cognizant®

The case for

AI-forward eCOA localization

A modernized linguistic validation framework for 2026

Contents

Abstract

The linguistic validation (LV) methodology codified by Wild et al. in 2005 established the gold standard for translating clinical outcome assessments (COAs) for use in global clinical trials—rigorous, sequential and built on the premise that nuanced cross-cultural language reasoning required human cognition. Modern large language models (LLMs) have rendered that premise obsolete. They now demonstrate translation quality, semantic consistency, cultural awareness and self-correction that equal or surpass human performance across clinical instrument translation tasks.

This paper proposes the AI-forward LV model: a redesigned framework in which AI agents perform all core translation, reconciliation, back-translation and verification steps, with humans retained at cognitive debriefing and final certification. The methodology—every step of Wild et al.—is preserved. What changes is who performs each step, and why.

> Applying a 2005 methodology unchanged in 2026 is not caution. It is inertia.

The 2005 framework: Sound methodology for a different era

Wild et al. established the foundational methodology for COA translation—a rigorous multistep process accepted as best practice by the FDA, EMA and the outcomes research community. Its logic was sound: translating nuanced clinical language requires human judgment, cultural knowledge and empathetic reasoning. In 2005, only human translators could provide this.

The framework established a sequential process—dual forward translation, reconciliation, back-translation, PM review, cognitive debriefing and certification—that remains largely unchanged in current practice (Fig. 1). Each step shares a common design logic: Human oversight is necessary because human cognition is the only available mechanism for the semantic, cultural and empathetic reasoning the process requires.

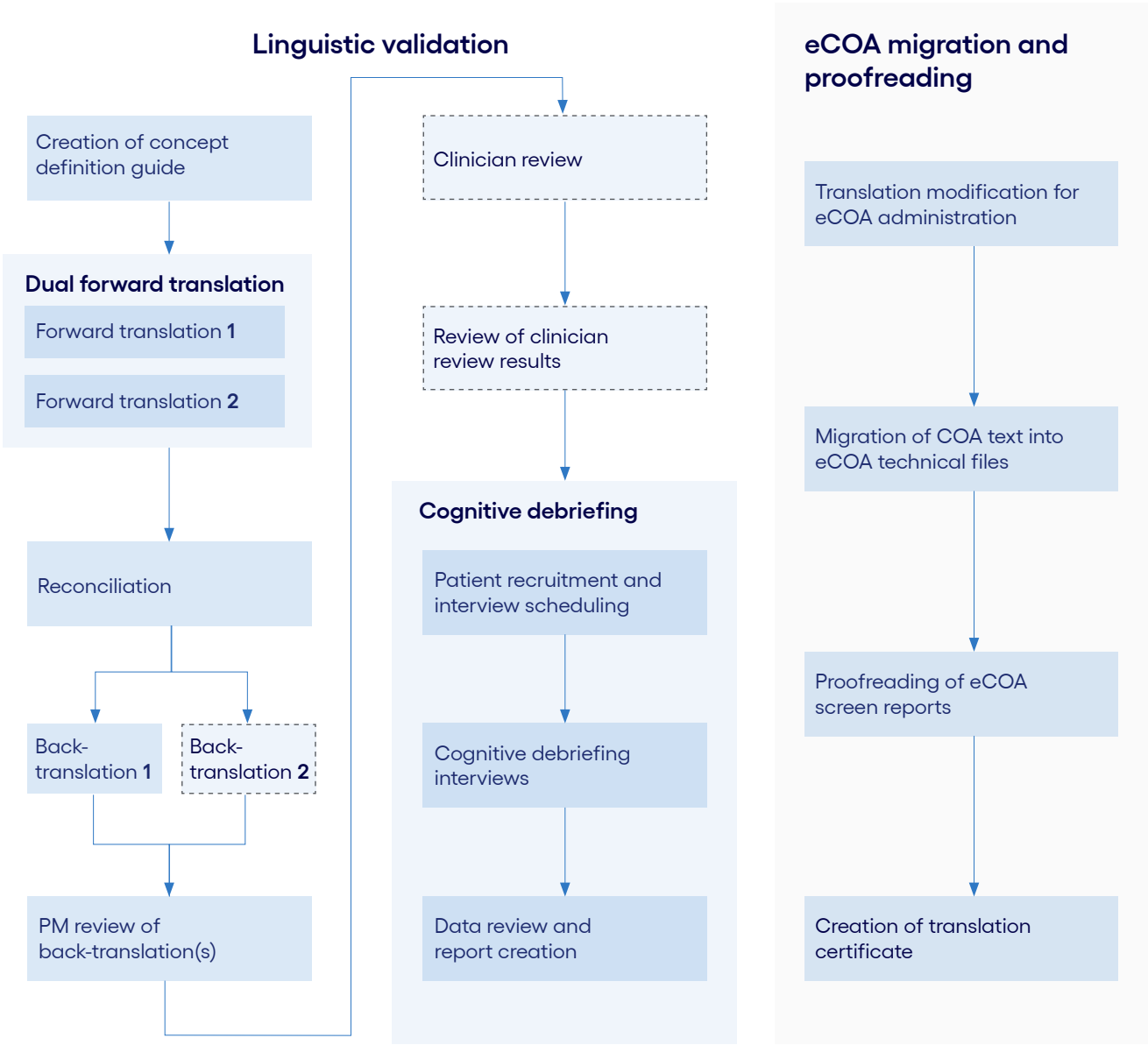

Fig. 1: Linguistic validation process map (reproduced from McKown et al.). Dotted lines indicate optional steps.

> The 2005 framework did not involve humans because humans were inherently superior translators. It involved humans because in 2005 there was no alternative capable of the required cognitive tasks. The premise has changed. The framework has not.

What LLMs can do in 2026 that they could not do in 2005

Translation quality

Lu et al., using METEOR scoring, demonstrated that LLMs produce translations statistically approaching human expert quality for high-resource languages on validated PRO measures across six languages. More significantly, Johnson et al. showed that GPT-4o performing comparative review of back-translations against source text actually outperformed human translators with over five years of experience. COA language is among the most demanding translation work that exists—precise enough for regulatory scrutiny, natural enough for patients with limited health literacy.

Cultural reasoning

Modern LLMs are trained on text from billions of people across hundreds of cultures—medical literature, patient forums, clinical trial reports and lay illness narratives in dozens of languages simultaneously. Their cultural knowledge is not experiential, but neither is that of most professional translators working outside their native tongue. Kunst and Bierwiaczonek noted that AI translations are already appropriate as primary translations for many language types.

Consistency, audit trails and error detection

On consistency, AI surpasses human capability categorically. A translator working on item 47 of a 60-item questionnaire at the end of a long day will not perform at the same level as on item one. An LLM will. A human reviewer may introduce confirmation bias from anchoring to the translation they helped produce. An AI reviewer has no such investment. Every AI decision is also logged, time-stamped and auditable—documenting not just what translation was chosen, but what alternatives were considered and what semantic concerns were flagged. No human translator produces this level of documentation as a byproduct of their work.

Capability	Human translators (2005–2026)	LLMs (2026)
Translation accuracy (common languages)	High; variable across individuals	High; approaching human benchmark
Cultural reasoning	High (native); moderate (non-native)	Moderate-high; improving rapidly
Rare language support	Inconsistent; specialist scarcity	Developing; improving faster than supply
Consistency across items	Variable (fatigue, context shift)	High; no fatigue or anchoring bias
Speed per language	Days to weeks	Minutes
Audit trail	Manual; incomplete	Complete; machine-generated
Back-translation comparison	Manual; subject to bias	Systematic; exceeds human in studies
Patient empathy (CD interviews)	Essential; irreplaceable	Not recommended; human required

Comparative capability assessment: Human translators vs. LLMs in COA translation tasks (2026)

The cost of inertia: Timeline and access

Timeline: Current LV takes six to 12 weeks per language pair. A global trial in 20 countries requires months of localization delay before a single patient can be enrolled. The AI-forward model compresses this to under two days per language—not a marginal efficiency gain, but a structural transformation in what is operationally possible.

Access: The 2005 framework was designed around high-resource languages with large translator pools. It works well for French, German and Spanish; it works poorly for Zulu, Malayalam and Tagalog—languages spoken by populations in the emerging markets that trials increasingly need to include. There are simply not enough qualified clinical translators in low-resource languages to run the current process at scale. LLMs do not face this constraint, and their low-resource language performance is improving far faster than the global supply of certified clinical translators.

The AI-forward LV model: A proposed framework for 2026

We propose a redesigned framework that retains every step of the Wild et al. methodology while inverting who performs each step. AI acts as primary agent for all translatable steps; humans are retained at two points of irreducible complexity—cognitive debriefing (requiring empathetic patient interaction that AI cannot replicate) and final certification (requiring professional accountability). AI generates a complete, machine-readable audit trail at every step, and AI use is disclosed in the translation certificate at step level.

AI-forward linguistic validation: Redesigned process

AI-primary | AI + human | Human-primary

01 AI-primary

Dual forward translation

AI role: Two AI agents with independent sessions and varied temperature → two forward translations

Human validation occurs in steps 5 and 6

02 AI-primary

Reconciliation

AI role: AI reconciliation agent—item-by-item comparison, selects preferred rendering, documents reasoning

Human role: Not needed

03 AI-primary

Back-translation

AI role: AI back-translates without source in context; AI comparison agent flags semantic drift

Human role: Not needed

04 AI + human

Review of back-translation

AI role: AI produces structured review report in eCOA platform; categorizes each item and proposes corrections

Human role: Human approves or overrides corrections; effort proportional to flagged items

05 Human-primary

Cognitive debriefing

AI role: AI generates interview guides, auto-transcribes and codes data, identifies patterns, drafts data report

Human role: Interviews entirely human-led; human expert signs off data review report

06 AI + human

Finalization and certification

AI role: AI auto-generates certificate, model versions, prompts, quality scores, audit log, flagged items

Human role: Qualified human reviews and signs; takes professional and legal accountability

AI use is explicitly disclosed at each step. All steps produce a verifiable audit log; a qualified human expert reviews and signs the final translation certificate.

Fig. 2: AI-forward LV model: AI-primary steps (blue) and human-led steps (light blue).

Process step	Wild et al. 2005	AI-forward LV 2026	What stays the same
Dual-forward translation	Human translators	AI agents (two independent sessions)	Two independent outputs; bias detection
Reconciliation	Human expert	AI agent	Single optimized output produced
Back-translation	Human translator (blind)	AI agent (blind session)	Source blindness; semantic verification
PM review	Human PM/reviewer	AI comparison + human sign-off	Discrepancies detected and resolved
Cognitive debriefing	Human interviewer + patients	Human-led; AI supports admin/ coding	Patient voice; comprehension verified
Certification	Human expert produces and signs	AI generates; human signs off in eCOA platform	Human accountability; full audit trail

Process ownership comparison: Wild et al. (2005) vs. AI-forward LV model (2026)

Steps 1–4 are AI-primary with human validation at critical handoff. Step 6 certification is AI-documented and human-signed, producing a more verifiable audit record than any purely human-generated certificate—logging not just what decision was made but why, with alternatives considered and quality scores at every step.

Regulatory pathway: Working with the guidelines, not against them

The FDA has not issued guidance prohibiting AI use in COA translation. Its guidance references methodological rigor, not human primacy. The FDA's 2023 AI guidance (with Health Canada and MHRA) emphasizes documentation, validation and transparent accountability—all hallmarks of the AI-forward model. The EMA's 2024 guidance on LLMs in regulatory science called for appropriate safeguards around data privacy and hallucination risks; it did not prohibit AI use. McKown et al. found that pharmaceutical sponsors considered regulatory rejection risk manageable through robust documentation and human oversight.

The AI-forward model offers regulators something the current model cannot: full process transparency and complete reproducibility. Current LV practice has well-documented quality variability—two nominally identical projects can produce substantially different outputs depending on which translators were engaged. An AI-generated certificate with machine-logged decision provenance at every step is more verifiable than a human-generated certificate that summarizes outcomes without documenting reasoning.

> **The argument is not about trusting AI because it is AI. It is about trusting this translation because every decision it contains is documented, justified, reviewable and reproducible.**

Recommended positioning: Frame AI as the production mechanism, not a deviation from methodology; disclose AI use completely in submission documentation; preserve human sign-off as the point of professional and legal accountability; and initiate pre-submission meetings with FDA and EMA to discuss AI-forward LV on a study-by-study basis. Guidance follows evidence.

Addressing the principal objections

"AI cannot understand what patients actually experience"

The vast majority of human translators have not experienced the conditions they translate for—they bring linguistic competence and cultural knowledge, not phenomenological expertise. The cognitive debriefing step, which the AI-forward model preserves unconditionally, is precisely the mechanism designed to catch failures of lived-experience translation. The process architecture accounts for this limitation.

"IP and data privacy risks are too high"

The AI-forward LV model requires deployment on dedicated, isolated AI infrastructure—not public-facing LLM services. All COA content must be processed within a secure, proprietary deployment, purged after each session, with data processing agreements in place consistent with GDPR, HIPAA and applicable regional regulations. McKown et al. confirmed this deployment model presents no greater IP risk than the current practice of sharing COA content with 10-plus human collaborators across multiple vendor organizations.

"This eliminates professional roles"

The model changes, not eliminates, LV work. Production translation volume decreases; quality review and AI oversight work increases. Roles evolve (production translator becomes AI output reviewer and quality auditor) and new ones emerge (AI systems specialist, LV-regulatory liaison). The skill profile of LV professionals will shift toward AI oversight and regulatory expertise, not toward obsolescence.

Implementation guidance for early adopters

01 Phase (1–3 months)

Run AI-forward LV in parallel with the existing process. Configure the dedicated AI deployment environment, develop standardized prompting protocols, build the quality evaluation framework and generate comparative data for publication. Do not use AI-forward outputs as primary deliverables.

02 Phase (4–6 months)

Introduce AI-primary production for forward translation and back-translation on internal or observational studies. Maintain human review of all AI outputs. Present interim data at ISOQOL, ISPOR and DIA forums to generate peer review and community engagement.

03 Phase (7+ months)

Following successful validation and regulatory engagement, deploy AI-forward LV as the primary methodology for qualifying studies. Maintain full human oversight at cognitive debriefing and certification. Publish validation findings in peer-reviewed literature to build the evidence base for industry-wide adoption.

Conclusion: Methodological courage in a regulated world

The Wild et al. framework was a landmark contribution. It deserves acknowledgment—and it deserves to be updated. The evidence is clear: LLMs now match or exceed human performance on the core translatable steps of LV, deliver complete audit trails that current practice cannot, and scale to language pairs that the human-centric model structurally cannot serve.

The question for the field is not whether AI-forward LV is theoretically sound. The question is whether the field has the methodological courage to validate it, publish it and present it to regulators—rather than perpetuating a 20-year-old framework because careers were built around it and no one has yet been required to justify it against the technology that now exists.

> In 2026, the best translation methodology produces the highest quality output in the shortest time with the most complete audit trail. That methodology puts AI at the front and the human at the gate.

The window for first-mover advantage in AI-forward LV methodology is open. It will not remain open indefinitely.

References

- Johnson, M. et al. “Gen AI Powered Precision: Revolutionizing Comparative Review in Clinical Outcome Assessment Localization.” Value in Health 27(12):S50. 2024. www.valueinhealthjournal.com/article/S1098-3015(24)03129-2/fulltext
- European Medicines Agency. “Guiding principles on the use of large language models in regulatory science and for medicines regulatory activities.” 2024. www.ema.europa.eu/en/documents/other/guiding-principles-use-large-language-models-regulatory-science-medicines-regulatory-activities_en.pdf
- Kunst J.R., and Kinga Bierwiaczonek. “Utilizing AI questionnaire translations in cross-cultural and intercultural research.” International Journal of Intercultural Relations 97:101888. 2023. www.sciencedirect.com/science/article/pii/S0147176723001360
- Lu, Sheng-Chieh, et al. “Can machine translation match human expertise? Quantifying the performance of large language models in translation of patient-reported outcome measures.” Journal of Patient-Reported Outcomes 9(1):94. 2025. pubmed.ncbi.nlm.nih.gov/40711496/
- McKown, Shawn, et al. “Use of AI within COA linguistic validation and eCOA migration processes: analysis and good practice recommendations.” Journal of Patient-Reported Outcomes 10:34. 2026. pmc.ncbi.nlm.nih.gov/articles/PMC12936314/
- Wild, Diane, et al. “Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation.” Value in Health 8:94–104. 2025. onlinelibrary.wiley.com/doi/full/10.1111/j.1524-4733.2005.04054.x
- Wolk, Krzysztof, et al. “Enhancing the Assessment of (Polish) Translation in PROMIS Using Statistical, Semantic, and Neural Network Metrics.” WorldCIST'18. 2018. link.springer.com/chapter/10.1007/978-3-319-77712-2_34

Authors

Kavitha Lokesh
Vice President, SBU Head,
ISG Life Sciences
R&D and Commercial

Bashir Ahmed
Program Delivery Director,
ISG Life Sciences R&D

This position paper was prepared by the Cognizant ISG LS R&D group. It represents the professional and analytical views of the authors and does not constitute regulatory guidance. Organizations considering changes to their LV methodology should consult with regulatory affairs specialists and conduct appropriate validation studies before adopting any new approach in submission-critical contexts.

WF 4741750