

From Cheating to Competence: A Pedagogical Framework for Reimagining Machine Translation in Higher Education

Angela Dumitran

Dimitrie Cantemir University, Bucharest, Romania

dumitranangela@gmail.com, [ORCID](#)

DOI: 10.63467/all17.art12

Abstract

The rapid advancement of Neural Machine Translation (NMT) and Large Language Models (LLMs) has sparked a crisis of academic integrity in Higher Education, particularly in Foreign Language departments. Traditional pedagogical models often categorize the use of tools like DeepL or ChatGPT as "cheating," a view that increasingly conflicts with the demands of the global professional market. This paper proposes a paradigm shift from prohibition to Machine Translation Literacy (MTL).

Using a "Human-in-the-Loop" (HITL) instructional design within English Linguistics and Practical Language courses, this study explores how the student's role evolves from a passive learner to a critical evaluator. The proposed framework replaces traditional rote translation with a three-tiered educational process: collaborative drafting, diagnostic post-editing (error taxonomy), and sociolinguistic critique. Preliminary observations indicate that this approach promotes high-level metacognitive skills, as students must justify linguistic choices that the AI fails to make. Rather than diminishing rigor, the integration of AI into the curriculum fosters "translingual competence," transforming the classroom into a laboratory for critical thinking. The paper concludes that by embracing AI, educators can move beyond the "cheating" narrative to cultivate the sophisticated analytical skills essential for the modern researcher.

Keywords: Machine Translation Literacy, Post-Editing, Academic Integrity, Higher Education Pedagogy, Applied Linguistics, Human-in-the-Loop.

1.Introduction

The emergence of high-fidelity Neural Machine Translation (NMT) and Large Language Models (LLMs) has fundamentally disrupted the epistemological foundations upon which Foreign Language (FL) departments have historically operated. For decades, the prevailing pedagogical consensus held that autonomous, unaided text production constituted the most reliable index of language competence - a paradigm that positioned the learner as an independent linguistic agent whose proficiency could be meaningfully measured in the absence of external technological intervention. This consensus has been rendered increasingly untenable by the rapid proliferation of AI-assisted translation tools now embedded in the professional and academic lives of language students worldwide. The current moment is therefore characterized by a fundamental tension: on one side stands an institutional reflex toward prohibition, in which the use of tools such as DeepL, ChatGPT, or Gemini in assessed work is routinely framed as academic dishonesty; on the other stands the sociotechnical reality of a global workplace in which proficiency with - and critical awareness of - these same tools is increasingly regarded as a core professional competency.

In translation studies and language faculties - where students are being trained as future translators, literary specialists, terminologists, and legal or medical language professionals - the stakes of this tension are particularly high. Raw LLM output, while impressively fluent, frequently falls short when it comes to rendering the cultural nuance of literary prose, the pragmatic force of speech acts, or the terminological precision demanded by legal contracts and medical records. Educators and students alike must develop what might be called *AI literacy for language professionals*: the ability to craft targeted, context-sensitive prompts, to critically evaluate and post-edit model output, and to intervene where necessary to restore the human judgment that automated systems inevitably lack. Professors bear a dual responsibility here - becoming proficient users of these tools themselves while designing pedagogical frameworks that guide students through the complexities of prompt engineering, ethical AI use, and domain-specific quality assessment. Yet many institutions continue to respond with prohibition rather than integration, a position that, as Selwyn

(2022) has argued, reflects institutional anxiety more than sound pedagogical reasoning, and that, as Fyfe (2023) has shown, is increasingly impossible to enforce. The challenge, then, is not policing, it is redesign. The institutional reflex of prohibition has proven both pedagogically counterproductive and practically unenforceable, leaving language faculties without coherent tools for developing the critical AI literacy that the contemporary professional landscape demands. The present study responds to this gap by proposing a structured pedagogical framework - Machine Translation Literacy - grounded in Human-in-the-Loop instructional design, which repositions the student as a critical evaluator of machine output rather than a passive recipient of it.

2. Theoretical Background

Neural Machine Translation tools have undergone dramatic development since the mid-2010s, when deep learning architectures began to displace earlier rule-based and phrase-based statistical systems. DeepL, a neural machine translation service developed by the German company DeepL SE and first launched in 2017, is a representative case in point. Built initially on a convolutional neural network architecture and subsequently enhanced with transformer-based models, it has become widely regarded among translators and linguists as the highest-fidelity publicly accessible MT system for European language pairs, consistently outperforming competing services on fluency and domain-specific accuracy benchmarks (Nurminen et al., 2023; Vanmassenhove et al., 2021).

Moreover, the Human-in-the-Loop paradigm originates in machine learning and human-computer interaction research, where it designates a design philosophy in which human judgment is preserved as an active, iterative component of an otherwise automated process (Monarch, 2021). Applied to the domain of translation, the HITL model reframes the student not as a passive recipient of machine-generated output, but as a trained evaluator who interrogates, revises, and assumes epistemic responsibility for that output. Rather than outsourcing cognition to the machine, the HITL translator deploys metalinguistic, cultural, and pragmatic knowledge to identify where automated output fails and why. It is this disposition, rather than the capacity for unaided production, that the present paper argues should constitute the new target competency in FL higher education.

On the other hand, the contemporary NMT landscape, exemplified by DeepL, Google Translate's Neural mode, and the multilingual capabilities of models such as GPT-4 and Claude, represents a qualitative departure from earlier MT paradigms. Output quality has improved to the point where errors are often subtle, stylistic, or register-dependent rather than grammatically catastrophic, which means that detecting and correcting them demands a higher, not lower, order of linguistic competence. For students at Faculties of Foreign Languages, this has consequential implications: exposure to NMT in the classroom is no longer a question of if, but how, and the nature of the competence it demands calls for a reconceptualization of what translation performance means in assessed contexts. Against this backdrop, the present study develops a pedagogical framework for what we term Machine Translation Literacy (MTL): a structured, competency-based approach to integrating NMT tools into FL curricula in ways that cultivate critical, reflective, and professionally relevant language use. Drawing on data from students enrolled in translation and language courses at the Faculty of Foreign Languages, the study examines how engagement with DeepL and comparable tools mediates translation performance, shapes metalinguistic awareness, and challenges prevailing assessment paradigms. The overarching argument is that the transition from policing AI to leveraging it pedagogically is not a concession to academic dishonesty, but an affirmative response to the demands of twenty-first-century language education.

The concept of Machine Translation Literacy (MTL) has emerged as a response to the widespread, often uncritical adoption of automated translation tools across both professional and academic contexts. Bowker & Ciro (2019) define MT literacy as the capacity to understand what machine translation can and cannot do, to evaluate its output critically, and to make informed decisions about when and how to use it. This definition foregrounds competence over prohibition: rather than treating NMT tools as threats to authentic language learning, MTL positions them as objects of study whose affordances and limitations must be systematically understood. For students enrolled in language faculties, acquiring MT literacy is not a supplementary skill, but a professional necessity, given that post-editing of machine-generated content has become a standard workflow in the translation and localization industries (Bowker, 2023). Post-editing - the process of reviewing and correcting machine-translated text to meet a defined quality standard - has been extensively theorized in translation studies as both a professional practice and, more recently, a pedagogical tool. Rather than having students produce translations from scratch - a

task that, as Pym (2013) notes, tends to prioritize output speed over metacognitive reflection, post-editing tasks require them to adopt a forensic orientation toward language, diagnosing error types, hypothesizing causes, and evaluating alternative formulations. This reorientation from producer to critical evaluator represents a significant and productive shift in pedagogical design.

The pedagogical value of post-editing has been empirically supported across several recent studies. Castilho et al. (2017) found that students who engaged with post-editing tasks demonstrated heightened awareness of syntactic divergence between source and target languages, while Koponen (2016) identified post-editing as an effective vehicle for developing error-recognition skills that are transferable across text types and domains. Of particular relevance for language faculties is the domain-specificity of NMT errors: the quality degradation that occurs when general-purpose MT systems encounter specialized terminology - whether legal, medical, or literary - provides rich material for discipline-grounded analysis. A legal translation course, for instance, can leverage the systematic miscalibration of NMT output in contractual or regulatory texts as a resource for teaching terminological precision and register awareness; a literary translation course can exploit the flattening of stylistic variation and the suppression of pragmatic nuance that characterizes machine-rendered fiction as an entry point into discussions of authorial voice and cultural mediation.

Beyond post-editing competencies, there is also the prompting engineering. Unlike the relatively passive interaction with standalone NMT engines, engaging with LLMs requires the user to actively formulate input - specifying the task, the target audience, the desired register, the relevant domain constraints - in ways that make the quality of the output directly contingent on the sophistication of the prompt. For language students, this dynamic is pedagogically significant in at least two respects. First, it renders visible the relationship between metalinguistic awareness and communicative outcome: a student who understands the difference between a legal brief and a policy summary, or between a literary translator's interpretive freedom and a technical translator's fidelity constraints, will formulate qualitatively different prompts and obtain qualitatively different results. Second, it positions the language specialist, rather than the technician or the data scientist, as the primary competent user of the tool, since the ability to specify and evaluate language quality is precisely the expertise that language education cultivates. Cope & Kalantzis (2024) have argued that AI-mediated writing environments call for a new form of "cybersocial literacy," in which the

learner must navigate collaboratively between human intention and machine execution; prompt engineering in applied linguistics contexts is a concrete instantiation of that broader shift.

Traditional language assessment has long operated on the premise that linguistic competence can be reliably measured under closed-book, controlled conditions - timed translation passages, unseen grammar tests, written examinations completed without reference materials - on the assumption that what a student produces independently reflects what they have genuinely internalized. This model rests on what might be called a stable epistemological contract between institution and learner: the examiner controls the conditions of production, and the output can therefore be read as transparent evidence of underlying knowledge. As Pym (2013) has argued, however, translation pedagogy has historically prioritized measurable output over the metacognitive processes that actually constitute professional competence, a tendency that leaves assessment frameworks structurally ill-equipped to respond when the conditions of production change. Large language models have changed them irrevocably. When a student can generate grammatically impeccable, register-appropriate, stylistically polished prose within seconds, the surface features that closed-book assessment traditionally used as proxies for linguistic knowledge, syntactic accuracy, lexical range, textual cohesion, fluency, are no longer reliable indicators of what the learner actually understands.

The institutional response to this disruption has, in many cases, taken the form of an academic integrity narrative: the student "cheated," the output is inauthentic, the submission must be penalized. This reaction, while understandable, conflates a pedagogical failure with a moral one. The deeper problem is not ethical but epistemological, and it maps directly onto what AI research has described as the "black box" problem: the opacity of machine-generated output whose internal reasoning cannot be audited or traced (Selwyn, 2022). A student who submits AI-generated text without critically engaging with it presents precisely the same interpretive challenge to the assessor: a polished surface with no window onto the processes that produced it, no evidence of syntactic decision-making, no trace of the sociolinguistic or pragmatic reasoning that Bowker & Ciro (2019) identify as the core of genuine machine translation literacy. The assessor cannot determine whether the student has grappled with the register demands of the target culture, the semantic ambiguities of the source text, or the terminological constraints of the domain, because the assessment instrument was never designed to make such processes visible. As Cope &

Kalantzis (2024) observe, AI-mediated writing environments demand a new form of cybersocial literacy in which the boundary between human intention and machine execution must be actively negotiated rather than simply policed; an assessment model built on prohibition cannot, by definition, cultivate that negotiation. The "cheating" narrative, in this light, is less a diagnosis of student misconduct than a symptom of institutional unpreparedness - a defensive response to the collapse of output-based measurement, and a signal that the linguistic rigor traditional assessment claimed to enforce was always more dependent on controlled conditions than on genuine insight into the learner's competence.

The challenges of teaching and learning a foreign language have never been straightforward, but the emergence of large language models as default translation and writing aids has introduced a qualitatively new layer of difficulty: one that is arguably more insidious than the errors produced by earlier generations of machine translation systems. The central paradox of the current moment is this: the more fluent and grammatically sophisticated AI output becomes, the harder it is for learners to detect its errors, and the more likely those errors are to be absorbed, reproduced, and eventually normalized. This section examines the specific mechanisms through which AI-assisted translation produces false analogies, collocation errors, and lexical interferences in the English-Romanian language pair, drawing on empirical evidence from the PhD thesis of Dumitran (2021) on lexical errors in machine translation of medical texts, from Dumitran's comparative study of false friends across human learners and AI systems, and from classroom observations of legal translation students at the faculty level. In the English-Romanian linguistic contact zone, false friends are structurally overdetermined: as Roibu (2014) has argued, the partial Latinity of English vocabulary, combined with the accelerating influence of English as a global *lingua franca* and primary intermediate language in Romanian educational settings, has created a large and growing inventory of deceptive cognates whose surface similarity actively misleads both human learners and automated systems. Where earlier MT systems produced dysfluent outputs that a moderately competent reader could identify as erroneous, current LLMs produce morphosyntactically well-formed, stylistically plausible sentences in which the false friend error is embedded within grammatically impeccable prose, rendering it effectively invisible to a learner who lacks the metalinguistic awareness to look past surface fluency.

3. Methodology

The present study adopts a mixed-methods design, combining quantitative corpus analysis with qualitative classroom observation, in order to capture both the measurable patterns of AI-generated translation errors and the pedagogical processes through which students learn to identify and correct them. The quantitative component draws on a corpus of 230 texts translated from English into Romanian using Google Translate, sourced from official WHO and Gavi communications produced during the COVID-19 pandemic. Each text was subjected to systematic error annotation following an adapted version of the ISO 18587:2017 post-editing quality framework, with errors coded across four primary categories: lexical collocation failures, false friend interference, syntactic transfer, and domain-specific terminological miscalibration. Frequency counts and proportional distributions were calculated for each error category, enabling comparison across domains: legal, medical, and administrative and across the English-Romanian language pair specifically. Additional quantitative data were drawn from Dumitran's (2021) doctoral corpus on lexical errors in machine translation of medical texts, which provided a validated baseline for error distribution figures, including the finding that collocation errors account for 31.7% of all lexical errors and false friend errors for 10.3%.

The qualitative component of the study was conducted within the framework of Human-in-the-Loop (HITL) instructional design, implemented across practical language and legal translation courses at the Faculty of Foreign Languages. Classroom observations were carried out over the course of one academic semester, during which students engaged with a three-tiered pedagogical sequence: collaborative AI-assisted drafting using DeepL and comparable LLMs, structured diagnostic post-editing guided by the error taxonomy developed from the corpus analysis, and sociolinguistic critique requiring students to justify their corrective choices with reference to target-language register conventions. Data collected qualitatively included field notes from classroom sessions, analysis of annotated student submissions, and informal reflective discussions in which students were asked to articulate why specific AI-generated forms - collocational calques, false cognate renderings, syntactically transferred structures - were professionally inadequate.

These classroom interactions provided the primary evidence for assessing whether and how the pedagogical framework fostered metalinguistic awareness and critical evaluative competence.

The integration of quantitative and qualitative methods was deliberate and theoretically motivated: the corpus data established the structural predictability and domain-specific distribution of AI translation errors, while the classroom observations revealed the conditions under which those errors become pedagogically productive rather than merely diagnostic. Taken together, the two strands of evidence address complementary research questions: *what* errors AI systems systematically produce, and *how* structured exposure to those errors develops the critical competencies that Machine Translation Literacy requires. The methodology was not designed to test a hypothesis in the experimental sense, but to construct, document, and iteratively refine a pedagogical model grounded in empirical evidence from both corpus linguistics and applied classroom practice. As such, it is best characterized as practitioner research with a mixed-methods orientation, situated at the intersection of translation studies, applied linguistics, and higher education pedagogy.

4. Results and Discussions

The evidence from the classroom corpus collected for the present study confirms this trajectory at every level of specialization. In legal translation practice, the term *expertise* is systematically rendered by AI systems - as well as by students who have internalized AI output as a model - as *expertiză*, a false cognate that carries a generalized meaning of "skill" or "competence" in Romanian (*DEX*: "pricepere, iscusință"), whereas the correct legal equivalent for the English *expertise* in the sense of a formal expert report or judicial assessment is *expertiză judiciară* or simply *raport de expertiză*. The semantic overlap between the English and Romanian forms is sufficient to generate the error, but the legal register demands terminological precision that the false cognate entirely fails to provide. Similarly, the English legal term *section* - referring to a numbered subdivision of a legislative text, a contract, or a statutory instrument - is recurrently mistranslated as *secțiune* by both students and AI systems, when Romanian legal convention consistently uses *articol* for statutory subdivisions and *clauză* for contractual ones. The substitution of *secțiune* for *articol* is not a random error: it is a predictable product of the formal similarity between the English *section* and the Romanian *secțiune*, and it illustrates precisely the

mechanism that Chuquet and Paillard (1989: 224) described as the core of the false friend phenomenon - "words that are close by form but partially or completely different by meaning." In a legal context, however, the stakes of such errors extend well beyond stylistic infelicity: a contract clause that references the wrong structural unit of a legislative text is not merely inelegant but potentially legally void.

The medical translation domain reveals the same pattern operating at scale. The analysis of 230 texts translated from English into Romanian by Google Translate, drawn from official WHO and Gavi communications during the COVID-19 pandemic, found that false friend errors accounted for 10.3% of total lexical errors. Representative cases include the rendering of *support* as the physical *suport* (a stand or rack) rather than the social and medical *sprijin* (assistance, backing), and the translation of *actual* as *de fapt* (in fact, actually) rather than the Romanian *actual* meaning "current" or "present" - a paradigmatic instance of bidirectional false-friend interference, since the English and Romanian terms are formally identical but semantically inverted. In the pandemic context, where accurate public health communication carried life-or-death implications, such errors were not merely academic: a Romanian reader encountering *suport* where *sprijin* was intended, or *actual* where "real" or "factual" was meant, received not an approximate message but a structurally misleading one.

If false friends represent the most structurally predictable category of AI-generated error in the English-Romanian pair, collocation errors are by far the most frequent and, from a pedagogical standpoint, the most consequential. Dumitran (2021) establishes that collocation errors constitute 31.7% of all lexical errors in the MT corpus, by a substantial margin the dominant error category, exceeding borrowing errors (17.2%), false friend errors (10.3%), and stylistic errors (10.3%) combined. This distribution reflects a fundamental architectural limitation of neural MT systems: as Zinsmeister and Heid (2004: 311) have argued, collocational knowledge - the language-specific conventions governing which words legitimately combine with which other words - encodes constraints that cannot be derived from general syntactic or semantic rules and must instead be acquired through extensive exposure to authentic language use.

The classroom evidence from legal translation courses at the faculty is particularly revealing in this regard. Students consistently produced and, when surveyed, frequently attributed to AI assistance collocations such as *consecințele adverse* for "adverse consequences" (the established

Romanian medical collocate being *efectele adverse*), *scădere abruptă* for "steep fall-off". In legal texts specifically, the collocation error pattern extends to the rendering of fixed legal phrases whose component words may individually be correctly translated but whose combination violates Romanian legal usage: *administrarea dovezilor* (management of evidence) appears in place of *administrarea probelor*; *cheltuielile financiare* (a semantically redundant construction that combines two expressions for the same concept) appears where *cheltuieli* alone is standard; and *atac bioteror* appears in place of the recognized Romanian legal and security term *atac biologic*. In each case, the AI system has processed the lexical components of the source expression correctly but has failed to retrieve the target-language collocate - the "lexical bond," as Dumitran (2021) describes it - that Romanian convention requires.

The false friend pairs documented across both the SLA and the AI corpora in Dumitran's research illustrate this diagnostic challenge concretely. Terms such as *eventual* (Romanian: "possible/perhaps") rendered as "eventually" (English: "in the end, finally"), *a avertiza* (Romanian: "to warn") confused with "to advertise" (English: "to promote"), *audiență* (Romanian: "a formal meeting or hearing") rendered as "audience" (English: "spectators, public"), and *evidență* (Romanian: "registry, record") confused with "evidence" (English: "proof") all produce output that is grammatically acceptable and superficially plausible, but that is semantically wrong in ways that matter acutely in legal, medical, and administrative contexts. A legal document in which *audiență* is used where "public hearing" is intended, or in which *evidență* is offered where "evidence" is required, does not merely fail stylistically: it communicates the wrong institutional reality, with potentially serious consequences for the parties it concerns.

This is the core pedagogical challenge that AI-era language teaching must confront. The errors that AI systems produce most reliably are not the errors that beginners make: phonetic substitutions, morphological misconstructions, elementary vocabulary confusions, but precisely the errors that distinguish an advanced learner from a competent professional: collocational precision, terminological appropriateness, register sensitivity, and pragmatic adequacy. By generating output that is advanced enough to pass as student work but flawed in the ways that professional formation is specifically designed to remediate, LLMs create a feedback loop that, if left pedagogically unaddressed, risks producing graduates who have never been forced to confront the precise competences they lack. As Koponen (2016) has argued, the detection and analysis of

MT errors is itself a high-order linguistic skill; the challenge for language faculties in the AI era is to ensure that this skill is systematically cultivated - not as an ancillary critical exercise, but as the central methodology through which professional translators and language specialists are formed. The theoretical framework of Machine Translation Literacy acquires its full pedagogical weight only when it is tested against the concrete difficulties that students encounter in actual translation practice: difficulties that, as the preceding section has established, are increasingly shaped not by the limitations of machine output but by its deceptive competence. The present subsection proposes a working error taxonomy derived from classroom observation in practical language and translation courses at the faculty level, organized around three analytically distinct, but empirically interrelated categories: syntactic interference, lexical collocation failure, and domain-specific terminological miscalibration. What unites these categories is not merely their frequency in student and AI-generated output, but the specific pedagogical trap each one sets: in each case, the error is structurally similar enough to a correct form to pass undetected by a learner who has not been trained to look for it, and sufficiently reinforced by AI output to acquire, over time, the false authority of a model.

The most immediately recognizable category of AI translation error in the English-Romanian pair, and yet, paradoxically, the one that students are least equipped to identify when it is embedded in otherwise fluent prose is the syntactic interference - the importation of source-language word order, clause structure, or morphosyntactic patterning into target-language output - represents the most immediately. Romanian and English diverge fundamentally in their syntactic organization: Romanian is a pro-drop language with rich inflectional morphology that allows considerable constituent order flexibility, governed by pragmatic principles of topic and focus distribution rather than by the relatively rigid subject-verb-object sequencing of English. Neural MT systems, trained on large parallel corpora in which English functions as either source or pivot language (Turovsky, 2016), systematically under-represent these pragmatic-syntactic distinctions, producing Romanian output in which the linear order of the English source is preserved at the expense of the discourse-functional organization expected in target-language text.

In classroom practice, this error manifests most visibly in nominal constructions and prepositional phrases. English favors premodification - the accumulation of attributive adjectives, noun adjuncts, and participial phrases before the head noun - a pattern that Romanian renders through

postmodification, using adjectival agreement, prepositional phrases, and relative clauses. AI-generated translations of English noun phrases such as "*evidence-based public health intervention*" or "*risk-stratified clinical management protocol*" - the type of compound nominal structure that characterizes legal, medical, and administrative discourse - routinely produce Romanian strings that preserve English-style pre-modification, generating output such as **intervenție bazată-pe-dovezi de sănătate publică* rather than the syntactically and collocationally appropriate *intervenție de sănătate publică bazată pe dovezi*. Students who have consulted AI output before producing their own translations reproduce these structures in their submitted work, not because they cannot recognize the Romanian norm when it is pointed out to them, but because the AI-generated form has been absorbed as plausible.

Verbal constructions present an analogous challenge. English aspectual distinctions between simple and progressive forms, and between active and passive voice in administrative and legal registers, are mapped by NMT systems onto Romanian equivalents that are formally adequate but pragmatically misaligned: the English passive, pervasive in legal and scientific writing, generates Romanian passive constructions (*este tradus, a fost aprobat*) where Romanian convention particularly in legal drafting - often prefers the reflexive-passive (*se traduce, s-a aprobat*) or an impersonal active construction. The student who has relied on AI output for a first draft and who lacks the legal or administrative register awareness to detect this misalignment will submit a translation that reads, to a native-speaking practitioner, as foreign - correct in its denotation, wrong in its institutional voice.

In legal translation practice, this distinction is acutely visible in the handling of verb-noun collocations that are fixed by convention in both English and Romanian but whose components do not align across the two systems. English legal usage requires *to bring an action* (not *to make* or *to take*); Romanian legal usage requires *a introduce o acțiune* or *a promova o acțiune* not *a aduce o acțiune*, which would be the literal calque of the English. AI systems consistently produce the calque, because they optimize for lexical correspondence rather than collocational convention. Similarly, English *to reach a verdict* maps in Romanian legal usage onto *a pronunța o hotărâre* or *a da o sentință*, not *a atinge un verdict*, but the latter is precisely what current NMT systems generate. The same mechanism operates in the medical discourse. Dumitran (2021) documents cases such as *doze timpurii* for "early doses" - where *timpurii* correctly translates "early" in a

temporal sense, but violates the medical collocational convention that requires *doze din primele tranșe* or *doze disponibile în faza inițială*; *descărcări din nas* for "nasal discharge" - where *descărcare* (electrical discharge, psychological unburdening) violates the medical collocate *surgeri nazale*; and *a prinde Covid-19* for "to catch Covid-19" - where the literal rendering of the English idiom produces a collocation that is pragmatically unacceptable in Romanian medical usage, which conventionally uses *a contracta* or *a lua*. In each case, the AI system has produced a string that is grammatically legal but collocationally wrong - and wrong in a way that specifically signals non-native professional competence to a practitioner in the field.

The foregoing analysis suggests that the specialized text: legal, medical, or administrative is not merely a challenging translation exercise, but a uniquely diagnostic pedagogical instrument: it is precisely in these domains that the limits of AI output become most clearly visible, and precisely here that the distance between surface fluency and professional adequacy is most measurable.

This is the pedagogical reorientation that the present study proposes for the classroom practice component of language faculty courses: rather than treating specialized texts as advanced exercises to be introduced only after general competence has been established, they should be used early and deliberately as diagnostic mirrors - texts through which students, confronted with AI output, are required to identify *why* a given rendering is wrong and *what competency* its correction demands. The question is not "is this a good translation?" - a question that students can often answer by feel, but "is this the right collocate, the right term, the right syntactic construction for this register, and what is the evidence for that judgment?" This shift from impressionistic evaluation to evidenced linguistic reasoning is, we argue, the core of what Machine Translation Literacy means in practice: not the capacity to avoid AI, but the capacity to audit it with the precision that the domains it serves require.

5. Conclusions

The argument developed across the preceding sections converges on a single, foundational claim: that the most urgent task facing language faculties in the age of AI is not the prevention of machine translation use, but the systematic mapping of its failure patterns and the deliberate integration of

that mapping into the design of translation and language pedagogy at every level. The evidence assembled in this study, drawn from the empirical corpus of Dumitran (2021), from classroom observation in legal and medical translation courses, and from the broader theoretical framework of Machine Translation Literacy, demonstrates that the errors produced by current NMT systems and large language models are neither random nor unpredictable. They are systematic. They follow identifiable structural logics: the logic of false-cognate interference, the logic of collocational calquing, the logic of syntactic transfer - that recur with sufficient regularity to constitute what this paper has called a grammar of machine deformation: a set of predictable distortions that AI systems impose on target-language text when the constraints of authentic linguistic convention are not externally enforced by a competent human evaluator. The pedagogical implication is direct and consequential: language faculties must develop and maintain structured error taxonomies - organized by error type, language pair, and domain - that function as shared diagnostic instruments for both instructors and students. Such taxonomies do not merely catalogue past errors; they generate expectations about where future errors will occur, enabling the instructor to design exercises that target the specific competency gaps that AI output systematically conceals. The classroom practice documented in this study - in which students are confronted with AI-generated translations of legal, medical, and administrative texts and required to identify, name, and correct errors by reference to established Romanian legal and medical convention - represents one model for how such a taxonomy can be operationalized. The specialized text, as argued in Section 3.2, is not simply a difficult exercise: it is a diagnostic mirror, and it is most useful precisely when AI output has made the error, because it forces the student to articulate why the machine was wrong and what genuine competence would look like in its place.

The rapid integration of Neural Machine Translation tools and Large Language Models into the daily academic and professional lives of language students has created an urgent crisis that existing pedagogical frameworks are ill-equipped to address. As AI-generated output grows increasingly fluent, the errors it produces - collocational failures, false-cognate interference, and syntactic transfer - become harder to detect and more likely to be absorbed by learners as authoritative models, particularly in high-stakes specialized domains such as legal and medical translation.

The second major conclusion of this study concerns the role of the educator in an AI-mediated learning environment and it is a conclusion that carries institutional as well as individual

implications. The teachers who ban AI tools from the translation classroom are not protecting linguistic rigor: they are removing the primary diagnostic instrument through which that rigor can be made visible, discussed, and cultivated. Conversely, the teacher who permits AI use without methodological scaffolding - who accepts AI-drafted translations without requiring students to account for their choices, defend their collocations, or justify their syntactic decisions - is not fostering technological fluency: they are abdicating the responsibility of formation. The position this paper has argued for throughout lies between these two failures, and it requires something more demanding than either prohibition or permissiveness: it requires the educator to become, themselves, a fluent and critical reader of AI output, capable of recognizing the deformations that NMT systems characteristically introduce and of designing learning tasks that make those deformations productive rather than harmful.

References

- Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing.
- Bowker, L. (2023). *De-mystifying translation: Introducing translation to non-translators*. Routledge.
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, 108(1), pp. 109-120. <https://doi.org/10.1515/pralin-2017-0013>.
- Dumitran, A. (2021). *Translation error in Google Translate from English into Romanian in texts related to coronavirus. Conference Proceedings of eLearning and Software for Education (eLSE)*, 17(2), pp. 37-43. Carol I National Defence University Publishing House.
- Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI & Society*, 38(4), pp. 1395-1405. <https://doi.org/10.1007/s00146-022-01397-z>.
- International Organization for Standardization. (2017). *ISO 18587:2017: Translation services — Post-editing of machine translation output — Requirements*. ISO.



<https://www.iso.org/standard/62970.html> ISO.

Kirana, A. (2024). The impact of using DeepL translator on EFL students' writing. In Repository UIN Sumatera Utara. UIN Sumatera Utara. Kirana, A., Kembaren, F. R. W., & Hz, B. I. R. (2024). The influence of DeepL Translator on EFL students' writing. *Jurnal Ilmu Sosial, Humaniora Dan Seni (JISHS)*, 2(4), pp. 746-753.

<https://doi.org/10.62379/jishs.v3i1.2059>.

Koehn, P. (2020). *Neural machine translation*. Cambridge University Press.
<https://doi.org/10.1017/9781108608480>.

Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, (25), pp. 131-148. <https://doi.org/10.26034/cm.jostrans.2016.303>.

Monarch, R. (2021). *Human-in-the-loop machine learning: Active learning and annotation for human-centered AI*. Manning Publications.

Nurminen, M., Brenner, J., Koponen, M., Latomaa, S., Mikhailov, M., Schierl, F., Ranasinghe, T., Culo, O., Groves, D., Ugan, R., Xu, H., & Yankovskaya, A. (Eds.). (2023). *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation.
<https://aclanthology.org/2023.eamt-1.0>.

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, 57(4), pp. 620–631. <https://doi.org/10.1111/ejed.12532>.

Vanmassenhove, E., Shterionov, D., & Way, A. (2021). Lost in translation: Loss and decay of linguistic richness in machine translation. *Machine Translation*, 35(4), pp. 495-523.
<https://doi.org/10.1007/s10590-021-09278-0>.

Yang, Y., Kandogan, E., Li, Y., Lasecki, W. S., & Sen, P. (2019). *HEIDL: Learning linguistic expressions with deep learning and human-in-the-loop*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 135–140). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P19-3023>.