

The Reliability of Large Language Models (LLMs) in Official Translation: An Analysis of Gemini's Capabilities and Challenges in Translating Arabic Administrative Documents into English

موثوقية النماذج اللغوية الكبيرة (LLMs) في الترجمة الرسمية: تحليل لقدرات وتحديات نموذج Gemini

في ترجمة الوثائق الإدارية العربية إلى اللغة الإنجليزية

Adil HAOUES

Badji Mokhtar - Annaba University, Algeria, adil.haoues@univ-annaba.dz

Received: 15/03/2026

Accepted: 26/04/2026

Published: 31/05/2026

Abstract:

The integration of Generative Artificial Intelligence (GenAI), specifically Google's Gemini, is creating a paradigm shift in the translation of official administrative documents. This article investigates the reliability of Gemini in translating sensitive Arabic administrative extracts (e.g., transcripts and degrees from an Algerian University) into English. By employing a comparative forensic gap analysis grounded in the Multidimensional Quality Metrics (MQM) framework, this study identifies critical reliability gaps in specialized legal terminology, the authentication of sovereign stamps, and data integrity. The paper argues that while Gemini serves as a powerful productivity catalyst, the Language Entrepreneur must exercise a Human-in-the-Loop auditing framework to mitigate algorithmic hallucinations and ensure legal validity within the Algerian administrative context.

Keywords: Gemini (LLM); Official Translation; Administrative Documents; Algorithmic Reliability; Arabic into English Translation.

ملخص:

يُحدث دمج الذكاء الاصطناعي التوليدي، وبالأخص نموذج جيميني من جوجل، تحولاً جذرياً في ترجمة الوثائق الإدارية الرسمية. يستقصي هذا المقال مدى موثوقية النموذج في ترجمة المستخرجات الإدارية الصادرة باللغة العربية إلى اللغة الإنجليزية، من خلال تطبيق تحليل فجوات مقارنة يستند إلى إطار مقاييس الجودة متعددة الأبعاد. هذه الدراسة فجوات موثوقية حرجة في المصطلحات القانونية المتخصصة، وتوثيق الأختام السيادية، وسلامة البيانات. حيث تخلص الورقة إلى أنه رغم كون جيميني محفزاً قوياً للإنتاجية، إلا أنه يجب على ريادي اللغة تطبيق إطار تدقيق يعتمد على العنصر البشري للحد من الهلوسة الخوارزمية وضمان الصلاحية القانونية في السياق الإداري الجزائري. **الكلمات المفتاحية:** جيميني؛ الترجمة الرسمية؛ الوثائق الإدارية؛ الموثوقية الخوارزمية؛ الترجمة من العربية إلى الإنجليزية

Corresponding author: Adil HAOUES, e-mail: adil.haoues@univ-annaba.dz

1. INTRODUCTION

1.1 General Context and the Rise of GenAI in Translation

The translation landscape has undergone a seismic shift with the advent of Generative Artificial Intelligence (GenAI). Unlike traditional Neural Machine Translation (NMT) which relied on static pattern matching, Large Language Models (LLMs) such as Google's Gemini employ sophisticated transformer architectures to predict context and intent (Massey & Ehrensberger-Dow, 2026). This technological leap has democratized access to high-speed linguistic transfer, positioning AI not merely as a dictionary but as a contextual co-pilot. However, the rapid adoption of GenAI in specialized fields—specifically official and administrative translation—has outpaced the development of reliability protocols. While LLMs excel in creative and general discourse, their application in zero-error environments remains contentious due to the inherent risk of algorithmic hallucinations (Bindels et al., 2025).

1.2 Digital Paradox: Translating between Algerian Legalism and GenAI

In Algeria, the national strategy for the digitization of sectors has necessitated the modernization of both academic and civil administrative workflows. However, the Algerian administrative apparatus remains deeply rooted in a rigid legalistic tradition. This rigidity manifests across a spectrum of official documentation, ranging from the LMD (Licence-Master-Doctorate) pedagogical framework to essential civil status records, such as Birth and Marriage Certificates. Whether originating from universities or municipal registries, these documents serve as vital legal instruments for citizens seeking opportunities abroad. In this context, the reliability of a translation is not a linguistic luxury but a legal requirement. The translation must accurately reflect the sovereign authority of the state, validated by the diverse stamps of the Ministry of Higher Education, the Ministry of Interior, and the Ministry of Foreign Affairs. Consequently, the Algerian translator operates at the intersection of archaic bureaucratic requirements and the hyper-modernity of GenAI tools like Gemini.

1.3 Research Problem and Literature Review

This Trust Deficit is most acute in the phonetic and orthographic reconstruction of North African onomastics. As Bouloudani (2026) demonstrates, the linguistic distance between French and Arabic creates a phonological gap that LLMs must navigate by acting as phonetic mediators. Without a human Audit, AI may fall victim to vowel blindness or fail to promote short vowels into the necessary long vowels (alif/waw/ya) required for correct Arabic pronunciation.

This challenge is a direct manifestation of the paradigm shift described by O'Hagan (2019), who argues that technology is no longer a peripheral tool but has become the very environment in which translation takes place. This total immersion in a digital environment necessitates what O'Hagan calls a re-evaluation of translator competence

and translation quality, especially when dealing with the sovereign nature of administrative documents.

Furthermore, the comparison of Algerian transcripts reveals a critical distinction in functionalist adequacy. While Gemini can effectively map vocabulary, it often overlooks the theological weight and historical norms that dictate strategy choice. For instance, the human certified translator recognizes that administrative seals are not merely text but cultural vessels requiring explicit formal description—such as *Follows a round seal*—to maintain legal validity abroad. In contrast, unmonitored LLM outputs risk pragmatic failure by treating these sovereign markers as mere data strings. By reducing these cultural vessels to raw data, AI-driven processes risk failing the modern quality standards which, as O'Hagan (2019) notes, must account for the complex interaction between human agency and algorithmic processing. Such a failure ultimately strips the document of its Islamic contextual authenticity and legal hierarchy.

Building on this problematic intersection between algorithmic fluency and administrative rigidity, the present study seeks to address the following research questions:

- **RQ1:** To what extent can Gemini accurately translate specialized Algerian administrative and pedagogical terminology compared to a human-certified translation?
- **RQ2:** How does the probabilistic nature of LLMs impact the integrity of non-textual legal markers, such as official seals, fiscal stamps, and manual signatures?
- **RQ3:** What is the specific Human Intervention Rate (HIR) required to transform an AI-generated draft into a legally valid document within the Algerian administrative framework?

1.4 Claims and Methodology of the Research Work

This research is built upon the central claim that while Large Language Models (LLMs), specifically Google's Gemini flash, demonstrate high linguistic fluency, they remain legally unreliable for the autonomous translation of official administrative extracts. The study posits that Gemini lacks the Institutional Intelligence required to decode the non-textual components of official documents—such as the hierarchical significance of state seals and the specific pedagogical nuances of the Algerian LMD system.

To validate these claims, the research employs a Comparative Case Study Methodology, utilizing empirical data from a standardized set of academic and civil records issued by an Algerian State University and governmental authorities. These include a Bachelor's (License) degree and its corresponding transcripts, as well as a Birth Certificate and a Marriage Certificate. The identities of the document holder and

the specific institution have been anonymized to focus on the structural and terminological challenges. The methodological framework consists of three analytical phases:

1. **Direct Processing Phase:** Subjecting the original Arabic administrative extracts to Gemini's translation engine using Zero-Shot prompting to observe its raw performance and its ability to recognize Algerian administrative conventions.
2. **Gap Analysis Phase:** Performing a rigorous auditing process where Gemini's output is subjected to a professional evaluation by the author in his capacity as an Algerian sworn certified translator. This phase moves beyond simple textual comparison; it functions as a forensic linguistic audit to identify reliability gaps in specialized terminology, data integrity, and the rendering of sovereign metadata (seals and signatures) that are essential for legal certification
3. **Evaluative Synthesis:** Developing a Reliability Matrix based on the findings to determine the Human Intervention Rate (HIR) required to transform an AI-generated draft into a legally binding document suitable for use abroad.

2. THEORETICAL FRAMEWORK: RELIABILITY CHALLENGE

The intersection of generative artificial intelligence and official translation necessitates a robust theoretical examination of how algorithmic processes align with legal mandates. While Large Language Models (LLMs) offer unprecedented linguistic speed, their application in administrative contexts introduces significant risks regarding data integrity and sovereign validity. This section establishes the conceptual boundaries of the study and explores the architectural limitations that lead to reliability gaps in automated legal translation.

2.1. Conceptual Definitions and Scope

To establish a clear analytical scope, this study defines three core concepts based on established translation and legal scholarship:

- **Official Translation:** Within the Algerian context, this refers to a **certified translation** performed by a *Traducteur-Interprète Officiel Agréé* (Sworn Translator). Beyond mere semantic transfer, official translation is a legal act of mediation intended to ensure **legal equivalence** between the source and target texts, thereby carrying the sovereign weight of the state (Pym, 2014; Šarčević, 1997).
- **Reliability (Algorithmic):** In this research, reliability is defined not as linguistic fluency, but as the degree of **adequacy and data integrity** in reproducing administrative data (Castilho et al., 2018). A reliable output must ensure that **hard data**—such as registration numbers, grades, and onomastics—

remain unchanged and accurate during the probabilistic processing of the LLM (Poibeau, 2017).

- **Legal Validity:** This term denotes the document's ability to be accepted by foreign or domestic judicial and administrative authorities. As Monks (2008) and Bowker (2023) suggest, legal validity in official contexts depends on fulfilling **paratextual and formal requirements**, such as seals, stamps, and certifications. The omission of these elements renders a document **functionally defective**, regardless of its linguistic correctness.

2.2 Algorithmic Hallucinations and Probability versus Fact Dilemma

The reliability gap in Large Language Models (LLMs) is fundamentally rooted in their architecture. Gemini operates on statistical probability—predicting the next token based on vast datasets—rather than verifying factual or legal accuracy. As Poibeau (2017) explains, *“The language model is there to ensure that the output looks like a valid sentence in the target language... It calculates the probability of a sequence of words”* (p. 144). Consequently, the model prioritizes linguistic structure over substantive truth.

- **Onomastic Inaccuracy (Proper Nouns):** LLMs often struggle with the phonological gap and specific orthographic traditions in North African names. As Bouloudani (2026, p. 17) argues, the *“transition from the Latin-based French script to the Arabic abjad involves profound phonetic adaptations' where the 'translator must act as a phonetic mediator, reconstructing the name to fit the phonotactic constraints' of the target language”*. For instance, Gemini might transliterate the name 'حواس' as 'Hawas', failing this role of mediation by ignoring the official administrative spelling in Algeria, which follows the French-influenced tradition 'Haoues'. Such a discrepancy in official identification, where the model fails to ensure the entity remains recognizable (Bouloudani, 2026) within a legal framework, can lead to the rejection of the document by foreign authorities, as names must match the passport and civil records exactly.
- **Table Reading and Data Misalignment:** Algerian academic transcripts are structurally complex documents. Gemini may treat sensitive identifiers—such as registration numbers—as linguistic patterns subject to statistical prediction. This probabilistic approach significantly increases the risk of *Data Misalignment*, where annual averages and grades may be erroneously linked to the wrong modules due to a failure in decoding the table's intersections.

2.3 Challenge of Non-Textual and Structural Elements

A significant hurdle for LLMs is the lack of "Institutional Intelligence" to decode elements that are not purely textual but carry immense legal weight.

- **Omission of Fiscal Stamps:** A critical failure in AI-generated translations is the systematic omission of **fiscal stamps** (e.g., ضريبة الدمغة or 10 DA stamps). While Gemini focuses on translating semantic content, it ignores these fiscal markers which verify that the administrative fee for the document has been paid. To a foreign consulate or legal entity, the absence of a reference to the fiscal stamp in the translation—usually described as Fiscal Stamp or Tax Stamp—suggests an incomplete or unofficial reproduction, rendering the document legally suspect.
- **Semantic Vacuum of Cursive Signatures:** Gemini is unable to interpret or describe **handwritten signatures** and manual annotations. A human translator identifies a signature (e.g., by algerian officials) as a *verification of intent*. Conversely, AI models often ignore these marks or fail to account for their position in the authentication chain.
- **Layout Fidelity and Structural Logic:** Official documents follow a rigid layout where the spatial arrangement of headers and stamps indicates a hierarchy of authority. Gemini typically generates a *stream of text*, which breaks the logical connection between a seal and the text it authenticates. Without preserving the document's layout, the translation fails the "functional adequacy" test required abroad.

2.4 Linguistic Fluency versus Administrative Precision: The Conflict

This creates a paradoxical situation where the output looks "correct" to a layperson but is "functionally defective" to an administrative official.

- **Fluency Trap:** This creates a paradoxical situation where the output looks 'correct' to a layperson but is 'functionally defective' to an administrative official. According to Bowker (2023, p. 11), the high linguistic fluency of Generative AI tools often creates a “*veneer of reliability that can be very seductive, leading users to believe that the translation is more accurate than it actually is*”. This deceptive quality ensures that Gemini’s ability to generate smooth English masks underlying failures in mapping specialized Algerian administrative terms. Consequently, the user is misled by the linguistic surface, failing to recognize that such fluency does not equate to the rigorous administrative precision required for legal validity.
- **Precision Requirements:** Official translation demands the exact mapping of titles, grades, and the complex authentication chain. Therefore, the value of a language professional is measured by their ability to "strategically audit" these algorithms to ensure legal validity.

2.5. Language Entrepreneur and Institutional Intelligence

The shift toward AI-mediated translation requires a redefinition of the translator's role and the cognitive assets they provide.

- **Language Entrepreneur:** As conceptualized by Veiga Díaz (2020), the language entrepreneur is a professional who transcends traditional bilingual transfer. In the context of this study, this role involves strategic risk management and the ability to audit algorithmic outputs. The language entrepreneur treats translation not as a product, but as a complex service requiring ethical oversight and cultural quality assurance to maintain legal standards.
- **Institutional Intelligence:** This concept refers to the translator's specialized knowledge of the hierarchical and administrative structures of a specific state. Institutional intelligence allows the professional to decode the sovereign significance of elements—such as the sequence of authentications between the Ministry of Higher Education and the Ministry of Foreign Affairs—which Gemini lacks due to its text-centric training. (Massey & Ehrensberger-Dow, 2026).

2.6. Quantifying Legal Validity: An MQM-Based Approach to Official Translation

To ensure a systematic and objective evaluation, this study adopts the Multidimensional Quality Metrics (MQM) framework as its primary methodological tool. This approach is essential to move beyond impressionistic assessments and provide an empirical analysis of Gemini's performance. By utilizing MQM, the study evaluates translation quality through a predefined hierarchy of errors, focusing specifically on the dimensions of Accuracy and Design. From the functionalist perspective of a sworn translator, the application of MQM allows for a rigorous quantification of legal fitness-for-purpose. Quality is thus measured by the model's ability to preserve critical administrative metadata, ensuring that the findings are grounded in a standardized, industry-recognized metric rather than purely stylistic preferences (Lommel et al., 2014).

3. METHODOLOGY AND CASE STUDY

To assess the reliability of Gemini in a real-world scenario, this study performed a gap analysis on a standardized set of official documents issued by the Algerian authorities. This collection included academic records issued by an Algerian state university, consisting of a Bachelor's (License) degree in Law and transcripts for various semesters, alongside civil status documents such as a Birth Certificate and a

Marriage Certificate. The analysis focused on how the model handles the transition from linguistic drafting to legal certification, highlighting the inherent risks of automated processing in official contexts, the sensitivity of the personal and legal data contained within these documents, and the structural limitations of LLM architectures in maintaining administrative fidelity (Massey & Ehrensberger-Dow, 2026).

3.1 Analytical Corpus

To ensure the empirical validity of this study, a purposive sampling method was employed to construct the analytical corpus. To assess the reliability of Gemini in a real-world scenario, this study performed a forensic gap analysis on a representative collection of academic and civil records issued by Algerian authorities. The data set consists of a Bachelor's (License) degree in Public Law and its corresponding annual academic transcripts covering the full degree cycle, alongside essential civil status documents, specifically a Birth Certificate and a Marriage Certificate.

The selection of these documents was based on their structural and legal complexity. While the academic records represent the Algerian LMD (Licence-Master-Doctorate) pedagogical system—requiring a nuanced understanding of modular nomenclature and credit systems—the civil status documents introduce a complex chain of physical authentications.

Collectively, the corpus includes:

- **Quantitative Data:** Tabular reports of grades, weighted annual averages (Moyenne), and credit distributions (Credits).
- **Qualitative Metadata:** Formal headers, registration numbers (Ref. N°), and multiple layers of authentication, including round state seals, faculty stamps, and manual registrar signatures.
- **Administrative Metadata:** Fiscal stamps and manuscript signatures that function as verifiable markers of official legal processing.

By integrating both pedagogical and civil status records, the study highlights the inherent risks of automated processing in diverse administrative contexts and the structural limitations of LLM architectures in maintaining legal fidelity.

The corpus was processed through Gemini 3 Flash using a zero-shot prompting technique to simulate a layperson's interaction with the AI. By adopting this approach, the study highlights the inherent risks of automated processing in official contexts and the structural limitations of LLM architectures in maintaining administrative fidelity. Subsequently, the generated outputs were cross-referenced against the Certified Professional Standard—established by the author's expertise as a sworn translator—to accurately measure the reliability gap. This comparative framework ensures that the evaluation is grounded in the actual legal and administrative requirements of the Algerian judicial system rather than a generic linguistic reference. To operationalize

this assessment, the gap analysis is structured around a predefined Typology of Errors derived from the MQM model, focusing on three primary dimensions: Accuracy (terminology, data integrity, and omissions), Design (layout and formatting), and Fluency (grammar and syntax). Finally, each identified error is assigned a Severity Level to determine its impact on the document's legal validity:

- **Critical:** Errors that invalidate the legal or sovereign status of the document (e.g., omission of seals).
- **Major:** Significant mistranslations that hinder institutional understanding (e.g., LMD nomenclature errors).
- **Minor:** Stylistic or linguistic preferences that do not compromise the document's factual integrity.

This systematic approach provides the empirical evidence necessary to calculate the **Human Intervention Rate (HIR)** and evaluate Gemini's reliability in zero-error administrative environments.

3.2 Ethical Implications

The use of Generative AI in translating official administrative extracts raises significant ethical dilemmas, primarily concerning Data Privacy (Local Algerian Law compliance) and Institutional Accountability.

- **Confidentiality and Data Sovereignty:** Uploading sensitive academic records containing full names, dates of birth, and unique identification numbers to a third-party LLM cloud poses a high risk of data leakage. Once ingested, this information can theoretically be used for model fine-tuning, violating the principle of professional secrecy inherent in legal translation (Massey & Ehrensberger-Dow, 2026).
- **Deception of Fluency:** Ethically, the *Fluency Trap* identified in LLMs creates a false sense of security for the end-user. If a student submits a hallucinated but professional-looking translation to a foreign university, the ethical burden of the error remains ambiguous: is it the fault of the AI, the user, or the institution? (Pym & Hao, 2024).
- **Algorithmic Bias and Erasure:** By omitting descriptions of state seals and official stamps, AI-driven translation inadvertently performs a form of *administrative erasure*, stripping the document of its sovereign legal weight and reducing it to mere information, which undermines the translator's ethical duty to preserve the document's legal integrity (Massey et al., 2023).

4. Systematic Error Analysis and Reliability Matrix

this study utilizes the MQM framework to quantify the failures of Gemini 3 Flash by classifying each error according to its nature (Typology) and its impact on the document's legal validity (Severity). The following Comparative Reliability Matrix

provides a synthesis of these empirical findings, evaluating the AI's performance against a Human Certified Reference (HCR) within the specific context of Algerian administrative and academic documentation. By contrasting the AI's probabilistic linguistic drafting with the professional's rigid adherence to legal and structural norms, this analysis highlights the specific reliability gaps that necessitate human intervention for official certification.

Table 1: MQM-Based Systematic Evaluation of Gemini 3 Flash Performance

MQM Dimension	Error Typology	Gemini 3 Flash Performance	Severity Level	Human Intervention required
Accuracy	Terminology: LMD Nomenclature	Inconsistent/Generic: Struggles with وحدة تعليم أساسية و وحدة تعليم منهجية	Major	Mandatory correction of pedagogical units.
Accuracy	Data Integrity: Grades & Numbers	Variable: Risks misaligning decimals (e.g., 10.41) and registration IDs.	Critical	Absolute verification against source grid.
Accuracy	Proper Nouns: Transliteration	Potential Inaccuracy: Risks failing to maintain orthographic consistency in personal and geographical names.	Critical	Manual verification against official civil records.
Accuracy	Omission: State Seals & Stamps	Systematic Omission: Ignores sovereign markers from Higher Ed & Foreign Affairs.	Critical	Descriptive restoration of all official stamps.
Accuracy	Omission: Manuscript Signatures	Total Disregard: Fails to recognize or denote the presence of handwritten official signatures.	Critical	Manual certification and descriptive translation of all signatory markers.
Design	Layout & Formatting	Linear Stream: Breaks the structural logic of tables and headers.	Major	Full restructuring to match official layout.
Fluency	Grammar & Syntax	High Proficiency: Almost natural and grammatically correct English syntax.	None	Minimal stylistic post-editing.
Accuracy	Omission: Fiscal Stamps	Ignored: Fails to recognize fiscal tax markers (e.g., 10 DA stamps).	Critical	Manual documentation of tax processing.

Note. Created by the author

The systematic evidence reveals a Polarized Reliability Pattern: while Gemini achieves high-level scores in Fluency (linguistic naturalness and syntax), it demonstrates a systemic failure in the Accuracy dimension, specifically regarding non-textual, numerical, and onomastic data. From the professional perspective of a sworn certified translator, the severity of an error is not determined by its linguistic frequency but by its legal consequence.

The 100% omission rate of sovereign markers—including official seals and manuscript signatures—combined with the potential inaccuracy in transliterating proper nouns and registration numbers, constitutes a Critical Reliability Gap. In the legal reality of official documentation, validity is binary: a document is either legally authentic or null. Consequently, Gemini's structural inability to apply Administrative Knowledge and ensure the exact orthographic consistency of names and figures necessitates a Human Intervention Rate (HIR) of 100% for the purpose of legal certification, regardless of the model's linguistic prowess.

5. PROPOSED STRATEGY: A CONSTRAINT-BASED PROMPTING MODEL

The empirical findings suggest that Gemini 3 Flash serves as a productivity catalyst rather than a replacement for legal responsibility. To bridge the identified reliability gap—specifically in the dimensions of Accuracy and Design—the study proposes a Constraint-Based Prompting model. This strategy empowers the Language Entrepreneur to pre-emptively mitigate algorithmic failures before the final human audit.

5.1. Tripartite Framework: Aligning AI with Administrative Rigidities

To ensure outputs align with the administrative rigidities of the Algerian system, professionals should employ a structured tripartite prompting framework. This approach addresses the necessity for users to actively manage the Machine Translation (MT) process from the outset. As noted by Kenny (2022, p. 20), successful interaction with translation technology requires that users “*understand how to interact with the system in a way that maximizes its potential while minimizing its risks.*” In the Algerian high-risk domain, where accuracy is non-negotiable, this strategic interaction facilitates Full Post-editing (FPE) and ensures the final output meets the requirements for legal certification.

5.1.1 Persona Prompting (Contextual Grounding):

- *Instruction:* Act as a certified Algerian translator specialized in the LMD pedagogical system.

- *Goal:* This forces the model to move away from generic linguistic patterns and adopt the specific professional persona required to translate academic degrees like the **License in Public Law**.

“The post-editor must have knowledge of the domain that the text deals with.” (Nitzke & Hansen-Schirra, 2021, p. 35).

5.1.2 Terminology Injection (Terminology Control):

- *Instruction:* Forcing the model to use Exploratory Unit for وحدة إكتشافية and Fundamental Unit for وحدة تعليمية أساسية.
- *Goal:* This mitigates the Fluency Trap by ensuring that critical LMD nomenclature is mapped accurately for international credit recognition.

“The terminology must be consistent and comply with domain-specific requirements.” (Nitzke & Hansen-Schirra, 2021, p. 35)

5.1.3 Auditing Prompt (Structural Integrity):

- *Instruction:* Review the output and ensure all Ministry of Foreign Affairs and Ministry of Higher Education stamps are described and all registration numbers are verified.
- *Goal:* This step addresses the systematic omission of non-textual markers, such as the **round state seals** and **handwritten signatures**, ensuring they are explicitly noted (e.g., Sgd or Follows a round seal).

“Full PE is defined as the process of post-editing to obtain a product comparable to a product obtained by human translation.” (Nitzke & Hansen-Schirra, 2021, p. 34).

5.2 Hybrid Synergy

By integrating these engineering techniques, the human professional transforms from a traditional translator into an *Algorithmic Auditor*. This hybrid approach ensures that sensitive data, such as annual averages like 10.41 or 10.85, are not only translated but correctly aligned with their respective modules, maintaining the document's legal and institutional authenticity abroad.

6. CONCLUSION

The integration of Large Language Models (LLMs) like Gemini into the workflow of official translation represents a permanent paradigm shift. However, this study demonstrates that linguistic fluency does not equate to administrative reliability.

Based on the systematic evaluation using the MQM framework, the study provides the following conclusions:

Regarding RQ1 (Terminology): While Gemini exhibits high-level semantic mapping, it fails to capture the institutional precision of the Algerian LMD system. The "Fluency Trap" often masks significant terminology inconsistencies that could jeopardize the pedagogical validity of academic records abroad.

Regarding RQ2 (Non-Textual Markers): The research identifies a Critical Reliability Gap in the perception of sovereign markers. The systematic omission of state seals, fiscal stamps, and manuscript signatures confirms that LLM architectures currently lack the "Institutional Intelligence" required to authenticate official documents.

Regarding RQ3 (Human Intervention Rate): Despite Gemini's role as a productivity catalyst, the Human Intervention Rate (HIR) remains at 100%. In the binary reality of legal documentation—where a text is either authentic or null—automated outputs remain functionally defective without a forensic human audit.

- **Strategic Recommendations for Practitioners**

To navigate this hybrid era and safeguard the integrity of official translations, the study proposes the following professional protocols:

Metadata Descriptive Auditing: Practitioners must manually ensure that all non-textual sovereign elements; specifically round seals, faculty stamps, and manuscript signatures are explicitly described and translated, as these critical markers remain invisible to current LLM architectures.

Strict Data-Grid Verification: To mitigate the risks of "vowel blindness" and numerical hallucinations, a side-by-side verification method should be utilized to check the integrity of onomastic data (proper nouns) and numerical values (grades and registration numbers) against the source document.

Advanced Constraint-Based Prompting: Before processing sensitive documents, translators should employ "Persona" and "Terminology" injection prompts. This proactive step injects the necessary institutional context, reducing the "Fluency Trap" and minimizing terminology inconsistencies from the outset.

Forensic Risk Management: Translators should treat AI-generated drafts as raw material requiring a 100% human-in-the-loop audit. Legal certification must never be granted based on linguistic smoothness alone, but rather on the total orthographic and structural alignment with the sovereign legal framework.

7. REFERENCES

- Bindels, J., Pluymaekers, M., & Dorst, A. G. (2025). (Re)defining machine translation literacy: From a competence-based to a process-based approach. *Revista Tradumàtica*, 23, 309–325. <https://doi.org/10.5565/rev/tradumatica.493>
- Bouloudani, M. (2026). Linguistic mechanisms and phonetic adaptation: Strategies for translating proper nouns between French and Arabic. *El-Tawassol Journal*, 32(1), 16–24. <https://asjp.cerist.dz/en/article/288133>
- Bowker, L. (2023). *De-mystifying translation: Introducing translation to non-translators*. Routledge. <https://doi.org/10.4324/9781003217718>
- Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). A retrospective on the role of human evaluation of machine translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 19–38). Springer. https://doi.org/10.1007/978-3-319-91241-7_2
- Kenny, D. (Ed.). (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Language Science Press. <https://doi.org/10.5281/zenodo.6653406>
- Lommel, A. R., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica: tecnologies de la traducció*, (12), 455–463. <https://doi.org/10.5565/rev/tradumatica.77>
- Massey, G., & Ehrensberger-Dow, M. (2026). *Translation competence in the age of generative AI: Debates, dilemmas, directions*. <https://doi.org/10.5281/zenodo.17641064>
- Massey, G., Piotrowska, M., & Marczak, M. (2023). Meeting evolution with innovation: An introduction to (re-)profiling T&I education. *The Interpreter and Translator Trainer*, 17(3), 325–331. <https://doi.org/10.1080/1750399X.2023.2237321>
- Monks, P. (2008). [Review of the book *Translating law*, by D. Cao]. *The Journal of Specialised Translation*, (9), 135–136. <https://doi.org/10.26034/cm.jostrans.2008.689>
- Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing* (Translation and Multilingual Natural Language Processing, Vol. 16). Language Science Press. <https://doi.org/10.5281/zenodo.5646896>
- O'Hagan, M. (Ed.). (2019). *The Routledge handbook of translation and technology* (1st ed.). Routledge. <https://doi.org/10.4324/9781315311258>

- Poibeau, T. (2017). *Machine translation*. MIT Press. https://ia802801.us.archive.org/9/items/ThierryPoibeauMachineTranslation/Thierry%20Poibeau_Machine%20Translation.pdf
- Pym, A. (2014). *Exploring translation theories* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315857633>
- Pym, A., & Hao, Y. (2024). *How to augment language skills: Generative AI and machine translation in language learning and translator training* (1st ed.). Routledge. <https://doi.org/10.4324/9781032648033>
- Šarčević, S. (1997). History of legal translation. In *New approach to legal translation* (pp. 23–53). Kluwer Law International.
- Veiga Díaz, M. (2020). Moving towards entrepreneurial translator education: A review of entrepreneurship competence in Spanish translator education programmes. *HERMES - Journal of Language and Communication in Business*, 60, 47–64. <https://doi.org/10.7146/hjlc.v60i0.121310>