LEAP: Learned Entity Alignment Prompting for Consistent Jargon Translation

Jared Ni*, Sibi Raja*, Aditi Raju*[*]
Massachusetts Institute of Technology
Harvard University

Abstract

We introduce Learned Entity Alignment Prompting (LEAP), a multi-step framework designed to enhance translation consistency in large language models (LLMs). LEAP utilizes named entity recognition and a translation oracle to construct domain-specific entity mappings, which are integrated into model prompts to ensure consistent translations across queries. To evaluate consistency, we propose a novel metric, Jargon Translation Consistency (JTC), alongside established metrics such as Jaccard Similarity and chrF++. Experiments conducted on domain-specific datasets (legal and medical) demonstrate that LEAP significantly improves consistency metrics compared to regular prompting methods by up to 2.1x. Our analysis reveals that LEAP's effectiveness depends on the linguistic capabilities of the LLM, and while it excels in maintaining domain-specific terminology, a trade-off exists between translation fluency and consistency. These findings underscore LEAP's potential to enhance the fidelity of machine translations in critical applications, such as medical and legal fields. Our code is available at https://github.com/jared-ni/LEAP

1 Introduction

Maintaining translation consistency—particularly for proper nouns and domain-specific terminology (jargon)—is a significant challenge in machine translation. Inconsistent translations, especially in fields that require high language accuracy (such as justice and medicine), can lead to confusion, lack of clarity, and incorrect interpretations. Existing approaches such as Delta (Wang et al., 2024) have attempted to address this issue using multi-level memory structures to store and retrieve contextual information. While they can enhance translation consistency, these methods introduce large memory and retrieval overhead, thus limiting their efficiency and scalability, especially as the translation task grows or when resources are constrained. We propose a novel framework, Learned Entity Alignment Prompting (LEAP), that uses named entity recognition (Sang and Meulder, 2003) and a translation oracle to generate domain-specific entity mappings to prompt engineer the model for greater translation consistency without reliance on model architectures and further overhead, thus reducing workload and ensuring translation consistency across languages and domains. See Figure 1 for an example of how named entity recognition works.

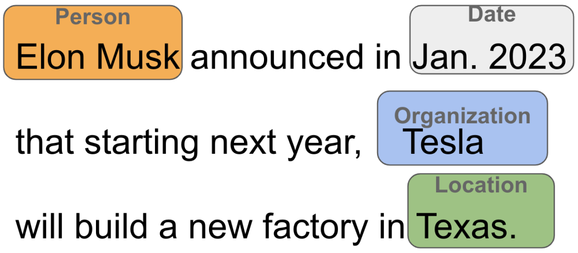


Figure 1: Named Entity Recognition example

2 Related Work

Large language models generally tend to struggle with generating consistent responses. "Are Large Language Models Consistent over Value-laden Questions?" shows that models give inconsistent responses to controversial questions and that fine-tuning models can actually increase inconsistency (Moore et al., 2024). Translation consistency is an even more specific challenge, which researchers have explored through various methodologies. Delta (Wang et al., 2024), for instance, utilizes multi-level memory components—including Proper Noun Records, Bilingual Summary, Long-Term Memory, and Short-Term Memory—to maintain consistency and enhance translation quality. Additionally, the paper "Adaptive Machine Trans-

[*]* indicates equal contribution.

lation with Large Language Model" (Moslem et al., 2023) uses in-domain sentence pairs and fuzzy matches to improve high-resource language translation. These research methods are effective but suffer from high memory requirements, hindering their scalability. Also, note that these models are limited to specific language pairs. We aim to use LEAP as an improvement by referencing a concrete entity mapping during the translation process, which would eliminate the need for maintaining complex memory hierarchies.

3 Methodology

Our LEAP workflow begins as follows: first, we take a corpus entry and apply named entity recognition to it to extract the key jargon terms. Once these jargon terms (entities) are extracted, they are passed into a translation oracle to get the entities translated into the target language. With these translations, a mapping from the source language to the destination language is constructed and appended to the prompt of the LLM. The LLM is queried k times, and ideally, our LEAP technique (the translated entity mapping) produces consistent translations across each of the k generations. See Figure 2 for an example of LEAP in action (note that the figure serves for demonstration purposes to showcase LEAP from Chinese to English, but we translate English to other languages in our actual experiments). And, see Figure 3 for a visualization of the LEAP workflow. Note that the following subsections explain components of the workflow in more detail.

3.1 Corpus

To test the consistency of translation, we focus on documents from the medical and legal domains. The PMC Patients dataset contains around 167,000 patient summaries extracted from case reports in PubMed Central (PMC) (Zhao et al., 2023). This dataset is a great way to test consistency in the medical domain, where it is crucial to be consistent in translations of medical terms to avoid ambiguity, misinterpretation, and misdiagnosis.

Similarly, CaseHOLD is a law dataset containing around 53,000 legal case contexts, based on EU legislative documents (Zheng et al., 2021). It is useful to evaluate legal document translation consistency as it contains pieces of text that contain several legal terms and phrases. Oftentimes, legal texts do include domain-specific jargon, and inconsistent translations can undermine the meaning and integrity of the interpretation, so this seems to be an interesting domain to test LEAP on.

Note that we test 500 examples from each dataset on each model and destination language, shortening each piece of text to the first 50 words.

3.2 Models

We use three large language models for our experiment: GPT 4o-mini (OpenAI, 2024), Claude 3.5 Haiku (Anthropic, 2024), and Mistral 7B (Jiang et al., 2023). We chose GPT, and Claude models as they are, within SOTA model families at the time we undertook this experiment. The choice of Mistral was fueled by wanting to test an open-source model that has fewer parameters than the GPT or Claude models, which are estimated to have hundreds of billions of parameters.

3.3 Languages

For our experiments, we set our source language to be English, and we set French and Simplified Chinese as our destination languages. Simplified Chinese is a form of Chinese that, as the name suggests, is simplified than regular Chinese, with fewer strokes being used on characters. Our choices for these two target languages were made because we wanted to explore how LEAP performs with different types of languages, as English and French have some inherent similarities within language roots, but Chinese is quite different from English.

3.4 Metrics for Evaluation

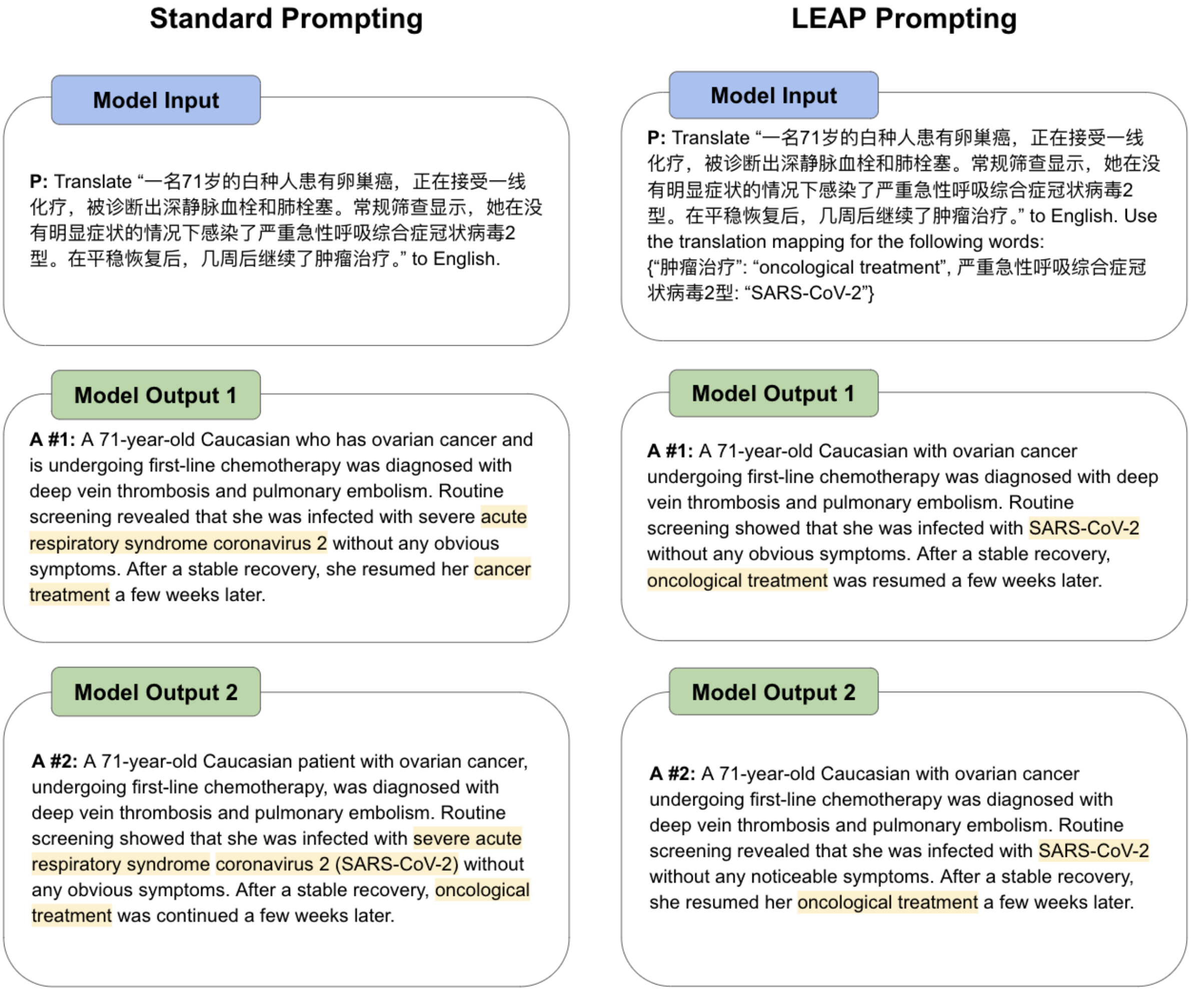
To evaluate the consistency and quality of translations, we employed three complementary metrics: Jargon Translation Consistency (JTC), Jaccard Similarity, and chrF++.

3.4.1 JTC

JTC is a novel metric that we contribute as part of this work in addition to the LEAP workflow. JTC aims to measure how consistently domain-specific terminology is translated across multiple translations. Note that domain-specific terminology refers to the jargon terms (i.e. entities). JTC is calculated as:

$$JTC = 1 - \frac{\sum_{e \in E}(K \cdot c_e - \sum_{t \in T} c_{t,e})}{K \cdot \sum_{e \in E} c_e} \quad (1)$$

In the above expression, E refers to the set of entities (the jargon terms), T is the set of translations,



P: Translate “一名71岁的白种人患有卵巢癌，正在接受一线化疗，被诊断出深静脉血栓和肺栓塞。常规筛查显示，她在没有明显症状的情况下感染了严重急性呼吸综合症冠状病毒2型。在平稳恢复后，几周后继续了肿瘤治疗。” to English.

A #1: A 71-year-old Caucasian who has ovarian cancer and is undergoing first-line chemotherapy was diagnosed with deep vein thrombosis and pulmonary embolism. Routine screening revealed that she was infected with severe acute respiratory syndrome coronavirus 2 without any obvious symptoms. After a stable recovery, she resumed her cancer treatment a few weeks later.

A #2: A 71-year-old Caucasian patient with ovarian cancer, undergoing first-line chemotherapy, was diagnosed with deep vein thrombosis and pulmonary embolism. Routine screening showed that she was infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) without any obvious symptoms. After a stable recovery, oncological treatment was continued a few weeks later.

P: Translate “一名71岁的白种人患有卵巢癌，正在接受一线化疗，被诊断出深静脉血栓和肺栓塞。常规筛查显示，她在没有明显症状的情况下感染了严重急性呼吸综合症冠状病毒2型。在平稳恢复后，几周后继续了肿瘤治疗。” to English. Use the translation mapping for the following words: {“肿瘤治疗”: “oncological treatment”, 严重急性呼吸综合症冠状病毒2型: “SARS-CoV-2”}

A #1: A 71-year-old Caucasian with ovarian cancer undergoing first-line chemotherapy was diagnosed with deep vein thrombosis and pulmonary embolism. Routine screening showed that she was infected with SARS-CoV-2 without any obvious symptoms. After a stable recovery, oncological treatment was resumed a few weeks later.

A #2: A 71-year-old Caucasian with ovarian cancer undergoing first-line chemotherapy was diagnosed with deep vein thrombosis and pulmonary embolism. Routine screening revealed that she was infected with SARS-CoV-2 without any noticeable symptoms. After a stable recovery, she resumed her oncological treatment a few weeks later.

Figure 2: Example of using LEAP to achieve consistent translations from Chinese to English

c_e represents the count of some entity e in source the text, $C_{t,e}$ represents the count of a translated entity in some translation t, and K is a hyperparameter represents the number of times we prompt the LLM (so K is the total number of translations as we have K translations to analyze the consistency of).

Note that higher JTC scores indicate better consistency as the second term essentially captures how inconsistent the translations are. Also, note that JTC penalizes both missing and extra occurrences of the translated terms.

3.4.2 Jaccard Similarity

Jaccard Similarity (Jaccard, 1901) is a common metric used to calculate similarity between two sets. We employ Jaccard between each pair of translations, and then take the average of all the Jaccard similarities:

$$Jaccard(T_1, T_2) = \frac{|T_1 \cap T_2|}{|T_1 \cup T_2|} \tag{2}$$

$$\text{JaccardAvg} = \frac{\sum_{i,j} \text{Jaccard}(T_i, T_j)}{\binom{n}{2}} \tag{3}$$

It is important to note though that the translated text in Simplified Chinese is preprocessed differently than the translated text in French because of the way the Chinese language inherently works. For Chinese text, we use Jieba segmentation instead of just splitting on words because some characters in Chinese represent whole words themselves. Thus, a simple splitting based on white spaces in between words is not sufficient. However, this does not change the way Jaccard is calculated, only ensuring that we correctly calculate Jaccard by splitting up on words.

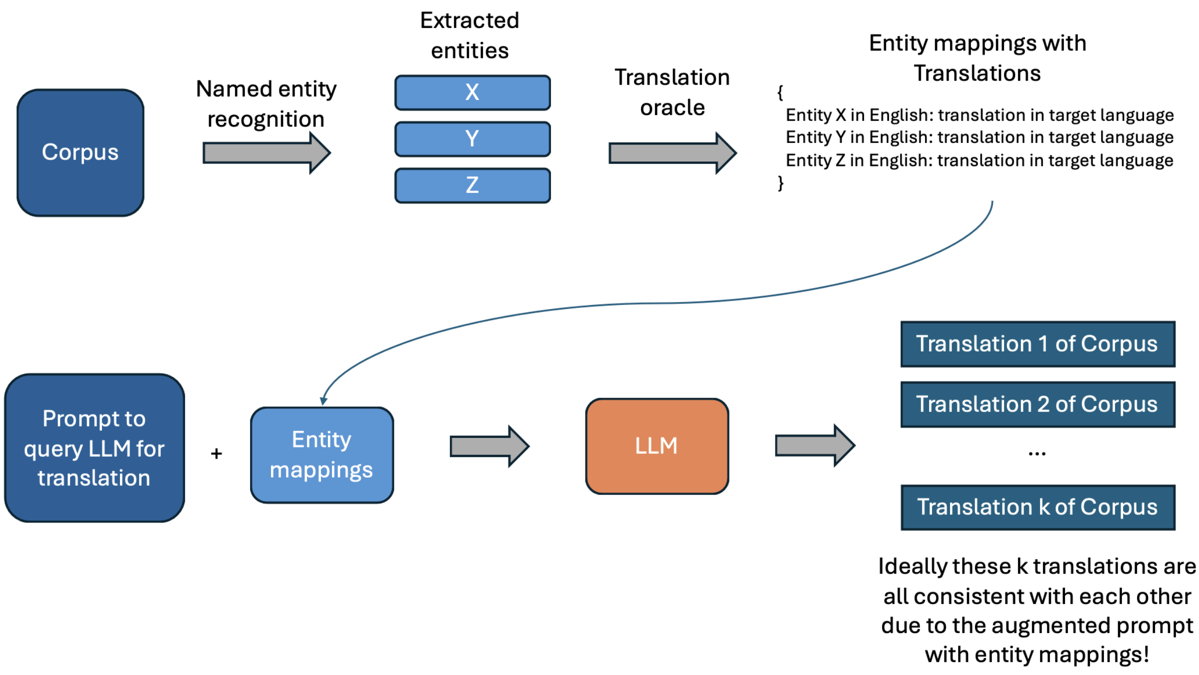

Corpus
Named entity recognition
Extracted entities
X
Y
Z
Translation oracle
Entity mappings with Translations
{
Entity X in English: translation in target language
Entity Y in English: translation in target language
Entity Z in English: translation in target language
}
Prompt to query LLM for translation
+
Entity mappings
LLM
Translation 1 of Corpus
Translation 2 of Corpus
...
Translation k of Corpus
Ideally these k translations are all consistent with each other due to the augmented prompt with entity mappings!

Figure 3: LEAP workflow

3.4.3 chrF++

chrF++ (Popović, 2015) is a character-level statistic that compares the similarity of text against a ground truth reference text. For the purposes of our experiment, we pass in the entire English corpus to the translation oracle and treat the output as the ground truth. See the equation below for how chrF++ is calculated:

$$chrF++ = (1+\beta^2)\cdot\frac{prec\cdot rec}{(\beta^2\cdot prec)+rec} \quad (4)$$

Here, $prec$ represents the n-gram precision, rec represents the character n-gram recall, β represents the weight of recall against precision (which is 2), and the n-gram size is 6.

3.5 Other

To implement our LEAP workflow, we used APIs to access each of the LLMs mentioned in section 3.3. We used Google Translate as the translation oracle for our LEAP workflow, specifically the Googletrans library (Community, 2024). Additionally, we used Spacy to implement named entity recognition, with spacy (AI, 2024) models already trained on both medical and law domains.

Because we call LLM APIs instead of running models locally, we required no significant amount of compute. In fact, no GPUs are required to run our experiments. Any modern CPU suffices.

Also, note that we set $K = 3$ for all our experiments for simplicity and low LLM API usage costs.

4 Experimental Results

We conducted jargon translation consistency experiments across three models (Claude, GPT4o, Mistral) and two langauges (French, Simplified Chinese). Table **??** displays average JTC scores, average Jaccard similarities, and average chrf++ scores for both regular and LEAP frameworks for each of the four experiments conducted with Claude, GTP4o, and Mistral: translating law data from English to French, medical data from English to French, law data from English to Chinese, and medical data from English to Chinese.

Figure 4 compares the average JTC score, Jaccard similarity, and chrf++ score between the regular and LEAP frameworks. Across all models, languages, and datasets, we observe that the JTC score is consistently higher for translations using the LEAP framework than for translations using the regular framework. Claude and GPT4o show significant improvements in JTC score, while Mistral shows less significant increases. For Jaccard similarity, we observe higher Jaccard similarity for translations using the LEAP framework and the GPT4o model, while Claude and Mistral translations show no significant improvement between

Average JTC Scores							
		Claude		GPT4o		Mistral	
		Regular	LEAP	Regular	LEAP	Regular	LEAP
En → French	Law	0.439	**0.824**	0.584	**0.838**	0.395	**0.419**
	Medical	0.606	**0.956**	0.637	**0.869**	0.455	**0.471**
En → Chinese	Law	0.396	**0.846**	0.499	**0.824**	0.265	**0.276**
	Medical	0.474	**0.809**	0.518	**0.859**	0.224	**0.239**
Average Jaccard Similarities							
		Claude		GPT4o		Mistral	
		Regular	LEAP	Regular	LEAP	Regular	LEAP
En → French	Law	**0.692**	0.577	0.820	**0.855**	**0.759**	0.735
	Medical	**0.771**	0.625	0.846	**0.904**	0.802	**0.813**
En → Chinese	Law	0.530	**0.576**	0.747	**0.785**	0.617	**0.619**
	Medical	0.593	**0.630**	0.762	**0.830**	0.590	**0.599**
Average chrf++ Scores							
		Claude		GPT4o		Mistral	
		Regular	LEAP	Regular	LEAP	Regular	LEAP
En → French	Law	**3.731**	2.425	**1.611**	1.611	**1.920**	1.904
	Medical	**1.647**	1.614	1.551	**1.579**	**1.925**	1.854
En → Chinese	Law	**5.336**	3.432	**3.829**	3.672	2.551	**2.722**
	Medical	**3.411**	2.353	**4.744**	4.126	2.005	**2.138**

Table 1: JTF, Jaccard, and chrf++ Scores: Regular vs. LEAP Frameworks

the LEAP framework and the regular framework. Finally, for chrf++ scores, translations using the regular framework have higher scores than LEAP framework translations.

Figure 5 displays a five-point summary of the JTC scores calculated for the regular and LEAP frameworks. We observe that across all models, languages, and datasets, the JTC score is consistently higher for translations using the LEAP framework than for translations using the regular framework. Although there is overlap in the interquartile range, the median and the inter-quartiles ranges of the JTC score are always higher with the LEAP framework.

Figure 6 compares the distributions of JTC scores for the regular and LEAP frameworks. We observe that across all models, languages, and datasets, the distribution of JTC scores peaks further to the right for the LEAP framework than the regular framework. This indicates that the LEAP framework translations produce higher JTC scores. Once again, the improvement in score distribution is very significant in the Claude and GPT4o models but less visible in Mistral.

5 Analysis and Discussion

This section analyzes the factors that contribute to the improvement in consistency performance, which we demonstrate via 1) the Jaccard similarity metric and 2) the Jargon Translation Consistency (JTC) score, our novel metric designed to evaluate the fidelity of text translations in the context of preserving domain-specific terminologies and named entities. In our experiments, LEAP prompting consistently outperformed regular prompting across all three LLM models (GPT4o-mini, Claude, Mistral) in realizing higher JTC scores in both legal and medical datasets, across both target languages we used (French, Simplified Chinese). These results suggest that integrating entity-aware mechanisms into translation workflows can significantly enhance the fidelity and utility of machine translations in critical fields.

5.1 Named Entity Preservation

One of the significant advantages of LEAP is the explicit handling of named entities. Domain-specific terms, especially in the context of legal references or medical jargon, are often mistranslated or paraphrased inconsistently by general-purpose models. This inconsistency is especially prevalent when multiple translations of the same jargon are required, across references in one or multiple translated documents. LEAP mitigates this issue by enforcing predefined mappings for these terms dur-

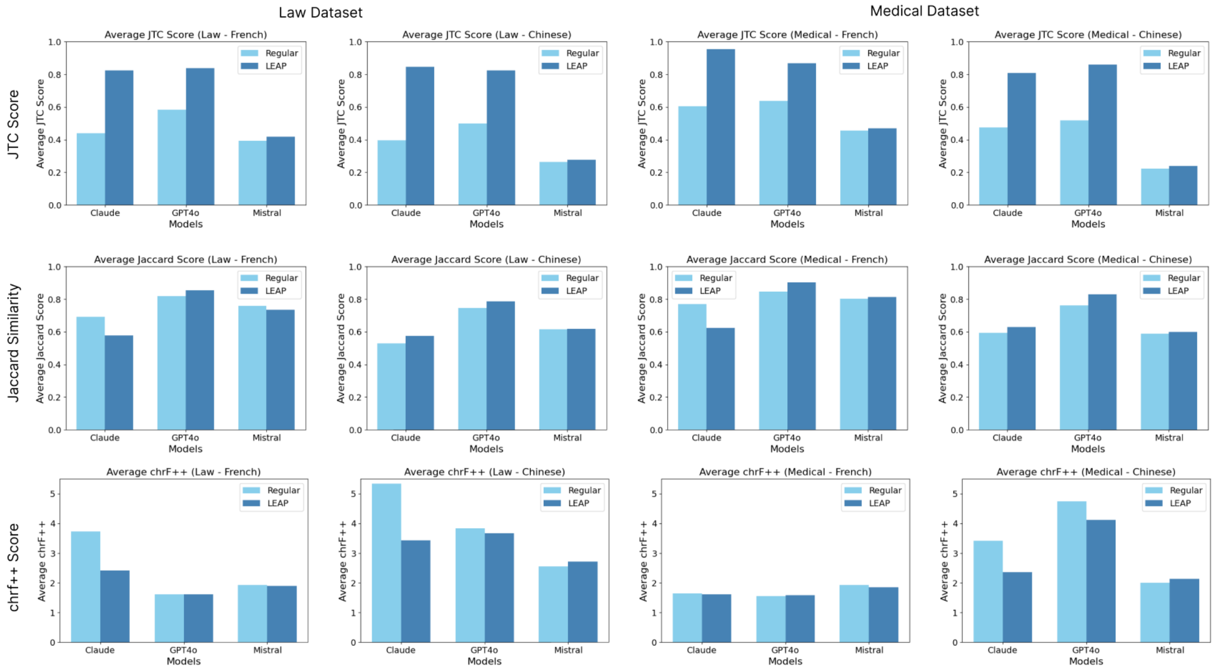


Figure 4: JTC, Jaccard, and chrf++ Scores: Regular vs. LEAP Frameworks

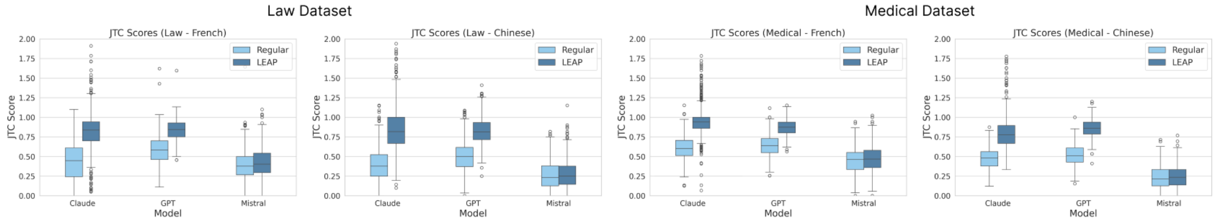


Figure 5: 5-Point Summaries of JTC Scores: Regular vs. LEAP Frameworks

ing the translation oracle-calling process. By calling named entity recognition models specifically trained for these domains such as Spacy (AI, 2024), we can accurately detect these words and phrases within their domain context. By calling the translation oracle, we can then find the ground truth values of these domain-specific languages, and accurately and consistently represent them in the target language.

5.2 Increased translation alignment

Legal and medical texts present unique challenges due to their reliance on precise terminology and complex jargon. LEAP's explicit mappings prevent the dilution of such terms, ensuring they are consistently and accurately represented in the target language. By establishing pre-defined ground truth values for all identified named entities as a mapping, LEAP avoids the variability inherent in a general-purpose probabilistic language model and ensures semantic consistency, reducing ambiguity within domain-specific language that requires high levels of precision across one or multiple translations, which increases the alignment of translated texts with each other and the source language.

The improvement of alignment between translations with each other and the source text is highlighted in the general improvement to the JTC score Jaccard similarity metrics after incorporating LEAP. LEAP reduces noise and improves the sensitivity across translated texts, and it's evident in the drastic improvements to the JTC score, as it penalizes mismatches between the ground-truth translation mapping for named entities and the varying translations. Across the k translations for each text, JTC's loss function punishes translation variance for each named entity. By establishing pre-defined named-entity mapping, LEAP drastically reduces variability, either when translating an entity within the same text or across multiple machine translations.

Jaccard similarity metric penalizes the set dif-

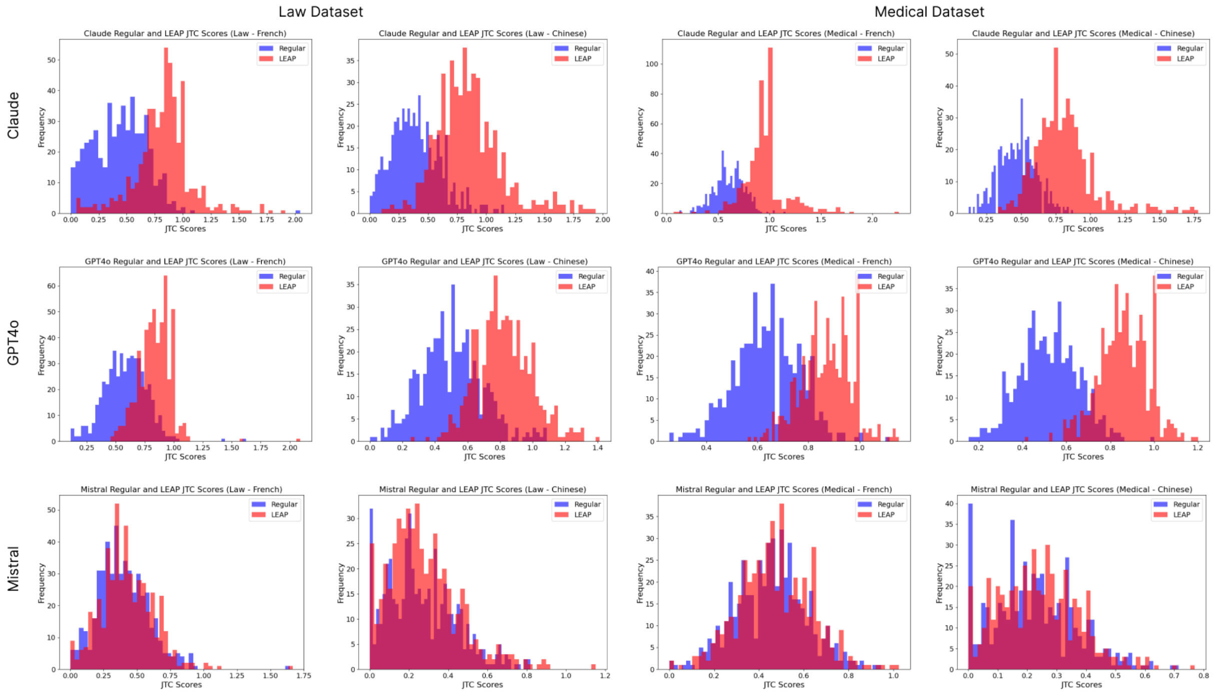


Figure 6: Distributions of JTC Scores: Regular vs. LEAP Frameworks

ferences between a baseline ground-truth translation of the entire source text and each of the k machine translations improving translation fidelity. Because of the increased consistency of the translations from the named entity oracle mapping, the set overlap between the ground-truth translation and the machine-generated translations increases, leading to higher Jaccard similarity scores. This demonstrates that LEAP's explicit handling of named entities reduces variability in how these entities are translated, ensuring greater uniformity across translations and with the source text. By minimizing inconsistencies and paraphrasing errors that typically arise in domain-specific texts, LEAP significantly enhances the fidelity of machine translations, aligning them more closely with the source text and improving both token-level consistency within domain-specific context and translations' semantic accuracy.

5.3 Context-Consistency Trade-off

The chrF++ metric, as observed from the results, does not show a consistent improvement after incorporating the LEAP framework. This indicates a potential trade-off between maintaining high-level fluency and semantic alignment (as measured by chrF++) and improving domain-specific consistency and precision (as measured by JTC and Jaccard metrics). While LEAP effectively preserves jargon and entity consistency, its focus on strict mapping of named entities without considering the full context of the other words in the source text might lead to rigid, or in a very small subset of cases, nonsensical translations in a given source context, which could impact metrics like chrF++ that uses an n-gram model ($n = 6$). This suggests that LEAP's strengths lie primarily in improving alignment for specialized domains, but further adjustments might be needed to balance domain-specific precision with broader text-level fluency, especially when targeting applications where both aspects are equally critical. The balance between entity consistency and language context would be an interesting future research direction beyond this work.

5.4 LLM Performance Limitations

One particular trend we identified in the data is that when the LLM model is more powerful, LEAP leads to greater performance variance compared to the performance of their regular prompting counterparts. For instance, in the case of the JTC score and Jaccard similarity metrics, Claude 3.5 Haiku and GPT-4o-mini models exhibited significantly greater improvements when applying LEAP compared to regular prompting, while Mistral showed smaller improvements. This suggests that LEAP performs better when the model has sufficient linguistic capacity to fully utilize the named entity mapping provided.

In contrast, for the chrF++ metric, models for whom regular prompting achieved the highest score on average (Claude for Law datasets, and GPT for Medical Dataset in Chinese) experienced a larger reduction in performance after applying LEAP than GPT or Mistral did. This indicates that the rigid nature of LEAP may have a greater contextual influence on translation quality when the LLM's inherent translation capabilities surpass the context-free translation oracle performance that LEAP enforces. Thus, the quality of the translation oracle and the need to enforce the linguistic context of named entities within the translations are interesting areas of future exploration.

6 Conclusion

In this paper, we propose the LEAP framework to improve the consistency of translations generated by large language models by using named entity recognition. We also contribute the JTC metric as a way to measure the consistency between multiple texts. Our experiments and results show that LEAP can greatly improve the JTC scores of translations, which indicates that consistency can improve as well. However, our methods do not perform as well with regards to Jaccard Similarity and chrF++, which signifies that more work may be needed before LEAP can consistently output consistent generations (no pun intended). Nonetheless, our work demonstrates a promising direction for improving translation consistency. A natural direction of future work could be to explore how high consistency can be maintained while simultaneously improving translation accuracy.

References

Explosion AI. 2024. spacy layout: A library for document layout processing and nlp. https://github.com/explosion/spacy-layout. Accessed: 2024-12-10.

Anthropic. 2024. Claude 3.5: Advancements in ai and computer use. https://www.anthropic.com/news/3-5-models-and-computer-use. Accessed: 2024-12-10.

Google Translate API Community. 2024. googletrans: Free google translate api for python. https://pypi.org/project/googletrans/. Accessed: 2024-12-10.

Paul Jaccard. 1901. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*, 37:241–272.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

J. Moore, T. Deshpande, and D. Yang. 2024. Are large language models consistent over value-laden questions? *arXiv*.

Y. Moslem, R. Haque, J. D. Kelleher, and A. Way. 2023. Adaptive machine translation with large language models. *arXiv*.

OpenAI. 2024. Gpt-4o: Enhanced optimization for language models. https://openai.com/index/hello-gpt-4o/. Accessed: 2024-12-10.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. *ACL Anthology*.

Y. Wang, J. Zeng, X. Liu, D. F. Wong, F. Meng, J. Zhou, and M. Zhang. 2024. Delta: An online document-level translation agent based on multi-level memory. *arXiv*.

Zhengyun Zhao, Qiao Jin, Fangyuan Chen, Tuorui Peng, and Sheng Yu. 2023. A large-scale dataset of patient summaries for retrieval-based clinical decision support systems. *Scientific Data*, 10:909.

Lucia Zheng, Neel Guha, Brandon R. Anderson, Peter Henderson, and Daniel E. Ho. 2021. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law (ICAIL '21)*, pages 159–168, São Paulo, Brazil. ACM.

7 Impact Statement

Our Learning Entity Alignment Prompting (LEAP) framework has great potential for both the advancement of machine translation technology and its numerous societal applications. Our work focused on applying LEAP to the medical and legal domains, both of which are domains where language and interpretation are critical. By improving translation consistency, LEAP could greatly improve the usage of large language models and generative AI as a whole in contexts where precision is absolutely critical. In the medical setting, consistent machine translations can greatly reduce the risk of miscommunication and medical errors especially when language barriers exist between patients and healthcare workers or among healthcare workers themselves.

For example, sometimes patients may receive treatment from different parts of the world to seek specialized doctors. In these situations, medical records may need to be translated, which requires great precision and consistency. The same can be said for legal documents, if non-native speakers are present in courtrooms or documents from other parts of the world (and thus in other languages) are presented to the courtroom as evidence in a trial).

However, we acknowledge that an increasing usage of machine translation could affect the jobs available for humans in these roles, such as human translators and interpreters. While machine translation tools can reduce costs and increase the accessibility of translation services, employability may be negatively affected, which could place economic and financial burdens on various communities. Furthermore, if machine translation systems enhanced by LEAP or other similar tools/frameworks help achieve high consistency but make occasional errors, users might place excessive trust in such systems and disregard double-checking generated outputs, which could lead to severe consequences, especially in medical or legal contexts.

To address the aforementioned concerns, we recommend a few initiatives. First, we encourage establishing clear guidelines regarding the appropriate usage of LEAP-like systems in critical domains, which include the acknowledgment of their limitations and human supervision. We also recommend collaborative research to occur between machine learning researchers and professionals across various industries (medical, legal, etc.) to mitigate potential risks that might arise with the use of machine translation technologies.