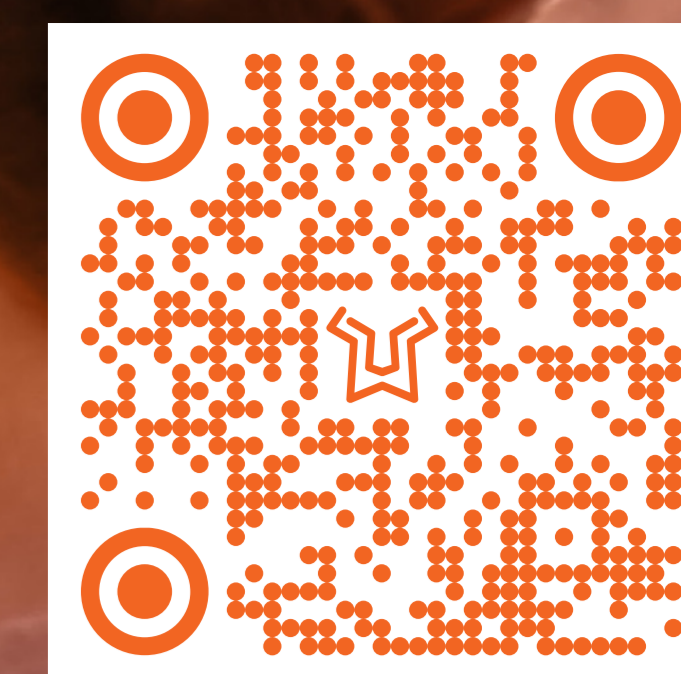


Il futuro della convalida linguistica è qui:

migliorare la conformità del processo di riconciliazione tramite GenAI



LIONBRIDGE

Pearson

Studio condotto da Lionbridge: Elisabet Sas Olesa, Karolina Elizondo, Kathryn Nolte, Nathalie Azuaje, Melinda Johnson

INTRODUZIONE

Questo studio analizza in che modo l'intelligenza artificiale generativa (GenAI) può supportare il processo di convalida linguistica, con particolare attenzione al ruolo della tecnologia nel migliorare l'efficienza della fase di riconciliazione delle traduzioni doppie, ovvero il passaggio in cui due o più traduzioni indipendenti vengono unite in una sola (Koller et al., 2012).

L'obiettivo principale dello studio è stato valutare la capacità della tecnologia di GenAI di individuare i casi di mancata conformità nelle riconciliazioni prodotte tramite metodi tradizionali.

EVOLUZIONE NELLA PROGETTAZIONE DI PROMPT

Il prompt iniziale era progettato per operare in combinazione con il risultato della riconciliazione tradizionale, confrontando Traduzione A e Traduzione B e generando quindi un esito binario Test superato/Test non superato per la decisione finale del responsabile della riconciliazione.

Tuttavia, con questo approccio sono emersi importanti limiti, in particolare nelle lingue diverse dall'inglese e in contesti altamente specializzati, come quelli legati a specifiche aree terapeutiche o condizioni cliniche comunemente esaminate con gli strumenti per la valutazione della qualità della vita. In risposta a quanto emerso, abbiamo applicato l'intelligenza artificiale generativa (GenAI) per valutare il risultato della riconciliazione, a supporto del nostro flusso di lavoro per il controllo della qualità.

RISULTATI

Risultati positivi

L'analisi degli output dell'intelligenza artificiale generativa, svolta da linguisti esperti, ha evidenziato tre aree principali di supporto:

- **Giustificazioni non valide o incomplete:** la GenAI ha rilevato costantemente commenti privi di giustificazione linguistica o concettuale adeguata (ad esempio "La Traduzione A è migliore" o "La Traduzione B è preferibile"). Ha inoltre rilevato le query non risolte o le domande senza risposta. Il consiglio è quindi quello di ottimizzare il prompt, per assicurare una corretta escalation per i casi di questo tipo.
- **Assenza di giustificazioni:** tutti gli esperti hanno osservato che la GenAI ha identificato in modo affidabile i casi in cui le giustificazioni erano assenti, incomplete o poco chiare, accelerando sensibilmente la revisione di controllo della qualità.
- **Efficienza:** l'elaborazione tramite GenAI è molto rapida e vengono analizzati fino a 300 segmenti di testo in pochi secondi.

Questa analisi preliminare consente ai linguisti esperti di filtrare rapidamente i risultati e decidere se un file è pronto per il passaggio successivo del processo di convalida linguistica.

METODI

Abbiamo condotto un'analisi pratica usando un campione Perfo (1.000-2.000 parole) che includeva una combinazione di lingue dell'area Asia Pacifico, lingue a coda lunga e varianti regionali. In base a questi parametri, abbiamo selezionato diversi file da un lavoro di localizzazione precedentemente completato dei test Clinical Evaluation of Language Fundamentals®, quinta edizione (CELF-5), un sistema flessibile di test individuali comunemente usato in ambito clinico per supportare la diagnosi accurata dei disturbi del linguaggio in bambini e adolescenti (NCS Pearson, Inc., 2013). Le lingue di destinazione selezionate per l'analisi sono state spagnolo (Argentina), spagnolo (Spagna), francese (Francia), armeno (Armenia), giapponese (Giappone) e cinese tradizionale (Taiwan).

Il prompt aggiornato è stato quindi integrato in *Aurora Clinical Outcomes*, la piattaforma proprietaria Lionbridge per la convalida linguistica end-to-end. L'uso della GenAI a supporto del processo decisionale interno fornisce anche visibilità sul ragionamento del responsabile della riconciliazione. Assicura che l'attività sia stata svolta nel rispetto degli standard necessari e secondo i requisiti del settore. Tali requisiti si riflettono negli OBIETTIVI DELLA RICONCILIAZIONE e sono indicati nelle istruzioni per l'attività fornite al responsabile della riconciliazione:

1

Equivalenza concettuale alla misura originale

2

Adattamento culturale

3

Accessibilità per i destinatari previsti per lo studio

4

Rilevamento di bias ricorrenti

Aree di miglioramento

Nonostante il suo potenziale, la natura intrinsecamente imprevedibile della GenAI ha introdotto anche alcuni limiti durante l'esecuzione dei prompt:

- **Comportamento non coerente:** i risultati della GenAI possono non essere coerenti quando vengono valutati commenti identici in segmenti diversi, anche nello stesso file. Ciò rende meno affidabile l'output della GenAI quando questa tecnologia viene usata come soluzione autonoma.
- **Complessità legate a riferimenti contestuali o meta riferimenti:** l'uso frequente di commenti e riferimenti vaghi, come "Vedere il commento sopra" è problematico, poiché la GenAI non è in grado di individuare il riferimento. Gli esperti raccomandano l'uso di riferimenti più concisi ed espliciti e la ripetizione delle spiegazioni nei vari segmenti da parte del responsabile della riconciliazione.
- **Mancanza di standardizzazione nei commenti per la riconciliazione:** i nostri esperti concordano sulla necessità di una definizione più precisa ed esplicita e dell'allineamento di tutte le parti in merito a quali siano le caratteristiche di un ragionamento accettabile e di una giustificazione dettagliata.

CONCLUSIONE

Questo studio dimostra il notevole potenziale della GenAI come strumento di supporto per il controllo qualità, per identificare lacune, mancanza di conformità e giustificazioni mancanti.

La GenAI consente di individuare più rapidamente i casi in cui è necessario ripetere il lavoro e permette al team di localizzazione di fornire feedback più mirati agli stakeholder, offrendo loro indicazioni utili per migliorare le prestazioni future. Tuttavia, l'esperienza umana resta fondamentale per interpretare i risultati, convalidare le decisioni tenendo conto delle sfumature e assicurare l'allineamento ai requisiti di progetto e normativi.

Un approccio comprensivo di intervento umano, che combina l'analisi effettuata tramite GenAI con la revisione a cura di esperti, migliora la qualità complessiva, ottimizza le prestazioni del responsabile della riconciliazione attraverso feedback più qualificati e supporta un'ottimizzazione continua del processo.

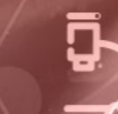
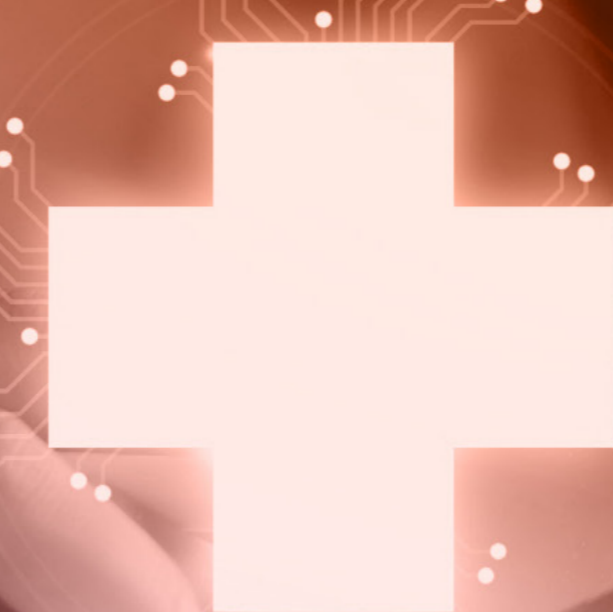
In definitiva, con l'ottimizzazione continua dei prompt e un addestramento mirato, questo approccio può offrire risultati di qualità superiore e una maggiore efficienza dei costi nei progetti di localizzazione di COA.

Citazioni:

Koller, M. et al. (2012). The process of reconciliation: evaluation of guidelines for translating quality-of-life questionnaires. Expert Review of Pharmacoeconomics and Outcomes Research 12(2), 189–197.
CELF-5 - Clinical Evaluation of Language Fundamentals | Quinta edizione. Copyright © 2013 NCS Pearson, Inc. Tutti i diritti riservati. Web: CELF-5 - Clinical Evaluation of Language Fundamentals | Quinta edizione | Pearson Assessments US

GenAI in azione:

studio sull'efficacia dell'output della revisione comparativa tramite AI



LIONBRIDGE

Pearson

Studio condotto da Lionbridge: Stephanie Casale

OBIETTIVI

La convalida linguistica è il processo attraverso il quale le valutazioni dei risultati clinici vengono localizzate e sottoposte a revisione per garantire una raccolta dati accurata e coerente nei mercati di riferimento. Si tratta di un processo lungo e complesso. Questo approccio assicura traduzioni di altissima qualità ed estremamente precise, ma tutto questo ha un costo. Per ridurre tempi e costi, il nostro studio si propone di abbreviare i tempi di consegna e i costi di outsourcing, mantenendo al contempo gli standard elevati richiesti dal processo.

Questo poster analizza la fattibilità dell'uso dell'intelligenza artificiale generativa per eseguire la revisione comparativa, un passaggio chiave del controllo qualità nel processo di convalida linguistica. Questa attività consiste nell'eseguire un confronto tra il testo di origine e quello ottenuto tramite traduzione inversa, per stabilire l'equivalenza concettuale. Poiché si tratta di un passaggio intermedio, i passaggi precedente e successivo vengono eseguiti da linguisti esperti, il che rende la revisione comparativa una candidata ideale per l'automazione. Questo approccio riduce al minimo il rischio che eventuali errori passino inosservati prima della finalizzazione.

L'obiettivo della ricerca era sviluppare un prompt che garantisse, almeno, la qualità attuale fornita dai revisori umani per la revisione comparativa.

METODI

Per prima cosa, abbiamo sviluppato un prompt in grado di produrre il risultato previsto di una revisione comparativa e un commento per la revisione, che fornisce ulteriori dettagli relativi al risultato. I risultati della revisione comparativa vengono suddivisi in tre categorie:



Segmenti identici - Questo risultato indica che il testo di origine e la traduzione inversa coincidono esattamente sotto ogni aspetto, incluse maiuscole, minuscole e punteggiatura.



Segmenti equivalenti - Questo risultato indica che, pur essendoci differenze nella scelta delle parole, nella struttura della frase o in altri dettagli, il significato dei segmenti rimane concettualmente equivalente. Il lettore ricaverebbe da entrambi i testi le stesse informazioni.



Necessità di revisione - Questo risultato indica che qualcosa nei due segmenti li rende concettualmente non equivalenti. Il lettore potrebbe fraintendere il testo tradotto, ricavandone un significato diverso da quello originale.

Successivamente, il prompt è stato configurato per generare un commento per la revisione comparativa, per qualsiasi risultato non identico. Questi commenti spiegano le differenze concettuali tra i due segmenti, includendo chiarimenti su possibili fraintendimenti da parte di un lettore non esperto. Al prompt è stato chiesto di ignorare le differenze di punteggiatura e uso delle maiuscole e minuscole, a meno che non influenzassero direttamente il significato e la comprensione. Il prompt doveva inoltre ignorare eventuale testo aggiuntivo non rilevante per il significato del testo di origine, ad esempio i tag di formattazione.

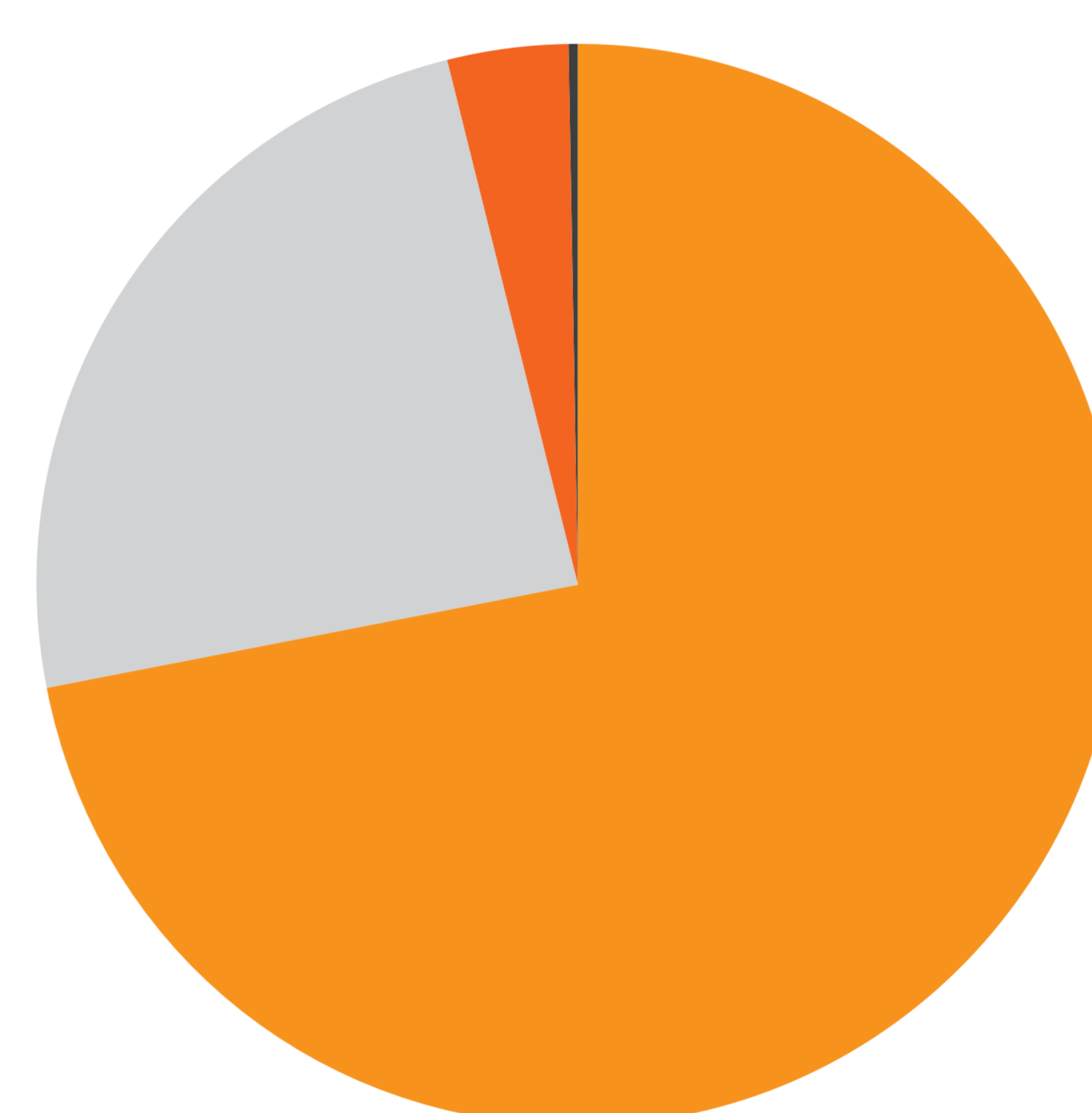
Per la raccolta dei dati, abbiamo usato una valutazione di Pearson localizzata in precedenza, il Delis-Kaplan Executive Function System (D-KEFS) - Trail Making Test, contenente circa 1.000 parole di origine. Le traduzioni inverse sono state eseguite da diversi linguisti madrelingua, in inglese o nella lingua di destinazione. Sono state condotte diverse revisioni comparative, sia da un revisore professionista che dai membri del nostro team di localizzazione di COA. I partecipanti avevano diversi livelli di esperienza nel processo di revisione comparativa ed erano parlanti nativi di una varietà di lingue, così da offrire un'ampia gamma di output da confrontare.

RISULTATI

I risultati iniziali sono incoraggianti, con descrizioni chiare e concise delle discrepanze tra la valutazione originale e la traduzione inversa e un tasso di accuratezza preliminare complessivo pari al 96,4%.

La ripartizione è mostrata nel grafico seguente. Include il 72,09% di corrispondenze esatte tra l'output umano e quello dell'AI, oltre a un ulteriore 24,3% di segmenti segnalati come discrepanze dall'AI, ma non dal revisore umano incaricato dell'analisi comparativa. I risultati hanno inoltre evidenziato il 3,5% di discrepanze rilevate dal professionista umano che l'AI aveva classificato come equivalenze. Infine, è stato registrato anche uno 0,17% di risposte identiche, considerate tuttavia rischiose dai ricercatori poiché la traduzione non era in caratteri latini e poteva quindi essere più facile non accorgersi delle discrepanze.

PERCENTUALE DI CORRISPONDENZA TRA OUTPUT DELL'AI E UMANO



Corrispondenza esatta 72%

Solo AI 24,3%

Solo revisione umana 3,5%

Rischio intrinseco per l'AI 0,2%

Percentuali rilevanti:

I risultati dell'intelligenza artificiale generativa analizzati dai nostri linguisti esperti hanno evidenziato tre aree principali in cui la GenAI può offrire un supporto efficace:

Rischio intrinseco per l'AI - 0,17%: questa percentuale include segmenti che potrebbero essere stati segnalati da un essere umano a causa dell'uso di un alfabeto non latino nella traduzione ma che non sarebbero stati rilevati dal prompt di AI.

Risposte non coerenti - 1,26%: durante la revisione, l'AI ha occasionalmente prodotto risposte diverse per lo stesso insieme di segmenti, ma questi casi rappresentano poco più dell'1% del totale dei dati.

CONCLUSIONE

L'AI generativa può aiutare a ridurre in modo significativo tempi e costi del processo di convalida linguistica.

Sono necessari ulteriori studi per ampliare il set di dati. I prossimi passaggi includeranno l'aggiunta di un modello di verifica per valutare gli effetti dell'uso di questo output durante la fase di riconciliazione della revisione comparativa. Un'ulteriore ottimizzazione del prompt potrebbe contribuire a ridurre alcuni rischi e incoerenze.

