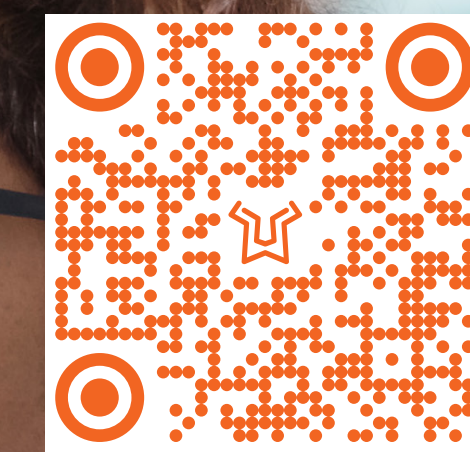


혁신과 적정선의 균형: 지적재산권 로컬라이즈 과정의 생성형 AI 사용에 대한 저작권자의 의견 관련 조사



LIONBRIDGE

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

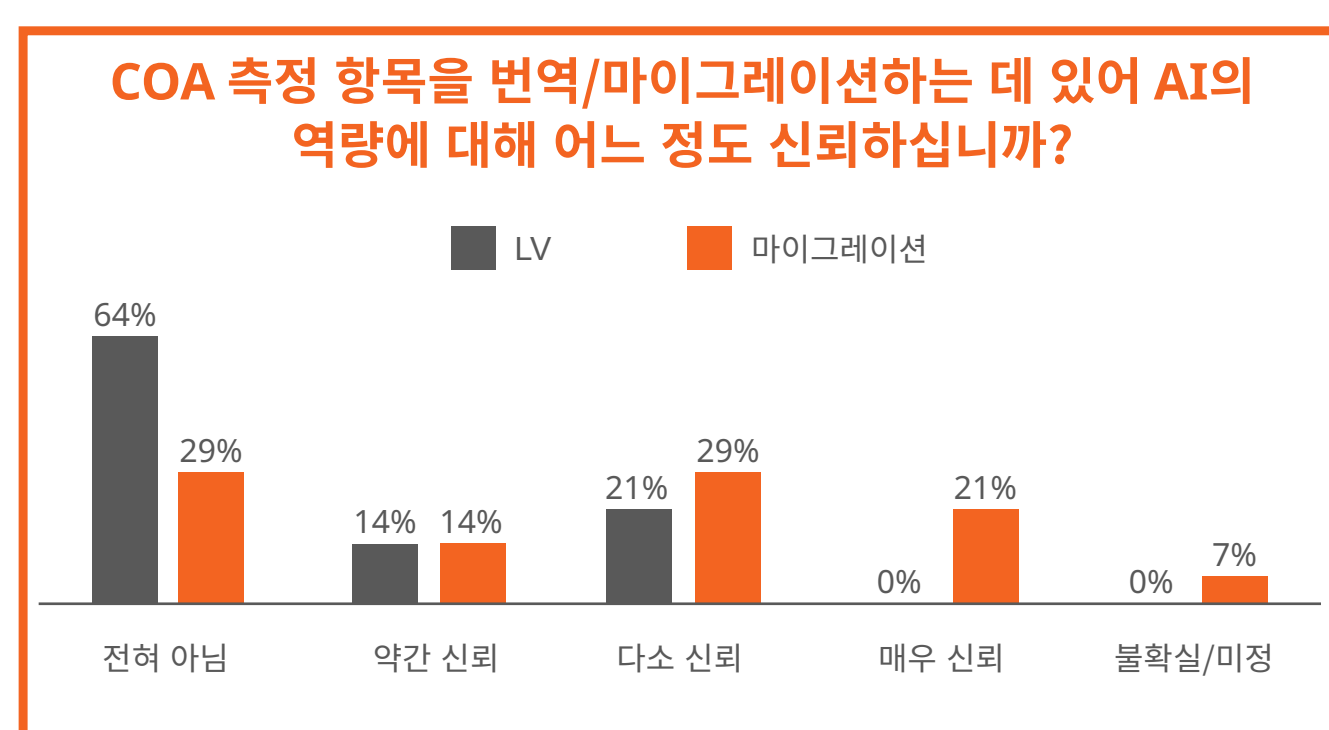
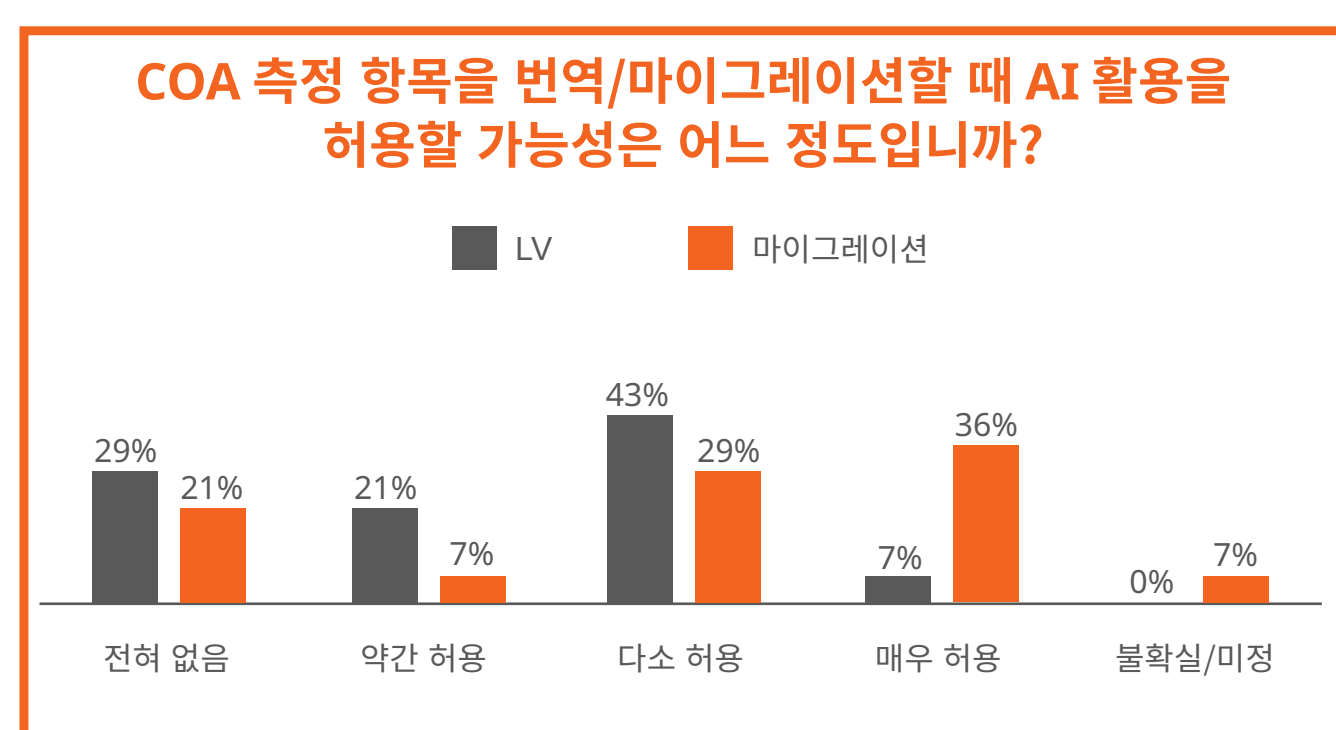
목적

생성형 AI(GenAI)를 임상시험 운영, 특히 임상결과평가(COA)에 통합할 수 있는지에 관한 의문은 GenAI의 COA 지적재산권(IP) 적용에 대한 명확한 지침이 없어 난관에 봉착해 있다. 이러한 불확실성은 환자 의견을 반영하는 데 핵심적인 역할을 하는 COA 언어검증 및 마이그레이션과 같은 중요 영역에서 GenAI의 도입을 지연시키고 있다. **본 연구는 향후 배포 전략에 참고할 수 있도록 이러한 과정에서 GenAI 사용에 대한 COA 저작권자들의 관점을 수집하고 공유하는 것을 목표로 한다.**

결과 및 해석

1) LV 및 마이그레이션에 AI 사용 시 허용 수준 및 신뢰도

a. 결과: 응답한 COA 저작권자 중 64%(9/14)는 언어검증 과정에서 COA 측정 항목을 번역하는 데 AI를 사용하는 것을 '전혀 신뢰하지 않는다'라고 답했다. 그러나 'AI 사용을 전혀 허용하지 않는다'라고 응답한 비율은 29%(4/14)에 불과하여 큰 차이를 보였다. 반면, eCOA 마이그레이션 과정에서의 AI 사용에 대한 신뢰도와 허용도 응답은 '전혀 아님'이 각각 29%(4/14)와 21%(3/14)로 큰 편차를 보이지 않았다.

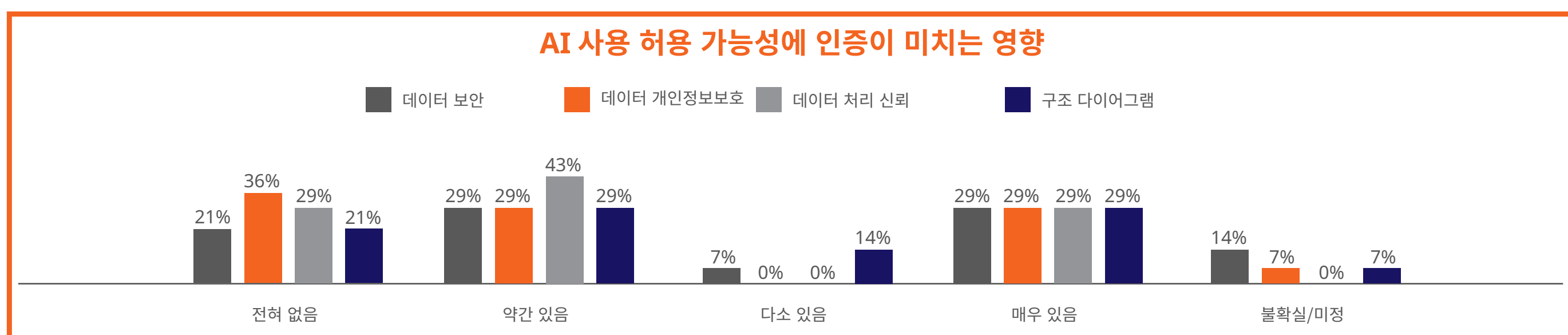


b. 해석: AI의 역량에 대한 저작권자의 신뢰도는 LV 및 마이그레이션에서의 사용 간에 차이가 있었으나, 이 두 워크플로 전반에서 AI 사용을 허용할 가능성은 LV의 경우 50%(7/14), 마이그레이션의 경우 65%(9/14)로 일관되게 긍정적이었다(중간 내지 매우 큼). 서술식 답변 분석에 따르면 전반적인 신뢰도는 낮은 편이지만 대부분의 응답자는 상세한 논의 후 AI 사용을 허용할 의사가 있었다.

2) 인증 및 신뢰

a. 결과: 저작권자에게 1) 데이터 보안 인증(예: ISO27001:2022), 2) 데이터 개인정보보호 인증(예: ISO27001:2019), 3) 데이터 처리 신뢰성 인증(예: SOC2 Type 2), 4) AI 아키텍처 다이어그램 등 여러 인증 및/또는 문서가 미치는 영향(즉, AI 사용 승인 가능성)을 확인했다. 다양한 인증/문서는 혼재된 결과를 보였으며, 특히 데이터 개인정보보호 및 데이터 처리 신뢰성 인증의 영향에 대해 회의적이라는 답변이 가장 높았다. 그러나 4가지 유형 모두에서 응답자의 29%(4/14)가 이러한 문서가 마련된 경우 AI 사용을 승인할 가능성이 매우 높다고 답했다.

b. 해석: 다양한 문서 유형(인증서/다이어그램), 데이터 보안, 개인정보 보호 및 AI 엔진 작동 방식에 대한 의심에도 불구하고, 저작권자의 최소 29%는 이러한 유형의 문서가 있을 때 IP 번역 시 AI 사용에 더 호의적인 태도를 보일 가능성이 더 높다.



결론

COA 측정 항목의 저작권자들은 다른 이해관계자들과 마찬가지로, 자사 IP를 로컬라이즈할 때 AI 사용에 대한 명확한 지침이 여전히 부족한 상황에서 업계를 탐색하고 있다. IP 로컬라이즈 과정에서 적절한 AI 사용에 대한 이해관계자들의 신뢰도를 높일 여지는 있지만, 이들은 **AI 엔진과 상호작용 시 IP 보안을 검증하기 위한 논의, 정책 수립, 라이선스 계약 수정, 문서 검토 등에 개방적인 태도를 보이고 있다.** 마지막으로, 본 연구는 응답자 수가 제한적이었으므로 이러한 결과를 검증하기 위해서는 추가 분석이 필요하다.

방법

이를 위해 아래 세 가지 주제에 대한 15개 문항으로 이루어진 전자 설문조사를 COA 저작권자 100명에게 발송했다. 14명의 저작권자가 주어진 기한인 3주 내에 이 익명 설문조사에 응답했다.

- 1) COA 번역(LV) 및 eCOA 마이그레이션(마이그레이션)에 AI 사용에 대한 허용 수준 및 신뢰도
- 2) 인증 및 신뢰
- 3) 현재 AI의 접근방식, 정책 및 라이선싱

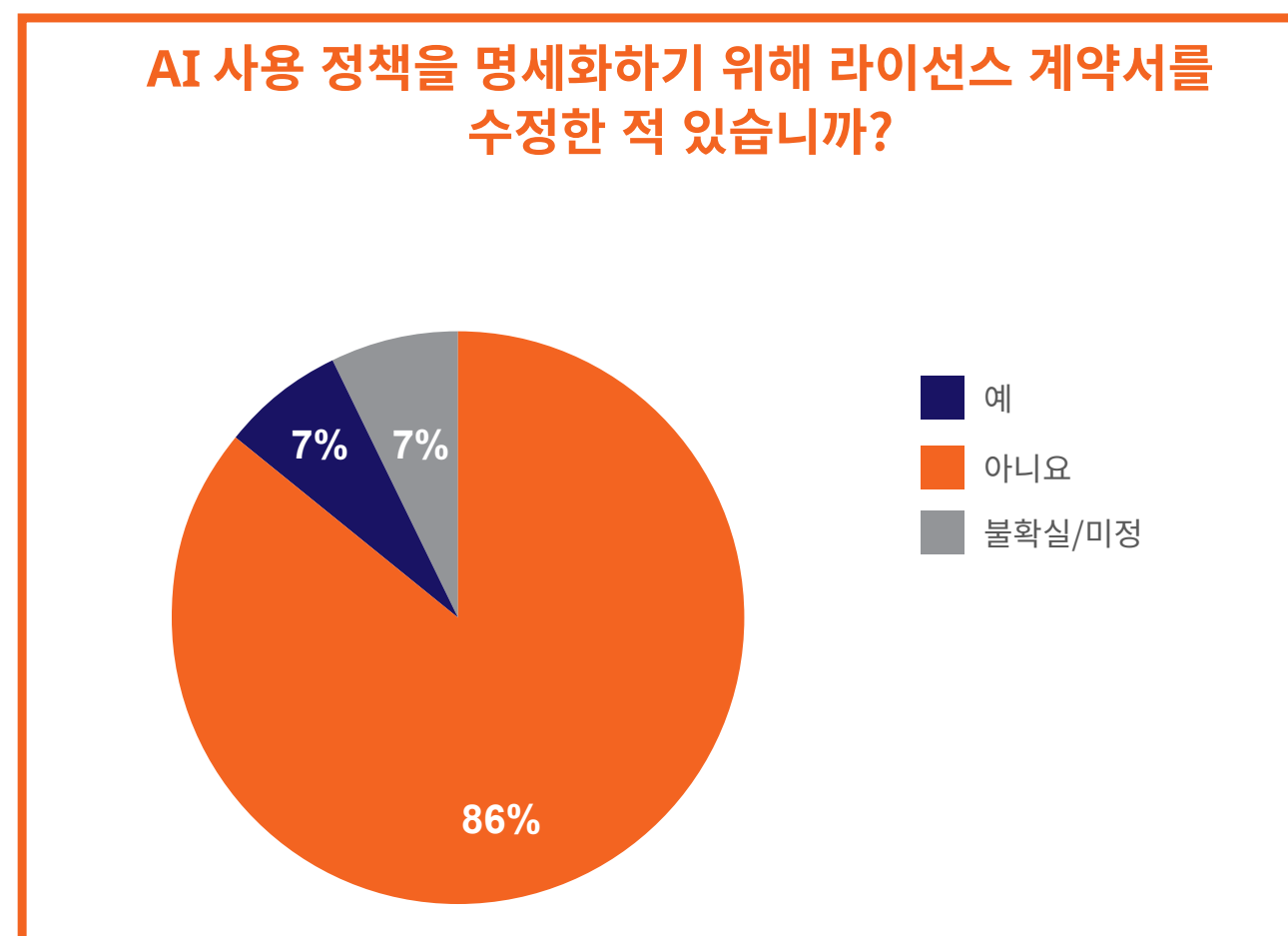
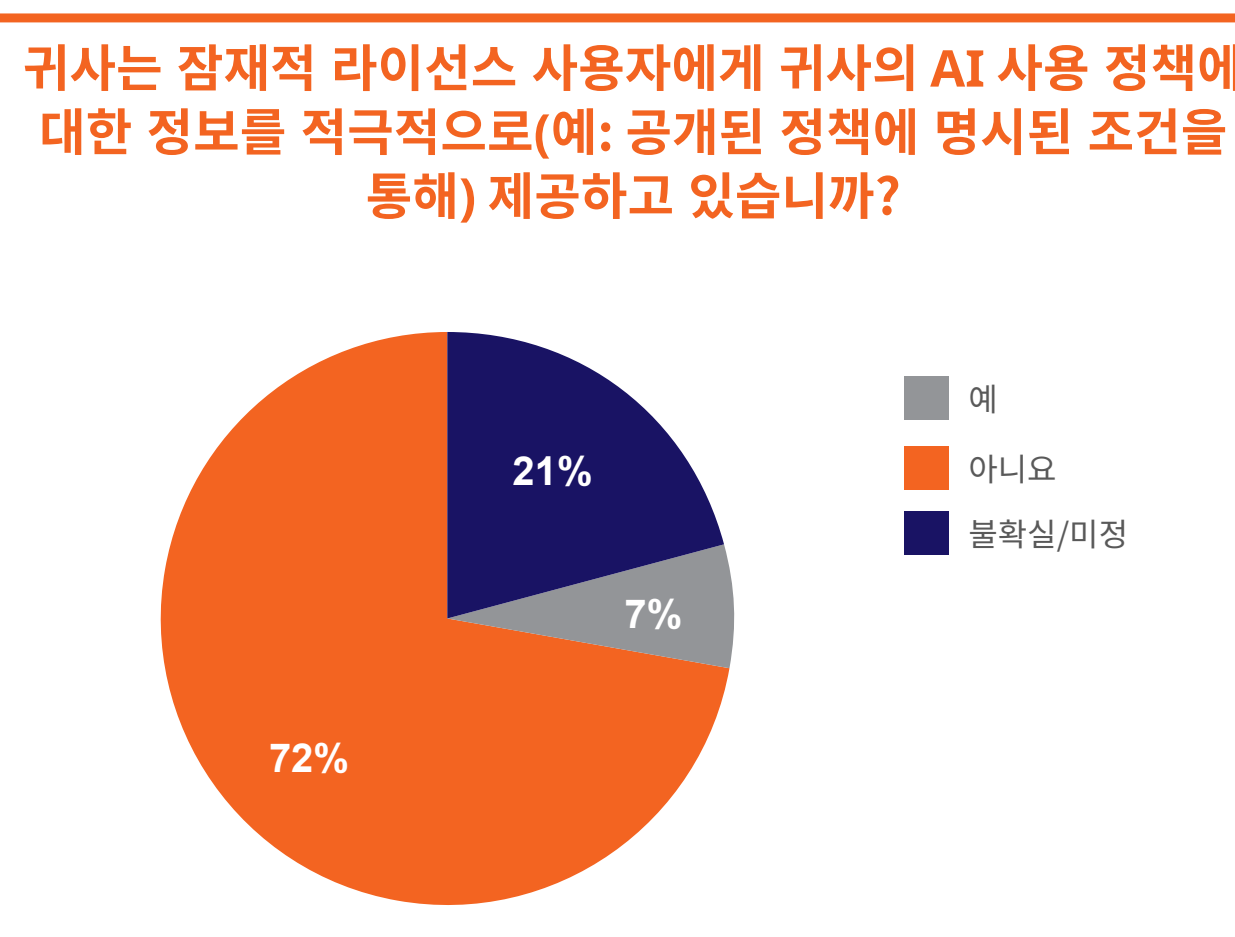
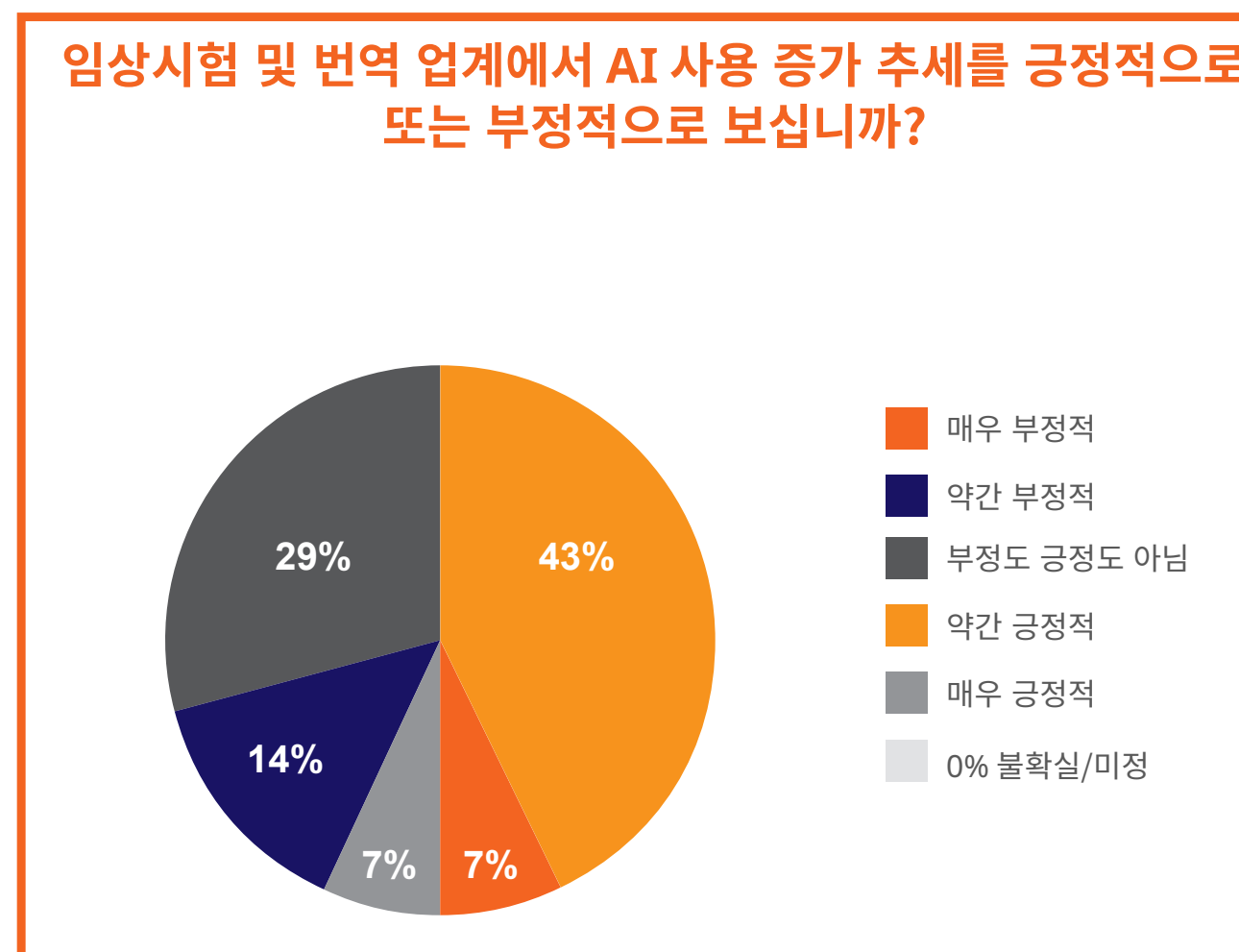
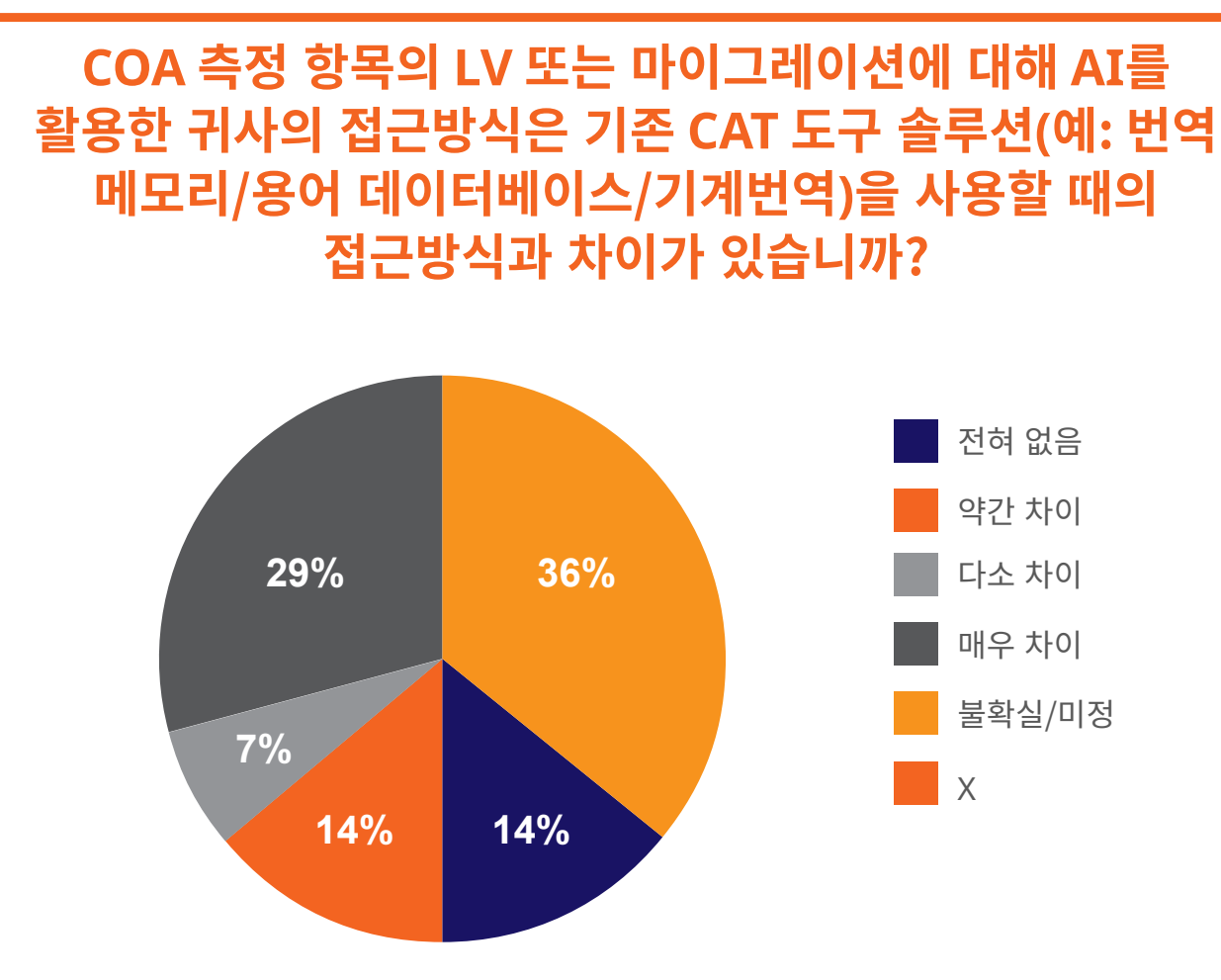
3) 현재 AI 접근방식

a. 결과: 이 주제에 대해 4가지 질문을 했다.

1) LV 및 마이그레이션에서 AI 사용에 대한 접근방식이 CAT 도구 솔루션 사용 접근방식과 얼마나 다른지 물었을 때 대다수(36%[5/14])가 '잘 모르겠음/미정'이라고 답했다. 이에 비해 28.5%(4/14)는 '매우 다름'이라고 답했다. 이 기술에 CAT 도구 기술과 다른 방식으로 접근하지 않거나 약간만 다르게 접근할 것이라고 답한 응답자는 28.5%(4/14)에 불과했다.

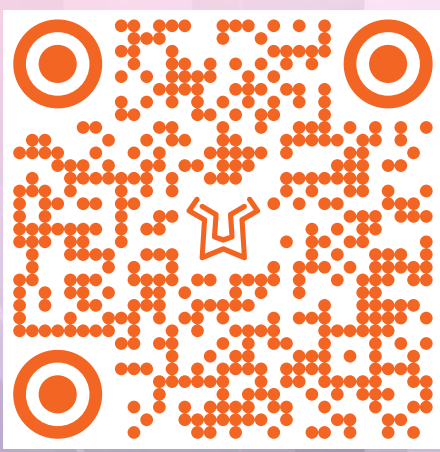
2) 임상시험 및 번역에서의 AI 활용에 대해 43%(6/14)는 '약간 긍정적'이라고 답했으며, 29%(4/14)는 '부정적이지도 긍정적이지도 않음'이라고 응답했다. '매우 부정적'과 '매우 긍정적'은 각각 7%(1/14)로 유사한 결과를 보였으며, 나머지 14%(2명)는 '약간 부정적'이라고 답했다.

3,4) 공개 포럼에서 AI 사용 정책을 적극적으로 제공하고 있는지 물었을 때 72%(10/14)가 '아니요'라고 답했다. 또한 86%(12/14)는 라이선스 계약에서 AI 사용 정책을 명시화하도록 수정하지 않았다.



b. 해석: 이 기술과 기존 번역 기술의 차이점, 그리고 이러한 차이점을 고려해 접근방식을 조정해야 하는 방식에 대해 다소 불확실성이 존재하는 것으로 보인다. 또한 업계 내에서 AI 사용에 대해 매우 다양한 의견이 존재한다. 그러나 이 주제에서, 그리고 전체 설문조사에서 가장 눈에 띄는 결과는 라이선스 계약서 문서조차도 잠재적 라이선스 사용자에게 자사 IP 사용에 대한 AI 정책을 적극적으로 알리지 않는다는 점이었다. 다만 서술식 응답에 따르면 이러한 상황도 곧 변화할 것으로 보이므로, 본 연구 그룹은 라이선스 사용자가 저작권자와 협상할 때 정보 부재는 의견이나 정책의 부재를 의미한다고 가정할 것을 권고한다.





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

서론

전자 임상결과평가(eCOA)의 번역, 마이그레이션 및 스크린샷 검토는 부분적으로 화면상에서 콘텐츠를 수정하기 어렵다는 점 때문에 이전부터 사람의 수작업에 크게 의존해 왔다. 스크린샷 검토 프로세스(SSR) 중에는 원문 마스터 화면 보고서 및 해당되는 경우 레거시 콘텐츠(즉, 원문 종이 설문지)와 대조 확인하여 대상 언어 화면 보고서를 작성한다.

라이온브리지(Lionbridge)는 단순 SSR과 복합 SSR의 두 가지 유형으로 스크린샷 검토를 수행한다. 복합 SSR은 모든 대상 레거시 콘텐츠에 대해 명시적으로 한 줄씩 대조 확인한다는 점에서 단순 SSR과 차이가 있다. 라이온브리지는 생성형 AI(GenAI) 및 광학문자인식(OCR) 기술을 활용했을 때 기존 스크린샷 검토 방식을 보완하고 비용 및 소요 시간(즉, 검토 횟수)을 줄이며 품질을 향상시켜 궁극적으로 환자 결과를 개선할 수 있는지 확인 검증했다. 본 연구는 라이온브리지의 독자적인 Aurora AI Clinical Outcomes™ 도구의 단순 및 복합 스크린샷 검토 기능을 활용하여 수행되었다.

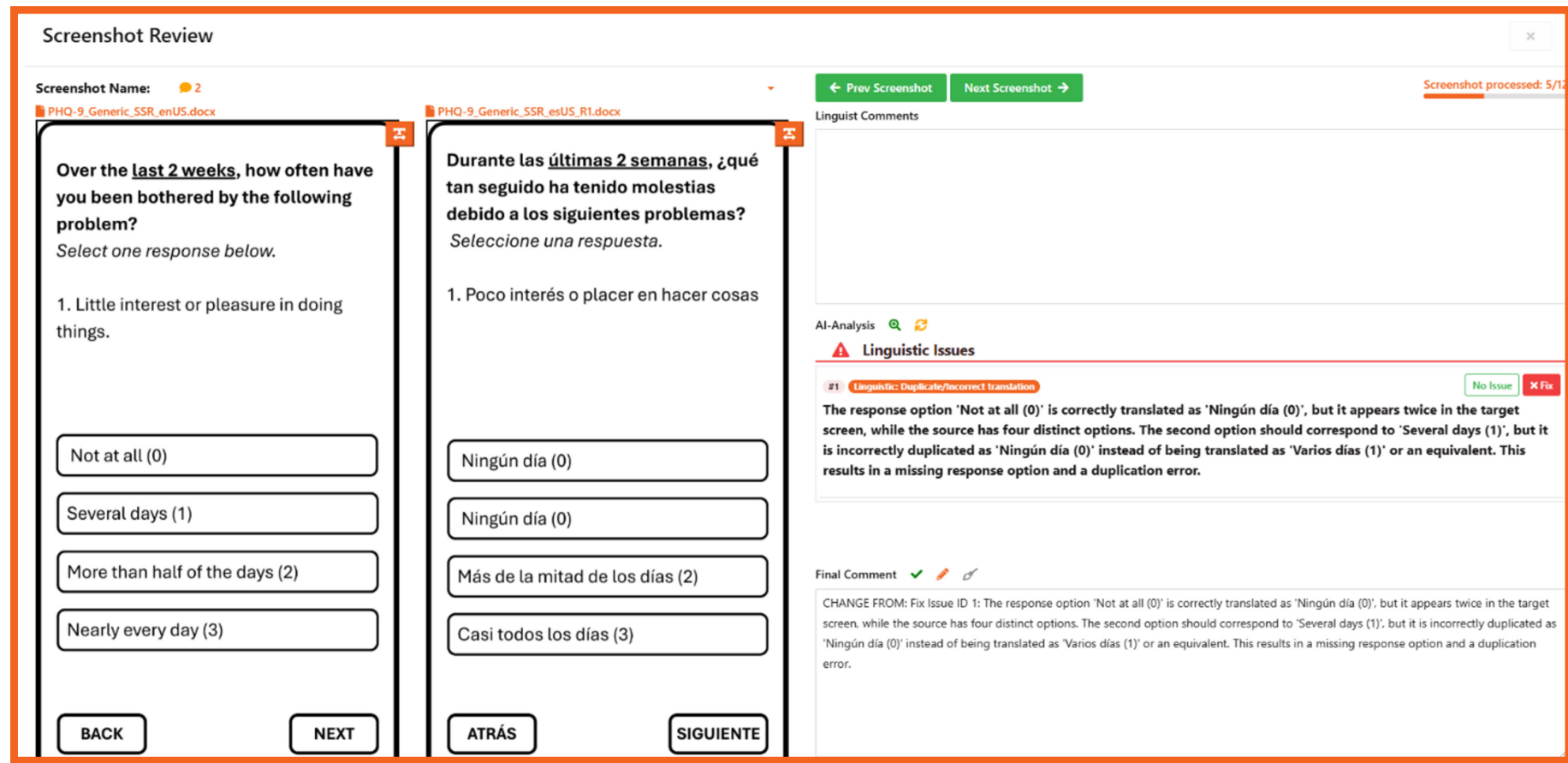
방법

라이온브리지는 보안이 강화된 GenAI 엔진을 활용하여 5가지 환자 대상 전자 임상결과평가(eCOA)의 대상 스크린샷 보고서에 대한 품질보증 피드백을 생성했다. 사용된 eCOA는 길이와 복잡도가 각각 달랐으며 테스트 목적으로 11가지 유형의 오류(누락된 콘텐츠, 번역되지 않은 콘텐츠, 오역, 누락된 eCOA 수정사항, 버전 번호 불일치, 서식 문제, 형태 문제, 태그 문제, 줄바꿈 문제, 대소문자 문제, 척도 기준점 문제)가 의도적으로 포함되었다. 대상 언어로는 불가리아어(불가리아), 폴란드어(폴란드), 루마니아어(루마니아), 그리스어(그리스), 스페인어(아르헨티나, 멕시코, 미국), 프랑스어(벨기에, 프랑스), 포르투갈어(브라질, 포르투갈), 헝가리어(헝가리), 터키어(터키), 한국어(한국), 중국어 번체(대만), 태국어(태국) 등 다양한 문자 체계와 언어군을 대표하는 총 16개의 언어가 테스트되었다. 최신 eCOA 산업 표준 및 확립된 관행에 부합하는 적절한 결과물을 얻을 때까지 프롬프트를 맞춤화했다. 동시에 동일한 eCOA 화면 보고서를 인적 검토를 위해 라이온브리지 언어전문가에게 전송했다. 다음으로 중립적인 검토자가 AI 및 사람의 두 결과물을 검증하고 정확성 및 완전성에 대해 평가했다.

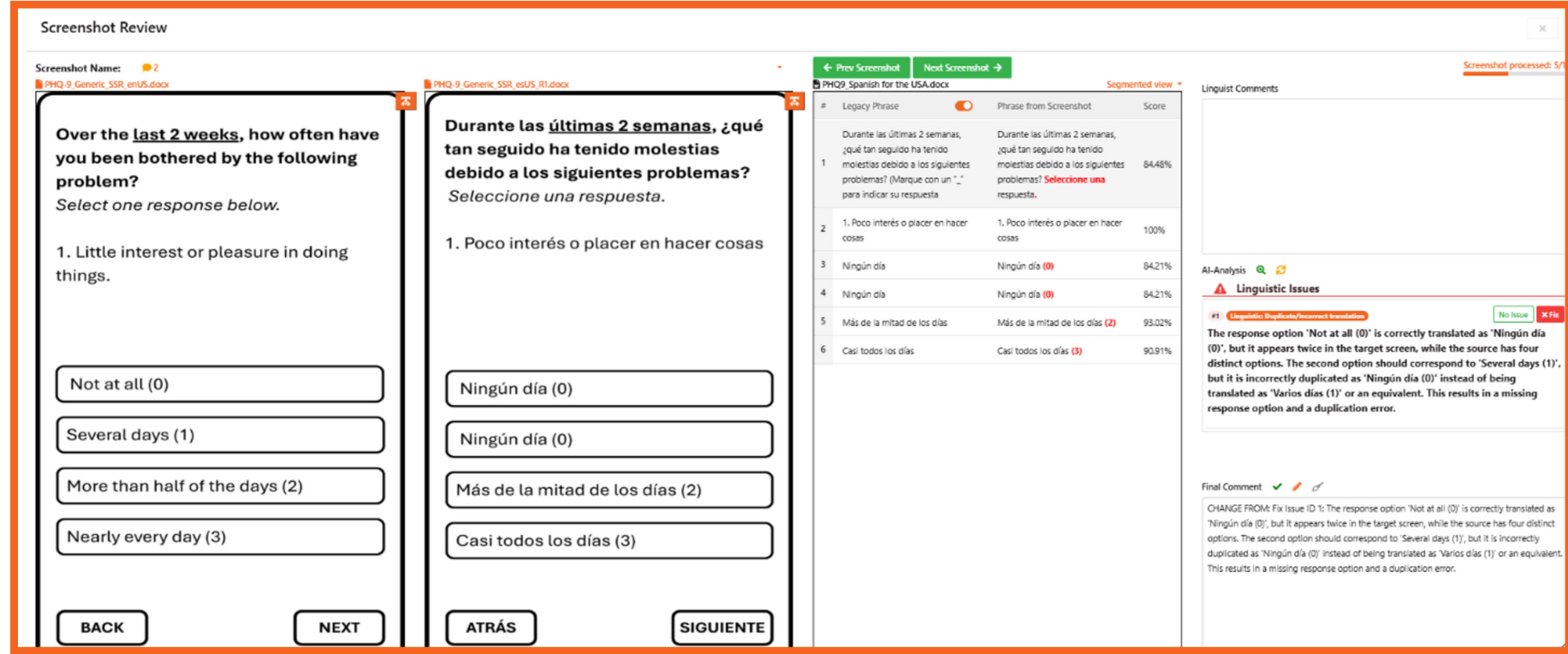
프롬프트 설계 진화:

본 연구의 첫 단계에서는 번역 워크플로(번역/편집 또는 순번역/역번역) 이후 원문 화면 보고서와 대상 화면 보고서를 비교하는 데 이상적인 '단순 스크린샷 검토'용 GenAI 프롬프트를 활용했다. 이 도구의 기능이 최종적으로 구축된 후에는 보다 정교한 기능인 복합 스크린샷 검토 GenAI 기능의 개발과 테스트에 집중했다. 단순 스크린샷 검토에서 한 단계 더 나아가 이 기능은 콘텐츠 마이그레이션에 더 적합했다. 복합 SSR 기능은 대상 화면 콘텐츠를 원문 마스터 화면 보고서와 비교하는 것뿐만 아니라 대상 화면 콘텐츠를 대상 레거시 콘텐츠와도 비교한다.

단순 스크린샷 검토 인터페이스:



복합 스크린샷 검토 인터페이스:



결론

본 연구는 eCOA 번역 및 마이그레이션의 스크린샷 검토에 있어 가장 효과적인 전략이 GenAI와 인적 검토의 통합이라는 것을 입증했다. 프롬프트는 시간 경과에 따라 더 많은 문제를 탐지할 수 있도록 지속적으로 개선된다. 그러나 이처럼 고도로 민감한 콘텐츠에서 최적 품질을 확보하기 위해서는 스크린샷 검토 과정에 여전히 사람의 개입이 권장된다. GenAI는 오류를 예비 식별함으로써 속도, 효율성, 정확성을 크게 향상시킬 수 있지만, AI가 놓칠 수 있는 문제를 식별하고 보다 세밀하게 의사결정을 내리기 위해서는 사람의 전문성이 여전히 필수적이다. 이러한 '휴먼인더루프' 방식을 통해 필요한 검토 횟수를 줄여 일정을 간소화하고, 언어전문가의 자체 품질 확인 시간을 줄여 비용을 절감하며, 무엇보다도 더 정밀하고 높은 품질로 환자의 의견을 포착함으로써 보다 신뢰할 수 있는 환자 중심 경험을 보장할 수 있다.

결과 및 해석

단순 SSR GenAI 기능에 대한 품질보증 피드백 분석 결과, 이 기능은 번역 및 마이그레이션 프로젝트 모두에서 사람의 피드백과 조합하여 활용할 수 있는 매우 유용한 QA 점검 도구로 나타났다. 단순 SSR 기능은 특히 기존 콘텐츠와의 교차 확인이 필요하지 않은 번역/스크린샷 검토 프로젝트에서 언어전문가들이 활용할 수 있는 보조 도구로 매우 적합하다. Aurora AI 임상결과 도구는 이 기능을 활용하여 대상 화면에서 11가지 오류 유형 모드를 여러 차례 포착했다. 그러나 일부 사례에서는 레거시 콘텐츠를 확인할 수 없다는 한계가 있었다.

복잡한 SSR GenAI 기능은 아직 개발 중이지만(2026년 1분기 배포 예정), 중간 결과에 따르면 이 기능은 단순 SSR 도구가 식별한 모든 오류를 포함해 더 많은 오류를 검출했다. 또한 단순 SSR 도구보다 상세한 피드백을 제공한다. 레거시 콘텐츠와의 교차 검증 기능을 갖추고 있으므로 아래 차트에 나타난 바와 같이 복합 SSR 도구는 현재 단순 SSR 도구가 생성하는 일부 오탐을 제거할 수 있다.

오류 유형	단순 SSR GENAI 기능: 참고사항 및 한계	복합 SSR GENAI 기능: 참고사항 및 한계	예
누락된 콘텐츠	대상 콘텐츠에서 누락된 부분에 표시 가능	레거시 버전에는 없으므로 의도적으로 누락된 텍스트인지 확인 가능	<div>Portuguese (Portugal)</div> <div>#4 [Redundancy/ Missing content] Target Paragraph: 1</div> <div>No Issue</div> <div>Fix</div> <div>"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."</div>
번역되지 않은 콘텐츠	영어로 남아 있는 텍스트를 표시하고 의도적인 경우인지 (예: 저작권 문구) 제안할 수 있음	의도적으로 영어로 남아 있는 텍스트인지 확인할 수 있음	<div>Bulgarian (Bulgaria)</div> <div>Linguistic Issues</div> <div>#1 [Linguistic/Untranslated content]</div> <div>No Issue</div> <div>Fix</div> <div>"The line '[scale name] - Items H17, BP1, N6, GE6' remains in English in the Bulgarian screen. This should be translated or localized for Bulgarian users unless these are standardized item codes that should remain in English. If these are not standard codes, provide a Bulgarian translation."</div>
오역	원문에서의 편차를 식별할 수 있음	원문과 일치하더라도 레거시와 차이가 있는 경우 오역임을 식별할 수 있음	<div>Portuguese (Portugal)</div> <div>#9 [Linguistic/ Inconsistent scale anchors]</div> <div>No Issue</div> <div>Fix</div> <div>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 'A little bit', and '2 - Mais ou menos' (More or less) does not 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.</div>
eCOA 수정사항 누락	단순 및 복합 기능에 대해 동일한 기능이 예상됨		<div>Portuguese (Portugal)</div> <div>#1 [Linguistic/ Incorrect translation] Source Paragraph: 6</div> <div>No Issue</div> <div>Fix</div> <div>Target Paragraph: 4</div> <div>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.</div>
버전 번호 불일치	원문과 대상 간의 버전 번호 불일치를 표시할 수 있음 (불일치가 의도적인 것일 수 있다는 경고 포함)	레거시를 확인하여 버전 번호 차이가 의도적인 것인지 확인할 수 있음	<div>Korean (Republic of Korea)</div> <div>#3 [Redundancy/ Version number mismatch]</div> <div>✓ No Issue</div> <div>Fix</div> <div>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.</div>
서식 문제	원문 마스터 화면 보고서와 서식이 다른 부분을 표시할 수 있음	원문과 서식이 다른 경우 레거시 콘텐츠를 기준으로 이러한 불일치가 의도적인 것인지 확인할 수 있음	<div>Portuguese (Portugal)</div> <div>Formatting/Layout Issues</div> <div>#1 [Formatting/Layout/ Inconsistent bolding/Underlining] Source Paragraph: 7</div> <div>No Issue</div> <div>Fix</div> <div>Target Paragraph: 5</div> <div>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.</div>
형식 문제	형용사의 예상치 못한 성별 형태를 감지할 수 있음(의도적인 것일 수 있다는 경고가 포함됨 [예: 환자군이 100% 여성으로 이루어진 경우 모든 형용사가 여성형일 수 있음]).	레거시와 대조 확인하여 예상되는 성별 형태에 대한 추가적인 컨텍스트를 제공할 수 있음	<div>Spanish (United States)</div> <div>Linguistic Issues</div> <div>#1 [Linguistic/ Inconsistent translation] Source Paragraph: 7</div> <div>No Issue</div> <div>Fix</div> <div>Target Paragraph: 5</div> <div>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.</div>
태그 문제	단순 및 복합 기능에 대해 동일한 기능이 예상됨		<div>Thai (Thailand)</div> <div>Formatting/Layout Issues</div> <div>#1 [Formatting/Layout/ Underlining/HTML Tag Display] Source Paragraph: 7</div> <div>No Issue</div> <div>Fix</div> <div>Target Paragraph: 7</div> <div>The target screen displays HTML tags (<u>) instead of rendering underlined text for 'nondisruptive', 'unfazed', and 'unobtrusive'. 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.</div>
줄바꿈 문제	단순 및 복합 기능에 대해 동일한 기능이 예상됨		<div>French (France)</div> <div>Formatting/Layout Issues</div> <div>#1 [Formatting/Layout/ Line break]</div> <div>No Issue</div> <div>Fix</div> <div>Source Paragraph: 6</div> <div>Target Paragraph: 4</div> <div>The target screen introduces a line break in 'habi-tuelles' 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 'habituels' appears on one line without a hyphen.</div>
대소문자 사용 문제	원문과 다른 예상치 못한 대소문자 패턴을 표시할 수 있음 (의도적인 것일 수 있다는 경고 포함)	대상 화면에서의 예상치 못한 대문자 사용이 레거시 버전과 일치하는지 확인할 수 있음	<div>Hungarian (Hungary)</div> <div>#2 [Linguistic/ Inconsistent capitalization]</div> <div>No Issue</div> <div>Fix</div> <div>Source Paragraph: 6</div> <div>Target Paragraph: 6</div> <div>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 részére'.</div>
척도 기준점 문제	단순 및 복합 기능에 대해 동일한 기능이 예상됨		<div>Portuguese (Portugal)</div> <div>#2 [Formatting/Layout/ Scale anchor placement]</div> <div>No Issue</div> <div>Fix</div> <div>Source Paragraph: 8</div> <div>Target Paragraph: 7</div> <div>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.</div>

인간 대 기계:

- GenAI는 대상 화면 콘텐츠 평가에 있어 인간보다 훨씬 빠르다. 주어진 척도의 모든 대상 화면에 대해 몇 초 만에 피드백을 생성한다.
- 라이온브리지 품질보증 전문가들은 Aurora 임상결과 도구 SSR 인터페이스에서 레거시 파일, 마스터 화면 보고서, 대상 언어 화면 보고서에 손쉽게 동시 접근할 수 있으므로 GenAI 과정의 시간과 효율이 향상되었다고 언급했다.
- GenAI는 특히 줄바꿈, 대소문자 사용, 잘못된 번역(중복 텍스트 포함), 척도 기준점 등 검토자인 사람이 놓친 여러 문제를 포착했다. GenAI는 광학문자인식(OCR)을 사용하여 사람의 눈으로는 놓치기 쉬운 편집 불가능한 이미지 콘텐츠 내 텍스트 불일치를 분석하므로 이러한 문제들은 포착하기 더 쉬울 수 있다.
- GenAI SSR에서 포착한 오류 중에는 언어 전반에서 일관성이 없을 때도 있었다. 예를 들어, 대부분의 언어에서 "items"의 번역이 누락된 것을 발견했지만 스페인어에서는 이를 검출하지 못했다. 이는 약센트 부호("items") 차이 때문일 가능성이 높다.
- 검토자인 사람은 문맥에 관련된 문제(예: 여성 환자군, 의도적인 버전 번호 불일치, 의도적으로 영어로 남겨진 저작권 문구 등)를 잘못 표시할 가능성이 더 낮았다.