



Time Travel: A Comprehensive Benchmark to Evaluate LMMs on Historical and Cultural Artifacts

Sara Ghaboura^{1†} Ketan More^{1†} Ritesh Thawkar¹ Wafa Alghallabi¹ Omkar Thawakar¹
Fahad Shahbaz Khan^{1,2} Hisham Cholakkal¹ Salman Khan^{1,3} Rao Muhammad Anwer^{1,4}

¹Mohamed bin Zayed University of AI, ²Linköping University, ³Australian National University, ⁴Aalto University

{sara.ghaboura, ketan.more, omkar.thawakar}@mbzuai.ac.ae

<https://mbzuai-oryx.github.io/TimeTravel/>

Abstract

Understanding historical and cultural artifacts demands human expertise and advanced computational techniques, yet the process remains complex and time-intensive. While large multimodal models offer promising support, their evaluation and improvement require a standardized benchmark. To address this, we introduce *TimeTravel*, a benchmark of 10,250 expert-verified samples spanning 266 distinct cultures across 10 major historical regions. Designed for AI-driven analysis of manuscripts, artworks, inscriptions, and archaeological discoveries, *TimeTravel* provides a structured dataset and robust evaluation framework to assess AI models' capabilities in classification, interpretation, and historical comprehension. By integrating AI with historical research, *TimeTravel* fosters AI-powered tools for historians, archaeologists, researchers, and cultural tourists to extract valuable insights while ensuring technology contributes meaningfully to historical discovery and cultural heritage preservation. We evaluate contemporary AI models on *TimeTravel*, highlighting their strengths and identifying areas for improvement. Our goal is to establish AI as a reliable partner in preserving cultural heritage, ensuring that technological advancements contribute meaningfully to historical discovery. Our code is available at: <https://github.com/mbzuai-oryx/TimeTravel>.

1 Introduction

In recent years, Large Multimodal Models (LMMs) have made significant strides in visual reasoning, perception, and multimodal understanding. Models such as GPT-4V (OpenAI, 2024) and LLaVA (Liu et al., 2023) have excelled in image caption-

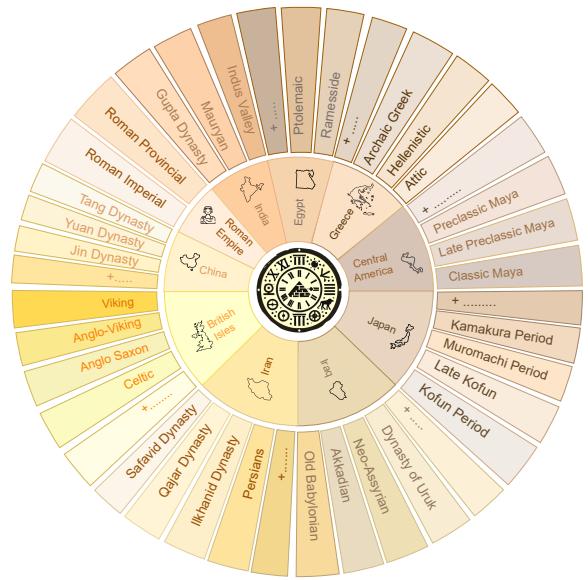


Figure 1: TimeTravel Taxonomy categorizes artifacts from 10 major civilizations, spanning diverse historical and prehistoric periods. It encompasses 266 distinct cultures and over 10k manually verified historical artifact samples, providing a structured framework for comprehensive AI-driven analysis.

ing, visual question answering (VQA), and complex visual reasoning, driving the development of benchmarks (Chiu et al., 2024; Nayak et al., 2024; Alwajih et al., 2024) to assess their capabilities. These benchmarks predominantly focus on modern objects, cultural landmarks, and textual sources, extending multimodal AI applications to domains such as medical imaging, remote sensing, and real-world scene understanding (Ghaboura et al., 2025). However, a critical gap remains—LMMs fail to address the historical dimension of visual data, particularly artifacts that shaped human civilization.

Historical artifacts, from ancient manuscripts and inscriptions to architectural ruins and cultural symbols, offer invaluable insights into the evolution

[†]Equal contribution.



Figure 2: **TimeTravel Samples.** Showcasing diverse cultural representations from various regions across the globe, these examples span multiple artifact categories, including coins, accessories, tools, and statues from ancient civilizations. Each artifact is accompanied by a detailed description, providing valuable contextual and historical insights. Additional TimeTravel examples can be found in Fig.7 and Fig.8.

of societies, artistic expression, and technological advancements. These artifacts preserve cultural heritage and serve as primary sources for understanding belief systems, trade networks, and socio-political structures of past civilizations. However, interpreting them requires deep contextual knowledge, which current LMMs struggle to achieve, particularly in non-English and non-Western historical contexts. While some models have been extended to low-resource languages to bridge cultural gaps (Heakl et al., 2025), they lack systematic capabilities to analyze artifacts from diverse civilizations. This limitation highlights the urgent need for a specialized benchmark that evaluates AI’s ability to process and understand historical artifacts with cultural and temporal awareness.

To address this challenge, we introduce TimeTravel, an open-source comprehensive benchmark (see Table 1) for evaluating LMM performance in historical artifact analysis across diverse civilizations. TimeTravel encompasses several major ancient and prehistoric civilizations across 10 distinct regions, spanning 266 cultural groups. It offers a structured taxonomy tailored for AI-driven historical research (see Fig. 1). Unlike existing benchmarks that focus on generic object recognition, TimeTravel prioritizes historical knowledge, contextual reasoning, and cultural preservation, making it a pioneering effort in multimodal AI evaluation. The benchmark consists of over 10k curated samples, each accompanied by high-quality images of manuscripts, inscriptions, sculptures, paintings, and archaeological discoveries. These samples assess key aspects of multimodal understanding, including visual perception, contextual reasoning, and cross-civilizational knowledge. Meticulously verified by historians and archaeologists, the dataset ensures accuracy, cultural relevance, and historical integrity. By evaluating both closed- and open-source LMMs

on TimeTravel, we aim to identify their strengths and limitations in handling historically significant artifacts, paving the way for AI models that contribute meaningfully to cultural heritage preservation and historical analysis.

Domain	British Museum	MMMU	Oracle-MNIST	Ithaca	KaoKore	HUST-OBS	TimeTravel (ours)
Hist. Artifact Recog.	✓	✗	✗	✗	✓	✗	✓
Geographic Region	✓	✗	✗	✓	✓	✗	✓
Ancient Artifacts	✓	✗	✗	✗	✗	✗	✓
Contextual History	✗	✗	✗	✗	✗	✗	✓
Image-Text Pairs	✓	✓	✗	✗	✓	✓	✓
Open-Source	✗	✓	✓	✗	✓	✓	✓

Table 1: The comparison of datasets and benchmarks for historical and cultural artifacts, evaluating features like **artifact recognition**, **geographic coverage**, **multimodal understanding**, and **metadata inclusion** with existing data such as British Museum (Tully, 2020), MMMU (Yue et al., 2024), Oracle-MNIST (Wang and Deng, 2022), Ithaca (Assael et al., 2022), KaoKore (Tian et al., 2020), HUST-OBS (Wang et al., 2024). TimeTravel stands out as the most comprehensive benchmark, uniquely integrating multimodal data, historical context, and a dedicated focus on ancient artifacts to support AI-driven cultural heritage research.

2 The TimeTravel Dataset

2.1 Data Collection

Our research is based on a well-structured and meticulously curated dataset sourced from museum collections, which houses an extensive collection of artifacts from diverse civilizations. From this vast repository, we compiled a dataset spanning 266 cultural groups, allowing the analysis of cultural, technological, and social developments over a broad historical timeline.

To ensure the integrity of our benchmark, we followed a systematic data collection process. We first identified key civilizations and historical periods relevant to our study, then collaborated closely with experts to validate the authenticity and completeness of each record. As a result, our dataset com-

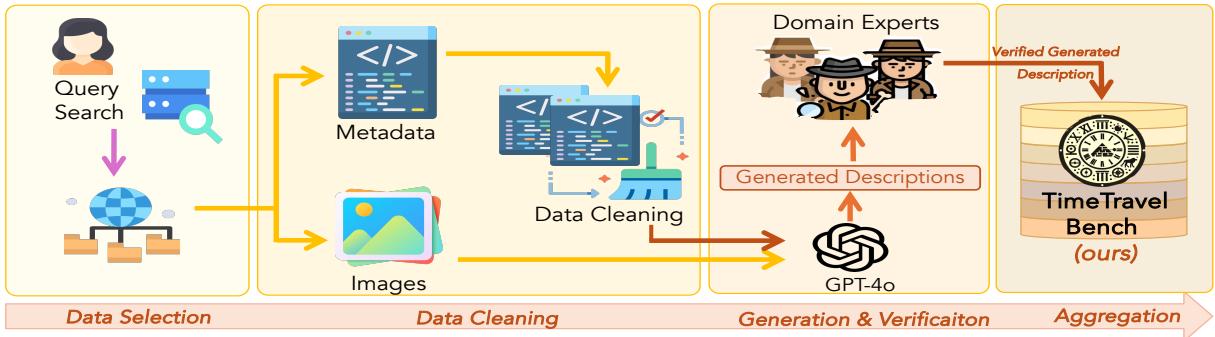


Figure 3: **TimeTravel Data Pipeline.** A structured workflow that collects image and text data from museum websites, cleans metadata, and integrates it with visual content. The GPT-4o model generates detailed, context-aware descriptions, which are refined by experts for accuracy before forming the TimeTravel Benchmark.

prises 10,250 carefully curated samples (see Fig 2). Each entry—ranging from artifacts and inscriptions to ancient manuscripts—was meticulously verified by historians and archaeologists, ensuring accuracy and reliability. By incorporating data from multiple civilizations, our benchmark provides a diverse and comprehensive perspective, avoiding the limitations of a single historical narrative while preserving the historical context for in-depth analysis. This meticulous approach allows us to reveal significant patterns in human history, offering valuable insights into the evolution of human history and civilizations over time.

2.2 Image-Text pair Generation

The dataset features a diverse range of historical objects, ensuring comprehensive documentation and contextual understanding. However, many metadata fields—such as title, iconography, and date—were missing or incomplete. To address this, we employed GPT-4o to generate detailed, context-aware textual descriptions based on the available metadata (see Fig. 5 and 6). To further enhance usability, we structured these descriptions into image-text pairs, ensuring that each artifact is not only visually documented but also enriched with contextual and cultural insights. By improving multimodal model compatibility and supporting digital archiving, this approach strengthens research in cultural heritage preservation while bridging gaps in existing records.

2.3 Data Filtering and Verification

To guarantee the accuracy and reliability of our dataset, we implemented a rigorous data filtering and verification process (Fig. 3). This process combined manual expert validation with automated techniques to eliminate inconsistencies, fill in miss-

ing details where possible, and authenticate historical records. During data cleaning, we addressed missing or incomplete metadata—such as titles, dates, and iconography—by cross-referencing museum archives, academic sources, and expert insights. Unavailable key information was transparently documented. Additionally, automated checks identified formatting inconsistencies, metadata mapping errors, and numerical anomalies, ensuring a structured and standardized dataset. For verification, we collaborated with historians, archaeologists, and museum curators to review each artifact’s description, cultural attribution, and historical significance. Expert validation ensured that generated textual descriptions were accurate, contextually relevant, and aligned with historical records. This rigorous process enhances the dataset’s credibility, making it a valuable resource for historical research, machine learning, and cultural heritage preservation while ensuring reliable insights into human history. Additional details are presented in Appendix (Sec. D).

3 TimeTravel Benchmark Evaluation

Evaluation Metric: To assess the quality, accuracy, and relevance of our generated textual descriptions, we employed a combination of traditional and advanced metrics. BLEU (Papineni et al., 2002) and ROUGE-L (Lin, 2004) evaluate linguistic fluency and structural similarity, ensuring syntactic alignment with reference texts. METEOR (Banerjee and Lavie, 2005) enhances this by incorporating synonym matching and paraphrasing, improving adaptability to human variations. SPICE (Anderson et al., 2016) assesses semantic accuracy through scene graph analysis, preserving object relationships and cultural context. Additionally, BERTScore (Zhang et al., 2019) offers a deep

	Model	BLEU	METEOR	ROUGE-L	SPICE	BERTScore	LLM-Judge
Closed	GPT-4o-0806 (OpenAI, 2024)	0.1758	0.2439	0.1230	0.1035	0.8349	0.3013
	Gemini-2.0-Flash (Reid et al., 2024)	0.1072	0.2456	0.0884	0.0919	0.8127	0.2630
	Gemini-1.5-Pro (Reid et al., 2024)	0.1067	0.2406	0.0848	0.0901	0.8172	0.2276
	GPT-4o-mini-0718 (OpenAI, 2024)	0.1369	0.2658	0.1027	0.1001	0.8283	0.2492
Open	Llama-3.2-Vision-Inst (Meta AI, 2024)	0.1161	0.2072	0.1027	0.0648	0.8111	0.1255
	Qwen-2.5-VL (Team, 2025)	0.1155	0.2648	0.0887	0.1002	0.8198	0.1792
	Llava-Next (Liu et al., 2024)	0.1118	0.2340	0.0961	0.0799	0.8246	0.1161

Table 2: Performance comparison of various closed and open-source models on our proposed TimeTravel benchmark.

	Model	India	Roman Empire	China	British Isles	Iran	Iraq	Japan	Central America	Greece	Egypt
Closed	GPT-4o-0806	0.2491	0.4463	0.2491	0.1899	0.3522	0.3545	0.2228	0.3144	0.2757	0.3649
	Gemini-2.0-Flash	0.1859	0.3358	0.2059	0.1556	0.3376	0.3071	0.2000	0.2677	0.2582	0.3602
	Gemini-1.5-Pro	0.1118	0.2632	0.2139	0.1545	0.332	0.2587	0.1871	0.2708	0.2088	0.2908
	GPT-4o-mini-0718	0.2311	0.3612	0.2207	0.1866	0.2991	0.2632	0.2087	0.3195	0.2101	0.2501
Open	Llama-3.2-Vision-Inst	0.0744	0.1450	0.1227	0.0777	0.2000	0.1155	0.1075	0.1553	0.1351	0.1201
	Qwen-2.5-VL	0.0888	0.1578	0.1192	0.1713	0.2515	0.1576	0.1771	0.1442	0.1442	0.2660
	Llava-Next	0.0788	0.0961	0.1455	0.1091	0.1464	0.1194	0.1353	0.1917	0.1111	0.0709

Table 3: Analysis of LLM-Judge evaluation of various models in describing archaeological artifacts across civilizations from different geographical locations. Additional comparisons are presented in Appendix (Table 4).

learning-based evaluation of semantic similarity, capturing contextual meaning beyond simple word overlap. LLM-Judge further enhances assessment by evaluating coherence, factual accuracy, and contextual appropriateness.

Results and Analysis: Our evaluation of closed-source and open-source models on the TimeTravel dataset reveals clear differences in their ability to generate historically accurate descriptions (see Table 2). Among closed-source models, GPT-4o-0806 achieved the highest BLEU (0.1758), ROUGE-L (0.1230), SPICE (0.1035), BERTScore (0.8349), and LLM-Judge score (0.3013), indicating superior semantic alignment and contextual richness. However, its lower METEOR score (0.2439) suggests that while it generates highly structured descriptions, they may lack word-level diversity and fluency. GPT-4o-mini-0718, despite scoring slightly lower in BLEU (0.1369) and ROUGE-L (0.1027), outperformed all models in METEOR (0.2658), highlighting its strength in producing more lexically diverse and well-formed outputs. Gemini-2.0-Flash and Gemini-1.5-Pro, while achieving moderate performance across all metrics, demonstrated weaker lexical alignment (BLEU: 0.1072, 0.1067) and semantic coherence (BERTScore: 0.8127, 0.8172), suggesting that they may struggle with historical specificity and structured descriptions. Among open-source models, Qwen-2.5-VL performed the best, achieving higher BLEU (0.1155), METEOR (0.2648), and SPICE (0.1002) compared to its counterparts. These scores indicate a better balance between fluency

and contextual accuracy, making it a strong contender despite being an open-source model. Llama-3.2-Vision-Inst and Llava-Next, however, showed lower SPICE (0.0648, 0.0799) and LLM-Judge scores (0.1255, 0.1161), suggesting difficulties in capturing object details and historical context.

Table 3 presents the LLM-Judge evaluation of models in describing archaeological artifacts across civilizations from different geographic regions. GPT-4o-0806 outperformed other models in describing archaeological artifacts, excelling in regions like the Roman Empire, Iran, Iraq, and Egypt, indicating strong contextual understanding. GPT-4o-mini-0718 and Gemini-2.0-Flash showed strengths in India, Central America, and China, but with some limitations. Among open-source models, Qwen-2.5-VL performed best in Iran, the British Isles, and Egypt, though overall, closed-source models provided more accurate historical descriptions. Additional analysis based on the METEOR score is presented in Appendix (Table 4).

Overall, closed-source models outperform open-source models in generating context-aware descriptions, but ongoing improvements in open-source models highlight opportunities for fine-tuning and dataset expansion. These findings will guide further model enhancements, advancing AI-driven historical analysis and cultural heritage preservation.

4 Conclusion

We present the TimeTravel dataset, a curated collection of historical artifacts from 10 cultural regions, extensively curated by domain experts. We

developed a rigorous data collection, filtering, and verification process, ensuring accuracy and completeness. Using GPT-4o, we generated detailed textual descriptions, making the dataset more accessible and valuable for AI-driven historical research. Our evaluation, using BLEU, METEOR, ROUGE-L, CIDEr, SPICE, BERTScore, and LLM-Judge, showed that closed-source models outperformed open-source alternatives, though open models are rapidly improving. Our analysis highlights the potential of LMMs in bridging gaps in historical records while maintaining academic integrity. By leveraging AI-driven methodologies, this work sets the foundation for advancing cultural heritage preservation and enhancing digital humanities research, ensuring greater accessibility and accuracy in historical documentation.

5 Limitations and Societal Impact

While this research demonstrates the potential of LMMs in enhancing historical documentation, the quality of generated descriptions depends on the completeness and accuracy of the input data. In cases where historical records are fragmented or ambiguous, AI-generated text may lack full contextual depth. Additionally, biases present in training data can influence how models interpret and describe cultural artifacts, necessitating continuous evaluation and expert validation to ensure historical accuracy and cultural sensitivity. Despite these challenges, this research contributes to cultural heritage preservation, educational accessibility, and AI-driven humanities research. By digitizing and enriching historical records, it enables wider public engagement with history, supports museum digitization efforts, and provides a foundation for future advancements in AI-assisted historical analysis, bridging the gap between technology and human expertise in understanding our collective past.

References

- Muhammad Farid Adilazuarda, Sagnik Mukherjee, Pradyumna Lavania, Siddhant Singh, Alham Fikri Aji, Jacki O'Neill, Ashutosh Modi, and Monojit Choudhury. 2024. Towards measuring and modeling "culture" in llms: A survey. *arXiv preprint arXiv:2403.15412*.
- Badr AlKhamissi, Muhammad N. ElNokrashy, Mai AlKhamissi, and Mona Diab. 2024. [Investigating cultural alignment of large language models](#). *ArXiv*, abs/2402.13231.
- Fakhraddin Alwajih, El Moatez Billah Nagoudi, Gagan Bhatia, Abdelrahman Mohamed, and Muhammad Abdul-Mageed. 2024. Peacock: A family of arabic multimodal large language models and benchmarks. *arXiv preprint arXiv:2403.01031*.
- Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. 2016. Spice: Semantic propositional image caption evaluation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14*, pages 382–398. Springer.
- Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, J. Prag, and Nando de Freitas. 2022. [Restoring and attributing ancient texts using deep neural networks](#). *Nature*, 603:280 – 283.
- Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.
- Fan Bu, Zheng Wang, Siyi Wang, and Ziyao Liu. 2025. An investigation into value misalignment in llm-generated texts for cultural heritage. *arXiv preprint arXiv:2501.02039*.
- Yu Ying Chiu, Liwei Jiang, Bill Yuchen Lin, Chan Young Park, Shuyue Stella Li, Sahithya Ravi, Mehar Bhatia, Maria Antoniak, Yulia Tsvetkov, Vered Shwartz, et al. 2024. Culturalbench: a robust, diverse and challenging benchmark on measuring the (lack of) cultural knowledge of llms. *arXiv preprint arXiv:2410.02677*.
- Sara Ghaboura, Ahmed Heakl, Omkar Thawakar, Ali Husain Salem Abdulla Alharthi, Ines Riahi, Abduljalil Saif, Jorma Laaksonen, Fahad Shahbaz Khan, Salman H Khan, and Rao Muhammad Anwer. 2025. Camel-bench: A comprehensive arabic lmm benchmark. *NAACL*.
- Ahmed Heakl, Sara Ghaboura, Omkar Thawakar, Fahad Shahbaz Khan, Hisham Cholakkal, Rao Muhammad Anwer, and Salman Khan. 2025. [Ain: The arabic inclusive large multimodal model](#).
- Cheng Li, Mengzhou Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024. Culturellm: Incorporating cultural differences into large language models. *arXiv preprint arXiv:2402.10946*.
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. 2024. [Llava-next: Improved reasoning, ocr, and world knowledge](#).
- Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. In *NeurIPS*.

- Shudong Liu, Yiqiao Jin, Cheng Li, Derek F Wong, Qingsong Wen, Lichao Sun, Haipeng Chen, Xing Xie, and Jindong Wang. 2025. Culturevlm: Characterizing and improving cultural understanding of vision-language models for over 100 countries. *arXiv preprint arXiv:2501.01282*.
- Meta AI. 2024. **Llama 3.2: Revolutionizing edge ai and vision with open, customizable models.**
- Shravan Nayak, Kanishk Jain, Rabiul Awal, Siva Reddy, Sjoerd van Steenkiste, Lisa Anne Hendricks, Karolina Stańczak, and Aishwarya Agrawal. 2024. Benchmarking vision language models for cultural understanding. *arXiv preprint arXiv:2407.10920*.
- OpenAI. 2024. **Gpt-4o mini: advancing cost-efficient intelligence.**
- OpenAI. 2024. **Gpt-4o system card.** *Preprint*, arXiv:2410.21276.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Aida Ramezani and Yang Xu. 2023. **Knowledge of cultural moral norms in large language models.** *ArXiv*, abs/2306.01857.
- Machel Reid, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lilllicrap, Jean baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittweis, and et al. 2024. **Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context.** *Preprint*, arXiv:2403.05530.
- David Romero, Chenyang Lyu, Haryo Akbarianto Wibowo, Teresa Lynn, Injy Hamed, Aditya Nanda Kishore, Aishik Mandal, Alina Dragonetti, Artem Abzaliev, Atnafu Lambebo Tonja, et al. 2024. Cvqa: Culturally-diverse multilingual visual question answering benchmark. *arXiv preprint arXiv:2406.05967*.
- Yan Tao, Olga Viberg, Ryan S Baker, and René F Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS nexus*, 3(9):pgae346.
- Qwen Team. 2025. **Qwen2.5-vl.**
- Yingtao Tian, Chikahiko Suzuki, Tarin Clanuwat, Mikel Bober-Irizar, Alex Lamb, and Asanobu Kitamoto. 2020. Kaokore: A pre-modern japanese art facial expression dataset. *arXiv preprint arXiv:2002.08595*.
- Caroline Tully. 2020. British museum. In *Encyclopedia of Global Archaeology*, pages 1618–1620. Springer.
- Michael EW Varnum, Nicolas Baumard, Mohammad Atari, and Kurt Gray. 2024. Large language models based on historical text could offer informative tools for behavioral science. *Proceedings of the National Academy of Sciences*, 121(42):e2407639121.
- Mei Wang and Weihong Deng. 2022. Oracle-mnist: a realistic image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:2205.09442*.
- Pengjie Wang, Kaile Zhang, Xinyu Wang, Shengwei Han, Yongge Liu, Jinpeng Wan, Haisu Guan, Zhebin Kuang, Lianwen Jin, Xiang Bai, et al. 2024. An open dataset for oracle bone script recognition and decipherment. *arXiv preprint arXiv:2401.15365*.
- Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, and Wenhui Chen. 2024. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In *Proceedings of CVPR*.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

A Appendix

In this appendix, we provide additional details to support our research, including related work, data statistics, and a comprehensive overview of archaeological samples from various cultures, civilizations, and dynasties. The related work section provides a review of existing research in AI-driven historical text generation, contextualizing our contributions within the broader field. The data statistics section offers a structured breakdown of collected samples, highlighting their geographical distribution and cultural significance. Additionally, the inclusion of archaeological records from diverse historical periods reinforces the depth and diversity of the dataset.

B Related Work

Recent years have seen significant progress in studying cultural representation in AI, particularly in behavioral patterns, food, landmarks, and historical knowledge. However, most works focus on misalignment and biases in AI models or modern cultural trends, rather than positioning artifacts within their historical context and era across ancient civilizations. Meanwhile, studies on cultural inclusion in LLMs highlight the challenges of capturing the contextual and multifaceted nature of culture, emphasizing the limitations of text-based models in representing underrepresented cultures and the need for more robust evaluation methods (Adilazuarda et al., 2024).

Research on cultural influences in AI has increasingly focused on biases and misalignment in language models, particularly how they reflect and perpetuate dominant cultural norms. Early research on cultural biases in LLMs revealed their alignment with Western norms, particularly in moral reasoning, historical narratives, and societal values. Ramezani et al. (2023) analyze how monolingual English language models tend to reflect Western moral norms more strongly than diverse cultural perspectives, limiting their applicability in cross-cultural ethical contexts (Ramezani and Xu, 2023). Tao et al. (2024) further highlight the overrepresentation of Anglo-American and Protestant European values in AI-generated content, often underrepresenting non-Western traditions and belief systems (Tao et al., 2024). Similarly, Bu et al. (2025) explore value misalignment in cultural heritage-related text generation, warning of historical inaccuracies, cultural identity erosion,

and oversimplification of complex narratives, with 65% of the generated content showing significant misalignment (Bu et al., 2025).

To mitigate these biases, several approaches have been proposed. AlKhamissi et al. (2024) introduce Anthropological Prompting, a method that encourages LLMs to reason like cultural anthropologists by incorporating both emic (insider) and etic (outsider) perspectives (AlKhamissi et al., 2024). Similarly, Li et al. (2024) propose CultureLLM, a fine-tuning approach designed to integrate cultural knowledge into LLMs, particularly for low-resource cultures (Li et al., 2024). While these techniques improve cultural alignment, their focus remains on modern cultural settings, leaving gaps in historical artifact contextualization across different time periods.

With the rise of Vision-Language Models (VLMs), cultural research has expanded to multimodal AI, revealing similar biases. Liu et al. (2025) introduce CultureVLM, a model designed to improve cultural understanding in VLMs, highlighting their inability to recognize non-Western cultural symbols, historical artifacts, and traditional gestures (Liu et al., 2025). Their work also presents CultureVerse, a large-scale multimodal dataset covering several cultural concepts, designed to evaluate VLMs' cultural reasoning. However, CultureVerse has a primary focus on modern cultural symbols, traditions, and everyday life. Additionally, Romero et al. (2024) develop CVQA, a multilingual and culturally diverse Visual Question Answering (VQA) benchmark, which reveals that state-of-the-art VLMs struggle with culturally grounded reasoning, particularly in non-Western contexts (Romero et al., 2024). However, these datasets primarily focus on present-day cultural contexts, even when historical artifacts are included, as they are often framed through the lens of modern nations rather than their original civilizations and historical epochs (Liu et al., 2025). This leaves a significant gap in representing artifacts within their authentic temporal and cultural contexts.

Efforts to bridge AI research with historical studies have led to the development of Historical Large Language Models (HLLMs), trained on historical texts to simulate past societies' psychology and value systems (Varnum et al., 2024). These models aim to provide insight into long-term cultural evolution, but their reliance on text-only representations limits their application in multimodal historical

Model	India	Roman Empire	China	British Isles	Iran	Iraq	Japan	Central America	Greece	Egypt
GPT-4o-0806 (OpenAI, 2024)	0.2566	0.2713	0.2324	0.2175	0.2486	0.2428	0.2269	0.2384	0.2441	0.2567
Gemini-2.0-Flash (Reid et al., 2024)	0.2478	0.2603	0.2183	0.2189	0.2432	0.242	0.2256	0.2264	0.2488	0.2588
Gemini-1.5-Pro (Reid et al., 2024)	0.2586	0.2596	0.2198	0.2203	0.2535	0.2524	0.2253	0.2218	0.2551	0.268
GPT-4o-mini-0718 (OpenAI, 2024)	0.2762	0.2731	0.2570	0.2531	0.2660	0.2640	0.2611	0.2741	0.2649	0.2741
Llama-3.2-Vision-Inst (Meta AI, 2024)	0.2128	0.2253	0.1867	0.1917	0.2115,	0.2078	0.1944	0.1979	0.2138	0.2182
Qwen-2.5-VL (Team, 2025)	0.2707	0.2815	0.2526	0.2464	0.2607	0.2631	0.2499	0.2587	0.2713	0.2827
Llava-Next (Liu et al., 2024)	0.2482	0.2527	0.2156	0.2192	0.2389	0.2321	0.2207	0.2196	0.2388	0.2427

Table 4: Analysis of METEOR Evaluation of various models in describing archaeological artifacts across civilizations from different geographical regions.

studies. Similarly, Assael et al. (2022) introduce Ithaca, a deep learning model designed to assist historians in restoring, geographically attributing, and dating ancient Greek inscriptions, significantly improving accuracy over traditional methods (Assael et al., 2022). While these works contribute to historical AI, they primarily focus on text-based reconstruction rather than multimodal representations of historical artifacts across civilizations.

TimeTravel fills this gap by providing a 10k historical artifact open-source dataset spanning 10 ancient world regions (prehistoric and historic), offering the first benchmark to evaluate LMMs on temporal-cultural understanding with expert verification. Unlike prior datasets focused on contemporary cultural knowledge, TimeTravel enables AI models to contextualize artifacts within their historical era, ensuring a more accurate representation of civilizations and their material culture. With domain expert verification, the dataset enhances reliability and authenticity, mitigating potential biases and inaccuracies in AI-generated interpretations. By integrating both textual and multimodal perspectives, TimeTravel advances research in historical-cultural AI, enabling AI systems to better understand and reason about artifacts in their original context.

C TimeTravel Samples Regional Distribution

Fig. 4 illustrates the balanced regional distribution of dataset samples based on archaeological provenance. Greece holds the largest share at 18%, followed by multiple regions, including the Roman Empire, China, British Isles, Egypt, Iraq, and Iran, each at 10%. Japan (9%), India (8%), and Central America (5%) contribute smaller yet significant portions. Overall, the dataset ensures diverse cultural representation without dominance by any single region.

Tables 5 to 14 present further details about

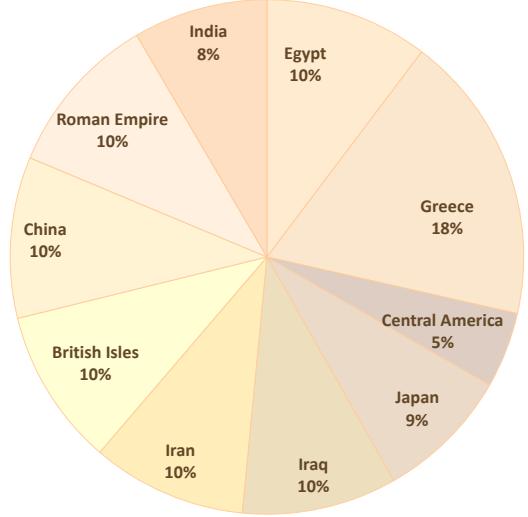


Figure 4: Regional distribution of dataset samples based on their archaeological provenance. Greece holds the largest share at 18%, with a balance-like distribution over regions.

sample counts categorized by region of discovery, section, and cultural affiliation.

The covered areas in our study are ordered as follows:

Tab. 5 → “Roman Empire”, Tab. 6 → “Greece”, Tab. 7 → “British Isles”, Tab. 8 → “Central America”, Tab. 9 → “Egypt”, Tab. 10 → “India”, Tab. 11 → “Iran”, Tab. 12 → “China”, Tab. 13 → “Japan”, and Tab. 14 → “Iraq”,

Place	Roman Empire
Section	Roman
Culture	Samples
Roman Imperial	610
Roman	3
Roman Provincial	436
Total	1049

Table 5: Culture Sample Counts from the Roman Empire.

D TimeTravel Benchmark Examples

Place Section		Greece Greek	
Culture	Sample	Culture	Sample
Greek; Hellenistic; Roman Imperial	4	Hellenistic; Roman Imperial	2
Attic	806	Middle Corinthian	5
Corinthian	41	East Greek; Classical Greek	1
Attic; Classical Greek	47	Transitional Corinthian	1
Middle Corinthian; Late Corinthian; Archaic Greek	7	Classical Greek; Attic	2
Proto-Corinthian	4	Classical Greek; Attic; Archaic Greek	1
Orientalising Period	14	East Greek Archaic II; Archaic Greek	1
Archaic Greek; Classical Greek	1	Attic; Western Greek	1
Archaic Greek	40	East Greek	23
Late Corinthian; Archaic Greek	11	Attic; Archaic Greek	318
Western Greek; Hellenistic	1	Attic; Archaic Greek; Classical Greek	12
Early Corinthian	8	Attic; Classical Greek; Archaic Greek	3
Laconian; Archaic Greek	10	Archaic Greek; East Greek	2
Classical Greek; Corinthian; Hellenistic	1	Rhodian	3
Late Helladic IIIB	2	Greek; Classical Greek	2
Transitional Corinthian; Archaic Greek	1	Early Corinthian; Archaic Greek	3
East Greek; Hellenistic	2	Middle Corinthian; Archaic Greek	11
Late Geometric IIA; Attic	1	East Greek; Orientalising Period	1
Archaic Greek; Attic	8	Late Minoan I; Late Minoan II	1
Late Minoan I	2	Archaic Greek; East Greek; North Ionian	1
Paestan	1	East Greek; Archaic Greek	237
Early Corinthian; Middle Corinthian; Archaic Greek	1	Greek; Hellenistic	2
Archaic Greek; East Dorian	1	Greek	3
Hellenistic	110	Western Greek	5
East Greek; Archaic Greek; Classical Greek	1	Roman; Hellenistic	3
East Dorian; Archaic Greek	2	Classical Greek	38
East Greek; East Dorian; Archaic Greek	11	Boeotian	25
Geometric Greek; Early Proto-Attic	1	Hellenistic; Classical Greek	2
East Greek; South Ionian	1	Geometric Greek	8
Greek; Classical Greek; Hellenistic	5	Hellenistic; Roman	4
Total			1869

Table 6: Culture Sample Counts from Greece (Greek Section).

Place Section	British Isles	
Culture	Samples	
Viking; Carolingian; Late Anglo-Saxon	1	
Viking; Early Anglo-Saxon; Mid. Anglo-Saxon	1	
Middle Anglo-Saxon Viking; Anglo-Saxon	1	
Celtic; Viking	14	
Viking; Late Anglo-Saxon	19	
Viking; Finno-Ugrian	1	
Anglo-Viking	52	
Viking	895	
Carolingian; Viking	1	
Viking; Medieval	1	
Late Anglo-Saxon; Viking	1	
Viking; Celtic	26	
Total	1013	

Table 7: Culture Sample Counts from the British Isles (Viking Section).

Place Section	Central America	
Culture	Samples	
Classic Maya; Classic	3	
Classic Maya; Late Preclassic Maya	64	
Formative (Pre-Classic); Early Classic Maya	8	
Late Classic Maya	23	
Olmec; Maya	1	
Classic Maya	275	
Preclassic Maya	10	
Classic Maya; Late Classic	2	
Classic Maya; Olmec	1	
Preclassic Maya; Classic Maya	2	
Maya	95	
Late Classic Maya; Late Classic	4	
Total	488	

Table 8: Culture Sample Counts from Central America (Maya Section).

Place	Egypt
Section	Ancient Egyptian
Culture	Samples
6 th Dynasty	1
Late Cypriot; 18 th Dynasty	1
26 th Dynasty; Archaic Greek; Punic	1
Late Period; 30 th Dynasty	1
30 th Dynasty; Ptolemaic	15
22 nd Dynasty	69
18 th Dynasty; 19 th Dynasty	2
New Kingdom; 19 th Dynasty; 20 th Dynasty	1
12 th Dynasty	1
26 th Dynasty; Archaic Greek; East Greek; Hellenistic New Kingdom	21
Late Predynastic; 1 st Dynasty	2
25 th Dynasty	7
30 th Dynasty	128
Middle Kingdom	1
Late Period	96
18 th Dynasty; 21St Dynasty	1
21 st Dynasty	171
19 th Dynasty; 20 th Dynasty	3
20 th Dynasty; 21St Dynasty	2
26 th Dynasty	257
19 th Dynasty	40
18 th Dynasty	95
1 st Dynasty	7
Ramesside	21
23 rd Dynasty	1
22 nd Dynasty; 23R rd Dynasty	3
26 th Dynasty; 27 th Dynasty	1
20 th Dynasty	25
Late Period; 26 th Dynasty	3
25 th Dynasty; Kushite	1
26 th Dynasty; Punic; Archaic Greek	1
27 th Dynasty	13
25 th Dynasty; 26 th Dynasty	1
Third Intermediate	47
Late Period; Archaic Greek	2
Late Period; Ptolemaic	5
29 th Dynasty	1
New Kingdom; Third Intermediate	1
Ancient Egypt	5
26 th Dynasty; 30 th Dynasty	1
Total	1056

Table 9: Culture Sample Counts from Egypt (Ancient Egyptian Section).

Place	India
Section	Mohenjo-Daro
Culture	Sample Count
Indus Valley Civil.	114
Section	Mauryan
Culture	Sample Count
Mauryan	17
Section	Gupta Dynasty
Culture	Sample Count
Gupta	737
Total	868

Table 10: Culture Sample Counts from India.

Place	Iran
Section	Persian
Culture	Samples
Inju Dynasty	3
Middle Islamic; Seljuq Dynasty; Persian	1
Safavid Dynasty; Mughal Dynasty	1
Persian; Islamic	11
Persian; Late Islamic	3
Samanid Dynasty	27
Safavid Dynasty	395
Timurid Dynasty; Islamic	1
Safavid Dynasty; Post-Medieval	1
Mughal Dynasty; Persian	1
Ilkhanid Dynasty; Persian	3
Turkman Dynasty	3
Early Sasanian; Safavid Dynasty	1
Islamic; Safavid Dynasty	1
Ilkhanid Dynasty	192
Middle Islamic; Persian	6
Islamic; Qajar Dynasty	2
Persian; Safavid Dynasty	1
Safavid Dynasty; Persian; Islamic	2
Mughal Dynasty; Safavid Dynasty	1
Qajar Dynasty	193
Safavid Dynasty; Islamic	4
Persian; Mughal Dynasty	1
Islamic; Persian	2
Timurid Dynasty	35
Persian	108
Total	999

Table 11: Culture Sample Counts from Iran (Persian Section).

Place	China
Section	Tang Dynasty
Culture	Samples
Tang Dynasty; Sui Dynasty	1
Tang Dynasty; Ming Dynasty	3
Tang Dynasty; Ming Dynasty; Jin Dynasty; Yuan Dynasty	1
Tang Dynasty; Song Dynasty	1
Song Dynasty; Tang Dynasty	1
Liao Dynasty; Tang Dynasty	2
Tang Dynasty; Northern Wei Dynasty	1
Six Dynasties; Tang Dynasty	5
Tang Dynasty	1
Northern Qi Dynasty; Sui Dynasty; Tang Dynasty	1
Tang Dynasty; Liao Dynasty	3
Six Dynasties; Sui Dynasty; Tang Dynasty	1
Tang Dynasty; Five Dynasties; Northern Song Dynasty	381
Five Dynasties; Tang Dynasty	4
Tang Dynasty	628
Sui Dynasty; Tang Dynasty	5
Total	1039

Table 12: Culture Sample Counts from China (Tang Dynasty Section).

Place	Japan
Section	Japanese
Culture	Samples
Momoyama Period	6
Genroku Era; Hoei Era	1
Asuka Period	1
Muromachi Period; Momoyama Period	2
Late Kofun; Nara Period	1
Nara Period	12
Middle Kofun	13
Yayoi Period	5
Middle Kofun; Late Kofun	34
Edo Period; Kamakura Period	1
Oei Era	2
Kyowa Era; Oei Era	1
Edo Period; Momoyama Period	1
Jomon Period	16
Kyowa Era	1
Bunka Era	1
Bun'An Era; Bunsei Era	1
Muromachi Period	40
Asuka Period; Nara Period	1
Heian Period	9
Muromachi Period; Momoyama Period; Edo Period	1
Muromachi Period; Buddhist	1
Meiji Era	1
Hakuho Period	1
Showa Era	13
Early Kofun; Middle Kofun	26
Nanbokucho Period	2
Kofun Period; Edo Period	1
Edo Period	24
Kamakura Period; Meiji Era	1
Kofun Period	419
Early Kofun	7
Wado Era	1
Late Kofun	179
Kofun Period; Asuka Period	5
Kamakura Period	26
Nara Period; Edo Period	1
Kofun Period; Nara Period	1
Kamakura Period; Muromachi Period	9
Heian Period; Kamakura Period	1
Total	869

Table 13: Culture Sample Counts from Japan (Japanese Section).

Place	Iraq
Section	Mesopotamian
Culture	Samples
Neo-Assyrian; Late Babylonian	9
Late Babylonian; Assyrian	1
Elamite; Third Dynasty Of Ur	1
Early Dynastic (Middle East)	1
Old Assyrian; Early Bronze Age III	1
Late Uruk	26
Isin-Larsa	3
Neo-Assyrian	406
Uruk	3
Late Uruk; Chalcolithic	1
Middle Babylonian; Neo-Babylonian	1
Dynasty	
Old Babylonian; Cypriot	1
Late Babylonian	20
Babylonian; Neo-Assyrian	1
Neo-Assyrian; Babylonian	1
Assyrian; Late Babylonian	2
Jemdet Nasr; Proto-Elamite	1
Halaf	38
Assyrian	7
Middle Assyrian	11
Jemdet Nasr	27
Third Dynasty Of Ur; Ubaid	1
Old Babylonian	41
Kassite	4
Babylonian	3
Neo-Babylonian Dynasty	2
Babylonian; Akkadian	1
Old Assyrian	2
Old Babylonian; Third Dynasty Of Ur	1
Ubaid	15
Early Dynastic (Middle East); Akkadian	2
Early Dynastic II	2
Isin-Larsa; Old Babylonian	1
Jemdet Nasr; Akkadian	1
Old Babylonian; Assyrian	1
Akkadian	102
Early Dynastic III; Akkadian	10
Old Babylonian; Old Assyrian	1
Isin-Larsa; Old Babylonian; Kassite	1
Uruk; Jemdet Nasr	3
Early Dynastic II; Early Dynastic III	1
Early Dynastic III	81
Mesopotamian	1
Late Babylonian; Neo-Assyrian	3
Assyrian; Ubaid	1
Third Dynasty Of Ur; Old Babylonian	2
Third Dynasty Of Ur	137
Neo-Assyrian; Phoenician	2
Middle Babylonian	11
Lagash II	1
Third Dynasty Of Ur; Isin-Larsa	4
Total	1000

Table 14: Culture Sample Counts from Iraq (Mesopotamian Section).



```
{
  "id": 4621,
  "Image": ".../Documents/2014_11/5_17/e572b5a1_adcb_4691_871c_a3da0123a3fa/preview_0122265_001.jpg",
  "Description": "Silver coin.; Bust of Skandagupta, facing to the right, wearing a decorated cap and collar.; Peacock, with outspread wings and tail, standing facing with head to left. Surrounded by a border of dots.",
  "Production date": "456-467 (about)",
  "Find spot": null,
  "Materials": "silver",
  "Technique": null,
  "Inscription": "Inscription type: inscription Inscription position: reverse Inscription language: Sanskrit Inscription script: Brahmi",
  "Subjects": "king/queen; peacock",
  "Assoc name": "Named in inscription & portrayed: Skandagupta",
  "Culture": "Gupta",
  "Section": "Gupta dynasty",
  "Place": "India"
}
```

for this artifact, we generated a description such as:

```
{
  "description": "This silver coin from the Gupta dynasty, produced approximately between 456-467 CE, features a distinguished portrait of Skandagupta on the obverse. Wearing an ornate cap and collar, Skandagupta is depicted facing right, encapsulating the regal aesthetic of the period. The reverse displays a majestic peacock, wings and tail fully spread, elegantly poised with its head turned to the left, bordered by a circle of dots adding to the coin's decorative appeal. An inscription in Sanskrit, using the Brahmi script, adorns the reverse, providing linguistic and cultural context. This artifact, emblematic of Gupta artistry and power, underscores the dynastic symbolism and reverence for nature characteristic of this era in Indian history."
}
```



```
{
  "id": 7443,
  "Image": ".../Documents/2015_9/4_10/b0f2152e_0ab5_4585_a7ca_a50900a611e8/preview_1997_q_1879_a.jpg",
  "Description": "Votive object in the form of polished piece of jade with six holes drilled in it.",
  "Production date": null,
  "Find spot": "Excavated/Findspot: Tzimin Kax (?); Excavated/Findspot: Hatzcap Ceel (?); Excavated/Findspot: Cahal Pichik (?)",
  "Materials": "jade",
  "Technique": "polished; drilled",
  "Inscription": null,
  "Subjects": null,
  "Assoc name": null,
  "Culture": "Classic Maya; Late Preclassic Maya",
  "Section": "Maya",
  "Place": "Central America"
}

for this artifact, we generated a description such as:

{
  "description": "This votive object, crafted by the Classic and Late Preclassic Maya cultures, is a testament to the artistry and spiritual practices of ancient Central America. Made from polished jade, a material revered for its symbolic and aesthetic value, this artifact features six precision-drilled holes, showcasing the sophisticated techniques employed by the Maya artisans. Although specific inscription details are absent, the object's polished surface and drilled holes suggest it may have had ceremonial or religious significance. Unearthed at find spots such as Tzimin Kax, Hatzcap Ceel, and Cahal Pichik, this piece provides valuable insights into the ritualistic and cultural engagements of the Maya civilization."
}
```

Figure 5: This entry represents a silver coin from the Gupta dynasty from India, featuring a distinguished portrait of Skandagupta on the obverse. GPT-4o generated a detailed, context-aware description based on the available metadata, highlighting its craftsmanship, ceremonial significance, and cultural context.

Figure 6: This particular entry represents a polished jade votive object from the Classic and Late Preclassic Maya. It features six precision-drilled holes, reflecting advanced craftsmanship and likely ceremonial significance. Unearthed at sites like Tzimin Kax, it offers insight into Maya rituals.

<p>Ancient Egyptian / Egypt</p> <p>"description": This exquisite shabti, crafted from green glazed composition, hails from the Late Period of Ancient Egypt. It was excavated at Tell el-Dab'a. It features a striped coiffed-wig, a finely modelled face with a beard, and crossed hands that hold a relief adze, hoe, and the stela of a scribe. The figure is depicted in a kneeling pose. Notably, the leg section is inscribed with nine rows of hieroglyphic text that identifies the figure as Diehulyirdis, the High Priest of Thoth, with the title of "Scribe of the Mortuary". The presence of a back-pillar and plinth adds to its structural elegance, embodying the funerary function and cultural significance of these figures as servants in the afterlife.</p> 	<p>Indus Valley Civilisation / India</p> <p>"description": This artifact is a fragmentary terracotta figure from the Indus Valley Civilization, dating approximately to 2500 to 2000 BCE. It depicts a female figure, specifically a pregnant woman, crafted in buff terracotta with a red slip finish, suggesting its original vibrant coloration. As part of the Mohenjo-daro collection, this piece belongs to one of the core urban centers of the ancient Indus Valley, underscoring the sophisticated craftsmanship and artistic expression of the society of the era. Discovered at the renowned archaeological site of Mohenjo-daro in present-day India, this artifact provides a tangible connection to one of the world's earliest urban settlements.</p> 	<p>Tang Dynasty / China</p> <p>"description": This exquisite earthenware cup, dating from late Tang dynasty through the Ming dynasty (1368-1644), is a remarkable example of Chinese craftsmanship. It is distinguished by its rounded form and a wide rim. The cup features a flared base and a slightly conical body. It is decorated with a light brown glaze, with some surface wear visible where marks, indicative of its handcrafted nature, and thread marks are evident on its base. Notably, the bowl base is inscribed with a date, a technique typical of the period, enhancing its rustic aesthetic. An inscription in English, referencing "Near Akasaka, Mino Kyomizu", adds another layer of depth to its historical intrigue. This artifact was discovered at Akasaka-cho, specifically near the Kyomizu dolmen, highlighting its archaeological significance within the region.</p> 	<p>Classical Greek / Greek</p> <p>"description": This classical Greek relief, dated from circa 490-470 BC, is a fragment of a larger scene. It depicts the mythological scene of Eos carrying off Kephalos. Crafted using a moulding technique, this artifact captures a moment of divine intervention, featuring Eos, the personification of dawn, and Kephalos, a mortal figure entrained in various forms of narratives. The plaque exhibits fine and delicate details of the period, with intricate detailing that evokes the ethereal beauty and narrative depth of Greek mythology. Unearthed from Greece, this fragment not only represents the aesthetic values of its time but also enriches our understanding of Greek funerary practices and mythological storytelling.</p> 
<p>Japanese / Japan</p> <p>"description": This shallow bowl, dating from the 9th to 14th centuries, originates from Japan's Kamakura Period. Crafted from high-fired stoneware, the bowl displays a conical shape with a slightly conical base. It features a light brown surface with visible wheel marks, indicative of its handcrafted nature, and thread marks are evident on its base. Notably, the bowl base is inscribed with a date, a technique typical of the period, enhancing its rustic aesthetic. An inscription in English, referencing "Near Akasaka, Mino Kyomizu", adds another layer of depth to its historical intrigue. This artifact was discovered at Akasaka-cho, specifically near the Kyomizu dolmen, highlighting its archaeological significance within the region.</p> 	<p>Maya / Central America</p> <p>"description": This artifact collection consists of ten stone tools of Maya origin, originating from the 3rd century BC to the 1st century AD. Although the precise production date remains unspecified, these implements serve as a testament to the technological and cultural practices of the Maya civilization. The tools, excavated from the region of Lubana, are crafted from stone and made from common materials used by the Maya for tool-making due to their durability and availability. Without inscriptions or specific iconographic elements, these tools speak primarily through their form and functionality, offering insights into daily life and survival strategies in ancient Maya communities.</p> 	<p>Mesopotamian / Iraq</p> <p>"description": This exquisite Neo-Assyrian gypsum tablet, dating to between 645-635 BC, exquisitely depicts a female musician in the lush setting of Ashurbanipal's garden. The musician is portrayed advancing to the right, wearing a long, flowing robe decorated with a headband, necklace, triple-knotted earrings, and bracelets. Her garment features a distinctively patterned edge on the sleeves, and she holds a long, thin staff. The tablet's opulent aesthetic and cultural sophistication of the Neo-Assyrian period. This fragment was excavated from the North Palace in Nineveh, present-day Iraq, providing a vivid glimpse into the artistic accomplishments and daily life in the heart of the Assyrian Empire.</p> 	<p>Tang Dynasty / China</p> <p>"description": This black earthenware vase, dating from the 8th to the 10th century, originates from the Tang dynasty in China. It features a bulbous design with a broad base and is adorned with two incised bands around the shoulder, showcasing the incised carving technique popular during this cultural period. Crafted from earthy clay, this artifact reflects the aesthetic and functional qualities typical of Tang dynasty ceramics. The artifact was sold at auction in 2018 and is historically significant for its role as a cultural crossroads along the Silk Road. This find spot underscores the vase's historical context within the exchange of art and ideas during the Tang era.</p> 
<p>Viking / British Isles</p> <p>"description": This Viking silver armband, crafted between the 9th and 11th centuries, exemplifies the intricate metalwork of the period. It features a repeating band and an expanded center adorned with a 6-pointed star created through a punched zigzag technique, while each terminal is marked by a salire, reflecting the motifs prevalent during the Viking period. A small, circular brooch, a token of personal adornment, offers insight into the cultural exchanges within the British Isles during this period, as evidenced by its discovery in the Republic of Ireland. The use of silver and intricate punched designs suggest it may have been a status symbol among its original owners, further underlining the interconnectedness of Viking and local cultures in the British Isles.</p> 	<p>Persian / Iran</p> <p>"description": This exquisite artifact from the Qajar dynasty, dating approximately from the 18th to 19th century, is a delicate figure of a dove crafted from steel and adorned with inlaid garnet and gold. Notably, it bears the inscription of its maker, Hajji 'Abbas, showcasing the intricate craftsmanship and meticulous attention to detail characteristic of Persian metalworking artistry. The use of fine materials and the inlay technique highlight the luxurious and refined aesthetic practices of the era, symbolizing both cultural and artistic values within the context of Qajar Iran. While the specific findspot remains undisclosed, the figure resonates with the rich heritage and artistic innovation prevalent in the region during this period.</p> 	<p>Ancient Egyptian / Egypt</p> <p>"description": This exquisite scarab, dating from 1504-1020 BCE, reflects the artistry and cultural symbolism of Egypt's 18th Dynasty. Crafted from glazed composition, the piece is intricately engraved with the hieroglyphic name of Amun-Ra, the sun disk, and the cartouche of the Pharaoh. The scarab is framed within a cartouche flanked by the protective Uraeus serpents and adorned with neb signs. Above these, smaller ureus and maat feathers are delicately embossed, accompanied by the hieroglyphic sign for gold, emerald, and the symbol for truth, representing Egyptian society. Although much of the original glaze has vanished, traces of the original detailing remain, showcasing the sophisticated mold-making and glazing techniques of the period. Used for suspension, this artifact may have served a ceremonial or decorative purpose. It was discovered around Maroni, further enriching its historical and geographic narrative tied to the ancient Egyptian realm.</p> 	<p>Gupta / India</p> <p>"description": This Gupta dynasty gold coin, dating approximately to 456-467 CE, epitomizes the opulence and artistry of its era. On one side, it features Skandagupta, the ruler of the Gupta Empire, seated in a highly decorated coat with a long sash and trousers, and encircled by a halo. He wields a bow in his left hand and an arrow in his right, with a Ganesha motif visible on the bow. On the other side, a queen, possibly Lata, is seated gracefully on a lotus, clad in lower garments, a shawl, and jewelry, and also haloed. Her right hand is outstretched holding a fillet, while her left supports the stem of a lotus; a symbol is present in the left field, and the scene is encircled by a border of dots.</p> 
<p>Neo-Assyrian; Late Babylonian / Mesopotamia</p> <p>"description": This artifact is a black stone conical pendant originally from the Neo-Assyrian or Late Babylonian period, dated between the 9th and 4th centuries BC. The pendant features a hole at the top, suggesting it was designed to be worn or suspended. Carved from dark stone, it depicts a centrally-forged figure oriented to the right, accompanied by a flower and the sun, which are emblematic of ancient Mesopotamian symbolic values. The figure's orientation and the sun may allude to religious or cultural motifs common in the region during this era. The pendant was excavated from the South East Palace in Nimrud, Iraq, providing valuable insight into the artifacts utilized in the heart of ancient Assyrian civilization.</p> 	<p>Mohenjo-daro / India</p> <p>"description": This terracotta figure represents a humped bull, crafted between 2500-2000 BCE by the Indus Valley Civilization. Exhibiting a buff hue, the bull figure likely served a symbolic purpose, possibly related to agricultural and cultural significance of cattle in the Indus Valley region. This artifact was unearthed from the ancient city of Mohenjo-daro, a major urban center of the civilization located in present-day Pakistan. It encapsulates the artistic and societal practices of one of the world's earliest urban cultures, offering a glimpse into their agricultural and spiritual beliefs.</p> 	<p>Mesopotamian / Iraq</p> <p>"description": This exquisite chalcedony cylinder seal, originating from the Neo-Assyrian to Late Babylonian period and possibly dating back to ancient Mesopotamia, is a captivating artifact. Both in form and function, this piece is a true work of art. Carved from a single piece of chalcedony, the seal depicts a dynamic scene characterized by a female worshipper with extended hands, standing opposite a beardless deity, presumably to a goddess adorned with a golden diadem and a tiered skirt. According to the legend of the mythical tablet, the bearded scorpion-man, whose form merges human traits with a bird or lion, is shown in hand-to-hand combat with the goddess. The seal is intricately detailed, with a distinct fish-tail handle and a lid adorned with vegetal signs and a chalice on the final tier. Further enhanced is a foot with a scalloped edge. The seal showcases a complex decorative scheme: a white slip underlies the transparent colorless glaze on the sides, while a lustrous overglaze accentuates the cobalt glaze on the shoulders.</p> 	<p>Persian / Iran</p> <p>"description": This exquisite ewer, originating from the Safavid dynasty in the 17th century, exemplifies the lusterware craftsmanship of pre-modern Iran. The artifact is crafted from stoneware with a flattened globular body, a stepped shoulder, and a flared base resting on a ring foot. Distinctive features include an elegant European brass fish-tail handle and a lid adorned with vegetal signs and a chalice on the final tier. Further enhanced is a foot with a scalloped edge. The ewer showcases a complex decorative scheme: a white slip underlies the transparent colorless glaze on the sides, while a lustrous overglaze accentuates the cobalt glaze on the shoulders.</p> 
<p>Greek / Greece</p> <p>"description": This Attic black-figure skyphos, crafted around 480 BCE, is a remarkable example of ancient Greek pottery. It features a central scene depicting a youth, likely Apollo, seated between palm trees. Made from pottery, its design exemplifies classical Greek artistry and the cultural significance of music and deities in ancient Greece. The skyphos was unearthed from grave 61 at Filakula, indicating its use in funerary contexts and providing valuable insight into the burial practices and artistic expressions of the period.</p> 	<p>Ancient Egyptian / Egypt</p> <p>"description": This exquisite 19th Dynasty Egyptian pectoral, dated circa 1275 BC, is a remarkable example of ancient artistry and religious symbolism. Crafted from glazed composition, it is shaped like a pyramid, mimicking the monumental structures found in Egyptian temples. The pectoral slopes down and a curved cartouche cornice, decorated with block patterns around its edges. Its vibrant color palette includes yellow, pink, and blue glass, with various precious materials like gold, lapis lazuli, cornelian, and turquoise. Central to the design is a dark blue heart scarab, representing lapis lazuli and depicted within a solar barque, flanked by the goddesses Isis.</p> 	<p>Tang Dynasty / China</p> <p>"description": This exquisite white ware cup from the Tang dynasty, dated between 618-905 CE, exemplifies the artistry of ancient China. Crafted from cream-glazed and carved earthenware, the cup is adorned with an intricate design featuring seven figures of Budhas, artfully splashed with green accents. The cup's form is a squat vessel with a flared base, reflecting the aesthetic sensibilities of the period, highlighting the profound influence Buddhism had on Chinese culture during the Tang dynasty. Discovered in China, this artifact provides valuable insight into the religious and aesthetic practices prevalent during one of the golden ages of Chinese history.</p> 	<p>Greek / Greece</p> <p>"description": This Attic black-figure kylix, dating back to circa 500 BC, is a testament to the skilled craftsmanship of ancient Greece, combining both painted and incised techniques. Made from pottery, this artifact features a striking design set against its iconic eyes: a scene illustrating a satyr in an act of assault on a fawn, a subject often reflected in the exploration of themes such as procreation and interplay between mythological creatures and nature. While no inscriptions are present, the visual narrative is rich and evocative. Additionally, it features a string of beads, remains undocumented, the kylix hails from Greece, providing insight into the cultural expressions and iconography of the period.</p> 

Figure 7: Cultural and material diversity of TimeTravel dataset samples across civilizations and historical periods. The dataset includes artifacts from Ancient Egypt, Greece, Mesopotamia, China, and Japan, spanning prehistoric to medieval times. A wide range of materials, including ceramics, metals, and stone, highlights artistic, technological, and societal influences, ensuring a comprehensive representation of historical craftsmanship and cultural heritage.

Roman Imperial / Roman Empire	GPT-4o	Gemini	Qwen	Llava-Next	Ground-Truth		
	This image features an ancient Roman coin, possibly a denarius. The coin has two sides, each with intricate designs typical of Roman craftsmanship. The Obverse (Front Side) "The obverse side depicts a profile of the Roman emperor. He is depicted with a laurel wreath on his head, a common symbol of victory and authority in Ancient Rome. The reverse side depicts another profile, which usually includes the emperor's name and title. These inscriptions are often abbreviations, such as "IMP TRAIANVS AVG PIVS FEL AVGVSTVS" on the obverse and "P M TR P COS V P P" on the reverse. There is another side of the coin depicting around him. There is another side of the coin, labeled "Back Side". The reverse side depicts a river god, possibly representing a specific river. The figure is semi-nude and holds a cornucopia overflowing with water and various symbols. The inscription around the image, often describing the scene or the emperor's achievements, is visible on both sides of the coin. The coin's detailed artwork and inscriptions provide insights into the political and cultural aspirations of the Roman Empire and its leaders, policies, and artistic expressions. Coins like this were not only used for commerce but also served as means of propaganda, promoting the emperor's achievements and divine favor.	A brief description of the image from GPT-4o: This image features an ancient Roman coin, possibly a denarius. The coin has two sides, each with intricate designs typical of Roman craftsmanship. The Obverse (Front Side) "The obverse side depicts a profile of the Roman emperor. He is depicted with a laurel wreath on his head, a common symbol of victory and authority in Ancient Rome. The reverse side depicts another profile, which usually includes the emperor's name and title. These inscriptions are often abbreviations, such as "IMP TRAIANVS AVG PIVS FEL AVGVSTVS" on the obverse and "P M TR P COS V P P" on the reverse. There is another side of the coin depicting around him. There is another side of the coin, labeled "Back Side". The reverse side depicts a river god, possibly representing a specific river. The figure is semi-nude and holds a cornucopia overflowing with water and various symbols. The inscription around the image, often describing the scene or the emperor's achievements, is visible on both sides of the coin. The coin's detailed artwork and inscriptions provide insights into the political and cultural aspirations of the Roman Empire and its leaders, policies, and artistic expressions. Coins like this were not only used for commerce but also served as means of propaganda, promoting the emperor's achievements and divine favor.	A detailed description of the image from Gemini: This image shows two sides of an ancient Roman coin. The coin appears to be made of bronze or a similar alloy, given its color and texture. The obverse side depicts a profile of the emperor Trajan, facing right. The reverse side depicts another profile, which likely belongs to the emperor himself, as it is common in Roman coinage. The inscription around the figure is in Latin and includes recognizable components ("IMP" for Imperator, "CAES" for Caesar, and "AVGVSTVS" for Augustus). The reverse displays a seated figure, possibly a river god or a personification of a river, holding a cornucopia overflowing with water and various symbols. The inscription around the figure and the reverse describes the emperor's reign and his victories, such as "AVGVSTVS AVG PIVS FEL AVGVSTVS" on the obverse and "P M TR P COS V P P" on the reverse.	The image presents two sides of an ancient Roman coin, each with its own unique design and inscription. The coin on the left is a Roman coin, identifiable by the Latin text cast around the edge. The inscription on the left side reads "AVGVSTVS AVG PIVS FEL AVGVSTVS," which translates to "Emperor Caesar, son of the invincible, Augustus." The reverse side shows a seated profile of a deity, possibly a river god, holding a cornucopia. The individual has a beard and wears a laurel wreath, which symbolizes victory and honor in Roman culture. The inscription on the reverse side includes "P M TR P COS V P P," which stands for "Pontifex Maximus, Tribune of the People, Consul for the Year." The coin on the right is a Greek coin, as indicated by the Greek text cast around the edge. The inscription on the left side of the Greek coin reads "TRAIANO AVG GER DAC P M TR P COS V P P," which translates to "Emperor Trajan, son of the invincible, Augustus, Pontifex Maximus, Tribune of the People, Consul for the Year." The reverse is inscribed with "S P Q R OPTIMO QVOD ET LONGA VITAE AVGVSTVS AVG PIVS C." This inscription suggests that the coin was minted during the reign of Emperor Trajan, who was known for his engineering projects and military conquests, particularly the building of the bridge across the Tigris River.	This Roman imperial copper alloy coin, produced between 104-111 CE, features a finely crafted bust of Emperor Trajan, laurele, facing right, with drapery wrapped around his shoulder. The reverse side depicts a river-god reclined on rocks from which water gushes, symbolized through an arched grotto supported by two columns. The coin rests on an all-hands clutch, suggesting his left on an armrest, while inscriptions grace both sides of this coin; the obverse is inscribed with "S P Q R OPTIMO QVOD ET LONGA VITAE AVGVSTVS AVG PIVS C" in the extenso. This coin not only reflects the artistic and cultural nuances of the Roman Empire but also emphasizes Trajan's engineering prowess and his successful campaigns in the east, though its precise findspot remains unknown, it is a testament to the widespread influence of Roman currency across the empire.		
	The image shows two sides of an ancient coin, possibly a denarius, from the Roman Empire. The coin is irregular in shape, indicating it might have been minted in ancient times when coinage technology was not as refined. The obverse (left side) shows a profile of the emperor Trajan, facing right. The reverse (right side) also shows a profile, likely of the river god Tiberinus. The coin is heavily worn and shows signs of significant age, suggesting it's been circulating for a long time. The details in the portrait are somewhat faded, but the overall composition follows the high artistic standards of Roman coinage. Roman coins like this were not only used for commerce but also served as means of propaganda, communicating the emperor's image and accomplishments to the public. The inscription around the edge reads "IMP TRAIANVS AVG PIVS FEL AVGVSTVS," which translates to "Emperor Caesar, son of the invincible, Augustus." The reverse side also shows a seated figure, possibly a river god or a personification of water, holding a cornucopia. The inscription around the figure and the reverse describes the emperor's reign and his achievements of Roman coinage.	A detailed description of the image from GPT-4o: This image shows two sides of an ancient coin, possibly a denarius, from the Roman Empire. The coin is irregular in shape, indicating it might have been minted in ancient times when coinage technology was not as refined. The obverse (left side) shows a profile of the emperor Trajan, facing right. The reverse (right side) also shows a profile, likely of the river god Tiberinus. The coin is heavily worn and shows signs of significant age, suggesting it's been circulating for a long time. The details in the portrait are somewhat faded, but the overall composition follows the high artistic standards of Roman coinage. Roman coins like this were not only used for commerce but also served as means of propaganda, communicating the emperor's image and accomplishments to the public. The inscription around the edge reads "IMP TRAIANVS AVG PIVS FEL AVGVSTVS," which translates to "Emperor Caesar, son of the invincible, Augustus." The reverse side also shows a seated figure, possibly a river god or a personification of water, holding a cornucopia. The inscription around the figure and the reverse describes the emperor's reign and his achievements of Roman coinage.	A detailed description of the image from Gemini: This image shows two sides of an ancient coin, possibly a denarius, from the Roman Empire. The coin is irregular in shape, indicating it might have been minted in ancient times when coinage technology was not as refined. The obverse (left side) shows a profile of the emperor Trajan, facing right. The reverse (right side) also shows a profile, likely of the river god Tiberinus. The coin is heavily worn and shows signs of significant age, suggesting it's been circulating for a long time. The details in the portrait are somewhat faded, but the overall composition follows the high artistic standards of Roman coinage. Roman coins like this were not only used for commerce but also served as means of propaganda, communicating the emperor's image and accomplishments to the public. The inscription around the edge reads "IMP TRAIANVS AVG PIVS FEL AVGVSTVS," which translates to "Emperor Caesar, son of the invincible, Augustus." The reverse side also shows a seated figure, possibly a river god or a personification of water, holding a cornucopia. The inscription around the figure and the reverse describes the emperor's reign and his achievements of Roman coinage.	The image depicts two ancient coins, both from the Roman Empire, showing a profile of the emperor Trajan on the left and the river god Tiberinus on the right. The coin on the left is a bronze coin, featuring a profile of Trajan facing right, with a laurel wreath and various inscriptions. The coin on the right is a silver coin, featuring a profile of Tiberinus facing left, holding a cornucopia and a laurel wreath. Both coins show significant wear and damage. The inscription on the left side of the coin reads "IMP TRAIANVS AVG PIVS FEL AVGVSTVS," and the reverse side shows "P M TR P COS V P P." The inscription on the right side of the coin reads "TIBERINVS AVG PIVS FEL AVGVSTVS," and the reverse side shows "P M TR P COS V P P."	This copper coin, originating from the Gupta dynasty between 3rd-5th CE, depicts a striking depiction of Chandragupta II. The king is portrayed standing and facing left, adorned with a waistcloth and jewelry, with his right hand raised in a greeting salutation and his left hand resting on his hip. The coin features a distinct central design on both sides, with the obverse showing the king's profile and the reverse showing the same profile in a different pose. The king's attire is highly detailed, with a dark greenish-blue patina, indicative of age and wear, with some areas that are smoother, possibly due to natural wear or intentional craftsmanship. The surface of the coin is rough and textured, with raised edges and recessed centers. The design is detailed, with the king's attire and the throne's ornamentation clearly visible. The inscription on the coin is in Brahmi script.		
	The image displays an artifact that appears to be an ancient sculpture. It is a small, seated figure, possibly of stone or ceramic material, depicting a figure in flowing robes, with detailed folds suggesting movement and grace, typical of traditional representations of deities or historical figures. The figure is adorned with what seems to be a headaddress or topknot, indicating a figure of significance, possibly a deity or a noble. The figure has a serene expression and is seated in a meditative or prayerful pose. The hands are positioned in front, possibly in a mudra or gesture, with fingers interlaced. The face has a gentle, contemplative expression, with a slight smile. The figure is wearing minimal clothing, consisting of a skirt or loincloth. The sculpture is in good condition, with some minor wear and discoloration. The background is plain, allowing the figure to stand out. The overall style is reminiscent of East Asian cultural art forms, specifically from Chinese Buddhist or Taoist traditions. The figure is depicted as a Bodhisattva, often depicted as bodhisattvas or sages. In the surface exhibits a weathered texture, indicating its age, and possibly suggesting it was once part of a larger structure. Its condition speaks to the passage of time and the enduring legacy of civilizations long past. In general, it's a well-preserved example of ancient craftsmanship.	A detailed description of the image from GPT-4o: This image displays an artifact that appears to be an ancient sculpture. It is a small, seated figure, possibly of stone or ceramic material, depicting a figure in flowing robes, with detailed folds suggesting movement and grace, typical of traditional representations of deities or historical figures. The figure is adorned with what seems to be a headaddress or topknot, indicating a figure of significance, possibly a deity or a noble. The figure has a serene expression and is seated in a meditative or prayerful pose. The hands are positioned in front, possibly in a mudra or gesture, with fingers interlaced. The face has a gentle, contemplative expression, with a slight smile. The figure is wearing minimal clothing, consisting of a skirt or loincloth. The sculpture is in good condition, with some minor wear and discoloration. The background is plain, allowing the figure to stand out. The overall style is reminiscent of East Asian cultural art forms, specifically from Chinese Buddhist or Taoist traditions. The figure is depicted as a Bodhisattva, often depicted as bodhisattvas or sages. In the surface exhibits a weathered texture, indicating its age, and possibly suggesting it was once part of a larger structure. Its condition speaks to the passage of time and the enduring legacy of civilizations long past. In general, it's a well-preserved example of ancient craftsmanship.	A detailed description of the image from Gemini: This image shows a small, seated ancient ceramic figure, likely a Bodhisattva or a deity. The figure is seated in a meditative pose with hands resting on their lap, and is a common form for depictions of deities. The face has a serene expression and is adorned with stylized hair and curves. The body is composed of a large central torso with legs crossed and hands resting in the lap. The figure is made of a light-colored material, possibly clay or stone. The figure is in good condition, with some minor wear and discoloration. The background is plain, allowing the figure to stand out. The overall style is reminiscent of traditional East Asian ceramic art, particularly from China or Japan.	The image depicts an ancient ceramic figurine, likely a Bodhisattva or a deity. The figure is seated in a meditative pose with hands resting on their lap, and is a common form for depictions of deities. The face has a serene expression and is adorned with stylized hair and curves. The body is composed of a large central torso with legs crossed and hands resting in the lap. The figure is made of a light-colored material, possibly clay or stone. The figure is in good condition, with some minor wear and discoloration. The background is plain, allowing the figure to stand out. The overall style is reminiscent of traditional East Asian ceramic art, particularly from China or Japan.	A detailed description of the image from Qwen: This image shows a small, seated ancient ceramic figure, possibly a Bodhisattva or a deity. The figure is seated in a meditative pose with hands resting on their lap, and is a common form for depictions of deities. The face has a serene expression and is adorned with stylized hair and curves. The body is composed of a large central torso with legs crossed and hands resting in the lap. The figure is made of a light-colored material, possibly clay or stone. The figure is in good condition, with some minor wear and discoloration. The background is plain, allowing the figure to stand out. The overall style is reminiscent of traditional East Asian ceramic art, particularly from China or Japan.	The image depicts two ancient coins, both from the Roman Empire. The coin on the left is a bronze coin, featuring a profile of Trajan facing right, with a laurel wreath and various inscriptions. The coin on the right is a silver coin, featuring a profile of Tiberinus facing left, holding a cornucopia and a laurel wreath. Both coins show significant wear and damage. The inscription on the left side of the coin reads "IMP TRAIANVS AVG PIVS FEL AVGVSTVS," and the reverse side shows "P M TR P COS V P P." The inscription on the right side of the coin reads "TIBERINVS AVG PIVS FEL AVGVSTVS," and the reverse side shows "P M TR P COS V P P."	This copper coin, originating from the Gupta dynasty between 3rd-5th CE, depicts a striking depiction of Chandragupta II. The king is portrayed standing and facing left, adorned with a waistcloth and jewelry, with his right hand raised in a greeting salutation and his left hand resting on his hip. The coin features a distinct central design on both sides, with the obverse showing the king's profile and the reverse showing the same profile in a different pose. The king's attire is highly detailed, with a dark greenish-blue patina, indicative of age and wear, with some areas that are smoother, possibly due to natural wear or intentional craftsmanship. The surface of the coin is rough and textured, with raised edges and recessed centers. The design is detailed, with the king's attire and the throne's ornamentation clearly visible. The inscription on the coin is in Brahmi script.
	The image depicts an intricately crafted gold neck ring, often associated with ancient Nordic cultures. This type of artifact is typically known as a torc or bangle. It is made of a single, continuous gold wire twisted into a complex, coiled pattern, resembling a thick, ornate bracelet. The torc is set against a dark blue background. The intricate pattern, showcasing skilled metalwork common in ancient jewelry. The torc is designed to be worn around the neck, with two handles, and is often used in Norse mythology. It is worn as symbols of status, wealth, or power in ancient societies, reflecting the sophisticated craftsmanship and high value of the metal. The gold material indicates that this could have been owned by someone of high status or tranquility.	A detailed description of the image from GPT-4o: The image depicts an intricately crafted gold neck ring, often associated with ancient Nordic cultures. This type of artifact is typically known as a torc or bangle. It is made of a single, continuous gold wire twisted into a complex, coiled pattern, resembling a thick, ornate bracelet. The torc is set against a dark blue background. The intricate pattern, showcasing skilled metalwork common in ancient jewelry. The torc is designed to be worn around the neck, with two handles, and is often used in Norse mythology. It is worn as symbols of status, wealth, or power in ancient societies, reflecting the sophisticated craftsmanship and high value of the metal. The gold material indicates that this could have been owned by someone of high status or tranquility.	A detailed description of the image from Gemini: The image depicts an intricately crafted gold neck ring, often associated with ancient Nordic cultures. This type of artifact is typically known as a torc or bangle. It is made of a single, continuous gold wire twisted into a complex, coiled pattern, resembling a thick, ornate bracelet. The torc is set against a dark blue background. The intricate pattern, showcasing skilled metalwork common in ancient jewelry. The torc is designed to be worn around the neck, with two handles, and is often used in Norse mythology. It is worn as symbols of status, wealth, or power in ancient societies, reflecting the sophisticated craftsmanship and high value of the metal. The gold material indicates that this could have been owned by someone of high status or tranquility.	A detailed description of the image from Qwen: The image depicts an intricately crafted gold neck ring, often associated with ancient Nordic cultures. This type of artifact is typically known as a torc or bangle. It is made of a single, continuous gold wire twisted into a complex, coiled pattern, resembling a thick, ornate bracelet. The torc is set against a dark blue background. The intricate pattern, showcasing skilled metalwork common in ancient jewelry. The torc is designed to be worn around the neck, with two handles, and is often used in Norse mythology. It is worn as symbols of status, wealth, or power in ancient societies, reflecting the sophisticated craftsmanship and high value of the metal. The gold material indicates that this could have been owned by someone of high status or tranquility.	The image depicts a gold bracelet with a distinctive design, reminiscent of ancient jewelry. The bracelet features a large, central decorative element, possibly a knot or a loop, which is surrounded by a series of smaller, circular motifs, possibly beads or ornaments. The bracelet is made of a thin, flexible metal wire, coiled and twisted to form the intricate pattern. It is set against a dark, neutral background. The design is symmetrical and balanced, with the central element being the focal point. The bracelet is made from a precious metal, likely gold, given its color and luster. The surface of the bracelet shows signs of age and wear, with some minor scratches and discoloration. The background is plain, allowing the bracelet to stand out. The overall style is minimalist and elegant, characteristic of ancient craftsmanship.	A detailed description of the image from Llava-Next: The image depicts a gold bracelet with a distinctive design, reminiscent of ancient jewelry. The bracelet features a large, central decorative element, possibly a knot or a loop, which is surrounded by a series of smaller, circular motifs, possibly beads or ornaments. The bracelet is made of a thin, flexible metal wire, coiled and twisted to form the intricate pattern. It is set against a dark, neutral background. The design is symmetrical and balanced, with the central element being the focal point. The bracelet is made from a precious metal, likely gold, given its color and luster. The surface of the bracelet shows signs of age and wear, with some minor scratches and discoloration. The background is plain, allowing the bracelet to stand out. The overall style is minimalist and elegant, characteristic of ancient craftsmanship.	A detailed description of the image from Ground-Truth: This Viking necklace, crafted between the 10th and 11th centuries, exemplifies the artistry and metalworking skills of the era. Composed of twisted silver wire, the necklace consists of two elongate, expanding terminals that are beaten flat and intricately ornamented with rows of diamond-shaped patterns and small circular holes, concluding in a functional hook. Such necklaces were likely symbols of wealth and status within Viking society, reflecting their religious and social preferences and technical prowess. This particular artifact was unearthed as part of the Haderup hoard, discovered in Denmark in the Viking era, providing valuable insights into Viking presence and influence in the region.
	This image features an ancient bronze vessel known as a Ding. The Ding is a tripod cauldron with a lid and three legs, often used in Chinese culture. The Ding is made of a dark metallic sheen, with a patina that hints at its antique quality. The handle on the lid is rounded and features a decorative pattern of swirling motifs and possible depictions of mythical creatures, which are common in ancient Chinese bronze vessels. The Ding is used for cooking and serving food, and its design elements suggest an artifact from a culture where personal adornment was not just decorative, but also carried deep social, religious, and political meaning. The gold material indicates that this could have been owned by someone of high status or tranquility.	A detailed description of the image from GPT-4o: This image features an ancient bronze vessel known as a Ding. The Ding is a tripod cauldron with a lid and three legs, often used in Chinese culture. The Ding is made of a dark metallic sheen, with a patina that hints at its antique quality. The handle on the lid is rounded and features a decorative pattern of swirling motifs and possible depictions of mythical creatures, which are common in ancient Chinese bronze vessels. The Ding is used for cooking and serving food, and its design elements suggest an artifact from a culture where personal adornment was not just decorative, but also carried deep social, religious, and political meaning.	A detailed description of the image from Gemini: This image features an ancient bronze vessel known as a Ding. The Ding is a tripod cauldron with a lid and three legs, often used in Chinese culture. The Ding is made of a dark metallic sheen, with a patina that hints at its antique quality. The handle on the lid is rounded and features a decorative pattern of swirling motifs and possible depictions of mythical creatures, which are common in ancient Chinese bronze vessels. The Ding is used for cooking and serving food, and its design elements suggest an artifact from a culture where personal adornment was not just decorative, but also carried deep social, religious, and political meaning.	A detailed description of the image from Qwen: This image depicts an ancient bronze vessel known as a Ding. The Ding is a tripod cauldron with a lid and three legs, often used in Chinese culture. The Ding is made of a dark metallic sheen, with a patina that hints at its antique quality. The handle on the lid is rounded and features a decorative pattern of swirling motifs and possible depictions of mythical creatures, which are common in ancient Chinese bronze vessels. The Ding is used for cooking and serving food, and its design elements suggest an artifact from a culture where personal adornment was not just decorative, but also carried deep social, religious, and political meaning.	A detailed description of the image from Llava-Next: This image depicts an ancient bronze vessel known as a Ding. The Ding is a tripod cauldron with a lid and three legs, often used in Chinese culture. The Ding is made of a dark metallic sheen, with a patina that hints at its antique quality. The handle on the lid is rounded and features a decorative pattern of swirling motifs and possible depictions of mythical creatures, which are common in ancient Chinese bronze vessels. The Ding is used for cooking and serving food, and its design elements suggest an artifact from a culture where personal adornment was not just decorative, but also carried deep social, religious, and political meaning.	A detailed description of the image from Ground-Truth: This bronze vessel, crafted in the form of a lamp with three legs and a lid, is one of the most famous examples of ancient Chinese bronze art. It is made of a dark metallic sheen and features intricate patterns etched into its surface. The Qajar dynasty in Iran influenced the design of this vessel, particularly the lid, which is shaped like a shallow bowl supported by three legs. The body of the vessel is cylindrical and has a flared base. The entire object is highly detailed, with intricate patterns and designs. The lid is decorated with a repeating geometric pattern of circles and squares. The body of the vessel is decorated with a more organic, swirling pattern of leaves and vines. The base of the vessel is decorated with a band of small, repeating characters. The overall design is highly sophisticated, reflecting the remarkable craftsmanship and cultural significance of ancient Chinese bronze vessels. This vessel is considered a masterpiece of ancient Chinese metalwork and is now housed in the National Museum of China in Beijing.	

Figure 8: Cross-model comparison of generated descriptions for TimeTravel dataset samples, highlighting variations in detail and accuracy. It illustrates differences in descriptive depth across open- and closed-source models, emphasizing the diversity in interpretative approaches and alignment with the ground truth.