

**EFFECTIVE USE OF AI-GENERATED
HISTORICAL RECONSTRUCTIONS IN REPRESENTING
MULTIVOCALITY IN HERITAGE**

I.G. Ravini Nimaya Wimalasuriya

Master of Arts in Heritage Studies

**Department of Art & Culture, History, and Antiquity
Faculty of Social Sciences and Humanities**

**Vrije Universiteit Amsterdam
The Netherlands**

27 June 2025



**EFFECTIVE USE OF AI-GENERATED
HISTORICAL RECONSTRUCTIONS IN REPRESENTING
MULTIVOCALITY IN HERITAGE**

**I.G. Ravini Nimaya Wimalasuriya
(2856411)**

Dissertation submitted in partial fulfilment of the requirements for the
Master of Arts Degree in Heritage Studies

**Supervised by:
Prof. dr. Gert-Jan Burgers
dr. Lea Krause**

**Department of Art & Culture, History, and Antiquity
Faculty of Social Sciences and Humanities**

**Vrije Universiteit Amsterdam
The Netherlands**

27 June 2025

Abstract

This research thesis explores the intersection of Cultural heritage and Generative Artificial Intelligence (Gen-AI), focusing on the potential of AI-generated historical image reconstructions to facilitate multivocality in heritage interpretation. While recognising that heritage is not merely the factual historical past, but a process of appropriation shaped by power dynamics and socio-cultural contexts, the study positions itself within critical heritage studies. Accordingly, the study underscores the importance of representing multiple, often underrepresented perspectives, not blindly but critically, to address any historical silences and foster inclusivity. Accordingly, this research evaluates three prominent AI image generation models using three heritage test cases with multiple perspectives. A total of 13 user prompts were transformed into 39 images, which were then linguistically analysed and followed by a systematic visual assessment by heritage experts, using predefined criteria. The study reveals both the strengths and limitations of current generative AI models in producing historically accurate and culturally/contextually sensitive (image) reconstructions.

Based upon the findings, a ‘Conceptual Framework cum Guideline for User Prompt Optimisation’ is proposed by this study to guide users in crafting effective prompts to obtain better AI-generated historical visualisations. Furthermore, it emphasises the necessity of cross-verification of AI outputs with credible historical sources and ethical considerations in generating images. As practical implications, the study suggests utilizing this simulation for visitor engagement, development of interpretive strategies and research. This thesis concludes by emphasizing that generative AI should not be regarded as a definitive source of historical truth, but rather as a supplementary tool for facilitating multivocality and enhancing public engagement with cultural heritage. Therefore, it is crucial that people use this technology in a meaningful, responsible, and critical manner to represent our human past in an inclusive and dynamic way.

Dedication

I am profoundly honoured to dedicate this thesis to Archt. (dr.) Nilan Cooray, whose continuous guidance and encouragement as a mentor have greatly shaped my academic and professional life up to this point. His dedication to the field of Cultural Heritage has instilled in me the values of integrity, intellectual curiosity, and perseverance. His insightful advice has always motivated me to push myself to reach my fullest potential. Although this small token is nothing compared to the depth of my gratefulness to you, may this thesis dissertation stand as a humble dedication to pay my gratitude. Thank you for believing in me!

I.G. Ravini Nimaya Wimalasuriya

27.06.2025

Acknowledgement

Undertaking my MA in Heritage Studies has been one of the most significant and transformative milestones of my life. It provided me with invaluable knowledge and exposed me to many new experiences. This journey prompted me to learn, and relearn, while challenging myself both inside and outside the classroom, extending into practical applications and multiple field visits, both domestic and international. I would like to take this opportunity to express my heartfelt thanks to the Head of the Department, the Course Coordinator, and all academic staff members of the MA in Heritage Studies program. Their continuous guidance and support have been instrumental in reaching this milestone in my research study.

I sincerely extend my deepest gratitude to my research supervisors, Prof. (dr.) Gert-Jan Burgers and dr. Lea Krause, for their unwavering dedication to guiding and supervising me throughout this research thesis. Their keen interest in my field of study and the trust they placed in me have been a constant source of motivation. Their insights, encouragement, patience, and mentorship were crucial to me in the successful completion of this study.

My sincere thanks go to the Ministry of Foreign Affairs of the Netherlands for granting me the opportunity to study in the Netherlands through the Cultural Heritage Scholarship Programme; without that support, I would not have been here. I am also thankful to Nuffic for coordinating our scholarship programme with such care and attention.

Furthermore, I also express my sincere thanks to all heritage experts who participated in the expert evaluation of my research data. Their willingness to share their knowledge and perspectives, despite their busy schedules, has been invaluable. I am further thankful to all the authors of the materials I referred during this study, whose work contributed significantly to shaping my research study.

Last but not least, I am forever grateful to my parents! Their unwavering love, encouragement, and patience have been the pillars of my life, continuously guiding me in achieving both my personal and professional goals.

Dankbetuiging

Het volgen van mijn masteropleiding Erfgoedstudies was een van de belangrijkste en meest transformerende mijlpalen in mijn leven. Het heeft me onschatbare kennis opgeleverd en me blootgesteld aan vele nieuwe ervaringen. Deze reis heeft me aangezet tot leren en herleren, terwijl ik mezelf zowel binnen als buiten de collegezaal uitdaagde, en me toelegde op praktische toepassingen en diverse veldbezoeken, zowel in binnen- als buitenland. Ik wil van deze gelegenheid gebruik maken om mijn oprechte dank uit te spreken aan het hoofd van de afdeling, de cursuscoördinator en alle academische medewerkers van de masteropleiding Erfgoedstudies. Hun voortdurende begeleiding en steun zijn van cruciaal belang geweest voor het bereiken van deze mijlpaal in mijn onderzoek.

Mijn diepste dank gaat uit naar mijn onderzoeksbegeleiders, prof. dr. Gert-Jan Burgers en dr. Lea Krause, voor hun niet aflatende toewijding aan het begeleiden en begeleiden van mij gedurende dit proefschrift. Hun grote interesse in mijn vakgebied en het vertrouwen dat ze in mij stelden, zijn een constante bron van motivatie geweest. Hun inzichten, aanmoediging, geduld en mentorschap waren cruciaal voor mij om deze studie succesvol af te ronden.

Mijn oprechte dank gaat uit naar het Ministerie van Buitenlandse Zaken van Nederland voor de mogelijkheid om in Nederland te studeren via het Cultureel Erfgoed Beurzenprogramma; zonder die steun was ik hier niet geweest. Ik ben ook Stichting Nuffic dankbaar voor de zorg en aandacht die zij ons beurzenprogramma heeft gecoördineerd.

Verder wil ik mijn oprechte dank uitspreken aan alle erfgoedexperts die hebben deelgenomen aan de deskundige evaluatie van mijn onderzoeksgegevens. Hun bereidheid om hun kennis en perspectieven te delen, ondanks hun drukke schema's, is van onschatbare waarde geweest. Ik ben ook dankbaar aan alle auteurs van de materialen die ik tijdens deze studie heb aangehaald, wier werk aanzienlijk heeft bijgedragen aan de vormgeving van mijn onderzoek.

Ten slotte ben ik mijn ouders eeuwig dankbaar! Hun onwrikbare liefde, aanmoediging en geduld zijn de pijlers van mijn leven geweest en hebben mij voortdurend begeleid bij het bereiken van zowel mijn persoonlijke als professionele doelen.

Table of Content

Declaration.....	i
Abstract.....	ii
Dedication.....	iii
Acknowledgement.....	iv
Dankbetuiging.....	v
Table of Content.....	vi
List of Figures.....	viii
List of Tables.....	x
List of Abbreviations.....	xi
Glossary.....	xii
List of Appendices.....	xiii
Chapter 1:.....	1
INTRODUCTION.....	
1.1 General Overview.....	1
1.2 Statement of Problem.....	2
1.3 Aims and Objectives of the Study.....	3
1.4 Significance of the Study.....	3
1.5 Methods, Scope and Limitations.....	4
1.6 Structure of the Dissertation.....	5
Chapter 2:.....	7
LITERATURE REVIEW.....	
2.1 Multivocality of Heritage.....	7
2.2 Generative AI: Image Generation.....	14
2.3 Related Pre-Research and Justification of the Research Gap.....	24
Chapter 3:.....	26
METHODOLOGY.....	
3.1 Review of Research Methodologies adopted in previous research.....	26
3.2 Research Methodology for this Study.....	28
3.3 Contexts of the Selected Test Cases.....	40

Chapter 4:	44
DATA ANALYSIS	
(EXPERIMENTATION OF TEST CASES)	
4.1 Linguistic Analysis examining the Conversion of User-given Keywords into System-generated Prompts	44
4.2 Image Analysis: Comparative Evaluation of the Historical Accuracy of the AI-generated Images	52
Chapter 5:	77
RESULTS AND DISCUSSION	
5.1 The Strengths and Limitations of AI models in generating historical (image) reconstructions	77
5.2 Conceptual Framework cum Guideline for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions	85
5.3 Potential Practical Application of this Research Simulation	89
5.4 Challenges and Limitations encountered during the Research	89
5.5 Directions and Recommendations for Future Studies	90
Chapter 6:	91
CONCLUSION	
Bibliography	94
Appendices	99

List of Figures

Figure	Description	Page
Figure 2.1	Temporary Exhibition titled ‘Legacies of Dutch Slavery: Nine Stories’ (Nov 2024) at the National Archives, The Hague	11
Figure 2.2	Exhibits of the Temporary Exhibition titled ‘Legacies of Dutch Slavery: Nine Stories’ (Nov 2024)	11
Figure 2.3	LLM Development Timeline	15
Figure 2.4	Phases of Natural Language Processing (NLP)	16
Figure 2.5	Several Image generation AI applications in the market	17
Figure 2.6	Sample AI images generated by user inputs and optimized prompts using Stable Diffusion	20
Figure 3.1	Procedure diagram of the study conducted by He et. al. (2025)	26
Figure 3.2	Comparative presentation table of the study conducted by Bansal et al. (2025)	27
Figure 3.3	DALL-E-3 (via ChatGPT-4o)	30
Figure 3.4	Stable Diffusion 3.5	30
Figure 3.5	Midjourney (V.6)	30
Figure 3.6	Aspects/ criteria to be included in AI-based historical (image) reconstructions	33
Figure 3.7	Sample of an expert evaluation form (Test case 03)	36
Figure 3.8	The process illustration of comparative image analysis of this study	37
Figure 3.9	Illustrated version of the current research’s design in brief	38
Figure 3.10	Sugar Plantation Catharina Sophia in Suriname, painted by Jacob Marius Adriaan Martini van Geffen (1860)	40
Figure 3.11	View of the Coffee Plantation Marienbosch in Suriname, painted by Willem de Klerk (1829 – 1876), made after a drawing by A.L. Brockmann	40
Figure 3.12	Sri Pada Mountain in Sri Lanka	41
Figure 3.13	Sri Pada Mountain summit at present	41
Figure 3.14	Siege of the Portuguese fortified city of Galle, painted by Johannes Vinckboons (circa 1669)	43

Figure 3.15	The Capture of Punto Gale by Willem Jacobsz. Coster; Artist unknown (1672)	43
Figure 4.1	Sample for a User-input prompt in keyword format and System-generated format	45
Figure 4.2	Test Case 01- Expert Rating for Representation of Random Visitor's Perspective	54
Figure 4.3:	Test Case 01- Expert Rating for Representation of Dutch Visitor's Perspective	55
Figure 4.4	Test Case 01- Expert Rating for Representation of Surinamese Visitor's Perspective	56
Figure 4.5	Test Case 01- Expert Rating for Representation of Perspectives in DALL-E	59
Figure 4.6	Test Case 01- Expert Rating for Representation of Perspectives in Stable Diffusion	59
Figure 4.7	Test Case 01- Expert Rating for Representation of Perspectives in Midjourney	59
Figure 5.1	Conceptual Process Framework for User Prompt Optimization for AI-generated Historical (Image) Reconstructions	87

List of Tables

Table	Description	Page
Table 2.1	Various capabilities of AI models in imaging domain	18
Table 2.2	Taxonomy of prompt modifiers	21
Table 3.1	Data Collection and Sampling Methods	29
Table 3.2	Model information of the selected gen-AI image generation models	30
Table 3.3	General aspects/categories of keywords to consider when crafting prompts in the selected models of this study	32
Table 3.4	Sample of Linguistic Transformation Table of the Portuguese Visitor's Perspective (Test case 03)	35
Table 4.1	Test Case 01- Linguistic Transformation Table (Random Visitor's Perspective)	46
Table 4.2	Test Case 01- Linguistic Transformation Table (Dutch Visitor's Perspective)	46
Table 4.3	Test Case 01- Linguistic Transformation Table (Surinamese Visitor's Perspective)	47

List of Abbreviations

Abbreviation	Description
AHD	Authorized Heritage Discourse
AI	Artificial Intelligence
AR	Augmented Reality
CoT	Chain of Thoughts
GAN	Generative Adversarial Network
Gen-AI	Generative Artificial Intelligence
GPT	Generative Pre-trained Transformer
LLM	Large Language Model
NLG	Natural Language Generation
NLI	Natural Language Inference
NLP	Natural Language Processing
NLU	Natural Language Understanding
ToT	Tree of Thoughts
UNESCO	United Nations Educational, Scientific and Cultural Organization
VAE	Variational AutoEncoder
VR	Virtual Reality
WHS	World Heritage Site

Glossary

Term	Definition
Generative AI (Gen-AI)	Generative AI, sometimes called Gen-AI, is a sub field of Artificial Intelligence (AI) that can create new content such as text, images, video, audio, or software code in response to a user's prompt or request
Large Language Models (LLM)	Large Language Models (LLM) are mainly created for text generation applications but some of them are multi-modal, hence support different types of content creation including images, videos, sound, and music, etc.
Natural Language Processing (NLP)	Natural Language Processing (NLP) is a branch of AI that enables humans to communicate with machines using human language, bridging the gap between human communication and machine learning.
Prompts	Prompts, or simply the natural language instructions, are the texts/phrases (input) that are written/used to guide AI models to synthesize/generate user's desired/intended output.
Prompt Engineering	Prompt Engineering is the process of designing high-quality prompts that guide Large Language Models (LLMs) to produce reliable and accurate outputs. This process involves crafting/writing the optimal prompt, by optimizing the prompt length, vocabulary, writing style, structure, clarity, specificity, and so on, in related to the desired task, with minimal iterations.

List of Appendices

Appendix	Description	Page
Appendix 1	The list of User-given keyword inputs and the System-generated prompts for image generation	100
Appendix 2	Linguistic Transformation Tables	105
Appendix 3	Analyzed Expert Ratings for Representation of a single perspective across three models	118
Appendix 4	Analyzed Expert Ratings for Representation of multiple perspectives of a single test case within a single model	122
Appendix 5	AI-generated images of this study	125
Appendix 6	List of Heritage Experts	145

Chapter 1:

INTRODUCTION

1.1 General Overview

Cultural heritage is often referred as the inheritance of the past, that is passed down from generation to generation. As per the general definitions provided by international organizations such as UNESCO and ICOMOS over the past decades, heritage comprises of tangible elements, such as artifacts, monuments, sites, and landscapes, as well as the associated intangible elements, such as rituals, beliefs, traditions, etc. However, in the late 20th century, a more critical approach to understanding and analysing heritage has emerged and, the notion of heritage has become less static, more debated and subjected to continuous reflection.

This critical approach towards heritage is rather focused on heritage values, power dynamics, and the socio-political and economic forces that shaped the appropriation of heritage. The critical heritage scholars have highlighted that traditionally heritage discourses reflected state/institutional authority. Therefore, these discourses often marginalised/silenced the diversity lens of the factual past (this will be further discussed in Chapter 2).

Accordingly, in a background where dominant discourse often focuses not on diversity, but on uniformity, emphasizing preservation and stability, the concept of ‘Multivocality’ (sometimes referred as ‘Polyvocality’) in heritage emerges, allowing multiple voices to be heard which were previously often overlooked and/or oppressed. The concept of multivocality is closely related to a more inclusive, diverse, and reflective approach. In support of this approach, heritage professionals and scholars are exploring various innovative methods and tools to incorporate this ‘Multivocality’ into heritage practice, and to promote inclusive storytelling.

Among the tools, the use of Artificial Intelligence, especially generative AI (gen-AI), in the field of cultural heritage is becoming increasingly popular. As a subset of AI, generative AI produces digital content (including text, images, audio, video, and even software codes) in response to user input prompts. By learning patterns and structures from extensive training datasets, gen-AI can produce dynamic, immersive, and personalized representations of digital content, making it a promising tool for historical (digital) reconstructions.

This thesis explores the potential of using generative AI in creating historical reconstructions that reflect multiple perspectives within heritage. It aims to examine how gen-AI can support

and enhance multivocality in heritage interpretation, by enabling more innovative and engaging ways of interacting with and understanding the past.

1.2 Statement of Problem

Interpretation of heritage often involves multiple, and sometimes conflicting perspectives, especially in the case of indigenous heritage, religious/sacred heritage, colonial and post-colonial heritage, military and conflict heritage, slavery heritage, multi-layered heritage, and so on. Heritage professionals and scholars have attempted to address the challenge of representing such diverse narratives in ways that are both comprehensive and respectful of the differing viewpoints.

Accordingly, in exploring various innovative methods and tools to incorporate this ‘Multivocality’ into heritage practice, gen-AI is one of the tools having potential use for this. The use of generative AI for creating digital reconstructions of cultural heritage has been viewed as an opportunity in recent years (Sukkar et al., 2024). As a branch of AI, Generative AI can produce historical reconstructions (as visual representations) in response to user prompts, based on the patterns and structures from it has learned from its training dataset. Since this tool enables the creation of dynamic, immersive, and personalized representations, it holds significant potential to contribute innovatively for heritage interpretation, particularly to address the previously mentioned needs in representing multivocality in heritage. However, despite being a useful tool for visualization, generative AI raises significant concerns regarding both the accuracy and cultural biases of the resulting representations (Cui et al., 2024; Foka & Griffin, 2024; He et al., 2025; Münster et al., 2024; Rane, 2023; Spennemann, 2024; Sukkar et al., 2024). Therefore, the scholars are actively working on addressing these challenges as much as possible, to ensure they are both accurate and sensitive to the relevant cultural context.

Due to the complex nature in representing multivocality in heritage, the interpretation and presentation in museums, archives, and heritage sites have always been challenging. Therefore, it is need not to say that ensuring the accuracy and cultural sensitivity of AI-generated content for multivocality representation is even more crucial. Hence, given the potential of generative AI in creating immersive and dynamic digital representations based on the user prompts, it is essential that this tool is used responsibly and respectfully for inclusive storytelling. This shall be done without distorting the accuracy and authenticity of history, while maintaining coherence and avoiding disruption. Nevertheless, the existing literature reveals that the

effective use of AI-generated historical reconstructions in interpreting multivocality in heritage remains less explored.

Accordingly, the main research questions for this study are:

- *How can the AI-generated historical reconstructions contribute to representing multivocality/ multiple perspectives in heritage interpretation?*
- *What key factors/ aspects should be considered when providing user prompts (textual) to improve the accuracy and cultural sensitivity of these AI-generated historical reconstructions?*

1.3 Aims and Objectives of the Study

Aim of the Study

The research aims to explore the potential of using AI-generated historical reconstructions to enhance the of multivocality in heritage interpretation. It focuses on how this technology can be used to produce more inclusive and diverse representations of narratives. At the same time, it addresses related technical challenges such as accuracy, bias, and cultural sensitivity.

Specific Objectives:

- To explore the potential of generative AI to foster meaningful dialogue on multivocality in heritage
- To assess the extent to which the gen-AI models either neutralize or emphasize multivocality in heritage interpretation
- To examine the differences between various gen-AI models in their representation of cultural elements such as architecture, religion, attire, etc.
- To identify key factors/ aspects to consider when crafting/ providing user prompts, to improve the accuracy and cultural sensitivity of AI-generated historical visual outputs

1.4 Significance of the Study

The research holds significant relevance for the field of Heritage Studies, as Artificial Intelligence, particularly generative AI, has an increasing potential in playing an influential role in digital reconstruction of cultural heritage, including sites, artifacts, and historical narratives. Therefore, the growing application of these technologies offers new possibilities for rethinking how multivocality in heritage can be interpreted and conveyed in more inclusive and engaging ways.

However, despite the potential, AI-generated historical reconstructions have technical limitations regarding the accuracy and the tendency to cultural biases, upon the algorithms and trained datasets. Therefore, ensuring the accuracy and contextual/ cultural sensitivity is essential, especially in heritage contexts with multivocality, where representation has long been contested. In the interpretation of multivocality, where representation is often politically and ethically charged, the responsible use of gen-AI becomes mandatory.

Although scholars are increasingly conducting research on the use of AI in cultural heritage, there remains a notable research gap in critically examining how AI-generated historical reconstructions can be used effectively and ethically to represent diverse narratives. Therefore, this study aims to address that gap by exploring how generative AI can be responsibly used to support the representation of multivocality in heritage and enhance the level of cultural heritage interpretation.

Accordingly, this research brings together Heritage, Artificial Intelligence (AI) and Inclusivity, an interdisciplinary combination that holds significant relevance for today's academic landscape. In addition, by focusing on both the possibilities and the challenges of AI in the context of multivocality, this research contributes to ongoing discussions on the future of digital heritage, offering practical insights into how emerging technologies can be responsibly aligned with inclusive and culturally sensitive heritage interpretation practices.

1.5 Methods, Scope and Limitations

The interdisciplinary nature of this study leads to a research design that integrates both technical experimentation with gen-AI models and interpretative analysis, through qualitative and quantitative methods. Accordingly, a 'Deductive Research Approach' with 'Mixed Research Method' has been adopted, to address this dual focus on AI-generated visual outputs and the interpretation of multivocality in cultural heritage. The research process involves user prompt-driven image generation, followed by expert-driven evaluation, and analytical procedures that combine qualitative and quantitative techniques. This approach allows for both depth of understanding and critical assessment of the research findings.

The study focuses on the potential of using generative AI as a tool for producing (digital) historical reconstructions that reflect multivocality in cultural heritage. It explores how such reconstructions can contribute to more inclusive and diverse heritage interpretation practices. Accordingly, the scope of this research sets around examining selected image generation AI-models and evaluating their outputs based on the predefined criteria related to accuracy,

contextual relevance and cultural sensitivity. For that, three (03) heritage test cases, each having multiple perspectives, are used to guide the analysis. Based on the user-given keywords (*zero-shot prompting*¹), the AI models generate visual content (images) through the system-generated prompts. The geographical/ cultural scope of the study is limited to three (03) test cases, to ensure a focused and in-depth analysis. Further, this study does not delve into the technical reasonings or model architectures behind the generated outputs. Rather, it focuses on their application and implications of this technology within the cultural heritage sector.

While this study aims to provide valuable insights, the following limitations are acknowledged:

- The generative AI models that are used for this historical image generation are generic and commercially available models and, they are not specifically trained and/or fine-tuned for cultural heritage purposes.
- Since the findings are based on a limited number (three) of generative AI models; the results may differ when using other platforms or future versions of the same selected models.

Despite these limitations, this research provides a foundation for further exploration into the responsible and effective use of generative AI in representing multivocality in heritage.

1.6 Structure of the Dissertation

This thesis is structured into seven chapters, each contributing to the exploration of how AI-generated historical reconstructions can be effectively used to represent multivocality in cultural heritage.

Chapter one outlines the background and context of the study, and presents the research problem, aims and objectives, significance, and scope, methods, and limitations of the study.

Chapter two critically reviews existing literature on ‘Multivocality in heritage’ and ‘Generative AI: Image generation’. Furthermore, it examines previous scholarly efforts in this area to highlight the key research gap that the current study aims to address, thereby positioning the research within the broader academic context. The objective of **chapter three** is to establish the methodological framework for this study. The initial part of the chapter reviews research methodologies employed by previous scholars. Based on it, the latter part formulates an appropriate methodological framework tailored to achieve the objectives of this research.

¹ ‘A zero-shot prompt is the simplest type of prompt. It only provides a description of a task and some text for the LLM to get started with. This input could be anything: a question, a start of a story, or instructions. The name zero-shot stands for ‘no examples’ (Boonstra, 2025, p. 13).

Accordingly, the details of the research design, methodological approach, and data collection and analysis methods employed in the study are discussed. In addition, this chapter also explains the selection of gen-AI models and criteria for evaluating the AI-generated historical reconstructions (images), followed by a brief introduction to the three (03) selected test cases.

Being the core of this study, *Chapter four* presents the data analysis process undertaken during the AI image generation experiments for selected heritage test cases with multiple perspectives, across different AI-image generation models. The analysis focuses on two key areas: the linguistic analysis and image analysis. Based on the chapter four, *Chapter five* discuss the research findings and the possibility of its practical application. This section addresses the primary aim of this research, by exploring the strengths and limitations of AI-generated historical reconstructions in responsibly representing multivocality in heritage. This is followed by a proposed conceptual guideline and possible practical application. Additionally, the latter part of this chapter addresses the challenges and limitations encountered in the research and provides directions and recommendations for future studies based on the current findings. Finally, the *Chapter six* concludes the study by summarizing the key findings and evaluating the extent to which this research study achieved its predefined objectives.

Chapter 2:

LITERATURE REVIEW

This chapter begins by reviewing the relevant theoretical context of multivocality in heritage, followed by the technical foundations of gen-AI image generation. The latter part focuses on examining previous scholarly efforts in this area to highlight the key research gap that the current study seeks to address, thereby positioning the current research within the broader academic context.

2.1 Multivocality of Heritage

2.1.1 Background

Cultural Heritage is commonly understood as the legacy of the past, passed down through generations. However, since the emergence of critical heritage studies in the late 20th century, the concept of heritage has been open to ongoing debates and interpretations. As argued by Deumert (2018, p. 4), there is a clear distinction between the inheritance and heritage. As a consequence of the continuation of humanity over thousands of years, '*we cannot not inherit*' the ideas, practices, artifacts, that existed before us. Therefore, Deumert's argument is that '*the past offers us various existential possibilities, from which we select, choose, and filter as we carry it into the present (and future)*'. In this study, the author uses her definition of heritage:

***'Heritage is the result of the choosing;
it is history-as-memory-work, not history-as-fact'***

(Deumert, 2018, p. 5)

Hence, when turning this complex, multiple and diverse inheritance into heritage, a process of appropriation and demarcation takes place and, only certain aspects of the past are borrowed (Derrida, 1994, p. 136, as cited by Deumert, 2018, p. 5).

In order to address this gap in appropriation of heritage, Critical Heritage scholars such as Denise Byrne, David C. Harvey and Laurajane Smith suggest to critically examine the values, power dynamics, and socio, political, and economic forces that influence what is considered/appropriated as 'heritage', and why certain elements of the past are prioritized for preservation (Zheng, 2023, p. 5). Often backed by ethnic or national identity (catering the needs of nationalism), the traditional heritage discourse often tends to see through the eyes of the state/authority, which often in opposition to the diversity lens of the factual/genuine past. Accordingly, this dominant discourse often focuses not on the diversity, but on the uniformity

or homogeneity, emphasising preservation and stability. In other words, the past is seen as a shared experience that form a foundation for a unified ethnic or national identity. Laurajane Smith (2006) refers to this as the '*Authorized Heritage Discourse (AHD)*', which emerged in Europe in the 19th century, and similarly Mbongeseni Buthelezi (2015) designates this as '*Heritage with capital H*' (Deumert, 2018, pp. 5–6; L. Smith, 2006).

Therefore, in this section of the chapter, the author expects to explore the concept of 'Multivocality' (sometimes referred as 'Polyvocality') in heritage, often forgotten and/or oppressed in traditional heritage discourse, which is important in understanding how heritage can be made more inclusive, diverse, and reflective of multiple viewpoints.

2.1.2 Concept and Evolution of Multivocality in Heritage

As per the Oxford dictionary reference, Multivocality means '*Presence of multiple voices*'². Kojan (2008) as cited by Zheng (2023, p. 6) also puts forward that Multivocality, as a social reality, refers to '*the existence of multiple interpretations*'. Although this looks fairly simple in meaning, it is extremely complex in practice. The idea of 'voices' can refer to *differences in opinion, ethnicity, class, education, gender, culture, and just about any other form of human identity and variation that exists* (C. Smith, 2020, p. 7479). Furthermore, Derrida (1994) as cited by Deumart (2018, p. 7) emphasises the importance of encouraging the idea of multivocality in heritage, as '*it allows us to listen not only to the multitude of audible voices, but also to the voices that have been silenced for forgotten in the past*'. In a global context, multivocality has mostly to do with lessening the power imbalances caused by centuries of domination by 'the West' (C. Smith, 2020, p. 7479). In other words, the truthful interpretation and presentation of heritage could involve multiple perspectives and voices, which can often be diverse and sometimes even conflicting.

Since the 1990s, heritage studies has been influenced by a broader range of humanities and social science disciplines, extending beyond the customary fields such as history, museology, and archaeology. As a result, the introduction of ethnographic methodologies drew attention to the importance of acknowledging non-expert knowledge (Mozaffari and Jones, 2019, p. 14, as cited by Zheng, 2023, p. 2). In Parallel, since 1980s, discussions on multivocality were initiated in the field of Archaeology as well, especially by archaeologists such as B.G. Trigger (1984)

² "Multivocality refers to '*the presence of multiple voices, perspectives, or narratives within a given context or text. It allows for diverse viewpoints to coexist and contributes to a richer and more complex understanding of a subject*' (Oxford University Press, 2025)'

and Ian Hodder (1986), who were inspired by pre-modern critique of scientific objectivity and various social movements advocating rights of marginalised groups (Zheng, 2023, p. 3). Accordingly, they highlighted the multivocal interpretations of cultural heritage sites, beyond the prevailing dominant archaeological narratives.

Since 2000s, with the introduction of the concept of ‘Authorized Heritage Discourse’ (AHD) by Laurajane Smith, she questions the notions and practices in heritage that privilege ideologies from certain cultures, classes and gender, by dismantling the power/knowledge interactions within the field of heritage (L. Smith, 2006; Zheng, 2023, p. 3). This initiative was further strengthened by additional scholars, drawing inspiration from Latour’s ‘Actor-Network Theory’ (ANT) and Deleuze and Guattari’s ‘Assemblage theory’ (Zheng, 2023, p. 3). These theories which helped to deconstruct the modernist understanding of heritage, paved the way to emphasizing the importance of recognising and acknowledging the multivocality in heritage.

In the field of archaeology, the differences between professional archaeologists (often considered as experts) and the diverse audiences (publics) they engage with, as well as the differences among those publics themselves, are often referred as the concept of multivocality. Additionally, it encompasses the ways in which these diverse individuals and groups assign value to, interpret, and derive meaning from an archaeological site, along with the narratives they construct about places and artifacts (C. Smith, 2020, p. 7480). Further, these differences do not necessarily align with predictable or straightforward categories such as ethnicity, training, gender, and similar factors. Therefore, multivocality is believed as one of several ideas that are central to the practice of post-processual archaeology.

According to Zheng (2023, p. 6), there are different scholarly perspectives on the concept of multivocality in heritage.

- *Empowering Disenfranchised Groups:* Some scholars (like Fawcett et al., Trigger and Hodder) argue that multivocality allows marginalized groups, who have been historically silenced, to have a voice in a field that is often dominated by narratives tied to nationalism, imperialism, and colonialism. The idea is that these groups can provide alternative perspectives that challenge the dominant, historically biased views.
- *Challenging Dominant Discourses:* Other scholars (such as Barnabas and Silberman) suggest that multivocality shouldn't just be about collecting different voices or perspectives, but it should actively challenge and question the dominant discourses in the

field. These scholars believe multivocality can be a way to provoke critical discussions and push against conventional narratives.

- *Risk of Naive Relativism:* There is also a concern from some scholars (e.g., Anthony and Trigger) that embracing multiple perspectives in an unrestrained manner could lead to naive relativism, the idea that all viewpoints are equally valid, even if they are contradictory or lack a strong basis. This could result in a situation where any claim, no matter how unsupported or unreliable, is given equal weight, leading to a loss of objectivity or critical analysis.

Further emphasising the complexity of this, C. Smith (2020, p. 7479) highlights that every individual has the capacity to express and identify with multiple voices, although they may prioritize one over another depending on the specific political or social context.

2.1.3 Current Practical Applications of Representing Multivocality in Heritage

Given the contemporary emphasis on facilitating multivocality in heritage, heritage institutions such as museums, archives, heritage sites, etc., often seem to put effort in researching and including previously unheard/silenced voices in the historical contexts and, marginalized voices in the contemporary communities associated with the heritage.

For instance, the ‘Migration Museum’ in London regularly features exhibitions co-curated with migrant communities, enabling the presentation of diverse, lived experiences often excluded from national histories (Migration Museum Project, 2025). Similarly, ‘Fenix’, a new art museum about migration was recently opened in Rotterdam in The Netherlands as well. ‘The Humanity House’ in The Hague hosted exhibitions focused on personal refugee experiences, the stories of displaced individuals behind disasters and conflicts (Humanity House, 2025). ‘Slavery and Freedom (1400–1877)’ exhibition and follow-ups in Washington DC by the Smithsonian National Museum of African American History and Culture reflected the previously marginalized African-American experiences and voices into mainstream narrative (National Museum of African American History and Culture, 2025). Similarly, the first ever ‘Slavery’ exhibition in 2021 by the Rijksmuseum in Amsterdam is another example of representation of personal and real-life stories from Brazil, Suriname, the Caribbean, South Africa and Asia (Rijksmuseum, 2025). The ‘Black Heritage Tours’ in Amsterdam, New York State and Brussels is another initiative. In Australia, institutions like the ‘Koorie Heritage Trust’ focus on empowering Aboriginal voices through oral histories, contemporary art exhibitions, and language revitalization programs (The Koorie Heritage Trust, 2025).

In a nutshell, to facilitate multivocality in practice, heritage institutions often employ a range of strategies, such as:

- Walk-in and participatory exhibitions that allow for community co-curation and interactive dialogue
- Oral history and storytelling initiatives, particularly capturing generational memory and lived experiences
- Podcast series that explore lesser-known heritage themes through first-person accounts and critical conversations
- Collaborative research and publishing projects that prioritize ethical engagement with underrepresented groups

Figure 2.1:

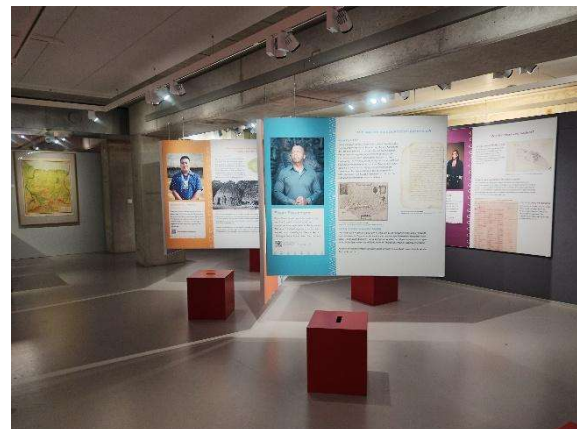
Temporary Exhibition titled 'Legacies of Dutch Slavery: Nine Stories' (Nov 2024) at the National Archives, The Hague



Note. Photograph taken by the author

Figure 2.2:

Exhibits of the Temporary Exhibition titled 'Legacies of Dutch Slavery: Nine Stories' (Nov 2024)



Note. Photograph taken by the author

2.1.4 Challenges in Recognising and Acknowledging Multivocality in Heritage

Enhancing multivocality in heritage contexts such as museums, archaeological projects, archives, and heritage sites is not as simple as it appears. It comes with a range of challenges that complicate its implementation. As Smith (2020, p. 7482) notes, ‘...even when archaeologists attempt to create multivocal projects at more mutually empowered end of the spectrum, the outcomes are not always positive. Each party involved must assess the truth claims of others and accept that multiple, even contradictory, truths can coexist within any given project, social, or political context’.

Advocating for multivocality is difficult for several reasons. As per Zheng (2023, p. 6), the desire to amplify voices beyond dominant grand narratives often proves difficult to achieve

than imagine. Barnabas (2016) and Atalay (2008, p. 37) as cited by Zheng (2023), point out that this challenge arises not only from the provocative and destabilizing potential of multivocality, but also because of most public audiences have been taught only to accept a singular/ univocal view and have not been trained to evaluate multiple perspectives. Moreover, according to Smith (2006), Authorised Heritage Discourse (AHD) possesses powerful assimilative forces that make it difficult to introduce alternative viewpoints. In addition, Neal (2015) as cited by Zheng (2023, p. 6), says since AHD is primarily shaped by Western European elite culture, hence resistance to it may be less noticeable in certain European countries compared to post-colonial contexts.

2.1.5 Critical Evaluation of Multivocality in Heritage

As of now, the scholars have identified three stages for the concept of multivocality in the heritage field, namely:

- (i) theorizing multivocality,
- (ii) operationalizing multivocality, and,
- more recently,
- (iii) evaluating multivocality which are often overlapping (C. Smith, 2020, p. 7481)

Evaluative and critical discussions of multivocality examine when it is effective, when it fails, and the reasons behind these outcomes, while avoiding the assumption that multivocality is always inherently beneficial (C. Smith, 2020, p. 7482). As an example, a group of archaeologists who are also members of indigenous groups have developed approaches that neither fully reject nor privilege Western mainstream methodologies. Instead, they have created new, hybrid approaches that recognize the possibility of multiple truths being equally valid, even if these truths are not scientifically commensurable (Atalay, 2008, as cited by C. Smith, 2020, p. 7479).

However, Zheng (2023, p. 7) notes that, as criticised by scholars such as Atalay (2008), Kojan (2008), Wylie (2008), it is important to understand that equalling multivocality with relativism is a pitfall and is indeed a misinterpretation of multivocality. Accordingly, it shouldn't be assumed that all accounts are equally valid simply because we seek to acknowledge the multivocal reality. Hence, scholars argue that while welcoming diverse and multiple perspectives/ voices, we must wise enough to critically evaluate each perspective for a truthful interpretation and presentation of heritage at the end.

2.1.6 Technological Mediation into Representing Multivocality

With the growing popularity of the subfield of ‘Digital Heritage’, heritage institutions are increasingly collaborating with academic researchers and industry experts to integrate digital techniques into heritage practices. These collaborative efforts actively support and enhance multivocality, making the marginalised/ underrepresented diverse narratives more visible and audible. Digital techniques such as Interactive websites, Mobile Apps, Digital Storytelling, Participatory GIS Mapping, Digital Archives, Virtual Reality (VR) and Augmented Reality (AR) notably contribute to these initiatives, by enhancing the accessibility and engagement.

For instance, projects such as ‘Virtual Harlem’, which reconstructed the historically African American neighbourhood of Harlem (in New York) during the 1920s-1930s era, using Virtual Reality (VR) technology has allowed visitors to engage with diverse narratives in an immersive environment (Electronic Visualization Laboratory, 2025). Similarly, Augmented Reality (AR) applications have empowered marginalized histories by overlaying alternative stories directly onto physical heritage site. A notable example is Museum of London’s ‘Street museum App’, which allows users to visualize historic photographs and hidden narratives (including suffragettes, hippies and Londoners) overlaid on contemporary everyday urban landscapes (Zhang, 2010).

Moreover, digital archives, like the ‘Digital Archive of Indigenous America’, ‘Endangered Archives Programme by the British Library’, ‘Digital Archive of NIOD (Institute for War, Holocaust and Genocide Studies of the Netherlands)’, document and facilitate access to digitalized resources (manuscripts, photographs, sound recordings, oral histories) of marginalized/ unheard narratives. Furthermore, Participatory GIS mapping and Crowdsourcing platforms also allow individuals and communities to tag personal histories and cultural markers to specify sites. In addition, the rise of social media platforms has also created opportunity to voice the unheard narratives beyond the dominant narratives of heritage. Moreover, ‘Pressing Matters Project’, part of the Cultural AI Lab (a collaborative initiative involving, Vrije Universiteit, Rijksmuseum, Wereldmuseum, etc.), explores how colonial objects can support reconciliation and its ‘Digital Traces’ component develops new digital methods for provenance research, focusing on diverse (polyvocal) perspectives in colonial heritage (Cultural AI Lab, 2021; Pressing Matters, 2025).

In this way, the integration of digital technologies is transforming the representation of multivocality in heritage, promoting greater inclusivity, dialogue, and participation with broader audience.

2.2 Generative AI: Image Generation

2.2.1 Introduction to Generative AI

Artificial Intelligence has been a field of research over the past few decades the IT industry, but Generative AI became a widely spoken topic among the general public since the arrival of ChatGPT by OpenAI in 2022 (Stryker & Scapicchio, 2024). “*Generative AI, sometimes called Gen-AI, is a sub field of Artificial Intelligence (AI) that can create new content such as text, images, video, audio or software code in response to a user’s prompt or request*” (Stryker & Scapicchio, 2024). Gen-AI rely on sophisticated machine learning models known as ‘deep learning algorithms’ that stimulate the learning and decision-making processes of the human brain. Accordingly, these models learn the underlying patterns and structures of their large-scale training dataset and use them to produce new content, based on the input, which often given in natural language prompts nowadays.

Gen-AI basically works in three phases namely:

- **Training** - creating foundation model to serve as the basis for multiple gen-AI apps
- **Fine tuning** - tailoring the foundation model to specific gen-AI applications
- **Generation, evaluation and retuning** - assessing gen-AI applications’ output and continually improving its quality and accuracy (Stryker & Scapicchio, 2024)

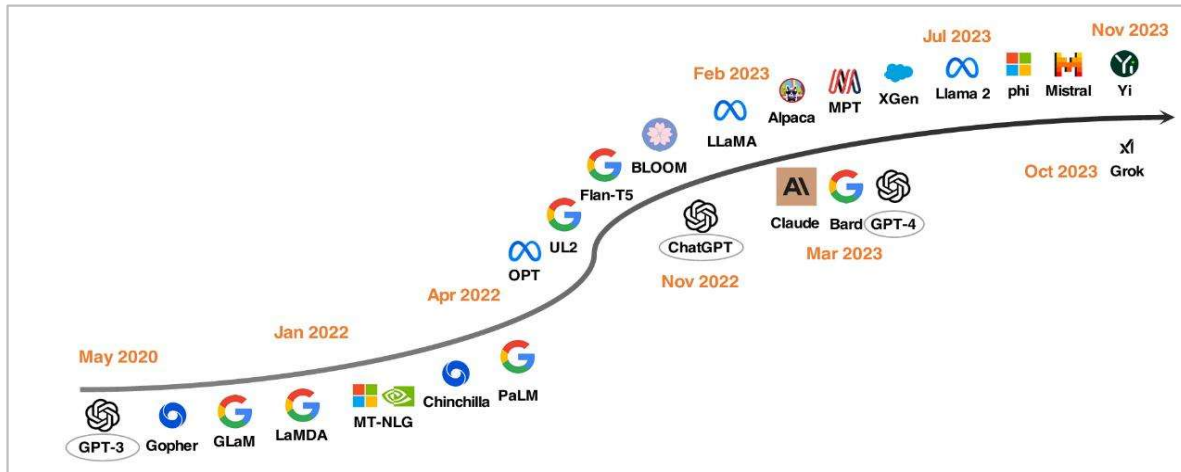
The most common foundation model used today is ‘*Large Language Models (LLM)*’, that are mainly created for text generation applications but some of them are multi-modal, hence support different types of content creation including images, videos, sound, and music, etc. Once the foundational model is trained, it is known as a ‘neural network’. A few of widely used LLMs are ChatGPT, GPT-4, Llama 4, Bard, Claude, Deep Seek, etc (see Figure 2.3).

Once the gen-AI model is setup, the content creation is done based on the model architecture. Accordingly, starting from milestone model architecture ‘*Variational AutoEncoders (VAEs)*’, today most of the widely known models are based on ‘*Generative Adversarial Networks (GANs)*’, ‘*Diffusion models*’ and ‘*Transformers*’. The first two architectures are widely used in high-quality image and video generation (i.e. DALL-E, Stable Diffusion, Midjourney). However, since the latter architecture (Transformers) excels at Natural Language Processing (NLP), it is the basis for GPTs (Generative Pre-trained Transformers) such as Open AI’s ‘ChatGPT’ and ‘GPT-4’, Microsoft’s ‘Copilot’ and Google’s ‘Gemini’ (formerly known as Bard). The major difference of these Gen-AIs from earlier AI applications are that former ones were primarily analytical and classificatory in nature, while gen-AI can produce outputs that

are not merely copied but synthesized based on learned patterns and representations. Hence, this technology is now being used in many industries including, entertainment, gaming and media, medicine, art and design, finance, marketing, and advertising, etc.

Figure 2.3:

LLM Development Timeline (The models below the arrow are closed-source while those above the arrow are open-source)



Note. From ‘ChatGPT’ s One-year Anniversary: Are Open-Source Large Language Models Catching up?’ by Chen et al., 2024 (<https://doi.org/10.48550/arXiv.2311.16989>)

2.2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of AI that enables humans to communicate with machines using human language, bridging the gap between human communication and machine learning. The history of NLP can be understood in three main eras, which were transformative: the ‘*rule-based approach*’ (1950s), the ‘*statistical methods era*’ (1980s), and the ‘*deep learning revolution*’ (mid-2010s to present) (Grammarly, 2024). Accordingly, since the mid-2010s, these modern deep learning techniques have enabled the computers to understand, translate and generate human language (text or speech) with a significant level of accuracy. Use of vast training data and advanced neural networks are two major advancements that contribute to drive this progress. NLP can be divided into sub fields:

- (i) Natural Language Understanding (NLU)
- (ii) Natural Language Generation (NLG)
- (iii) Natural Language Inference (NLI)

Despite the complexity of languages, including context, slang, and evolving usage, NLU focuses on enabling machines to comprehend and interpret human languages by extracting meaning, intent and context from text or speech. NLG is about creating human-like text by machines, allowing them to generate coherent and contextually relevant responses. NLI is the

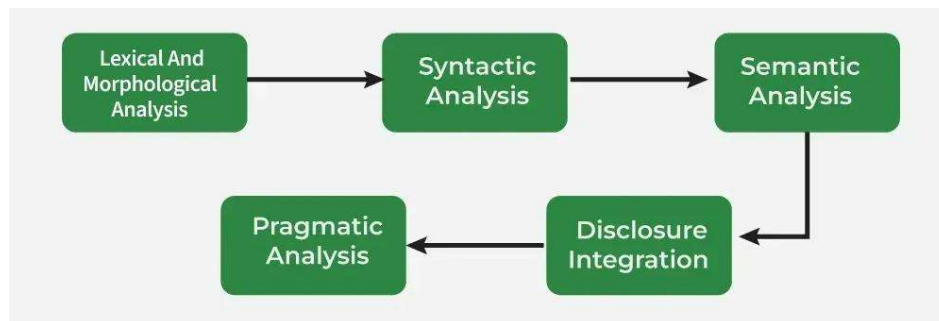
task of determining the logical relationship between a pair of sentences. A *premise* and a *hypothesis*, which is essential for tasks such as reading comprehension and question answering. Together these NLU, NLG and NLI enable machines to engage in more natural and meaningful interactions with users (Grammarly, 2024).

Phases of NLP

In order to understand, analyse and interact with human languages, modern deep learning systems follow a series of interconnected phases that process linguistic input and uncover its structure and meaning.

Figure 2.4:

Phases of Natural Language Processing (NLP)



Note. From ‘Phases of Natural Language Processing (NLP)’ by GeeksforGeeks, 2025
(<https://www.geeksforgeeks.org/phases-of-natural-language-processing-nlp/>)

Modern NLP involves several key phases to understand and generate human language effectively. First, ‘Lexical and Morphological Analysis’ focuses on breaking text into tokens and understanding word structures through techniques like *stemming* and *lemmatization*. Next, ‘Systematic Analysis’ examines sentence structure to ensure grammatical correctness and resolve ambiguities. Followingly, ‘Semantic Analysis’ goes deeper to interpret the meaning of words and sentences in context, using methods such as *named entity recognition* and *word sense disambiguation*. The phase of ‘Discourse Integration’ ensures coherence across multiple sentences by resolving references and maintaining context. Finally, ‘Pragmatic Analysis’ interprets the speaker’s intent and non-literal meanings, such as idioms or sarcasm (GeeksforGeeks, 2025). Together, these phases aid machines to understand and interact with human language in a more accurate, natural, and context-aware manner.

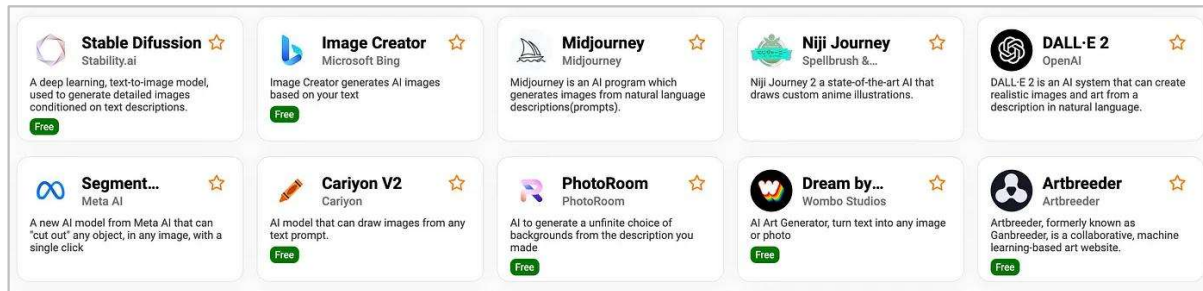
2.2.3 AI Image Generation

Image generation is a major application of gen-AI, which involves synthesizing realistic or stylized visual content from various user input forms, including natural language prompts,

noise vectors and structured data. Accordingly, gen-AI applications like DALL-E, Midjourney, Stable Diffusion, Imagen, etc. can create realistic images or original art, by performing style transfers, image-to-image translation, and other image editing/ enhancement tasks. Figure 2.5 shows a few of image generation AI applications available in the market at present.

Figure 2.5:

Several Image generation AI applications in the market







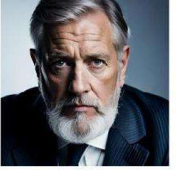
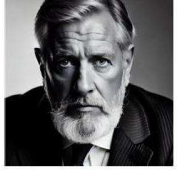
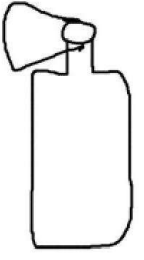







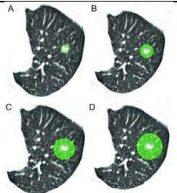


Note. From 'Top 5 Best AI Art Generators in 2024 (Free & Paid)' by Seven Sky Writes, 2024 (<https://medium.com/@sevensky823/top-5-best-ai-art-generators-in-2024-free-paid-f0793024aa70>)

At present, this image generation process has been advanced due to the development of neural network model architectures such as Generative Adversarial Networks (GANs) and Diffusion models, as described in Chapter 2.2.1. GANs consist of two components: a 'generator' that creates images and a 'discriminator' that evaluates their realism, resulting in more and more lifelike outputs. Therefore, the AI models created on these networks are applied in diverse domains ranging from art and fashion to medicine.

Unlike GANs, the diffusion models, the latest innovation in generative image modelling, starts with random noise and iteratively denoise the image over a sequence of steps guided by learned probability distributions (Imagen-Team-Google et al., 2024). Models such as DALL-E, Stable Diffusion, Midjourney, Imagen, etc. have demonstrated better performances in generating high-fidelity images from textual prompts (Imagen-Team-Google et al., 2024). Therefore, these models have the capability to render complex scenes with precise object arrangements, lighting, and stylists' control, which is particularly important for recreating historical environments for this study.

In addition to text-to-image generation (image synthesis), these models are capable of facilitating 'Style transfer', 'Data Augmentation', 'Image-to-image translation', 'Super-resolution', 'Anomaly Detection', 'Face Generation', 'Medical Imaging' and 'Data Compression' (Bansal et al., 2024, pp. 4–5) (See Table 2.1).

Table 2.1:*Various capabilities of AI models in imaging domain*

Category	Type	Input 1	Type	Input 2	Model	Output
Image Synthesis	Text	"A Bee flying out of a glass jar in a green and red leafy basket"	-	-	Diffusion	
Style Transfer	Image		Image		GANs, VAEs	
Data Augmentation	Image		Text	"Change type to black and white"	Image manipulation methods	
Image-to-Image Translation	Image		Text	"Transform doodle to complete image"	GANs	
Super-Resolution	Image		Text	"Enhance the resolution of the given image while preserving its content and maintaining natural details."	Diffusion	
Anomaly Detection	Image		Text	Detect the anomaly in the given image	VAEs	
Face Generation	Text	"A handsome boy with moustache having brown hair and blue eyes wearing rounded goggles"	-	-	GANs	
Medical Imaging	Image		Text	Automatic tumour detection	GANs, VAEs	
Data Compression	Image		Text	"Compress the image to reduce its file size while retaining essential visual information and minimizing loss in image quality."	GANs	

It specifies input types (Type), the primary input (Input 1), a secondary input (Input 2), and the resulting image (Output). Each model's utilization is categorized by the Imaging domain.

Note. From 'Revolutionizing Visuals: The Role of Generative AI in Modern Image Generation' by Bansal et al., 2024 (<https://doi.org/10.1145/3689641>)

Moreover, these image generation AI models are trained on large datasets comprising images, texts, and associated annotations. To ensure quality and safety standards, the models usually employ multi-stage filtering processes, including removing unsafe, violent, or low-quality images. Describing Google's Imagen model specifically say that their model removes AI-generated images to prevent model from learning artifacts or biases commonly found in such images, and a deduplication pipelines and down-weight similar images to minimize the risk of outputs overfitting particular elements of training data (Imagen-Team-Google et al., 2024).

2.2.4 Prompts (Natural Language Instructions)





'Prompts', or simply the natural language instructions, are the texts/phrases (input) that are written/used to guide AI models to synthesize/generate user's desired/intended output. This is a fundamental component of gen-AI models, especially in text-to-text and text-to-image applications, and this was made possible with the introduction of 'Transformer' neural network architecture in 2017. Earlier, the models had to be fine-tuned on a specific task, but with this Transformer architecture, users can instruct the pre-trained models (instruction-tuned) with prompts with natural languages that describe the expected task. Therefore, compared to fine-tuned models, it enables users to interact with AI models in a flexible and natural way to turn the written descriptions into visual, textual, or multimodal outputs, being the primary interface between human intention and machine interpretation. This interactive design made gen-AI more accessible and user friendly, particularly for non-technical users worldwide (IBM, 2025; *Prompt Basics*, 2025).

2.2.5 Prompt Engineering

Prompt Engineering is the process of designing high-quality prompts that guide Large Language Models (LLMs) to produce reliable and accurate outputs (Boonstra, 2025, p. 7). This process involves crafting/writing the optimal prompt, by optimizing the prompt length, vocabulary, writing style, structure, clarity, specificity, and so on, in related to the desired task, with minimal iterations. Since an AI-model's ability to respond to a prompt depends on its training data, architecture, and interpretive mechanisms/configurations, understanding the nuances of language, context, as well as how the model analyses the textual input is important. In simple terms, the effectiveness of a prompt is not necessarily determined by its length, but rather on how specific and clear the instruction is, as slight alterations in phrasing or structure can also significantly influence the output (see Figure 2.6).

Figure 2.6:

Sample AI images generated by user inputs and optimized prompts using Stable Diffusion

User Input	Optimized Prompt
A rabbit is wearing a space suit	A rabbit is wearing a space suit, digital Art, Greg rutkowski, Trending cinematographic artstation
	
Several railroad tracks with one train passing by	several railroad tracks with one train passing by, hyperdetailed, artstation, cgsociety, 8 k
	

Note. From ‘Optimizing Prompts for Text-to-Image Generation’ by Hao et al., 2023
(<https://arxiv.org/abs/2212.09611>)

Prompting Techniques

Although LLMs are tuned to follow instructions and are trained on large amount of data to understand a prompt and generate an output, they are not always optimal without well-structured prompts. The clearer the input prompts are more relevant and reliable the generated output will be. Therefore, understanding different prompting techniques is vital to craft effective prompts.

- ‘Zero-shot prompting’ involves providing a prompt without examples, relying solely on the model’s pre-trained knowledge.
- ‘One-shot’ and ‘few-shot prompting’ introduce one or a few examples within the prompt to guide the model’s behaviour.
- ‘Step-back prompting’ enables the models to temporarily zoom out and reconsider a problem before addressing it, which often leads to more thoughtful responses.
- ‘System prompting’ involves setting high-level instructions or constraints at the beginning to shape the model’s overall behaviour.
- ‘Contextual prompting’ focuses on embedding relevant background information within the prompt to provide better awareness on the situation.
- ‘Role prompting’ assigns the model a specific role or persona (i.e., “You are a legal expert...”) to navigate the responses in a desired style or domain.

- ‘Chain of Thought (CoT) prompting’ guides the model to reason step-by-step, improving performance on complex tasks.
- ‘Tree of Thought (ToT) prompting’, builds on CoT by exploring multiple reasoning paths simultaneously and selecting the most promising one.
- ‘Multimodal prompting’ integrates input across different formats (i.e., text, images, audio), broadening the versatility of model interactions (Boonstra, 2025, pp. 13–64).

Additionally, ‘Automatic prompt engineering’ uses algorithms or models to generate or optimize prompts for improved output quality, to reduce the reliance on manual prompt design.

Prompt Modifiers in Text-to-Image Generation

Prompt modifiers are key elements in the practice of prompt engineering, particularly within text-to-image generation such as DALL-E, Midjourney, and Stable Diffusion. These modifiers (short keywords or phrases) are added to prompts to influence the style, subject, quality, or creativity of the generated image. Oppenlaender (2024) proposes a taxonomy comprising six distinct modifier types: ‘subject terms’, ‘style modifiers’, ‘image prompts’, ‘quality boosters’, ‘repeating terms’, and ‘magic terms’ (see Table 2.2). Each of these modifiers play a unique role in shaping the output of generative models. Therefore, according to Oppenlaender, understanding and applying these modifiers is often an iterative, community-driven process, reflecting the evolving and collaborative nature of human-AI co-creation.

Table 2.2:

Taxonomy of prompt modifiers

Modifier	Description
Subject term	Denotes the subject
Style modifier	Indicates an artistic style
Image prompt	Indicates a style or subject via an image
Quality booster	A term intended to improve the quality of the image
Repeating term	Repetition of subject terms or style terms with the intention of strengthening this subject or style
Magic term	A term that is semantically different from the rest of the prompt with the intention to produce surprising results

Note. From ‘A taxonomy of prompt modifiers for text-to-image generation’ by Oppenlaender, J., 2024 (<https://doi.org/10.1080/0144929X.2023.2286532>)

Oppenlaender (2024)’s argument of an iterative and community-driven process encourages the present study, as the study attempts to explore the behaviour of AI-models in text-to-image generation through zero-shot prompting, for the heritage contexts related to historical scenarios with multiple narratives, a context which remains relatively less explored and applied.

Overall, prompts transform the process of image-making into a collaborative dialogue between human creativity and machine capability. Hence, by understanding and mastering the techniques of effective prompting, one can unlock the wide potential of gen-AI models to obtain the desired outputs across a range of domains.

2.2.6 Current Use of AI and Generative AI in the field of Cultural Heritage

Artificial Intelligence and its subfield Generative AI is emerging as a powerful tool in the cultural heritage domain, that enables innovative approaches to restoration, reconstruction, education, public engagement and many more. Accordingly, in the areas of preservation and restoration, AI models and AI-based image enhancement applications are used to upscale and restore damaged artworks and artifacts by virtually reconstructing lost visual details (Altaweel et al., 2024; Gupta et al., 2021; Kumar & Gupta, 2023). In addition, AI-driven monitoring systems help in detecting early signs of material degradation in heritage buildings through sensor data and image analysis. In addition, language models accessible from the platforms like ‘Hugging Face’ support researchers in heritage sector for advanced image and video generation, text mining, etc.

For documentation and cataloguing in museums and archives, applications such as ‘Transkribus’ automate the transcription and recognition of handwritten or printed historical texts, significantly fast tracking the archival processes (Capurro et al., 2023; Dagenais, 2025; Jaillant, 2022, pp. 180–205; Kirmizialtin & Wrisley, 2020). Further, gen-AI contributes to the prediction of missing or incomplete data. AI also plays a pivotal role in virtual reconstruction and simulation, where models such as Midjourney, DALL-E, Stable Diffusion, and Imagen, can generate visual projections of destroyed or incomplete heritage sites, such as reconstructing ancient temples or war-damaged buildings, enhancing education and museum experiences (Arzomand et al., 2024; Sukkar et al., 2024). Moreover, gen-AI has been using in creating audio-visual storytelling experiences, including immersive mixed reality (AR and VR) (Gemeinhardt & Zöllner, 2024; Stephanidis et al., 2024; Wu et al., 2024), on historical narratives, as seen in projects like Time Machine Europe (Euronews, 2019). Furthermore, AI is gradually being aided in tracking and identifying stolen cultural objects through image recognition, helping to combat illicit trafficking (Abate et al., 2023; Nandi & Cheriyeedath, 2024). Accordingly, the use of AI and gen-AI, open new opportunities to the field of cultural heritage and suggests a paradigm shift in how cultural heritage is visualized, interpreted, conserved, and shared in this digital age.

2.2.7 Gaps and Limitations in Generative AI-based Historical Image Generation

Despite the growing popularity of AI and gen-AI in cultural heritage field, the gen-AI based historical image generations encounter several critical gaps and limitations. One of such main deficiencies is accuracy and historical/cultural sensitivity of the outputs. Particularly in historical image generations, the gen-AI models often synthesise details to fill in gaps with plausible but historically/contextually irrelevant elements, based on the existing patterns in the training dataset. This results in visual outputs look convincing as an artwork/production but may misinterpret historical facts (Foka & Griffin, 2024; Hao et al., 2023; He et al., 2025; Liu et al., 2025). For instance, efforts to reconstruct historical middle Eastern or Asian palaces using text-to-image models might produce architectural elements of contextually inaccurate, due to the biases in the training datasets, which are often disproportionately sourced from contemporary Western art and media (Bu et al., 2025; Foka & Griffin, 2024; Liu et al., 2025). These inaccurate reconstructions might not only mislead public understanding but also cause harm to communities whose heritage is being portrayed. This is linked with the overrepresentation of dominant cultural narratives since the large-scale web data that the models are trained on have often omitted or misrepresented the underrepresented or marginalized cultures. In addition, these same models for creative historical reconstructions can be guided to produce deepfakes or pseudo-historical narratives (false historical imagery), for ideological and politically-charged purposes. (Hao et al., 2023).

Moreover, the lack of provenance tracking and transparency in generated images also poses challenges for scholarly verification. Unlike the traditional projections/reconstructions that are produced based on archaeological and scholarly evidence, AI-generated visuals, especially closed-source gen-AI models, rarely provide the details of sources or logic/decision-making process used to synthesize the visuals. This makes the situation more problematic employing such AI-generated historical visuals for academic and professional use. (Hao et al., 2023; Liu et al., 2025).

In addition, ethical considerations also arise regarding the authorship and cultural appropriation of the output visuals. When gen-AI models reproduce artworks of masters, copyrighted, sacred, culturally restricted, or community-owned designs (i.e., indigenous) without context or consent of the owners or descendants, that risks violating cultural norms and intellectual property rights. From a regulatory perspective, a clear legal framework for generative AI in heritage context is absent. Tools such as image watermarking, blockchain-based content verification, and

explainable AI-frameworks have been proposed as solutions, but their adaptation within heritage workflows is limited (Hao et al., 2023).

Therefore, the scholars are actively working on addressing the above challenges as much as possible, to ensure the gen-AI based historical image generations are both accurate, sensitive to the relevant cultural context, reliable and ethical (Bansal et al., 2024; Cui et al., 2024; Foka & Griffin, 2024; He et al., 2025; Münster et al., 2024; Rane, 2023; Spennemann, 2024; Sukkar et al., 2024). In author's standpoint, the addressing these issues requires interdisciplinary collaboration across computer science, heritage studies, ethics, and law. At the same time, users should be responsible for their practices with generative AI image generation, to enhance our collective understanding of history, rather than distort it.

2.3 Related Pre-Research and Justification of the Research Gap

As described in Chapters 2.1 and 2.2, numerous research studies have been conducted on the multivocality of heritage and use of generative AI for (historical) image generation. However, due to the interdisciplinary nature of the present study combining two areas, research work on this specific intersection, which is the use of gen-AI for representing multivocality, particularly via historical reconstructions (images), has noticeable absence. Hence, this evident gap in the scholarly literature justifies the relevance and necessity, and paves way for the current study.

Nevertheless, despite the limited number of directly related studies, the author acknowledges the relevance of He et al. (2025)'s work, *"I Recall the Past": Exploring How People Collaborate with Generative AI to Create Cultural Heritage Narratives*, which holds certain connection to the themes explored in this research. In their study, they have explored how individuals use generative AI to create cultural heritage narratives. In their study, participants have been asked to use Stable Diffusion, text-to-image AI model, to generate visual representations of both familiar and unfamiliar cultural heritage sites. Their goal was to investigate how gen-AI could support participatory heritage storytelling by individuals, with varying levels of narratives or technical expertise.

In author's point of view, this study is relevant to the present study for a few reasons. First, it addresses the challenge of Authorized Heritage Discourse (AHD) by allow emerging alternative and community-based perspectives. By encouraging participants to narrate personal and emotional connections to familiar sites, their study also engages with the concept of multivocality in heritage, which is the foundation of the present study as well. Secondly, He et

al. points out how AI-generated images can serve as a tool to interpret personal narratives, allowing to evoke memory, inspire creativity, and challenge dominant representations. In addition, their comparison between familiar and unfamiliar heritage sites highlights how the user's local knowledge influences their interaction with the gen-AI models and the outputs. Finally, they present the limitations of gen-AI: its default bias to western aesthetics, struggle with culturally/contextually specific features, and misinterpretation of abstract prompts.

Accordingly, the present author finds that this pre-research provides a foundation for the current study. Furthermore, the issue highlighted by them are critical to the current study's aim as well, which is exploring the potential this technology in responsibly representing multivocality, while addressing its inherent challenges including accuracy, bias, and contextual/cultural sensitivity.

Chapter 3:

METHODOLOGY

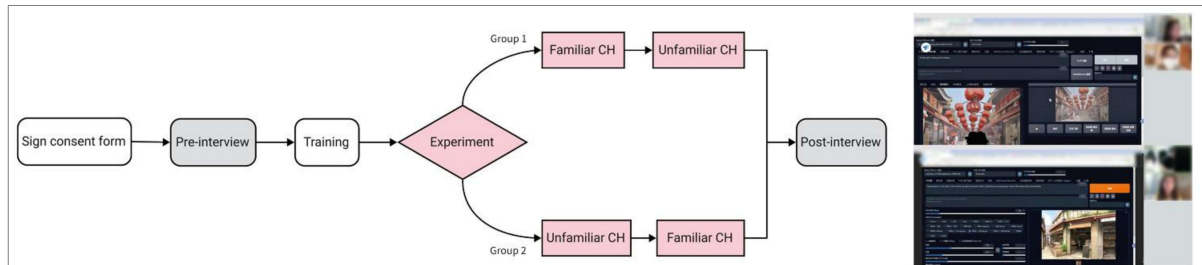
This chapter outlines the research design employed in the study aiming to systematically achieve the research objectives. It begins with a review of research methodologies used by previous researchers, leading to the formulation of an appropriate methodology tailored to meet this study's aims and objectives. The chapter details the chosen research approach, strategy, and methodological choices, including data collection and sampling methods. Further the chapter explains the techniques used for data analysis and presentation, providing justifications for these decisions. Finally, the chapter covers the ethical clearance necessary for the research.

3.1 Review of Research Methodologies adopted in previous research

In order to determine the most suitable research design for this study in line with the aims and objectives, the research methodologies employed by previous researchers were reviewed. As a result, the following components were identified as potentially adoptable with necessary customization in the current study. Accordingly, as further discussed in Chapter 2.3, the study conducted by He et. al. (2025) holds a certain relevance to the present study.

Figure 3.1:

Procedure diagram of the study conducted by He et. al. (2025)



Note. From “*I Recall the Past*”: Exploring How People Collaborate with Generative AI to Create Cultural Heritage Narratives’ by He et al., 2025 (<https://arxiv.org/abs/2212.09611>)


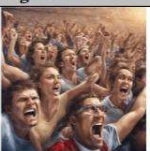
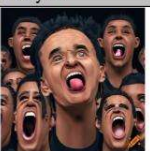


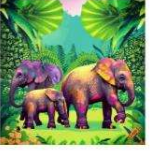



As visualized in Figure 3.1, He et. al. (2025) has conducted a workshop with the participants and have used Stable Diffusion (text-to-image generation AI-model), to create a cultural narrative and generate a series of images related to familiar cultural heritage sites, from their hometowns or other locations they spend considerable time. Following that, the participants had been requested to create narratives on unfamiliar cultural heritage sites given to them, to compare and explore how the level of familiarity with heritage sites influenced their creations. In the subsequent semi-structured interviews, participants have shared their creative themes, insights, and challenges encountered during AI-image creation process.

While appreciating the substantial efforts of He et.al, the present author acknowledges the importance of engaging with the general public to genuinely understand their perspectives on heritage. In addition, the present author expects to adopt their factor of how the user’s level of familiarity with heritage sites influences the creations be examined in the present study. Moreover, in line with the aims and objectives of the present study, the author finds the importance of obtaining feedback/evaluation on the generated image outputs from heritage experts, an aspect which is absent in the work of He et al.

Although not related to my line of research, the research study by Bansal et al. (2024) has worked on practically comparing visual images generated by selected different gen-AI models. Though their objective was not relevant to historical image generations, they have aimed at examining and evaluating the performance and artistic capabilities of each model by subjecting them to three (03) distinct and challenging case scenarios (Human faces, Nature and Tech). The said scenarios had been thoughtfully selected to represent a wide range of visual content, ensuring a well-rounded assessment of the model’s capabilities, upon pre-defined criteria.

Figure 3.2:

Comparative presentation table of the study conducted by Bansal et al. (2025)

Theme	Prompt	Midjourney	Stable Diffusion	Nightcafe	Craiyon
Human faces	"Generate a realistic image featuring 4-5 individual faces of people in a crowded stadium, showcasing their expressions while cheering and shouting with enthusiasm."				
Nature	"Generate a vivid and realistic image of a lush jungle landscape teeming with joyful elephants and vibrant, colorful animals, set amidst a dense forest filled with flowering trees."				
Tech	"A robot inside a dimly lit, cluttered workshop, sparks flying from a welding operation, computer screens displaying schematics, the robot's body covered in welding residue."				

Note. From ‘Revolutionizing Visuals: The Role of Generative AI in Modern Image Generation’ by Bansal et al., 2024 (<https://doi.org/10.1145/3689641>)

Recognising and appreciating the significant study efforts of Bansel et al., the present author adopts their presentation format of examining and comparing the capabilities of multiple models, by subjecting them to multiple test case scenarios. In addition, the present study also ensured that the selected test scenarios represent a wide spectrum, enabling a comprehensive assessment of the selected models’ capabilities in representing diverse historical narratives.

3.2 Research Methodology for this Study

Given the interdisciplinary nature of this study, the methodology was carefully designed to integrate both technical experimentation with AI models and interpretive analysis through qualitative and quantitative methods. To address this dual focus on AI-generated visual outputs and the interpretation of multivocality in heritage, the '*Mixed Research Method*'³ was adopted. The process of user-prompt-driven image generation combined with expert review, required the use of both qualitative and quantitative data collection techniques and analytical procedures. Hence, the integration of these methods enabled to meet the predefined aims and objectives of this research study, providing both depth of insight and validation of findings.

3.2.1 Research Approach

In order to achieve the aim of this research study, which is exploring the potential of using AI-generated historical reconstructions to enhance the interpretation of multivocality in heritage, the current study adopted '*Deductive Approach*'⁴. Accordingly, based on a review of the academic literature, a set of variables was tested through a carefully designed research strategy, to examine the verification or falsification of the relationships and impacts of these variables in generating accurate and culturally sensitive historical (AI-image) reconstructions.

3.2.2 Research Strategy

To align with the aforesaid research approach, it was necessary to examine the behaviour of different image generative AI models across range of heritage scenarios. Therefore, a dual research strategy was adopted, integrating both '*Experimental Strategy*'⁵ and '*Case Study-based Strategy*'. Accordingly, the experimental strategy tested how different user prompts generate varying historical reconstructions across different AI-models. The case study-based strategy enabled to analyse how these AI-generated historical reconstructions could facilitate/represent multivocal narratives. By combining user-prompt-driven image generation

³ "*Mixed methods research is the branch of multiple methods research that integrates the use of quantitative and qualitative data collection techniques and analytical procedures in the same research project. It is therefore based on philosophical assumptions that guide the collection and analysis of data and the mixing of quantitative and qualitative collection techniques and analysis procedures* (Molina-Azorin et al. 2017 as cited by Saunders et al., 2019, p. 181)"...... "*Inmixed methods research, quantitative and qualitative techniques are combined in a variety of ways that range from simple, concurrent forms to more complex and sequential forms* (Saunders et al., 2019, p. 182)".

⁴ "*If your research starts with theory, often developed from your reading of the academic literature, and you design a research strategy to test the theory, you are using a deductive approach* (Saunders et al., 2019, p. 153)"

⁵ "*The purpose of an experiment is to study the probability of a change in an independent variable causing a change in another, dependent variable* (Saunders et al., 2019, p. 190)"

with expert review, the adopted research strategy supported both creative production and critical analysis at the same time. Consequently, this combined approach enabled a systematic evaluation of AI model performance while contextualizing findings within specific heritage-related contexts.

3.2.3 Research Data Collection and Sampling

Table 3.1:

Data Collection and Sampling Methods

Data Collection Tool/Technique	Objective/ Need	Sampling Method	Sample Size
Review of existing literature	<ul style="list-style-type: none"> • To explore the general factors/ aspects to be considered when crafting user prompts in AI-models (prompt engineering) 	N/A	N/A
Brief informal/ unstructured discussions (with general public)	<ul style="list-style-type: none"> • To collect keywords to be used for the user prompts 	Convenience Sampling	07
Image generation AI-models	<ul style="list-style-type: none"> • To generate images upon the given user keywords (natural language prompts) 	Purposive Sampling	03 (see chapter 3.2.5)
Evaluation forms (from heritage experts)	<ul style="list-style-type: none"> • To obtain rated scores and textual remarks/ comments on the accuracy of the AI-generated historical reconstructions (image outputs) in relation to the pre-defined criteria 	Purposive Sampling	03

3.2.4 Selection of AI-Models, Heritage Test cases and Cultural Parameters

Selection of AI Image generation models

To experiment with the representation of multivocality, ‘purposive sampling strategy’ was employed to ensure the selection of relevant and suitable AI-models for the study. To ensure that the image generation quality is high enough to facilitate the intended analysis, several image generation AI-models were initially sample-tested. Based on these preliminary evaluation, three (03) leading and widely used image generation models were selected for the research (see Table 3.2 for detailed information).

Figure 3.3:*DALL-E-3 (via ChatGPT-4o)***Figure 3.4:***Stable Diffusion (3.5)***Figure 3.5:***Midjourney (V.6)**Note.* From Public domain**Table 3.2:***Model information of the selected gen-AI image generation models*

Description	DALL-E 3 (via ChatGPT 4o interface)	Stable Diffusion (V.3.5)	Midjourney (V.6)
<i>Model Information</i>	Developed by OpenAI; integrated into ChatGPT 4o; Proprietary model	Developed by Stability AI; open-source community supported	Proprietary model by Midjourney; focuses on artistic image synthesis
<i>Training process</i>	Reinforcement learning + instruction tuning with human feedback	Supervised training on large-scale LAION datasets + fine-tuning	Trained on proprietary data, details not disclosed
<i>Training data size</i>	Not disclosed, but likely similar to GPT-4 scale (hundreds of billions of tokens)	Trained on LAION-5B and related image-caption datasets	Not disclosed; estimated multi-billion image-text pairs
<i>Model architecture</i>	Diffusion-based transformer, integrated with GPT-4 for natural language alignment	Latent diffusion model architecture (LDM)	Exact specifications unknown (likely transformer-enhanced diffusion model)
<i>Multi-lingual capabilities</i>	Strong multilingual understanding via GPT 4o integration	Moderate multilingual prompt support via CLIP text encoder	Limited multilingual capabilities; English prompts perform best
<i>Visual storytelling</i>	High- excels at narrative scenes and complex prompt chaining	High- adaptable via ControlNet and prompt engineering	Very high- excels at stylized and cinematic storytelling
<i>Zero-shot generation</i>	Strong zero-shot capability across styles and subjects	Effective with good prompts and models; fine tuning improves output	Strong zero-shot generation with creative composition abilities

Note. Adapted From ‘Revolutionizing Visuals: The Role of Generative AI in Modern Image Generation’ by Bansal et al., 2024 (<https://doi.org/10.1145/3689641>)

In addition, the performance/ advanced capabilities in generating detailed content based on textual prompts of these three selected models were further validated by scholarly research, which had been tested against technical evaluation criteria: including *Performance and Robustness*, *Customization and Control*, *Ethical considerations and Accessibility*, *User Experience and Handling* and *Other Technical Aspects* (Bansal et al., 2024), and *Prompt-image alignment*, *Visual Appeal*, *Numerical Reasoning*, and *Image quality* (Imagen-Team-Google et al., 2024). These validations confirmed the suitability of the selected models for this research.

To ensure the geo-regional representation among the AI model origins, the study also explored the possibility of using the ‘*Janus Pro 7B model*’ by Deepseek.ai (China). However, the image generation quality of this model was noticeably lower than that of other three, when tested with the same sample user-prompt. Particularly, the generation of human figures was of too low quality, hence the said AI-model was excluded from this study due to its inadequate image quality to facilitate the desired analysis of this research.

Selection of Heritage Test cases

Due to the need of experimenting the representation of multivocality with gen-AI models, ‘purposive sampling strategy’ was employed to ensure the selection of relevant and suitable heritage test cases for the study. In selecting the heritage test cases, heritage sites/histories characterised by the presence of multiple, and often conflicting, perspectives were at researcher’s interest, as such could reveal the real potential of generative AI in the representation of multivocality. Therefore, the study deliberately included heritage types that are commonly associated with diverse interpretations, such as religious/sacred heritage, colonial and post-colonial heritage, military and conflict-related heritage, slavery heritage, and multi-layered heritage. Additionally, each test case was incorporated with the perspective of a random visitor (which is also a common real-world scenario), to examine how AI models respond to user input that reflects limited familiarity with the subject matter.

Accordingly, following three (03)⁶ diverse and historically complex test cases were chosen for the experimentation of image generation with selected AI models.

- (i) Surinamese plantations during the Dutch colonial period – 03 perspectives
- (ii) Beliefs associated with heritage site of Sri Pada Mountain, Sri Lanka – 06 perspectives
- (iii) Dutch occupation of Portuguese Galle Fort in 1640 (in Sri Lanka) – 04 perspectives

⁶ The three (03) test cases are described in detail in Chapter 3.3

Selection of Cultural parameters/ criteria for testing historical accuracy

As defined in the research objectives and discussed in chapter 2.2.7, ensuring the historical accuracy of the AI-generated historical images is a substantial challenge. Unlike standard image generations tasks, that are often evaluated based on technical qualities, these images must also be culturally/contextually sensitive. Hence, it was first necessary to identify the key factors/aspects to consider when generating historical images using AI. This is also important when crafting user prompts, to maintain both accuracy and cultural sensitivity in the resulting visual outputs. Table 3.3 below outlines the general aspects/categories of keywords that should be considered when writing effective prompts, based on the basic prompt guidelines of the three selected gen-AI models of the study, to achieve high-quality and accurate results.

Table 3.3:

General aspects/categories of keywords to consider when crafting prompts in the selected models of this study

Stable Diffusion	Midjourney	DALL-E
1. Topic	1. Subject (Who and what?)	Not Available <i>Nevertheless, ChatGPT-4o answered as,</i> 1. Subject Matter (What) 2. Style/ Aesthetic (How) 3. Setting/ Environment (Where) 4. Mood/ Atmosphere (Feel) 5. Technical Attributes (Extras): Lighting, Composition, etc.
2. Material	2. Medium (In what form?)	
3. Style	3. Environment (Where?)	
4. Artist Reference	4. Lighting (What kind?)	
5. Website Influence	5. Colour (In what shades?)	
6. Sharpness	6. Mood (Feelings to evoke?)	
7. Extra details	7. Composition (How is it framed?)	
8. Colour and Shade		
9. Lighting and Brightness		
10. Negative prompts		

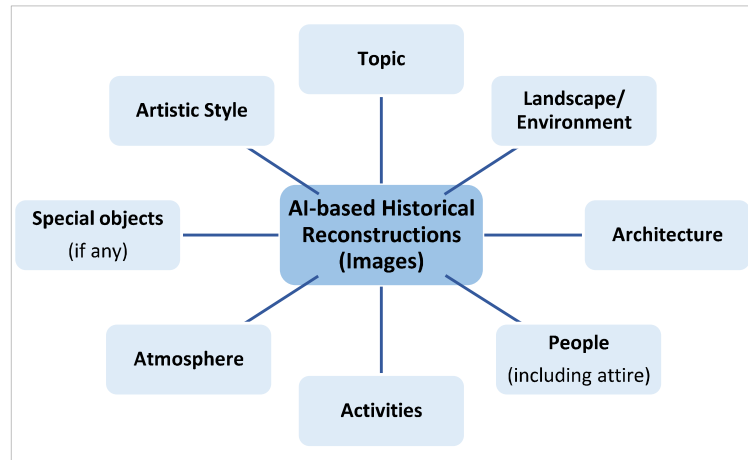
Note. From ‘Prompt Basics’ by Midjourney, 2025 (<https://docs.midjourney.com/hc/en-us/articles/32023408776205-Prompt-Basics>) and ‘Stable Diffusion Prompt Guide’ by Stable Diffusion, 2024 (<https://stablediffusion.com/guides/prompt-guide>)

Accordingly, based on the basic prompt-writing guidelines provided by the three selected models, the author of this study identified the following aspects/criteria to be included in AI-based historical reconstructions (images). Simultaneously, this same set of criteria has been established as cultural parameters for evaluating the historical accuracy of the generated images (see Figure 3.6).

Figure 3.6:

Aspects/ criteria to be included in AI-based historical (image) reconstructions

- (i) Topic
- (ii) Landscape/ Environment
- (iii) Architecture
- (iv) People (including attire)
- (v) Activities
- (vi) Atmosphere
- (vii) Special Objects (if any)
- (viii) Artistic Style



Note. Developed by the author for the present study

3.2.5 Data Analysis Techniques

The data analysis of this study comprises of three key aspects, namely:

- (i) Linguistic Analysis - examining the conversion of user-given keywords into system-generated prompts
- (ii) Image Analysis - evaluating the historical accuracy of the AI-generated images

Linguistic Analysis

Although in name, the AI-generated historical images in this study are based on user's input (user-given keywords), in technically, the AI model does not use these inputs directly. Instead, the AI-model enhances and transforms the user input into a system-generated prompt, which guides the image generation process. Hence, the author identified the importance of examining this intermediate conversion, as it plays a critical role in resulting the generated image output.

This aspect of prompt conversion has not been thoroughly addressed in the pre-research reviewed in Chapter 2. Hence, the author takes the liberty in adopting the below approach, based on the 'Phases of Natural Language Processing (NLP)' discussed under Chapter 2.2.2. According to that framework, Large Language Models (LLMs) process human languages across multiple layers: *lexical and morphological*, *syntactic*, *semantic*, *discourse* and *pragmatic*. Drawing from these NLP phases, the following linguistic operations were identified recurring in the transformation of user input into system-generated prompts. Hence, those operations formed the basis of the linguistic analysis in this study:

- *Synthesis/ Generation* - creating entirely new content where user input is absent or minimal

- *Specification* - adding relevant context information to clarify broad or vague terms
- *Expansion* - including additional related information to expand the original input
- *Enrichment* - adding descriptive depth to enhance the already existing content
- *Clarification/ Disambiguation* - resolving ambiguities in user input
- *Merging* - merging several user-given elements/concepts into one
- *Substitution* - replacing user terms with system-preferred alternatives (either right/ wrong)
- *Omission* - intentionally excluding user-given elements/concepts
- *Rephrase* - reforming the user input for clarity or language fluency
- *Visual mapping* - preparing the user-given terms to a set of visual concepts to support visualising process
- *Style-context matching* - relating the output to reflect the intended (historical) context

Accordingly, each of the 13 prompts (representing the total of 13 perspectives of 3 heritage test cases), was examined for linguistic transformations using the afore linguistic operations. This enabled the present author to systematically assess how the user's expectations (input) have been interpreted and transformed by the AI-model. Such analysis is crucial in understanding the final image outputs and their historical/contextual accuracy. Table 3.4 shows a sample of linguistic transformation assessment of a single perspective of a test case.

Image Analysis: Historical Accuracy (Expert Verification)

As discussed in chapter 3.1, and in alignment with the aims and objectives of this study, the author recognized the importance of obtaining expert feedback on the historical accuracy of the generated image outputs. Therefore, a custom-designed evaluation form was provided to three (03) heritage experts (see Appendix 6 for details), each of whom has substantial expertise in the heritage contexts represented in the test cases.

The evaluation form has been designed to obtain both quantitative ratings and qualitative comments for each of the 39 generated images. Accordingly, the images are assessed for their historical accuracy, based on the 8-criteria, such as People, Landscape/Environment, Activities, etc., as explained in Chapter 3.2.4. Each criterion included a 1-5 rating scale, a Likert-scale, allowing experts to assign a numerical value to each image per criterion. In addition, space has been provided next to each rating box for textual remarks, enabling the experts to elaborate on their evaluations. This qualitative feedback was particularly valuable in cases where images received lower scores, as it helped to identify the reasons behind the ratings (see Figure 3.7 for a sample of expert evaluation form).

Table 3.4:




A sample of Linguistic Transformation Table- Portuguese Visitor's Perspective (Test case 03)

Criterion	Portuguese Visitor's Perspective		
	User Input Keyword(s)	Prompt Phrase(s)/ Expansions	Linguistic Operation(s) Observed by the Author
1) Topic	Historical Recreation of Dutch Occupation of Galle Fort from the Portuguese (during 17 th century) in Sri Lanka- Siege of Galle (1640)	...a fierce daytime siege at a coastal fortress during the Age of Sail	-Substitution -Omission -Rephrase
2) Landscape/ Environment	Galle Fort, Sri Lanka, Indian Ocean, Bay of Galle	...coastal fortress with the Indian Ocean and Bay of Galle in the background	-Synthesis
3) Architecture	Galle Fort	stone fort walls with bastions and towers, some damaged	-Expansion -Enrichment -Specification (<i>adding context</i>)
4) People	Portuguese troops, Dutch troops	Portuguese defenders in red tunics and armour stand firm, with a commanding officer leading the charge...Dutch assailants, with pikes and muskets	-Expansion -Enrichment -Specification (<i>adding context</i>) -Rephrase (<i>Tone adding- "assailants", "defenders"??</i>)
5) Activities	Portuguese troops defending the Dutch attacks to secure their Galle Fort, under the leadership of Captain Lourenco Ferreira de Brito (Portuguese Commander of the Galle fort)	Portuguese defenders.... with a commanding officer leading the charge...Dutch assailants scaling walls and advancing with pikes and muskets...	-Rephrase (<i>Extraction of core meaning</i>) -Generation -Specification (<i>adding context</i>)
6) Atmosphere	Battlefield	...fierce daytime siege...dynamic battle scene marked by rich, warm earth tones, smoke, and fire	-Expansion -Enrichment -Visual mapping/ grounding -Style-context matching
7) Special Objects (if any)	Cannons, weaponry	Red tunics and armour...pikes and muskets	-Omission -Generation (<i>adding context</i>)
8) Artistic Style	Any style suitable for a historical scenery	...oil painting style with rich, warm earth tones	-Style-context matching

Note. Developed by the author for the present study

Figure 3.7:

A Sample of an expert evaluation form (Test case 03)

		Dutch Occupation of Galle Fort in 1640									
#	Criteria	Image	1	2	3	4	5	Remarks			
Portuguese DALL-E											
1	Topic	 <p>How well does the image or video match the description?</p> <p>1. Does not match at all. 2. Does not match at all. 3. Has several minor discrepancies. 4. Has several minor discrepancies. 5. Matches exactly.</p>						Acceptable to certain extent			
2	Landscape/ Environment								Partly matches		
3	Architecture								Partly matches, but its depiction as a battle between two forts is wrong		
4	People								Acceptable to certain extent, but the color of the uniform of the soldiers is wrong		
5	Activities								Partly matches		
6	Atmosphere								Partly matches		
7	Special objects (if any)								Partly matches		
8	Artistic Style								Partly matches		
Portuguese Stable Diffusion											
1	Topic	 <p>How well does the image or video match the description?</p> <p>1. Does not match at all. 2. Does not match at all. 3. Has several minor discrepancies. 4. Has several minor discrepancies. 5. Matches exactly.</p>						Partly matches			
2	Landscape/ Environment								Do not match		
3	Architecture								Do not match		
4	People								Medieval uniforms		
5	Activities								Do not match		
6	Atmosphere								Do not match. Portuguese should attack the Dutch from the top of the fortifications		
7	Special objects (if any)								There are only boats and no ships, only the arrows and no cannons and guns		
8	Artistic Style								Looks more medieval		
Portuguese MidJourney											
1	Topic	 <p>How well does the image or video match the description?</p> <p>1. Does not match at all. 2. Does not match at all. 3. Has several minor discrepancies. 4. Has several minor discrepancies. 5. Matches exactly.</p>						Partly matches			
2	Landscape/ Environment								Partly matches		
3	Architecture								Partly matches		
4	People								Partly matches, but the color of the uniforms do not match with those of a Dutch soldier		
5	Activities								Partly matches		
6	Atmosphere								Partly matches		
7	Special objects (if any)								Only the swords and related weapons,		
8	Artistic Style								More medieval		

Note. Developed by the author for the present study

Accordingly, in the expert evaluation form, the 1-5 Likert scale is defined as follows:

- 1- Does not match
- 2- Has significant discrepancies
- 3- Has several minor discrepancies
- 4- Has a few minor discrepancies
- 5- Matches well

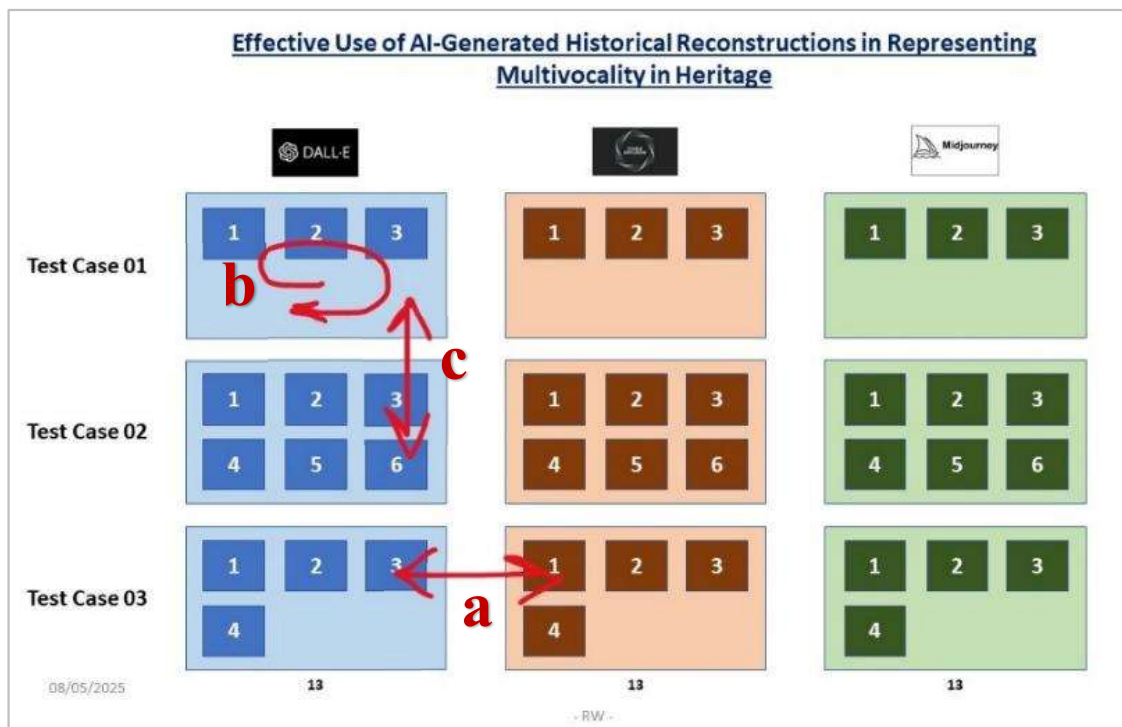
Image Analysis: Comparative Evaluation

Once the expert feedback was received for all the generated images, the 39 images have been comprehensively cross analysed by the author in the following three (03) methods (in the chapter 4), and that process is illustrated in Figure 3.8.

- a) Comparison of single perspective of each test case across three (03) models
(*13 perspectives/ occasions = 3 + 6 + 4)
- b) Comparison of multiple perspectives of a single test case within a single model
(*9 occasions = 3 x 3)
- c) Comparison of multiple test cases within in a single model
(*3 occasions)

Figure 3.8:

The process illustration of comparative image analysis of this study

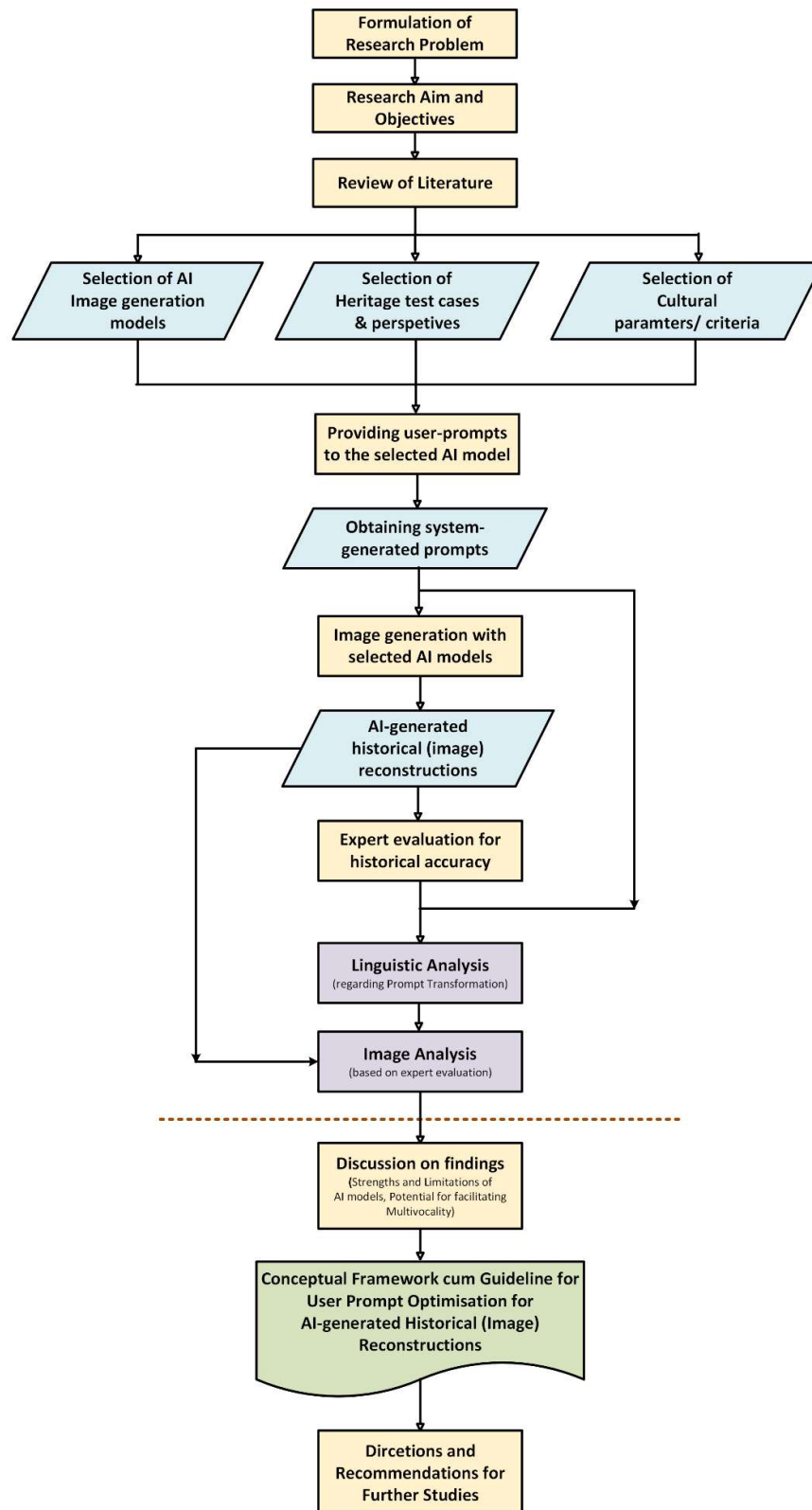


Note. Developed by the author for the present study

3.2.6 Research Design in brief (graphically)

Figure 3.9:

Illustrated version of the current research's design in brief



Note. Developed by the author for the present study

3.2.7 Data Presentation Techniques

- In Chapter 4 (Data Analysis), the linguistic analysis examining the conversion of user-given keywords into system-generated prompts, has been presented in the form of *tables* (see Appendix 2).
- The analysis of numerical ratings (Likert scale) assigned by the heritage experts to each image per each criterion, as well as the overall ratings assigned to each model based on their representation of perspectives, have been illustrated using *bar charts* (see Appendices 3 and 4, respectively).
- One of the key research outputs, the Conceptual Process Framework cum Guideline, proposed in Chapter 5.2, has been presented as a detailed *text document*, while its process flow is visually illustrated through a *flowchart diagram*.

Accordingly, these data presentation techniques have been purposefully selected to ensure critical analysis and make the findings comprehensible to diverse audiences, including subject-experts to general public.

3.2.8 Ethical Clearance

This research study did not involve any human participants providing personal or sensitive information, and nor it requires direct interaction with vulnerable groups. However, ethical considerations were carefully taken into account throughout the research process to ensure that responsible and respectful use of cultural heritage information and AI-generated visual outputs.

Additionally, expert opinions were obtained solely for the purpose of evaluating the generated images. They all were made aware of the study's aims and the voluntarily nature of their involvement.

Accordingly, this chapter has outlined the methodological approach adopted by this research to explore the potential of generative AI in facilitating multivocality through historical (image) reconstructions.

3.3 Contexts of the Selected Test Cases

In order to explore the capabilities and limitations of gen-AI models in generating images to facilitate multivocality in heritage, three diverse and historically complex test cases were chosen for the study. As detailed in Chapter 3.2.4, these test cases span across different geographies, time periods, and cultural contexts, each presenting unique challenges for visual accuracy, cultural representation, and ethical sensitivity.

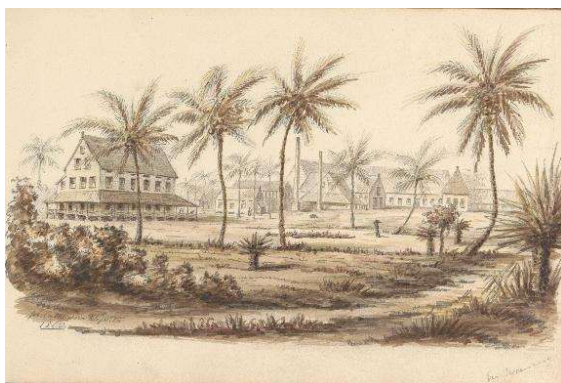
3.1 Surinamese Plantations during the Dutch Colonial Period

The first test case focuses on the plantations in Suriname during the Dutch colonial period, which is a historical context marked with forced labour, colonial exploitation, and Afro-Surinamese cultural identity. This test case highlights the difficulty of visualising painful and contested histories, while respecting the dignity and voices of historically marginalized communities.

Suriname, formerly known as Dutch Guiana, became a hub for plantation agriculture in the 17th and 18th centuries. After seizing from the English by the Dutch in 1667, the colony's economy was revolved around the production of sugar, coffee, cocoa, and cotton industries, which heavily depended on enslaved African labour. These enslaved individuals endured brutal working conditions, violence, and cultural erasure, yet they formed resilient communities (maroon communities by the escaped slaves) that preserved African linguistic and spiritual traditions. Despite the abolition of slavery in 1863, exploitative labour systems persisted. The

Figure 3.10:

Sugar Plantation Catharina Sophia in Suriname



Note. 'Gezicht op de suikerplantage Catharina Sophia', by Jacob M.A.M. van Geffen, 1860. Rijksmuseum (Object no. RP-T-1994-281-20). Public domain. (<https://id.rijksmuseum.nl/200362664>)

Figure 3.11:

View of the Coffee Plantation Marienbosch in Suriname



Note. 'Gezicht op de koffieplantage Marienbosch in Suriname', by Willem de Klerk (after A.L. Brockmann), ca.1829-1876. Rijksmuseum (Object no. SK-A-4087). Public domain. (<https://id.rijksmuseum.nl/20011810>)

Afro-Surinamese population, along with later shifted indentured labour groups from India, Java, and China, formed the foundation of Suriname's multicultural society (Encyclopaedia Britannica, 2025).

Accordingly, visualising this context is challenging, as the context has always been contested and sensitive. Therefore, the attempt to visualise '*Surinamese Plantations during the Dutch Colonial Period*' through AI-imagery while facilitating multivocality is also challenging. Hence, this study decided on testing three associated perspectives: 'Random visitor's perspective', 'Dutch visitor's perspective' and 'Surinamese visitor's perspective', to observe the behaviour of the gen-AI models in representing this contested and sensitive context.

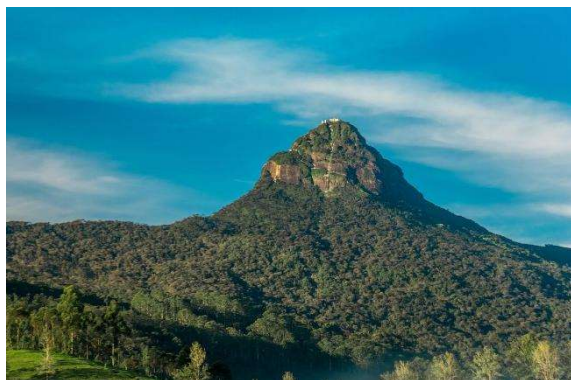
3.2 Beliefs Associated with the Heritage Site of Sri Pada Mountain in Sri Lanka

The second test case focuses on the Sri Pada Mountain (also known as Adam's Peak) in Sri Lanka, which is one Sri Lanka's most spiritually significant heritage sites, revered by multiple religious communities in the country: Buddhists, Hindus, Islamic, and Christians. Every religion considers this mountain sacred relating it to their own religious beliefs, hence making it a rare and powerful symbol of interfaith reverence and shared cultural heritage site. The summit stands at a height of 2243m (7359ft) with a stunning view of the surrounding landscape.

Documented in the ancient Sri Lankan chronicle *Mahavamsa*, Sri Pada Mountain has been a pilgrimage site for over two thousand years since the ancient *Anuradhapura* Kingdom period (1st century BC). At the mountain summit lies a footprint-shaped impression, interpreted variously by different religions, the Lord Buddha's (by Buddhists), Lord Shiva's (by Hindus),

Figure 3.12:

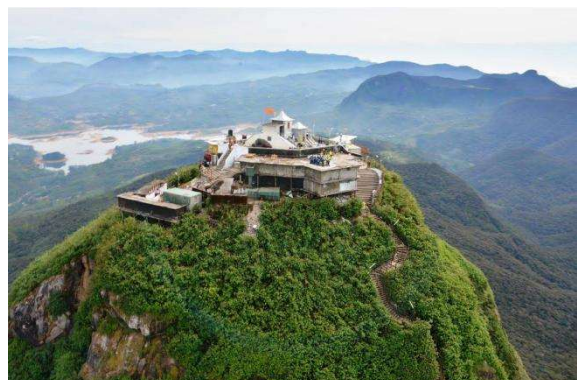
Sri Pada Mountain in Sri Lanka



Note. From "8 Epic Hiking Trails in Sri Lanka", by Matador Network, 2019
(<https://matadornetwork.com/read/8-epic-hiking-trails-sri-lanka/>)

Figure 3.13:

Sri Pada Mountain summit at present



Note. From "Sri Pada- The Most Sacred Mountain in Sri Lanka", by Ceylon Guide, 2020.
(<https://ceylon.guide/2020/04/22/sri-pada-the-most-sacred-mountain-in-sri-lanka/>)

Prophet Al Rohun's (Adam in Islamic tradition), Adam's (by some Christians), and St. Thomas the Apostle's (by some Christians). The pilgrimage to the peak occurs annually during the '*Sri Pada season*' (December – May) and devotees ascend store stairways and forest paths to reach the mountain summit to perform spiritual observance (Manjula, 2024).

Accordingly, the site's deep spiritual pluralism makes it an idea context for evaluating how generative AI models manage overlapping religious narratives. In addition, since the AI models are often trained on Western-centric datasets, it was expected to explore whether depicting South Asian religious iconography, ritual settings might pose a challenge to the models. Therefore, the attempt is to visualise '*Historical Recreation of Sri Pada Mountain summit in Sri Lanka*' through AI-imagery. Hence, this study decided on testing six associated perspectives: 'Buddhist pilgrim's perspective', 'Hindu pilgrim's perspective', 'Islamic pilgrim's perspective', 'Christian pilgrim's perspective-I', 'Christian pilgrim's perspective-II' and 'Random visitor's/Hiker's perspective', to observe the behaviour of the gen-AI models in representing this multireligious heritage context.

3.3 Dutch Occupation of Portuguese-held Galle Fort: The Siege of Galle in 1640 (Sri Lanka)

The third test case focuses on the Dutch occupation of Portuguese-held Galle Fort in Sri Lanka in 1640, a significant turning point in Sri Lanka's colonial history. This event marked the beginning of the transfer of control of from Portuguese to Dutch hands, reshaping the political and cultural dynamics of the island's coastal regions.

Originally built by the Portuguese in the late 16th century as a fortified trading post, Galle Fort was strategically vital for controlling the maritime spice trade. Therefore, the Portugues fortified it with bastions, established churches, and administrative buildings. In 1640, with the assistance from the Kingdom of Kandy, the Sri Lankan kingdom, which was resisting against the Portuguese expansion, the Dutch East India Company (VOC) successfully captured the fort. The Dutch expeditionary force was led under Commodore Willem Jacobsz. Coster, while the Portugues troops were led by Captain Lourenco Ferreira de Brito (Portuguese Commander of Galle Fort). This event inaugurated a new phase of colonialism, with the Dutch swiftly reinforcing and expanding the fort's architecture according to their own military and urban planning principles. Accordingly, the fort was restructured with robust ramparts, new gates, street grids, and Dutch-style buildings that still stand today as a UNESCO World Heritage Site.

Accordingly, this test case expects to represent how AI imagery tools can visualise a moment of conflict, colonial power battle among European forces and hybrid architecture in a South

Asian land. Therefore, the attempt is to visualise ‘*Historical Recreation of Dutch Occupation of Galle Fort from the Portuguese (during 17th century) in Sri Lanka- Siege of Galle (1640)*’ through AI-imagery. Hence, this study decided on testing four associated perspectives: ‘Dutch visitor’s perspective’, ‘Portuguese visitor’s perspective’, ‘Sri Lankan visitor’s perspective’, and ‘Random visitor’s perspective’, to observe the behaviour of the gen-AI models in representing this battle of European colonial powers in a South Asian land.

Figure 3.14:

Siege of the Portuguese fortified city of Galle in 1640



Note. ‘*Siege of the Portuguese fortified city of Galle*’, by Johannes Vinckboons, circa 1669. Het Scheepvaartmuseum, Amsterdam (Object no. 2009.2158). Public domain. (<https://collectie.hetscheepvaartmuseum.nl/Details/collect/663704>)

Figure 3.15:

The Capture of Galle Fort by Willem Jacobsz. Coster



Note. ‘*The Capture of Punt Gale by Willem Jacobsz. Coster, 13 March 1640*’, Print in Philippus Baldaeus, *Naauwkeurige beschryvinge van het machtige eyland Ceylon*, 1672 (‘A true and exact description of the great island of Ceylon’), p. 60. Artist not known, c. 1672. Rijksmuseum, Amsterdam, inv. nr. 307-A-13 [library]

Overall, this chapter outlined the methodological approach adopted to explore the potential of generative AI in representing multivocality through historical (image) reconstructions. It described the research design, data collection and analysis methods, evaluation criteria, and the ethical considerations of the study. Accordingly, the next chapter presents the analysis of the collected data of thirteen (13) distinct perspectives drawn from the three (03) selected heritage test cases, each diverse and historically complex in interpretation.

Chapter 4:

DATA ANALYSIS

(EXPERIMENTATION OF TEST CASES)

Being the core of this study, this chapter focuses on the gradual process of data analysis involved with the experimentation of AI image generation for selected heritage test cases with multiple perspectives, across selected AI-models. Accordingly, the generated AI images are analysed in different aspects to explore and understand the behaviour of image generation process for historical reconstructions.

- **Linguistic Analysis** examining the conversion of user-given keywords into system-generated prompts
- **Image Analysis** evaluating the historical accuracy of the AI-generated images

4.1 Linguistic Analysis examining the Conversion of User-given Keywords into System-generated Prompts

In line with the aim and objectives of this study, the research was designed to explore the potential of gen-AI for historical image generation, enabling the general public/visitors to express how they perceive a particular heritage, facilitating the representation of multivocality in heritage interpretation.

Therefore, the research employed a single AI model to process user-provided keywords and generate its corresponding system-generated prompts for image creation. Considering its user-friendly chat interface, high performance in NLP, widespread use, and general public familiarity, ChatGPT-4o, integrated with the selected DALL-E model, was chosen for generate prompts that are used for image generation across the three selected models in this study.

The keywords collected from the participants to define their desired image outputs were entered by the author into the system as user prompts, based on the predefined 8 key-criteria described in Chapter 3.2.4, such as Topic, Landscape/Environment, People, etc. (see Figure 4.1)

Accordingly, a total of 13 user and system-generated prompts, representing 13 distinct perspectives across three heritage test cases (see Appendix 1), have been examined for linguistic transformation (see Appendix 2) using the predefined linguistic operations described in Chapter 3.2.5, such as Specification, Enrichment, Omission, etc.

Figure 4.1:

Sample for a User-input prompt in keyword format and System-generated format

User-input (keyword format)		System-generated Prompt		
SURINAMESE PLANTATIONS - 2D Image		SURINAMESE PLANTATIONS		
1	Random visitor	#	Perspective	Image Code
Topic:	Surinamese Plantation during the colonial period	1	Random visitor	(1)S-D-001
Landscape/Environment:	Surinamese land, maybe flat land I guess			
Architecture:	Colonial architecture, Surinamese architecture			
People:	Surinamese people, Dutch people			
Activities:	Surinamese people working in the land, Dutch overseeing it			
Atmosphere:	daytime			
Special objects (if any):	I don't know			
Artistic Style (optional)	Any style suitable for a historical scenery			
			ChatGPT Prompt (AI-generated)	
			A colonial-era oil painting showcases a day on a Surinamese plantation, with enslaved African workers toiling under the watch of Dutch overseers. In the foreground, laborers engage in agricultural tasks, while a grand plantation house with Dutch architectural elements dominates the background, surrounded by lush tropical vegetation. The piece employs earthy tones, rich greens, and subtle splashes of color to create a balanced composition, focusing on the contrast between the workers and their overseers.	

As depicted in the linguistic transformation tables in Appendix 2, this analytical process enabled the author to systematically assess how users' expectations (inputs) have been interpreted and transformed by the model. On one hand, this allowed the author for an in-depth examination of the AI model's role and influence in representing multivocality in heritage. This addresses the second specific objective of this research (see Chapter 1.3). On the other hand, this analysis was essential for Assessing the historical accuracy and cultural/contextual sensitivity of the generated image outputs, as discussed in the image analysis in Chapter 4.2. Accordingly, the linguistic analysis results presented in Appendix 2 have been comparatively analysed for each test case in the following chapter sections 4.1.1 to 4.1.3.

4.1.1 Test Case 01: Surinamese Plantations during the Colonial Period

(Comparative Analysis of Linguistic Transformation across Visitor Perspectives)

As pointed out in the three linguistic transformation tables (see Table 4.1- 4.3), it is noticeable how differing user perspectives, Random, Dutch and Surinamese, has affected the shaping of system-generated AI-prompts used for image generation. Each table begins with the common theme/topic, "*Surinamese plantation during the colonial period*". However, the system prompts have expanded and detailed in varied ways in the remaining criteria depending on the user-input given, in each different perspectives/positional standpoints of the visitors, as evidence from the linguistic operations.

Across all the three perspectives, several linguistic operations such as 'Specification', 'Enrichment' and 'Expansion' are consistently applied. It is apparent that they have been used to refine the vague and limited inputs into a more comprehensive prompt supporting the models to generate more relevant images. For instance, general term "*Surinamese land*" has been expanded and enriched with descriptions such as "*lush tropical vegetation*", "*towering palm trees*", and "*cleared land for cultivation*".

Table 4.1:

Test Case 01- Linguistic Transformation Table (Random Visitor's Perspective)

Random Visitor's Perspective			
Criterion	User Input Keyword(s)	System-generated Prompt Phrase(s)/ Expansions	Linguistic Operation(s) Observed by the Author
1) Topic/ Subject	Surinamese Plantation during the colonial period	a day on a Surinamese plantation during the Dutch colonial period	-Specification (<i>adding historical context, 'Dutch'</i>)
2) Landscape/ Environment	Surinamese land, maybe flat land	flat, tropical terrain... lush tropical vegetation... cleared land for cultivation	-Expansion -Enrichment -Clarification/Disambiguation
3) Architecture	Colonial architecture, Surinamese architecture	a grand plantation house with Dutch architectural elements	-Generation -Specification
4) People	Surinamese people, Dutch people	enslaved African workers... Dutch overseers... laborers... overseers on horseback	-Substitution -Specification (<i>Role assignment?</i>)
5) Activities	Surinamese people working in the land, Dutch overseeing	workers toiling... engaged in agricultural tasks... Dutch overseers watching	-Expansion -Generation
6) Atmosphere	Daytime	daytime scene... earthy tones... balanced composition	-Style-context matching -Visual mapping/ grounding
7) Special Objects (if any)	'I don't know'	agricultural tools, water barrels, wooden carts	-Generation
8) Artistic Style	Any style suitable for historical scenery	colonial-era oil painting... earthy tones, rich greens, subtle splashes of colour... historically illustrative	-Style-context matching, -Visual mapping/ grounding

Table 4.2:

Test Case 01- Linguistic Transformation Table (Dutch Visitor's Perspective)

Dutch Visitor's Perspective			
Criterion	User Input Keyword(s)	Prompt Phrase(s)/ Expansions	Linguistic Operation(s) Observed by the Author
1) Topic	Surinamese Plantation during the colonial period	a day on a Suriname sugar plantation during the colonial era	-Specification (<i>adding historical context</i>)
2) Landscape/ Environment	Surinamese land	surrounded by towering palm trees and lush greenery	-Expansion -Enrichment
3) Architecture	Colonial plantation bungalow, labour quarters/ camps	a white plantation house with a red-tiled roof and wraparound veranda	-Specification -Enrichment -Omission

4) People	Dutch plantation owners, Surinamese workers	enslaved Africans... two European men in period attire	-Substitution -Specification
5) Activities	Slaves are working, and Dutch Masters overseeing it	toil in the fields... overseen by...	-Rephrase
6) Atmosphere	Daytime	...captures a day	<i>(no significant observation)</i>
7) Special Objects (if any)	Plantation-related objects	baskets, wheelbarrow, agricultural tools	-Generation
8) Artistic Style	Any style suitable for a historical scenery	A traditional oil painting	-Style-context matching

Table 4.3:

Test Case 01- Linguistic Transformation Table (Surinamese Visitor's Perspective)

Surinamese Visitor's Perspective			
Criterion	User Input Keyword(s)	Prompt Phrase(s)/ Expansions	Linguistic Operation(s) Observed by the Author
1) Topic	Surinamese Plantation during the colonial period	A historical 2D reconstruction of a Surinamese plantation during the colonial period	<i>(no significant observation)</i>
2) Landscape/ Environment	Surinamese land	set in the lush tropical Surinamese landscape	-Expansion -Enrichment
3) Architecture	Colonial plantation bungalow, Surinamese architecture	a colonial-style plantation bungalow with Surinamese architectural elements	-Synthesis -Omission
4) People	Surinamese workers, Dutch Plantation owners	Surinamese workers ... Dutch plantation owners in colonial attire	-Enrichment
5) Activities	My people (Surinamese) people working hard, Dutch plantation owners are tightly overseeing	Surinamese workers labouring in the fields...engaged in plantation activities such as harvesting and carrying loads... Dutch plantation owners ... sternly overseeing the work	-Rephrase (<i>Tone Neutralization?</i>) -Enrichment -Expansion -Generation
6) Atmosphere	daytime	under the hot daytime sun	-Expansion -Enrichment
7) Special Objects (if any)	plantation-related objects	typical plantation crops, tools, and equipment of the time	-Generation
8) Artistic Style	Any style suitable for a historical scenery	a realistic yet painterly tone suitable for a historical scene... emphasizing the atmosphere and social dynamics of the colonial era	-Style-context matching

However, it is noticeable that, under the criterion Architecture, the input “...*Colonial architecture*” has been expanded into “*grand plantation houses*” and “*white plantation houses with wraparound verandahs*” in Random and Dutch visitor perspectives, while completely omitting the user-requested input “*Surinamese architecture*”. Nevertheless, in Surinamese visitor perspective, it suddenly appears as “*colonial style plantation bungalow with Surinamese architectural elements*”. Furthermore, under the criterion People, “*Surinamese people/workers*” have been transformed into “*enslaved Africans, laborers*” in Random and Dutch visitor perspectives, but in contrast, the Surinamese visitor input (“*Surinamese workers*”) has been retained without any changes.

Moreover, under the criterion Activities, the conversion of “*My people working hard*” into “*Surinamese workers labouring...*” is clearly a prominent transformation in the Surinamese visitor prompt. In situations where, user is unfamiliar with the context, for instance Random visitor’s input stating “*I don’t know*”, under the criterion Special objects, the model itself has generated/synthesised relevant inventory of plantation-related objects for the prompt. Nevertheless, the Artistic style in all three perspectives match the historical context, leaning towards “*traditional oil painting*” for Random and Dutch visitor prompts, and “*realism*” for Surinamese visitor prompt.

4.1.2 Test Case 02: Beliefs associated with the heritage site of Sri Pada Mountain (Comparative Analysis of Linguistic Transformation across Visitor Perspectives)

This section presents a comparative analysis of six linguistic transformation tables (see Appendix 2- 108-113 pgs.), each representing perspectives on the sacred heritage site of Sri Pada Mountain in Sri Lanka, namely, the Buddhist pilgrim, Hindu pilgrim, Islamic pilgrim, Christian pilgrim (two variations), and a random visitor or hiker.

As pointed out in the six linguistic transformation tables, it is noticeable how a single heritage site is represented differently in accordance with the differing user perspectives. Unlike in the previous Surinamese test case, the common theme/topic (“*Historical Recreation of Sri Pada Mountain in...*”) has been rephrased by the system directly adding contextual information (attaching religious figures’ names) in Hindu and two Christian perspectives.

Following that, the linguistic operations ‘Specification’, ‘Enrichment’, ‘Expansion’ and “Visual Mapping”, have been extensively used as a recurring pattern across all perspectives, more than in the Surinamese test case. It is apparent that those operations have served as

mechanisms for refining vague and limited inputs into more elaborative prompts, to support image generation. Nevertheless, these elaborations are distinctively noticeable in terms of expanding spiritual and emotional dimensions of the scene, particularly in describing the criteria ‘Landscape/Environment’ and ‘Atmosphere’, such as, *“soft golden light”*, *“mist-covered summits”* and *“forested hills”* and *“cool mountain vistas”*. Further, these enrichments are quite poetic descriptions enhancing the sacred tone of the prompt across every perspective, especially when describing criterion People (i.e., religious figures), such as *“calm and radiant”*, *“benevolent”*, *“serene expression”*, *“solitary”*, *“solemn”* and *“soft golden halo”*.

However, in terms of the criterion ‘Architecture’, it is interesting to note that when the user states, *“no special buildings on the mountain summit”*, the prompts generated for Hindu, Islamic and Christian-I pilgrim perspectives describe a ‘large distinct indentation of a footprint on the summit’. This implies the existence of a pre-existing footprint before the arrival of the particular religious figure, which is an inaccurate addition to the prompt. Moreover, when the Random visitor responds with *“I don’t know”* in the same context, the system generated *“a small Buddhist temple atop the peak with a stupa and prayer flags”*. While this introduction of architectural elements is not inherently inaccurate, it implies that the AI model has drawn on the documented history of this site, ranging from early Anuradhapura Kingdom period of Sri Lanka to its present status, rather than considering the multiple intangible beliefs associated with the site.

Similarly, the rewriting of user inputs, especially the omission of original user input and synthesis of new activities by the system in relation to the criterion ‘Activities’, has added inaccuracies in the Islamic and both Christian perspectives. These additions imply that those respective religious figures were not placing their footprints, as suggested by user belief, but were merely gazing at an existing footprint or engaging in unrelated or passive behaviour. This misinterpretation has distorted the user’s intended narrative and the associated religious interpretations.

Nevertheless, the artistic styles proposed by the system for user’s input of *“any style suitable for a historical scenery”*, are notably different, though they aligned with the intended sacred and serene context. For instance, the prompt for the Buddhist perspective applies *“realism”*, while the Hindu perspective termed as *“earthy tones”*. The Islamic, Christian-II and Random visitor perspectives adopt a *“traditional style (oil or acrylic) painting”*, while the Christian-I perspective is proposed in a *“romantic-style oil painting”*.

4.1.3 Test Case 03: Dutch occupation of Portuguese Galle Fort in 1640 (in Sri Lanka) ***(Comparative Analysis of Linguistic Transformation across Visitor Perspectives)***

The comparative analysis of the four linguistic transformation tables (see Appendix 2, pp. 114-117 pgs.), developed for each perspective, the Dutch, Portuguese, Sri Lankan, and Random visitor, reveals both common and specific applications of linguistic transformations across the eight criteria.

While the theme provided by the user remains consistent across all four perspectives, the model has applied noticeable linguistic transformations in the Dutch and Portuguese visitor prompts. In the Dutch visitor prompt, the user's input has been rephrased as "*...the dramatic siege of Galle in...*", reflecting a tonal addition. Meanwhile, the Portuguese visitor prompt has entirely omitted the user's phrasing and has substituted it with "*...a fierce daytime siege at a coastal fortress during the Age of Sail*", which is vague and less specific without the required contextual details.

Similar to the previous two test cases, the linguistic operations of 'Expansion', 'Enrichment', and 'Specification', are prominently evident, especially in the criteria of Landscape/Environment, Architecture, People and Activities.

A comparison of the 'Landscape/Environment' criterion illustrates how the model has processed the same user input across the four perspectives. The Dutch visitor prompt is notably brief and general, using the phrase "*...set along Sri Lanka's southwestern coast*", and omitting more keywords specific to the context. This prompt expression has gradually enriched with contextual details across the perspectives, in which finally in the Random visitor's prompt, which is highly descriptive, states "*curved shoreline*", "*palm trees swaying*", "*coastal stone ramparts*", all of which accurately reflect the real-world setting.

This gradual descriptive trend is similarly apparent in the Architecture criterion as well. The user's input "*Galle Fort*" has been transformed into "*...fort in ruins*" in the Dutch prompt, while the Portuguese prompt adds further elements with, "*...stone fort walls with bastions and towers, some damaged*". With such a gradual expansion, the Random visitor prompt states a lengthy and highly descriptive account of the fort, incorporating factual architectural elements.

Furthermore, it is significant to note that in Dutch and Sri Lankan visitor prompts, no significant modifications can be observed under the People criterion, as "*Dutch troops/soldiers and Portuguese troops*" have remained almost the same with no enrichment. In contrast, the Portuguese and Random visitor prompts have been significantly enriched and specified such

as: *“Portuguese defenders in red tunics and armour stand firm”*, *“Dutch assailants, with pikes and muskets”*, *“Dutch soldiers in 17th century European armour and wide-brimmed hats”*, and *“Portuguese defenders also in period armour with red sashes and steel morions”*.

Moreover, another significant feature is the model’s reaction to the criterion Activities. The Dutch and Portuguese user inputs are highly detailed, including the names of the commanders leading each side. However, the model has rephrased these details, either omitting them entirely or simplifying the input by extracting only the core meaning, without including specific names or additional information.

Conversely, the user input by the Sri Lankan and Random visitors, which are initially very brief (i.e. *“Dutch troops attacking and Portugueses defending”*), have been extensively expanded and enriched, by adding information regarding troop positioning, the nature of the attack, defensive mechanisms, and more. Similarly, the most elaborated inventories of battlefield-related objects appear in these same two prompts.

The transformation of the user input *“battlefield”* under the criterion Atmosphere also reflects expansion, enrichment, and visual mapping, in almost every prompt. The prompts have been enriched with sensory details such as *“...smoky sky”*, *“vivid interplay of light and shadow”*, *“...rich, warm earth tones, smoke, and fire”*, and *“golden-yellow tropical sunlight filtered through battle smoke”*. Regarding the criterion Artistic style, all for prompts have been generated with *“classical historical painting”*, with some being extra specific, such as *“vibrant oil painting”*. However, the Random visitor prompt optionally suggests flexibility, stating *“oil painting or digitally rendered in classical historical style”*.

Overall, as presented in tabular format in Appendix 2 and comparatively analysed in this Chapter 4.1, it is evident that the user input prompts have undergone extensive transformations in the system-generated prompts, which were later translated into images. Hence, it is important to note that these intermediate linguistic transformations happened between the user and system-generated prompts have had both positive and negative impacts on the resulting images. These effects will be discussed in detail in Chapters 4.2 and 5.1.

4.2 Image Analysis: Comparative Evaluation of the Historical Accuracy of the AI-generated Images



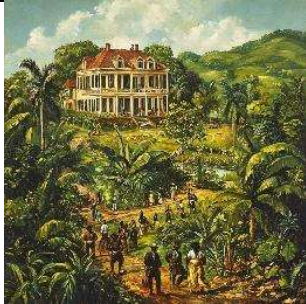
As discussed in Chapter 4.1, once users input prompts in keyword format, the selected AI model (ChatGPT 4o) created system-generated prompts, which served as the basis for image generation. Accordingly, 13 such prompts across three heritage test cases, were input into the three selected AI models (DALL-E, Midjourney, and Stable Diffusion), generating a total of 39 images (13 x 3).


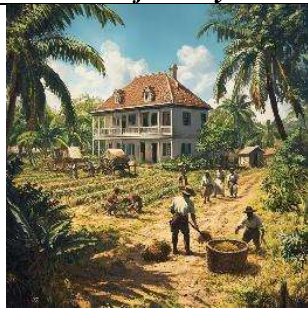
These 39 images were then evaluated using the eight criteria defined in Chapter 3.2.4 (i.e., Landscape/Environment, People, Activities, etc.), to assess how accurately each element has been represented. The objective was to determine how effectively the models have translated the essence of the textual prompts into semantically meaningful and contextually relevant visual outputs. The initial evaluation was conducted by three heritage experts, as detailed in Chapter 3.2.5. Based on the expert feedback regarding historical accuracy, a comparative analysis was conducted by the author, as presented in this Chapter 4.2.

As further detailed in Chapter 3.2.5, the 39 images have been comprehensively cross analysed to evaluate the following:

- Representation of a single perspective across three (03) models
- Representation of multiple perspectives of a single test case within a single model
- Representation of multiple test cases within in a single model

4.2.1 Test Case 01: Surinamese Plantations during Colonial period

Random Visitor's Perspective		
<p><i>"A colonial-era oil painting showcases a day on a Surinamese plantation, with enslaved African workers toiling under the watch of Dutch overseers. In the foreground, laborers engage in agricultural tasks, while a grand plantation house with Dutch architectural elements dominates the background, surrounded by lush tropical vegetation. The piece employs earthy tones, rich greens, and subtle splashes of color to create a balanced composition, focusing on the contrast between the workers and their overseers."</i></p>		
DALL-E	Stable Diffusion	Midjourney
		

Dutch Visitor's Perspective		
<p><i>"A traditional oil painting captures a day on a Suriname sugar plantation during the colonial era. At its heart is a white plantation house with a red-tiled roof and wraparound veranda, surrounded by towering palm trees and lush greenery, while enslaved Africans toil in the fields nearby, overseen by two European men in period attire."</i></p>		
DALL-E	Stable Diffusion	Midjourney
		
Surinamese Visitor's Perspective		
<p><i>"A historical 2D reconstruction of a Surinamese plantation during the colonial period, set in the lush tropical Surinamese landscape. In the scene, a colonial-style plantation bungalow with Surinamese architectural elements stands prominently. Surinamese workers are seen laboring in the fields under the hot daytime sun, engaged in plantation activities such as harvesting and carrying loads. Dutch plantation owners in colonial attire are sternly overseeing the work. The environment features typical plantation crops, tools, and equipment of the time. The artistic style reflects a realistic yet painterly tone suitable for a historical scene, emphasizing the atmosphere and social dynamics of the colonial era."</i></p>		
DALL-E	Stable Diffusion	Midjourney
		

Representation of a single perspective across three (03) models

Random Visitor's Perspective

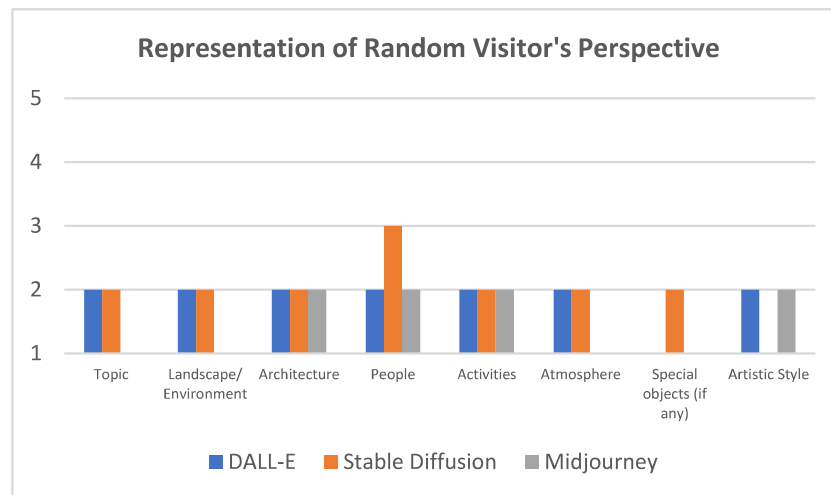
(*Please see Figure 4.2 for the numerical rating (Likert scale) assigned by the heritage expert to each image per criterion)

- a. Topic:** The images generated by DALL-E and Stable Diffusion partly represent the user-requested theme, while the Midjourney image does not represent the theme accurately.
- b. Landscape/Environment:** The DALL-E and Stable Diffusion images are only partly acceptable in terms of landscape/environment depiction, due to several discrepancies in the

images. For example, the plantations in Suriname were typically located close to rivers, a key feature missing in both images. Additionally, the landscape appears overly open and expansive. The Midjourney image is contextually inaccurate, portraying hills and overly dense vegetation, which do not align with the actual geography of Suriname.

Figure 4.2:

Test Case 01- Expert Rating for Representation of Random Visitor's Perspective



Note. Developed by the author for the present study

- c. Architecture:** The architectural elements in all three images are only slightly acceptable. The plantation house in each case does not resemble Surinamese structures. According to expert feedback, the prompt “*large plantation house*”, may have resulted in buildings that look more like Southern U.S. mansions, or overly grand houses from Paramaribo placed in an unlikely plantation setting, both of which are historically inaccurate.
- d. People:** The portrayal of people in the DALL-E and Midjourney images is only partly acceptable. In the DALL-E image, the Dutch figures are overdressed for the tropical climate and appear too formal, which is contextually inaccurate. Meanwhile, the Surinamese people are dressed in unlikely colours and types of clothing, and appear noticeably well-fed, which is also inaccurate representation given the contemporary scarcity of food. In the Midjourney image, the figures are too vague to be clearly identified, and Dutch individuals are barely visible. The representation of people in the Stable Diffusion image is moderately acceptable, though it still lacks Dutch figures.
- e. Activities:** All three images are partly acceptable in depicting activities. However, the Midjourney image does not portray the activities well enough as the user has intended.
- f. Atmosphere:** The atmosphere in the DALL-E and Stable Diffusion images is only partly acceptable, while the Midjourney image does not match the context. As per the expert

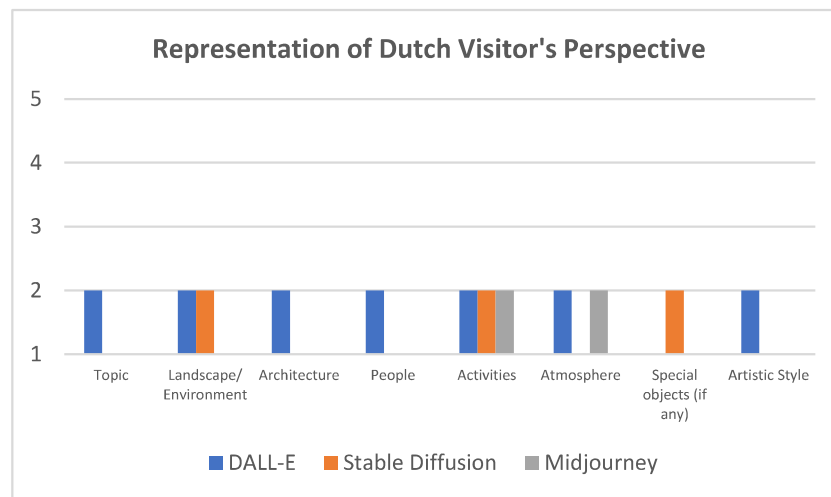
feedback, the scenes are romanticised and lack any indication of the violence and hardship historically associated with the setting.

- g. Special objects (if any):** The images of DALL-E and Midjourney do not clearly represent special objects, while the Stabel Diffusion image is somewhat acceptable in this regard but could have portrayed more context-specific details.
- h. Artistic Style:** The Artistic styles used in the DALL-E and Midjourney images are partly acceptable for representing this historical context, however, the style used in the Stable Diffusion image, does not align with the theme and the context.

Dutch Visitor's Perspective

Figure 4.3:

Test Case 01- Expert Rating for Representation of Dutch Visitor's Perspective



Note. Developed by the author for the present study

- a. Topic:** In terms of theme representation, the DALL-E image is partly acceptable. However, the Stable Diffusion and Midjourney images do not align well with the intended topic.
- b. Landscape/Environment:** The DALL-E image and Midjourney images are partly acceptable, although the Stable Diffusion image shows an overly diverse range of vegetation. Moreover, the Landscape/Environment representation in the Midjourney image is not acceptable, as it includes visible hills, which are inaccurate for Suriname's land. The absence of rivers near the plantations is also a noticeable omission.
- c. Architecture:** The Architectural structures in the DALL-E image are partly acceptable. Notably, it has included representations of Surinamese people's huts, which had not appeared in previous images. However, the architectural elements in the Stable Diffusion

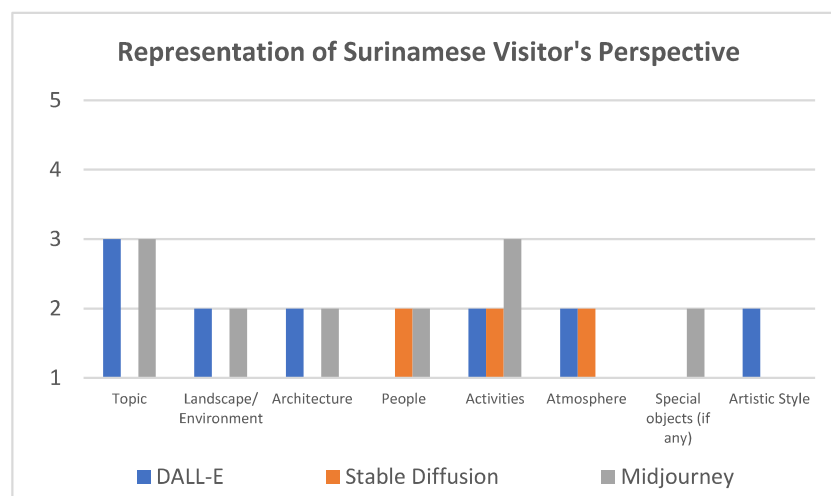
and Midjourney images are not acceptable, as they resemble large house models from Paramaribo or the Southern US, which are historically inaccurate for a plantation context.

- d. People:** The representation of people in the DALL-E image is partly acceptable. However, the overdressing of Dutch figures and well-fedness of the plantation workers is noticeably inaccurate to the context. The representations in the Stable Diffusion and Midjourney images are not acceptable, particularly due to vague and unclear depictions, inaccurate clothing, and the absence of Dutch individuals.
- e. Activities:** All three images partly acceptable in terms of depicting relevant activities. However, in the Stable Diffusion has not depicted the Dutch people overseeing plantation work. The Midjourney image lacks both the presence of Surinamese workers working in the fields, and Dutch people supervising them.
- f. Atmosphere:** The atmosphere is partly acceptable in DALL-E and Midjourney images, however, the Stable Diffusion image does not convey the atmosphere of a colonial plantation setting and is therefore not acceptable in this regard.
- g. Special objects (if any):** Special objects are not well represented in the DALL-E and Midjourney images. The Stable Diffusion image is only partly acceptable in this regard.
- h. Artistic Style:** The artistic style of the DALL-E image is partly acceptable. However, the styles used in the Stable Diffusion and Midjourney images are not matching enough for this historical context.

Surinamese Visitor's Perspective

Figure 4.4:

Test Case 01- Expert Rating for Representation of Surinamese Visitor's Perspective



Note. Developed by the author for the present study

- a. Topic:** Both the DALL-E and Midjourney image are reasonably acceptable in terms of theme representation. However, the Stable Diffusion image is not acceptable, as it does not align with the intended context.
- b. Landscape/Environment:** The DALL-E and Midjourney images are partly acceptable in terms of landscape/environment, while the Stable Diffusion image is not acceptable, as it resembles more Southeast Asian character, which is geographically and culturally inaccurate for Suriname.
- c. Architecture:** Similarly, the architectural elements in the DALL-E and Midjourney images are partly acceptable. The representation in the Stable Diffusion image, however, does not match and is therefore not acceptable.
- d. People:** The representation of both Dutch and Surinamese people in the DALL-E image is inaccurate. The clothing styles are incorrect, and the workers appear unrealistically well-fed, which interprets the historical conditions. In contrast, the portraits in Stable Diffusion and Midjourney are partly acceptable.
- e. Activities:** While the plantation activities are partly well represented in the DALL-E and Stable Diffusion images, the Midjourney image offers a relatively well depiction of such activities. Hence, it is considerably acceptable in this regard.
- f. Atmosphere:** Both DALL-E and Stable Diffusion images are partly acceptable in this regard, while the Midjourney image is more romanticized instead of an ambience of a colonial plantation.
- g. Special objects (if any):** Special objects are not clearly represented in the DALL-E and Stable Diffusion images. However, the Midjourney image is partly acceptable in this regard.
- h. Artistic Style:** The artistic styles of DALL-E image is acceptable. However, the styles used in Stable Diffusion and Midjourney images are not well-suited to the historical context.

[This space has been intentionally left blank]

Representation of multiple perspectives of a single test case within a single model

Please see Figure 4.4 - 4.6 or Appendix 2, p. 118, for the bar charts displaying the Likert scale ratings assigned by heritage experts to each model, based on the eight criteria (i.e., Topic, Architecture, etc.), in relation to the representation of three perspectives (Random, Dutch, and Surinamese) of heritage test case 01. Those ratings are analysed in this section below.

DALL-E

The chart titled '*Representation of Perspectives (DALL-E)*' (Figure 4.5), reveals a consistent rating of 2 across many criteria (Landscape, Architecture, Activities, Atmosphere, and Artistic Styles) for all three perspectives. The notable exceptions are for the criterion Topic, in which the Surinamese perspective received a slightly higher score of 3. Moreover, for the criterion Special Objects, all three perspectives have been received a minimal score of 1 (meaning, 'does not match'). Accordingly, the chart depicts an overall slightly fair performance in general across many perspectives.

Stable Diffusion

As per the chart titled '*Representation of Perspectives (Stable Diffusion)*' (Figure 4.6), across many of the criteria, such as Landscape/Environment, Activities, Atmosphere, and Special Objects, the rating scores as 2, particularly with Random visitor perspective. However, the criterion People of Random visitor perspective standouts with a slighter score of 3 as a noticeable feature. Moreover, the criterion Artistic Style has received the minimal score (score of 1) across all three perspectives, meaning the represented artistic style 'does not match'. In addition, the except the criterion Activities, the remaining criteria have at least one or two perspectives that have been scored as 1. Overall, this chart depicts an overall low performance across many perspectives.

Midjourney

In the chart titled '*Representation of Perspectives (Midjourney)*' (Figure 4.7), the Surinamese perspective shows slightly higher scores compared to other perspectives in general, particularly by receiving a score of 3 for the criteria Topic and Activities, and 2 for other criteria, except in Atmosphere and Artistic Style. However, the Random visitor perspective has been rated only for Architecture, People, Activities, and Artistic style, with a score of 2. The Dutch perspective is rated only in the criteria Activities and Atmosphere, with a score of 2. It is notable that criteria such as Topic, Landscape/Environment, Atmosphere, Special Objects and Artistic Style largely remain with minimal rating, across two perspectives each. Accordingly, this chart depicts an overall low performance across many perspectives.

Figure 4.5:

Test Case 01- Expert Rating for Representation of Perspectives in DALL-E

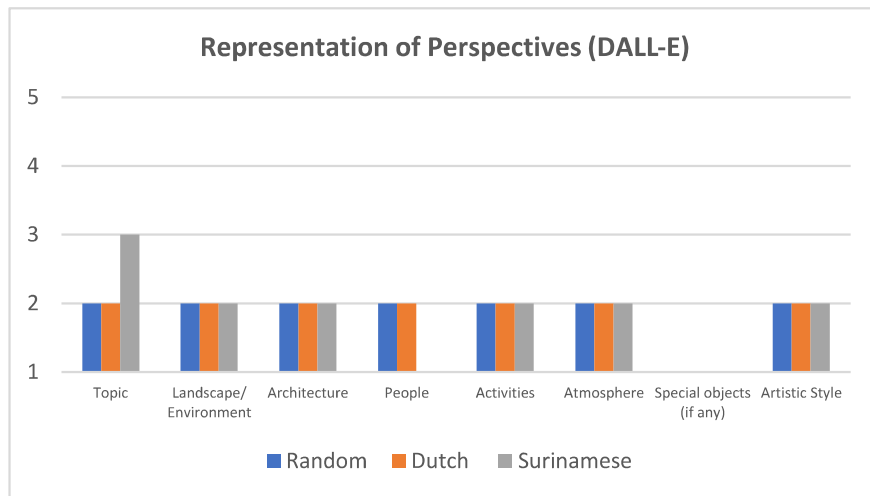


Figure 4.6:

Test Case 01- Expert Rating for Representation of Perspectives in Stable Diffusion

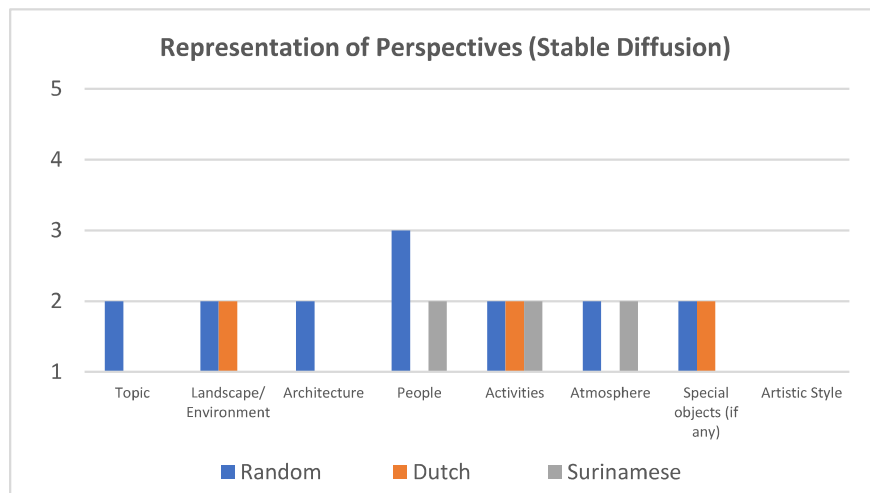
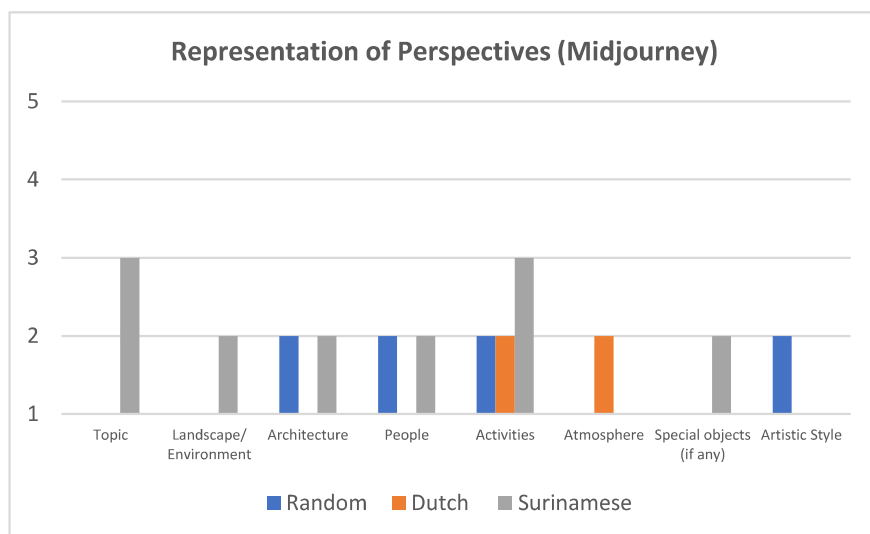







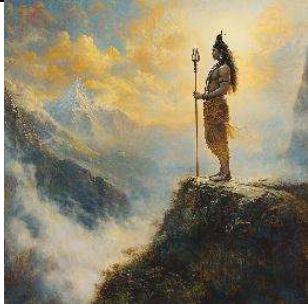
Figure 4.7:

Test Case 01- Expert Rating for Representation of Perspectives in Midjourney

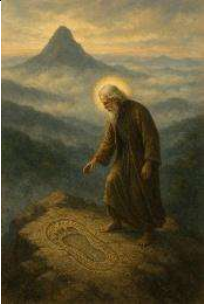
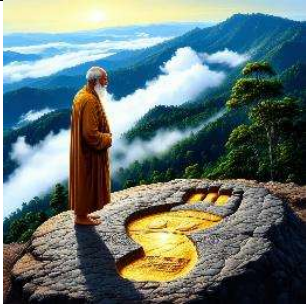
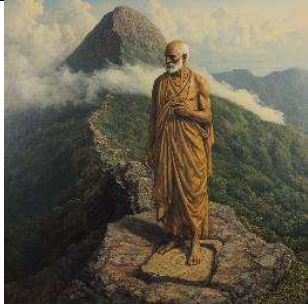


Note. Developed by the author for the present study

4.2.2 Test Case 02: Beliefs associated with the heritage site of Sri Pada Mountain




Buddhist Pilgrim's Perspective		
<p><i>“A historical recreation of the sacred moment on the summit of Sri Pada (Adam's Peak), Sri Lanka, during ancient times. The mountaintop is bare of any modern architecture—just a rocky summit surrounded by distant, misty forest-covered mountains and floating clouds. At the center stands Lord Buddha, calm and radiant, gently placing his sacred footprint on a rock. Beside him, God Sumana Saman, depicted as a benevolent guardian deity in traditional Sinhalese attire, respectfully invites and watches over the act. The atmosphere is divine and sacred, bathed in soft, golden morning light. The scene is peaceful and reverent, with no crowds—just a quiet, spiritual moment at the mountain summit. Style: realistic with subtle painterly touches, suitable for historical illustration.”</i></p>		
DALL-E	Stable Diffusion	Midjourney
		
Hindu Pilgrim's Perspective		
<p><i>“The painting portrays Lord Shiva standing proudly on a mountain peak, placing his right foot into an indentation in the rock, surrounded by soft, billowy clouds and a verdant, mist-filled landscape below. Capturing both serenity and strength, Shiva, with a trident in one hand and a serene expression, is enveloped in warm golden light as the cool mountain vistas fade into the distance, creating a beautifully balanced composition in earthy tones.”</i></p>		
DALL-E	Stable Diffusion	Midjourney
		
Islamic Pilgrim's Perspective		
<p><i>“A traditional-style painting, possibly executed in oil or acrylic, depicting the summit of Sri Pada (Adam's Peak) in Sri Lanka during a sacred moment. The scene shows a solitary, elderly man representing Al-Rohun (Adam from Islamic tradition), dressed in a flowing earth-toned robe, standing barefoot on the rocky summit. His face is solemn, illuminated by a soft golden halo. Before him is a large, deeply etched footprint on the stone, stylized with ornate, symbolic carvings. The mountain is surrounded by mist and distant clouds, with</i></p>		

forested hills visible far below. The atmosphere is sacred, peaceful, and filled with divine light, evoking the moment Adam first stepped on Earth."

<i>DALL-E</i>	<i>Stable Diffusion</i>	<i>Midjourney</i>
		

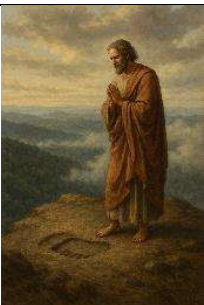
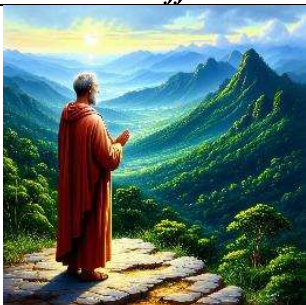
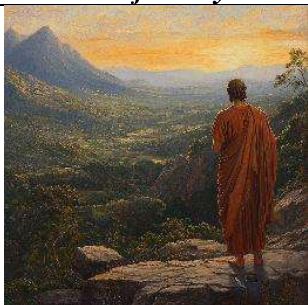
Christian Pilgrim's Perspective - I

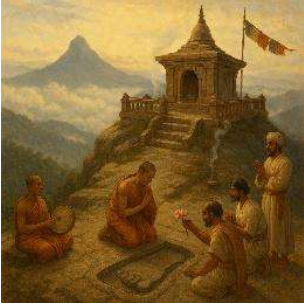


"This Romantic-style oil painting on canvas captures Adam on the summit of Sri Pada (Adam's Peak) in Sri Lanka during the day, his sorrowful gaze fixed upon a distinct footprint in the rocky ground beneath him. The scene is bathed in soft golden light, highlighting his rugged, muscular form while the mist-covered mountain backdrop fades into layers of green and blue—emphasizing both Adam's isolation and the expansive landscape surrounding him."

<i>DALL-E</i>	<i>Stable Diffusion</i>	<i>Midjourney</i>
		

Christian Pilgrim's Perspective - II

"This traditional oil painting presents St. Thomas the Apostle standing atop Sri Pada Mountain in first-century AD Sri Lanka, gazing at a noticeable rectangular footprint imprinted in the rocky terrain. The scene, set against a backdrop of lush hills and a glowing sky, conveys a solemn atmosphere, with meticulous details of St. Thomas' praying posture, textured garments, and the surrounding natural landscape blending warm earth tones with cooler greens and blues in the distance."

<i>DALL-E</i>	<i>Stable Diffusion</i>	<i>Midjourney</i>
		

Random Visitor's Perspective		
<p><i>"This traditional oil or tempera painting depicts Sri Pada (Adam's Peak) in Sri Lanka at dawn, showcasing a small Buddhist temple atop the peak with a stupa and prayer flags adorning the scene. Pilgrims gather around the significant stone footprint, engaged in prayer and meditation, while the serene landscape beneath misty mountains is illuminated by soft, golden light, skillfully capturing the depth and textures of the environment."</i></p>		
DALL-E	Stable Diffusion	Midjourney
		

Representation of a single perspective across three (03) models

Buddhist Pilgrim's Perspective

- a. **Topic Representation:** Based on expert feedback, the images generated by all three models, DALL-E, Stable Diffusion and Midjourney, align with the overall user-expected topic, although certain internal components may vary/mismatch. Notably, the expert has commented on the DALL-E image with the remark, '*Recreation or Re-creation*', questioning whether it represents an AI-generated recreation or a re-interpretation of a pre-existing artwork or concept, such as a famous style or a known depiction.
- b. **Landscape/ Environment:** While the landscape depicted in the DALL-E image is deemed acceptable, the Stable Diffusion image is only partly acceptable due to its resemblance to Far-Eastern rather than Sri Lankan environments. The Midjourney image contains notable discrepancies, particularly the central background mountain, which disrupts the compositional balance of the scene.
- c. **Architecture:** The representation of architectural elements in both the DALL-E and Stable Diffusion images are considered acceptable. However, the Midjourney image portrays architectural elements that are significantly inaccurate, as the depiction suggests that the mountain summit had previously been occupied. That is in contrary to the narrative that it remained virgin and was made holy by the Lord Buddha's footprint. Therefore, preserving the untouched sanctity of the mountain summit is necessary.
- d. **People:** The human representations in the DALL-E image are generally acceptable, although the deity appears slightly too stout. The Midjourney image is also acceptable,

though its figures are relatively vague in representation. However, the Stable Diffusion image portrays the deity *Sumana Saman* in a way that resembles an itinerant monk, making it only partly acceptable in terms of accuracy.

- e. **Activities:** Both the DALL-E and Midjourney images matches the expected activities involved in the belief of the user. While agreeing to the existing portrayal, the expert suggests an (optional) enhancement: depicting deity *Sumana Saman* in a kneeling posture, which is the much common posture appears in Sri Lankan mural paintings of deities and from Buddhist texts. The activity depicted in the Stable Diffusion image is barely acceptable, as it appears as a conversation rather than the central expected activity of placing Lord Buddha's footprint.
- f. **Atmosphere:** The Stable Diffusion image is acceptable in terms of conveying the sacred atmosphere that was intended to create. However, the DALL-E and Midjourney images are only partly acceptable, as they lack sufficient expression of holiness. The expert suggests enhancing the DALL-E image by making Lord Buddha's halo more radiant and visible to reflect his state of enlightenment and, adding foreground clouds in the Midjourney image to suggest the sacredness of the setting.
- g. **Special objects (if any):** All three images are acceptable in their representation of special or symbolic objects.
- h. **Artistic Style:** The artistic styles of the DALL-E image is partly acceptable. The expert notes a similarity to the style of renowned Sri Lankan artist *M. Sarlis* and suggests exploring whether the style of *Solius Mendis* (another famous Sri Lankan artist) would be more appropriate. The artistic styles of Stable Diffusion and Midjourney are acceptable, with a recommendation to consider using the styles found in the human figures of *Tivanka Pilimage* (Sri Lanka) and *Solius Mendis's* paintings at *Kelaniya temple* (Sri Lanka) as reference prototypes.

(*Please see Appendix 2, p. 119 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Hindu Pilgrim's Perspective

- a. **Topic:** All three images align with the overall user-expected topic. Additionally, the DALL-E image has been commented by the expert as a 'Re-creation'.
- b. **Landscape/ Environment:** While all three images are partly acceptable, landscape in the Stable Diffusion image appears somewhat foreign and not distinctly Sri Lankan. The

Midjourney image, in particular, appears overly alien and does not reflect a Sri Lankan environment.

- c. Architecture:** Architectural elements in all three images are considered acceptable.
- d. People:** The representation of people in all three images are only partly acceptable. A key omission across all is the absence of the god's vehicle, the bull Nandi, who according to Hindu texts, always accompanies the deity.
- e. Activities:** The DALL-E image is partly acceptable, but the deity is depicted in an anatomically inconsistent posture: meaning that when the right leg is raised, the left leg should be straight, whereas in the image the left leg is slightly bent. Additionally, it would be more accurate if the deity's right hand were raised slightly higher, so that the holy trident appears just above the head. In the Stable Diffusion image, the activity of imprinting the right foot on the rock lacks visual prominence, though it should be the central focus. The Midjourney image appears action-less and static, rather than dynamically engaging in the act of making the footprint. Therefore, both Stable Diffusion and Midjourney images are only partly acceptable in terms of activity portrayal.
- f. Atmosphere:** All three images are only partly acceptable in capturing the divine atmosphere. Hence, the expert suggests adding a red divine glow or light, since red is closely associated with this deity, would enhance the DALL-E and Stable Diffusion images.
- g. Special objects (if any):** The DALL-E image is acceptable in this regard. However, the Stable Diffusion and Midjourney images are only partly acceptable as the depiction of the footprint is absent, which is the central and most significant symbolic element.
- h. Artistic Style:** The artistic styles in the Stable Diffusion and DALL-E images are acceptable. The Midjourney image is only partly acceptable, as the deity (Lord Siva) appears disproportionately small in relation to the overall picture frame.

(*Please see Appendix 2, p. 119 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Islamic Pilgrim's Perspective

- a. Topic:** All three images are acceptable in terms of representing the intended theme.
- b. Landscape/ Environment:** Although the DALL-E image is acceptable in this regard, the Stable Diffusion image is only partly acceptable, as the environment appears too foreign and lacks Sri Lankan characteristics. The Midjourney image is also only somewhat

acceptable, the depiction of another mountain peak higher than Sri Pada Mountain and visually connected to it in background is both inaccurate and contextually inappropriate.

- c. Architecture:** All three images are acceptable with regard to architectural elements.
- d. People:** The representation of Prophet Al-Rohun (Adam in the Islamic tradition) is acceptable in both the DALL-E and Stable Diffusion images, However, the Midjourney image is only partly acceptable, as the figure appears significantly old.
- e. Activities:** The DALL-E image effectively conveys the intended activity, while the Stable Diffusion image is only partly acceptable, as the religious figure appears to be observing an existing footprint rather than actively placing one. Similarly, in the Midjourney image, Prophet Al-Rohun is portrayed looking away and does not appear to be engaged in the act of printing his foot, which is an inaccurate depiction of the narrative.
- f. Atmosphere:** The Midjourney image is acceptable in this regard, while the atmosphere in the DALL-E image requires slight improvement in terms of brightness, to emphasize the significance of Prophet Al-Rohun. However, the sacred atmosphere in the Stable Diffusion image appears foreign, hence is only partly acceptable.
- g. Special objects (if any):** All three images are only partly acceptable in terms of special objects. As per the expert suggestion, including a contextually relevant item, such as holding the holy Qur'an (book), would have enhanced the symbolic depth.
- h. Artistic Style:** The artistic style in all three images is acceptable.

(*Please see Appendix 2, p. 119 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Christian Pilgrim's Perspective -I

- a. Topic:** All three images are acceptable in terms of representing the intended theme. The expert considers the DALL-E and Stable Diffusion images as '*Re-creations*'.
- b. Landscape/ Environment:** The DALL-E image is only partly acceptable due to its overly dark background, which would have been better if more light was added to enhance the visibility. The Stable Diffusion image is acceptable, however, the landscape representation in Midjourney image appears too foreign and do not resemble a Sri Lankan setting.
- c. Architecture:** Architectural elements in all three images are acceptable.
- d. People:** The depiction of Adam in DALL-E image is only partly acceptable, as he resembles a primitive hunter-gatherer than a biblical figure. As per the expert's suggestion, a more appropriate portrayal would reflect the youthful, energetic Adam seen in High-Renaissance

paintings, such as works by Michelangelo or Raphael. Similarly, the Stable Diffusion image is only partly acceptable, as Adam appears more like a modern young man rather a historical or religious figure. Similarly, the Midjourney portrayal is also only partly acceptable as the figure of Adam is minimized, while the mountain in the middle ground dominates the composition. In addition, it is noticeable that Midjourney avoids showing the face of Adam.

- e. **Activities:** All three images are only somewhat acceptable in this regard, as DALL-E and Stable Diffusion images do not portray Adam actively setting his foot after arriving from the heaven. The Midjourney image is particularly misaligned, as Adam is shown in a sitting posture rather than engaged in the act of placing his footprint, which is not in accordance with the scenario.
- f. **Atmosphere:** The represented atmosphere in both the DALL-E and Stable Diffusion images is acceptable. However, the Midjourney image is only partly acceptable, as its overall mood and tone are less appropriate for the intended theme.
- g. **Special objects (if any):** All three images are acceptable in this regard.
- h. **Artistic Style:** The artistic styles used in all three images are only partly acceptable, as they do not highly suit the religious and historical tone of the themes. A High-renaissance style would be more fitting as per the expert's point of view.

(*Please see Appendix 2, p. 119 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Christian Pilgrim's Perspective -II

- a. **Topic:** All three images are acceptable in terms of representing the intended theme, and the expert considers all three images as '*Re-creations*'.
- b. **Landscape/ Environment:** The DALL-E Image is acceptable, while the remaining two are only partly acceptable. The landscaped in the Stable Diffusion image appears foreign and not representative of a Sri Lankan setting. In the Midjourney image, the Sri Pada Mountain is minimized due to the presence of disproportionately large background mountains, making it appear less like a summit, which is inaccurate in the real context.
- c. **Architecture:** Architectural elements in all three images are acceptable.
- d. **People:** The depiction of people is acceptable in all three images. However, it is noticeable that Midjourney has avoided showing the face of St. Thomas.
- e. **Activities:** All three images are only partly acceptable in this regard. In both the DALL-E and Stable Diffusion images, St. Thomas appears to be worshipping an already existing

footprint, rather than actively imprinting one. In the depiction by Midjourney, the sacred figure is turned back to the viewer and appears to be simply enjoying the view, rather than engaging on the act of imprinting the foot.

f. Atmosphere: The representation of atmosphere in all three images is acceptable.

g. Special objects (if any): All three images are acceptable in this regard.

h. Artistic Style: The artistic style of all three images is acceptable.

(*Please see Appendix 2, p. 120 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Random Visitor/ Hiker's Perspective

a. Topic: All three images are acceptable in terms of representing the intended theme, and the expert considers the DALL-E image as a '*Re-creation*'.

b. Landscape/ Environment: The DALL-E image is acceptable, while the remaining two are only partly acceptable. The Stable Diffusion image appears overly artificial and does not resemble a Sri Lankan landscape. In the Midjourney image, the Sri Pada Mountain is visually diminished due to a taller mountain in the background, which misinterprets the actual setting.

c. Architecture: All three images are acceptable in terms of architectural representation, however, the shrine depicted in Stable Diffusion image appears to have been influenced by architectural elements of Hindu shrines rather than Buddhist shrines.

d. People: The depiction of people in all three images is acceptable. However, the people in Stable Diffusion and Midjourney images can only be vaguely identified compared to the DALL-E image.

e. Activities: All three images are acceptable with regard to the activities shown, yet the depictions are vague and ambiguous due to the lack of prominence given to the representation of activities, praying and meditation.

f. Atmosphere: The DALL-E and Midjourney images are acceptable, while the Stable Diffusion image is only partly acceptable, as it appears somewhat unnatural.

g. Special objects (if any): All three images are acceptable in their depiction.

h. Artistic Style: The artistic style of all three images is acceptable. The expert notes that the Midjourney image resembles the style of 19th century landscape painting.

(*Please see Appendix 2, p. 120 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Representation of multiple perspectives of a single test case within a single model

Please see Appendix 2, p.123, for the bar charts displaying the Likert scale ratings assigned by heritage experts to each model, based on the eight criteria (i.e., Topic, Architecture), in relation to the representation of six perspectives (i.e., Buddhist, Hindu) of the heritage test case 02.

DALL-E

In the chart titled '*Representation of Perspectives (DALL-E)*', the overall ratings are consistently high across all perspectives and criteria. Criteria, including Topic, Architecture, and Special Objects, have been received almost highest score (5). The People and Activities criteria show slightly more variation, with the Hindu, Christian-I and Christian-II perspectives, by receiving scores of 3 and 4. The criteria Landscape/Environment, Atmosphere and Artistic Style also depicts such slight variations with Hindu and Christian-I perspective, while Atmosphere criteria depicts slight variations for the images generated for Buddhist, Hindu, and Islamic perspectives. Despite these slight differences, the model has maintained a strong overall performance in the image generations for the heritage test case 02.





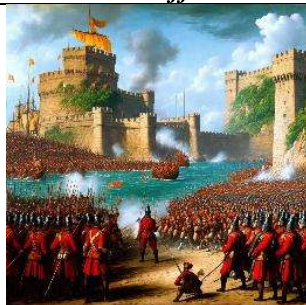
Stable Diffusion







The chart titled '*Representation of Perspectives (Stable Diffusion)*' shows a mixed performance with noticeable strengths in criteria such as Topic, Architecture, Special Objects, Atmosphere, and Artistic Styles, in which have received high ratings (either 4 or 5), across all the perspectives. However, the chart reveals a drop in performance for criteria such as People, Activities, and Landscape, in which mostly the ratings range from 2 - 4, depending on the perspective. For instance, the Activities criteria has been rated particularly low in Buddhist, Islamic and Christian-I perspectives. Despite these moderate differences, the model has maintained a stable overall performance in the image generations for heritage test case 02.

Midjourney

The chart titled '*Representation of Perspectives (Midjourney)*' illustrates that the Midjourney model performs strongly in the criteria Topic and Special Objects with most perspectives with either 4 or 5 as scores. However, its performance is notably less consistent in Landscape/Environment, in which the scores have significant drops to 2-3 in almost every perspective. The criteria Activities, Atmosphere, People and Artistic Style, show moderate variations, scoring between 3 and 5, depending on the perspective. It is noteworthy that for the criterion Architecture, all perspectives have been scored with 5, except for the Buddhist perspective, which scored significantly lower with a score of 1. Despite these differences, the model has maintained a durable performance for the heritage test case 02.

4.2.3 Test Case 03: Dutch occupation of Portuguese Galle Fort in 1640 (in Sri Lanka)

Dutch Visitor's Perspective		
<p>“This vibrant oil painting captures the dramatic Siege of Galle in the 17th century, with Dutch soldiers attacking the Portuguese-held fort under a smoky sky. Set along Sri Lanka's southwestern coast, the piece skillfully portrays the detailed confrontation between the forces, with the fort in ruins, Dutch ships in the distance, and a vivid interplay of light and shadow.”</p>		
DALL-E	Stable Diffusion	Midjourney
		
Portuguese Visitor's Perspective		
<p>“This oil painting vividly depicts a fierce daytime siege at a coastal fortress during the Age of Sail. On one side, Portuguese defenders in red tunics and armor stand firm, with a commanding officer leading the charge, while on the other, Dutch assailants, scaling walls and advancing with pikes and muskets, create a dynamic battle scene marked by rich, warm earth tones, smoke, and fire.”</p>		
DALL-E	Stable Diffusion	Midjourney
		
Sri Lankan Visitor's Perspective		
<p>“A historical 2D painting-style reconstruction of the Siege of Galle Fort in 1640, depicting the Dutch attack on the Portuguese-held fort in Sri Lanka. The setting is Galle Fort, located on the southern coast of Sri Lanka, surrounded by the Indian Ocean and the Bay of Galle. The scene features colonial-era architecture of Galle Fort, with stone ramparts and bastions under assault. Portuguese troops are positioned on the fort walls, defending with cannons and muskets, while Dutch soldiers, some arriving by wooden sailing ships in the harbor, launch an aggressive attack from both land and sea. Cannons fire, smoke rises, and flags of the Dutch Republic and Portugal are visible. The environment is intense and chaotic, capturing the atmosphere of a 17th-century battlefield. Artistic style should resemble a classical historical painting, rich in detail and drama.”</p>		

<i>DALL-E</i>	<i>Stable Diffusion</i>	<i>Midjourney</i>
		
Random Visitor's Perspective		
<p><i>“A historical recreation of the 1640 Siege of Galle Fort in Sri Lanka, during the Dutch-Portuguese War. The scene shows a fierce battle between Dutch and Portuguese troops in and around the coastal stone ramparts of Galle Fort. Dutch soldiers in 17th-century European armor and wide-brimmed hats advance from ships anchored in the Indian Ocean bay, firing muskets and cannons. Portuguese defenders, also in period armor with red sashes and steel morions, retaliate from fortified bastions and crumbling stone walls. The backdrop shows the curved shoreline of the Bay of Galle, palm trees swaying, smoke from cannon fire drifting across the sky. The architecture reflects the colonial European stonework of the Galle Fort — tall bastions, arched gateways, and gun emplacements. The atmosphere is tense and dramatic, with a golden-yellow tropical sunlight filtered through battle smoke. Historical accuracy is emphasized. Artistic style: oil painting or digitally rendered in classical historical illustration style.”</i></p>		
<i>DALL-E</i>	<i>Stable Diffusion</i>	<i>Midjourney</i>
		

Representation of a single perspective across three (03) models

Dutch Visitor's Perspective

- Topic Representation:** Based on expert feedback, the image generated by DALL-E is considered the most aligned with the user-expected topic from a Dutch Visitor's perspective. The other two images only partially reflect the historical scenario at Galle Fort during the 17th century.
- Landscape/ Environment:** DALL-E has been able to render the land-sea interface of the Galle fort up to a reasonably acceptable level. Midjourney provides a partial match, while

the image by Stable Diffusion, showing the fort on an elevated highland terrain facing the sea, is historically and geomorphologically inaccurate.

- c. Architecture:** In terms of fortification architecture, Midjourney offers a partial match to actual features of Galle Fort. However, the images produced by DALL-E and Stable Diffusion do not accurately reflect the historical architectural details.
- d. People:** The representation of people (soldiers) is acceptable only in DALL-E image; however, the attire does not accurately reflect the historical reality. While the depiction of people in Stable diffusion is incorrect in terms of clothing and other contextual details, the Midjourney's rendering is so vague that the figures are nearly unidentifiable.
- e. Activities:** While the DALL-E image is generally acceptable, the orientation and actions of some soldiers appear unrealistic and lack of clear meaning. The division between the opposing forces in the battle is also not very clear. The Stable Diffusion image partially aligns with historical activities but lacks accuracy based on historical records. Midjourney's depiction, showing the troops attacking the fort immediately after disembarking from a ship, is historically inaccurate.
- f. Atmosphere:** All three models partially succeed in capturing the atmosphere of a battlefield, though none are highly accurate in this regard.
- g. Special objects (if any):** The depiction of war-related objects such as weapons, swords, cannons, Dutch ships, and boats, is generally acceptable in DALL-E. The representations in the other two models are only partially acceptable and offer very limited details.
- h. Artistic Style:** While the DALL-E and Midjourney images are generally acceptable, the Stable Diffusion output has a more medieval appearance, lacking an alignment with the style of 17th century paintings.

(*Please see Appendix 2, p. 121 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Portuguese Visitor's Perspective

- a. Topic Representation:** In terms of the user-expected topic/subject representation, the image generated by DALL-E aligns moderately well. However, the other two images, Midjourney and Stable Diffusion, only partially reflect the historical battle, lacking with full depiction of both battle parties, but highlighting only the Portuguese side.

- b. Landscape/ Environment:** The fort and the land-sea interface partially match in the DALL-E and Midjourney images. In contrast, the Stable Diffusion image fails to represent this aspect correctly, showing a landscape that does not align with the Galle Bay setting.
- c. Architecture:** Regarding the architectural details of the Galle Fort, both DALL-E and Midjourney provide partially accurate depictions. However, DALL-E erroneously portrays the event as a battle between two forts, which is historically incorrect. The image from Stable Diffusion is completely inaccurate, misrepresenting the fortification style.
- d. People:** The representation of people (soldiers) in DALL-E image is somewhat acceptable, but notably biased, showing only Portuguese forces and omitting/ barely depicting Dutch soldiers, although the prompt refers to both. Midjourney also follows a similar pattern to DALL-E by only focusing on Portuguese forces. The Stable Diffusion image is inaccurate, by both parties been shown wearing identical red uniforms with medieval designs, deviating significantly from the historical records.
- e. Activities:** Both DALL-E and Midjourney images partly capture the battle activities expected by the prompt. However, it is noticeable that they primarily highlight the Portuguese actions while barely addressing Dutch soldiers' efforts, which is deviating from the prompt text. The prompt text mentions on both Dutch and Portuguese soldiers attacking and defending respectively. The Stable Diffusion image is clearly incorrect, portraying Portuguese soldiers attacking from the beach rather than from the fortifications.
- f. Atmosphere:** Both DALL-E and Midjourney have managed to create the battlefield ambience to some extent, however, the Stable diffusion image lacks the appropriate tone and mood to convey the gravity and ambience of the scene.
- g. Special objects (if any):** The expected battle-related objects are moderately well represented in the DALL-E image, but both Stable Diffusion and Midjourney show very limited objects, not up to a satisfactory level in this regard.
- h. Artistic style:** The DALL-E image partly matches a vivid 17th century oil painting style aligning somewhat with the historical period. In contrast, the Midjourney and Stable Diffusion images apparent more of generalized or medieval style, missing the era-specific aesthetic.

(*Please see Appendix 2, p. 121 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Sri Lankan Visitor's Perspective

- a. Topic:** The DALL-E image provides a generally acceptable depiction of the user-requested topic, aligning reasonably well with the historical theme. The Midjourney image also partially captures the intended subject but lacks clarity. Although Stable Diffusion image has made a general attempt, it is not acceptable to the given title.
- b. Landscape/Environment:** The landscape/environment is generally well-represented in the DALL-E image and somewhat acceptable in Midjourney. Both have attempted to incorporate elements resembling Galle coastline. Although the natural environmental and geomorphological elements slightly resemble, the overall landscape depiction is incorrect.
- c. Architecture:** While not fully accurate, the architectural portrayal of Galle Fort in both DALL-E and Midjourney is partially acceptable. In contrast, the Stable Diffusion image presents a completely inaccurate architectural setting that has no resemblance to Galle fort.
- d. People:** The representation of people in DALL-E image is partly acceptable, yet the colour of the uniform is not accurate as noted by the experts. Midjourney image is also somewhat acceptable, yet its vague portrayal of soldiers with minimal details makes historical verification difficult. The Stable Diffusion representation is also vague and not acceptable.
- e. Activities:** Depictions of battle-related activities in DALL-E are generally acceptable and show a reasonable level of alignment with the historical context. The Midjourney image provides a somewhat acceptable depiction, though missing details. The Stable Diffusion image is not acceptable as it does not reflect the correct historical battlefield activity.
- f. Atmosphere:** The battlefield atmosphere in DALL-E is moderately acceptable, with elements like lighting, smoke, and composition, that contributes to the sense of conflict. Midjourney also creates somewhat fitting ambience but not fully captured the tension and intensity. Stable Diffusion has not been able to create that necessary battlefield atmosphere for this historical scene.
- g. Special objects (if any):** DALL-E image includes a generally acceptable range of warfare-related objects, such as weapons, cannons, ships, and flags, necessary for the historical scene. Midjourney portrays a limited but has notably omitted guns and ammunition. The Stable Diffusion image includes only a very few, but they are contextually inaccurate.
- h. Artistic style:** The artistic style of the DALL-E image is partly acceptable with a moderate resemblance to 17th century aesthetics. While Midjourney also presents a slightly acceptable style, it requires more refinements to align with historical artistic resemblance. The Stable Diffusion style does not meet the stylistic expectations, and hence is not acceptable.

(*Please see Appendix 2, p. 121 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Random Visitor's Perspective

- a. Topic:** The DALL-E image partially represents the user-expected historical theme. The Midjourney image slightly addresses the topic but lacks accuracy. The depiction by Stable Diffusion is not acceptable for representing the historical scene.
- b. Landscape/ Environment:** The environmental elements in all three models are slightly convincing in terms of the geomorphology of Galle Bay. However, regarding the landscape, specifically the coastal terrain and fort-adjacent setting, only DALL-E offers a partially acceptable representation, and the remaining two are not acceptable in this regard.
- c. Architecture:** The image by DALL-E is partly acceptable in terms of capturing the fortification features of the Galle Fort, though it is not entirely accurate. However, the depictions by Midjourney and Stable Diffusion depict structures that resemble medieval European castles rather than Galle Fort, making the portrayals inaccurate.
- d. People:** DALL-E portrays people with more clarity and details, especially in terms of fine facial detailing of the frontmost figures, but the historical accuracy of costumes and other elements is only partially acceptable. Midjourney's depiction of people lacks distinct facial details and has issues with historical inaccuracies in attires. Stable Diffusion's portrayal is vague, with soldiers dressed in generic medieval military attire rather than clothing from the colonial era. Some human figures are disproportionate and incomplete.
- e. Activities:** The portrayal of battlefield activities in DALL-E is partially accurate, showing both Dutch attack and Portuguese defence. However, aspects such as the distance between opposing forces, their positioning, and the proportion of figures relative to the fort require refinement. Stable Diffusion's image does not align with the historical story, while Midjourney's image shows some resemblance, yet historically inaccurate.
- f. Atmosphere:** DALL-E image has partially captured the battlefield ambience through elements like colour tones, smoke and feeling of tension, creating the sense of conflict. Midjourney also slightly conveys the ambience, while the image by Stable Diffusion lacks the atmosphere of conflict, and the smoke and fire depicted does not really make a sense.
- g. Special objects (if any):** DALL-E includes a fair number of relevant objects accurate to the historical period. Midjourney includes some but has missed key objects like cannons and

guns. Additionally, it's ships carry a wrong flag instead of a Dutch flag. The Stable Diffusion depiction is not acceptable as it lacks many major elements.

- h. Artistic style:** The artistic style of DALL-E image partly reflects of a 17th century art piece resembling to the historic feeling yet requires improvements. Midjourney's style does not match to the intended historical period, while the Stable Diffusion's image does not fit into the period-accurate style nor any a style making a feeling of a historic painting.

(*Please see Appendix 2, p. 121 for the Likert scale rating assigned by the heritage expert to each image per criterion)

Representation of multiple perspectives of a single test case within a single model

Please see Appendix 2, p. 124, for the bar charts displaying the Likert scale ratings assigned by the heritage experts to each model, based on the eight criteria (i.e., Topic, Architecture, etc.), in relation to the representation of four perspectives (i.e. Dutch, Portuguese, etc.) of the heritage test case 03. Those ratings are analysed in this section below.

DALL-E

In the chart titled '*Representation of Perspectives (DALL-E)*', the model has overall received relatively high score in criteria such as People, Atmosphere, and Special Objects, with Dutch, Portuguese, and Sri Lankan perspectives, with a score of 4. In contrast, the same criteria in the Random perspective have been received lower ratings, particularly for the people criterion (a score of 2). Score of 3 is the most common rating score received across many criteria by many perspectives. The criteria Topic and Activities show slight variations, with Dutch and Random perspectives scoring slightly lower than the Portuguese and Sri Lankan perspectives. Overall, the Random visitor perspective appears to have lower scores compared to other perspectives within the DALL-E model. Despite these differences, the model has maintained a durable performance in the image generations for the heritage test case 03.

Stable Diffusion

The chart titled '*Representation of Perspectives (Stable Diffusion)*' illustrates that the overall ratings for the historical accuracy of the generated images are significantly low. Only the Dutch and Portuguese perspectives of the criterion Topic have been rated with a score of 2, while all other perspectives of that criterion have received the minimal score of 1 (meaning, '*does not match*'). In addition, only the Dutch perspective alone has received the marginal score of 2 across several criteria, including Activities, Atmosphere, Special Objects and Artistic Style. It is notable that, apart from the aforementioned ones, all the other perspectives (Portuguese, Sri

Lankan and Random visitor) across every criterion have been received only a score of 1 (meaning, '*does not match*'). Accordingly, this chart depicts an overall low performance across many perspectives.

Midjourney

In the chart titled '*Representation of Perspectives (Midjourney)*', the overall ratings are moderate, with no rating exceeding 3 on the Likert scale. The Sri Lankan perspective has been received consistent scores (either 2 or 3) for all the criteria. Except for the criteria special objects and Artistic Style, the Portuguese perspective has also been received a score of 2, and a score of 3 for the Atmosphere criterion only. The Dutch perspective has also been received a score of 2, except for the criteria People and Activities. The most notable feature is the Random perspective has been only received, a minimal score of 1 (meaning, '*does not match*'), for every criterion. In general, the chart depicts an overall slightly fair performance across many perspectives.

With that, the step-by-step data analysis process of this study, began with a linguistic analysis to examine how the user-given keywords were converted into system-generated prompts, followed by an image analysis, evaluating the historical accuracy of the AI-generated images, concludes here, paving the way for the results and discussion of the study in the next chapter.

Chapter 5:

RESULTS AND DISCUSSION

Building on the data analysis presented in Chapter 4, this chapter presents the key findings derived from the research and provides a detailed discussion of the research results. These outcomes are critically examined in relation to the predefined research objectives of the study. In addition to interpreting the results, this chapter also highlights the potential practical applications in the heritage field and suggests directions for future research.

5.1 The Strengths and Limitations of AI models in generating historical (image) reconstructions

According to the research design, the three selected generative AI models were tested in three different challenging historical/heritage contexts, each having multiple perspectives and interpretative complexities. Based upon the extensive data analysis conducted, the study identified the following strengths and limitations of these AI models in generating historical (image) reconstructions.

5.1.1 General Observations from the Data Analysis

Test Case 01: Surinamese Plantations during the Colonial period

The first challenging context given to the AI models involved generating images of a colonial plantation, a setting deeply intertwined with the history of slavery, a subject that is inherently sensitive and contested in interpretation, let alone visualisation. Accordingly, the test case comprised of three perspectives: a Random visitor, Dutch visitor, and a Surinamese visitor.

Based on the extensive data analysis, combining a linguistic and visual image assessments (including expert feedback), it was revealed that all three AI models struggled to produce historically accurate representations of this heritage scenario. According to the cumulative scores from expert ratings: DALL-E has received 42 points, while both Stable Diffusion and Midjourney scored 34, out of a possible 120. Numerically, DALL-E appears to perform slightly better than other two models, but the margin is very negligible.

Moreover, the bar charts of the numerical ratings presented in Appendix 4 (p. 122) clearly illustrate that even DALL-E model's scores are consistent in overall, it remains low receiving only a score of 2 points for each, indicating many discrepancies. None of the models have been scored higher than 3 for any criterion, across any of the perspectives. This is clearly explained under Chapter 4.2.1 in the criterion-wise image analysis as well.

Therefore, while each model demonstrates certain strengths and weaknesses, none can be considered clearly better in performance in the context of Surinamese colonial plantations. However, amidst this overall low performance, it is noticeable that the images generated for the Random visitor's perspective has been received relatively higher scores across most of the eight evaluation criteria, across all three models.

Test Case 02: Beliefs associated with the heritage site of Sri Pada Mountain

The second challenging context tasked to the AI models was to generate images for a sacred heritage site characterised by religious pluralism, requiring the models to manage multiple overlapping religious narratives. In addition, given that most AI models are typically trained on predominantly Western-centric datasets, it was initially assumed that depicting a South Asian religious setting would pose a significant challenge to the models. Accordingly, this test case included six perspectives: Buddhist pilgrim, Hindu pilgrim, Islamic pilgrim, Christian pilgrim-I, Christian pilgrim-II and Random visitor/hiker.

However, surprisingly, the models have performed well in this challenging context, especially when compared to the Surinamese colonial context. Based on the comprehensive data analysis, all three models have produced reasonably accurate (image) representations across all perspectives with more or less discrepancies in each. According to the cumulative expert ratings, DALL-E has achieved a total score of 225, Stable Diffusion 210, and Midjourney 202, out of a possible 240. Numerically, DALL-E has consistently performed better in almost every criterion across every perspective, than other two models. Nevertheless, the other two models also have managed to maintain relatively close scores.

Nevertheless, some of the Stable Diffusion images had clearly not represented a landscape/environment relatable to Sri Lanka (South Asia), instead resembling settings of a Far East Asia. Overall, although DALL-E leads in performance, all three models have demonstrated a commendable level of success in representing the diverse pilgrim/visitor perspectives in general.

Test Case 03: Dutch Occupation of Portuguese-held Galle Fort in 1640 (Sri Lanka)

The third challenging context given to AI models to visualise was a major historical battle in Sri Lanka, more specifically, a power struggle between European colonial powers on South Asian soil. The test case comprised of four perspectives: Dutch visitor, Portuguese visitor, Sri Lankan visitor and Random visitor.

Unlike the previous two test cases, this scenario revealed significant performance differences among the models. Based upon the comprehensive data analysis, all three models produced representations for each perspective but with varying degrees of historical accuracy. According to the cumulative expert ratings, DALL-E received a total score of 106, Midjourney 57, and Stable Diffusion 38, out of a possible 160.

Numerically, DALL-E has performed the best across all the perspectives, while Midjourney comes second, and Stable Diffusion low performance showing the lowest performance. Notably, the performance gaps are substantial, with each model's score nearly the half of the preceding one. Although DALL-E takes the lead in performance by generating more historically plausible visualisations, it still has weak points that need to be addressed.

5.1.2 Specific Observations from the Data Analysis

- **User's unfamiliarity to the context:**

The Linguistic analysis revealed that when vague, missing, or ambiguous keywords were provided by the user (i.e., "*I don't know*", "...*maybe?*"), the AI-model autonomously synthesises information to fill these gaps, which could be either accurate or inaccurate.

- **Imbalance in architectural representation:**

In the Surinamese test case, colonial plantation bungalows have been consistently and prominently depicted across all models. However, the user-requested Surinamese architectural structures, such as labour houses, have been largely overlooked. Only the DALL-E images for the Random and Dutch visitor perspectives have depicted these elements, while the remaining seven images in this test case have failed to represent Surinamese architecture or labour camps.

- **Sympathizing and neutralising narratives:**

In the Surinamese test case, the intermediate linguistic transformation process tended to have sympathised/neutralised with the Surinamese perspective, even without any explicit instructions do so. For instance, the user prompts for the Random and Dutch visitor perspectives have more directly reflected the factual historical realities (i.e., Conversion of "*Surinamese people...*" into "*enslaved African labour*" ...). In contrast, the user prompt for Surinamese visitor perspective has been transformed into a neutral description ("*My people (Surinamese) people working hard...*" into "*Surinamese workers labouring in the fields, engaged in plantation activities such as harvesting and carrying loads...*") This suggests that the AI model (ChatGPT-4o) has demonstrated a tendency to soften or neutralise

references to factual violent aspects of colonial plantation history. The author assumes that the model may have built-in capabilities for handling empathetic/sensitive language inputs.

- **Romanticisation of violent contexts:**

According to expert feedback, the images generated for the Suriname test case are often romanticised and lack any indication to reflect the historical violence and hardship inherent to the setting.

- **Emphasis on spiritual and emotional tone:**

In the Sri Pada Mountain test case, the intermediate linguistic transformation has prominently enhanced the spiritual and emotional aspects of the site. The system-generated prompts have often adopted poetic language that enhanced the sacred and serene tone of the images, particularly in the criteria of Landscape/Environment, Ambience, and Artistic style. This increases the viewers' emotional and spiritual connection with the images.

- **Errors in religious beliefs representation:**

A critical linguistic transformation error is identified in the Sri Pada Mountain test case, causing it to generate inaccurate image representations. Although the user intention was to depict the act of imprinting the footprint by the respective religious figures, the system generated prompts for the Islamic, Christian-I and Christian-II perspectives instead describes a pre-existing footprint at the time of their arrival to the summit (i.e., “...*He stands barefoot., before him is a large, deeply etched footprint on the stone...*”, “...*gaze fixed upon a distinct footprint in the rocky ground beneath him...*”, “...*gazing at a noticeable rectangular footprint imprinted in the rocky terrain...*”). This misinterprets the religious beliefs as well as leading to historically inaccurate image outputs.

- **Recreation and Re-creation in DALL-E images:**

In the Sri Pada Mountain test case, the heritage expert noticed that except for the Islamic pilgrim's perspective, all the other five perspectives (Buddhist, Hindu, Christian-I, Christian-II and Random visitor) appeared to be more of artistic re-creation of scenes, rather generative recreations. Additionally, he notices that some images seemed stylistically influenced by well-known Sri Lankan artists.

- **Imbalance in depiction:**

Particularly, the images generated for the third test case (Siege of Galle) often shows imbalance in depiction in Portuguese and Dutch forces, often with a marginal representation of the Dutch side, across all models. Surprisingly, the linguistic transformations in prompts do not reflect such an imbalance in description. For instance, while the Portuguese visitor

perspective often reflects imbalance in image depiction, the images depicting Dutch visitor perspective still highlights the Portuguese forces. This pattern confuses the author with an impression that the gen-AI models may be inclined to empathise with the defeated or weaker party when visualising historical conflicts.

- **Difficulties in depiction:**

Further, it is noticeable that across the three test cases, generating images for factual historical events (the first and third test cases) has been more challenging for the models than representing beliefs and spiritual narratives (the second test cases). However, the author assumes that this challenge is not due to the historical nature of the depictions, but rather the highly detailed nature of the first and third test cases, while the second test case is relatively less descriptive in its representation expectations.

- **Portrayal of people:**

Notable differences can be observed in how each model has portrayed human figures. Stable Diffusion has produced vague depictions of people in general, while the Midjourney has often avoided showing faces of people by presenting their back views. In contrast, DALL-E images have consistently portrayed people with clear and detailed faces, prominently placed in the foreground, with noticeable facial expressions.

- **Use of colour and style:**

Despite prompts requesting traditional oil painting styles and historical scenes, Stable Diffusion images have consistently used bright, high contrast colours, which experts remark as unrealistic and less effective for historical scene depictions.

- **Overall impact of intermediate linguistic transformation on historical accuracy:**

After analysing the linguistic transformations between user and system-generated prompts, as well as the generated images, it is important to emphasise that a key factor behind the lower scores for historical accuracy of the generated images is, the errors and misinterpretations occurred during the intermediate linguistic transformation process. This includes omissions or inaccurate reinterpretations of user-provided keywords, which is consistently observed across all perspectives in all three test cases.

- **Influence of Experts' perspectives in evaluation:**

The expert's evaluations have also largely influenced the results and observations regarding the historical accuracy and cultural/contextual sensitivity of the AI-generated image outputs. For a fully balanced assessment across the three heritage case studies, it would have been ideal if the same expert/ team of experts evaluate all the test cases to ensure consistency.

However, since the study involved varied nature of test cases, different experts were required based on their respective specialisations. This naturally adds varying levels of criticality and personal interpretive perspectives, for instance, tendencies of romanticising certain representations. Therefore, this variability is acknowledged as an important factor in interpreting the results and has been reflected in the data analysis as well.

5.1.3 Additional Variables

In addition to the main experimental design, the author made an effort to identify some additional factors that may influence the historical accuracy and cultural/contextual sensitivity of the historical (image) reconstructions generated by AI models.

User's familiarity with the historical/ heritage context

Based upon the general understanding of how generative AI functions, and as evidenced in the data analysis and findings presented in Chapters 4.1, 4.2, and 5.1.2, it is clear that both the user prompt and the system-generated prompt significantly affect the quality and historical accuracy of AI-generated images. As He et al. (2025) also points out, the user's familiarity with the historical/ heritage context is vital, as it directly impacts the relevance and specificity of the input prompt. Naturally, a user with contextual knowledge is more likely to provide precise and contextually appropriate keywords, leading to producing images that are more historically and culturally/contextually accurate. In converse, as observed across all three test cases for the 'Random visitor perspective', when user input vague or ambiguous keywords such as "*I don't know*", "*...maybe?*", the AI-model autonomously synthesise information for these void filling. Hence, there is a possibility that such generated content may be partially or entirely inaccurate.

Training dataset of the AI model

Another crucial factor that influences the quality and accuracy of historical reconstructions is the training dataset on which the AI model has been developed. Even when the users provide precise and contextually accurate input prompts, the resulting system-generated prompts and the corresponding images are generated based on the dataset that the model is trained on. Scholarly studies have noted that many generative AI models are not balanced in representation and often predominantly trained on western-centric datasets (He et al., 2025; Liu et al., 2025). Moreover, they generally cover major cultures of the world, while underrepresenting minor or regional cultures. This imbalance affects the accuracy of the generated output, even when user inputs are highly detailed and correct.

Non-existence of accessible digital data sources

Moreover, it is important to note that, not all historical data sources in the world exist in digital format. Many libraries, archives, and museums house vital records that remain undigitized, limiting the AI model's ability to access such sources. Unlike the modern knowledge, historical information in many countries are still stored using traditional, non-digital means. Moreover, some forms of knowledge, such as intangible cultural beliefs, may not be formally recorded, posing an additional challenge for AI models in understanding and accurately representing heritage/historical scenarios. For instance, the gen-AI models may not have many sources to make references in representing human beliefs such as in the test case 02 of this study.

User's input language

As described in Chapter 4.1, this study used ChaptGPT-4o to process user-provided keywords and generate its corresponding system-generated prompts for image creation. Since the supports multilingual input, the author experimented a few of the same user input keyword prompts in author's native language, instead in English, beyond the main experimental design. Although the resulting images are generally acceptable, the author observed several errors in those images and those inaccuracies have stemmed from the intermediate, unseen step when the user input was translated into English before generating the system prompt. The minor translation errors at this stage have led to misphrasings and, subsequently to inaccuracies in the final images. Although this factor was tested only with the author's native language, it should be considered as an additional variable that may influence image generation for any user.

Providing additional data sources as references (One-shot or Few-shot prompting)

While this study is focused on zero-shot text-to-image generation, it is worth noting that the users may prefer to enrich their input prompts with additional reference materials, such as images, maps, weblinks, or academic articles. Incorporating such additional information could substantially influence the accuracy and contextual relevance of the generated images. However, this approach needs further research in future to systematically assess its impact.

Accordingly, these additional variables demonstrate that the level of accuracy of AI-generated historical (image) reconstructions depends not only on the model's capabilities but also on different other factors. Therefore, the author suggests that acknowledging these influences is essential for improving prompt design and, also for developing more contextually reliable AI applications for cultural heritage field.

5.1.4 Potential of Generative AI in fostering meaningful dialogue on multivocality in heritage

Despite the limitations and challenges showcased by the gen-AI image models in this study, the findings reveal that generative AI holds significant potential as a tool for encouraging dialogue on multivocality in heritage contexts. As demonstrated across the three test cases, the models indeed often struggled with historical accuracy. Nevertheless, they were able to visualise multiple perspectives/viewpoints within a single heritage scenario. Hence, it offers a novel opportunity to engaging audiences/communities with complexities in heritage interpretation.

For example, in the Sri Pada Mountain test case, the AI models were able to capture the distinct spiritual atmospheres for Buddhist, Hindu, Islamic, and Christian pilgrims, highlighting the layered and overlapping beliefs that are associated with that sacred heritage site. Similarly, even in the more challenging Surinamese and Galle Fort contexts, the AI generated prompts and visuals reflected varying standpoints of different visitors, including local, colonial, and neutral perspectives. Although they are indeed imperfect, it demonstrates how generative AI can serve as a starting point for critical conversations about how complex histories can be viewed, interpreted, and remembered.

Moreover, the tendency of AI models to simplify, romanticise, or neutralise sensitive aspects, which was clearly evident in the Surinamese plantation test case, and to some extent in Siege of Galle Fort test case, emphasise the need for scholarly reflections on how to interpret these generated visuals and to what extent these can be incorporated into academic dialogues. This could lead to discussions on how these contested historical narratives have been framed and whose voices are amplified or marginalised.

Therefore, while generative AI alone cannot resolve issues regarding historical biases or representation of unheard voices, its ability to produce multivocal visual narratives may provide heritage professionals, educators, and communities with a valuable tool for participatory engagement. Hence, if used critically and in combination with expert input, generative AI can foster richer and more inclusive discussions around complex and contested pasts and plural heritage meanings.

5.2 Conceptual Framework cum Guideline for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions

One of the objectives of this study is to identify key factors to consider when providing user prompts, to improve the historical accuracy and cultural/contextual sensitivity of AI-generated historical visual outputs. Therefore, the research assessed 8-criteria throughout the research process (i.e., People, Landscape/Environment) across three heritage test cases and three AI models. Based on the findings from the extensive data analysis, this research study proposes the following Conceptual framework cum guideline for optimising user prompts in this context.

Conceptual Process Framework cum Guideline for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions

OBJECTIVE:

The objective of this framework cum guideline is to assist the users to optimise their user prompt to obtain AI-generated historical (image) reconstructions with improved historical accuracy and cultural/contextual sensitivity.

TARGET AUDIENCE:

- General Public
- Professional Users (Heritage Professionals, Educators)

PROCESS FRAMEWORK:

1. Identify the exact requirement:

Clearly think and define the type of image you want to create and its intended purpose/use.

2. Describe input criteria/ variables:

While users may provide prompts in paragraph form, it is recommended to use a clear, structured keyword format (a sample template is provided in the guidelines sections of this document). Key criteria to cover include:

- Topic:** Describe the event/ scenario that is needed to be depicted (i.e., a battle scene, a historical marketplace, etc.), specifying the time period as precisely as possible (i.e., 1640 AD, 17th century, or at least ‘Medieval period’, ‘Colonial era’, etc.)
- Landscape/Environment:** Include geographical features, time of the day/season, lighting conditions, particular flora, and fauna species if any, and, if possible, details about the foreground, middle ground, and background.

- (iii) **Architecture:** Mention notable structures, construction materials used, distinctive design styles, and any culturally relevant architectural details.
- (iv) **People:** Describe clothing, specific ethnicities, social roles (i.e., soldiers, kings, farmers), and other distinguishing features.
- (v) **Activities:** Specify any particular actions or interactions expected to be depicted, beyond general scene setting.
- (vi) **Atmosphere/Ambience:** Define the desired mood or emotional tone of the image.
- (vii) **Special Objects:** Include relevant tools, furniture, models of transportation, machinery, or any other distinctive objects
- (viii) **Artistic Style:** Indicate any preferred artistic style that aligns with the historical context (i.e., traditional oil painting, realistic painting, etc.)

3. Generate the prompt using the model:

Once the user keywords are defined, input them into a generative AI model with NLP capabilities and integrated image generation (i.e., ChatGPT-4o, Gemini, Bing AI) A sample input template is provided in the guidelines section below.

4. Review and refine the system-generated prompt:

After providing the initial user prompt, the model will generate a refined prompt suitable for input into any image generation tool. If the first attempt does not meet your expectations, you may iterate by adjusting the keywords or interacting with the AI through a conversational interface (if available). As mentioned in step 3 above, AI models with strong NLP and chat-based interfaces are particularly useful for this prompt refinement.

5. Generate images using gen-AI imagery models:

Use the final prompt with your preferred AI image generation tool (i.e., DALL-E, Midjourney, Imagen, Stable Diffusion, etc.). It is recommended to experiment with multiple image generation models simultaneously to compare results and select the most accurate and contextually appropriate output.

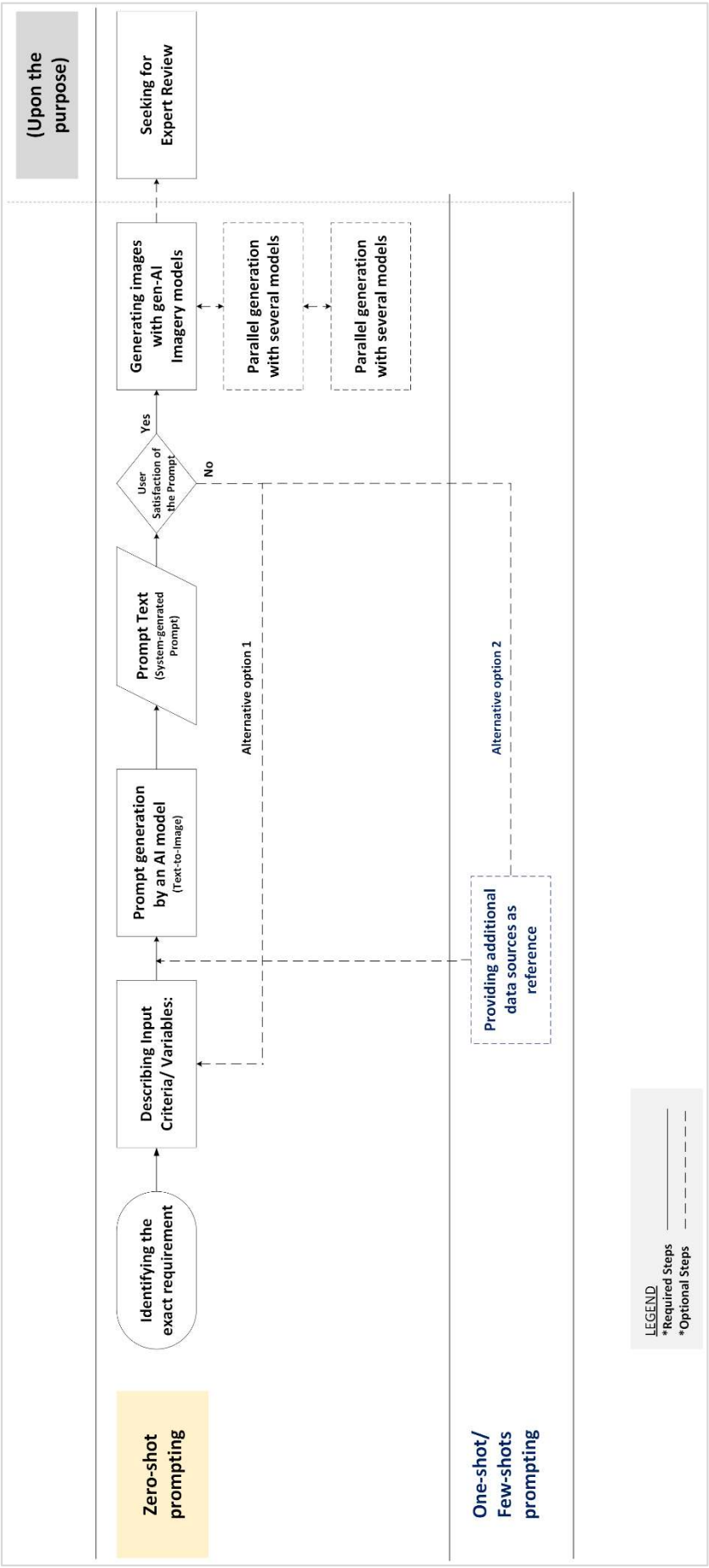
6. Seek expert review, if applicable:

Depending on the intended purpose/use (especially for professional or educational purposes), it is highly recommended to seek expert review of the generated images. Expert feedback helps you verify historical accuracy and contextual appropriateness.

See Figure 5.1 for the flowchart of the Conceptual Process Framework for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions

Figure 5.1:

Conceptual Process Framework for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions



Note. Developed by the author for the present study

GUIDELINES:

Do's:

- Be as clear, specific, and detailed as possible when providing keywords. The more precise your input, the better the output visuals will be.
- Respect the cultural and historical significance of the heritage content you are recreating. Please remember that heritage images may represent identities, memories, and beliefs that are sensitive to various communities.
- Review your prompts before submission to image generation, to ensure they do not contain ambiguous or contradictory information.

Don'ts:

- Avoid using vague or overly broad terms as keywords, as they can lead to inaccurate or misleading interpretations to the model
- Do not use these historical (image) reconstructions to distort or disrespect historical events, cultural groups, or sacred beliefs of any community.
- Do not rely solely on AI outputs without cross-checking with credible historical reference or expert input, especially if you use these for professional or educational purposes.

Sample Template for a User-input prompt:

“...Generate a prompt to create a 2D Image to depict a historical scenario based on the below criteria:

- (i) **Topic** :
- (ii) **Landscape/Environment** :
- (iii) **Architecture** :
- (iv) **People** :
- (v) **Activities** :
- (vi) **Atmosphere/Ambience** :
- (vii) **Special Objects** :
- (viii) **Artistic Style** :

-----End of the document-----

5.3 Potential Practical Application of this Research Simulation

This research simulation has a significant potential for practical application, particularly as a visitor engagement tool within museums and heritage sites. For instance, it could be implemented as an interactive activity, inviting visitors to create AI-generated historical images based on their own perspectives and interpretations after experiencing museum galleries or heritage sites.

Such an initiative would enable museum and heritage site staff to gain valuable insights into visitors' perceptions and interests, which could in return, be used to take informed decisions on more effective and audience-responsive interpretation strategies.

Furthermore, the collected data, including the user-generated prompts, system-generated prompts, and the resulting AI images, could serve as a valuable resource for further research. This dataset could support studies on the behaviour and limitations of gen-AI models in the cultural heritage context, contributing to ongoing efforts to refine AI tools for historically accurate and culturally/contextually sensitive applications.

5.4 Challenges and Limitations encountered during the Research

As a student pursuing an MA in Heritage Studies, my academic and professional background has primarily focused on cultural heritage. However, I have always been drawn to the intersection between heritage and digital technologies, which led me to continuously explore how digital techniques can enhance heritage management and interpretation. Accordingly, my undergraduate thesis focused exploring effective use of digital techniques for enhancing heritage interpretation and resulted with developing a Conceptual Framework for On-site Digital Interpretation Developments in Cultural Heritage Sites. Consequently, my current MA thesis focuses on how AI-generated imagery can facilitate multivocality in heritage interpretation.

Although, my interest lies firmly in this area, my limited technical background in AI presented me certain challenges, particularly limiting me in conducting a technical analysis relevant to this historical (image) generation process, which could have further enriched the depth of this research study and its results interpretation. Nevertheless, I am deeply grateful to both of my supervisors for patiently supporting me and guiding me, which helped me navigate through this challenge and limitation, and to complete this research.

5.5 Directions and Recommendations for Future Studies

- As explained in Chapter 5.1.2, one key factor behind the lower scores for historical accuracy of the generated images of this study is, the errors and misinterpretations occurred during the intermediate linguistic transformation process (between user-prompts and system-generated prompts). Therefore, to address this, future studies are encouraged to investigate on the technical mechanisms underlying this linguistic transformation phase, to better prevent such errors and ultimately improve image accuracy.
- This research focused on zero-shot text-to-image generation. Therefore, future research should explore one-shot and few-shot prompting, which involves supplementing prompts with additional sources (images, weblinks, academic articles, etc.). Systematically studying how the generative AI models respond to such enriched prompting can provide deeper insights into improving the AI-generated historical (image) reconstructions.
- This study examined only three models and three heritage test cases. Therefore, expanding the scope by conducting more systematic studies will further contribute to enhancing the performance of AI models in generating more historically accurate and culturally/contextually sensitive image outputs.
- Finally, the future researchers are encouraged to investigate methods for improving cultural understanding of AI models. For instance, the study by Liu et al. (2025) presents an initiative called '*CultureVLM: Characterizing and Improving Cultural Understanding of Vision-Language Models for over 100 countries*', which demonstrates advancements in this area.

Accordingly, by revisiting the predefined objectives of this research, it is clear that based on the data analysis presented in Chapter 4, that the first and second objectives have been addressed through the discussions in Chapter 5.1.4. The third objective has been extensively discussed across Chapters 5.1.1, 5.1.2, and 5.1.3. Finally, the fourth objective has been fulfilled through the development of a conceptual framework cum guideline, as outlined in Chapter 5.2. Hence, the study has been able to achieve all of its predefined research objectives.

Chapter 6:

CONCLUSION

The legacy or inheritance of the past that has been passed down through generations is often understood as heritage. However, the emergence of critical heritage studies in the late 20th century, led to view, comprehend, and interpret heritage in a broader and more critical manner. As extensively discussed in Chapter 2.1, heritage is not merely a static record of historical facts, but rather “.... *the result of the choosing; it is history-as-memory-work...*” (Deumert, 2018, p. 5), shaped by choices, power dynamics, and cultural contexts. This means our human past is inherently complex, multifaceted, and open to contestation, and it highlights the importance of addressing cultural appropriation and representing diverse perspectives.

One significant approach to acknowledging this complexity is the concept of ‘Multivocality (or Polyvocality)’, which emphasises the inclusion of multiple voices and interpretations. In the field of cultural heritage, this concept is used to redress the historical silences/overlooking, in the past and, to ensure that the voices of present-day underrepresented perspectives are provided a platform in heritage interpretation and presentation. In practice, heritage institutions, such as museums, archives, and heritage sites often employ different strategies to foster multivocality, such as oral history projects, storytelling initiatives and participatory exhibitions, and so on.

Motivated by a deep interest in the intersection of heritage and digital technologies, this research study explored the potential of generative AI, particularly AI-generated historical (image) reconstructions, to meaningfully support the representation of multivocality in heritage. While focusing on how this technology can be used to produce more inclusive and diverse representations of narratives, the associated technical challenges, primarily the historical accuracy and cultural/contextual sensitivity in representations were meant to address as well.

As detailed in Chapter 3, the study selected three prominent image generation models and tested them against three challenging heritage scenarios (test cases), having multiple perspectives. Accordingly, 13 user prompts were provided to a selected AI model and obtained system-generated prompts required for image creation. In total, the prompts of 13 distinct perspectives were transformed into 39 images, across the three models.

Proceeded by a linguistic analysis to observe the transformations of user prompts into system-generated prompts, these 39 images were systematically evaluated using eight criteria defined in Chapter 3.2.4 and 3.2.5 (i.e., People, Landscape/Environment, Activities, etc.) with the support of heritage experts. Based on their evaluation, these visual outputs were cross analysed to examine:

- How well each model represented a single perspective
- How a single model handled multiple perspectives within one heritage test case, and
- How each model performed across different heritage test cases.

Based upon this extensive data analysis, the study highlighted both strengths and limitations of current generative AI models in producing historically accurate and culturally/contextually sensitive historical reconstructions. General observations pointed to the recurring inaccuracies, while Specific observations highlighted case-specific accomplishments and shortcomings.

Fulfilling one of the predefined research objectives, the study then discussed how gen-AI models have contributed in neutralising or emphasising multivocality, based upon the research findings and, the potential of gen-AI image generation in fostering meaningful dialogue on multivocality in heritage.

Another research objective of the study was identifying the key factors/aspects of providing effective user prompts to generate historically accurate and culturally/contextually sensitive historical image outputs. In response to that, the study proposed a ‘Conceptual Framework cum Guideline for User Prompt Optimisation for AI-generated Historical (Image) Reconstructions’. While guiding users to craft clear and more effective prompts, this guideline reminds the users of the importance of,

- Not using these historical (image) reconstructions to distort or disrespect historical events, cultural groups, or sacred beliefs of any community, and
- Not relying solely on AI outputs without cross-checking with credible historical reference or expert input, especially they intend to use them for professional or educational purposes.

Beyond the theoretical scope, the research also explored how this simulation could be practically applied, for instance, as an interactive visitor engagement activity and in museums and heritage sites, to gather insights into visitors’ interpretations, to make informed decisions on interpretive strategies, and for further research purposes. Finally, the study offers clear directions and recommendations for future research.

In conclusion, this research underscores that, despite several limitations discussed in the study, generative AI has potential in fostering multivocality in heritage interpretation. However, it must be employed responsibly, ethically, and critically. As noted in Chapter 2.1, embracing multiple perspectives in an uncritical manner can risk naïve relativism, the belief that all viewpoints are equally valid, even when unsupported with evidence or contradictory. Hence, scholars argue that, while multivocality enriches understanding, it must be accompanied by critical evaluation and contextual knowledge to ensure truthful and respectful heritage representation (C. Smith, 2020, p. 7482).

Ultimately, this study contributes to the growing discourse on digital heritage by demonstrating that, when used thoughtfully and with appropriate safety measures, generative AI can serve as an innovative tool for facilitating diverse voices and enriching the interpretation of our human history.

-----END-----

Bibliography

- Abate, D., Agapiou, A., Toumbas, K., Lampropoulos, A., Petrides, K., Pierdicca, R., Paolanti, M., Di Stefano, F., Felicetti, A., Malinverni, E. S., & Zingaretti, P. (2023). Artificial Intelligence to fight Illicit trafficking of Cultural property. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023*, 3–10. <https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-3-2023>
- Altaweel, M., Khelifi, A., & Zafar, M. H. (2024). Using Generative AI for Reconstructing Cultural Artifacts: Examples Using Roman Coins. *Journal of Computer Applications in Archaeology*, 7(1), 301–315. <https://doi.org/10.5334/jcaa.146>
- Arzomand, K., Rustell, M., & Kalganova, T. (2024). From ruins to reconstruction: Harnessing text-to-image AI for restoring historical architectures. *Challenge Journal of Structural Mechanics*, 10(2), 69. <https://doi.org/10.20528/cjsmec.2024.02.004>
- Bansal, G., Nawal, A., Chamola, V., & Herencsar, N. (2024). Revolutionizing Visuals: The Role of Generative AI in Modern Image Generation. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(11), 1–22. <https://doi.org/10.1145/3689641>
- Boonstra, L. (2025). *Prompt Engineering*. Google. <https://www.kaggle.com/whitepaper-prompt-engineering>
- Bu, F., Wang, Z., Wang, S., & Liu, Z. (2025). *An Investigation into Value Misalignment in LLM-Generated Texts for Cultural Heritage* (No. arXiv:2501.02039). arXiv. <https://doi.org/10.48550/arXiv.2501.02039>
- Capurro, C., Provatorova, V., & Kanoulas, E. (2023). Experimenting with Training a Neural Network in Transkribus to Recognise Text in a Multilingual and Multi-Authored Manuscript Collection. *Heritage*, 6(12), 7482–7494. <https://doi.org/10.3390/heritage6120392>
- Cui, C. de lian, Curra, E., & Fioravanti, A. (2024). AI-powered built heritage: Enhancing interpretation and recovery processes with generative AI models. *Catalogo Dei Prodotti Della Ricerca*, 117–126.
- Cultural AI Lab. (2021). *Projects: Pressing Mattes*. Cultural AI Lab. <https://www.cultural-ai.nl/projects/pressing-matter>
- Dagenais, F. (2025). *Artificial intelligence in museums: Three Transkribus case studies*. Transkribus. https://www.transkribus.org/blog/ai-in-museums?utm_source=chatgpt.com

- Deumert, A. (2018). The multivocality of heritage – moments, encounters and mobilities. In A. Creese & A. Blackledge (Eds.), *The Routledge Handbook of Language and Superdiversity* (1st ed., pp. 149–164). Routledge.
<https://doi.org/10.4324/9781315696010-12>
- Electronic Visualization Laboratory. (2025). *Virtual Harlem*. Electronic Visualization Laboratory (EVL). <https://www.evl.uic.edu/art/1775>
- Encyclopaedia Britannica. (2025). *History of Suriname*. Encyclopaedia Britannica.
<https://www.britannica.com/place/Suriname/History>
- Euronews. (2019). *Walking through time: How AI is rebuilding centuries-old Europe*. Euronews. <https://www.youtube.com/watch?v=KDVa6xC184I>
- Foka, A., & Griffin, G. (2024). AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage*, 7(11), 6125–6136.
<https://doi.org/10.3390/heritage7110287>
- GeeksforGeeks. (2025). *Phases of Natural Language Processing (NLP)*. GeeksforGeeks.
<https://www.geeksforgeeks.org/phases-of-natural-language-processing-nlp/>
- Gemeinhardt, J., & Zöllner, M. (2024). WebXR Cultural Heritage Tour with Generative AI Characters. *Eurographics Workshop on Graphics and Cultural Heritage*.
<https://doi.org/10.2312/GCH.20241272>
- Grammarly. (2024). *What Is NLP? How Machines Understand and Generate Human Language*. Grammarly Blog: Understanding AI.
<https://www.grammarly.com/blog/ai/what-is-natural-language-processing/?msckid=25e0c9fc785e6f3c0ea0ddcd79ce6eb1>
- Gupta, V., Sambyal, N., Sharma, A., & Kumar, P. (2021). Restoration of artwork using deep neural networks. *Evolving Systems*, 12(2), 439–446. <https://doi.org/10.1007/s12530-019-09303-7>
- Hao, Y., Chi, Z., Dong, L., & Wei, F. (2023). Optimizing Prompts for Text-to-Image Generation. *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*. <https://arxiv.org/abs/2212.09611>
- He, Z., Su, J., Chen, L., Wang, T., & LC, R. (2025). *‘I Recall the Past’: Exploring How People Collaborate with Generative AI to Create Cultural Heritage Narratives*. 28th ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW). <https://doi.org/10.48550/arXiv.2501.00359>
- Humanity House. (2025). *What We Did*. The Humanity House, The Hague.
<https://humanityhouse.org/wat-we-deden/>

- IBM. (2025). *Prompt Engineering*. IBM. <https://www.ibm.com/think/topics/prompt-engineering>
- Imagen-Team-Google, Baldrige, J., Bauer, J., Bhutani, M., Brichtova, N., Bunner, A., Castrejon, L., Chan, K., Chen, Y., Dieleman, S., Du, Y., Eaton-Rosen, Z., Fei, H., Freitas, N. de, Gao, Y., Gladchenko, E., Colmenarejo, S. G., Guo, M., Haig, A., ... Zwols, Y. (2024). *Imagen 3* (No. arXiv:2408.07009). arXiv. <https://doi.org/10.48550/arXiv.2408.07009>
- Jaillant, L. (Ed.). (2022). *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections* (1st ed., Vol. 2). Bielefeld University Press / transcript Verlag. <https://doi.org/10.14361/9783839455845>
- Kirmizialtin, S., & Wrisley, D. J. (2020). *Automated Transcription of Non-Latin Script Periodicals: A Case Study in the Ottoman Turkish Print Archive*. <https://arxiv.org/pdf/2011.01139>
- Kumar, P., & Gupta, V. (2023). Artwork restoration using paired image translation-based generative adversarial networks. *ITM Web of Conferences*, 54, 01013. <https://doi.org/10.1051/itmconf/20235401013>
- Liu, S., Jin, Y., Li, C., Wong, D. F., Wen, Q., Sun, L., Chen, H., Xie, X., & Wang, J. (2025). *CultureVLM: Characterizing and Improving Cultural Understanding of Vision-Language Models for over 100 Countries* (No. arXiv:2501.01282). arXiv. <https://doi.org/10.48550/arXiv.2501.01282>
- Manjula, H. A. K. L. (2024). *Sri Pada (Adam's Peak)*. Lanka Pradeepa. <https://www.lankapradeepa.com/2021/03/sri-pada-adams-peak.html>
- Migration Museum Project. (2025). *Exhibitions*. Migration Museum, London. <https://www.migrationmuseum.org/>
- Münster, S., Maiwald, F., Di Lenardo, I., Henriksson, J., Isaac, A., Graf, M. M., Beck, C., & Oomen, J. (2024). Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe. *Heritage*, 7(2), 794–816. <https://doi.org/10.3390/heritage7020038>
- Nandi, S., & Cheriyaedath, S. (2024). *AI Fights Illegal Cultural Goods Trafficking*. AZOAI. <https://www.azoai.com/news/20240711/AI-Fights-Illegal-Cultural-Goods-Trafficking.aspx>
- National Museum of African American History and Culture. (2025). *Exhibitions: Slavery & Freedom 1400–1877*. The Smithsonian National Museum of African American History and Culture's Searchable Museum. <https://www.searchablemuseum.com/slavery-and-freedom/>

- Oppenlaender, J. (2024). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, 43(15), 3763–3776. <https://doi.org/10.1080/0144929X.2023.2286532>
- Oxford University Press. (2025). *Multivocality*. Oxford Reference. <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100216225>
- Pressing Matters. (2025). *Pressing Matters: About*. Pressing Matters: Ownership, Value and the Question of Colonial Heritage in Museums. <https://pressingmatter.nl/about/>
- Prompt Basics*. (2025). Midjourney. <https://docs.midjourney.com/hc/en-us/articles/32023408776205-Prompt-Basics>
- Rane, N. (2023). Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Arts and Humanities. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4603208>
- Rijksmuseum. (2025). *Past Exhibitions: Slavery*. Rijksmuseum. <https://www.rijksmuseum.nl/en/whats-on/exhibitions/past/slavery>
- Saunders, M., Thornhill, A., & Lewis, P. (2019). *Research Methods for Business Students* (Eighth Edition). Pearson Education Limited. <https://dr-notes.com/research-methods-for-business-students-eighth-edition-v3f>
- Smith, C. (Ed.). (2020). *Encyclopedia of Global Archaeology*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30018-0>
- Smith, L. (2006). The Discourse of Heritage. In *Uses of Heritage* (first edition, pp. 11–43). Routledge. <https://doi.org/10.4324/9780203602263>
- Spennemann, D. H. R. (2024). Generative Artificial Intelligence, Human Agency and the Future of Cultural Heritage. *Heritage*, 7(7), 3597–3609. <https://doi.org/10.3390/heritage7070170>
- Stephanidis, C., Antona, M., Ntoa, S., & Salvendy, G. (Eds.). (2024). *HCI International 2024 Posters: 26th International Conference on Human-Computer Interaction, HCII 2024, Washington, DC, USA, June 29 – July 4, 2024, Proceedings, Part III* (Vol. 2116). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-61950-2>
- Stryker, C., & Scapicchio, M. (2024). *What is Generative AI?* IBM. <https://www.ibm.com/think/topics/generative-ai>
- Sukkar, A. W., Fareed, M. W., Yahia, M. W., Abdalla, S. B., Ibrahim, I., & Senjab, K. A. K. (2024). Analytical Evaluation of Midjourney Architectural Virtual Lab: Defining Major Current Limits in AI-Generated Representations of Islamic Architectural Heritage. *Buildings*, 14(3), 786. <https://doi.org/10.3390/buildings14030786>

- The Koorie Heritage Trust. (2025). *Koorie Heritage Trust*. Koorie Heritage Trust.
<https://kht.org.au/>
- Wu, R., Gong, C., Chen, L., & Su, J. (2024). *The Present in the future is the past*:
https://isea-archives.siggraph.org/wp-content/uploads/2025/01/2024_Wu_THE_PRESENT_IN_THE_FUTURE_IS_THE.pdf?utm_source=chatgpt.com
- Zhang, M. (2010). *Museum of London Releases Augmented Reality App for Historical Photos*. PetaPixel. <https://petapixel.com/2010/05/24/museum-of-london-releases-augmented-reality-app-for-historical-photos/#:~:text=Streetmuseum%20is%20a%20new%20%28and%20free%29%20augmented%20reality,around%20London%20using%20either%20the%20map%20or%20GPS.>
- Zheng, N. (2023). *Coming to grips with diverse voices in participatory heritage initiatives* (p. HFNZG90020230404) [PhD, Vrije Universiteit Amsterdam].
<https://doi.org/10.5463/thesis.146>

Appendices