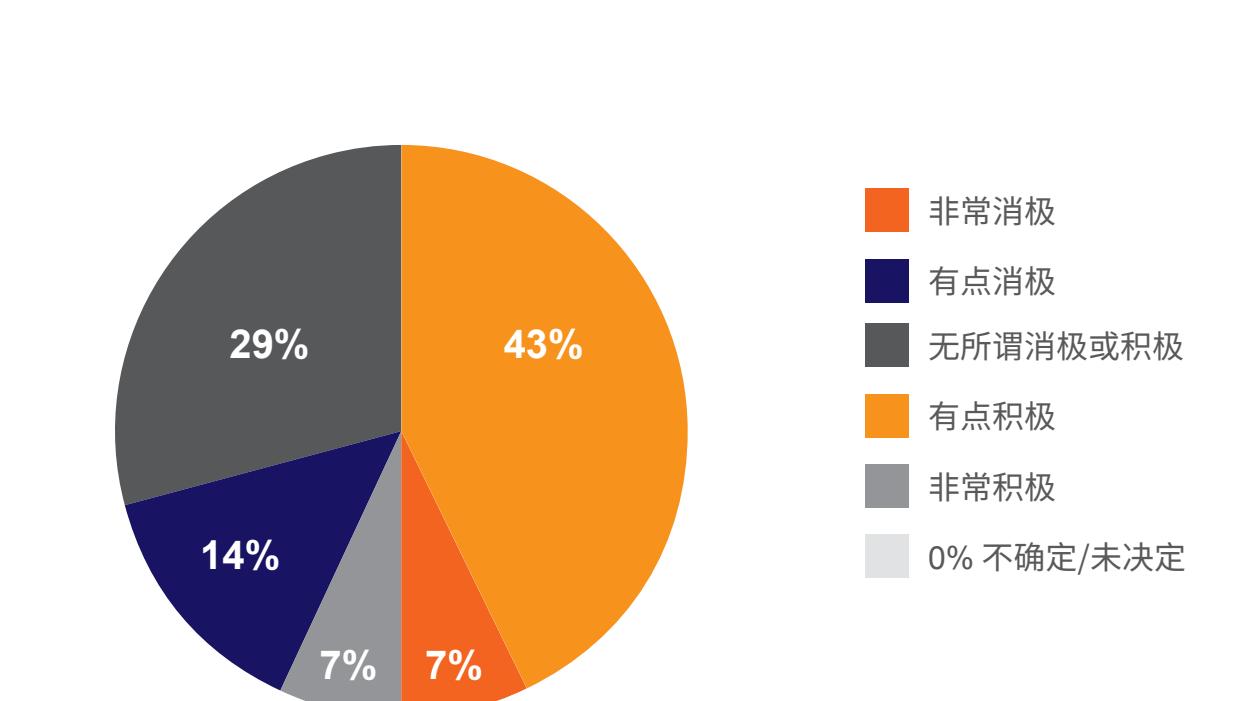
2) 43% (6/14) 的受访者认为在临床试验和翻译行业中使用 AI 是一种“有点积极”的趋势, 还有 29% (4/14) 的受访者认为“无所谓消极或积极”。认为“非常消极”和“非常积极”的比例相同, 均为 7% (1/14), 其余 14% (2) 的受访者认为“有点消极”。

3,4) 当被问及是否在公开论坛中主动提供 AI 使用政策时, 72% (10/14) 的受访者回答“否”。此外, 86% (12/14) 的受访者尚未修改其许可协议以明确其关于 AI 使用的政策。

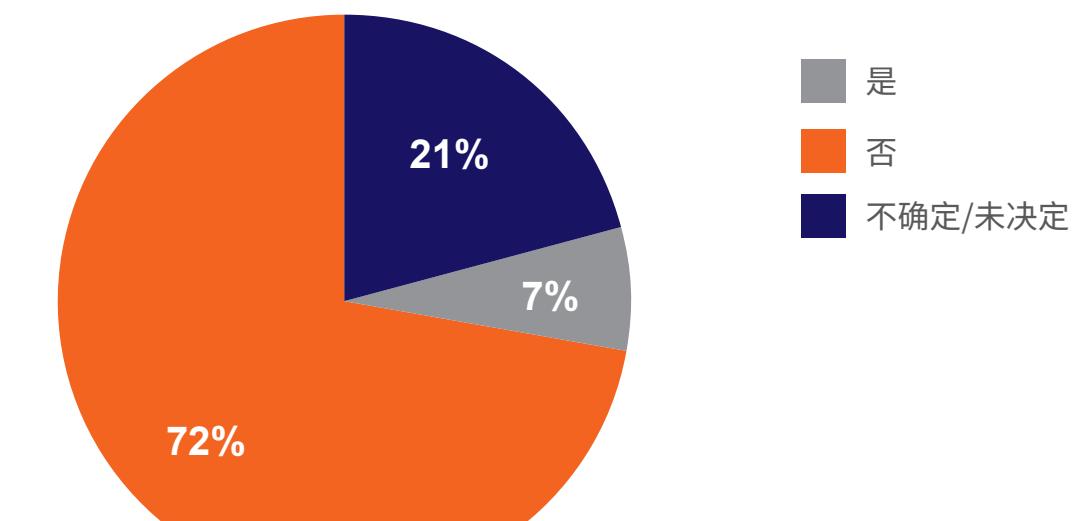
您在对 COA 衡量标准进行语言验证或迁移时使用 AI 的方法与以前使用 CAT 工具解决方案(例如, 翻译记忆库/术语库/机器翻译)的方法是否有所不同?



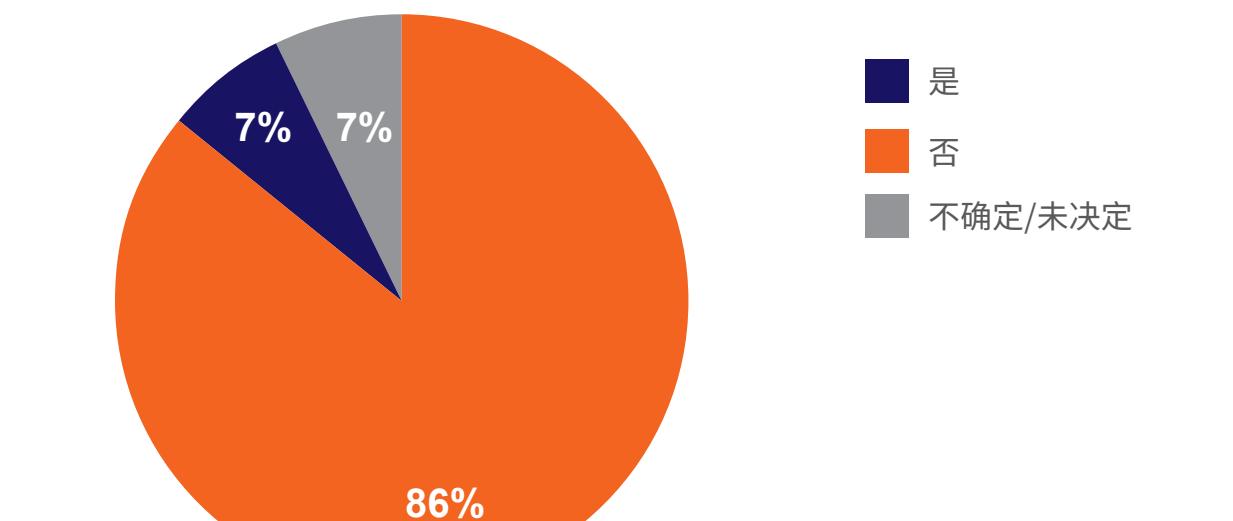
您认为在临床试验和翻译行业中越来越多的使用 AI 是一种积极的趋势还是消极的趋势?



您是否会主动向潜在被许可方提供关于 AI 使用政策的信息(例如, 通过公开政策中的条款)?



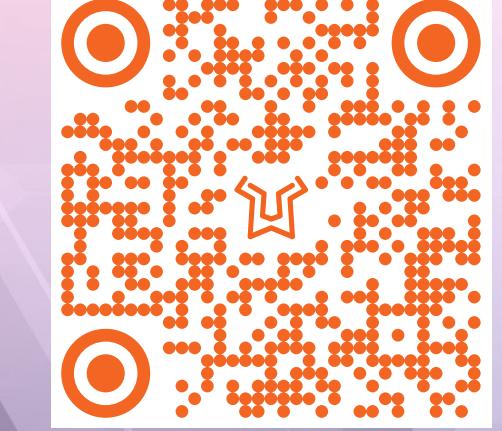
您是否调整了许可协议以明确关于 AI 使用的政策?



b. 解读: 人们对于这项技术相较于过往翻译技术的差异程度, 以及如何基于这种差异调整应用方法, 似乎存在一定的不确定性。业内对于 AI 的使用也存在广泛的意见分歧。然而, 针对本主题(甚至可能整个调查中)最突出的发现是: 版权所有者没有主动向潜在被许可方告知与使用其 IP 相关的 AI 政策, 即使在许可协议本身中也没有相关说明。不过, 根据对于开放式问题的文本回复, 这一状况预计将很快发生变化。因此, 本研究小组建议, 被许可方在与版权所有者接洽时, 若对方未能提供相关信息, 应假定这意味着他们尚未形成明确意见或政策。



# 评估 GENAI 与人工屏幕截图审查在 ECOA 本地化中的输出效果： GenAI 能否成为提升反馈质量的关键？



LIONBRIDGE

作者:Kathryn Nolte, Karolina Elizondo Jimenez, Rupali Kadam, Melinda Johnson

## 背景介绍

一直以来，电子临床结果评估(eCOA)的翻译、迁移和屏幕截图审查严重依赖人工操作，部分原因在于屏幕内容不可编辑。在屏幕截图审查(SSR)流程中，需将目标语言的屏幕报告与源语言主屏幕报告和原始内容(即原始纸质问卷，如果有)进行比对。

Lionbridge 执行的屏幕截图审查分为两种类型：简单审查和复杂审查。复杂 SSR 与简单 SSR 的区别在于，前者需要逐行检查所有目标原始内容。Lionbridge 认识到生成式人工智能(GenAI)和光学字符识别(OCR)技术有潜力补充传统屏幕截图审查方法、降低成本和缩短时间(即减少审查轮次)、提高质量并最终改善患者治疗效果，并对这种潜力进行了测试。本研究利用了 Lionbridge 专有的 Aurora AI Clinical Outcomes™ 工具中的简单与复杂屏幕截图审查功能。

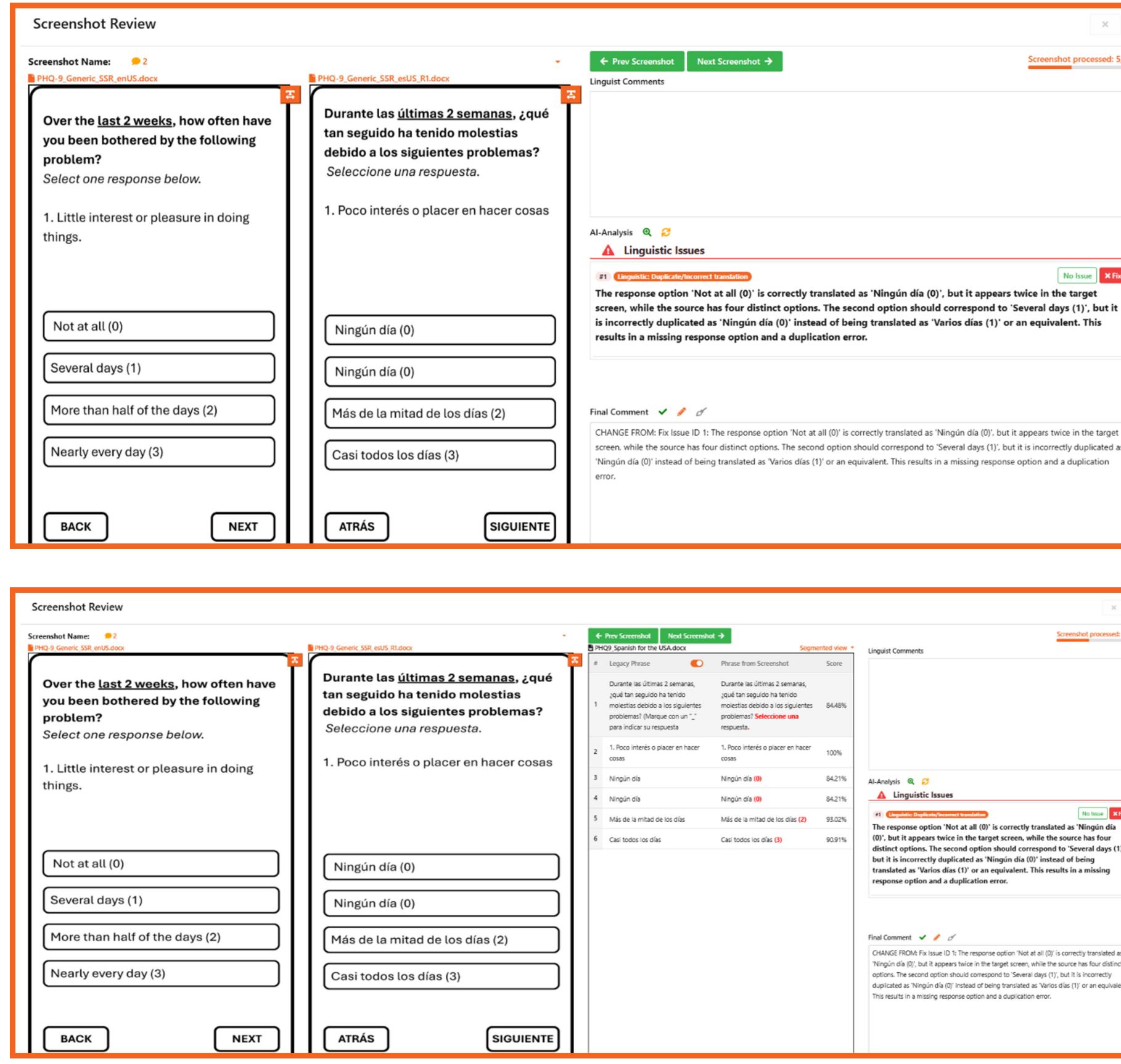
## 方法

Lionbridge 利用安全的 GenAI 引擎，针对 5 份面向患者的电子临床结果评估(eCOA)的目标语言屏幕截图报告生成质量保证反馈。这些 eCOA 在长度和复杂度上各有不同，并且特意加入了 11 种类型的错误以进行测试：内容缺失、内容未翻译、翻译错误、缺少 eCOA 编辑记录、版本号不匹配、格式问题、词形问题、标签问题、换行符问题、大小写问题以及量表锚点问题。测试共涵盖 16 种目标语言，涉及多种字母体系与语系：保加利亚语(保加利亚)、波兰语(波兰)、罗马尼亚语(罗马尼亚)、希腊语(希腊)、西班牙语(阿根廷、墨西哥、美国)、法语(比利时、法国)、葡萄牙语(巴西、葡萄牙)、匈牙利语(匈牙利)、土耳其语(土耳其)、韩语(韩国)、繁体中文(台湾)及泰语(泰国)。我们根据最新的 eCOA 行业标准和既定实践，对提示进行了定制，直至获得合适的输出。同时，我们将相同的 eCOA 屏幕报告发送给 Lionbridge 认可的语言专家进行人工审查。之后，我们安排一位公正的审校人员来验证这两种输出(AI 和人工审查)，并对准确性和完整性进行评分。

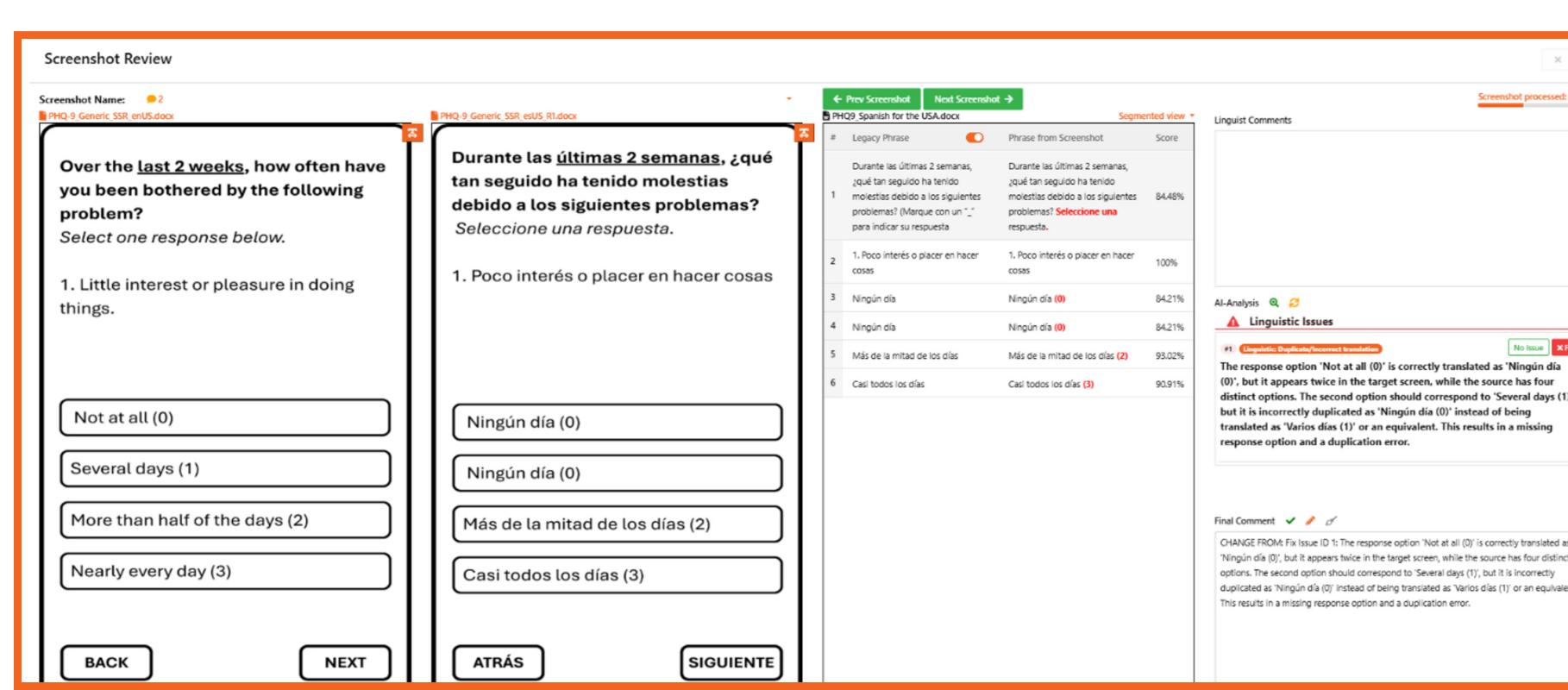
### 提示设计演进过程：

研究的第一阶段采用了 GenAI 的简单屏幕截图审查提示，该提示非常适合在翻译工作流程(翻译/编辑或正向翻译/反向翻译)后比较源语言屏幕报告和目标语言屏幕报告。该工具的这一功能完成后，我们便转向开发和测试更细致的复杂屏幕截图审查 GenAI 功能，该功能更适合已迁移的内容，因为它比简单功能更进一步。除了将目标语言屏幕内容与源语言主屏幕报告进行对比外，复杂功能还会将目标语言屏幕内容与目标原始内容进行对比。

### 简单屏幕截图审查界面：



### 复杂屏幕截图审查界面：



## 总结

我们的研究表明，在翻译和迁移 eCOA 的过程中，将 GenAI 与人工审查相结合是进行屏幕截图审查最有效的策略。提示会不断完善，以便随着时间的推移检测出越来越多的问题。然而，对于此类高度敏感的内容，目前仍建议在屏幕截图审查过程中保留人工参与，以确保最佳质量。GenAI 可以识别初步错误，从而显著提高速度、效率和准确性，而人类的专业知识对于进行细致的决策和识别 AI 可能遗漏的问题仍然至关重要。这种“人机协作”方法能够通过减少所需的审查轮次来简化流程，通过节省语言专家进行质量检查的时间来降低成本，最重要的是，它能够增强我们更精准、更高质量地捕捉患者心声的能力，从而确保提供更可靠、以患者为中心的服务体验。

## 结果与解读

对简单 SSR GenAI 功能的质量保证反馈分析表明，该功能是一项非常实用的质量保证(QA)检查工具，可在翻译和迁移项目中与人工反馈协同使用。简单审查功能特别适合支持语言专家处理无需交叉核对原始内容的翻译/屏幕截图审查项目。利用此功能，Aurora AI Clinical Outcomes 工具成功捕捉到了目标语言屏幕中反映出的全部 11 种错误类型的多个实例。但该功能在某些情况下也存在局限性，那就是无法核对原始内容。

尽管复杂 SSR GenAI 功能目前仍在开发中(预计将于 2026 年第一季度完成部署)，但阶段性结果显示，它不仅能识别出简单 SSR 工具所识别的所有问题(甚至更多)，提供的反馈也更细致。复杂 SSR 功能可以交叉核对原始内容，这将消除当前简单 SSR 工具产生的部分误报，具体如下表所示。

| 错误类型         | 简单 SSR GenAI 功能：<br>备注及局限性                                      | 复杂 SSR GenAI 功能：<br>备注及局限性      | 示例  |
|--------------|---|---------------------------------|---|
| 内容缺失         | 可以标记目标内容是否缺失  | 可以确认目标文本是否故意缺失，因为在原始内容中不存在      | Portuguese (Portugal)<br>#4 <b>Inconsistency: Missing content</b> Target Paragraph: 1<br>The statement '[scale name] is subject to [copyright holder]’s Terms of Use.' is present in the source but missing in the target. All content should be retained and translated.<br><br>No Issue ✘ Fix   |
| 内容未翻译        | 可以标记保留英文的文本，并提示这可能是故意为之(例如，版权声明文本)                              | 可以确认文本是否故意保留英文                  | Bulgarian (Bulgaria)<br>⚠ <b>Linguistic Issues</b><br>The line '[scale name]' items H17, BP1, N6, EG6' remains in English in the Bulgarian screen. This should be translated or localised for Bulgarian users unless these are standardised item codes that should remain in English. If these are not standard codes, provide a Bulgarian translation.<br><br>No Issue ✘ Fix   |
| 翻译错误         | 可以识别与源文的偏差  | 可以识别偏离原始内容的错误译文，即使它们与源文相符       | Portuguese (Portugal)<br>#3 <b>Linguistic: Inconsistent scale anchors</b><br>The response options in the target text do not consistently match the gradation of the source. For example, '1 - Muito pouco' (Very little) is stronger than '2 - Mais ou menos' (More or less). The scale anchors should directly correspond to 'Somewhat', '4 - Muito' (A lot) is not equivalent to 'Very much'. The scale anchors should be consistently translated to preserve the intended gradation.<br><br>No Issue ✘ Fix   |
| 缺少 eCOA 编辑记录 | 简单功能和复杂功能的功能性应相同  |                                 | Portuguese (Portugal)<br>#1 <b>Linguistic: Incorrect translation</b> Source Paragraph: 6<br>Target Paragraph: 6<br>The instruction 'Please select one number per statement to indicate your response as it applies to the past 7 days.' is translated as 'Faça um círculo ou marque um número por afirmação para indicar a sua resposta no que se refere aos últimos 7 dias.' The phrase 'Faça um círculo ou marque um número' introduces the instruction to circle or mark a number, which is not present in the source text. The source only instructs to select a number, not to circle or mark. The translation should not introduce additional instructions not present in the source.<br><br>No Issue ✘ Fix |
| 版本号不匹配       | 可以标记源语言与目标语言版本号不匹配的问题，但需要注意的是，这些差异可能是故意为之                       | 可以通过检查原始内容来确认版本号差异是否是故意为之       | Korean (Republic of Korea)<br>#3 <b>Inconsistency: Version number mismatch</b><br>The version number in the source is 'v1.1', while in the target it is 'v1.0'. The version number should match between source and target unless there is a justified reason for the difference.<br><br>No Issue ✘ Fix  |
| 格式问题         | 可以标记不同于源语言主屏幕报告的格式  | 可以根据原始内容确认与源语言的格式差异是否是故意为之      | Portuguese (Portugal)<br>#1 <b>Formatting/Layout: Inconsistent bolding/underlining</b> Source Paragraph: 5<br>Target Paragraph: 5<br>In the source screen, the phrase 'past 7 days' is bolded, while in the target screen only 'últimos' is bolded. The bolding should match the source, with 'últimos 7 dias' in bold in the target.<br><br>No Issue ✘ Fix   |
| 词形问题         | 可以检测形容词中意外的性别形态，但需要注意的是，这可能是故意为之(例如，由于患者群体全部为女性，所有形容词都可能采用阴性形式) | 可以通过与原始内容进行对比，提供有关预期性别形态的更多背景信息 | Spanish (United States)<br>⚠ <b>Linguistic Issues</b><br>#2 <b>Linguistic: Incorrect gender form</b> Source Paragraph: 7 Target Paragraph: 5<br>The source uses 'I feel fatigued,' which is gender-neutral. The target uses 'Me siento agotada,' which is feminine. For a general patient-facing questionnaire, the translation should be gender-neutral ('Me siento fatigado/a' or 'Me siento fatigada o fatigado') unless the instrument is specifically for female patients.<br><br>No Issue ✘ Fix   |
| 标签问题         | 简单功能和复杂功能的功能性应相同  |                                 | Thai (Thailand)<br>⚠ <b>Formatting/Layout Issues</b><br>#1 <b>Formatting/Layout: Underlining/HTML Tag Display</b> Source Paragraph: 7<br>Target Paragraph: 7<br>The target screen displays HTML tags (<u>) instead of rendering underlined text for 'นักเรียนที่เข้าร่วม', 'เพื่อสุขภาพ', and 'รายงานผล'. The source uses underlining for emphasis, but the target shows raw tags, which is incorrect formatting. The underlining should be properly rendered in the target.<br><br>No Issue ✘ Fix  |
| 换行符问题        | 简单功能和复杂功能的功能性应相同  |                                 | French (France)<br>⚠ <b>Formatting/Layout Issues</b><br>#1 <b>Formatting/Layout: Line break</b> Source Paragraph: 6 Target Paragraph: 4<br>The target screen introduces a line break in 'habituées' that splits the word across two lines with a hyphen, which is not present in the source screen. This disrupts readability and should be corrected so that 'habituées' appears on one line without a hyphen.<br><br>No Issue ✘ Fix   |
| 大小写问题        | 可以标记与源语言不符的意外大小写模式，但需要注意的是，这些差异可能是故意为之                          | 可以确认目标语言屏幕上的意外大小写是否与原始内容一致      | Hungarian (Hungary)<br>#2 <b>Linguistic: Inconsistent capitalization</b> Source Paragraph: 6<br>Target Paragraph: 6<br>The target text uses lowercase for 'magyar verzió', while the source uses title case 'English version'. For consistency and professionalism, the Hungarian should use 'Magyar verzió Magyarország része'.<br><br>No Issue ✘ Fix  |
| 量表锚点问题       | 简单功能和复杂功能的功能性应相同  |                                 | Portuguese (Portugal)<br>#2 <b>Formatting/Layout: Scale anchor placement</b> Source Paragraph: 8<br>Target Paragraph: 7<br>The scale anchor text in the target is longer and may not sit directly under the endpoint of the scale, potentially causing layout misalignment. Ensure the anchor text does not spread into the center and remains directly under the endpoint.<br><br>No Issue ✘ Fix   |

### 人机对比：

- 在评估目标语言屏幕内容时，GenAI 的速度比人类快得多，它可在数秒内针对指定量表的所有目标语言屏幕生成反馈。
- Lionbridge 的质量保证专家指出，GenAI 流程显著缩短了时间、提高了效率，因为 Aurora Clinical Outcomes 工具的 SSR 界面使他们能够轻松地同时访问原始文件、主屏幕报告和目标语言屏幕报告。
- GenAI 发现了人工审校人员遗漏的几个问题，特别是换行符、大小写、翻译错误(包括重复文本)和量表锚点问题。GenAI 更容易发现这些问题，因为该工具使用了光学字符识别(OCR)技术来解析不可编辑图像内容中的文本差异，而这些差异人眼很容易忽略。
- GenAI SSR 在不同语言间的错误检测有时会存在不一致。例如，虽然 GenAI 在大多数语言中都发现了未翻译“items”的问题，但在西班牙语中却未能发现。这可能是因为两者之间的唯一区别在于一个重音符号(“items”)。
- 人工审校人员不太可能错误地标记与上下文相关的问题(例如，女性患者群体、故意的版本号差异、故意保留英文的版权声明文本等)。