



Prompt-Driven Augmentation in AI Translation: Thematic Alteration and Hallucination

Samin Salajegheh* 

PhD Candidate in Translation Studies, English
Translation Studies Department, Faculty of
Persian Literature and Foreign Languages,
Allameh Tabataba'i University, Tehran, Iran

Abstract

This study examines how prompt-driven expansion affects the content and source-groundedness of AI-generated translations. It introduces “augmented translation”, translation prompted to exceed the source text in length, and examines changes in textual features and hallucinations across augmentation levels. The study analyzes 100 English tourism paragraphs from *TripAdvisor* and their Persian ChatGPT translations with 20%, 50%, and 100% augmentation. Using qualitative thematic and descriptive analysis, the study identifies five patterns of change: lexical and stylistic amplification, content alteration, pragmatic and relational shift, omission and reduction, and structural and syntactic shift. Lexical and stylistic amplification was most frequent, especially through adjective addition and persuasive enhancement. Counterintuitively, 20% augmentation showed the most content deletion, while 50% favored adjective-based elaboration and 100% increasingly produced unsupported claims and persuasive enhancement. Hallucinations also shifted across augmentation levels: lower augmentation produced calculation-related errors, whereas higher augmentation increasingly generated unsupported claims. Hallucinations also tended to emerge toward the later portions of augmented translations, suggesting that pressure to sustain additional content may push generation beyond source-grounded information. The findings highlight prompt design as a consequential factor in AI-assisted translation and reveal a tension between elaboration and informational reliability.

Keywords: AI translation, augmented translation, hallucination, thematic alteration, translation expansion, translation prompt

*Corresponding author: samin_salajegheh@atu.ac.ir

Cite this article: Salajegheh, S. (2026). Prompt-driven augmentation in AI translation: Thematic alteration and hallucination. *Translation and Interpreting Research*, 3(10), 43-57. <https://doi.org/10.22054/tir.2026.96206.1096>



© 2026 – Allameh Tabataba'i University Press

Print ISSN: 3092-6599

Online ISSN: 3060-7019

Publisher: Allameh Tabataba'i University Press

Published: 30.06.2026

Accepted: 14.06.2026

Received: 08.05.2026

1. Introduction

The emergence of artificial intelligence (AI), particularly large language models such as ChatGPT, has significantly reshaped the contemporary translation landscape (S. Sun et al., 2025). Although language models are widely used for translation, they repeatedly generate untrustworthy and factually incorrect outputs, a phenomenon broadly referred to as hallucination (Förster & Skop, 2025). This concern is particularly important in translation because the target text is generally expected to remain grounded in the source while fulfilling the communicative requirements of the target language and context.

Research on AI translation hallucination has largely examined conventional translation, in which the primary expectation is that the target text should reproduce the content of the source. However, translation is not always restricted to close reproduction. Depending on the communicative purpose and task requirements, translators may paraphrase, adapt, elaborate, or otherwise expand source material. This creates a less straightforward context for identifying hallucinations. When additional material is intentionally requested, an increase in target-text content is not necessarily erroneous; rather, the critical question is whether such additions remain grounded in information available in the source.

This distinction has received some attention in research on AI-generated paraphrased translation. For example, Gogoulou et al. (2025) examined AI hallucinations in producing paraphrased translation and demonstrated that hallucination can take multiple forms even when the task permits some degree of reformulation. Their findings are particularly relevant to translation tasks in which departure from the source is intentional. Nevertheless, paraphrasing and deliberate length expansion represent different conditions. In paraphrasing, the principal objective is to reformulate existing content, whereas a length-constrained task explicitly requires the system to generate additional textual material. The latter therefore creates a particular challenge: the model must increase the amount of text while maintaining a meaningful relationship with the information provided in the source.

The present study addresses this relatively unexplored condition through the concept of “augmented translation”, defined here as a translation deliberately requested to contain a greater number of words than its source text. Unlike conventional translation, in which differences in length may arise naturally from linguistic and structural differences between languages, augmented translation makes increased textual length an explicit instruction. Such an instruction gives the AI system greater freedom to elaborate the source, but it might also create more opportunities for unsupported additions. The boundary between legitimate elaboration and hallucination consequently becomes central to the task. This issue is particularly relevant because the effect of a length instruction may extend beyond the simple addition of words. When no specific content is identified for augmentation, the AI system must independently determine what aspects of the source to elaborate. It may intensify descriptions, introduce evaluative language, provide additional details, establish relationships between existing pieces of information, or add material that is not supported by the source. Thus, increasing translation length may affect not only the quantity of the output but also the thematic and informational composition of the translation.

The present study therefore investigates how AI-generated translations respond to progressively increasing requests for textual expansion or augmentation. It examines which aspects of the source text are most frequently augmented when the requested target length increases and identifies the thematic patterns through which such augmentation is realized. Particular attention is given to whether the additional material elaborates information already available in the source, modifies its presentation, omits existing information while compensating through other additions, or introduces hallucinated content.

The study further examines the relationship between requested augmentation and hallucination. Rather than treating every departure from the source as hallucination, it distinguishes between augmentation that remains grounded in the source and augmentations that cannot be adequately supported by it. This distinction is important because intentional transformation is an increasingly relevant feature of AI-assisted translation, while source-groundedness remains essential when translated content is expected to represent source information reliably.

More specifically, the study addresses three questions:

1. What types of thematic and textual changes are introduced when AI is instructed to produce translations that are progressively longer than their source texts?
2. Which aspects of the source are most frequently expanded at different levels of requested augmentation?
3. How does the occurrence and type of hallucination vary as the requested degree of translation augmentation increases?

By examining these questions, the study moves beyond whether AI translation can hallucinate to investigate how deliberate textual expansion shapes AI-generated translations. It connects translation hallucination research with prompt-based control and source-groundedness, examining whether AI can elaborate source content without introducing unsupported information.

2. Literature Review

This study aimed to reflect on the hybridity of translating an English book into Persian, which is written by an Iranian author, so there are many numbers of hybrid items in the text. The starting point to examine the hybridity of the text is to define the meaning of hybridity itself, “A hybrid text is a text that results from a translation process. It shows features that somehow seem ‘out of place’/ ‘strange’/ ‘unusual’ for the receiving culture, i.e., the target culture” (Schäffner & Adab, 2001, p. 169).

The transformative potential of artificial intelligence across a wide range of industries, particularly language-related fields, is increasingly evident (Tseng & Warschauer, 2023). AI-assisted systems can rapidly process large volumes of text, but their growing use also increases the need for human revision, quality assessment, and post-editing. Consequently, translators increasingly need to combine AI-related skills with their linguistic and analytical expertise (Amini et al., 2024).

These benefits are accompanied by important limitations. AI-generated language may reduce cultural specificity and homogenize stylistic features, while prioritizing fluency over accuracy and thus failing to faithfully convey source information (Shi et al., 2024). Moreover, AI systems may reproduce biases embedded in training data and generate factually inaccurate content (Heinisch, 2025). Such deviations may extend beyond isolated factual errors, as AI-generated content can potentially distort how information and reality are represented, thereby raising broader concerns about epistemic reliability (Krüger, 2025). This problem is particularly evident in the phenomenon commonly referred to as “hallucination”, whereby text-generation systems produce content that appears coherent and plausible but is not sufficiently grounded in the available information (Maleki et al., 2024). Hallucination has become a particularly important concern in generative AI because generated content may appear coherent and plausible while deviating from the user’s intent, containing factual inaccuracies, or being inappropriate to the context (Liu & Lin, 2026). Despite advances in automated translation, current systems can still struggle to maintain coherence and source alignment beyond the sentence level, creating opportunities for hallucinated or linguistically unsupported output (Bogucki, 2026).

More generally, AI hallucination, also referred to as artificial hallucination or confabulation, describes generated responses that are not factually accurate or sufficiently grounded in real-world data. Such outputs may emerge when models operate with insufficient contextual information or constraints and generate unexpected combinations based on patterns learned from large datasets. Errors in training data, model overfitting, and insufficient understanding of underlying concepts have all been identified as contributing factors (Jamaluddin et al., 2023). Because hallucinations can have ethical and legal consequences and negatively affect decision-making, their prediction and mitigation have become active areas of research (Athaluri et al., 2023). For example, Jesson et al. (2024) explored approaches to estimating hallucination rates and predicting their occurrence based on uncertainty surrounding the meaning of generated responses. Importantly, hallucination is not limited to general text generation but can also emerge in AI-assisted translation, making source-groundedness a central concern in evaluating AI-generated translations.

Research on translation hallucination has consequently sought to define the phenomenon and distinguish its different manifestations. Xu et al. (2023) characterized translation hallucinations as source-unrelated outputs and distinguished between “detached hallucinations”, in which substantial portions of the output diverge from the input, and “natural hallucinations”, which may arise from corpus-level noise. From a broader perspective, hallucinations in translation have been associated with model architecture, biases in training data, and weaknesses in inference algorithms (Naser, 2026). Wu et al. (2025), however, argue that hallucination in translation involves more than simply departing from the source text. Models may also fail to follow the instructions governing the translation task. Their taxonomy therefore distinguishes “Instruction Detachment” from “Source Detachment”. Instruction Detachment includes phenomena such as untranslated content and incorrect target language, whereas Source Detachment encompasses extraneous additions and repetition. On this basis, they developed HalloMTBench, a multilingual, human-verified benchmark containing 5,435 hallucination instances across 11 English-to-X translation directions.

Other studies have further demonstrated that translation hallucination can take diverse forms depending on the linguistic and communicative context. Yilmaz (2026), examining Turkish-Arabic LLM translation across literary, legal, and medical texts, distinguished “intrinsic hallucinations”, involving distortions or shifts in source meaning, from “extrinsic hallucinations”, involving information absent from the source. The study identified manifestations including interpretive expansion, source-external aesthetic additions, task-irrelevant explanations, repetition, terminological deviations, and unsupported additions. Their distribution also varied across domains: literary translation was particularly associated with interpretive expansion, aesthetic additions, repetition, and irrelevant explanations, whereas legal translation involved terminological deviation, semantic narrowing, and source-external additions; medical translation showed limited semantic shifts, terminology problems, and irrelevant explanations. Similarly, Bang et al. (2023) found that ChatGPT, particularly when translating low-resource languages, was susceptible to extrinsic hallucination. Jiao et al. (2025) likewise categorized hallucinations in ChatGPT translation into semantic, pragmatic, syntactic, and cultural types.

The occurrence of hallucination also varies according to the model and language pair. Benkirane et al. (2024) found that model selection is important for mitigating hallucinations, although even the best-performing systems continued to generate hallucinated content. Working across more than 100 language pairs, Guerreiro et al. (2023) similarly found that hallucinations were more prevalent in low-resource and non-English translation directions and could sometimes reflect toxic patterns embedded in training data. These findings demonstrate that hallucination cannot be treated simply as a uniform translation error; rather, its occurrence and form may depend on the interaction among the model, language pair, source material, and translation task.

Given these challenges, researchers have investigated ways of reducing hallucination. Tang et al. (2025) reviewed post-hoc approaches that detect hallucinated translations and subsequently retranslate them, while proposing an alternative approach that seeks to mitigate hallucination intrinsically during model training rather than correcting it only after generation. Their approach achieved a 96% success rate across. Prompt engineering has also been examined as a means of controlling hallucination. A prompt is a natural-language question, statement, or instruction that provides context for an AI-generated response (Paradowska, 2025), whereas prompt engineering involves designing and refining prompts to optimize AI outputs (Bozkurt, 2024). Tang et al. (2025) found that zero-shot prompts reduced hallucinations by an average of 89% across three previously unseen target languages. Similarly, Anh-Hoang et al. (2025) reported that vague prompts produced the highest hallucination rate (38.3%), whereas Chain-of-Thought (CoT) prompting reduced the rate to 18.1%. Their findings further suggested that hallucinations may be prompt-dominant, model-dominant, or mixed in origin. Although structured prompting generally reduced hallucination, it did not eliminate model-intrinsic errors; notably, CoT prompting could produce longer responses that remained incorrect when the model lacked the necessary knowledge.

An important complication, however, emerges when translation itself does not require strict reproduction of the source. Translation is not always expected to preserve the source in a conventional, closely corresponding format. Depending on the communicative purpose, translators may adapt, paraphrase, elaborate, or otherwise modify the source. This creates a particularly important problem for the study of hallucination: if additional material is intentionally requested, not every addition can automatically be treated as hallucinated. The distinction between legitimate transformation and unsupported content generation therefore becomes less straightforward. Addressing this issue, Gogoulou et al. (2025) examined hallucinations in paraphrased translation and identified 11 categories, including additions, named-entity changes, number and date errors, gender changes, pronoun changes, antonym substitutions, tense changes, negations, conversions, and natural hallucinations. Their work demonstrates that hallucination research needs to account for translation tasks in which some degree of departure from the source is intentional.

Translation length can naturally increase because of language-specific structural requirements and translation-related textual changes. Differences in morphology and syntax may require additional words to convey equivalent meanings, while translation itself can introduce further textual material (Frankenberg-Garcia, 2009). However, deliberately instructing an AI system to substantially increase word count creates a distinct condition: although expansion is intentional, the added content must remain grounded in the source. Greater textual freedom may therefore increase opportunities for unsupported information, making translation length an important context for examining hallucination.

3. Methodology

The present study combines qualitative content analysis with descriptive quantitative analysis. It investigates how progressively increasing the requested length of an AI-generated translation affects the content and distribution of textual additions and whether such expansion gives rise to hallucinated information. More specifically, the study examines whether asking an AI system to produce a translation that is longer than the source text (augmented translation) results primarily in elaboration of information already present in the source or whether the additional length is achieved through modifications or the generation of information unsupported by the source. The study further explores how these changes vary across three levels of requested augmentation and identifies the types and frequency of any hallucinations that emerge.

The corpus consisted of 100 English-language tourism paragraphs randomly selected from [TripAdvisor](#) entries describing tourist attractions in Iran. The tourism domain was selected because tourism texts

commonly combine factual, descriptive, cultural, and evaluative information, thereby providing a suitable context for examining whether an AI system expands a translation by elaborating existing content or by introducing information that cannot be supported by the source text. TripAdvisor was used as the source because it provides user-generated tourism content concerning individual attractions and destinations. All texts were collected in July 2026 to establish a fixed temporal point for corpus compilation, given that online content and search results may change over time. The selected paragraphs were copied into Microsoft Word to constitute the initial source corpus, and the word count of each paragraph was recorded before translation.

To minimize the possible influence of previous interactions with the AI system, a new account on the ChatGPT (free version available in July 2026) was created specifically for the study, with no prior interaction history. Each source paragraph was submitted to ChatGPT. The same source paragraph was translated under three controlled augmentation conditions, with the requested target length progressively increased relative to the source. Thus, each paragraph generated three Persian translations: a first-level augmented translation requested to be 20% longer than the English source, a second-level augmented translation requested to be 50% longer, and a third-level augmented translation requested to be 100% longer. The prompts were deliberately kept identical in wording across conditions except for the specified degree of augmentation, allowing the requested length to function as the principal experimental variable.

The following prompt was used: “Translate the following English tourism text into Persian. Produce a translation that is 20% longer than the source text”. For the second level, the same prompt was used with the length constraint changed to 50%. For the third level, the length constraint was changed to 100%.

The three sets of AI-generated translations were subsequently entered into Microsoft Word tables alongside their corresponding English source paragraphs to facilitate systematic source-target comparison.

The analysis proceeded at two complementary levels. First, the additional or altered material introduced at each augmentation level was identified and subjected to thematic analysis. This analysis examined the thematic domains in which AI-generated augmentations occurred and compared their frequency across the 20%, 50%, and 100% augmentation conditions. Particular attention was paid to whether the altered material constituted elaboration of information already available in the source text or introduced new content.

Second, all additions that were not adequately supported by the source text were examined for possible hallucination. Hallucinations were identified through direct comparison of each Persian translation with its English source and were coded according to the comprehensive hallucination taxonomy proposed by Sun et al. (2024). The taxonomy organizes hallucination-related errors into first-level and second-level categories and draws on several previously proposed classifications. Because some instances could potentially correspond to more than one category, overlap between categories was possible in both the thematic and hallucination analyses. In such cases, the category that most closely captured the primary function or defining characteristic of the instance in its textual context was selected as the final classification. Each instance was therefore assigned one primary category to avoid multiple counting within the respective analysis.

The hallucination taxonomy used in the present study is provided in Table 1.

Table 1. Taxonomy of AI Hallucination (Sun et al., 2024)

No	First-Level Hallucination Type	Second-Level Hallucination Type	Definition
1	Overfitting	Illusions of Confidence	Presenting unsupported judgments or conclusions with excessive certainty while overlooking possible errors or alternative information.
		Falling into Traps	Producing content as a result of being misled or drawn into a trap contained in the prompt or context.
		Flattery	Adding exaggerated, favorable, or one-sided content to satisfy or please the intended audience.
2	Logic Errors	Causal Uncorrelation	Establishing a causal connection between events or information when the source provides no such relationship.
		Contradictions	Producing statements that conflict with one another or are internally inconsistent.
3	Reasoning Errors	Spatial Reasoning Errors	Misrepresenting the spatial relationships among people, objects, or locations.
		Temporal Reasoning Errors	Misrepresenting the timing, duration, sequence, or chronological relationship of events.
		Physical Reasoning Errors	Misrepresenting physical objects, interactions, or processes in the real world.
		Psychological Reasoning Errors	Incorrectly attributing or predicting human thoughts, intentions, emotions, or behavior.
		Satire	Misinterpreting or reproducing satirical content as factual information, or failing to recognize its humorous or critical intent.
		Metaphorical Errors	Misunderstanding or incorrectly interpreting metaphorical meaning.
4	Mathematical Errors	Conceptual Errors	Misunderstanding mathematical concepts, relationships, or properties.
		Measurement Units Errors	Misinterpreting or incorrectly representing units of measurement.
		Calculation Errors	Performing or reporting incorrect mathematical calculations or numerical operations.
5	Unfounded Fabrication	False Health Information	Generating unsupported or false health-related claims, explanations, cases, or opinions.
		Fake Reviews	Creating fabricated reviews or attributing experiences or opinions to consumers without a genuine basis.
		Fake News	Fabricating, substantially distorting, or deceptively presenting news-related information as factual.
		Pseudoscience	Generating scientifically unsupported or scientifically dubious claims.
		False Academic Information	Inventing academic works, references, links, or other apparently scholarly information that does not exist or is irrelevant.
		Conspiracy Theory	Generating unsupported accounts that explain events through alleged secret plots without an established factual basis.
6	Bias and Discrimination	Hate News	Generating derogatory content targeting particular ethnic, gender, or other social groups.
		Discrimination	Introducing content that reflects unequal or prejudicial treatment based on race, gender, religion, ethnicity, social status, or similar characteristics.
		Biased or One-Sided	Presenting information in an excessively partial manner by favoring a particular person, group, position, or situation.
7	Factual Errors	Common Sense Mistakes	Producing information or conclusions that conflict with generally accepted knowledge or basic common sense.
		Objective Fact Errors	Incorrectly reporting verifiable facts such as names, places, dates, times, or numerical information.
		Author's Work Errors	Misrepresenting works, authors, characters, relationships among works, or content associated with particular authors or works.

No	First-Level Hallucination Type	Second-Level Hallucination Type	Definition
8	Text Output Errors	Repetition and Redundancy	Adding repetitive, unnecessary, or excessively redundant material that was not requested.
		Programming Code Errors	Producing incorrect or ineffective programming code.
		Translation Errors	Producing a target text meaning that does not correspond to the meaning expressed in the source language.
		Grammar Errors	Producing language that violates grammatical or linguistic conventions, including problems with syntax, tense, vocabulary, or expression.
		Spelling Errors	Producing incorrectly spelled words or otherwise failing to generate the expected written form.
9	Misleading Error	Misleading Connections	Combining individually plausible information through an incorrect context, relationship, source attribution, or background.
		Propaganda	Generating information intended to shape public perceptions or opinions in favor of a particular person, organization, or government.
		Clickbait	Using exaggerated or misleading wording or presentation to attract attention despite an otherwise factual basis.
		Manipulation	Altering or transforming genuine information in a way that supports a deceptive or false narrative.
		Parody	Presenting or interpreting absurd or exaggerated claims in a way that can make them difficult to distinguish from factual information.
		Trolling	Producing deliberately provocative, offensive, or inflammatory content intended to provoke reactions or disrupt interaction.
10	Other Errors	Hoax	Presenting relatively extensive fabricated or inaccurate information as legitimate fact, potentially causing harm or material loss.
		Restrictive filtering	Ignoring particular words or statements in the input or refusing to respond to certain questions.
		Self-awareness	Representing or interpreting the system as possessing an individual identity, thoughts, feelings, personality, or self-understanding.

4. Results

This section presents and analyzes one example from the dataset, followed by the overall results across the entire corpus.

The example is a text introducing the *Iran Science Park* on [TripAdvisor](#):

The Iran Science Park is the largest science center in I.R.Iran located in Tehran book garden. It's a dynamic destination where families, adults and children can explore the wonders of science through interactive exhibits. It also has a planetarium with a domed wraparound movie theater. Let these films take you on scientific adventures so big.

ChatGPT's 20% expanded translation:

پارک علم ایران، بزرگ‌ترین مرکز علمی جمهوری اسلامی ایران، در باغ کتاب تهران واقع شده است. این مرکز، مقصدی پویا و جذاب برای خانواده‌ها، بزرگسالان و کودکان به شمار می‌رود تا شگفتی‌های علم را از طریق نمایشگاه‌های تعاملی و جذاب کشف نمایند. همچنین دارای افلاکنمایی با سینمای گنبدی فراگیر است که تجربه‌ای منحصر به فرد را ارائه می‌دهد. بگذارید این فیلم‌ها شما را به ماجراجویی‌های علمی عظیم و فراموش‌نشدنی ببرند.

In this augmented translation, ChatGPT has added the adjective «جذاب» (“attractive”) to describe both the “dynamic destination” and the “interactive exhibits”. It has also introduced (“unique”) to describe

the movie theater experience, and «فراموش‌نشدنی» (“unforgettable”) to modify the scientific adventures. The pattern of augmentation here is the systematic addition of adjectives, which collectively serve to make the text more persuasive and appealing to the reader. However, no hallucination is identified in this instance, as these adjectival additions do not introduce content that is unsupported or contradicted by the source text.

ChatGPT’s 50% expanded translation:

پارک علم ایران، بزرگ‌ترین مرکز علمی کشور، در باغ کتاب تهران قرار دارد. این مقصد پویا و جذاب به خانواده‌ها، بزرگسالان و کودکان امکان می‌دهد از طریق نمایشگاه‌های تعاملی، شگفتی‌های علم را کشف کنند. این مجموعه همچنین دارای افلاک‌نمایی با سالن سینمای گنبدی و فراگیر است. فیلم‌های این سالن، بازدیدکنندگان را به ماجراجویی‌های علمی گسترده می‌برند و به همین دلیل، نمایشگاه‌های تعاملی پارک موجب افزایش علاقه بازدیدکنندگان به علم می‌شوند و آنان را به تماشای فیلم‌ها ترغیب می‌کنند و در نتیجه، تجربه‌ای آموزشی و جذاب فراهم می‌شود.

This augmented translation adds adjectives such as «جذاب» (“attractive”), following the same pattern of adjective addition. However, it also introduces an unsupported causal relationship through the phrase «به همین دلیل» (“for this reason”). Specifically, the statement «به همین دلیل، نمایشگاه‌های تعاملی پارک موجب افزایش علاقه بازدیدکنندگان به علم می‌شوند» (“For this reason, the park’s interactive exhibits increase visitors’ interest in science”) establishes a causal connection between the presence of interactive exhibits and increased visitor interest in science. The source text merely states that visitors can explore the wonders of science through interactive exhibits and does not claim that these exhibits increase their interest in science. This constitutes a causal uncorrelation hallucination, which falls under the broader category of Logical errors.

ChatGPT’s 100% expanded translation:

پارک علم ایران، به عنوان بزرگ‌ترین مرکز علمی جمهوری اسلامی ایران، در باغ کتاب تهران واقع شده است و یکی از مهمترین مقاصد علمی-تفریحی کشور محسوب میشود. این مرکز، مقصدی پویا، زنده و جذاب برای خانواده‌ها، بزرگسالان و کودکان به شمار میرود تا بتوانند شگفتی‌های شگفت‌انگیز علم و فناوری را از طریق نمایشگاه‌های تعاملی، hands-on و جذاب کشف نمایند. همچنین دارای افلاک‌نمایی مجهز با سینمای گنبدی فراگیر و همه‌جانبه است که تجربه‌ای منحصر به فرد و تماشایی را برای بازدیدکنندگان به ارمغان می‌آورد. بگذارید این فیلم‌های مهیج و دیدنی، شما را به ماجراجویی‌های علمی عظیم، شگرف و فراموش‌نشدنی ببرند که تاکنون نظیر آن را تجربه نکرده‌اید.

In this augmented translation, ChatGPT added «یکی از مهم‌ترین مقاصد علمی-تفریحی کشور محسوب می‌شود» (“it is considered one of the most important scientific-recreational destinations in the country”), which asserts a nationwide significance not present in the source text. This constitutes an unjustified claim addition, creating a hallucination classified as Illusion of Confidence under the category of Overfitting. Adjectives such as «زنده و جذاب» (“lively and attractive”), «همه‌جانبه» (“immersive”), and «منحصربه‌فرد و تماشایی» (“unique and spectacular”) continue the same pattern of adjective addition. The addition of «که تاکنون نظیر آن را تجربه نکرده‌اید» (“the likes of which you have never experienced”) serves the purpose of persuasion enhancement. However, the most significant problem is the addition of “hands-on”, an English word embedded within the Persian translation. This word does not appear anywhere in the English source text. It appears that when the AI was asked to produce a translation with 100% more words, it first expanded the source into a longer text and then converted that into Persian, leaving one word untranslated. This source-text lexical intrusion constitutes a form of hallucination based on Sun et al.’s (2024) taxonomy, which is called a translation error under the category of text output errors.

4.1. Thematic Alteration in AI-Augmented Translations

The thematic analysis examined how the progressively increased length requested from ChatGPT was realized in the Persian translations and, more specifically, whether the additional textual material primarily elaborated information already present in the English source or introduced changes and content unsupported by it. Comparison of the source texts with the three levels of augmented translation revealed five overarching themes: Lexical and Stylistic Amplification, Content Alteration, Pragmatic and Relational Shift, Omission and Reduction, and Structural and Syntactic Shift.

1. **Lexical and Stylistic Amplification:** encompasses additions that primarily modify the descriptive, rhetorical, emotional, or stylistic qualities of the translation. This theme includes adjective addition, literary enhancement, persuasive enhancement, and quantifier addition. Among these, adjective addition emerged as the most prominent pattern, reflecting the AI system's tendency to make descriptions more elaborate, vivid, and appealing. Literary enhancement refers to the use of more poetic, figurative, or aesthetically elaborate expressions than those found in the source text, whereas persuasive enhancement involves additions that strengthen the rhetorical and promotional appeal of the tourism discourse. Quantifier addition concerns the introduction of numbers, amounts, degrees, or other expressions of quantity that are absent from the source text.
2. **Content Alteration:** encompasses changes that modify, expand or otherwise affect the informational and propositional content of the source text. This theme includes unjustified claim addition, factual detail addition, cultural adaptation expansion, temporal reference addition, and source-text lexical intrusion. Unjustified claim addition refers to the introduction of assertions, interpretations, or relationships that are not supported by the source text and was therefore particularly relevant to the identification of hallucinated information. Factual detail addition involves the introduction of new factual information that is absent from the source. Cultural adaptation expansion refers to additions that localize or culturally adapt the source content for the presumed target audience. Temporal reference addition concerns the introduction of dates, periods, or other temporal information not specified in the source. Source-text lexical intrusion refers to the inappropriate retention or introduction of source-language lexical material into the target text.
3. **Pragmatic and Relational Shift:** encompasses changes that affect the communicative relationship established between the text, its subject matter, and its intended reader. This theme includes informational redundancy and interpersonal relationship shift. Informational redundancy refers to additions that increase the length of the translation without introducing substantive new information, such as repetitive, filler-like, or unnecessarily elaborative expressions. Interpersonal relationship shifts concerns changes in the degree of formality, politeness, interpersonal distance, or reader orientation established by the source text.
4. **Omission and Reduction:** encompasses changes involving the removal or reduction of information available in the source text. This theme consists of content deletion, which covers the omission of source-text information, including general, historical, and geographical details. Although the translation task explicitly required the AI system to produce longer target texts, these instances demonstrate that increased target-text length did not necessarily entail the complete preservation of source-text information. In other words, augmentation could occur alongside the deletion of existing content.
5. **Structural and Syntactic Shift:** encompasses changes in the linguistic structure and syntactic realization of the source text. This theme includes translation error, passive structure increase, and readability reduction. Translation error refers to cases in which the meaning of a source-text segment was inaccurately rendered or misinterpreted at the sentence level. Passive structure increase concerns the use of passive constructions where the source employed a different

structural realization. Readability reduction refers to instances in which the augmented translation became more difficult to process because of increased syntactic complexity or less accessible sentence construction.

Table 2 presents the complete distribution of the identified themes and their original codes.

Table 2. Distribution of Changes Identified in AI-augmented Translations

Overarching theme	Original codes	Frequency	Percentage
Lexical and Stylistic Amplification	Adjective Addition (51); Persuasive Enhancement (29); Literary Enhancement (14); Quantifier Addition (9)	103	37.32%
Content Alteration	Unjustified Claim Addition (44); Cultural Adaptation Expansion (15); Temporal Reference Addition (8); Factual Detail Addition (4); Source-Text Lexical Intrusion (2)	73	26.45%
Pragmatic and Relational Shift	Informational Redundancy (35); Interpersonal Relationship Shift (11)	46	16.67%
Omission and Reduction	Content Deletion (28)	28	10.14%
Structural and Syntactic Shift	Translation Error (13); Passive Structure Increase (9); Readability Reduction (4)	26	9.42%
Total	—	276	100%

The distribution of the overarching themes shows that Lexical and Stylistic Amplification was the most prevalent type of change, indicating that increased translation length was most often achieved through modifications to the style and presentation of the source content. Content Alteration was the second most frequent theme, suggesting that expansion also commonly involved changes to the informational content itself. Pragmatic and Relational Shift occurred less frequently, while Omission and Reduction and Structural and Syntactic Shift were the least prevalent themes and occurred at equal frequencies. Overall, the findings indicate that AI-driven augmentation was predominantly realized through lexical, stylistic, and content-level changes rather than structural changes.

Table 3 presents the most frequent changes identified at each level of requested augmentation, highlighting how the predominant type of change shifted as the required target-text length increased.

Table 3. Most Frequent Changes Across Levels of Requested Augmentation

Level of requested augmentation	Most frequent change(s)	Observed tendency
Level 1 (20% augmentation)	Content Deletion, particularly historical and geographical information	Information loss despite the instruction to produce a longer translation
Level 2 (50% augmentation)	Adjective Addition	Descriptive and stylistic elaboration
Level 3 (100% augmentation)	Unjustified Claim Addition; Persuasive Enhancement	Augmentation through unsupported content and increased promotional force

4.3. Hallucination Occurrence Across Augmented Levels

The second phase of the analysis focused on the occurrence and distribution of hallucinations across the three levels of requested augmentation. Table 4 presents the frequency and predominant type of hallucination identified at each level of requested augmentation.

Table 4. Predominant Hallucination Types Across Augmentation Levels

Level of requested augmentation	Number of hallucinations	Most frequent hallucination type
Level 1 (20% augmentation)	28	Mathematical Error: Calculation
Level 2 (50% augmentation)	63	Overfitting: Illusions of Confidence
Level 3 (100% augmentation)	105	Overfitting: Illusions of Confidence
Total	196	—

Overall, the findings indicate a gradual increase in hallucination occurrence as the requested target-text length increased. This pattern suggests that greater expansion requirements may provide more opportunities for the AI system to move beyond source-based elaboration and generate information that is not adequately grounded in the source text. The nature of the predominant hallucination also differed across augmentation levels, indicating that the relationship between requested augmentation and hallucination was not limited to an increase in frequency but also involved changes in the forms that hallucinated content assumed. At the highest levels of augmentation, the observed hallucinations were particularly associated with information presented with unwarranted certainty.

Across the three augmentation levels, hallucinated material was also generally concentrated toward the later portions of the generated translations. This distributional pattern suggests that unsupported content tended to emerge as the AI system continued generating material beyond the core information available in the source, particularly when it was required to produce substantially longer target texts.

5. Conclusion

The present study aimed to investigate whether requesting AI to produce a translation with a higher word count than the source text results in systematic expansion of particular types of content and, consequently, leads to an augmented translation. More specifically, the study examined which aspects of the source text are most frequently expanded as the requested translation length increases. It also sought to determine whether such augmentation remains grounded in and elaborates upon information available in the source text or instead introduces unsupported information, thereby giving rise to hallucination. The findings show that AI-generated augmentation followed multiple patterns. The added or altered material clustered into five overarching themes: Lexical and Stylistic Amplification, Content Alteration, Pragmatic and Relational Shift, Omission and Reduction, and Structural and Syntactic Shift. Lexical and Stylistic Amplification was the most frequent, particularly through adjective addition and persuasive enhancement, indicating that increased length often prompted more elaborate and rhetorically enhanced translations. Notably, the persuasive dimension can also be linked to the nature of the texts, which was promotional tourism material. Content Alteration was the second most frequent theme and included unjustified claims, cultural adaptation, temporal references, factual details, and source-text lexical intrusion. Notably, greater length did not necessarily result in more complete source-text transfer, as content deletion also occurred. The progression across augmentation levels revealed qualitative differences in expansion. At the 20% level, content deletion was particularly noticeable, while adjective addition predominated at 50%. At 100%, expansion was more strongly associated with unjustified claims and persuasive enhancement. This progression suggests that increasing the requested degree of expansion does not simply produce more

of the same type of elaboration; rather, it can shift the qualitative nature of the generated material from relatively low-risk stylistic enhancement toward more substantial content-level intervention.

Hallucination increased progressively with the requested level of augmentation. At Level 1, Mathematical Error: Calculation was the most recurrent hallucination type, whereas Overfitting: Illusions of Confidence predominated at Levels 2 and 3. Hallucinated content also tended to occur toward the later portions of the generated translations, particularly at higher levels of augmentation. Taken together, these findings suggest that prompting AI to move beyond straightforward translation and to deliberately manipulate or expand the source text increases the likelihood of introducing unsupported information. Thus, while expanded prompting may produce more elaborate and rhetorically enriched translations, it may also progressively weaken their grounding in the source text. For translation tasks in which fidelity and source-groundedness are primary concerns, conventional translation prompts may therefore provide a more controlled approach than prompts that explicitly encourage substantial textual expansion.

The study was not without limitations. First, the analysis was restricted to promotional tourism texts, whose persuasive and descriptive characteristics may have contributed to the prominence of lexical amplification and persuasive enhancement. Different patterns may emerge in other genres, particularly in domains such as medical, legal, or technical translation, where hallucinations may carry more serious consequences. This genre-specificity is important in light of Yilmaz's (2026) finding that hallucination manifestations varied across literary, legal, and medical translation, suggesting that the type of source material may shape not only the amount but also the form of AI-generated deviation. Second, the study examined the output of only one AI system, ChatGPT; other systems may respond differently to expansion-oriented prompts and may exhibit different patterns of augmentation and hallucination.

Despite these limitations, the study contributes to AI-assisted translation by demonstrating that requested output length can influence not only the quantity but also the nature and reliability of translated content. It draws attention to prompt design as an important factor in shaping AI-generated translations and shows that instructions encouraging substantial expansion may introduce a tension between elaboration and source-groundedness. The findings may also help translators and other users of AI translation systems make more informed decisions about when and how to request elaboration, particularly when factual fidelity is important. More broadly, the categorizations of augmentation and hallucination patterns provide an analytical basis for examining how AI modifies source content when instructed to produce increasingly longer translations.

For future research, it would also be valuable to compare different prompt formulations, such as requests for elaboration, stylistic enhancement, persuasion, fluency, or strict source fidelity, to determine which types of instructions are more likely to trigger hallucinations. Further studies could additionally examine human post-editing of augmented AI translations to investigate whether hallucinated or unjustified additions can be reliably detected and removed, and whether translators recognize the increasing risk of hallucinations.

References

- Amini, M., Ravindran, L., & Lee, K.-F. (2024). Implications of using AI in translation studies: Trends, challenges, and future direction. *Asian Journal of Research in Education and Social Sciences*, 6(1), 740–754. <https://doi.org/10.55057/ajress.2024.6.1.67>
- Anh-Hoang, D., Tran, V., & Nguyen, L. M. (2025). Survey and analysis of hallucinations in large language models: Attribution to prompting strategies or model behavior. *Frontiers in Artificial Intelligence*, 8. <https://doi.org/10.3389/frai.2025.1622292>

- Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., Duddumpudi, R. T. S., Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*, 15(4). <https://doi.org/10.7759/CUREUS.37432>
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *Proceedings of the 13th International Joint Conference on Natural Language Processing*, 1, 675–718. <https://doi.org/10.18653/v1/2023.ijcnlp-main.45>
- Benkirane, K., Gongas, L., Pelles, S., Fuchs, N., Darmon, J., Stenertorp, P., Adelani, D. I., & Sánchez, E. (2024). Machine translation hallucination detection for low and high resource languages using large language models. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 9647–9665). Association for Computational Linguistics (ACL). <https://doi.org/10.18653/v1/2024.findings-emnlp.564>
- Bogucki, Ł. (2026). *Toward a generative AI turn in translation and interpreting studies*. Springer Nature. <https://doi.org/10.1007/978-3-032-12926-0>
- Bozkurt, A. (2024). Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI. *Open Praxis*, 16(2), 111–118. <https://doi.org/10.55982/openpraxis.16.2.661>
- Förster, S., & Skop, Y. (2025). Between fact and fairy: Tracing the hallucination metaphor in AI discourse. *AI & Society*, 41(3), 1685–1698. <https://doi.org/10.1007/S00146-025-02392-W>
- Frankenberg-Garcia, A. (2009). Are translations longer than source texts? A corpus-based study of explicitation. In *Corpus Use and Translating* (pp. 47–58). John Benjamins. <https://www.jbe-platform.com/content/books/9789027291066-btl.82.05fra>
- Gogoulou, E., Zahra, S., Guillou, L., Dürlich, L., & Nivre, J. (2025). Can LLMs detect intrinsic hallucinations in paraphrasing and machine translation? In *Proceedings of the Fourth Workshop on Generation, Evaluation and Metrics (GEM²)* (pp. 161–177). Association for Computational Linguistics. <https://aclanthology.org/2025.gem-1.13/>
- Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). Hallucinations in large multilingual translation models. *Transactions of the Association for Computational Linguistics*, 11, 1500–1517. https://doi.org/10.1162/tacl_a_00615
- Heinisch, B. (2025). Text-generative large language models in specialized translation: Not only suitable for translation. In K. Walter & M. Agnetta (Eds.), *Applying Artificial Intelligence in Translation* (pp. 103–116). Taylor & Francis.
- Jamaluddin, J., Gaffar, N. A., & Din, N. S. S. (2023). Hallucination: A key challenge to Artificial Intelligence-Generated writing. *Malays Fam Physician*, 18, 68. <https://doi.org/10.51866/LTE.527>
- Jesson, A., Beltran-Velez, N., Chu, Q., Karlekar, S., Kossen, J., Gal, Y., Cunningham, J. P., & Blei, D. (2024). Estimating the hallucination rate of generative AI. *Advances in Neural Information Processing Systems*, 37, 31154–31201. <https://doi.org/10.48550/arXiv.2406.07457>
- Jiao, H., Li, X., Ding, J., & Zhang, X. (2025). Investigating ChatGPT translation hallucination from an embodied-cognitive translatology perspective. *International Conference on Brain Inspired Cognitive Systems*, 15498 LNAI, 117–126. https://doi.org/10.1007/978-981-96-2885-8_12

- Krüger, R. (2025). Prelude: The birth of artificial intelligence out of the spirit of translation? In K. Walter & M. Agnetta (Eds.), *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena*. Taylor & Francis.
- Liu, C., & Lin, C. (2026). Are AI hallucinations the same as error information? A qualitative exploration of hallucination tolerance and resistance strategies. *Online Information Review*, 50(2), 390–408. <https://doi.org/10.1108/OIR-04-2025-0233>
- Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI hallucinations: A misnomer worth clarifying. 2024 *IEEE Conference on Artificial Intelligence (CAI)*, 133–138. <https://doi.org/10.1109/CAI59869.2024.00033>
- Naser, M. Z. (2026). Hallucinations in generative artificial intelligence and large language models: Tests, datasets, detection and correction methods. *Language Resources and Evaluation*, 60(64), 64–. <https://doi.org/10.1007/s10579-026-09938-4>
- Paradowska, U. (2025). Using LLMs and NMT to identify and transcreate culture-specific items from English to Polish. In K. Walter & M. Agnetta (Eds.), *Applying Artificial Intelligence in Translation* (pp. 61–80). Taylor & Francis.
- Shi, Y., Xu, H., Kwok, H. L., & Liu, K. (2024). ChatGPT in professional translation: A double-edged sword—insights from Chinese translators on capabilities, concerns, and future prospects. In S. Sun, K. Liu, & R. Moratto (Eds.), *Translation and Interpreting in the Age of Artificial Intelligence*. Taylor & Francis.
- Sun, S., Liu, K., & Moratto, R. (2025). Introduction: Navigating the paradigm shift—Translation Studies in the age of AI. In S. Sun, K. Liu, & R. Moratto (Eds.), *Translation studies in the age of artificial intelligence* (pp. 1–17). Taylor & Francis. <https://doi.org/10.4324/9781003482369-1>
- Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03811-x>
- Tang, Z., Chatterjee, R., & Garg, S. (2025). Mitigating hallucinated translations in large language models with hallucination-focused preference optimization. *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 3410–3433. <https://doi.org/10.18653/v1/2025.naacl-long.175>
- Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, 3(2), 258–262. <https://doi.org/10.1515/jccall-2023-0008>
- Wu, X., Liu, H., Zhou, J., Zhao, X., Xu, L., Wang, L., Luo, W., & Zhang, K. (2025). Challenging multilingual LLMs: A new taxonomy and benchmark for unraveling hallucination in translation. *ArXiv Preprint ArXiv*. <https://arxiv.org/pdf/2510.24073>
- Xu, W., Agrawal, S., Briakou, E., Martindale, M. J., & Carpuat, M. (2023). Understanding and detecting hallucinations in neural machine translation via model introspection. *Transactions of the Association for Computational Linguistics*, 11, 546–564. https://doi.org/10.1162/tacl_a_00563
- Yılmaz, S. (2026). Hallucination in Turkish–Arabic LLM translation outputs: A text-type based translation studies analysis. *Qədim Diyar*, 8(8), 16–33. <https://doi.org/10.36719/2706-6185/62/16-33>