AI and Translation: Human Creativity, Machine Output, and Ownership.

Mounir KHEDAR.

- University of Mohamed BOUDIAF; M'sila (Algeria).
- Email*: mounir@kheddar@univ-msila.dz*

Received:02/02/2026; Accepted:25/04/2026; Published:15/06/2026.

Abstract:

The rapid rise of AI and neural machine translation has transformed translation, raising legal and ethical issues about authorship and intellectual property. It highlights the conflict between traditional copyright laws and machine involvement, examines translation as a derivative work, and addresses rights attribution and the use of human translation data in training systems. To achieve these goals, the study adopts an interdisciplinary analytical framework combining doctrinal legal analysis, conceptual analysis, and a comparative perspective. The doctrinal legal method is used to examine existing copyright laws, legal doctrines, and regulatory frameworks governing authorship, ownership, and intellectual property in the context of artificial intelligence and machine translation. In parallel, conceptual analysis is employed to clarify key notions such as originality, creativity, and the legal status of AI-generated and hybrid translations. Furthermore, the study incorporates a comparative dimension by referring to selected international and national legal developments, including institutional guidelines and regulatory texts. Legal documents and institutional reports are analyzed through a critical interpretive method, focusing on how these sources conceptualize authorship, ownership, and liability in AI-related contexts. This methodological approach enables the

identification of normative gaps, inconsistencies, and emerging legal challenges within existing legal framework.

Keywords: Artificial intelligence; machine translation; translation rights; intellectual property; authorship.

Introduction:

Translation, as a central linguistic, cultural, and intellectual activity, has undergone a profound transformation in recent decades due to the accelerated development of artificial intelligence (AI) technologies and machine learning applications. Translation is no longer limited to the human act of understanding, interpreting, and reconstructing meaning; machines have become essential actors in producing translated texts, particularly with the emergence of neural machine translation systems capable of simulating human performance with advanced accuracy and harmony.

However, this technological transformation has not been neutral in its effects. It has generated deep legal and ethical challenges, most notably concerning authorship, the ownership of meaning, and exploitation rights in contexts where translations—and sometimes the original texts themselves—are produced with minimal or no direct human intervention. This reality raises fundamental questions about the role of the translator, the limits of human creativity, and the legitimacy of using human translations as training data for intelligent systems.

The problem addressed in this study revolves around the tension between traditional copyright systems, which are founded on human centrality, and a new technical reality in which AI systems actively participate in producing translated texts and the meanings they convey. With fully or partially automated translations, critical questions emerge: Can machine translation outputs be considered protected intellectual works? Who holds the rights to use them—the programmer, the operator, or the user—or are such rights inherently absent due to the lack of human creativity? These questions are further complicated when human translations,

translation memories, and bilingual corpora are used to train AI systems without clear legal guidance, raising concerns about potential violations of the literary and economic rights of translators.

In light of these challenges, this article seeks to address the following research questions:

1. How is translation legally classified as a derivative intellectual work under the use of AI technologies?

2. What is the legal status of translations produced wholly or partly by machine translation systems?

3. Who can be considered the author or owner of translation rights in the absence of direct human intervention?

4. How legitimate is the use of human translations and translation memories in training AI systems?

5. What are the ethical and cultural implications of converting human translations into training data?

6. To what extent can existing legal frameworks ensure a fair balance between protecting rights and promoting technological innovation?

Accordingly, the objectives of this study are to:

- Analyze the legal framework governing translation rights as derivative intellectual works in the context of AI;
- Highlight emerging problems related to authorship and ownership of meaning in machine translation;
- Discuss the position and role of the human translator under full and partial machine translation;
- Examine legal and ethical issues associated with using human translations and translation memories to train AI systems;
- Contribute to the academic debate on reformulating intellectual property concepts in light of contemporary technological transformations.

This study derives its significance from addressing a highly current topic at the intersection of law, translation, and artificial intelligence. It highlights the challenges faced by

translators amid digital transformation, contributes to the debate on protecting intellectual property in a rapidly evolving technological environment, and opens theoretical horizons for rethinking authorship, ownership, and the nature of meaning in the digital age. To achieve these goals, the study adopts an analytical-descriptive approach, examining international and national legal texts, comparing different legislative positions, and drawing on theoretical literature in Translation Studies and AI ethics to clarify the legal and ethical dimensions of these emerging challenges.

2. Translation as a Derivative Work: Protecting Rights in the AI Era:

According to international legal frameworks, translation constitutes a form of transformation of the original work, which requires prior authorization from its author. The Berne Convention for the Protection of Literary and Artistic Works (1886) enshrined this principle in its second article, paragraph 3, stating that: "Translations, adaptations, arrangements of music and other alterations of a literary or artistic work shall be protected as original works without prejudice to the copyright in the original work."

Accordingly, the right to authorize the translation of a work remains vested solely in the original author, who holds the legal authority to permit or prohibit the use of their text in any translation process. This principle is explicitly reaffirmed in the Berne Convention, Article VIII, under the heading "Right of Translation" , which states: "The authors of literary and artistic works protected by this Agreement have the exclusive right to translate their works or to allow their translation for the duration of the established protection of the original works."

National legislation similarly enshrines the principle of exclusive rights to license the translation of literary works while simultaneously recognizing the independence of the translation as a new intellectual creation. For instance, in France, Article L112-3 of the Intellectual Property Code

provides that authors of translations, quotations, and adaptations enjoy legal protection, provided that such protection does not prejudice the rights of the original author.

This legal framework demonstrates how French law balances the protection of derivative creativity with the preservation of the literary and economic rights of the first author: "Authors of translations, adaptations, transformations, or arrangements of works of the mind shall enjoy the protection established by this Code without prejudice to the rights of the author of the original work."

Comparative legislation categorizes translations among a series of legal transformations of intellectual works, alongside revisions, updates, commentaries, abstracts, excerpts, and musical arrangements. These forms of derivative transfer confer compensable financial rights on their owners, either as a percentage of proceeds or as a lump-sum remuneration.

In this context, Morkens and Lewis emphasize that the exploitation of a translation requires obtaining a license from the rights holder, and that the original author may not use the translation as the basis for a subsequent translation without prior consent. They further underline that translation rights remain enforceable even for works that have fallen into the public domain, reflecting the strict protection granted to translators as creators of derivative works:

The translation can't be used without authorization from the translation copyright owner (…) the original author may not use the translation as the basis for a further translation without permission, and the translation copyright owner retains rights even over translation of a text that is out of copyright." (Moorkens & Lewis, 2020, p. 472)

This principle stems from the recognition of translation as an independent intellectual work. Although it belongs to the category of derivative or dependent works, the translation requires prior authorization from the original rights holder—except in cases where the law permits use without permission,

such as works that have entered the public domain or legislative and official texts. Nevertheless, the translation itself remains legally protected as an independent creation, and the translator, as its author, is granted full literary and economic rights over the text produced.

3. Authorship and Ownership in Fully Automated Machine Translation:

The development of machine translation (MT), particularly through the integration of artificial intelligence (AI), has profoundly transformed the translation landscape. The advent of neural networks and machine learning techniques has significantly enhanced the quality of translations. In this context, Morkens and Lewis observe that:

«The application of neural networks to translation means that Machine Translation (MT) is now considered an application of machine learning. Define machine learning as ' an automated process that extracts patterns from data '. . . the application of machine learning to translation has led to a leap forward in MT quality and claims that MT is approaching human parity (Moorkens & Lewis, 2020 : p 469)

This technological evolution has led to two central issues requiring careful consideration. The first concerns the ownership of translations generated by automated systems, while the second addresses the use of these translations as data to train and improve machine translation algorithms, raising questions about authorship, rights, and ethical use. Translation memories (TMs) may belong to the translator, the language service provider managing the process, or the translator who produced the texts using the tool.

Morkens and Lewis further emphasize the ethical dimension, questioning the legitimacy of sharing these memories beyond their original purpose or reusing them in different contexts or projects: "ownership and sharing of translation data became a point of discussion among translators and translation buyers. For example, Topping (2000) discussed whether T[ranslation[M[emorie]s should

belong to the end client, the L[anguage]S[ervice]P[rovider], or the translator (as owner of the TM tool); whether data sharing was ethical, and whether bespoke TMs created for one purpose would be useful for another " . They are naturally referring to the data generated by the act of translating, which are useful for the training of intelligent translation systems, especially those equipped with neural networks, as neural MT (NMT)" (Moorkens & Lewis, 2020, p. 472)

These issues extend beyond professional contracts, touching upon fundamental legal and intellectual questions regarding the nature and ownership of translation in the digital age . As AI-based translation tools advance, it becomes necessary to re-examine the legal frameworks governing this activity. Translation, as a derivative intellectual work, occupies a delicate intersection between human creativity and technical reproduction. International conventions and national legislation recognize the translator's inherent rights, provided the literary and economic rights of the original author are respected. This framework establishes a careful balance between protecting the original work and valuing the translator's creative contribution.

However, this balance is increasingly challenged by machine translation systems capable of producing outputs approaching human quality. These developments raise unprecedented legal questions concerning the ownership of machine-generated texts, their authenticity, and the identity of the author in the absence of direct human involvement. Moreover, the use of human and machine translations to train AI models introduces additional concerns regarding secondary rights, the protection of linguistic data, and equitable distribution of the value derived from such data. There is, therefore, a pressing need to redefine authorship, ownership, and legitimate use to ensure fairness across the linguistic value chain—from the original author to the translator and the AI system developer.

Despite technological advances, translation remains fundamentally a human activity, as it relies on the interpretation, understanding, and reconstruction of meaning within multidimensional cultural contexts. Fully automated machine translation, however, raises several critical questions:

Who owns the legal rights to the resulting text?

Who controls its exploitation?

Should creativity be attributed to the human operator providing instructions, the system developer, or does the text lack a traditional author altogether?

This dual challenge threatens to erase the human creative footprint while simultaneously undermining classical concepts of intellectual property and copyright. Ethically, AI translation can distort meaning by neglecting the cultural and cognitive contexts of the original text, producing interpretations that may misrepresent the author's intent.

The European Parliament resolution of October 20, 2020, addressing intellectual property in the era of AI, underscores the need to distinguish between two types of creativity:

1. Works produced by humans with AI as a supportive tool.

2. Works generated entirely by AI without direct human intervention.

In paragraph 14 of the resolution, the Parliament noted that fully automated creations present novel legal challenges regarding authorship, ownership, compensation, and the concentration of AI-driven knowledge within corporate monopolies, potentially disrupting the balance between human and machine contributions:"AI-assisted human creations and AI-generated creations, with the latter creating new regulatory challenges for IPR protection, such as questions of ownership, inventorship and appropriate remuneration, as well as issues related to potential market concentration".

Translations produced entirely by machine, without human involvement in drafting or stylistic decisions, clearly fall into this second category of pure automated creativity , raising fundamental legal questions regarding authorship and rightful ownership. While human translators continue to play a role in current practices, the growing sophistication of AI systems necessitates a re-examination of authorship and ownership , potentially redefining these concepts to align with new technical realities.

Consider the 2021 case where a major European publishing house commissioned a generative AI system to translate a contemporary novel. The AI was trained on previous translations and public-domain texts, producing a full draft. A human editor made minor stylistic adjustments. A dispute arose over ownership between the AI developers, the publishing house operating the system, and the human editor. In the UK, under CDPA Article 9.3, ownership could be granted to the operator (publishing house), whereas in other jurisdictions, direct human involvement might be necessary to claim rights, potentially recognizing the human editor as the holder of literary rights. This example illustrates the real legal ambiguity surrounding fully automated MT and highlights the importance of clear authorship and ownership frameworks.

Some researchers argue that property rights should be assigned to a human legal entity responsible for overseeing the process, not as the author of the text, but as the legal owner of the rights resulting from system use. For instance, UK law (CDPA, Article 9.3) considers the operator of a computer system to be the "author" in cases of computer-generated works: "In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken."

The key legal question is: who is closest to the "product of the machine" and thus entitled to claim rights? Fernand

Carballo-Calero identifies four potential candidates for IP rights in fully machine-generated works:

1. System programmer – creator of the software and algorithms.
2. Actual user – the individual who operates the system and directs output.
3. The program itself – legally rejected due to lack of legal personality.
4. No one – the work is considered authorless and enters the public domain.

System programmer: The programmer designs the algorithms and controls the program's functioning, providing the structural basis for the work. However, many scholars argue that while the programmer enables the process, they do not participate in the creative act itself and therefore should not be considered the "author" under traditional copyright principles.

Actual user: The user operates the system, providing instructions, controlling context, and potentially editing output. While closer to authorship, the user's creative contribution varies, and often they act merely as an operator, making legal attribution of authorship controversial.

The program itself: Although it executes the creative act, the program cannot hold legal rights, as it lacks legal personality.

No one : Fully automated works may enter the public domain due to the absence of human creativity, yet this approach risks reducing incentives to develop AI technologies.

In the specific field of computer-generated works, the potential candidates are four: (1) the author of the program; (2) the user of the program; (3) the program; and (4) none (Lacruz Mantecón, 2023, p. 79).

The current legal debate centers on the duality between the programmer and the actual user: determining who is the rightful author or owner of AI-generated content. For

instance, consider a case where an international publishing house contracts a company to translate a contemporary novel using a generative AI system. The company trains the system with previous translations and public-domain texts, produces a full draft, and a human editor makes minor stylistic adjustments. A dispute arises over ownership among:

1. The programmers (designers of the system).
2. The company (operator and investor in the translation process).
3. The human reviewer (minor linguistic adjustments).

Under certain jurisdictions (e.g., UK CDPA 9.3), ownership may be granted to the operator company, whereas in others, a direct human contribution is required, potentially recognizing the human reviewer or literary editor as the holder of the literary rights to the final version.

4. Human Translation as Data: Legal and Ethical Implications for AI Training

Modern machine translation (MT) systems rely heavily on vast sets of linguistic data during the training and learning phase. Most commonly, these datasets take the form of bilingual corpora , which consist of sentences in a source language paired with their corresponding translations in a target language.

These collections are constructed through a careful process of sentence alignment , ensuring that each original sentence matches its correct translation. This alignment allows the system to learn automatically how to generate coherent and accurate sentences in the target language based on their counterparts in the source language. Through such parallel data , intelligent models can identify linguistic patterns and semantic relationships across languages, significantly enhancing their ability to simulate human translation.

Bilingual data can also be exploited indirectly, for instance, to improve traditional translation systems or to develop models capable of functioning with minimal

dependence on parallel corpora, via self-learning or transfer learning technologies:

According to Bardet (2021), training automatic translation systems requires the use of data corpora. As he explains:

"Pour l'apprentissage de systèmes de traduction automatique, nous avons besoin de corpus de données. Ils peuvent prendre plusieurs formes. Généralement, ils sont bilingues […] De cette façon, le système peut apprendre à générer, à partir d'une phrase source, la phrase cible qui est sa traduction. Il est aussi possible d'exploiter des données monolingues, avec des techniques pour améliorer un système classique ou en essayant directement de traduire sans données parallèles." (Bardet, 2021, p. 10).

This means that for training purposes, corpora can be bilingual—composed of a source and a target language—and aligned sentence by sentence so that the system learns to generate the correct target sentence from a source sentence. Monolingual data can also be used to enhance traditional systems or to attempt translation without parallel corpora.

However, this mechanism, which positions human translation as the cornerstone of machine learning , raises complex legal and ethical questions regarding the ownership and source of these data. Most parallel corpora used for training large-scale AI models derive from years of human translation work, whether produced by institutions, academic projects, or publicly available content (e.g., EU documents, encyclopedias, multilingual websites). This gives rise to the question: who owns these translations? Is it the individual translators who produced them, the entities that funded their production, or the platforms that collected and organized them?

The integration of human translations into open or commercial databases further blurs the boundaries between individual and collective ownership, particularly in the absence of clear licensing mechanisms or attribution

practices. The issue becomes even more acute when such data is employed to train commercial MT systems that generate profit without providing recognition or compensation to the original contributors.

In this sense, human translation has shifted from a cognitive activity grounded in creativity and cultural understanding to a "raw material" that fuels AI systems and enables them to learn. Consequently, the translator's relationship with the translated text changes: the text no longer necessarily belongs to the translator but becomes part of the collective intelligence upon which AI capabilities are built.

This transformation calls for a reconsideration of copyright in the digital age, recognizing translators as producers of valuable data whose economic and scientific worth is inseparable from their human skill. It opens the door to developing new legal frameworks that balance the right to technological advancement with the translator's right to preserve intellectual property and receive recognition for their contribution to the construction of linguistic artificial intelligence.

4.1 Legal and Ethical Challenges of Using Human Translations in AI Training

The widespread reliance on bilingual data in training neural machine translation (NMT) systems raises a host of complex issues, extending far beyond technical concerns regarding the availability or quality of the data. Central among these are questions of legality, cultural impact, and ethical responsibility.

In many cases, training datasets are derived from human translations produced by professional translators or specialized institutions. This gives rise to acute legal questions concerning the respect of translators' literary and financial rights, and whether their work is treated as protected intellectual property or merely as extractable data in a digital environment.

Beyond legal considerations, significant cultural and ideological concerns emerge. Bilingual datasets often encode linguistic and cultural representations that carry implicit biases or stereotypes. When these data are reproduced automatically by AI systems, they perpetuate these biases without any awareness of the contextual nuances, thereby reinforcing patterns of cultural exclusion and reproducing unequal representations across languages and societies.

From this perspective, it is essential to reconsider the relationship between the translator and the data they produce. Translation is not merely a technical or functional operation; it is a cultural and interpretative act that engages with authority, meaning, and identity. Treating human translations merely as "digital raw material" risks stripping the activity of its creative and symbolic dimension , undermining the rights of translators as cultural actors.

Even when translations are publicly available on the Internet, this does not automatically place them in the public domain. Most translations remain subject to intellectual property laws or licensing agreements that define permissible reuse. This underscores the need for clear legal frameworks governing the use of such materials in AI training, ensuring that both individual and collective rights are respected.

In this context, the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) is particularly significant. It represents the first international normative document to adopt a holistic approach , linking technological development with respect for human rights , social justice, and the preservation of cultural and linguistic diversity .

Adopted by 193 countries, the Recommendation reflects broad international consensus on the necessity of regulating technological innovation according to ethical principles that protect humans and ensure the fair and responsible use of AI.

Two principles outlined by UNESCO are particularly relevant to the use of human translation in AI:

1. Respect for authors' and intellectual property rights: Technological progress cannot be divorced from the responsibility to acknowledge those who produce knowledge, language, and cultural content. UNESCO emphasizes the importance of recognizing these rights, especially in fields such as translation, where works are frequently used without consent or proper attribution.

2. Transparency and representation: UNESCO calls for full disclosure of data sources, including whether data are drawn from open sources, copyrighted publications, or human-generated works such as translations. This requirement is increasingly critical in the context of large-scale language models (LLMs), which are trained on massive datasets encompassing human translations and literary or cultural works under legal protection. Using such materials without licenses or proper attribution constitutes a violation of authors' and translators' rights and raises serious ethical concerns regarding fairness and the concealment of human contributions behind technological interfaces.

Several recent studies have warned that ignoring these considerations undermines the principles of fairness and transparency, which are fundamental to the ethical governance of AI. UNESCO emphasizes that recognizing the human source of data, whether authors or translators, is a prerequisite for genuine transparency and ensures that technical innovations do not become tools for erasing cultural identities or exploiting individuals' efforts without compensation.

In light of these challenges, there are growing calls for the introduction of explicit legal exceptions within intellectual property laws addressing "machine learning and translation." Such reforms would aim to balance the need for access to data for technological development with the protection of creators, particularly translators, whose crucial role in the knowledge production chain is frequently overlooked or marginalized.

For example, in 2022, a major AI company trained a neural machine translation system using a combination of publicly available literary translations and proprietary translations licensed from a language service provider. After the system produced translations for commercial clients, several professional translators raised concerns that their work had been used without explicit consent or additional compensation. This case highlights the legal ambiguity surrounding the use of human translations in AI training and underscores the ethical necessity of transparent licensing agreements and recognition of translators' rights.

4.2 Ownership and Ethics of Bilingual Corpora in AI Translation:

Modern machine translation (MT) systems rely on advanced statistical and predictive technologies aimed at improving both the accuracy and efficiency of translations.

One of the most prominent tools these systems utilize is translation memories (TMs)—databases containing previously completed human translations for specific entities or institutions. These memories are later reused as training data for machine translation systems, whether statistical or neural-based.

Forcada emphasizes that these systems operate on the principle of machine learning, trained on large quantities of linguistic examples in which each sentence in the source language is paired with its corresponding translation in the target language. Occasionally, these datasets are supplemented with monolingual corpora to enhance the system's understanding of the linguistic and stylistic structures of each language independently:

"... each of which contains a source-language sentence paired with its translation in the target language, and sometimes also from an additional corpus of monolingual target-language sentences" (Forcada, 2023, p. 84).

Bilingual corpora (parallel corpora) used in training neural translation systems require vast amounts of data, often

ranging from hundreds of thousands to millions of sentence pairs. The preparation of such corpora involves highly professional translations, followed by a sentence alignment process, which—although often automated—still requires human supervision to ensure accuracy and conformity between source and target texts.

Modern MT systems utilize segmentation methods, dividing texts into smaller linguistic units such as sentences or phrases. When integrated with translation memories, automated search tools—such as concordancing tools—identify previously translated passages, improving efficiency and consistency within the final output. Segments are analyzed to estimate the most probable translation for each unit, while predicting the most harmonious sequences in the target language. Neural MT has surpassed the limitations of prior statistical models by analyzing language at granular levels, down to individual words or letters, enabling the production of more natural and fluent texts. However, this advancement has increased the demand for large-scale parallel data, making corpus preparation both technically and cognitively challenging.

In addition to bilingual corpora, other linguistic resources, such as dictionaries and grammars, remain important, particularly in rule-based systems. Although reliance on these resources has decreased in favor of neural approaches, they continue to support training code formation and quality control of outputs. The preparation and adaptation of such resources for computer processing requires specialized linguistic and technical expertise and is subject to copyright protection. When organized into executable programs, they also fall under software license regulations governing use and distribution.

Translation memories exemplify language computing applications, consisting of bilingual units (segments) that include sentences or phrases previously translated. These modules, organized into machine-retrievable databases, are

legally classified as databases. Accordingly, the structure and organization of a TM are protected under database legislation, which generally requires obtaining a license before extraction or widespread reuse of its content.

However, protection does not usually extend to the abstract linguistic data within the memory—such as individual sentences or words—unless they reflect personal creativity. The legal value lies primarily in the organization and investment required to assemble these units into a structured, machine-processable format. In this sense, translation memories are considered collective intellectual and technical products, whose protection focuses on form and structure rather than the linguistic content itself:

"The TM file as a by-product of a translation effort is usually… handed over to the hiring agency or end client… In practice, some translators keep copies… In the absence of an enforceable contractual agreement… their legal status is often ill-defined in national and international laws and is thus unclear." (Moorkens & Lewis, 2020, p. 472)Morkens and Lewis note that TM files are often transferred to the client, and unless otherwise specified in the contract, the rights to their content typically belong to the client. Forcada further emphasizes that translators should receive additional compensation for secondary uses of their work, particularly when extracted from the internet via crawling technologies to generate new value in machine translation or computer-assisted translation: "and this has led some authors to ask whether authors and translators should get additional compensation for this unintended use of their work to generate new value downstream." (Forcada, 2023, p. 49).

The legal structure of translation memories involves multipe stakeholders, complicating the determination of rights and ownership. These stakeholders include:

1. Clients or end customers: Owners of the original texts, who under translation contracts usually hold exclusive exploitation rights for the resulting translations. These rights

are separate from the original copyright, which may expire according to national legislation (e.g., seventy years after the author's death in Spain).

2. Translators: Who hold copyright over their translations, particularly the moral right to be attributed. However, segmentation into sentences or passages complicates proving these rights or ensuring fair compensation.

3. Database compilers: Individuals or entities who collect source–target pairs and organize them into machine-processable structures. They are legally protected if their effort demonstrates sufficient creativity, as provided in Article 10 of the TRIPS Convention, provided that the texts were obtained legitimately.

4. Public institutions and government agencies: Often provide open-source translation corpora, particularly in legal and regulatory domains.

5. Academic or official data collection agencies: Typically obtain texts from translators, clients, public domain works, or resources under open licenses permitting reuse.

6. Private institutions: That collect linguistic data and make it available for public use, through either internal efforts or open projects.

These corpora have high knowledge value, serving not only translation purposes but also forming the foundation for linguistic AI development. Yet, the legal framework governing their use remains debated, particularly regarding overlapping rights between original authors, translators, and MT developers.

The use of such resources to train MT systems raises fundamental ethical and legal questions, notably concerning fair compensation when translations are repurposed for contexts not anticipated during their original creation. For instance, consider a translation agency contracting freelance translators for legal and commercial texts. The contract grants the agency the right to use translations solely for the agreed

project while maintaining the translators' moral rights. Years later, the agency develops a proprietary MT system using its accumulated TMs, feeding millions of language pairs into the model without new consent. The system succeeds commercially, generating substantial profits.

This scenario raises a dual problem:

1. Legal: Original translations are protected works, and using them for new purposes requires additional licenses or compensation.

2. Ethical: The automated system benefits from the translators' expertise without recognition or reward, despite their foundational role in training the model.

The issue becomes more complex if the corpora include translations from works that are public domain or open-licensed, blurring distinctions between legally legitimate and disputed uses. This underscores the need for transparent licensing policies to manage translation data, ensuring respect for all stakeholders and limiting unauthorized exploitation.

5. Discussion:

The rapid development of artificial intelligence (AI) in translation has fundamentally reshaped the traditional understanding of authorship, ownership, and the role of human translators. On the one hand, translations have historically been recognized as derivative works that require the prior consent of the original author, granting the translator moral and economic rights over their creative effort. This framework, as established by the Berne Convention (1886) and reinforced in national legislations such as France's L112.3 of the Intellectual Property Code, emphasizes a delicate balance between protecting the original author and acknowledging the creativity inherent in translation.

On the other hand, neural machine translation (NMT) and other AI-based systems have introduced profound challenges to this balance. Translations can now be generated with minimal or even no human intervention, raising pressing legal and ethical questions regarding:

The ownership of machine-generated translations;

The attribution of creativity and authorship—whether to the system programmer, the operator, or potentially no human actor;

The use of human translations as training data, often without the consent or compensation of the original translators;

The reproduction of cultural biases embedded in training data, which may perpetuate stereotypes across languages and societies.

These developments suggest that translation is no longer merely a human cognitive act, but also a source of raw material for AI systems, creating new dynamics in the production, reuse, and monetization of linguistic knowledge. The current legal frameworks, while robust for human-generated translations, are increasingly inadequate to address fully automated or hybrid translation processes, as evidenced by debates over translation memories, bilingual corpora, and proprietary AI systems.

From an ethical perspective, the reliance on human translations to train AI without adequate recognition threatens to devalue the role of the translator, reduce transparency, and undermine the cultural significance of translation as an interpretative and creative act. The UNESCO Recommendation on AI Ethics (2021) provides a timely normative reference, emphasizing respect for authors' rights, transparency, and responsible use of AI-generated outputs. These principles highlight the necessity of reconceptualizing authorship, copyright, and ownership in the AI era, not only to safeguard legal rights but also to ensure ethical and culturally sensitive practices.

In conclusion, the convergence of human creativity and algorithmic production in translation requires a re-evaluation of existing legal, ethical, and professional frameworks. Translators, authors, AI developers, and users must navigate a complex landscape in which human input, machine learning,

and data ownership intersect. Policy interventions may include:

1. Clear legal definitions of authorship and ownership for AI-generated translations;

2. Licensing frameworks ensuring fair compensation for human translators whose work feeds AI training;

3. Ethical guidelines to prevent cultural bias and ensure that machine-generated translations respect linguistic and cultural diversity;

4. Transparency mechanisms for data sourcing, particularly for large-scale language models.

Ultimately, while AI offers unprecedented efficiency and the potential to approximate human translation quality, it cannot replace the interpretative, cultural, and ethical dimensions of human translation. Maintaining a balance between technological innovation and the recognition of human intellectual and creative contributions is essential to preserve both the integrity of translation as a cultural practice and the equitable distribution of the economic and symbolic benefits it generates.

6. Conclusion:

The convergence of human creativity and artificial intelligence in the field of translation has generated profound legal, ethical, and professional challenges. While traditional frameworks recognize translation as a derivative work that respects both the original author's rights and the translator's creative contribution, AI-powered systems—especially neural machine translation—blur these boundaries, raising critical questions about authorship, ownership, and the use of human-generated data.

Human translators, whose work often serves as the foundation for training AI models, face the risk of having their intellectual and creative contributions undervalued or exploited. At the same time, fully automated translations challenge the very notion of authorship, prompting a reconsideration of copyright laws and ethical standards.

Furthermore, the use of human translations as "training material" without clear licensing mechanisms threatens to commodify human creativity, potentially reducing translators to mere data providers while enriching corporate entities controlling AI systems.

To address these challenges, it is essential to reform legal and ethical frameworks to:

1-Ensure fair recognition and compensation for translators;

2-Clarify ownership and rights in AI-generated translations;

3-Promote transparency and accountability in the use of linguistic data;

4- Safeguard cultural and linguistic diversity in translation outputs;

5-Develop international standards for the ethical integration of AI in translation, ensuring that technological progress does not override human intellectual rights or perpetuate cultural biases.

Ultimately, the human element remains central to translation, not only as a source of creativity and interpretative judgment but also as a guarantor of ethical, culturally aware, and meaningful communication.

Balancing technological innovation with the protection of human intellectual and creative rights is not merely a legal necessity but a cultural imperative, ensuring that translation continues to enrich global knowledge, foster cross-cultural understanding, and preserve the dignity of human creativity in the AI era.

Bibliography:

1- Amrousi, A. M. (2024). *The Impact of Artificial Intelligence on the Management and Protection of Intellectual Property Rights*. ***Journal of Intellectual Property and Innovation Management***, 7(1), 89–134. DOI:10.21608/jipim.2024.362437.

2- Bardet, A. (2021). ***Architectures neuronales multilingues pour le traitement automatique des langues naturelles***. PhD thesis for obtaining a PhD degree in computer science. Université du Mans.

3- ***Berne Convention for the Protection of Literary and Artistic Works***. (1886). Paris: World Intellectual Property Organization.
4- European Parliament. (2020). ***European Parliament resolution on intellectual property rights for the development of artificial intelligence technologies (2020/2015(INI))***. Brussels: European Parliament.
5- Forcada, M. L. (2023). *"Licensing and usage rights of language data in machine translation"*. ***In*** H. Moniz & C. Parra Escartín (Eds***.). Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation***. Heidelberg: Springer International Publishing.
6- Gaidartzi, A., & Stamatoudi, I. (2025). *Authorship and Ownership Issues Raised by AI Generated Works: A Comparative Analysis*. ***Laws***, 14(4), 57. https://doi.org/10.3390/laws14040057.
7- Lacruz Mantecón, M. L. (2023). *"Authorship and rights ownership in the machine translation era"*. ***In*** H. Moniz & C. Parra Escartín (Eds.). ***Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation***. Heidelberg: Springer Nature.
8- Moorkens, J., & Lewis, D. (2020). *"Copyright and the reuse of translation as data"*. ***In*** M. O'Hagan (Ed.). ***The Routledge Handbook of Translation and Technology***. London: Routledge.
9- Sbihi, S. (2025). *Artificial Intelligence and Intellectual Property Rights: A Correlative Legal Study*. ***Africa University Journal of Scientific Research***, 19(1), 212–226.
10- UNESCO. (2021). ***Recommendation on the Ethics of Artificial Intelligence***. Paris: UNESCO.
11- United Kingdom. (1988). ***Copyright, Designs and Patents Act 1988***, Section 9(3). London: HMSO.
12- Venuti, L. (1995). ***The Translator's Invisibility: A History of Translation***. London: Routledge.