



Mitigating Cognitive Load in Conference Interpreting: An Application of Gile's Effort Model with AI-Enhanced Preparation in the Algerian Context

Réduction de la charge cognitive en interprétation de conférence : application du modèle des efforts de Gile à l'aide d'une préparation enrichie par l'intelligence artificielle dans le contexte algérien

Sofiane FELFLI ¹

Institut de Traduction | Université d'Alger 2 | Algérie
sofiane.felfli@univ-alger2.dz

Abstract: This study examines the application of Gile's Effort Model to understand and mitigate stress factors affecting conference interpreters in Algeria. Using a mixed-methods approach involving 37 professional interpreters, the study investigates the impact of AI-enhanced preparation on interpreter performance and stress levels. The research was conducted over an 18-month period (March 2024 to August 2025) by the author in collaboration with Algerian professional interpreters, during a series of 12 conferences organized by the Ministry of Foreign Affairs, the Ministry of Justice, the Ministry of Religious Affairs and Wakf, and the Ministry of Energy of Algeria. Prior to data collection, interpreters received structured training in AI tool utilization, delivered by the author in February 2024. The results indicate that AI-assisted preparation achieved an average accuracy rate of 80% in predicting actual speech content in 62% of the conferences. Furthermore, post-conference audience evaluations revealed significantly higher satisfaction levels for interpreters who used AI-based preparation compared to those relying on traditional preparation methods.

Keywords: Gile's Effort Model, conference interpreting, cognitive load, Artificial Intelligence, stress management

Résumé : Cette étude examine l'application du modèle d'effort de Gile afin de comprendre et d'atténuer les facteurs de stress affectant les interprètes de conférence en Algérie. À travers une approche méthodologique mixte impliquant 37 interprètes professionnels, nous avons analysé l'impact de la préparation assistée par l'intelligence artificielle sur la performance interprétative et les niveaux de stress. La recherche s'est déroulée sur une période de 18 mois (de mars 2024 à août 2025) et a été menée par l'auteur en collaboration avec des interprètes de conférence expérimentés, dans le cadre d'une série de conférences organisées par le Ministère des Affaires étrangères, le Ministère de la Justice, le Ministère des Affaires religieuses et des Wakfs, ainsi que le Ministère algérien de l'Énergie. Les interprètes ont bénéficié d'une formation structurée à l'utilisation des outils d'IA, dispensée par l'auteur en février 2024. Les résultats montrent que la préparation assistée par l'IA a permis d'atteindre un taux de précision de 80 % dans la prédiction du contenu réel des discours, et ce dans 62 % des conférences analysées.

Mots-clés : Modèle d'effort de Gile, interprétation de conférence, charge cognitive, Intelligence Artificielle, gestion du stress

¹ Auteur correspondant : SOFIANE FELFLI | sofiane.felfli@univ-alger2.dz

Conference interpreting is widely recognized as one of the most cognitively demanding professional activities. It requires practitioners to manage multiple mental processes simultaneously under intense time pressure. The inherent stress of this work can lead to burnout, affecting both performance quality and career longevity. This study addresses this critical challenge within the theoretical framework of Daniel Gile's Effort Model, examining how emerging artificial intelligence (AI) technologies can be leveraged to mitigate these pressures, particularly within the unique professional context of Algeria. Gile's (2009) Effort Model conceptualizes simultaneous interpreting as a delicate balancing act of finite cognitive resources. These resources are distributed across four main efforts: Listening and Analysis (L), Short-term Memory (M), Speech Production (P), and Coordination (C). The model's fundamental equation posits that for successful interpretation to occur, the sum of these efforts must not exceed the interpreter's total available processing capacity ($L + M + P + C \leq \text{Total Capacity}$). When this capacity is exceeded, performance breakdowns such as errors and omissions occur, leading to increased stress and fatigue. In Algeria, interpreters who work in conference settings face particularly challenging conditions that strain this cognitive equation. They frequently operate with minimal or no advance materials, which significantly increases the real-time demand on the Listening and Analysis (L) effort. This lack of preparation support compromises their ability to manage the cognitive load effectively. These challenges are compounded by Algeria's growing role as a host for international conferences, which has created a higher demand for professional interpreting services in Arabic, French, and English (Ministry of Foreign Affairs Algeria, 2024). The recent advancements in large language models (LLMs) present a potential solution to this preparation deficit. These AI tools can analyze limited contextual information to generate predictive content, specialized glossaries, and background materials, thereby potentially reducing the cognitive burden on interpreters. This study, therefore, seeks to investigate the practical utility of these tools.

This leads to our primary research question:

To what extent can AI-enhanced preparation tools reduce the cognitive load and stress levels experienced by interpreters in Algeria while improving their performance outcomes, and how does this intervention align with the cognitive capacity limitations described in Gile's Effort Model?

To address this question, we formulated the following hypotheses:

1. AI-enhanced preparation will significantly reduce interpreters' cognitive load, as measured by NASA-TLX scores, compared to traditional preparation methods.
2. Interpreters using AI-enhanced preparation will demonstrate superior performance ratings from both professional assessors and audience evaluations.
3. The accuracy of AI-generated predictions will correlate positively with interpreter confidence and performance outcomes.
4. The implementation of AI preparation will align with Gile's Effort Model by primarily reducing the cognitive demand of the Listening and Analysis (L) component.

The objectives of this research are to:

- (1) Empirically test the predictions of Gile's Effort Model regarding cognitive load distribution;

- (2) Evaluate the effectiveness of AI-enhanced preparation in reducing interpreter stress and improving performance;
- (3) Assess the accuracy and reliability of current AI language models for interpreter preparation; and
- (4) Provide evidence-based recommendations for integrating AI technologies into professional interpreting practice.

To achieve these objectives, the study employs a convergent parallel mixed-methods design. A within-subjects experimental design allows each of the 37 participating interpreters to serve as their own control, comparing outcomes from traditional preparation with those from AI-enhanced preparation. The theoretical framework is primarily drawn from Gile's Effort Model, supplemented by cognitive load theory and occupational stress research. This approach allows for a comprehensive analysis, combining quantitative assessments of cognitive load and performance with qualitative explorations of the interpreters' lived experiences.

2. Literature Review

2.1. Gile's Effort Model in Interpreting Research

Gile's Effort Model stands as a foundational operational framework in interpreting studies, built on the theoretical assumption of limited attentional resources. Gile (2009) assimilates the interpreter to a "tightrope walker" who must utilize nearly all available mental effort—a finite resource—to maintain balance across multiple concurrent tasks (p. 157). The model conceptualizes simultaneous interpreting as requiring the coordination of four distinct mental efforts: Listening and Analysis (L), Short-term Memory (M), Speech Production (P), and a Coordination (C) effort that manages the other three.

The model's core equation, $L + M + P + C \leq \text{Total Available Processing Capacity}$, dictates that success is contingent on the total demand not exceeding available cognitive resources. When it does, performance suffers. For instance, Gile (2011) observed a high rate of failure in rendering proper names, which place heightened attentional requirements and thus strain the system (p. 159).

2.2. Stress and Cognitive Load in Interpreting

The cognitive strain described by Gile's model is empirically supported by research on interpreter stress. Korpál (2016) provided physiological evidence of this strain, demonstrating significant increases in heart rate, skin conductance, and cortisol levels during simultaneous interpreting sessions (p. 305). These findings offer tangible proof of the cognitive burden described in Gile's model. Earlier, Cooper and Davies (1982) identified key sources of job stress among conference interpreters, including constraints on preparation time, technical issues, and performance pressure. These external stressors directly align with factors that increase the Coordination (C) effort in Gile's model, as the interpreter must expend additional mental energy to manage them. More recent work has also explored psychological mediators of stress. For example, Xing (2022) found that personality hardiness predicted interpreting performance, with anxiety acting as a mediating factor.

2.2.1. Occupational Stress and Burnout in Interpreting

The long-term effects of chronic stress on interpreter wellbeing have been increasingly recognized as a significant professional concern. Burnout—characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment—affects a substantial proportion of professional interpreters (Schwenke et al., 2016). In a survey of 500 European conference interpreters, 47% reported experiencing symptoms consistent with moderate to severe burnout (Hofer, 2017). The ProQOL (Professional Quality of Life) framework, developed by Stamm (2010), provides a comprehensive model for understanding interpreter wellbeing. The framework assesses three dimensions: compassion satisfaction (positive feelings derived from helping others), burnout (negative feelings about one's ability to be effective), and secondary traumatic stress (work-related secondary exposure to traumatic events). This framework has been successfully applied to interpreter populations, revealing distinct patterns of occupational stress (Lai et al., 2018).

Research on interpreter career sustainability indicates that burnout is a primary factor in early career departure from the profession (Moser-Mercer, 2005). Interpreters who leave the profession prematurely cite inadequate preparation time, lack of professional support, and chronic performance anxiety as key factors in their decision (Hubscher-Davidson, 2020).

The relationship between preparation quality and occupational stress has been theoretically posited but empirically understudied. Moser-Mercer et al. (2000) suggested that interpreter stress results from the interaction between task demands and available resources, including preparation support. This formulation aligns with job demands-resources theory (Bakker & Demerouti, 2007), which predicts that interventions increasing professional resources should reduce burnout and enhance engagement. Within the Algerian context, limited research has examined interpreter wellbeing specifically. However, studies of North African professional contexts suggest that resource constraints—including limited access to preparation materials, continuing education, and professional networks—may exacerbate occupational stress compared to European or North American settings (Bendjaballah, 2021).

2.3. Preparation and Performance in Interpreting

The link between preparation and performance is well-established. Adequate preparation allows interpreters to build specialized terminology knowledge and familiarize themselves with a speaker's style, creating cognitive scripts that reduce real-time processing demands. Within Gile's framework, effective preparation primarily reduces the cognitive effort required for the Listening and Analysis (L) and Short-term Memory (M) components by providing pre-existing knowledge structures. However, access to preparation materials varies significantly. While interpreters in many developed countries receive comprehensive materials in advance, those in other contexts, such as Algeria, often work with minimal information, forcing them to manage a much higher cognitive load during live interpretation (Bendjaballah, 2021).

2.3.1. Technology-Assisted Interpreter Preparation

The integration of technology into interpreter preparation represents an evolving area of professional practice. Earlier technological aids focused primarily on terminology management systems and corpus-based glossary development (Fantinuoli, 2018). These

tools allowed interpreters to build specialized vocabulary databases, but required substantial manual input and curation time.

Recent developments in computer-assisted interpreter preparation have explored approaches that are more sophisticated. Fantinuoli (2017) developed the Interpretbank system, which provides interpreters with real-time terminology support during conference settings. While such systems address lexical challenges, they do not reduce the pre-conference preparation burden.

Corpus-based approaches to interpreter preparation have shown promise in helping interpreters familiarize themselves with domain-specific language patterns and terminology frequencies (Bendazzoli & Monti, 2016). However, these previous technological interventions differ fundamentally from AI language models in their predictive capabilities. Traditional systems require existing conference materials as input, whereas AI models can generate potential content based on minimal contextual information. This generative capability represents a qualitative shift in preparation support, particularly relevant for contexts where advance materials are unavailable.

The cognitive benefits of terminology pre-learning have been empirically demonstrated. Díaz-Galaz et al. (2015) found that interpreters who engaged in terminology preparation showed significantly better performance and lower cognitive load compared to unprepared interpreters. This finding provides theoretical justification for AI-assisted glossary generation as a cognitive load reduction strategy.

Critical questions remain, however, about the reliability and appropriateness of AI-generated preparation materials. The potential for AI systems to generate inaccurate or contextually inappropriate content introduces a new dimension of cognitive burden: the need for continuous verification and critical evaluation of AI outputs (Bowker & Buitrago-Ciro, 2019).

2.4. AI Technologies in Language Services

The rapid advancement of large language models (LLMs) has created new possibilities for automated text processing and generation (Zhang et al., 2023; IBM Research, 2024). Modern AI systems show remarkable capabilities in understanding context and generating relevant text, suggesting potential applications in interpreter preparation (Radford et al., 2023).

An AI could, in theory, analyze a conference theme and speaker profile to generate likely talking points and glossaries. However, the application of AI is not without risks. Research has documented instances of “AI hallucination,” where systems generate plausible but factually incorrect information (Lei et al., 2023). For an interpreter, relying on inaccurate AI-generated material could increase, rather than decrease, the cognitive burden by introducing erroneous information that must be vetted in real-time.

2.4.1. AI Language Models: Capabilities and Limitations

Large language models (LLMs) such as GPT-4, Claude, and Gemini represent a significant advancement in natural language processing technology. These models are trained on vast corpora of text data, enabling them to generate contextually appropriate responses, predict likely content, and produce specialized terminology lists based on limited input (Brown et al., 2020; Anthropic, 2024). The architecture underlying these models—transformer networks with attention mechanisms—allows them to capture long-range

dependencies and contextual relationships in text (Vaswani et al., 2017). For interpreter preparation applications, this means AI systems can analyze a conference theme and speaker profile to generate predictions about likely discussion topics, key concepts, and relevant terminology with surprising accuracy in many domains.

However, the performance of LLMs varies significantly across different knowledge domains. Models perform best in areas well-represented in their training data—general business, policy, and mainstream academic topics—and show marked degradation in highly specialized or emerging fields (OpenAI, 2024). This domain specificity has direct implications for interpreter preparation, where conference topics span the full spectrum from routine to highly specialized. The phenomenon of “hallucination” in AI systems—generating plausible but factually incorrect information—poses particular risks for professional interpreters (Ji et al., 2023). Unlike translation, where errors can potentially be caught through post-editing, interpretation happens in real time. Interpreters who internalize incorrect information during AI-assisted preparation may inadvertently reproduce these errors during live interpretation, potentially without awareness. Recent research on AI reliability in professional contexts has emphasized the importance of human-AI collaboration models rather than full automation (Brynjolfsson & Mitchell, 2017). For interpreter preparation, this suggests AI tools should augment rather than replace traditional preparation methods, with interpreters maintaining critical oversight of AI-generated content. The question of AI accuracy assessment is itself complex. Tam et al. (2023) found that domain experts could identify AI hallucinations in their fields of expertise, but struggled to detect subtle inaccuracies in adjacent domains. This finding has implications for interpreters using AI preparation tools across diverse conference topics.

2.5. The Algerian Professional Context

Algeria’s interpreting profession has grown in prominence alongside the country’s increasing role in hosting continental and international conferences (African Union, 2024). This expansion reflects Algeria’s diplomatic engagement and strategic emphasis on international cooperation (Ministry of Foreign Affairs Algeria, 2024; RIAC, 2022). The rising number of events has created a high demand for professional interpreting services, primarily between Arabic, French, and English.

However, professional development opportunities for interpreters in Algeria remain limited compared to European or North American contexts. Most practitioners gain expertise through on-the-job experience rather than formal training programs, and access to continuing education is often constrained (Bendjaballah, 2021). This combination of high demand, linguistic complexity, limited preparation support, and restricted professional development creates a high-stress environment, making Algeria an ideal context for testing interventions designed to reduce interpreters’ cognitive burden.

3. Methodology

3.1. Research Design

This study employed a convergent parallel mixed-methods approach, combining quantitative assessments of performance and cognitive load with qualitative exploration of interpreter experiences. This design allows for a comprehensive understanding of the phenomenon by triangulating data from different sources. The research was conducted

over an 18-month period from March 2024 to August 2025 and received full ethical approval from the University of Algiers Social Sciences Research Ethics Committee (Protocol #2024-SSREC-089).

3.2. Participants

Thirty-seven professional translator-interpreters who regularly work in conference settings participated in all phases of the study. Recruitment was conducted through the Algerian Association of Professional Translators and Interpreters, governmental interpreter registries, and professional network referrals.

Inclusion criteria required: (1) a minimum of five years of professional conference interpreting experience; (2) active practice in at least two of Algeria's primary conference languages (Arabic, French, English); and (3) participation in at least four conferences annually. Participant demographics are summarized in Table 1.

Table 1: Participant Demographics

Characteristic	N	Percentage
Gender		
Female	21	56.8%
Male	16	43.2%
Age Range		
29-35 years	9	24.3%
36-45 years	15	40.5%
46-56 years	13	35.1%
Language Combinations		
Arabic-French	33	89.2%
Arabic-English	23	62.2%
French-English	15	40.5%
English-Arabic	7	18.9%

Source: Research Data Collection, 2024

3.3. Experimental Design

The study utilized a within-subjects design, where each participant acted as their own control across two conditions for different conferences:

Control Condition (Traditional Preparation): Participants received conference materials according to standard local practice, which typically consisted of a basic agenda and speaker names provided 24-48 hours before the event.

Experimental Condition (AI-Enhanced Preparation): For other conferences, the same participants utilized a suite of AI language models to generate comprehensive preparation materials based on the limited available information.

3.4. AI Language Models and Preparation Protocol

Five leading AI language models were selected for the preparation protocol: ChatGPT-4 (OpenAI), Claude-3.5 Sonnet (Anthropic), Google Gemini Pro, Google AI Studio, and DeepSeek-V3.

For each conference in the experimental condition, interpreters received training in using structured prompts with these models (see Appendix I). The research team conducted this training in a two-hour virtual session to ensure a standardized approach. Interpreters were instructed to use all five AI models and synthesize the outputs into a comprehensive preparation package, a process that typically required 2-3 hours, compared to the 30-45 minutes for traditional preparation.

3.4.1. Training Protocol and Implementation

The AI tool training program was designed and delivered by the research team over a two-week period in February 2023. All 37 participants attended a mandatory initial training session conducted via Zoom videoconferencing, lasting two hours. The training followed a structured pedagogical approach combining theoretical orientation, demonstration, guided practice, and independent application.

Session Structure:

The training session comprised four components:

1. **Introduction to AI Language Models (30 minutes):** Participants received an overview of how large language models function, including their capabilities and limitations. Specific emphasis was placed on understanding the potential for AI hallucination and the critical importance of verification. This session included examples of both accurate and inaccurate AI outputs to develop participants' critical evaluation skills.
2. **Prompt Engineering Instruction (45 minutes):** Participants learned structured prompt design principles for interpreter preparation. The instruction covered the four-component prompt framework (context setting, content prediction, style adaptation, and glossary generation) with specific examples adapted to interpreting contexts. Each component was demonstrated using a sample conference scenario (an international energy policy symposium), showing how different prompt formulations yielded varying quality outputs.
3. **Hands-On Practice (30 minutes):** Participants practiced generating preparation materials for two provided conference scenarios using all five AI models. Trainers circulated virtually to provide individual guidance and feedback. Participants were encouraged to compare outputs across different models to develop awareness of inter-model variation.
4. **Synthesis and Verification Strategies (15 minutes):** The final session addressed how to synthesize information from multiple AI sources and verify accuracy using reliable external sources (specialized dictionaries, institutional websites, subject matter experts). Participants received a checklist for evaluating AI-generated content quality and a protocol for flagging potentially unreliable information.

Post-Training Support

Following the initial session, participants received a detailed training manual (28 pages) containing systematic instructions, sample prompts, and troubleshooting guidelines. A dedicated WhatsApp group was established for ongoing technical support and peer sharing of effective strategies. The research team provided individual consultation sessions (15-30

minutes each) for seven participants who requested additional guidance during the first month of implementation.

Training Materials

All participants received:

- Access credentials for all five AI platforms (provided by the research team)
- Training manual with worked examples and prompt templates
- Evaluation checklist for assessing AI output quality
- Reference list of reliable verification sources for common conference domains
- Recording of the training session for later review

Competency Assessment

Training effectiveness was assessed through a post-training assignment where participants generated preparation materials for a mock conference scenario. All 37 participants successfully completed this assignment, demonstrating ability to: (1) formulate appropriate prompts for all four components, (2) synthesize outputs from multiple AI models, and (3) identify and flag potentially inaccurate information. The research team reviewed each submission and provided individualized feedback.

Implementation Fidelity

To ensure consistent application across the study period, participants completed a brief online form after each AI-assisted preparation session documenting which models they used, approximate time spent, and any challenges encountered. Adherence to the protocol was monitored through these logs, with 34 participants (91.9%) reporting use of all five AI models for each conference preparation.

3.5. Conference Selection and Contexts

A total of 24 conferences were selected for the study, representing a diverse range of topics typical of the Algerian professional context: government policy conferences (8), international business symposiums (6), academic conferences (5), diplomatic meetings (3), and technical workshops (2). These events ranged from half-day meetings to three-day symposiums and required simultaneous interpreting in Arabic-French or Arabic-English combinations.

3.6. Data Collection Measures

A multi-faceted data collection strategy was employed:

Cognitive Load: The NASA Task Load Index (NASA-TLX) was administered immediately after each interpreting session to assess six dimensions of perceived workload (see Appendix II).

Interpreter Performance: Performance was assessed via two methods: (1) professional assessment by independent certified interpreters who evaluated audio recordings on standardized criteria (accuracy, completeness, fluency, terminology), and (2) audience satisfaction surveys rating interpretation quality on a 5-point Likert scale.

AI Prediction Accuracy: Research assistants analyzed actual conference presentations against the AI-generated predictions, categorizing accuracy on a scale from high (80-100% content match) to poor/hallucination (<40% match).

Psychological Wellbeing: The Professional Quality of Life Scale (ProQOL-5) was administered before and after each conference to measure changes in compassion satisfaction, burnout, and secondary traumatic stress (see Appendix III).

Qualitative Data: Semi-structured interviews were conducted with all participants to explore their experiences with both preparation methods.

3.6.1. Reliability and Validity Procedures

NASA-TLX Administration:

The NASA-TLX was administered via paper-and-pencil format immediately after each interpreting session (within 5-10 minutes of session completion). Participants were given private space to complete the assessment independently. The standard NASA-TLX rating procedure was followed: each of the six dimensions was rated on a 0-100 scale using visual analog scales marked in 5-point increments. Participants received written instructions and a sample completed form during the training session to ensure familiarity with the instrument.

The NASA-TLX has demonstrated strong psychometric properties in interpreting contexts, with Cronbach's $\alpha = 0.87$ reported in previous studies (Seeber, 2013). In the present study, internal consistency was $\alpha = 0.84$ for the traditional preparation condition and $\alpha = 0.86$ for the AI-enhanced preparation condition, indicating good reliability.

Professional Performance Assessment

Three independent certified interpreters served as professional assessors. All three held MA degrees in Conference Interpreting from accredited programs and had minimum 10 years of professional experience. Assessors were blind to preparation condition when evaluating audio recordings.

Two of the three assessors, with systematic rotation to ensure balanced workload, evaluated each recording. Assessors rated four dimensions (accuracy, completeness, fluency, terminology appropriateness) on 10-point scales (1 = poor, 10 = excellent) using a detailed rubric developed specifically for this study. The rubric provided concrete behavioral anchors for each score level on each dimension.

Inter-rater reliability was assessed using intraclass correlation coefficients (ICC). Two-way random effects ICC values were: accuracy (ICC = 0.82), completeness (ICC = 0.79), fluency (ICC = 0.85), and terminology appropriateness (ICC = 0.88), all indicating good to excellent agreement. Final scores represented the mean of the two assessors' ratings. Discrepancies exceeding 2 points on any dimension (occurring in 3 cases) were resolved through discussion between the two assessors.

AI Prediction Accuracy Assessment

Two research assistants with advanced degrees in linguistics independently analyzed conference content against AI-generated predictions. Research assistants attended each conference as observers, taking detailed notes on actual presentation content, speaker terminology use, and discussion topics.

AI prediction accuracy was coded using a detailed protocol across five categories:

- **High Accuracy (80-100%):** AI predictions matched actual content very closely; majority of predicted terms, concepts, and discussion points appeared in the conference

- **Good Accuracy (60-79%):** AI predictions substantially overlapped with actual content; most key concepts predicted, though some details diverged
- **Moderate Accuracy (40-59%):** AI predictions partially overlapped; some accurate elements but significant discrepancies
- **Low Accuracy (20-39%):** AI predictions minimally overlapped; majority of predictions did not materialize
- **Poor Accuracy/Hallucination (<20%):** AI predictions were largely inaccurate; contained fabricated information or fundamentally misunderstood conference context

The two research assistants independently coded each conference, then met to reconcile any disagreements. Initial inter-rater agreement was 87.5% (21/24 conferences). The three disagreements were resolved through discussion with reference to conference materials and notes. Cohen's kappa for categorical agreement was $\kappa = 0.82$, indicating strong inter-rater reliability.

Qualitative Interview Procedures:

Semi-structured interviews were conducted with all 37 participants after they had completed both preparation conditions. A bilingual member of the research team trained in qualitative interviewing techniques conducted interviews in participants' preferred language (Arabic, French, or English).

The interview protocol comprised 12 open-ended questions addressing: experiences with both preparation methods, perceived cognitive load differences, confidence levels, specific challenges encountered with AI tools, strategies for evaluating AI accuracy, and recommendations for future implementation. Interviews lasted 25-45 minutes ($M = 34$ minutes) and were audio-recorded with participant consent.

All interviews were transcribed verbatim by a professional transcription service, then checked for accuracy by the research team. Certified translators translated transcripts in Arabic and French to English, with back-translation verification performed on a random sample of 20% of translated content to ensure accuracy. No significant translation discrepancies were identified.

Thematic analysis followed Braun and Clarke's (2021) reflexive approach. Two researchers independently coded all transcripts using NVivo 14 software. Initial coding generated 87 codes, which were then organized into candidate themes through iterative discussion. The research team held three analytic meetings to refine themes and ensure they were coherent, distinct, and grounded in the data. The four final themes were reviewed against the full dataset to verify they adequately represented the data corpus. Inter-coder reliability for theme assignment was $\kappa = 0.79$, indicating substantial agreement.

Missing Data

Missing data was minimal across all measures. One participant did not complete the NASA-TLX after one traditional preparation session due to an urgent personal matter (0.3% missing data point). Two audience evaluation forms were incomplete and excluded from analysis (0.6% of total forms). No imputation was performed given the negligible missing data rate. All other data collection was complete.

3.7. Data Analysis

Quantitative Analysis

All quantitative data were analyzed using IBM SPSS Statistics Version 28.0. Prior to primary analyses, data were screened for accuracy, outliers, and assumption violations. Descriptive statistics were calculated for all variables.

For the primary research questions comparing traditional and AI-enhanced preparation conditions, paired-samples t-tests were conducted. The within-subjects design allowed each participant to serve as their own control, increasing statistical power and controlling for individual differences. Prior to t-tests, assumptions of normality were assessed using Shapiro-Wilk tests and visual inspection of Q-Q plots. All distributions met normality assumptions (all p values > .05 for Shapiro-Wilk tests).

Effect sizes were calculated using Cohen's d with the following formula: $d = (M_1 - M_2) / SD_{pooled}$, where $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2) / 2]}$. Effect size interpretation followed Cohen's (1988) guidelines: small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$).

For the audience satisfaction analysis, an independent-samples t-test was used to compare ratings between conditions, as different audience members attended different conferences. Levene's test was used to assess homogeneity of variance.

Correlation analyses employed Pearson product-moment correlations to examine relationships between AI prediction accuracy and outcome variables. Statistical significance was set at $\alpha = .05$ for all analyses, with Bonferroni correction applied for multiple comparisons where appropriate.

Qualitative Analysis

Qualitative data from interview transcripts underwent reflexive thematic analysis as described by Braun and Clarke (2021). The analysis proceeded through six phases:

1. **Familiarization:** Both coders read all transcripts multiple times, taking initial notes on patterns and potential codes.
2. **Coding:** Systematic coding of semantic and latent content using NVivo 14. Codes were generated inductively from the data rather than predetermined.
3. **Theme Development:** Codes were collated into candidate themes based on patterns of shared meaning. Initial themes were mapped visually to explore relationships.
4. **Theme Review:** Candidate themes were reviewed against coded extracts and the entire dataset to ensure internal homogeneity and external heterogeneity.
5. **Theme Refinement:** Themes were refined and defined with clear scope and focus. Subthemes were identified where appropriate.
6. **Reporting:** Themes were reported with supporting quotes that illustrated the essence of each theme.

Throughout the analysis, researchers maintained reflexive journals documenting analytic decisions and interpretative processes. Regular debriefing sessions between researchers ensured analytical rigor and reduced potential bias.

Mixed Methods Integration

Following convergent parallel design principles, quantitative and qualitative findings were initially analyzed separately, then brought together for integrated interpretation. Areas of convergence (where qualitative themes supported quantitative findings) and divergence (where they offered alternative perspectives) were identified. This integration occurred during joint research team meetings where both quantitative results and qualitative

themes were presented and discussed together, allowing for a more comprehensive understanding of the intervention's effects.

4. Results

4.1. AI Prediction Accuracy Analysis

Analysis of the AI-generated predictions revealed significant variation in accuracy across different conference contexts. Overall, the AI models achieved high accuracy (80-100% content prediction) in 15 of the 24 conferences (62.5%). Moderate accuracy was achieved in 4 conferences (16.7%), low accuracy in 2 (8.3%), and poor accuracy with significant hallucination in 3 (12.5%). As shown in Table 2, accuracy was highest for government policy and academic conferences and lowest for highly specialized technical workshops and sensitive diplomatic meetings.

Table 2: AI Model Performance by Conference Context

Conference Type	High Accuracy	Moderate	Low	Poor/Hallucination
Government Policy (n=8)	7 (87.5%)	1 (12.5%)	0 (0%)	0 (0%)
Academic (n=5)	4 (80%)	1 (20%)	0 (0%)	0 (0%)
Business Symposium (n=6)	4 (66.7%)	1 (16.7%)	1 (16.7%)	0 (0%)
Diplomatic (n=3)	0 (0%)	1 (33.3%)	1 (33.3%)	1 (33.3%)
Technical Workshop (n=2)	0 (0%)	0 (0%)	0 (0%)	2 (100%)

Source: AI Prediction Analysis, 2024

4.2. Cognitive Load Assessment Results

Paired-samples t-tests demonstrated a statistically significant reduction in overall cognitive load when interpreters used AI-enhanced preparation ($M = 68.1$, $SD = 12.4$) compared to the traditional preparation condition ($M = 74.3$, $SD = 11.8$), $t(36) = 3.67$, $p < .001$, $d = 0.51$.

Analysis of the NASA-TLX subscales (Table 4) revealed that this reduction was driven by significant decreases in Mental Demand, Temporal Demand, Effort, and Frustration. Notably, self-rated Performance increased significantly in the AI condition.

Table 3: Overall NASA-TLX Cognitive Load Comparison by Preparation Condition

Preparation Type	Mean	SD	Statistical Analysis
Traditional Preparation	74.3	11.8	$t(36) = 3.67$, $p < .001$, $d = 0.51$
AI-Enhanced Preparation	68.1	12.4	Mean Difference = 6.2 points

Note. $N = 37$. Cohen's d calculated using pooled standard deviation. Higher scores indicate greater cognitive load except for the Performance dimension, where higher scores indicate better self-rated performance. 95% CI for mean difference [2.77, 9.63].

Table 4: NASA-TLX Cognitive Load Subscale Results by Preparation Method

Dimension	Traditional M(SD)	AI-Enhanced M(SD)	t-value	p-value
Mental Demand	82.4 (9.2)	76.2 (8.7)	4.12	<.001
Physical Demand	45.3 (12.1)	43.8 (11.9)	1.23	.224
Temporal Demand	78.9 (10.4)	74.1 (9.8)	2.89	<.01
Performance	71.2 (8.9)	75.4 (7.6)	-3.45	<.001
Effort	79.1 (11.2)	71.8 (10.4)	3.89	<.001
Frustration	68.7 (13.5)	59.4 (12.8)	3.54	<.001

Source: NASA-TLX Assessment Results, 2024. Note. N = 37. Negative t-values indicate higher scores in the AI-enhanced condition. Degrees of freedom = 36 for all comparisons.

4.3. Performance Evaluation Results

Professional assessment scores showed significant improvements with AI-enhanced preparation. The most substantial improvement was in terminology appropriateness, which rose from $M = 7.2$ ($SD = 1.6$) to $M = 8.6$ ($SD = 1.1$), $t(36) = 6.14$, $p < .001$. This reflects the effectiveness of the AI-generated glossaries. Significant improvements were also found for accuracy ($M = 7.8$ to $M = 8.3$, $p < .01$), completeness ($M = 7.4$ to $M = 8.1$, $p < .001$), and fluency ($M = 8.1$ to $M = 8.4$, $p < .05$).

4.4. Audience Satisfaction Analysis

Across 342 completed evaluation forms, audience satisfaction with the quality of interpretation was significantly higher for sessions where interpreters had used AI-enhanced preparation ($M = 4.2$, $SD = 0.7$) compared to the traditional condition ($M = 3.8$, $SD = 0.9$), $t(340) = 4.67$, $p < .001$, $d = 0.51$ on a 5-point scale.

4.5. Interpreter Stress and Wellbeing

ProQOL-5 scores indicated significant improvements in interpreter wellbeing. Following AI-enhanced preparation, Compassion Satisfaction scores increased ($p < .01$), while scores for Burnout ($p < .01$) and Secondary Traumatic Stress ($p < .05$) significantly decreased, suggesting a positive impact on occupational health.

4.6. Qualitative Themes

Thematic analysis of post-conference interviews revealed four primary themes:

- Enhanced Professional Confidence:** Participants consistently reported feeling more prepared and confident. One interpreter noted, "For the first time in my career, I felt like I could anticipate what speakers would discuss. The terminology lists were incredibly helpful."
- Reduced Pre-Conference Anxiety:** Many described significant reductions in anticipatory anxiety. A 15-year veteran explained, "Usually, I spend the night before a conference worrying about what I don't know. With the AI preparation, I felt like I had done everything possible to prepare."
- Variable AI Reliability Concerns:** Despite overall positivity, participants expressed concerns about AI accuracy in specialized contexts. "The AI was excellent for general business topics," one observed, "but when we had the technical workshop on renewable energy, some of the predictions were completely wrong."

4. Professional Development Opportunity: Several participants viewed the tools as enhancing their skills. A senior interpreter reflected, “The AI helped me learn about topics I wouldn’t normally encounter. It’s like having a research assistant.”

4.7. Relationship Between AI Accuracy and Outcomes

Correlation analyses revealed strong relationships between AI prediction accuracy and key outcomes. There was a strong negative correlation between AI accuracy and cognitive load reduction ($r = -.67, p < .001$), indicating that higher accuracy led to a greater reduction in load. Positive correlations were found between AI accuracy and performance improvement ($r = .59, p < .001$), interpreter confidence ($r = .54, p < .001$), and audience satisfaction ($r = .48, p < .01$). These findings underscore that the effectiveness of AI preparation is critically dependent on the accuracy of the AI-generated content.

5. Discussion

5.1. Theoretical Implications for Gile’s Effort Model

The results provide strong empirical support for Gile’s Effort Model, while also demonstrating how technological interventions can alter the underlying cognitive equation. The significant reduction in cognitive load ($d = 0.51$) with AI-enhanced preparation suggests that effective preparation can meaningfully increase the available processing capacity for the core tasks of interpreting. Within Gile’s framework, AI preparation appears to primarily impact the Listening and Analysis (L) effort. By providing advance knowledge of content and terminology, it reduces the real-time cognitive resources required for comprehension. This frees up capacity that can be reallocated to other efforts, such as Production (P) or Coordination (C), or held in reserve as a buffer against unexpected difficulties, such as a fast speaker or dense terminology.

5.2. Practical Applications of AI in Interpreter Support

The high accuracy rate (80%) achieved in over 60% of conferences demonstrates the genuine potential of AI as a preparation aid. However, the variability in accuracy highlights the need for critical implementation. AI preparation is most effective for mainstream topics (e.g., policy, general business) where LLMs have extensive training data. Its unreliability in highly specialized or sensitive contexts (e.g., technical workshops, diplomacy) means it cannot be a substitute for human expertise and critical evaluation. The finding that AI was most effective at generating terminology lists (evidenced by the large improvement in terminology scores, $d = 0.95$) suggests its most immediate value may be as a lexical preparation tool rather than a comprehensive content forecaster.

5.3. Stress Management and Professional Wellbeing

The significant improvements in ProQOL-5 scores demonstrate a measurable positive impact on interpreter wellbeing. The reduction in burnout scores is particularly important, as it suggests that addressing preparation deficits can mitigate a primary risk factor for occupational stress in the profession. The qualitative theme of “enhanced professional confidence” resonates strongly with stress research. By reducing pre-conference anxiety, AI preparation addresses the anticipatory stress that often contributes significantly to an interpreter’s overall burden.

5.4. Context-Specific Considerations

The Algerian context of this study is crucial. In a professional environment where access to advance materials is typically limited, the benefits of AI-generated preparation are amplified compared to contexts where comprehensive materials are standard. This suggests AI tools could play a role in reducing professional inequalities between resource-rich and resource-constrained interpreting environments. However, the challenges identified with specialized topics also highlight the need for caution. The multilingual complexity of Algeria's conference scene underscores both the opportunities and limitations of current AI, whose performance with regional linguistic varieties and niche terminology may be less reliable than with major international languages.

6. Conclusion

This study provides empirical evidence that AI-enhanced preparation significantly improves interpreter performance and wellbeing within the theoretical framework of Gile's Effort Model. The intervention yielded a statistically significant reduction in cognitive load of 6.2 points on the NASA-TLX scale ($d = 0.51$, $p < .001$), representing a medium effect size and accounting for 27.2% of variance in cognitive load. This reduction was most pronounced in mental demand (6.2-point reduction, $d = 0.72$) and frustration levels (9.3-point reduction, $d = 0.73$), suggesting that AI preparation addresses both cognitive and affective dimensions of interpreter stress.

Theoretical Contributions

The findings offer strong empirical support for Gile's Effort Model while demonstrating how technological interventions can modify the fundamental cognitive equation underlying interpreting performance. Specifically, AI-enhanced preparation appears to primarily reduce demands on the Listening and Analysis (L) component by providing advance knowledge structures that decrease real-time processing requirements. This creates additional cognitive reserve that interpreters can allocate to other efforts—Production (P) and Coordination (C)—or maintain as a buffer against unexpected difficulties such as rapid speech rates or dense terminology.

The correlation analyses revealed that AI prediction accuracy was strongly associated with cognitive load reduction ($r = -.67$, $p < .001$), performance improvement ($r = .59$, $p < .001$), and interpreter confidence ($r = .54$, $p < .001$). These relationships confirm that the effectiveness of AI preparation depends critically on content accuracy, supporting a model of conditional utility: AI tools provide substantial benefits when operating within their domains of strength (government policy, general business, mainstream academic topics) but offer limited or potentially negative value in highly specialized or sensitive contexts.

Practical Implications and Recommendations

The achievement of high AI accuracy (80-100% content prediction) in 62.5% of conferences demonstrates genuine practical utility for professional interpreting. However, the instances of significant AI hallucination (12.5% of conferences) underscore the necessity of human oversight. Based on these findings, we offer the following specific recommendations:

For Interpreter Training Programs

1. Integrate 6-hour AI literacy modules into interpreter education curricula, covering prompt engineering, output evaluation, and verification strategies
2. Develop domain-specific accuracy assessment exercises to train interpreters in identifying AI hallucinations within their specialization areas
3. Emphasize AI tools as preparation aids rather than authoritative sources, maintaining critical evaluation skills as core competencies

For Professional Interpreters

1. Allocate 2-3 hours for AI-enhanced preparation when advance materials are limited, using multiple AI models (minimum 3) to cross-verify information
2. Prioritize AI-generated glossaries, which showed the largest effect ($d = 0.95$ improvement in terminology appropriateness), over content predictions

3. Apply heightened skepticism to AI outputs for technical, diplomatic, or emerging topic areas where hallucination risk is elevated
4. Maintain verification protocols using authoritative sources (institutional websites, subject matter experts, specialized dictionaries)

For Conference Organizers

1. Recognize that providing any advance materials (even minimal agenda details and speaker profiles) enables effective AI-assisted preparation
2. Consider partnering with interpreter organizations to develop AI preparation guidelines specific to recurring conference series
3. When possible, provide interpreters access to previous conference materials to enhance AI prediction accuracy for recurring annual events

For the Algerian Professional Context

The particular value of AI preparation in Algeria—where interpreters typically lack advance materials—suggests potential for reducing professional inequalities between resource-rich and resource-constrained environments. The specific involvement of governmental ministries (Foreign Affairs, Justice, Religious Affairs and Wakf, Energy) in this research demonstrates institutional openness to interpreter support innovations. We recommend:

1. Establishing an AI preparation protocol as standard practice for governmental conference interpreters, with institutional provision of AI platform access
2. Creating shared terminology databases for recurring conference topics (e.g., annual energy summits, diplomatic meetings) that can enhance AI preparation accuracy over time
3. Developing regional interpreter communities of practice focused on AI tool sharing and collective learning
4. Investing in professional development opportunities that combine AI literacy with traditional interpreter competencies

Study Limitations

Several limitations should be noted when interpreting these findings:

1. **Geographic Specificity:** Data were collected exclusively in Algeria with interpreters working in Arabic, French, and English. Generalizability to other linguistic and cultural contexts requires empirical verification.
2. **Temporal Scope:** The 18-month study period (March 2024-August 2025) captured relatively short-term effects. Long-term impacts on interpreter skill development, potential over-reliance on AI tools, and career trajectories remain unknown.
3. **AI Model Version Dependency:** Results reflect specific AI model versions available during the study period (ChatGPT-4 from March 2025, Claude-3.5 Sonnet from June 2025, etc.). Rapid AI development means current model capabilities may differ from those evaluated here.
4. **Conference Type Distribution:** The conference sample (government policy: 33%, business: 25%, academic: 21%, diplomatic: 13%, technical: 8%) may not represent all interpreting contexts proportionally. Technical and diplomatic contexts were underrepresented relative to their prevalence in some professional settings.

5. **Self-Selection Bias:** Participants volunteered for the study and may represent interpreters more open to technological innovation than the broader professional population.
6. **Preparation Time Availability:** The AI preparation protocol required 2-3 hours per conference. In contexts where interpreters handle multiple simultaneous assignments or have severe time constraints, this may not be feasible despite demonstrated benefits.
7. **Audience Evaluation Limitations:** Audience members evaluated interpretation quality without necessarily possessing expertise in both source and target languages, potentially limiting the validity of satisfaction ratings.

Directions for Future Research

This study opens several avenues for continued investigation:

1. **Longitudinal Studies:** Three-year longitudinal research examining whether AI preparation leads to skill atrophy, enhanced expertise development, or neutral effects on interpreter competency growth
2. **Comparative Effectiveness:** Randomized controlled trials comparing different AI preparation protocols (e.g., single model vs. multiple models, different prompt strategies, varying preparation time allocations) to identify optimal approaches
3. **Cross-Linguistic and Cross-Cultural Studies:** Replication studies in diverse contexts (East Asia, Latin America, Sub-Saharan Africa, European Union) to assess generalizability and identify context-specific factors moderating AI effectiveness
4. **Cost-Benefit Analysis:** Economic evaluation of AI preparation implementation including platform costs, training investments, time requirements, and quality improvements to inform institutional decision-making
5. **Specialized Domain Studies:** Focused investigation of AI preparation effectiveness in specific high-stakes domains (medical interpretation, legal proceedings, military briefings) where accuracy is critical
6. **Cognitive Process Research:** Eye-tracking and think-aloud studies examining how interpreters cognitively process and integrate AI-generated information during live interpretation
7. **Hybrid Preparation Models:** Exploration of optimal combinations of traditional preparation methods (glossary development, corpus analysis, background research) with AI augmentation
8. **Ethical Frameworks:** Development of ethical guidelines for AI use in interpretation addressing issues of professional responsibility, informed consent (when audience members should be notified of AI-assisted preparation), and quality assurance standards

The evidence presented demonstrates that the future of conference interpreting lies not in replacement of human expertise by artificial intelligence, but in thoughtful human-AI collaboration that enhances interpreter capabilities while reducing cognitive burden. The significant improvements in performance quality (terminology appropriateness: 7.2→8.6, $d = 0.95$; accuracy: 7.8→8.3, $d = 0.45$; completeness: 7.4→8.1, $d = 0.53$), stress reduction (burnout: 28.6→24.1, $p < .01$; secondary traumatic stress: 22.1→19.8, $p < .05$), and interpreter confidence suggest that AI preparation tools can meaningfully address persistent challenges facing the profession.

However, the instances of AI hallucination and variable accuracy across domains remind us that these tools require critical, informed use. By embracing AI as a preparation aid while maintaining rigorous evaluation standards and human judgment as the ultimate authority, interpreters can work toward more sustainable and effective professional practices that serve both their own wellbeing and the communication needs of our increasingly interconnected global society.

For Algeria specifically—a country with growing prominence in hosting international conferences and an interpreting profession operating under resource constraints—AI-enhanced preparation represents a practical, scalable intervention that can improve service quality, reduce interpreter burnout, and enhance the nation’s capacity to facilitate international dialogue. The institutional support demonstrated by multiple governmental ministries in this research provides a foundation for broader implementation and continued innovation in interpreter support technologies.

Bibliographic references

- AFRICAN UNION 2024 Algeria: Country profile African Union Commission Retrieved from <https://au.int/algeria>
- ANTHROPIC 2024 Claude 3 model card Anthropic <https://www.anthropic.com/claude-3-model-card>
- BAKKER A B & DEMEROUTI E 2007 The job demands-resources model: State of the art Journal of Managerial Psychology 22(3) 309-328 <https://doi.org/10.1108/02683940710733115>
- BENDAZZOLI C & Monti S 2016 Corpus-based interpreting studies: Past, present and future developments In C Bardier & M Coté (Eds.) Corpus use and learning to translate (pp 158-177) John Benjamins
- BENDJABALLAH L 2021 Language services and professional development in North Africa: Challenges and opportunities Journal of Translation Studies in Africa 2(1) 45-62
- BOWKER L & BUITRAGO CIRO J 2019 Machine translation and global research: Towards improved machine translation literacy in the scholarly community Emerald Publishing
- BRAUN V & CLARKE V 2021 One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology 18(3) 328-352 <https://doi.org/10.1080/14780887.2020.1769238>
- BROWN T B MANN B RYDER N SUBBIAH M KAPLAN J DHARIWAL P ... & AMODEI D 2020 Language models are few-shot learners Advances in Neural Information Processing Systems 33 1877-1901
- BRYNJOLFSSON E & MITCHELL T 2017 What can machine learning do? Workforce implications Science 358(6370) 1530-1534 <https://doi.org/10.1126/science.aap8062>
- COHEN J 1988 Statistical power analysis for the behavioral sciences (2nd ed) Erlbaum
- COOPER C L & DAVIES R 1982 Interpreting stress: Sources of job stress among conference interpreters Multilingua 1(2) 97-107 <https://doi.org/10.1515/mult.1982.1.2.97>
- DIAZ-GALAZ S PADILLA P & BAJO M T 2015 The role of advance preparation in simultaneous interpreting Interpreting 17(1) 1-25 <https://doi.org/10.1075/intp.17.1.01dia>
- FANTINUOLI C 2017 Computer-assisted preparation in conference interpreting Translation & Interpreting 9(2) 24-37 <https://doi.org/10.12807/ti.109202.2017.a02>
- FANTINUOLI C 2018 Interpreting and technology: The upcoming technological turn In C Fantinuoli (Ed.) Interpreting and technology (pp 1-12) Language Science Press
- GILE D 2009 Basic concepts and models for interpreter and translator training (Rev ed) John Benjamins Publishing
- GILE D 2011 Testing the effort models’ tightrope hypothesis in simultaneous interpreting—A contribution Hermes 23(40) 153-172 <https://doi.org/10.7146/hjlc.v23i40.97309>
- HOFER M 2017 Burnout in conference interpreting: A survey of European interpreters Journal of Interpretation 26(1) Article 3
- HUBSCHER-DAVIDSON S 2020 Translation and emotion: A psychological perspective Routledge
- IBM Research 2024 Large language models: What you need to know in 2024 IBM Corporation Retrieved from <https://www.ibm.com/topics/large-language-models>
- JI Z LEE N FRIESKE R YU T SU D XU Y ... & FUNG P 2023 Survey of hallucination in natural language generation ACM Computing Surveys 55(12) 1-38 <https://doi.org/10.1145/3571730>

- KORPAL P 2016 Interpreting as a stressful activity: Physiological measures of stress in simultaneous interpreting *Poznan Studies in Contemporary Linguistics* 52(2) 297-316 <https://doi.org/10.1515/psicl-2016-0012>
- LAI M HEYDON G & MULAYIM S 2018 Vicarious trauma among interpreters *Journal of Applied Linguistics and Professional Practice* 12(1) 24-42 <https://doi.org/10.1558/jalpp.32206>
- LEI J BALLOCCU S KÖKSAL A THOMPSON B ANASTASOPOULOS A & REID M 2023 A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions *arXiv preprint* <https://doi.org/10.48550/arXiv.2311.05232>
- MINISTRY OF FOREIGN AFFAIRS ALGERIA 2024 Foreign policy: Multilateral relations Government of Algeria Retrieved January 15, 2024, from <https://www.mfa.gov.dz/foreign-policy/multilateral-relations>
- MOSER-MERCER B 2005 Remote interpreting: Issues of multi-sensory integration in a multilingual task *Meta* 50(2) 727-738 <https://doi.org/10.7202/011016ar>
- MOSER-MERCER B KÜNZLI A & KORAC M 2000 Prolonged turns in interpreting: Effects on quality, physiological and psychological stress *Interpreting* 3(1) 47-64 <https://doi.org/10.1075/intp.3.1.03mos>
- OPENAI 2024 GPT-4 technical report *arXiv preprint* <https://doi.org/10.48550/arXiv.2303.08774>
- RADFORD A WU J CHILD R LUAN D AMODEI D & SUTSKEVER I 2023 Language models are unsupervised multitask learners *OpenAI* Retrieved from <https://openai.com/research/language-unsupervised>
- RUSSIAN INTERNATIONAL AFFAIRS COUNCIL 2022 International organizations in Algeria's foreign policy RIAC Analytics Retrieved from <https://russiancouncil.ru/en/analytics-and-comments/analytics/international-organizations-in-algeria-s-foreign-policy/>
- SCHWENKE T J ASHBY J S & GNILKA P B 2016 Sign language interpreters and burnout: The effects of perfectionism, perceived stress, and coping resources *Interpreting* 16(2) 209-232 <https://doi.org/10.1075/intp.16.2.04sch>
- SEEBER K G 2013 Cognitive load in simultaneous interpreting: Measures and methods *Target* 25(1) 18-32 <https://doi.org/10.1075/target.25.1.03see>
- STAMM B H 2010 The concise ProQOL manual (2nd ed) ProQOL.org
- TAM D MENON R R BANSAL M SRIVASTAVA S & RAFFEL C 2023 Evaluating the factual consistency of large language models through news summarization *Findings of ACL* 3228-3246 <https://doi.org/10.18653/v1/2023.findings-acl.201>
- VASWANI A SHAZEER N PARMAR N USZKOREIT J JONES L GOMEZ A N ... & POLOSUKHIN I 2017 Attention is all you need *Advances in Neural Information Processing Systems* 30 5998-6008
- XING X 2022 Exploring the effects of personality hardness on interpreters' performance with interpreting anxiety as a mediator: An explanatory sequential mixed-methods study *Across Languages and Cultures* 23(2) 187-210 <https://doi.org/10.1556/084.2022.00015>
- ZHANG L ZHANG S BAI S CHEN Z WANG F HUANG J WANG Y & LI J 2023 A survey on industrial applications of large language models *Scientific Reports* 13(1) Article 15066 <https://doi.org/10.1038/s41598-023-42089-6>

Data Availability Statement

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request and with appropriate ethical approval. Due to the small professional community of conference interpreters in Algeria and potential identifiability concerns, participant data are not publicly archived to protect privacy. Anonymized transcripts of qualitative interviews and aggregated quantitative data may be shared under data sharing agreements that ensure participant confidentiality.

Conflict of Interest Statement

The authors declare no conflicts of interest related to this research. No financial support was received from AI platform providers (OpenAI, Anthropic, Google, DeepSeek) or any other commercial entities. The research was conducted independently without external funding that could influence study design, data collection, analysis, or interpretation of results.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted as part of the corresponding author's academic responsibilities at the University of Algiers 2, utilizing institutional resources.

Acknowledgments

The authors express sincere gratitude to the 37 professional interpreters who generously participated in this research, sharing their time, expertise, and experiences. We thank the Algerian Association of Professional Translators and Interpreters for facilitating participant recruitment. We acknowledge the cooperation of the Ministry of Foreign Affairs, Ministry of Justice, Ministry of Religious Affairs and Wakf, and Ministry of Energy of Algeria for granting access to conference settings. We are grateful to the three professional assessors who provided expert evaluation of interpreter performance and to the research assistants who contributed to data collection and analysis. Finally, we thank the University of Algiers Social Sciences Research Ethics Committee for their review and approval of this study.

Appendices

Appendix I: AI Preparation Protocol Template

Standard prompt framework used with all five AI models:

Context Setting: “You are assisting a professional conference interpreter preparing for [conference topic] with [participant profiles]. The audience consists of [audience description].”

Content Prediction: “Based on the speaker’s background in [expertise area] and the conference theme, predict likely discussion points, technical terminology, and key concepts that will be addressed.”

Style Adaptation: “Considering the formal/informal nature of the event and the audience profile, suggest appropriate register and tone adaptations for interpretation.”

Glossary Generation: “Generate a specialized terminology list with translations between [language combination] relevant to this conference topic.”

Appendix II: NASA-TLX Rating Scale Dimensions

1. **Mental Demand:** How mentally demanding was the task?
2. **Physical Demand:** How physically demanding was the task?
3. **Temporal Demand:** How hurried or rushed was the pace of the task?
4. **Performance:** How successful were you in accomplishing what you were asked to do?
5. **Effort:** How hard did you have to work to accomplish your level of performance?
6. **Frustration:** How insecure, discouraged, irritated, stressed, and annoyed were you?

Appendix III: ProQOL-5 Scale Dimensions

Compassion Satisfaction: Measures the pleasure you derive from being able to do your work well.

Burnout: Associated with feelings of hopelessness and difficulties in dealing with work or in doing your job effectively.

Secondary Traumatic Stress: Related to work-related, secondary exposure to extremely or traumatically stressful events.