Artificial Intelligence and the Translation Profession: A Critical Integrative Review of Labour Market Change, Translation Quality, and Professional Reconfiguration (2022 to 2025)

Irina S. Egorova
Higher School of Economics, Moscow, Russian Federation

ABSTRACT

Background. The rapid diffusion of large language models has intensified claims that professional translation is approaching technological redundancy. The available evidence, however, is dispersed across labour economics, computational linguistics, translation studies, industry research, and translator education, and it does not support a uniform conclusion for the profession as a whole.

Purpose. This article examines how artificial intelligence has affected demand for translation labour, identifies the language pairs and communicative domains in which machine-generated translation may support substitution, and analyses how professional practice and translator education are being reorganised in response.

Materials and Methods. A critical integrative review was conducted using studies published from January 2022 through December 2025, together with earlier methodological and theoretical works needed to interpret current developments. Searches covered Scopus, Web of Science, Google Scholar, ACL Anthology, and eLibrary.ru. Bibliographic metadata, publication status, and the fit between cited findings and the claims attributed to them were checked against official journal pages, proceedings records, institutional reports, or working papers. The evidence was appraised according to design fit, outcome validity, and transferability, and was synthesised by distinguishing technological exposure, translation quality, organisational adoption, workflow change, and realised labour outcomes.

Results. Causal and quasi causal labour studies identify measurable pressure in online freelance markets and in United States local labour markets with greater machine translation adoption, but these findings cannot be generalised to all forms of translation employment. Quality studies show strong performance for well resourced language pairs and general content while also documenting persistent variation across languages, linguistic phenomena, document contexts, literary style, specialised domains, and evaluator expertise. The evidence therefore indicates that technical quality is a boundary condition for substitution rather than a sufficient cause of job loss. Professional work is being segmented among routine content with a low consequence of error, expert workflows supported by artificial intelligence, and human led translation where contextual judgement and responsibility remain indispensable. No Russian longitudinal or econometric study comparable to the strongest international labour evidence was identified.

Conclusion. The review develops the concept of conditional substitution. Technical adequacy produces an occupational consequence only when organisations redesign workflows, clients accept the relevant quality threshold, responsibility can be reassigned, and market arrangements permit costs and risks to be transferred to language professionals. Translator education should therefore integrate machine translation and data literacy without weakening advanced language competence, domain knowledge, research ability, revision, ethical judgement, and responsibility for the final text.

KEYWORDS

artificial intelligence; translation profession; machine translation; large language models; labour market; translation quality; postediting; translator education; integrative review

For citation:
Egorova, I. S. (2025). Artificial Intelligence and the Translation Profession: A Critical Integrative Review of Labour Market Change, Translation Quality, and Professional Reconfiguration (2022 to 2025). *Journal of Employment and Career, 4*(4), 27–43. https://doi.org/10.56414/jeac.2025.4.438945

Correspondence:
Irina S. Egorova
irina.egorova@hse.ru

Data Availability Statement:
Not applicable. The review is based on publicly available sources cited in the article.

Received: 08.09.2025
Revised: 21.12.2025
Accepted: 30.12.2025

© Egorova, I. S., 2025

Declaration of Competing Interest:
none declared.

Искусственный интеллект и профессия переводчика: критический интегративный обзор изменений рынка труда, качества перевода и профессиональной реконфигурации (2022–2025)

И. С. Егорова

Национальный исследовательский университет «Высшая школа экономики», г. Москва, Российская Федерация

АННОТАЦИЯ

Введение. Быстрое распространение больших языковых моделей усилило представление о приближающейся технологической избыточности профессии переводчика. Однако имеющиеся данные распределены между экономикой труда, компьютерной лингвистикой, переводоведением, отраслевыми исследованиями и работами по подготовке переводчиков и не позволяют сделать единый вывод в отношении всей профессиональной сферы.

Цель. Статья направлена на установление того, как искусственный интеллект повлиял на спрос на переводческий труд, в каких языковых парах и коммуникативных областях качество машинного перевода создает условия для замещения человека, а также каким образом меняются профессиональная практика и подготовка переводчиков.

Материалы и методы. Проведен критический интегративный обзор исследований, опубликованных с января 2022 по декабрь 2025 г., с привлечением более ранних методологических и теоретических работ, необходимых для интерпретации современных изменений. Поиск выполнялся в Scopus, Web of Science, Google Scholar, ACL Anthology и eLibrary.ru. Каждое содержательное утверждение проверялось по официальной странице журнала, материалам конференции, институциональному отчету или рабочей статье. Доказательства оценивались с учетом соответствия исследовательского дизайна заявленному выводу, валидности показателя и переносимости результата. При синтезе разграничивались технологическая подверженность, качество перевода, организационное внедрение, изменение рабочего процесса и фактические последствия для занятости.

Результаты. Исследования с причинным и квазипричинным дизайном выявляют измеримое давление на занятость и доходы в цифровом фрилансе и на локальных рынках труда США с более широким использованием машинного перевода, однако эти результаты нельзя переносить на все формы переводческой занятости. Исследования качества подтверждают высокий уровень перевода для ресурсообеспеченных языковых пар и текстов общего назначения, но одновременно фиксируют устойчивую зависимость результата от языка, лингвистического явления, объема контекста, литературного стиля, предметной области и квалификации эксперта. Следовательно, техническое качество определяет границы возможного замещения, но не является достаточной причиной сокращения рабочих мест. Профессиональная деятельность разделяется на работу с рутинным контентом, где цена ошибки невысока, экспертные процессы с поддержкой искусственного интеллекта и переводы, в которых контекстуальное решение и профессиональная ответственность сохраняют определяющее значение. Российских продольных или эконометрических исследований, сопоставимых с наиболее сильными зарубежными данными о рынке труда, не выявлено.

Для цитирования:
Егорова, И. С. (2025). Искусственный интеллект и профессия переводчика: критический интегративный обзор изменений рынка труда, качества перевода и профессиональной реконфигурации (2022–2025). *Журнал работа и карьера, 4*(4), 27–43. https://doi.org/10.56414/jeac.2025.4.438945

Корреспонденция:
Ирина Сергеевна Егорова
irina.egorova@hse.ru

Заявление о доступности данных: не применимо. Обзор основан на общедоступных источниках, приведенных в списке литературы.

Поступила: 08.09.2025
Поступила после рецензирования: 21.12.2025
Принята к публикации: 30.12.2025

© Егорова И. С., 2025

Конфликт интересов: автор сообщает об отсутствии конфликта интересов.

Заключение. Для объяснения полученных результатов предложено понятие условного замещения. Техническая пригодность системы преобразуется в трудовой эффект только тогда, когда организация меняет рабочий процесс, заказчик принимает соответствующий порог качества, ответственность может быть перераспределена, а устройство рынка позволяет перенести издержки и риски на специалиста. Поэтому подготовка переводчиков должна включать грамотное использование машинного перевода и данных, не ослабляя углубленную языковую подготовку, предметную специализацию, исследовательские навыки, редактирование, этическую оценку и ответственность за итоговый текст.

КЛЮЧЕВЫЕ СЛОВА

искусственный интеллект; профессия переводчика; машинный перевод; большие языковые модели; рынок труда; качество перевода; постредактирование; подготовка переводчиков; интегративный обзор

INTRODUCTION

Debate about artificial intelligence and translation has often been organised around a question that is rhetorically powerful but analytically weak: will machines replace translators? The question compresses several different processes into a single imagined event. A system may be able to generate fluent target language text without being reliable for a particular communicative purpose. A client may adopt automated translation without removing human work, because the workflow still requires terminology management, verification, revision, documentation, and final responsibility. Conversely, a client may reduce paid human involvement even when output quality remains uneven, if price, speed, or procurement practice carries more weight than linguistic adequacy. A credible account of professional change must therefore separate technical capability, task level adequacy, organisational adoption, workflow allocation, and labour market outcome. The present review begins from this distinction because it prevents capability demonstrations from being treated as employment evidence and prevents professional anxiety from being dismissed merely because aggregate employment has not collapsed.

Translation is particularly suitable for examining this distinction because it combines unusually high technological exposure with unusually heterogeneous work. The occupational category includes platform freelancers producing short commercial texts, staff translators working within regulated institutions, literary translators responsible for voice and interpretation, localisation specialists operating in integrated content systems, and revisers who validate texts produced by humans or machines. The same technology can affect these groups in different ways because the relevant quality threshold, consequence of error, client relationship, and allocation of legal or professional responsibility differ. Eloundou et al. (2024) identify translation among occupations with high exposure to large language model capabilities, but exposure measures describe the share of tasks that a technology may affect, not the number of jobs that will disappear. This limitation is not incidental. It identifies the central inferential problem of the literature: evidence about what a system can do must be connected, rather than equated, with evidence about what organisations and markets actually do.

Recent evidence relevant to occupational change falls into three main domains. Labour studies examine realised changes in employment and earnings under different conditions of technological exposure and adoption. Industry surveys describe reported changes in workflows, rates, post-editing, and professional expectations, while translation-quality research identifies the language, genre, contextual, and evaluative conditions under which automated output may or may not be adequate for professional use. Professional studies complement these strands by examining workflow allocation, responsibility, trust, and working conditions. Because these designs answer different questions, technical capability, reported experience, and realised labour effects must be synthesised without treating one as a proxy for another.

The literature remains fragmented because each disciplinary tradition observes only part of this process. Labour economics measures employment and earnings but seldom models translation quality in domain specific terms. Computational linguistics compares systems but rarely traces how quality estimates enter procurement and workplace decisions. Translation studies analyses professional experience and competence but often relies on small samples or cross sectional perceptions that cannot determine causal labour effects. A critical integrative review is therefore required not simply to collect findings, but to identify which inference each design can support and how the findings can be connected without exceeding their evidentiary limits. The review addresses three questions: how artificial intelligence has measurably affected demand for translator labour; how translation quality defines the realistic boundary of substitution; and how professional practice and translator education are being reconfigured. Its principal contribution is the concept of conditional substitution, developed through the synthesis rather than assumed in advance.

REVIEW DESIGN AND ANALYTICAL FRAMEWORK

Scope and Review Logic

The study uses a critical integrative review design. Integrative reviewing is appropriate when a problem cannot be answered through a single research design because the relevant evidence includes experiments, econometric studies, surveys, qualitative interviews, technical evaluations, standards, and conceptual work (Whittemore & Knafl, 2005). The method is not treated here as permission for an impressionistic narrative. Following the logic proposed by Torraco (2005) and Snyder (2019), the review seeks to produce an explanatory synthesis by comparing assumptions, units of analysis, supported inferences, and limitations across research traditions. A meta analysis was not appropriate because the outcomes are not commensurable: employment coefficients, automatic translation scores, human error judgements, survey attitudes, and educational outcomes cannot be combined into a meaningful common effect size. A scoping review would map the field effectively, but the present problem also requires critical appraisal and resolution of apparent contradictions, which extends beyond descriptive mapping (Munn et al., 2018).

The main temporal window extends from January 2022 to December 2025 because this period captures the public diffusion of generative large language models and the most consequential recent changes in translation workflows. Earlier sources were retained when they supplied a methodological foundation, a labour theory required for interpretation, or evidence on neural machine translation and professional practice that directly shaped the post 2022 debate. The review is deliberately problem centred rather than technology centred. Studies were included when they addressed at least one link between system capability and occupational consequence: realised labour outcomes, professional market conditions, comparative translation quality, work organisation, competence, or education. Purely technical papers were used only when their evaluation design permitted a defensible inference about professional translation. Vendor claims, market forecasts without transparent methods, commentary based solely on anecdote, and studies that treated fluency as a complete measure of translation quality were excluded from the analytical core.

Search, Verification, and Selection of Evidence

Searches were conducted iteratively in Scopus, Web of Science, Google Scholar, ACL Anthology, eLibrary.ru, CyberLeninka, and official publisher or institutional repositories. The final search update was conducted on September 30, 2025. English search combinations connected *artificial intelligence*, *large language models*, *machine translation*, *translation quality*, *employment*, *earnings*, *professional practice*, *postediting*, *competence*, and *translator education*. Russian searches combined *искусственный интеллект*, *машинный перевод*, *профессия переводчика*, *рынок труда*, *качество перевода*, *постредактирование*, *компетенции*, and *обучение переводчиков*. Backward and forward citation searching was used around the strongest empirical studies and recent specialised issues. Because the review was repeatedly updated as new evidence became available during 2025, it is not presented as a closed PRISMA style systematic review with a fixed search date and a reproducible count of every record encountered. Claiming such a protocol retrospectively would create procedural precision that the review process did not possess. Transparency is instead provided through explicit inclusion logic, source verification, evidence tables, and careful restriction of each conclusion to the design that supports it.

For every source used to support a substantive empirical claim, bibliographic metadata, publication status, and the fit between the reported finding and the attributed claim were checked against an official journal page, proceedings record, report, or institutional working paper. Particular attention was given to publication status and to the difference between peer reviewed articles, conference papers, preprints, working papers, and industry reports. This step led to the removal of several records included in the earlier manuscript because their bibliographic details or claimed findings could not be confirmed. It also changed the treatment of several valid sources. Eloundou et al. (2024), for example, are cited as evidence of task exposure, not displacement. Brynjolfsson et al. (2025) are used to demonstrate that generative assistance may raise productivity in a different occupation, not to infer a translator employment effect. ELIS reports are interpreted as structured evidence of reported experience and expectation, with their own explicit warning about representativeness, rather than as objective measures of market size or causal impact. This verification is essential because a review about a rapidly changing technology can easily become more confident than its evidence.

Appraisal and Synthesis

The evidence was appraised through three questions. The first concerned design fit: whether the method could answer the claim attributed to it. The second concerned outcome validity: whether the measure represented exposure, technical quality, adoption, work organisation, employment, earnings, or professional judgement. The third concerned transferability: whether findings from a particular platform, language pair, genre, country, or educational cohort could be generalised beyond that setting. This approach does not reduce heterogeneous studies to a single quality score. It makes their inferential boundaries visible. Table 1 summarises the principal evidence types used in the review and clarifies what each can and cannot establish. The distinction is subsequently maintained throughout the findings and discussion, so that evidence becomes cumulative without becoming artificially homogeneous.

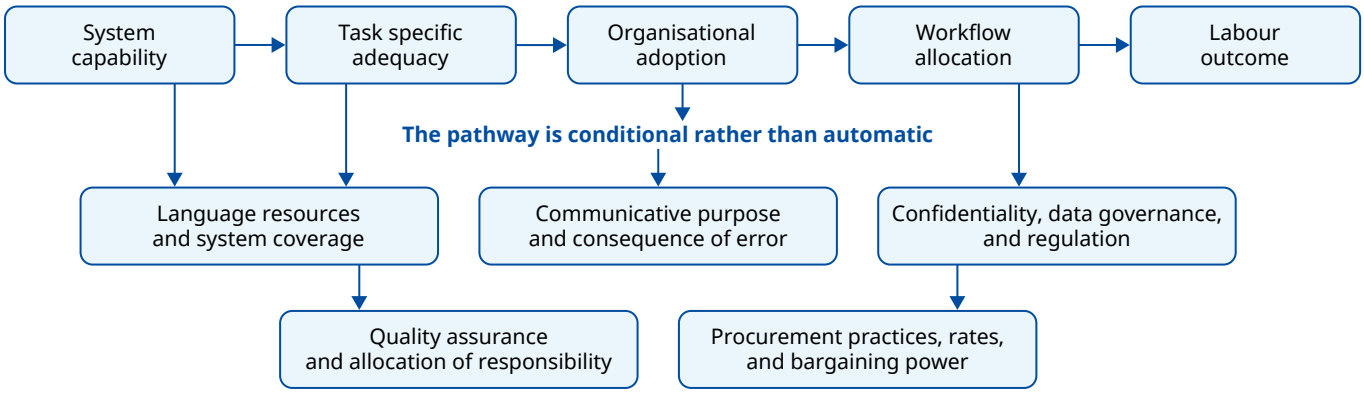
Table 1

Evidence Types and Their Inferential Boundaries

Evidence type	Typical unit of analysis	Supported inference	Inference not supported by the design alone
Occupational exposure models	Tasks within occupations	Which tasks appear technically susceptible to current systems	Actual job loss, wage change, adoption, or long term occupational decline
Causal or quasi causal labour studies	Workers, contracts, platforms, or local labour markets	Changes in employment, earnings, or demand associated with technology adoption in the studied setting	Uniform effects across all translation segments, countries, and forms of employment
Industry surveys	Respondents from companies, institutions, education, and freelance work	Reported conditions, practices, expectations, stress, and perceived change	Independent market totals or causal attribution to artificial intelligence
Translation quality evaluations	Segments, documents, systems, language pairs, and evaluators	Relative performance under the specified task, metric, and evaluation conditions	General human parity or direct labour market consequences
Professional and educational studies	Translators, trainees, teachers, projects, and curricula	How work, competence, trust, and learning are experienced or organised	Population wide employment effects or universal curricular outcomes

Figure 1

From Technical Capability to Occupational Consequence



The synthesis proceeded by tracing a pathway from technical capability to occupational consequence and by identifying the conditions that interrupt or accelerate movement along that pathway. Figure 1 represents this analytical architecture. It does not posit a deterministic sequence in which better technology inevitably removes labour. Instead, each transition is treated as an empirical question. Adequate output for one task may not satisfy the quality requirement of another. Organisational adoption may increase professional oversight rather than eliminate it. Workflow redesign may reduce the volume of drafting while increasing revision, terminology, or risk management. A labour outcome emerges only after these technical and organisational changes interact with procurement, rates, bargaining power, confidentiality, regulation, and the consequence of error. This architecture makes it possible to explain why strong technical progress, local displacement, professional adaptation, and continued demand for expert translators can coexist.

Labour Market Change: From Exposure to Realised Effects

Exposure and Productivity Do Not Determine Employment Outcomes

The labour literature becomes coherent once occupational exposure is separated from realised substitution. Exposure studies estimate whether current or foreseeable systems can perform or accelerate tasks contained in an occupation. Eloundou et al. (2024) place translators among highly exposed occupations because language models can affect a large share of text transformation tasks. This is important evidence of technological relevance, but it does not reveal whether clients adopt the systems, whether human review remains necessary, or whether productivity gains are captured through higher output, lower prices, reduced labour, or a different service mix. Autor et al. (2003) and Autor (2015) provide the broader theoretical

reason for caution. Technologies alter task composition rather than eliminating occupations in a one to one fashion, and automation can coexist with new complementary work. Translation is likely to follow this task based pattern, but the balance between automation and complementarity must be established empirically rather than inferred from exposure alone.

Evidence from other knowledge intensive occupations demonstrates why productivity cannot be read directly as displacement. Brynjolfsson et al. (2025) studied 5,172 customer support agents during the staggered introduction of a generative assistant and found an average productivity increase of 15 per cent, with larger gains among less experienced workers. The study shows that generative systems can diffuse stronger patterns of performance to workers who have less accumulated task knowledge. It does not show that translation behaves identically, but it identifies a plausible mechanism for professional stratification. If language technology reduces the performance gap between novice and experienced producers for routine content, clients may become less willing to pay for undifferentiated experience while continuing to value expertise that concerns responsibility, specialised knowledge, and difficult judgement. These findings support a hypothesis that generative assistance may alter the value attached to undifferentiated experience and reshape early career progression before any aggregate employment effect becomes visible. Whether this mechanism operates in translation requires direct empirical testing.

Evidence from Online Labour Markets and Local Labour Markets

Hui et al. (2024) provide the strongest direct evidence of short term labour pressure associated with generative artificial intelligence in an online labour market. Their quasi experimental analysis compares freelancers in occupations with different levels of exposure after the public release of generative systems. Employment and earnings declined more in highly exposed categories, and high prior performance did not provide complete protection. The design is valuable because it observes realised transactions rather than intentions, but its setting also defines its boundary. Platform work has low switching costs, granular contracts, transparent price competition, and limited institutional commitment between client and worker. These conditions make rapid substitution more feasible than in organisations where translation is embedded in security procedures, legal accountability, subject specific knowledge, or long term relationships. The study should therefore be interpreted as evidence that displacement can occur quickly under low friction market conditions, not as an estimate for the entire profession.

Frey and Llanos-Paredes (2025) extend the analysis beyond a single platform by examining machine translation adoption across 695 local labour markets in the United States after the release of Google Translate as a mobile application. They report a negative relationship between adoption and translator employment, supported by an instrumental variable strategy and placebo regressions, and also identify reduced demand for several foreign language skills outside the translation occupation. This study is especially important because it concerns dedicated machine translation rather than only the generative systems that became widely visible after 2022. It suggests that the current disruption has a longer technological history and that the demand effect may extend from professional translation to the broader economic value assigned to foreign language tasks. The result remains a relationship estimated within United States labour markets and does not identify which translation domains or employment arrangements account for the change. It nevertheless establishes that technical adoption can be associated with a measurable reduction in labour demand beyond platform contracting.

The available evidence therefore supports a qualified conclusion. Artificial intelligence and machine translation have already produced negative labour effects in identifiable settings, but the magnitude and distribution of those effects depend on how translation is purchased and governed. Table 2 places the main labour sources in a common inferential frame. Read together, they show a progression from potential exposure to realised pressure, while also demonstrating why no single study can settle the future of the profession. The evidence is strongest for segments where output is modular, quality requirements can be standardised, and human responsibility is weakly institutionalised. It is less direct for staff translators, regulated domains, interpreting, transcreation, and long form literary work.

Industry Surveys and the Distribution of Pressure

Repeated ELIS reports give the labour studies professional and organisational context. The 2025 survey reported that machine translation was used in more than half of professional translation work by both language companies and independent professionals (ELIS Research, 2025). Among independent professionals, 24% of projects were pre-translated by clients, predominantly language companies; 69% reported using machine translation on their own initiative, and 29% used it in at least half of their projects (ELIS Research, 2025). These patterns support the proposition that market intermediation matters. Large providers are better positioned to standardise workflows, impose discounts, and separate production from quality assurance. Direct clients and public bodies may retain more varied expectations or place greater weight on continuity and responsibility. The survey thus reveals not a single adoption curve, but an institutional distribution of automation.

Table 2

Central Evidence on Labour Market Change and Industry Conditions

Source	Design and context	Supported finding	Principal limitation
Eloundou et al. (2024)	Task based exposure assessment using occupational task data	Translation belongs to the group of occupations with high technical exposure to large language model capabilities	Exposure is not adoption, displacement, or an observed labour outcome
Hui et al. (2024)	Quasi experimental analysis of an online freelance labour market after the release of generative systems	Freelancers in more exposed occupations experienced relative reductions in employment and earnings	One platform and a short observation period; translation specific estimates are not isolated
Frey and Llanos-Paredes (2025)	Adoption of Google Translate across 695 United States local labour markets, with instrumental variable and placebo analyses	Greater adoption is negatively related to translator employment and to demand for several foreign language skills	Working paper; national context; limited visibility of affected domains and work arrangements
Brynjolfsson et al. (2025)	Staggered introduction of a generative assistant among 5,172 customer support agents	Generative assistance increased productivity, with larger benefits for less experienced workers	Different occupation; productivity effects do not determine employment or translation quality
ELIS Research (2024, 2025)	Repeated European surveys of companies, independent professionals, institutions, universities, and students	Respondents report deteriorating sentiment, increasing automated translation use, extensive postediting, and price pressure	Self selected perceptions and expectations; not an objective or globally representative market measure
Humlum and Vestergaard (2025)	Representative Danish surveys linked to administrative context	Adoption is uneven and shaped by workplace conditions and training	Evidence concerns general occupational use, not translator specific displacement

The same report records a serious deterioration in confidence among independent professionals, including concern about financial sustainability, time pressure, reduced personal contact, rigid conditions, and the replacement of translation by postediting. Only a small minority of independent respondents identified positive effects of artificial intelligence on their own activity, while companies were more divided because some also associated the technology with productivity and new service development (ELIS Research, 2025). These findings cannot prove that artificial intelligence caused every reported business difficulty. The report itself warns that responses reflect market mood and that the sample is not globally representative. They nevertheless show how technological change is experienced within a market already characterised by outsourcing and unequal bargaining power. This professional experience is part of the occupational effect because a profession can be weakened through rate compression, reduced autonomy, and deteriorating entry conditions even when the number of people performing language work does not immediately fall.

The distributional interpretation resolves the apparent contrast between local negative effects and the absence of a universal collapse. Platform freelancers and suppliers working through large intermediaries face rapid conversion of automated output into procurement rules and discounts. Institutional translators may experience technology first as workflow redesign, productivity expectation, or internal pressure to justify the department. Specialists in high responsibility domains may retain demand while spending more time validating machine generated material or defending the conditions under which automation is inappropriate. The central labour question is consequently not whether translation employment rises or falls in the aggregate, but which actors can decide the workflow, define acceptable quality, and retain the productivity gain. This shift from occupation totals to the distribution of control provides the bridge from labour evidence to translation quality.

The Russian Speaking Translation Market: An Evidentiary Gap

A targeted search of Russian and international databases identified studies of machine translation quality, literary translation, competence, and education, but did not identify a longitudinal, platform based, administrative, or econometric study of translator employment and earnings in the Russian speaking market comparable to Hui et al. (2024) or Frey and Llanos-Paredes (2025). This statement concerns the results of the search rather than a claim that no relevant data exist anywhere. Some information may remain in commercial platforms, professional associations, tax records, vacancy databases, or internal organisational datasets that are not publicly accessible or indexed. The absence of retrievable evidence is nevertheless consequential because it prevents verification of claims about the scale, timing, and distribution of labour change.

Direct extrapolation from Western evidence would be methodologically weak. The Russian speaking market has its own mix of state institutions, industrial translation, education, publishing, freelance platforms, domestic technology providers, procurement practices, and cross border constraints. Language technology adoption may also differ by direction, domain, access to systems, confidentiality requirements, and the availability of Russian language data. A negative effect found on a United States platform cannot be assumed to have the same magnitude in Russian legal translation, scientific communication, audiovisual localisation, or regional freelance work. Nor can European survey sentiment represent the Russian-speaking market: ELIS 2025 included only nine respondents from Russia, including one language company, five independent professionals, two training-institute respondents, and one student (ELIS Research, 2025).

A credible evidence base would connect several forms of data rather than relying on a single opinion survey. Repeated surveys of translators and employers could establish changes in rates, volumes, task composition, and working time. Vacancy and platform data could identify shifts in demand for translation, postediting, localisation, language technology, and domain expertise. Organisational case studies could show how ministries, universities, publishers, technology firms, and industrial companies allocate responsibility when adopting automated translation. Quality studies should examine Russian in specialised and high responsibility domains, not only general or literary examples. These strands would allow researchers to determine whether apparent decline reflects reduced demand, hidden transfer of translation to other employees, changed job titles, movement into adjacent services, or deterioration in remuneration without numerical job loss.

Until such evidence is produced, professional and educational recommendations for the Russian speaking context must remain explicitly provisional. International research can identify mechanisms and methodological models, but it cannot supply local effect sizes. This boundary should not be presented as a weakness that can be overcome through confident interpretation. It is a substantive finding of the review. The lack of labour data makes it impossible to know whether the segments most exposed internationally are also the segments changing most rapidly in Russia, and it limits the ability of universities to align curricula with verified rather than presumed labour demand.

Translation Quality as a Boundary Condition for Substitution

Quality is a Relation between a Text and its Purpose

Translation quality is often invoked as though it were a stable property that can be measured independently of use. Professional translation does not permit such a simplification. A text may be adequate for information access and unacceptable for publication, legally dangerous despite fluent wording, or terminologically accurate while pragmatically unsuitable for its audience. Quality therefore concerns fitness for a specified communicative purpose under an explicit consequence of error. This principle also explains why an average comparison between machine and human translation is of limited value. Human translations vary in quality, machine systems vary across languages and prompts, and the evaluation method may privilege surface fluency over meaning, voice, or document coherence. International standards for translation and postediting define process requirements and competence more readily than they define a universal output threshold, which reflects the contextual nature of the judgement (International Organization for Standardization, 2015, 2017).

Evaluation research has repeatedly shown that the unit and method of assessment alter the conclusion. Läubli et al. (2018) demonstrated that claims of human parity can weaken when evaluators assess documents rather than isolated sentences. Freitag et al. (2021) showed that expert evaluation and document context materially affect comparative judgement. Zhang et al. (2025) extend this point to literary translation by showing that complex schemes and insufficient evaluator expertise can misclassify published human translations, while simpler comparative procedures used by appropriate evaluators distinguish them more reliably. Their corpus includes more than 2,000 translations and over 13,000 evaluated sentences across four language pairs, and their overall evaluation finds that published human translations consistently outperform large language model translations, which are more literal and less diverse. The study does not establish a universal hierarchy for every genre, but it shows that the evaluator is part of the measurement instrument.

Language Resources and Linguistic Phenomena

The most stable finding across recent technical evaluations is resource asymmetry. Robinson et al. (2023) evaluate ChatGPT based translation across 204 languages and find that performance is competitive for some high resource languages but weak for many low resource languages. Jiao et al. (2023) and Hendy et al. (2023) similarly report stronger results for widely represented language pairs and less reliable performance as linguistic distance, data scarcity, or evaluation complexity increases. This pattern has direct occupational significance because a technology that performs well for English centred commercial communication may leave other language markets with a different balance of costs, risks, and professional need. It also warns against treating progress in a small group of benchmark languages as evidence about multilingual communication in general.

Fine grained evaluation further shows that aggregate scores conceal systematic weaknesses. Manakhimova et al. (2023)

compare large language models and neural machine translation on linguistically motivated phenomena in English, German, and Russian. Performance varies by phenomenon and direction, and no system family is uniformly superior. A model may produce highly fluent text while mishandling tense, punctuation, omission, named entities, or a construction that matters disproportionately in the target domain. For professional work, such patterns are often more important than a small difference in a global metric because errors are not equally consequential. The Russian results in this study are valuable, but they do not provide a general evaluation of Russian legal, medical, scientific, or literary translation. They establish that Russian belongs within contemporary comparative evaluation while simultaneously exposing the shortage of domain specific evidence.

Context, Genre, and the Persistence of Critical Error

Large language models possess an important advantage over earlier sentence based systems because they can process broader context and respond to detailed instructions. Castilho et al. (2023) show that context handling differs across online machine translation systems and a generative model, while Chen et al. (2024) demonstrate that iterative refinement can improve translations. These capabilities make the systems more useful, but they also relocate part of the work into instruction, selection, and verification. Prompting does not eliminate professional judgement; it changes the point at which that judgement enters the process. A user must recognise what context is relevant, formulate constraints, evaluate alternatives, and detect when fluent output violates the source or communicative situation. The system can assist these operations, but it cannot assume responsibility for defining the purpose against which its own output should be judged.

Literary translation makes this dependence particularly visible. Karpinska and Iyyer (2023) find that large language models can exploit document context, yet critical errors persist in literary translation. Zhang et al. (2025), using more than 2,000 translations and over 13,000 evaluated sentences across four language pairs, show that published human translations consistently outperform LLM translations when evaluation is conducted with appropriate methods and expertise; LLM outputs are also more literal and less diverse. Russian case-study evidence likewise compares expert, ChatGPT, Google, and Yandex translations of literary fragments and concludes that cognitive and creative components of literary translation remain a major obstacle to full automation (Norets & Reynova, 2024). Case studies cannot define the performance of all models or genres, but their convergence with larger evaluations undermines the claim that fluent prose constitutes autonomous literary translation.

Table 3

Selected evidence on translation quality and its professional interpretation

Source	Languages or domain	Main result relevant to professional use	Interpretive boundary
Robinson et al. (2023)	204 languages	Large language model translation is competitive for some high resource languages but weak for many low resource languages	Results demonstrate resource asymmetry, not a universal ranking of systems
Manakhimova et al. (2023)	English, German, and Russian; linguistically motivated test suite	Relative performance changes across linguistic phenomena and language directions	Aggregate scores can conceal systematic errors that matter in specific domains
Castilho et al. (2023)	Online systems and a generative model; contextual phenomena	Systems differ in their ability to use context	Context sensitivity is task dependent and does not establish final fitness for purpose
Chen et al. (2024)	Iterative refinement with large language models	Repeated prompting and feedback can improve output	Improvement depends on instructions and evaluation, which introduces expert work
Karpinska & Iyyer (2023)	Literary translation with document context	Document context helps, but critical errors persist	The finding concerns literary material and selected language pairs
Zhang et al. (2025)	Literary corpus, four language pairs, human and automated evaluation	Published human translations consistently outperform large language model versions under appropriate evaluation	Evaluation design and evaluator expertise strongly affect the result
Norets & Reynova (2024)	Russian literary case study	Artificial intelligence can assist drafting but does not reproduce the full interpretive and stylistic work of literary translation	Small case study; cannot support profession wide conclusions

Specialised translation presents a related problem through responsibility rather than aesthetic voice. Legal and institutional translation requires jurisdictional interpretation, terminological consistency, traceable sources, confidentiality, and accountable revision. Prieto Ramos (2024) argues that artificial intelligence changes the configuration of competence in these domains but does not remove the need for legal and institutional expertise. A similar mechanism may be relevant in medical, technical, and other safety-critical communication, but this extension requires domain-specific evidence. A system may achieve high average adequacy while remaining unsuitable for unsupervised use because a rare error can carry disproportionate consequences. In such settings, improved technical performance may increase adoption and simultaneously increase the need for qualified validation. Quality therefore defines the form of human involvement, not merely its presence or absence.

Why Quality Does Not Directly Determine Employment

The quality evidence defines where substitution is technically plausible, but it does not determine whether substitution will occur. Clients may accept lower quality for ephemeral or internal content because speed and cost dominate. They may require human translation for reputational reasons even when machine output performs well. Organisations may use automated drafts while retaining the same number of specialists because the volume of multilingual content expands. They may also reduce external purchasing while shifting hidden review work to bilingual employees whose labour is not recorded as translation. The relationship between quality and employment is therefore mediated by organisational choice and by the visibility of quality costs. Errors that are detected only after publication, litigation, user harm, or reputational damage may not enter procurement calculations at the moment of adoption.

This mediation explains why technical evaluations should not be interpreted as labour forecasts. Better output expands the set of tasks for which substitution can be considered, but the final decision depends on communicative responsibility, risk tolerance, confidentiality, workflow integration, and bargaining power. Figure 2 combines technical adequacy with communicative responsibility to distinguish four workflow regions. The lower right region contains routine, low consequence content for which system performance is adequate and direct substitution pressure is highest. The upper right region contains work in which systems can contribute substantially but qualified professionals remain responsible for validation and sign off. The left side contains tasks where poor performance preserves human production, with high responsibility further strengthening the need for specialist translation and independent review. The matrix is not a permanent taxonomy because models and domains change. Its value lies in showing that higher technical adequacy moves work horizontally, while responsibility determines whether that movement leads to autonomous automation or to AI assisted expertise.

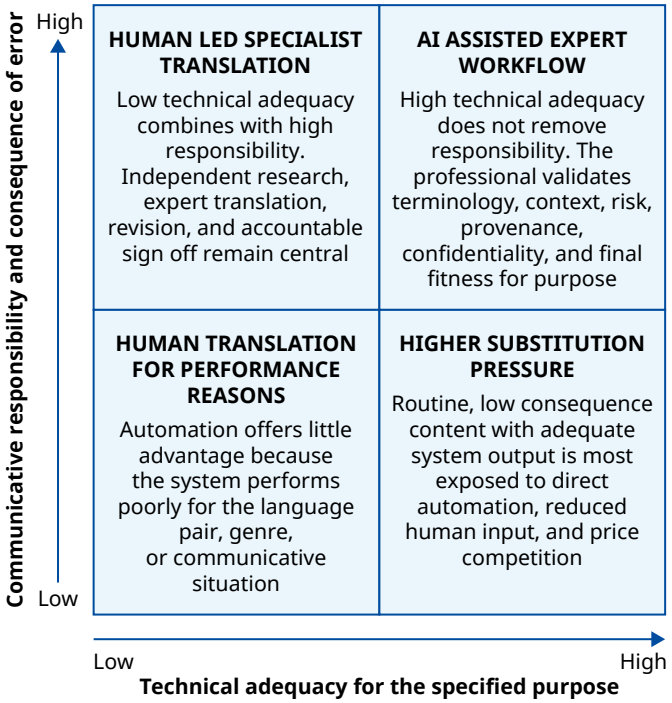
Professional Reconfiguration: Work, Responsibility, and Conditions

From Production to Decision and Validation

The professional response to artificial intelligence is often described as a transition from translation to postediting. This description captures an important increase in revision of machine generated output, but it is too narrow to explain the emerging division of labour. Postediting begins after a system and workflow have already been selected, a source has been prepared, data have been transmitted, a quality threshold has been specified, and responsibility has been allocated. Professional work increasingly enters at each of these earlier stages. Translators assess whether automation is appropriate, determine terminology and reference materials, protect confidential data, identify domain risks, decide how much context the system requires, and document unresolved uncertainty. The final text remains only one output of a larger decision process.

Figure 2

Translation workflow according to technical adequacy and communicative responsibility



Source: developed by the author on the basis of the reviewed evidence

Table 4

Evidence on professional practice, working conditions, competence, and education

Source	Focus and method	Contribution to the synthesis	Caution
Kirov & Malamin (2022)	Survey of translators concerning artificial intelligence and professional change	Concern involves market entry, price pressure, and work structure as well as output quality	Attitudinal and cross sectional evidence
Fırat et al. (2024)	Platform translators examined through a decent work framework	Adaptation is shaped by remuneration, security, and control, not only by technical skill	Context specific and not an employment effect estimate
Rico Pérez (2024)	Critical reconsideration of postediting guidelines	Light and full postediting categories are insufficient for contemporary ecosystems	Conceptual contribution rather than outcome evaluation
Nitzke et al. (2024)	Interview study of decisions in machine translation and postediting projects	Adoption depends on project value, risk, quality expectations, and working conditions	Small qualitative sample suited to mechanism, not prevalence
Rivas Ginel & Moorkens (2025)	Mixed method study of translators trust and distrust in generative artificial intelligence	Trust is connected to quality, privacy, job security, and governance	Reported attitudes do not directly measure market effects
Guerberof-Arenas et al. (2024)	Experiment with eleven translation students before and after postediting training	Training affected confidence and willingness to attempt creative shifts, but did not show a simple quantitative creativity gain	Very small educational sample
Prieto Ramos (2024)	Competence analysis in legal and institutional translation	Artificial intelligence increases the importance of domain specific judgement, quality control, and accountability	Focused on high responsibility institutional domains
Evteev et al. (2025)	Survey of 99 teachers and 373 students in Russian translator education	Supports systematic teaching of machine translation and postediting within professional preparation	Single institutional context and self reported views

Rico Pérez (2024) argues that the traditional distinction between light and full postediting has become inadequate because contemporary systems and projects do not fit a simple continuum from rough comprehension to publishable quality. An ecosystem approach is needed in which human factors, organisational conditions, quality expectations, and technology are considered together. Nitzke et al. (2024), drawing on interviews with participants in machine translation and postediting projects, similarly show that adoption decisions are strategic and contextual. Risk, value, client requirements, and working conditions shape the workflow. These studies support a broader conception of professional competence in which the translator is not merely the final corrector of a machine draft, but the person capable of connecting technological possibilities to communicative obligations.

This broader role also clarifies why advanced translation competence remains necessary. Machine output is often most difficult to evaluate when it is fluent, because plausibility reduces the perceptual signal that normally prompts closer checking. A professional must compare source and target meanings, trace terminology, recognise unsupported additions, detect omissions, evaluate register, and understand the consequences of a subtle error in the relevant domain. These operations require language competence, subject knowledge, research skill, and revision experience. Artificial intelligence may reduce the time required for initial drafting while increasing the importance of diagnostic reading. The value of expertise consequently shifts from visible text production toward the less visible control of uncertainty.

Trust, Autonomy, and Working Conditions

Professional distrust of generative systems should not be interpreted automatically as resistance to innovation. Rivas Ginel and Moorkens (2025) find that translators' trust is shaped by observed limitations, data privacy, professional security, and the conditions under which systems are introduced. Kirov and Malamin (2022), writing before the current wave of generative adoption, likewise show that concern is directed not only at output quality but at the entry of non specialists, price competition, and changes in market structure. These findings locate trust within professional experience. A system may be technically useful and institutionally untrustworthy if its training data, confidentiality, error profile, or contractual consequences are opaque.

Working conditions determine whether technological assistance becomes professional augmentation or work intensification. Fırat et al. (2024) use the International Labour Organization's decent work perspective to examine translators in the platform economy and show how economic insecurity, remuneration, and control shape the meaning of adaptation. ELIS 2025 reports stronger price pressure, lower work volumes, and a generalised shift from human translation to discounted and less-valued post-editing among independent professionals; 53% questioned their future as freelance professionals and 23% were considering ending freelance activity (ELIS Research, 2025). The same report shows that language companies were more divided in their assessment of artificial intelligence: up to

half of the companies implementing AI reported positive effects through efficiency gains, lower costs, or new service offerings, while 28% reported no significant impact (ELIS Research, 2025). When a client expects human level quality while procurement assumes a substantial productivity discount, the translator absorbs the mismatch through unpaid checking, faster work, or increased liability. The relevant issue is therefore not post-editing as a technique itself, but the potential mismatch between the remuneration assigned to the task and the level of responsibility attached to the final text.

Research on human factors reinforces this interpretation. Ragni and Vieira (2022) review changes associated with neural machine translation and show that the effects on effort, agency, and experience cannot be inferred from output quality alone. Vieira (2020) situates automation anxiety within longer standing concerns about business practices and remuneration. Moorkens et al. (2024) propose evaluating translation automation through a triple bottom line that includes people, performance, and environmental consequences. Together these contributions shift the normative question from whether professionals should use artificial intelligence to the conditions under which its use is sustainable. A workflow that reduces drafting time but transfers risk downward, obscures authorship, or makes skilled labour economically unviable cannot be judged successful solely because it increases throughput.

Professional Identity and Accountable Expertise

The transformation of professional identity is therefore more complex than a change in tools. Translation has traditionally combined authorship, mediation, service, and expert responsibility. Automated drafting weakens the visibility of authorship but intensifies mediation and responsibility, particularly when the professional must explain why a fluent system output is unsafe or inappropriate. In this review, the term accountable language expertise is used to denote a professional role in which language and domain competence are combined with responsibility for workflow selection, evidence evaluation, uncertainty management, and the final communicative outcome. It includes the capacity to justify workflow selection, evaluate evidence, communicate uncertainty, and accept or refuse final responsibility. Such expertise is valuable precisely because the system cannot define the social purpose of the translation or bear the consequences of failure.

The profession may nevertheless become more segmented. Routine production can move toward autonomous systems or low cost review, while specialists concentrate in high consequence domains and in functions that combine language, technology, quality governance, and client consultation. This segmentation creates an entry problem. Junior translators have traditionally developed expertise through substantial volumes of comparatively routine work. If that work is automated, new entrants may encounter fewer paid opportunities in which to build domain knowledge and revision judgement. The productivity mechanism identified by Brynjolfsson et al. (2025) may help novices perform some tasks, but assistance is not equivalent to a professional development pathway. A sustainable profession must therefore address how expertise is formed when technology removes or transforms the work through which expertise was previously acquired.

Translator Education after Generative Artificial Intelligence

Technological Literacy Cannot Replace Translation Competence

Translator education faces a false choice when technological integration is presented as an alternative to traditional competence. Students require direct experience with machine translation and generative systems because professional ignorance would leave them unable to evaluate tools, negotiate workflows, or recognise data risks. They also require advanced command of source and target languages, genre, subject matter, research, terminology, revision, and ethical responsibility because these are the resources through which system output becomes assessable. The European Master's in Translation Competence Framework[1] 2022 already treats technology as one component within a wider professional architecture rather than as the organising centre of the curriculum (European Commission, 2022). The rapid change in tools strengthens rather than weakens that design principle. Software specific instruction dates quickly, while the ability to formulate a translation purpose, investigate a domain, evaluate evidence, and justify a decision remains transferable.

Machine translation literacy provides a useful educational concept because it concerns informed and critical use rather than operational familiarity. Bowker (2020) shows that such literacy is relevant beyond professional translator programmes because users need to understand both the capabilities and limitations of machine translation. Mellinger (2017) identifies knowledge and skill gaps that arise when translator pedagogy does not engage adequately with machine translation. Hackenbuchner and Krüger (2023) extend this perspective through data literacy, drawing attention to the origin, selection, preparation, ownership, and quality of data used in machine translation systems. Generative artificial intelligence adds further requirements: students must assess confidentiality, recognise that a fluent response can be unsupported, preserve traceability, and distinguish assistance from unacknowledged delegation of responsibility.

[1] European Master's in Translation (EMT) Competence Framework 2022. https://commission.europa.eu/system/files/2022-11/emt_competence_fwk_2022_en.pdf

Postediting Should be Taught as Diagnosis and Decision

A narrow course in postediting can inadvertently train students to accept the machine draft as the default representation of the source. Effective instruction must instead teach diagnostic comparison and workflow choice. Students should learn when a draft is worth editing, when translation from scratch is more efficient, how error patterns differ across systems and language pairs, and how the required quality threshold changes the amount of checking. Latorraca (2023) shows that trainees' perceptions of their translation and postediting performance do not necessarily correspond to actual difficulty, which makes reflective calibration an educational objective. Guerberof-Arenas et al. (2024) find no simple quantitative evidence that postediting training improves creativity, although students became more willing to attempt creative shifts and more confident in addressing machine translation problems. The result suggests that education should not promise automatic gains from tool exposure. It should create repeated opportunities for comparison, explanation, and independently reviewed judgement.

Assessment practices must also change. A final target text alone cannot reveal whether the student selected an appropriate workflow, verified terminology, protected data, or corrected an initially plausible error. Process evidence is therefore required. Students can document sources, prompts, system versions, rejected alternatives, quality criteria, and reasons for human intervention. Such documentation serves an analytical and pedagogical function: it makes the student's decision process visible and supports assessment of whether the final text was produced through an appropriate and accountable workflow. It develops the professional ability to make reasoning auditable without reducing translation to a mechanical checklist. It also discourages the uncritical substitution of generated prose for learning, because students remain accountable for demonstrating how the final text was produced and why it is fit for purpose.

Russian Translator Education and the Need for Evaluated Curricular Change

Russian scholarship has begun to address the educational consequences of machine translation and artificial intelligence, but the available evidence remains concentrated in conceptual proposals, institutional surveys, and case based analysis. Evteev et al. (2025) report responses from 99 teachers and 373 students and argue for systematic inclusion of machine translation, postediting, and critical evaluation in translator preparation. Zubanova and Brager (2023) discuss the renewal of translator education in relation to professional requirements, while Norets and Reynova (2024) provide material that can support critical literary translation exercises. These studies demonstrate active curricular concern and offer locally relevant examples. They do not yet establish which instructional designs improve graduate performance, employment, or long term professional judgement.

The next stage should therefore move from curricular advocacy to comparative evaluation. Programmes can examine whether students who receive integrated instruction in translation, machine translation literacy, data governance, and revision perform better on authentic tasks than students who receive tool demonstrations without a broader competence framework. Outcomes should include error detection, decision quality, source use, time, confidence calibration, and the ability to explain when automation is inappropriate. Graduate follow up is equally important because immediate classroom performance may not predict workplace adaptation. Russian education will be best served not by copying a rapidly changing international tool list, but by developing evidence about how technology can be incorporated without weakening the linguistic and intellectual basis of professional responsibility.

DISCUSSION

Conditional Substitution as the Central Synthesis

The combined evidence supports neither technological determinism nor a reassuring narrative of uncomplicated augmentation. It supports conditional substitution. This concept describes a process in which artificial intelligence reduces or reorganises paid human translation only when several conditions align. The system must produce output that is adequate for the specified purpose. An organisation or client must adopt it and redesign the workflow. The residual human tasks must be separable, measurable, and purchasable under new conditions. Responsibility for error must be accepted, transferred, or obscured. Finally, the market must allow productivity gains, quality risks, and adjustment costs to be distributed in a way that makes reduced human input economically attractive. A failure at any point can preserve human translation, create AI assisted expert work, or prevent adoption altogether.

This synthesis explains the uneven findings across disciplines. High exposure scores identify the first condition, technical reach. Quality evaluations refine the second condition by showing that adequacy differs across languages, genres, and evaluators. Labour studies observe the final outcome in markets where adoption and workflow redesign have already occurred. Professional studies reveal the intermediate mechanisms, including discounts, postediting, trust, control, and responsibility. The disciplines do not contradict one another when their units of analysis are kept distinct. They describe different stages of the same process. The apparent disagreement arises when a result at one stage is treated as proof of another, such as when high benchmark performance is converted into a job forecast or when poor market sentiment is treated as a causal estimate.

Conditional substitution also clarifies why quality is necessary but insufficient. Poor output can block automation,

particularly where errors are visible and costly. Adequate output only makes automation possible. Whether it becomes substitution depends on procurement and governance. A client purchasing internal gist translation may accept a lower threshold than a court, publisher, regulator, or patient information provider. A large language company may standardise postediting and negotiate discounts more readily than a direct client with a long standing specialist. An institution may adopt generative systems but retain expert translators because the institution, not the system, remains accountable. The same technical performance can therefore generate different labour outcomes because quality is interpreted through organisational responsibility.

Implications for the Profession and for Policy

For professional practice, the findings suggest that adaptation cannot be reduced either to a shift toward creative tasks or to the acquisition of prompt-writing skills. Creative domains are not automatically protected, and prompt competence without domain judgement is fragile because interfaces and models change quickly. Durable professional value lies in combining language expertise with a clearly defined area of responsibility. Translators who can evaluate specialised meaning, manage terminology, protect data, document sources, revise at document level, and explain risk occupy a different market position from providers of undifferentiated text production. This does not guarantee protection from price pressure, but it gives clients and institutions a defensible reason to retain qualified human control.

For organisations, the evidence requires a more complete accounting of productivity. Time saved in drafting should be measured together with time spent preparing data, prompting, checking, correcting, documenting, and resolving failures. Quality should be evaluated at the level of the intended document and purpose, with independent review in high consequence contexts. Procurement should specify who is responsible for the final text and should avoid imposing a discount based on assumed machine productivity before the actual error profile and review burden are known. Where a workflow expects human level quality, the professional must receive the time, information, and authority required to deliver it. Otherwise automation creates an unrecorded quality debt that appears only when errors escape.

For public and educational policy, the main risk is responding to uncertainty with either prohibition or premature substitution. Blanket bans prevent the development of informed professional use, while uncritical adoption can erode training, confidentiality, and accountability. Policy should require transparent disclosure of automated assistance where it affects authorship, risk, or data processing; protect confidential material; preserve qualified human responsibility in high consequence communication; and support collection of labour data that distinguishes translation from postediting and adjacent language services. These measures do not assume that human translation is always superior. They recognise that institutions need evidence about when a system is adequate and a responsible actor when it is not.

Directions for Empirical Research

The next generation of labour research should connect employment data to the actual content of translation work. Platform and vacancy studies can identify whether declines are concentrated in particular genres, language pairs, experience levels, or client categories. Administrative and longitudinal surveys can follow individuals as they move between translation, postediting, localisation, quality assurance, language data work, and non language employment. Such designs would reveal whether apparent occupational decline represents unemployment, reduced earnings, changed job titles, or movement into adjacent tasks. They should also examine who captures productivity gains and whether technology narrows or widens differences between junior and senior professionals.

Quality research should move from generic system ranking toward purpose specific risk. Evaluations need document level context, professional evaluators, transparent prompts and system versions, and error weighting that reflects the consequences of failure. Russian and other underrepresented languages require domain specific studies in legal, medical, scientific, technical, public service, and literary communication. Longitudinal evaluation is also necessary because model updates can improve one phenomenon while introducing new patterns of instability. Research that combines quality data with observed editing time, confidence calibration, and final responsibility would connect computational performance more directly to professional workflow.

Educational research should test coherent competence models rather than isolated tool workshops. Comparative interventions can examine whether integrated teaching of translation analysis, machine translation literacy, data literacy, postediting, and professional ethics improves authentic performance. The most informative outcomes will include error detection, independent judgement, source verification, workflow selection, and the ability to justify a refusal to automate. Graduate follow up can then determine whether these capabilities improve employment and reduce vulnerability to low value platform work. Such research would transform the current educational literature from a persuasive call for reform into evidence about which reforms actually work.

Limitations

The review integrates heterogeneous evidence, and that strength creates a limit. Labour coefficients, industry perceptions, quality evaluations, interviews, and educatio-

nal studies do not measure the same object. The synthesis therefore depends on conceptual comparison rather than statistical pooling. Although sources were verified and appraised through design fit, outcome validity, and transferability, the appraisal does not produce a standardised risk of bias score across all designs. The concept of conditional substitution should consequently be understood as an explanatory framework supported by converging evidence, not as a quantitatively estimated model. Its individual pathways require further testing with data that connect system use, workflow change, quality, and employment within the same organisations or markets.

The technological environment also changes more rapidly than conventional publication. Studies frequently evaluate a named system version that may be replaced before the article appears, while proprietary systems provide incomplete information about training data and updates. This review addresses that problem by emphasising recurring conditions, such as resource asymmetry, context dependence, evaluator expertise, and responsibility, rather than treating a temporary benchmark ranking as permanent. Even so, some quality findings will date quickly. The review includes peer reviewed research, conference proceedings, preprints, working papers, and industry reports because no single publication type covers the whole problem. Publication status is identified in the references and the argument gives greater weight to transparent designs, but the evidence base remains uneven.

Geographic coverage is another important limitation. The strongest labour evidence concerns the United States, online platforms, Denmark, and European industry respondents. The targeted search did not identify comparable Russian labour studies, and the available Russian literature is stronger on education and textual analysis than on employment. This imbalance prevents a direct estimate of change in the Russian speaking profession and limits global generalisation. The review also concentrates on written translation. Interpreting, audiovisual accessibility, subtitling, and multimodal localisation interact with artificial intelligence in related but distinct ways and require separate analysis. These boundaries are not reasons to withhold conclusions. They specify where the conclusions apply and where further evidence is required.

CONCLUSION

This review set out to determine what can be established, rather than merely predicted, about artificial intelligence and the translation profession. The first research question concerned labour market effects. The evidence shows that negative effects are already measurable in particular settings. Online freelancers in more exposed occupations have experienced relative reductions in employment and earnings, and greater machine translation adoption across United States local labour markets is negatively related to translator employment. These results are serious, but they do not describe every segment of translation. They are strongest where work is modular, procurement is price sensitive, and organisational barriers to substitution are low. Industry surveys add evidence of widespread postediting, deteriorating confidence, and pressure on rates and working conditions, while also requiring cautious interpretation because they measure reported experience rather than causal market change.

The second question concerned comparative quality and the realistic scope of substitution. Artificial intelligence now produces useful and sometimes strong translations for many well resourced language pairs and general purposes. It also remains uneven across languages, linguistic phenomena, document contexts, specialised domains, and literary tasks. Evaluation method and evaluator expertise can change the apparent result, which makes general claims of human parity methodologically unsound. Quality is therefore a boundary condition. It determines where substitution can be considered, but organisational adoption, responsibility, risk tolerance, and procurement determine whether it occurs. This distinction is the foundation of conditional substitution and the principal theoretical contribution of the review.

The third question concerned professional and educational adaptation. Translation work is not simply becoming postediting. It is being reorganised around workflow selection, verification, terminology, risk, data governance, revision, documentation, and final responsibility. In routine low consequence content, human input may contract substantially. In high responsibility settings, the same technological improvement can produce an AI assisted expert workflow in which professional judgement becomes more rather than less important. Translator education must prepare graduates for both conditions. It should integrate machine translation and data literacy while protecting the advanced language, research, domain, writing, revision, and ethical competences required to recognise when fluent output is wrong and to defend a responsible professional decision.

The future configuration of translation work will depend less on an abstract comparison between human and machine performance than on concrete decisions about automation, acceptable quality, workflow allocation, distribution of productivity gains, and responsibility for communicative failure. Research and policy should make those choices visible. For the Russian speaking market, that task begins with producing reliable longitudinal evidence on employment, earnings, workflow, and quality rather than importing conclusions from foreign settings. A profession can adapt to powerful technology, but adaptation is professionally sustainable only when expertise, responsibility, and the conditions of skilled work are preserved together.

REFERENCES

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics, 118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Bowker, L. (2020). Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship, 25*(1-2), 25–43. https://doi.org/10.1080/08963568.2020.1794739

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Castilho, S., Mallon, C. Q., Meister, R., & Yue, S. (2023). Do online machine translation systems care for context? What about a GPT model? In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 393–417). European Association for Machine Translation. https://aclanthology.org/2023.eamt-1.39/

Chen, P., Guo, Z., Haddow, B., & Heafield, K. (2024). Iterative translation refinement with large language models. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation* (Volume 1) (pp. 181–190). European Association for Machine Translation. https://aclanthology.org/2024.eamt-1.17/

ELIS Research. (2024). *European Language Industry Survey 2024: Trends, expectations and concerns of the European language industry*. https://elis-survey.org/wp-content/uploads/2024/03/ELIS-2024-Report.pdf

ELIS Research. (2025). *European Language Industry Survey 2025: Trends, expectations and concerns of the European language industry*. https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science, 384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

European Commission. (2022). *European Master's in Translation* (*EMT*) *competence framework 2022*. https://commission.europa.eu/system/files/2022-11/emt_competence_fwk_2022_en.pdf

Evteev, S. V., Voevoda, E. V., & Goncharova, N. A. (2025). Machine translation in translator training. *Periodical Scientific and Methodological Electronic Journal "Concept", 7*, 162–175. (In Russ.). https://doi.org/10.24412/2304-120X-2025-11136

Fırat, G., Gough, J., & Moorkens, J. (2024). Translators in the platform economy: A decent work perspective. *Perspectives, 32*(3), 422–440. https://doi.org/10.1080/0907676X.2024.2323213

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics, 9*, 1460–1474. https://doi.org/10.1162/tacl_a_00437

Frey, C. B., & Llanos-Paredes, P. (2025). Lost in translation: Artificial intelligence and the demand for foreign language skills [Working paper]. Oxford Martin School, University of Oxford. https://www.oxfordmartin.ox.ac.uk/publications/lost-in-translation-artificial-intelligence-and-the-demand-for-foreign-language-skills

Guerberof-Arenas, A., Valdez, S., & Dorst, A. G. (2024). Does training in post-editing affect creativity? *The Journal of Specialised Translation, 41*, 74–97. https://doi.org/10.26034/cm.jostrans.2024.4712

Hackenbuchner, J., & Krüger, R. (2023). DataLitMT - Teaching data literacy in the context of machine translation literacy. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 285–293). European Association for Machine Translation. https://aclanthology.org/2023.eamt-1.28/

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H. (2023). *How good are GPT models at machine translation? A comprehensive evaluation*. arXiv. https://doi.org/10.48550/arXiv.2302.09210

Hui, X., Reshef, O., & Zhou, L. (2024). The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science, 35*(6), 1977–1989. https://doi.org/10.1287/orsc.2023.18441

Humlum, A., & Vestergaard, E. (2025). The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences*, *122*(1), e2414972121. https://doi.org/10.1073/pnas.2414972121

International Organization for Standardization. (2015). *Translation services – Requirements for translation services* (ISO Standard No. 17100:2015). https://www.iso.org/standard/59149.html

International Organization for Standardization. (2017). *Translation services - Post-editing of machine translation output – Requirements* (ISO Standard No. 18587:2017). https://www.iso.org/standard/62970.html

Jiao, W., Wang, W., Huang, J.-T., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator*? *Yes with GPT-4 as the engine*. arXiv. https://doi.org/10.48550/arXiv.2301.08745

Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 419–451). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.41

Kirov, V., & Malamin, B. (2022). Are translators afraid of artificial intelligence? *Societies, 12*(2), 70. https://doi.org/10.3390/soc12020070

Latorraca, R. (2023). Lost in post-editing: An exploratory study on translation trainees' perceived EN to IT post-editing versus translation performance. *Ampersand, 11*, 100144. https://doi.org/10.1016/j.amper.2023.100144

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4791–4796). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1512

Manakhimova, S., Avramidis, E., Macketanz, V., Lapshinova-Koltunski, E., Bagdasarov, S., & Möller, S. (2023). *Linguistically motivated evaluation of the 2023 state-of-the-art machine translation: Can ChatGPT outperform neural machine translation?* In *Proceedings of the Eighth Conference on Machine Translation* (pp. 224–245). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.23

Mellinger, C. D. (2017). Translators and machine translation: Knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer, 11*(4), 280–293. https://doi.org/10.1080/1750399X.2017.1359760

Moorkens, J., Castilho, S., Gaspari, F., Toral, A., & Popović, M. (2024). Proposal for a triple bottom line for translation automation and sustainability: An editorial position paper. *The Journal of Specialised Translation, 41*, 2–25. https://doi.org/10.26034/cm.jostrans.2024.4706

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*, 143. https://doi.org/10.1186/s12874-018-0611-x

Nitzke, J., Canfora, C., Hansen-Schirra, S., & Kapnas, D. (2024). Decisions in projects using machine translation and post-editing: An interview study. *The Journal of Specialised Translation, 41*, 127–148. https://doi.org/10.26034/cm.jostrans.2024.4715

Norets, M. V., & Reynova, A. V. (2024). The use of artificial intelligence in the translation of a literary text (on the example of the novel "Fire and Blood" by George R. R. Martin). Scientific Notes of V. I. Vernadsky Crimean Federal University. *Philological Sciences*, *10*(3), 149-158. (In Russ.). https://doi.org/10.29039/2413-1679-2024-10-3-149-158

Prieto Ramos, F. (2024). Revisiting translator competence in the age of artificial intelligence: The case of legal and institutional translation. *The Interpreter and Translator Trainer, 18*(2), 148–173. https://doi.org/10.1080/1750399X.2024.2344942

Ragni, V., & Vieira, L. N. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives, 30*(1), 137–158. https://doi.org/10.1080/0907676X.2021.1889005

Rico Pérez, C. (2024). Re-thinking machine translation post-editing guidelines. *The Journal of Specialised Translation, 41*, 26–47. https://doi.org/10.26034/cm.jostrans.2024.4696

Rivas Ginel, M. I., & Moorkens, J. (2025). Translators' trust and distrust in the times of GenAI. *Translation Studies, 18*(2), 283–299. https://doi.org/10.1080/14781700.2025.2507594

Robinson, N., Ogayo, P., Mortensen, D. R., & Neubig, G. (2023). *ChatGPT MT: Competitive for high-resource languages, but not low-resource languages*. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 392–418). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.40

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Vieira, L. N. (2020). Automation anxiety and translators. *Translation Studies, 13*(1), 1–21. https://doi.org/10.1080/14781700.2018.1543613

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Zhang, R., Zhao, W., & Eger, S. (2025). How good are LLMs for literary translation, really? *Literary translation evaluation with humans and LLMs*. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 10961–10988). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.548

Zubanova, S. G., & Brager, A. A. (2023). On the issue of training translators at the present stage: Domestic traditions and the challenges of education renewal (case study of an aviation institute). *Society: Sociology, Psychology, Pedagogics, 5*, 99–108. (In Russ.). https://doi.org/10.24158/spp.2023.5.13