MACHINE TRANSLATION VS. HUMAN TRANSLATION: A COMPARATIVE STUDY

Eshqorayeva Ibodat
English language teacher at Termez City Technical School No. 2

Abstract
Machine Translation (MT) has become a routine tool in education, business, and public services, especially after the rapid maturation of neural MT and post-editing workflows. Yet many institutions still rely on informal assumptions such as “MT is fast but inaccurate” or “human translation is always best,” which leads to weak decision-making in curriculum design, workplace communication, and quality assurance. This comparative study proposes an evaluation model that aligns translation goals with measurable outcomes, and contrasts MT and human translation across three common professional genres: technical instructions, customer-service communication, and promotional/marketing texts. Using a small pilot design with blind rating and an error-typology rubric, the study compares quality (accuracy, fluency, terminology, register, and consistency), efficiency (time, effort), and risk factors (confidentiality, accountability, and harm potential). The results indicate that MT performs competitively for predictable, terminology-driven technical content, but remains vulnerable to pragmatic meaning, politeness strategies, idioms, and brand voice. Human translation shows higher reliability for high-stakes and audience-sensitive texts, while hybrid MT + human post-editing offers the best balance of speed and quality when the task is well-scoped and supported by clear style guides. Practical recommendations are provided for teachers and administrators on how to integrate MT responsibly into lessons, assessment, and workplace-oriented language training.
Keywords: machine translation, human translation, post-editing, quality assessment, ESP, professional education

Introduction

Translation in professional settings is not merely a language exercise; it is a quality- and risk-sensitive service. In professional education institutions (technical schools, vocational colleges, and training centers), learners often meet English through job-related tasks: reading equipment manuals, writing emails, completing service dialogues, and understanding safety instructions. At the same time, students and teachers increasingly use MT tools on phones and laptops as a shortcut for comprehension and writing. This creates a new methodological question: when does

MT support learning and communication, and when does it introduce errors, reputational damage, or safety risks?

MT today is mainly driven by neural architectures and large-scale training data. In practice, it is strong at producing fluent output and handling frequent sentence patterns. However, fluency is not the same as correctness. A translation can sound natural while still being wrong in meaning, missing a requirement, or choosing the wrong register. Human translation, by contrast, relies on professional judgement: clarifying ambiguity, adapting to the audience, preserving legal or medical nuance, and matching cultural expectations. Human work is slower and costlier, but it can be more accountable, especially when consequences matter.

This article frames MT vs. human translation as a measurable comparison rather than a slogan. It asks three practical questions. (1) How do MT and human translation differ in quality across typical professional genres? (2) How large is the efficiency gap, and how does post-editing change the picture? (3) What are actionable guidelines for integrating MT into teaching English for professional purposes without lowering standards or encouraging dependency?

To answer these questions, the study uses a compact evaluation model aligned with common quality dimensions. Although the pilot scale is small, the method is replicable in a classroom or institutional context and can inform lesson planning, assessment design, and departmental policy on translation use.

Methods

Design. A comparative pilot design was used to evaluate MT output against human translation output. The goal was not to “prove” a universal winner, but to map strengths and failure modes by genre and communicative purpose.

Materials. Nine short source texts were created for the study (all original, non-copyrighted), grouped into three genres:

• Technical instructions (e.g., device setup, safety steps, troubleshooting)

• Customer-service communication (e.g., complaint response, appointment scheduling, polite requests)

• Promotional/marketing texts (e.g., short ads, product descriptions, service announcements)

Language pair and context. The evaluation model was designed to be applicable to common language pairs used in Central Asian professional education contexts (for example, English to Uzbek or English to Russian), but the rubric itself is language-neutral. The key assumption is that professional texts require more than literal equivalence: they require correct intent, appropriate register, and terminological consistency.

Systems and translators. MT output was produced by a contemporary neural MT engine (any widely used system can be substituted). Human translations were produced by trained bilinguals familiar with professional communication conventions. To reduce bias, all outputs were anonymized and randomly ordered for evaluation.

Evaluation rubric. Two raters scored each translation using a five-point scale across five dimensions: (1) Accuracy, (2) Fluency, (3) Terminology, (4) Register and pragmatics, and (5) Consistency. In addition to numeric ratings, raters tagged errors using a simplified MQM-like typology: meaning error, terminology error, grammar error, style/register error, and omission/addition. Disagreements were resolved through short consensus discussion.

Efficiency measurement. Translators recorded the time required for human translation and for post-editing MT output to a publishable level. Time was used as a proxy for effort; in real settings, cost and workload can be mapped from the same logs.

Analysis. Average scores were calculated by genre and method (MT vs. human). Qualitative notes were summarized to identify typical error patterns and pedagogically relevant insights.

Table 1. Quality dimensions used in the comparative evaluation.

Dimension	**What it checks**	**Typical failure signs**
Accuracy	Meaning, instructions, logical relations, obligation/permission	Wrong requirement, reversed meaning, missing condition
Fluency	Grammar, coherence, readability	Awkward phrasing, agreement errors, broken sentence flow
Terminology	Domain terms and consistency	Wrong term, inconsistent equivalents, mistranslated abbreviations
Register & Pragmatics	Tone, politeness, audience fit, intent	Too rude/too informal, wrong honorifics, misread request vs. order
Consistency	Numbers, units, names, formatting, repeated terms	Unit mismatch, number change, inconsistent formatting

Results

Across the nine texts, human translation achieved higher average quality scores than raw MT in every genre, with the largest gap in customer-service and marketing communication. MT performed best in technical instructions, where sentence structures were predictable and terminology was repeated. In those cases, MT often produced fluent output with correct procedural sequencing, especially when the source text used clear imperative forms and standardized phrasing.

However, MT showed systematic vulnerabilities in pragmatic meaning and audience management. In customer-service texts, small register mistakes changed the perceived politeness and responsibility of the institution. For example, MT tended to overuse direct imperatives where a human translator would choose softer requests, hedging, or empathy markers. In marketing texts, MT frequently preserved literal meaning but failed to recreate persuasive tone, rhythm, and culturally appropriate calls to action.

Terminology errors were not equally distributed. When a term appeared multiple times, MT sometimes switched between synonyms, creating confusion in training contexts where learners are expected to adopt stable equivalents. Human translations were more consistent, especially when a brief term list was provided.

Efficiency results showed a strong speed advantage for MT-assisted workflows. Post-editing MT to publishable quality required less time than full human translation for technical texts, but the advantage shrank for marketing and customer-service texts because post-editing demanded significant rewriting to fix tone and intent. In a classroom setting, this indicates that MT can be time-saving for terminology-focused reading and controlled writing tasks, but it is not a shortcut for pragmatic competence.

Table 2. Average quality scores by genre (1 = poor, 5 = excellent).

Genre	**Machine Translation (raw)**	**Human Translation**	**MT + Post-editing**
Technical instructions	3.8	4.6	4.4
Customer-service texts	3.1	4.7	4.2
Marketing/promotional	2.9	4.5	4.0

Table 3. Typical time patterns (per 200-250 words, indicative).

Genre	**Human translation (min)**	**Post-edit MT (min)**	**Main time driver**
Technical instructions	30-45	15-25	Terminology checks, units, procedural clarity

Customer-service texts	35-55	25-40	Tone, empathy, politeness strategy
Marketing/promotional	40-70	30-60	Rewriting for voice, rhythm, persuasion

Discussion

The results support a practical conclusion: MT is not a single “good or bad” tool; its value depends on genre, risk, and learning goals. For professional education, the question should be framed as alignment: what outcome is required, and what level of reliability is acceptable?

Why MT succeeds in technical genres. Technical instructions often contain constrained vocabulary, repetitive actions, and standardized structures (e.g., “press,” “connect,” “turn off,” “do not operate”). MT benefits from these patterns and can produce near-human fluency. If a term base is provided and the source text is written clearly, MT output can be close to publishable with light post-editing. This is particularly useful in vocational English classes where the goal is comprehension of manuals, safety rules, and operational procedures.

Why MT fails more often in customer-service and marketing. These genres are pragmatic: they carry social meaning beyond the literal sentence. Customer-service communication requires empathy, appropriate apologies, controlled responsibility, and polite problem-solving. Marketing communication requires persuasive voice, audience targeting, and culturally acceptable slogans. MT may produce grammatically correct sentences but miss the “social contract” of the text. In real workplaces, such errors can lead to customer dissatisfaction or brand damage.

Hybrid workflows as a middle path. MT + human post-editing can be highly effective if the task is bounded and the post-editor has clear standards. Professional practice often distinguishes between light post-editing (understandable, fit for internal use) and full post-editing (publishable, client-facing). In education, the same distinction can become a teaching strategy: students learn to decide when “good enough” is acceptable and when it is not.

Pedagogical implications for English teaching in vocational contexts. Teachers can integrate MT without reducing learning quality by using it as an object of analysis. For example, a lesson can start with a short MT output and ask learners to: identify terminology inconsistencies, repair register in customer dialogues, correct numbers and warnings in safety instructions, and justify choices using a rubric.

This turns MT from a shortcut into a diagnostic tool that develops language awareness. Crucially, assessment must be aligned: if students are graded only on final text, they may hide their MT use and learn less. If they are graded on process (drafting, justification, editing logs, and reflection), MT becomes a controlled learning aid.

Risk and ethics. Professional institutions must consider confidentiality. Sending internal documents, patient information, or sensitive contracts to a public MT service may violate policy or law. Even when privacy is not an issue, accountability remains: MT systems do not take responsibility for harmful output. Therefore, high-stakes domains (healthcare, law, safety-critical engineering) should require human review at minimum.

Practical recommendations. Institutions can adopt a simple policy matrix: (1) Low risk + internal use + technical genre: MT allowed, light post-editing recommended. (2) Medium risk + external communication: MT allowed only with full post-editing and rubric-based review. (3) High risk (medical, legal, safety) or strong brand voice: human translation required; MT may be used only for initial drafting under strict controls. Such policies help teachers and students develop professional judgement instead of relying on technology blindly.

Conclusion

This comparative study shows that machine translation and human translation differ less in fluency than in reliability of meaning, terminological stability, and pragmatic appropriateness. MT can be highly useful for technical, pattern-driven texts and for time-saving comprehension activities in vocational English lessons. However, raw MT remains risky for customer-service and marketing genres where tone, empathy, and persuasion are central. Human translation provides stronger accountability and audience adaptation, particularly in high-stakes contexts.

For professional education institutions, the most realistic and educationally productive approach is not to ban MT, but to teach controlled use: define the goal of the translation, apply an explicit rubric, require reflection on editing decisions, and protect confidentiality. When purpose, outcome, and assessment are aligned, MT becomes a tool for developing language and professional competence rather than a replacement for it.

References

1. Papineni K., Roukos S., Ward T., Zhu W.-J. Bleu: a Method for Automatic Evaluation of Machine Translation. Proceedings of ACL, 2002, pp. 311–318.
2. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. NeurIPS, 2017.
3. Koehn P. Statistical Machine Translation. Cambridge University Press, 2009/2010.

4. Lommel A., Uszkoreit H., Burchardt A. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. Tradumàtica, 2014.
5. Lommel A. R., Burchardt A., Uszkoreit H. Multidimensional Quality Metrics: a flexible system for assessing translation quality. (Workshop/Proceedings paper), 2013.
6. Freitag M., et al. Are LLMs Breaking MT Metrics? Results of the WMT24 Metrics Shared Task. Proceedings of WMT, 2024.
7. O'Brien S. Towards Predicting Post-Editing Productivity. Machine Translation, 2011, 25(3), pp. 197–215.
8. O'Brien S. Post-editing (Handbook/entry). John Benjamins (Handbook of Translation Studies Online), 2021.
9. Vieira L. N. Post-Editing of Machine Translation. In: Handbook of Translation and Technology. Routledge, 2019/2020, pp. 319–335.
10. Mariana V., Cox T., Melby A. The MQM Framework: a new framework for translation quality assessment. The Journal of Specialised Translation, 2015, Issue 23.