Visualizing Migration Narratives with Generative AI: *A Multi-Modal Approach to Cultural and Emotional Storytelling*

Haofan Wang
Beijing 101 Middle School, Beijing, China
DOI: 10.65285/HSS.V2I4.007

Abstract: Although there's been a worldwide increase in the number of people migrating due to different social factors, their stories and voices have been largely left unheard by the general public. Existing visualizations often generalize, overlooking emotion, metaphor, and context at the individual level. This study tackles the challenges of personal stories and public outreach by utilizing Large Language Models (LLMs)—AI systems designed to understand and generate human-like text. By combining LLMs, sentiment analysis, and image generation to analyze migration stories, the study uncovers social shifts, emotional trends, and cultural themes. This work contributes a multimodal language-to-visual pipeline and a replicable reporting template, focusing on emotional resonance in migration narratives. Retrieval-Augmented Generation (RAG) allows the LLM to access specific information before responding. Through prompt engineering and text-to-image synthesis, the study creates symbolic visualizations and develops an evaluation system in which the LLM assesses whether prompts effectively convey emotion, culture, and meaning. A survey was conducted to evaluate the system's effectiveness. Over 95.9% of respondents (n > 120) reported that at least one AI-generated image clearly evoked feelings related to migration, with an average of 73.2% per image feeling it effectively conveyed immigration concepts. This project increases public awareness of identity, belonging, loss, adaptation, and cultural survival, and explores multimodal generative AI as a tool for cultural interpretation and inclusive social development.
Key words: migration narratives; multimodal generative AI; Large Language Models; texttoimage synthesis; symbolic visualization

1. Design and Implementation of an AI-Powered Visual Migration Storytelling Project (Introduction)

Today, migration stands out as one of the most urgent and complex global issues. Yet, with the emergence of large-scale social media communication, digital platforms may deviate from factual migration-related narratives by fostering algorithm-driven "echo chambers" and accelerating the dissemination of emotional or misleading content (Levi, Arias, & Gehlbach, 2025). Moreover, digital natives differ strikingly from previous generations in their reading preferences, showing a strong shift toward visual and interactive media. People (especially young people) are increasingly turning to online media—especially social media—for faster, more accessible, and engaging content at the expense of traditional print texts and narratives (Dahlan, 2022).

Given this shift and the rapidly evolving era of artificial intelligence technologies, particularly LLMs and multimodal AI, this research is dedicated to exploring how AI tools should enhance the visibility and analyzability of complex human experiences (e.g., migrations) through visual means, especially for younger audiences who tend to engage more deeply with images rather than lengthy texts. Notably, visual text is particularly valuable for groups that may be socially marginalized, including immigrant populations, because it supports identity expression and cross-cultural understanding in a way that traditional text-based models usually fail to do (Farrar, Arizpe, & Lees, 2024).

The rapid development of AI technologies, especially vision-language models (VLMs) and multimodal AI, enables deeper analysis of migration and makes public outreach more effective (Agüera y Arcas, 2022; Jones & Walton, 2018). Fregly, Barth, and Eigenbrode (2023) emphasized that multimodal generative AI comes infinitely closer to artificial general intelligence (AGI) by augmenting a model's contextual understanding and cross-modal learning capabilities to expand the range of use cases and tasks. Multimodal generative AI is a step towards modeling real-world complexity by not only handling different data formats but also transfer learning (Bommasani, Hudson, Adeli, Altman, Arora, et al., 2021). Specifically, picture creation performs far better when it comes to creative issue solving, with wireless prospects, but still faces challenges (Oppenlaender, 2022). While there has been considerable research on visual anthropology and technology-mediated empathy, as well as human-computer interaction (HCI), few researchers have explored the intersection of migration and AI-driven visualization tools.

This project aims to conduct in-depth research on visualizing migration narratives through generative AI tools, selecting the use of symbolic AI-generated images to represent people's emotional and cultural experiences of migration as a specific application scenario. The project also incorporates RAG techniques and prompt engineering, which help each output capture context and emotion precisely and support the structuring of a comprehensive and comprehensible system for story visualization and symbolic image generation. The study introduces a multimodal framework that synthesizes narrative analysis, sentiment recognition, and symbolic image generation related to human migration, addressing the generally homogeneous and weakly readable nature of existing migration narratives. By applying generative artificial intelligence, the study highlights emotional and cultural expressions in the migration stories. These dimensions are often marginalized in broader data-driven research. Meanwhile, traditional textual storytelling continues to receive limited public attention.

Redefining AI as both a technological agent and cultural mediator, this study enhances public inclusion and understanding of mobility, identity, and cultural change, while opening new possibilities for engagement in migration studies.

2. Literature Review Status of Multimodal Anthropology and Human-AI Collaboration

Sensory ethnography has emerged as a trendy new term in visual anthropology in the past decade (Nakamura, 2013). Cultural anthropologists—also referred to as social anthropologists—focus on how humans create meaning through daily and cultural practice in both literate and nonliterate societies, whose work rests on lived experience and social progress within specified cultural contexts, have tried to expand beyond the word, especially when dealing with affective experiences, because while words alone may have their limitations, the emotions and imagery they evoke can transcend those boundaries. For instance, Paul Stoller (1989) urged anthropologists to consider how sensory experiences like taste, sight, and sound influence their ethnographic work.

Sensory ethnography has emerged as a trendy new term in visual anthropology in the past decade (Nakamura, 2013). Cultural anthropologists—also known as social anthropologists—study how humans create meaning through everyday practices within specific cultural contexts (Mead, 1952). They have tried to go beyond the single narrative approach, especially when dealing with affective experiences. The reason for them to seek multimodal presentation is that while words alone may have their limitations, the emotions and imagery they evoke can transcend those boundaries. For example, Paul Stoller (1989) urged anthropologists to consider how sensory elements such as taste, sight, and sound shape ethnographic knowledge.

Dhawka, Perera, and Willett (2024) point out that while earlier studies on anthropographics largely centered on their capacity to elicit prosocial responses in support of humanitarian issues, recent scholarship has increasingly emphasized the value of multimodal approaches in reshaping traditional disciplines across the humanities and social sciences (Westmorelan, 2022; Li et al., 2024; Olier & Spadavecchia, 2022). In anthropology, Westmoreland (2022) argues that multimodality is "reshaping the epistemological, methodological, and representational boundaries" of the field, expanding beyond the text to embrace "more-than-textual mediations of sensorial research experiences" (p. 173). This shift enables anthropology to address its enduring "logocentric practices" by fostering collaborative, experimental, and publicly engaged forms of knowledge production.

The movement toward multimodality is likewise evident in the interdisciplinary use of digital technologies and visual analytics in AI-related fields. Li et al. (2024) propose a framework of human-AI collaboration in data storytelling, delineating roles such as "creator," "assistant," "optimizer," and "reviewer" across four storytelling stages: analysis, planning, implementation, and communication. They stress the need to understand how these roles interact across stages to "leverage the complementary strengths of humans and AI" while addressing challenges in control, autonomy, and coordination.

Status of AI Visualization Tools in Migration

Complementing this methodological evolution, Olier and Spadavecchia (2022) demonstrate the analytical power of interdisciplinary AI-based approaches to uncover visual stereotypes and disproportions in media portrayals of migrants. Their work reveals how the emotional, demographic, and compositional characteristics of media images reproduce and reinforce structural biases, with portrayals often "aligning with emotional and gender stereotypes" and reflecting "power asymmetries" between groups such as "migrants" and "expats" (p. 1). Together, these studies illustrate the critical role of multimodal research and human-AI collaboration in reframing representation and challenging disciplinary conventions in both content and method.

However, in the case of AI-powered visualization tools for migration-related images, there are complex challenges that require careful consideration of cultural and social factors. The tool should not only generate accurate and representative images that can help the public better understand both the journeys themselves and the emotional dimensions they encompass, but also avoid perpetuating stereotypes or biases. Generative text-to-image models can streamline the process of creating migration images, but risk perpetuating designer biases in addition to stereotypes latent in the models. As we pursue the design of diverse and data-driven anthropographics, it is crucial to recognize that text-to-image models carry significant risks, which may disproportionately impact marginalized and underrepresented populations (Dhawka et al.2024). To mitigate this, it is essential to use diverse and representative training data and to implement bias-mitigation techniques during the model development process.

Studies Summary and Research Implications

The reviewed literature highlights a growing interdisciplinary interest in combining AI technologies with anthropological practices, particularly through multimodal and visual approaches that engage audiences more deeply with cultural and emotional narratives. While sensory ethnography and multimodality have expanded the methodological horizons of anthropology, recent studies demonstrate that AI-powered tools—especially generative text-to-image models—offer both opportunities and challenges in the field of public communication. Building on that, we can combine the latest methodology in anthropology and the great potential of generative AI to create a multimodal-powered pipeline that aims to enrich public understanding of migration by making complex stories more accessible and emotionally resonant. Also, this model raises serious ethical concerns regarding bias, representation, and the potential reinforcement of harmful stereotypes. These insights underscore the need for careful, critical, and culturally sensitive design when employing AI-generated visuals to represent human migration. Together, these findings lay the groundwork for a more detailed examination of how AI-powered visualization tools can be leveraged to gain a deeper understanding of the cultural and social patterns embedded in migration.

3. System Design and Implementation(Methodology) Overall System Design

This project combines large language models and AI with text-to-image conversion capabilities to generate images, providing new perspectives and methods for showcasing human migration stories. Figure 1 shows the system's workflow. The input layer includes the migration narrative input and the RAG model, which expands the knowledge base of the LLMs to produce more accurate outputs that align with specific migration stories. The processing layer features the LLM for emotion extraction and summarization. It also has a prompt engineering module to craft prompts for the image-generating AI.

Additionally, it incorporates a self-evaluation system that scores prompts on various dimensions. Finally, the image generation model produces the visual content. The output layer combines text and images to present human migration stories in a multimodal format.

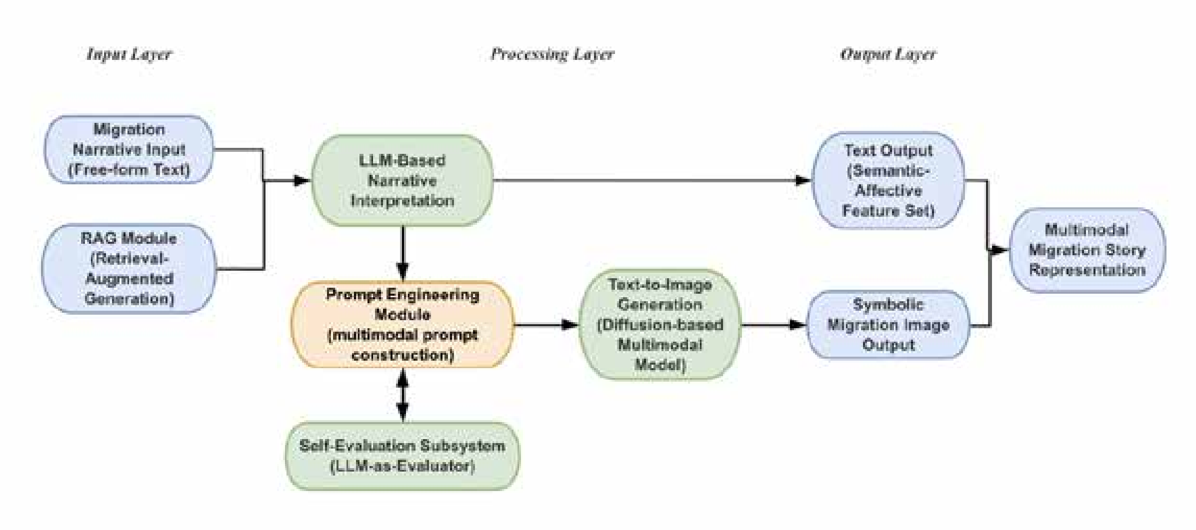

Figure 1 Architecture of the Generative AI System for Multimodal Migration Visualization

Note. The system architecture consists of three layers: (1) The Input Layer collects migration narratives and supplements them with external knowledge via a retrieval module. (2) The Processing Layer interprets narratives, constructs multimodal prompts, and refines them via self-evaluation before image generation. (3) The Output Layer combines text and images into a unified multimodal presentation of migration stories.

Key Technology Implementation

Figure 2 shows the module structure and information flow of the Prompt Engineering section, the technical core of this study, which is divided into three parts: signal extraction zone, interpretive synthesis core, and command assembly output.

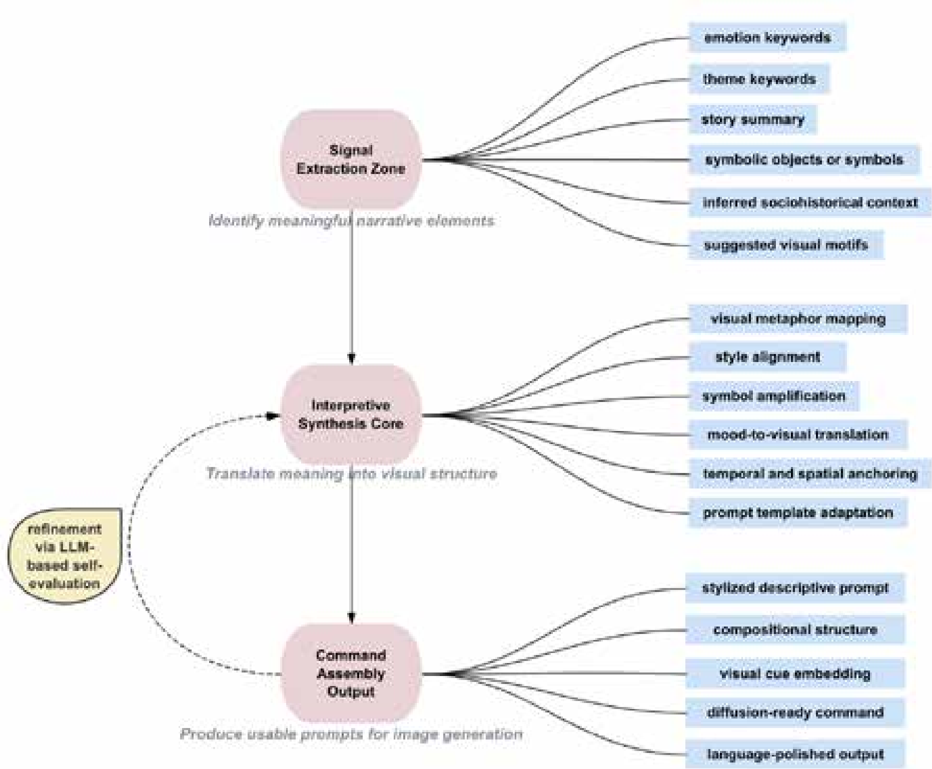

Figure 2 Modular Workflow of Prompt Engineering Module for Multimodal Generation

Note. The workflow consists of three stages: (1) Signal Extraction Zone: Used to identify what is emotionally and symbolically relevant from the narrative and to simplify the story; (2) Interpretive Synthesis Core: responsible for transforming extracted signals into visual concepts and structured prompt fragments; (3) Command Assembly Output: refine and finalize structured prompts that can be used directly in image generation models. In addition, the feedback loop in the figure builds on the involvement of a self-evaluation subsystem (not shown in this figure, but mentioned previously), allowing the generated prompt to be optimized iteratively.

In the first phase (Signal Extraction Zone), the LLM-driven text understanding initially recognizes emotional keywords (e.g., fear, excitement, hope, loss, confusion) and subject keywords (e.g., refugee, climate change, family reunion) within the text. It also performs context-aware analysis to extract symbolic objects or visual cues (e.g., packed suitcases, bouquets on the way out) from the natural language, while capturing the story overview and context (e.g., religion, war). The second stage (Interpretive Synthesis Core) handles vision-language modeling, translating emotions into visual representations on images (e.g., depicting the emotion of "isolation" as a "drifting boat"). This section also leverages prior emotion analysis to match the image style and scene, ensuring that the emotion is conveyed and the narrative remains coherent.

The third component (Command Assembly Output) processes the previously analyzed "image logic" into structured commands that can be directly input into the image generation model. It produces stylized image descriptions using a stable prompt syntax for image models, embedding visual cues, and giving compositional suggestions. The output instructions are optimized for diffusion models by adjusting language rhythm, emphasis, and punctuation (DALL-E was used in this case).

Finally, a separate but essential module within the overall prompt project is Self-Evaluation for Iterative Refinement. This module ensures prompt quality before image generation. In symbolic prompt generation, early model output often carries problems, including a lack of emotional depth, weak cultural relevance, and low symbolic alignment with the original migration story. To solve this, I implemented a loop-based quality control mechanism driven by a secondary LLM evaluator. With this loop, the system does not directly accept the first generated prompt; by contrast, it performs automated self-critique. Each prompt is evaluated along four key dimensions:

1. The main emotions
2. The cultural context
3. The symbolic meaning
4. Overall symbolic alignment

And each dimension is strictly scored on a 1–5 scale, and the system regenerates the prompt until all scores meet a minimum threshold of 4.5 or the maximum attempt limit(10) is reached. This prompt quality optimization loop enhances the prompt's expressiveness and relevance before it is processed by the image model.

The Prompt Engineering Module's workflow implements a decoding-encoding process from narrative input to diffusion-ready output. It is the core of the entire system, directly affecting the effectiveness of the final presentation of AI-generated images related to migration.

System Implementation

The interactive digital platform delivers migrationrelated stories to users. The content combines AI-generated images and brief text descriptions, making these human experiences more visible and understandable—especially for people in the digital age who relate to and empathize with visual communication more than lengthy reports (Paul, 2024).

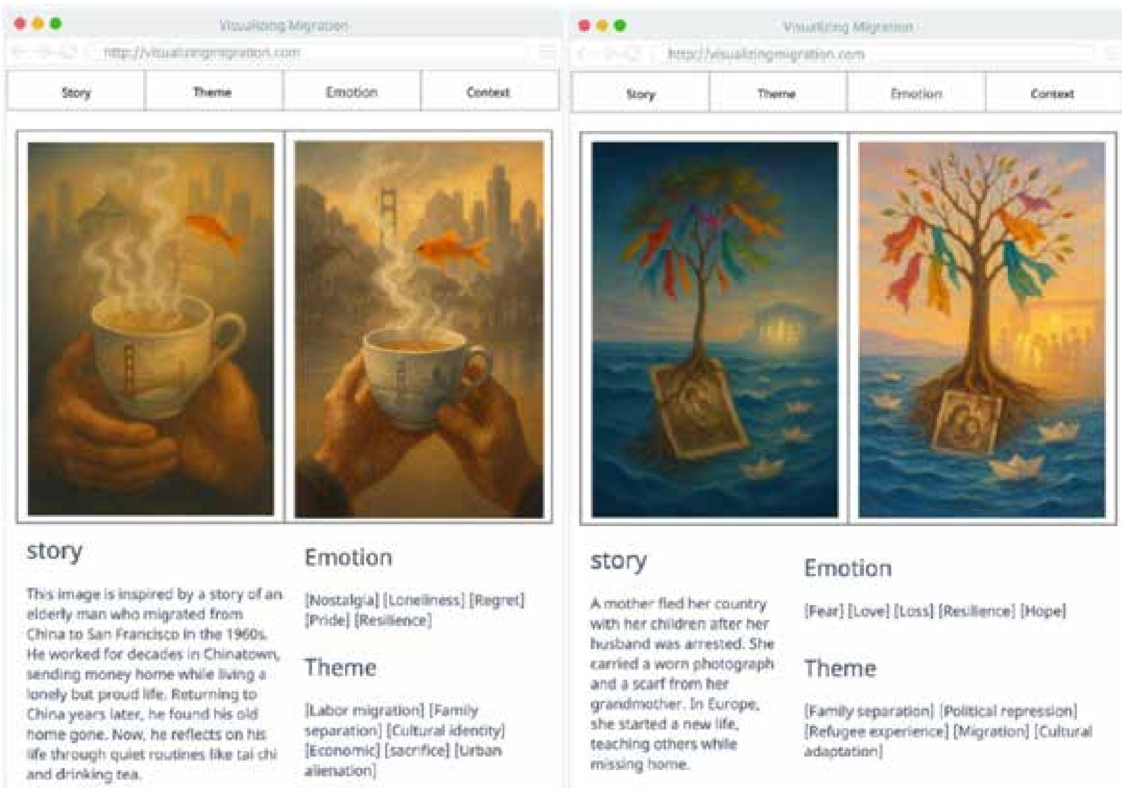

Figure 3 Demonstration Interface of the AI-Powered Migration Visualization Prototype

Note: These two screenshots show a simulated interface of the Migration Visualization System, with simple text-assisted pictures telling different stories. These two examples demonstrate how the system can translate migration stories into multimodal representations of expressions—visual and textual—that can be used to promote cross-cultural understanding and develop emotional resonance.

4.Results and Discussion

Results

According to a questionnaire distributed for this study, which received 123 valid responses (Participants were recruited through a combination of purposive and snowball sampling to achieve a diverse group with various cultural and migration backgrounds. The survey link was shared via international school networks and extended social circles that included local residents, foreigners, and overseas students from different professional fields and elder family acquaintances. This recruitment strategy allowed the study to adopt an intergenerational scope and cross-cultural range rather than focusing on just one demographic group. In this survey, respondents from more than three different cultural backgrounds covering the 10-80 age group), about 69 percent of the respondents indicated that they or their families had experienced migration. This is a good indication that migration experiences are widespread in society and that there is a large audience base for analyzing and visualizing related stories. Among those who have not experienced migration, 87% of the respondents are interested in migration-related stories. Multiple forms of migration storytelling can be a bridge to connect people from different backgrounds, opening new doors for emotional resonance and cognitive understanding.

In the same questionnaire, the study gained user feedback on the effectiveness of AI-generated images on the theme of migration in conveying emotions. The data showed that the migration story images had a strong ability to convey emotion—most viewers could feel the emotional atmosphere and migration elements from them. Even though each image may not be effective for everyone, almost all people can feel the emotional message it conveys from at least one image. These responses prove the social value of AI-generated images in carrying and conveying the emotions of subjects in migration narratives.

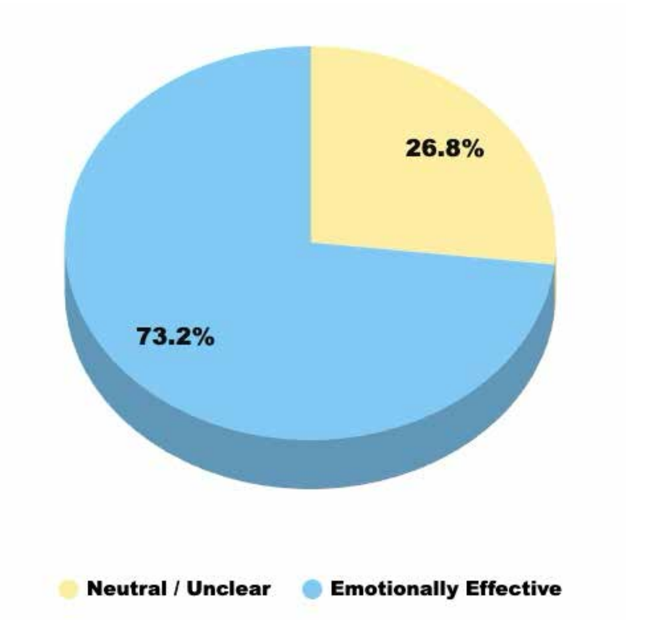

Figure 4 (left) Emotional Effectiveness Evaluation of AI-Generated Migration Images
Figure 5 (right) Aggregate Emotional Recognition in AI-Generated Migration Image Set

Note. Figure 4 (left) demonstrates an assessment of the emotional effectiveness of the AI-generated migrated images. We report per-image resonance (73.2%) and set-level resonance (95.9%) with k=5 images; definitions, sample sizes, and randomized presentation are specified to facilitate replication. They indicate a strong overall resonance across the entire image set. Together, these two graphs demonstrate the breadth of emotions evoked by AI-generated visual images.

Discussion

It is important to recognize that public feedback reveals both the strengths and limitations of using AI-generated images to tell stories and evoke emotions. While more than 70% of people generally believe they can feel emotional messages, some of these feelings are not intuitive or are even abstract. Additionally, over 20% of respondents felt that the images did not trigger any emotional response or perception of information, or that the expression was unclear.

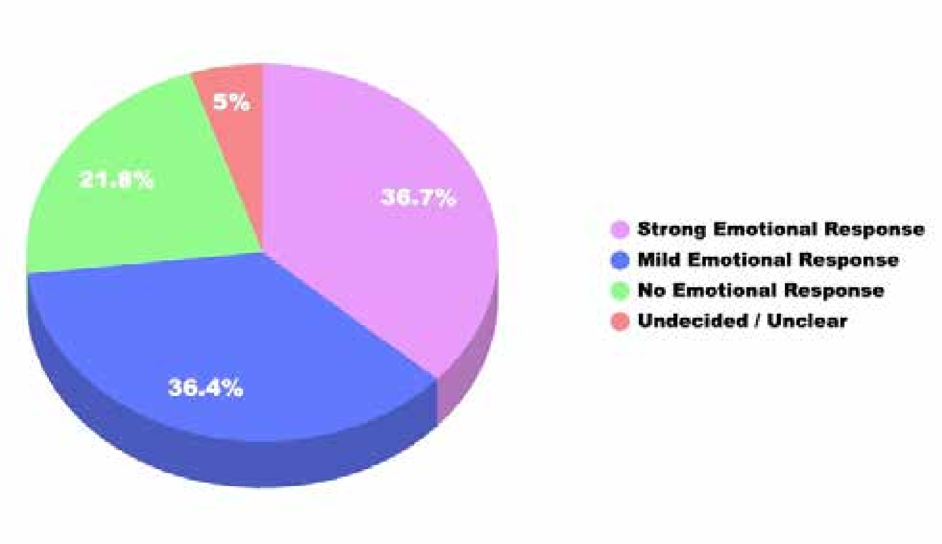

Figure 6 Overall User Evaluation Results of the AI-Generated Migration Visual System

Note. The questionnaire gave the respondents four options in terms of how they felt about the emotional and migratory elements of each picture (clearly felt the emotion; somewhat felt the emotion; did not feel the emotion; not sure), and this pie chart presents the average feedback across the five images corresponding to the five stories.

Five limitations of this study should be noted.

1. Since the migration visualization system is built on LLMs and multimodal AI, there is an inevitable presence—and sometimes amplification—of bias from the training data (Guo et al., 2024). Stereotyped verbal and visual outputs may be produced, affecting the representativeness and fairness of the results.
2. This study relies on a single model (GPT-4o and DALL·E) and lacks multiple architectural approaches or cross-validation. The results are limited by model-specific training preferences.
3. The limited sample size prevents the study from collecting feedback from people of all backgrounds, which reduces the model's generalizability.
4. Subjectivity in emotion and image expression may lead to risks of misinterpretation.
5. The migration visualization system includes several high-consumption modules (LLM, image generation, scoring), which make it expensive to run and limit accessibility.

Summary

The AI Migration Visualization System shows promise in enhancing emotional expression and public understanding of migration stories. It uses text summarization, emotion recognition, and multimodal generation to achieve this. Together, these methods offer a more efficient and flexible way to tell migration experiences. However, more research is needed to test its adaptability across cultures and its long-term impact on social awareness.

5. Conclusion

The study investigates whether an AI-powered migration visualization system can improve emotional expression and boost public understanding of migration stories. Our findings suggest the feasibility of narrative-to-visual generation. Feedback from questionnaires showed that over 95% of respondents connected emotionally with at least one AI-generated image, highlighting the value of symbolic visualization in migration communication.

Nonetheless, there are still limitations in narrative scope and model dependence, and improvements could be made in the following aspects in future work. The study could be strengthened by incorporating a larger and more diverse corpus of migration narratives, which would help generate a more accurate and richer output. Also, the system could evolve from a back-end pipeline into a participatory front-end platform, allowing migrants to write comments and co-design the visual outputs, which would more truly convey their emotional and cultural experience. Meanwhile, it would be better for the system to introduce an audio or video element to make it more appealing through multiple senses. Moreover, the future work should explore ethical safeguards (such as avoiding bias and protecting individual dignity) and transparency mechanisms, which would prevent AI models from emotional distortion or symbolic misrepresentation. The directions mentioned above will collectively enhance the system's cultural accountability, cultural and emotional symbol transmission accuracy, and social practice value of the system.

To conclude, the Generative AI System for Multimodal Migration Visualization expands possibilities for cross-cultural storytelling and digital humanities, providing new ways to depict immigrants' emotions and cultural experiences and addressing the gaps in traditional narratives.

References

[1] Agüera y Arcas, B. (2022). Do Large Language Models Understand Us? Daedalus, 151(2), 183–197.

[2] Dahlan, D. (2022). Social media and the reading culture revolution among digital natives. COMMICAST, 3(1), 105-112.

[3] Farrar, J., Arizpe, E., & Lees, R. (2024). Thinking and learning through images: A review of research related to visual literacy, children's reading, and children's literature. Education 3–13, 52(7), 993–1005.

[4] Fregly, C., Barth, A., & Eigenbrode, S. (2023). Generative AI on AWS: Building context-aware multimodal reasoning applications. O'Reilly Media, Inc.

[5] Guo, Y., Guo, M., Su, J., Yang, Z., Zhu, M., Li, H., ... & Liu, S. S. (2024). Bias in large language models: Origin, evaluation, and mitigation. arXiv preprint arXiv:2411.10915.

[6] Jones, N. & Walton, R. (2018). 10 Using Narratives to Foster Critical Thinking about Diversity and Social Justice. In A. Haas & M. Eble (Ed.), Key Theoretical Frameworks: Teaching Technical Communication in the Twenty-First Century (pp. 241-267). Utah State University Press.

[7] Nakamura, K. (2013). Making sense of sensory ethnography: The sensual and the multisensory. American Anthropologist, 115(1), 132–144.

[8] Levi, E., Bayerlein, M., Grimalda, G., & Reggiani, T. (2025). Narratives of Migration and Political Polarization: Private Preferences, Public Preferences, and Social Media. IZA-Institute of Labor Economics.

[9] Li, H., Wang, Y., & Qu, H. (2024). Where are we so far? Understanding data storytelling tools from the perspective of human-AI collaboration. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24) (pp. 1–19). ACM.

[10] Mead, M. (1952). The training of the cultural anthropologist. American Anthropologist, 54(3), 343-346.

[11] Olier, J. S., & Spadavecchia, C. (2022). Stereotypes, disproportions, and power asymmetries in the visual portrayal of migrants in ten countries: An interdisciplinary AI-based approach. Humanities and Social Sciences Communications, 9(410).

[12] Oppenlaender, J. (2022). The creativity of text-to-image generation. In Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek 2022) (pp. 1–11). ACM.

[13] Paul, S. (2024, April 26). The picture-superiority effect: Harness the power of visuals. Nielsen Norman Group.

[14] Bommasani, Rishi, et al. 2021. On the Opportunities and Risks of Foundation Models. CoRR abs/2108.07258 (2021), 212 pages. arXiv:2108.07258.

[15] Stoller, Paul. 1989. The Taste of Ethnographic Things: The Senses in Anthropology. Philadelphia: University of Pennsylvania Press.

[16] Westmoreland, M. R. (2022). Multimodality: Reshaping anthropology. Annual Review of Anthropology, 51, 173–194.