Assessment of integrating LLM in website Localisation Service

Evgeniya Svintsova
School of Computing and Digital Media- London Metropolitan University London, UK
evs0187@my.londonmet.ac.uk

Sahar Al-Sudani
School of Computing and Digital Media- London Metropolitan University London, UK
s.al-sudani@londonmet.ac.uk

Mohmed Gahnem
School of Computing and Digital Media- London Metropolitan University
m.ghanem@londonmet.ac.uk

***Abstract*— This paper examines the integration of Large Language Models (LLMs) into website localisation services through a prompt-engineering-first approach. Traditional localisation methodologies - human translation, machine translation, and hybrid approaches - have established tradeoffs between quality, cost, and efficiency. The emergence of advanced LLMs offers a potential paradigm shift in addressing these challenges. The research presented explores how focusing on prompt engineering rather than post-translation editing can transform localisation workflows while maintaining high-quality outputs.**

Drawing on recent research in prompt engineering and machine translation efficacy, the paper establishes a theoretical framework for LLM implementation in localisation services. A practical case study involving the localisation of a cross-platform application demonstrates the implementation of this approach, including technical architecture, prompt design strategies, and testing methodologies. The findings indicate that LLM-powered localisation with well-engineered prompts can deliver comparable quality to specialised translation services while offering advantages in maintaining marketing tone, reducing implementation complexity, and supporting broader content creation needs across languages.

The analysis extends beyond theoretical considerations to provide a decision framework for selecting appropriate localisation tools based on specific project requirements. The research concludes that for projects with marketing-focused content and moderate translation volume, the LLM approach with focused prompt engineering represents an optimal solution compared to traditional translation services and dedicated localisation platforms.

Keywords—Localisation, Large Language Models , Prompt Engineering.

I. Introduction

Website localisation has become increasingly crucial for organisations looking to expand their global reach and engage with diverse audiences in their native languages. Traditional approaches to website localisation have typically followed three established methodologies: human translation, machine translation, and hybrid approaches. Each comes with inherent tradeoffs between quality, cost, and scalability that organisations must navigate based on their specific requirements and resources.

The emergence of advanced Large Language Models (LLMs) in recent years has introduced a potential paradigm shift in how localisation can be approached. These sophisticated AI systems offer new possibilities for achieving high-quality translations while potentially addressing many of the limitations associated with traditional methods. Unlike conventional machine translation systems that often struggle with context, nuance, and brand-consistent tone, LLMs demonstrate remarkable capabilities in understanding and generating natural language across multiple languages while maintaining stylistic consistency.

This paper examines how integrating LLMs into website localisation services can transform the traditional localisation workflow through a prompt-engineering-first approach. Rather than focusing on post-translation editing to correct machine-generated content, this methodology prioritises the refinement of input prompts to achieve higher quality initial translations. The approach represents a fundamental shift from output correction to input optimisation, with significant implications for efficiency, quality, and resource utilisation.

Drawing on recent research in prompt engineering and machine translation, this paper provides a comprehensive assessment of how LLMs can be effectively leveraged for website localisation. The analysis includes both theoretical foundations and a detailed case study of a cross-platform application implementation, offering insights for organisations considering LLM integration for their localisation needs. Specifically, this research utilises state-of-the-art LLM models such as ChatGPT (GPT-4), which has demonstrated superior contextual understanding and natural language generation capabilities compared to earlier machine translation systems. The prompt engineering approach we explore uses techniques including zero-shot prompting with rich contextual information, custom terminology glossaries, and iterative refinement strategies to optimise translation outputs before they are generated, rather than correcting them afterwards. This approach was implemented and tested in a real-world case study involving a React/TypeScript cross-platform application available on

Web, iOS, and Android platforms, using the i18next framework integrated with ChatGPT through the Crowdin platform. By examining the technical architecture, prompt design methodologies, and quantitative outcomes of this implementation, we provide a research-based framework for evaluating the potential of LLM-powered localisation for diverse project requirements.

By examining the advantages, challenges, and implementation considerations of this approach, the paper aims to provide a framework for evaluating whether an LLM-powered localisation strategy aligns with specific project requirements and organisational objectives. As digital marketplaces become increasingly global, effective localisation strategies that balance quality, efficiency, and scalability will continue to be essential competitive differentiators for businesses across sectors.

II. RELATED RESEARCH

Website localisation has traditionally followed three main methodologies as identified in the critical evaluation of related products and solutions. Human translation is widely recognised as the highest quality translation method, but it is the most time-consuming and expensive approach (O'Hagan and Ashworth, 2002). Machine translation is the cheapest and fastest method but can be less accurate, especially when dealing with specific (professional) or structurally complex language (Hutchins and Somers, 1992). Hybrid approaches combine the previous methods to provide the most accurate translated text, though they can be the most expensive as companies must pay for both approaches (Doherty, 2016).

The selection of an appropriate localisation strategy has historically depended on multiple factors including the company budget, business needs, text volume, project timeline, update frequency, technical requirements, and development team resources. These considerations were already recognised by Sandrini (2005) nearly two decades ago, when a framework for localisation decision-making was established that emphasised similar concerns. Sandrini recommended addressing crucial questions at the beginning of any localisation project, including market need, product affordability, delivery logistics, customer support resources, localisation and maintenance costs, system compatibility, and regulatory issues (Sandrini, 2005, p. 136-137). Sandrini established the importance of detailed translation briefs that specify the client's purpose explicitly at the project's outset. The research emphasises that documenting the client's specific goals for the localised website is a critical first step to ensure successful outcomes. Sandrini's research highlights that in website localisation, the function or purpose that the content serves for its target audience should take precedence over preserving the exact form or structure of the original content. The work suggests that successful localisation prioritises meeting the company's predefined objectives for the foreign language website over strict adherence to the source text. According to this perspective, the target text and its intended function should be the focus, with literal correspondence to the original content being secondary to achieving the desired outcomes in the new language environment. This principle justifies approaches that prioritise communicative intent over strict linguistic equivalence.

Recent research has established prompt engineering as a crucial digital competence in effectively leveraging generative AI technologies. Korzynski et al. (2023) describe prompt engineering as a practice centred on human language where individuals craft prompts by combining specific textual elements or tokens. This perspective positions prompt engineering as distinct from automated processes, situating it within human-computer interaction. According to their research, prompt engineering primarily focuses on the manual creation and refinement of prompts, which aligns it more closely with human-centred disciplines like conversational AI and human-computer interaction rather than traditional machine learning (Korzynski et al., 2023). This characterisation highlights how prompt engineering remains accessible to professionals without technical backgrounds in AI development.

Pourkamali and Sharifi (2024) conducted extensive research on how prompt engineering influences machine translation quality with LLMs. Their work revealed that "*the style, arrangement, and generally the configuration of prompts affect the quality of outputs*" (Pourkamali and Sharifi, 2024, p. 7). Through systematic experimentation with different prompting techniques, they tested both n-shot prompting and specialised translation-tailored frameworks, discovering that both prompt templates and example selection substantially impact results. Interestingly, they found that combining structured prompting frameworks with in-context learning examples often reduced translation efficiency due to increased input complexity, except in cases where models could handle diverse data patterns and lengthy texts. This suggests that simpler, more focused prompting approaches typically yield better results in translation scenarios, as most LLMs struggle when simultaneously processing both structured frameworks and multiple examples, becoming confused rather than enhanced by the additional complexity.

Yamada (2023) investigated ChatGPT's adaptability for Machine Translation (MT) through prompt engineering, identifying a critical limitation in traditional machine translation systems, their inability to incorporate translation specifications from the pre-production phase. Their research demonstrated that "*MTs have not incorporated functions to set or adjust the specifications of the translation...making it challenging for current MT to fit into practical translation workflows*" (Yamada, 2023, p. 197). By integrating specific parameters like purpose and target audience into prompts, Yamada found that LLMs could maintain marketing tone and transform culturally dependent expressions into appropriate equivalents that preserve communicative intent rather than literal wording. These findings establish prompt engineering as a practical method to achieve what translation theory has advocated for but conventional machine translation has struggled to implement, particularly in maintaining stylistic consistency across diverse content types.

Wang et al. (2024) provide comprehensive insights into how large language models can enhance translation quality through interactive methods. Their analysis highlights research by Chen et al. on iterative translation refinement using LLMs, which "*despite sometimes lowering traditional metric scores, showed improved quality in neural assessments and human evaluations. The process effectively reduced 'translationese' and demonstrated the LLM's capacity to enhance translations iteratively*" (Wang et al., 2024, p. 16). Their work also highlights significant advancements for domain-specific translation, noting how Moslem et al.'s methodology doubled the incorporation of technical terminology across multiple language pairs, and Zhang et al.'s findings on the critical importance of both quantity and quality of prompt examples for achieving accurate translations. These collective insights demonstrate the versatility of LLM-based approaches for addressing specialised translation challenges across various content types.
Another research highlights the growing significance of prompt engineering skills, with notable industry forecasts suggesting their substantial future impact. Korzynski et al. (2023) reference a striking prediction from Robin Li, Baidu's co-founder and CEO, who suggests that "*in ten years, half of the world's jobs will involve prompt engineering, and those who cannot write prompts will become obsolete*" (p. 26). This anticipated demand indicates the importance of developing prompt engineering capabilities as an essential digital competence for professionals working with generative AI technologies in localisation services. The research also suggests that organisations would benefit from investing in these capabilities, as properly designed prompts enable systems to deliver more refined and contextually appropriate outputs. The evolving research in this field points to a shift in localisation methodology that emphasises refining input approaches rather than correcting output, with meaningful implications for efficiency, quality, and resource utilisation in website localisation services.

III. System Implementation

The implementation of LLM-based localisation was conducted on a cross-platform productivity application Emery developed using React/TypeScript for web, iOS, and Android platforms. This application provided an ideal test case for implementing an AI-powered localisation strategy that prioritises prompt engineering over traditional post-editing workflows.
The localisation project encompassed approximately 1,247 text strings totaling around 2,000 words across the application interface, including navigation elements, feature descriptions, error messages, and marketing content. Implementation began in November 2024 with English as the source language and initially targeted French, German, and Spanish, and then expanded to include Russian, Rumanian, and Portuguese.
Key implementation goals included achieving high translation accuracy, maintaining consistent brand voice across all languages, reducing implementation and maintenance costs compared to traditional translation services, and creating a sustainable localisation workflow that could scale additional languages without linear resource increases. This scalable approach is economically justified, as traditional translation models typically require proportional increases in resources when adding each new language, whereas this LLM-powered system established reusable infrastructure that minimises additional setup costs for new languages. The ability to rapidly expand language offerings without proportional cost increases provided strategic flexibility and faster market entry capabilities, creating a competitive advantage. Furthermore, this approach aligned with broader sustainability principles by designing systems that could grow without corresponding growth in resource consumption, making it both environmentally and economically more sustainable in the long term.
This implementation represents a practical demonstration of the theoretical prompt-engineering-first approach, testing its efficacy in a production environment with measurable outcomes against established quality and performance metrics. Figure-1 depicts the designed system architecture.

Figure 1. System Architecture

The system architecture for the Emery application's localisation implementation was designed to create a seamless, maintainable workflow that prioritises prompt engineering while integrating with modern development practices. Each technology was selected for specific advantages it brings to the localisation process:

A. *Frontend Layer:*

- React/TypeScript: Selected for its component-based architecture which allows for modular internationalisation implementation and strong typing support that helps prevent localisation-related errors.

- CSS Modules: Utilised to ensure styling remains consistent across languages with different text lengths and character sets, preventing layout issues in localised versions.
- i18next: Chosen as the internationalisation framework due to its robust support for React applications, pluralisation handling, and flexible string management capabilities.

B. Localisation Management:

- Crowdin Platform: Implemented as the central translation management system due to its seamless integration with GitHub project repository and support for custom machine translation engines. The GitHub synchronisation was particularly valuable, allowing translation strings to stay automatically synchronised with code changes whilst also enabling completed translations to be pushed back to the codebase without manual intervention.
- ChatGPT Integration: Leveraged through Crowdin to provide AI-powered translations using custom prompt engineering approach.

C. Testing Environment:

- Storybook: Deployed to isolate and test UI components across different languages without requiring a full application build.
- Custom Language Environment: Created to allow testers and developers to toggle between languages easily during development and testing phases.
- Visual Validation: Implemented to detect layout issues across languages with different text lengths and character sets.

This architecture supports the prompt-engineering-first approach by creating a continuous feedback loop where prompt improvements can be tested immediately across all languages, enabling efficient refinement of translation quality through input optimisation rather than output correction.Unlike traditional localisation approaches that rely heavily on post-translation editing, the implemented system uses a prompt-engineering-first methodology as shown in Figure 2.

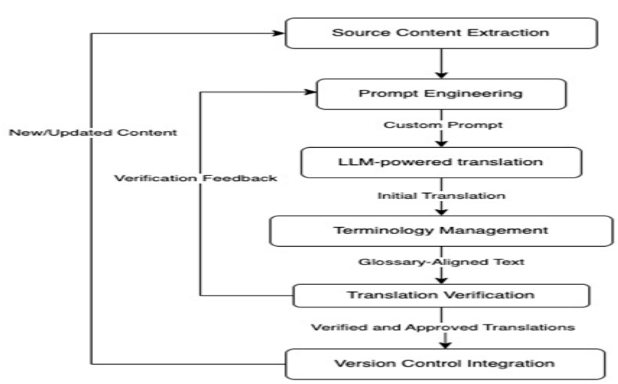

Figure 2. Emery localisation Workflow

The localisation workflow for Emery implements a prompt-engineering-first approach that shifts focus from traditional post-translation correction to input refinement through an iterative cycle. Each stage in the process serves a specific purpose in creating high-quality translations efficiently:

Source Content Extraction - All UI text elements are extracted from the application codebase into standardised JSON files. This extraction process creates a structured repository of content that can be processed consistently across languages. The JSON format was selected for its compatibility with both the frontend framework and translation management systems, providing a seamless data flow between development and localisation.

Prompt Engineering - This critical stage represents the cornerstone of the approach. Custom prompts are developed specifically for ChatGPT to ensure consistent tone, terminology, and marketing voice across all target languages. Rather than fixing translations after generation, significant effort is invested in refining these prompts to produce optimal initial results. This created a powerful cumulative effect where each prompt refinement benefited all languages simultaneously. Researchers categorised issues by type (terminology, tone, technical accuracy), allowing targeted prompt improvements. This approach proved significantly more efficient than traditional post-editing as the project expanded from three to seven languages, creating a self-improving system that enhanced translation quality across the entire application with each iteration.

LLM-Powered Translation - Translations are generated by feeding the engineered prompts along with source content to ChatGPT via the Crowdin integration. This process leverages advanced AI capabilities while maintaining control through carefully crafted prompts. The system generates translations for all required languages using consistent parameters, ensuring stylistic and terminological alignment across the multilingual content.

Terminology Management - A custom glossary is maintained to ensure consistent handling of application-specific terms across all languages. This glossary defines which terms should remain untranslated (e.g., product features or brand names) and which require specific translations. Specific lexemes from the glossary bypass the iterative refinement process, maintaining consistency for critical terminology.

Translation Verification - Generated translations are verified for accuracy and appropriateness. Unlike traditional workflows, when issues are identified, the focus shifts back to prompt refinement rather than directly editing the outputs. This creates an iterative cycle where prompt improvements lead to better initial translations in subsequent iterations. For example, if Russian translations contain stylistic inconsistencies, the prompt is refined to better capture Russian linguistic nuances rather than simply correcting the specific instances.

Version Control Integration - Once translations have been verified and approved through the iterative prompt refinement process, they are integrated into the version control system.

Seamless synchronisation between GitHub and Crowdin ensures translations stay current with application development. When new features are added or existing content is modified, the localisation workflow automatically captures these changes and initiates the appropriate translation processes.

The implementation of LLM-powered localisation for Emery involved a comprehensive approach that leveraged both technical infrastructure and carefully designed prompt.
The localisation workflow began by extracting all UI text elements into standardised JSON files. These files were then synchronised with Crowdin through its GitHub integration, creating a continuous pipeline for translation updates.
A custom prompt was developed that guided ChatGPT to produce translations that maintained the application's specific tone and terminology. The prompt was designed to provide rich context about the application domain and expected output style. For example, key phrases were added: "It's UI of a daily planning productivity app", "use shorter words when possible", "Use JSON keys to get additional information", "strictly refer to the translation glossary" (Figure 3).

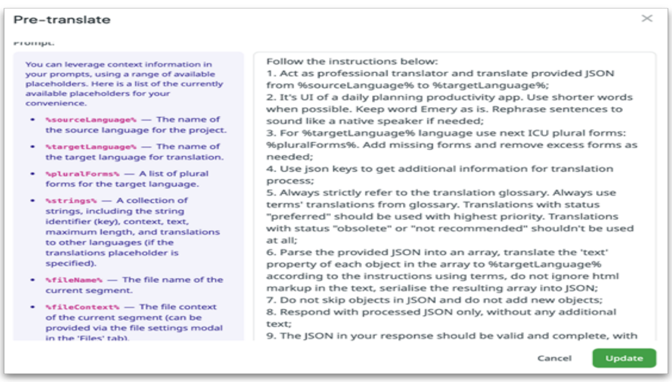

Figure 3. Screenshot of prompt from Crowdin platform

To ensure consistency in terminology across languages, a custom glossary was implemented in Crowdin. This glossary defined application-specific terms and their approved translations in each target language. The glossary served as an "authoritative reference" for ChatGPT, ensuring that technical terms maintained consistent translations throughout the application

Figure 4 shows a comparison of the same interface element in English and German, demonstrating how the localisation preserves both meaning and layout while adapting to language-specific characteristics.

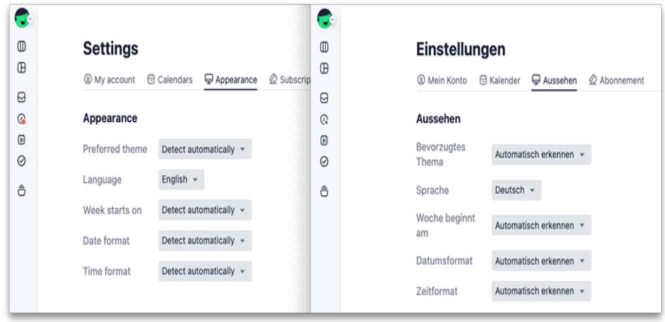

Figure 4. Side-by-side screenshot comparison showing the same UI elements in English and German

IV. Analysis and Discussion of Findings

The implementation of LLM-based localisation for the Emery application yielded significant findings regarding translation quality, efficiency, and comparative advantages over traditional localisation approaches. This chapter presents the key results from the implementation and provides a comparative analysis of existing market solutions.

The prompt-engineering-first approach demonstrated measurable improvements in translation quality through iterative refinement. Each iteration of the prompt addressed specific translation challenges identified during verification. Table 1 presents data showing the impact of prompt refinements on translation quality across multiple languages.

The data shows that the most significant quality improvement (18%) occurred when moving from the base prompt to Version 2, which added specific application context. This contextual information helped the LLM better understand the nature of the content being translated. Further refinements targeting specific aspects of translation such as plural forms of the words and enrichment glossary yielded additional quality gains, resulting in a total improvement of 44 percentage points from base to final version. All translations were verified by native speakers to ensure prompt effectiveness. These approved translations now serve as a contextual foundation for the LLM when adding new languages, significantly reducing the need for extensive native speaker proofreading in future localisation efforts.

Table 1. Translation Quality Improvement Through Prompt Refinement

Prompt version	Description	Average quality (%)	Russian (%)	Romanian (%)	Spanish (%)
Base prompt	Standard Crowdin prompt	56%	56%	54%	58%
Version 2	Added context: "It's UI of a daily planning productivity app"	74%	74%	72%	76%
Version 3	Added context: "Rephrase sentences to sound like a native speaker if needed"	86%	86%	83%	86%
Version 4	Added context: "Use Intensive Care Unit (ICU) plural forms for target language"	97%	97%	97%	97%
Version 5	Added context: "Always strictly refer to the translation glossary"	100%	100%	100%	100%

Following the initial evaluation of translation tools, it is valuable to contextualise these findings within the broader landscape of commercial localisation solutions. This section examines how the LLM-based approach compares to established market solutions.

The localisation market offers various specialised tools with different capabilities and target users. Each approach represents a different philosophy toward solving the localisation challenge, with varying implications for implementation complexity, resource requirements, and ongoing maintenance needs. This fragmentation in the market creates significant decision complexity for development teams seeking optimal solutions for their specific project requirements.

Table 2 provides a comprehensive comparison between the evaluated LMM approach and leading market solutions.

Feature	LLM Implementation	Weglot	IPinfo + i18next	Localazy	Geo Targetly
Auto-Translation	Yes	Yes	No	Yes	No
Translation Quality	Excellent	Good (requires proofreading)	Requires manual translation	Good (requires verification)	Requires manual translation
Translation Style	Marketing-focused through prompt engineering	Formal	Developer-controlled	Formal	Developer-controlled
Currency Handling	Yes (with code)	Yes	Yes (manual)	Yes	Yes (manual)
Implementation Complexity	Moderate	Low	High	Moderate	Moderate
Cost Structure	Usage-based (<$1/mo for medium projects)	Subscription (€15+/mo)	Free tier + subscription	Free tier + subscription	Starts at $25/mo
Multi-functionality	High	Low	Moderate	Low	Location-focused
Translation Improvement Method	Prompt engineering	Post-editing	Manual translation	Translator verification	Manual translation
Content Creation Capabilities	Extensive (blogs, articles, ads)	Limited	None	Limited	None
Future Content Strategy Support	High	Low	Low	Moderate	Low

This analysis shows that while dedicated localisation platforms offer streamlined implementation for non-technical users, the LLM approach provides unique advantages through its focus on prompt engineering rather than post-translation editing and its ability to support broader content creation needs. This translation strategy represents an important paradigm shift in localisation methodology and content strategy planning.

Based on comparative analysis, the decision framework shown in Table 3 can guide technology selection for website localisation projects.

Table 3. Localisation strategy selection framework based on business requirements

Requirement	Recommended Approach
No-code solution with minimal developer involvement	Weglot (with budget for proofreading)
Developer-controlled translation with maximum flexibility	i18next + manual translation source
Location-based content targeting without translation	IPinfo
Dynamic content (blogs, CMS) with minimal technical debt	Localazy + translator verification
User redirection based on geographical region	Geo Targetly
Marketing-focused content with brand voice consistency	LLMs with prompt engineering
Limited development resources with quality requirements	LLMs with prompt engineering
Multi-channel content strategy across languages	LLMs with prompt engineering
Future expansion to content marketing	LLMs with prompt engineering

For projects with specific requirements of a marketing-friendly tone, limited development resources, and moderate translation volume, the LLM approach with focused prompt engineering represents an optimal solution despite the availability of free alternatives that would require more post-translation editing effort.

V. Conclusion and Future Work

Integrating LLMs for website localisation demonstrates a promising approach that leverages prompt engineering rather than traditional post-translation editing. This methodology represents a paradigm shift in how machine translation can be applied to website localisation, particularly for projects with limited resources but high-quality requirements.

LLM translations provide a superior marketing-oriented tone compared to traditional translation services, maintaining brand voice consistency across languages. Focusing on prompt engineering rather than translation editing offers scalable efficiency as additional languages are added, creating a more sustainable localisation workflow. The cumulative improvement effect, where each new language benefits from the previous translation context, represents a significant advantage over traditional machine translation approaches. The multi-functional capabilities of LLMs extend beyond simple translation to support broader content creation needs, offering strategic advantages for comprehensive content strategies.

This implementation suggests that for certain project profiles - particularly those with marketing content and moderate translation volume - LLMs with properly engineered prompts may offer an optimal balance of quality, efficiency, and resource utilisation compared to traditional translation services and dedicated localisation platforms.

By focusing resources on prompt refinement rather than extensive post-editing, this approach creates a more sustainable and scalable localisation workflow that aligns with established translation theory while leveraging the unique capabilities of modern LLMs.

Future work should explore the potential for automated prompt optimisation, integration with continuous deployment workflows, and quantitative evaluation of user engagement metrics across different language implementations to validate this approach further. Additionally, research into combining LLM-based translation with traditional localisation frameworks could provide insights into hybrid approaches that leverage the strengths of both methodologies.

References

[1] Doherty, S. (2016) “The impact of translation technologies on the process and product of translation,” International Journal of Communication, 10, pp. 947–969.

[2] Hutchins, W.J. and Somers, H.L. (1992) An Introduction to Machine Translation. Academic Press.

[3] Korzynski, P. Mazurek, G., Krzypkowska, P., Kurasinski, A. (2023) “Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT,” Entrepreneurial Business and Economics Review, 11(3), pp. 25–37. Available at: https://doi.org/10.15678/eber.2023.110302.

[4] O'hagan, M. and Ashworth, D. (2002) Translation-mediated Communication in a Digital World: Facing the Challenges of Globalization and Localization. Multilingual Matters.

[5] Pourkamali, N. and Sharifi, S.E. (2024) "Machine translation with large language models: Prompt engineering for Persian, English, and Russian directions," arXiv [cs.CL]. Available at: https://doi.org/10.48550/ARXIV.2401.08429.

[6] Sandrini, P. (2005) "Website localisation and translation", EU-High-Level Scientific Conference Series MuTra 2005 - Challenges of Multidimensional Translation: Conference Preceedings (p. 131 -138). Available at: https://www.researchgate.net/publication/237570958_Website_Localization_and_Translation

[7] Wang, Y. Zhang, J., Shi, T. Deng, D. ; Tian, Y., Matsumoto, T. (2024) "Recent advances in interactive machine translation with large language models," IEEE access: practical innovations, open solutions, 12, pp. 179353–179382. Available at: https://doi.org/10.1109/access.2024.3487352.

[8] Yamada, M. (2023) "Optimizing Machine Translation through prompt engineering: An investigation into ChatGPT's customizability," arXiv [cs.CL] [Preprint]. Available at: https://doi.org/10.48550/ARXIV.2308.01391.