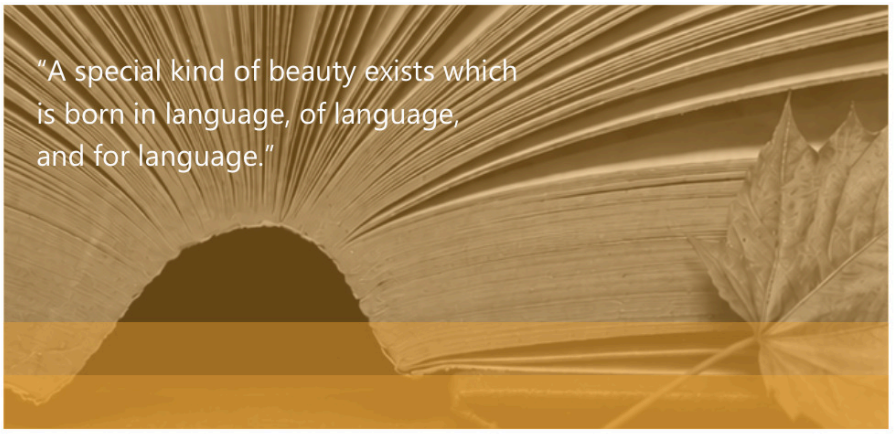
VOLUME 10, OCTOBER 2026
ISSN: 3060-4958







The Lingua Spectrum



UZBEKISTAN STATE WORLD
LANGUAGES UNIVERSITY

MORE INFO:

www.lingvospektr.uz

t.me/thelinguaspectrum

Jalolova G.M.	Structural-semantic classification and etymological layers of astrological vocabulary in English and Uzbek	146
Abduxakimova G.	A Lexico-Semantic Comparative Analysis of Euphemisms Related to the Body, Health and Private Life in Contemporary English and Uzbek Media Discourse (Social Media, Online News and Blogs)	153
Dalieva M.Kh.	Lexical and pragmatic features of English-language media headlines: a corpus study of world health organization news releases	161
Baynazarov Z.M.	Language policy in the Timurid period and the status of the Turkic language	169
Im S.B.	Gallicisms and the Ways They Entered the Russian Language	181
Im S.B.	Karamzin as the Creator of the French Culture of Sensibility in the Russian Language	189
Siddikova F.O.	The representation of hospitality in English and Uzbek: A Comparative Linguocultural Analysis	195
Sadinova M.A.	Binary conceptual oppositions in the linguistic worldview: The life – death model	204
Umarova J.Zh.	Evaluative Language and Metaphor in Museum Discourse: Assessing AI-Based Analysis of the Savitsky Museum Catalogue	228
Xojanov Sh.B.	The study of phraseological units based on modern linguistic theories in K. Karimov's novel "Agabiy"	239
Azimova D.A.	The System of Non-Canonical Antonym Pairs in English Presidential Speeches and Proverbs (Non-Canonical Antonyms as Powerful Rhetorical Instruments)	246
Nasirova R.A.	Phraseological and Paremiological Conceptualization of Moral Values in English and Uzbek	254
Kambarova M.A.	Universal and National-Specific Features in English and Russian Proverbs (based on Proverbs about Homeland)	260
Djalilov M.M.	Neologisms and abbreviations: the interaction of word-formation processes in modern linguistics	267
Sabirova B.U.	The lexical-semantic analysis of Uzbek pragmalinguistic terms "lingvistik kontekst, vaziyat konteksti, ijtimoiy konteks va kognitiv kontekst"	273
Karimova Z.G'.	Illocutionary features of modal and interjection word-sentences in typologically different languages	282
Borisova Ye.V.	Linguistic representation of blame and responsibility in public corporate apologies	289
Yahyaeva E.V.	Vocative words and vocative sentences in Azerbaijani	301
Musayeva S.	Structural and semantic features of phraseological units in the form of word combinations and sentences in Azerbaijani	310

The Lingua Spectrum
Cogito, ergo sum

Jurnal Oliy Attestatsiya Komissiyasi tomonidan 2025-yil 4-yanvar kuni 366/5-sonli qaror bilan tavsiya etilgan milliy ilmiy nashrlar ro'yxatiga kiritilgan

Evaluative Language and Metaphor in Museum Discourse: Assessing AI-Based Analysis of the Savitsky Museum Catalogue

Umarova Jazira Zhalgasbai kizi
zhaziraumarova2001@gmail.com
Master's Degree Student,
Karakalpak State University named after Berdakh

Annotation *Museum catalogue texts do more than describe works of art; they also express judgements about artistic value and use figurative language to guide interpretation. This study focuses on evaluative language and metaphor in selected English-language descriptions from the I.V. Savitsky State Museum of Art of the Republic of Karakalpakstan. Four catalogue entries are examined: Caravan, Northern Lights, Houses, and The Old and the New. The analysis considers how individual expressions function in context and how evaluative meaning can overlap with metaphorical or symbolic interpretation. ChatGPT is used as an analytical tool to identify relevant expressions and explain their possible meanings. Its interpretations need to be checked against the original catalogue texts, since a plausible explanation does not necessarily mean that a classification is linguistically justified. The study also considers the difficulty of distinguishing figurative language from literal description, particularly when a phrase performs more than one function. By focusing on specific examples from museum discourse, the research explores the usefulness of AI-assisted linguistic analysis while recognising the continuing importance of careful human interpretation.*

Keywords *Museum discourse, evaluative language, metaphor, ChatGPT, AI-assisted analysis, Savitsky Museum*

Muzey diskursida baholovchi til vositalari va metafora: Savitskiy muzeyi katalogining sun'iy intellekt yordamida tahlil qilinishini baholash

Umarova Jazira Zhalgasbai qizi
zhaziraumarova2001@gmail.com
Magistrant,
Berdaq nomidagi Qoraqalpoq Davlat Universiteti

Annotatsiya *Muzey kataloglari matnlari nafaqat san'at asarlarini tasvirlaydi, balki ularning badiiy qiymatiga baho beradi va talqinni yo'naltirish uchun obrazli tildan ham foydalanadi. Ushbu tadqiqot Qoraqalpog'iston Respublikasi I.V. Savitskiy nomidagi Davlat san'at muzeyining ingliz tilidagi ayrim tavsiflarida baholovchi til vositalari va metaforalarni o'rganishga bag'ishlangan. Tadqiqot doirasida to'rtta katalog maqolasi tahlil qilinadi: Caravan, Northern Lights, Houses va The Old and the New. Tahlilda alohida ifodalarning kontekstdagi vazifasi hamda baholovchi ma'noning metaforik yoki ramziy talqin bilan qanday uyg'unlashishi ko'rib chiqiladi. ChatGPT tegishli ifodalarni aniqlash va ularning mumkin bo'lgan ma'nolarini tushuntirish uchun tahliliy vosita sifatida qo'llanadi. Model talqinlarini katalogning asl matnlari bilan tekshirish zarur, chunki ishonarli izoh har doim ham tasnifning lingvistik jihatdan asoslanganligini anglatmaydi. Tadqiqotda, ayniqsa, bitta ifoda bir nechta*

vazifani bajargan hollarda, obrazli tilni so‘zma-so‘z tavsifdan ajratish qiyinligi ham ko‘rib chiqiladi. Muzey diskursidan olingan aniq misollar asosida tadqiqot lingvistik tahlilda sun’iy intellektdan foydalanish imkoniyatlarini yoritadi va inson tomonidan sinchkovlik bilan talqin qilish muhimligini ta’kidlaydi.

Kalit so‘zlar *Muzey diskursida baholovchi til vositalari va metafora: Savitskiy muzeyi katalogining sun’iy intellekt yordamida tahlil qilinishini baholash*

Оценочная лексика и метафора в музейном дискурсе: оценка анализа каталога музея Савицкого с помощью искусственного интеллекта

Умарова Жазира Жалгасбай кызы
zhaziraumarova2001@gmail.com
Магистрант,
Каракалпакский Государственный Университет имени Бердаха

Аннотация *Тексты музейных каталогов не только описывают произведения искусства, но и выражают оценки их художественной ценности, а также используют образный язык, направляющий интерпретацию. Настоящее исследование посвящено оценочной лексике и метафоре в отдельных англоязычных описаниях из Государственного музея искусств имени И. В. Савицкого Республики Каракалпакстан. Рассматриваются четыре каталожные статьи: «Караван», «Северное сияние», «Дома» и «Старое и новое». Анализ направлен на изучение функционирования отдельных выражений в контексте, а также на выявление взаимосвязи оценочного значения с метафорической или символической интерпретацией. ChatGPT используется в качестве аналитического инструмента для выявления соответствующих выражений и объяснения их возможных значений. Интерпретации модели необходимо проверять по оригинальным текстам каталога, поскольку правдоподобное объяснение не всегда означает, что классификация лингвистически обоснована. В исследовании также рассматривается сложность разграничения образного языка и буквального описания, особенно в случаях, когда одно выражение выполняет несколько функций. На примере конкретных фрагментов музейного дискурса исследование раскрывает возможности применения искусственного интеллекта в лингвистическом анализе, подчёркивая сохраняющуюся важность внимательной интерпретации человеком.*

Ключевые слова *Музейный дискурс, оценочная лексика, метафора, ChatGPT, анализ с использованием искусственного интеллекта, музей Савицкого*

Museum catalogue descriptions do more than provide factual information about artworks. Through evaluative words and metaphorical expressions, they highlight artistic features, convey emotional and cultural significance, and suggest particular

interpretations. These linguistic choices can shape readers' perceptions of a painting even before they examine it in detail. This makes museum discourse a relevant area of linguistic research, especially given the interpretive role of relatively short texts.

One important aspect of museum communication is evaluative language, through which words and expressions convey judgments about an artwork's artistic qualities, emotional impact, or cultural significance (Lazzeretti, 2025). Another relevant feature is metaphor, which connects visible elements of an artwork with abstract concepts or experiences. Ntzani (2015) examines how conduit and container metaphors shape the museum experience. In the present study, these concepts provide a basis for analysing how evaluative and metaphorical meanings interact in museum catalogue descriptions and how their interpretation depends on context.

The growing use of artificial intelligence has created new opportunities for linguistic analysis, including the classification and interpretation of textual expressions. However, ChatGPT may produce plausible interpretations that are not explicitly supported by the source text, making human verification essential. Research on large language models for qualitative coding has also raised questions about the relationship between automated classification and human judgement (Zhang et al., 2024). This limitation is particularly relevant when analysing evaluative language and metaphor, as their interpretation depends on the context in which they appear.

The English-language catalogue of the I.V. Savitsky State Museum of Art of the Republic of Karakalpakstan provides a context for examining how AI analyses museum discourse. Its descriptions combine factual information, evaluative statements, and interpretive language. The present study examines how ChatGPT identifies and interprets evaluative language and metaphorical expressions in selected catalogue descriptions, with particular attention to the distinction between evaluation, metaphor, symbolic interpretation, and literal description. It also considers how ChatGPT explains its classifications and whether its interpretations are supported by the source texts

The study addresses three research questions:

1. How consistently does ChatGPT identify evaluative language in the selected museum descriptions?
2. How does ChatGPT distinguish metaphorical expressions from symbolic interpretations and literal descriptions?
3. What types of agreement and disagreement emerge between ChatGPT-generated analysis and manual linguistic annotation?

By comparing AI-generated analysis with manual annotation, the study examines the potential and limitations of ChatGPT in analysing evaluative and metaphorical language in museum discourse, particularly where interpretation requires contextual judgement.

Literature review

Evaluative Language in Museum Communication. Museum descriptions combine factual information with language that conveys judgements about an artwork's artistic quality, originality, emotional impact, and cultural significance. Evaluative language therefore provides a way to examine how museum texts communicate not only information but also assessments of artistic value. Lazzeretti (2025) examines evaluation across different forms of museum communication, drawing on discourse analysis, corpus linguistics, and multimodal analysis. She identifies evaluative parameters such as novelty, quality, quantity, exclusivity, and newsworthiness. This framework is relevant to the present study because evaluation may be expressed not only through explicitly emotional adjectives but also through statements about an artist's originality, a work's importance, or its place within an artistic tradition. Identifying such expressions

therefore requires attention to their contextual function rather than reliance on a predetermined list of evaluative words. In this study, evaluative language is distinguished from relatively neutral description according to whether an expression conveys a judgement about an artwork rather than merely identifies its observable features.

Metaphor and Interpretation in Museum Discourse Metaphor offers a way to examine how museum descriptions connect visible features of artworks with more abstract meanings. According to conceptual metaphor theory, abstract concepts can be understood through more concrete experiences and conceptual structures (Lakoff & Johnson, 1980). In museum discourse, such connections may influence how readers interpret particular visual elements. Ntzani (2015) examines conduit and container metaphors in museum communication, highlighting the role of metaphor in shaping relationships between museum objects, spaces, and visitors. Although her focus differs from the analysis of individual expressions in catalogue descriptions, her research provides a useful theoretical basis for treating metaphor as a means of constructing meaning rather than merely a stylistic feature. For the present study, metaphorical language must be distinguished from symbolic interpretation. A statement that an object symbolises an abstract concept does not necessarily contain a metaphorical expression, while figurative meaning may emerge from context without an explicit metaphor marker. The analysis therefore considers both the wording of candidate expressions and the explanations provided by ChatGPT, assessing whether proposed classifications are supported by the source text. Where the distinction remains uncertain, the interpretation is treated as open to discussion rather than assigned to a single category without sufficient evidence.

Large Language Models and Qualitative Coding Large language models offer new possibilities for analysing textual data by identifying patterns, assigning categories, and explaining classification decisions. However, plausible explanations do not necessarily guarantee accurate or consistent classifications. Zhang et al. (2024) explore AI-assisted qualitative coding through QualiGPT, examining how language-model-based coding compares with traditional manual analysis. Their work provides a methodological reference for the present study, which evaluates ChatGPT-generated classifications against manual linguistic annotation. Qualitative coding in museum discourse requires attention to context, as evaluative and metaphorical meanings may be implicit or open to interpretation. ChatGPT may overlook relevant expressions, misclassify them, or produce explanations that go beyond the evidence in the source text. The present study therefore examines agreement and disagreement between AI-generated analysis and manual annotation, considering omissions, unsupported interpretations, and differences in category assignment without automatically treating every disagreement as proof of error.

Research Gap and Focus of the Present Study

Taken together, previous research provides a basis for examining how ChatGPT identifies and interprets evaluative language and metaphor in English-language museum catalogue descriptions, particularly where classification requires contextual judgement.

Materials and Methods

Research Design

The study adopts an exploratory qualitative design combining close linguistic analysis with a comparison of ChatGPT-generated classifications and manual annotation. The analysis draws on Martin and White's (2005) Appraisal framework to examine evaluative language and the Metaphor Identification Procedure proposed by the Pragglejaz Group (2007) to identify metaphorical expressions. Saldaña (2021) provides a broader methodological reference for making qualitative coding decisions explicit and examining them in relation to the data.

Research Material and Corpus Selection

The research material consists of selected English-language descriptions of artworks from the I.V. Savitsky State Museum of Art of the Republic of Karakalpakstan. Four entries were selected for close analysis: Alexander Volkov's *Caravan* (1926), Alexey Kuprikov's *Northern Lights* (1914), Liubov Popova's *Houses* (1916), and Solomon Nikritin's *The Old and the New* (1935). Purposive sampling was used to select descriptions offering opportunities to examine evaluative language, metaphorical expressions, and the relationship between visual description and interpretation across different artworks. The unit of analysis is the English-language descriptive passage associated with each artwork; titles, dates, dimensions, and medium are excluded unless they form part of an evaluative or figurative expression.

Analytical Categories and Annotation Criteria

The analysis distinguishes three principal categories: evaluative language, metaphorical expressions, and literal or primarily symbolic descriptions. Evaluative language refers to expressions that convey judgements about an artwork's quality, originality, significance, emotional impact, or cultural value. Metaphorical expressions are identified by examining whether a contextual meaning contrasts with a more basic meaning, following the Metaphor Identification Procedure of the Pragglejaz Group (2007). Literal descriptions primarily identify visible features or actions, while symbolic interpretations connect them with broader cultural, emotional, or philosophical meanings.

Each candidate expression is recorded with its exact wording, proposed category, contextual justification, and confidence label. High confidence indicates strong textual support, medium confidence reflects plausible alternative interpretations, and low confidence indicates that the classification depends on assumptions not clearly established by the passage. These labels represent the clarity of individual analytical decisions rather than statistical probabilities.

ChatGPT-Based Analysis

ChatGPT (GPT-6) was used to analyse the four selected catalogue descriptions on 18 October 2026. Each passage was processed using the same instructions, which required the model to identify evaluative expressions and possible metaphors, explain its classifications, distinguish textual evidence from interpretation, and flag uncertain cases. The model was instructed to quote the source text exactly, avoid external information, and not introduce examples absent from the supplied passage. The responses were assessed at the level of individual expressions according to three criteria: identification of relevant expressions, linguistic justification of their classification, and consistency with the evidence in the source text. Omissions, unsupported interpretations, and uncertain classifications were also recorded. The analysis considers only the outputs obtained under these conditions and does not establish reproducibility across independent sessions or model versions, nor does it support generalisations about all ChatGPT versions or large language models.

Manual Annotation and Comparative Procedure

Manual annotation was planned using the same catalogue passages and principal categories as the ChatGPT-based analysis. The comparison examines candidate expressions, category assignments, and the textual support provided for each interpretation. Disagreements are distinguished as differences in expression identification, classification, or interpretation; borderline cases are retained for contextual discussion rather than automatically treated as evidence of model error. Manual annotation serves as a reference for comparison, not as an unquestionable ground truth.

Data Organisation and Interpretation

The findings are organised around examples relevant to the research questions,

focusing on the identification, classification, and interpretation of evaluative and metaphorical expressions. The analysis uses qualitative comparison to examine agreement, disagreement, and borderline cases, with interpretations linked to the wording of the source texts. The conclusions are limited to the selected descriptions and the specific analytical conditions of the study.

Results

The selected catalogue descriptions contain evaluative expressions that vary in their linguistic form and function. In *Caravan* (1926), "a tremendous influence" conveys a strong assessment of artistic significance, while "an endless quest for a new artistic language" presents artistic development as a continuing search. In *Northern Lights* (1914), "a celestial explosion of colours" combines figurative language with an evaluative effect. *Houses* (1916) includes descriptions of visual form alongside theoretical interpretation, while *The Old and the New* (1935) contains the comparative judgement "the most advanced museum." These examples provide a basis for examining how ChatGPT distinguishes evaluation from literal description and metaphor from other forms of interpretation.

Identification of Evaluative Language

Evaluative language in the selected descriptions conveys judgements about artistic significance, originality, and visual impact. In *Caravan*, "a tremendous influence on his fellow artists in Central Asia" evaluates the artist's historical importance, while "an endless quest for a new artistic language" frames artistic development as a continuing search for innovation. In *Northern Lights*, "extraordinary natural phenomenon" and "fantastic landscape" express positive assessments, whereas "a celestial explosion of colours" combines an evaluative effect with figurative wording. In *The Old and the New*, "the most advanced museum of the 1920s" presents an explicit comparative judgement.

By contrast, expressions in *Houses* such as "black and white crystals" and "an irregular square" primarily describe visual features rather than assess artistic quality. The comparison between these examples allows the analysis to examine whether ChatGPT distinguishes evaluative language from relatively neutral description based on contextual function rather than grammatical form alone. Its actual classifications and explanations must be reported from the responses obtained for these passages.

Metaphor, Symbolism, and Literal Description

The selected descriptions contain expressions that require contextual analysis to distinguish metaphorical meaning from literal description and symbolic interpretation. In *Caravan*, "an endless quest for a new artistic language" represents artistic development as a search, while "the caravan drivers are preparing and loading packs onto the camels" describes an action depicted in the painting. The former is a candidate for metaphorical classification because its contextual meaning extends beyond a literal journey. In *Northern Lights*, "a celestial explosion of colours" describes visual intensity through the image of an explosion, combining figurative wording with an evaluative effect. Similarly, "creative push" in *The Old and the New* may represent creative activity as a force that moves something forward. Under the criteria adopted in this study, these expressions require justification based on the relationship between contextual and more basic meanings, rather than on vividness alone. The discussion of Suprematist theory in *Houses* illustrates a different issue: an interpretation of an artwork's artistic or conceptual significance does not necessarily constitute a metaphor. These examples establish the linguistic distinctions relevant to the study, but conclusions about ChatGPT's actual classifications require examination of its recorded responses.

Comparison of Manual Annotation and ChatGPT-Based Analysis

The comparison is organised around three dimensions: identification of candidate

expressions, assignment of analytical categories, and justification of interpretations. Table 1 presents representative cases from the selected descriptions. The classifications summarise the linguistic analysis of the quoted expressions; they should not be interpreted as a completed statistical measurement of ChatGPT's accuracy.

Artwork	Expression from the catalogue	Proposed linguistic classification	Main analytical consideration
Caravan	"an endless quest for a new artistic language"	Evaluation; metaphorical candidate	Artistic development is framed as an ongoing search; "endless" also intensifies the description
Caravan	"a tremendous influence on his fellow artists in Central Asia"	Explicit evaluation	"Tremendous" intensifies a claim about artistic influence
Caravan	"the most vivid and enduring symbols of Central Asia"	Evaluation, explicit symbolic reference	"Vivid" and "enduring" evaluate the symbol, but the reference to a symbol is not itself proof of lexical metaphor
Caravan	"preparing and loading packs onto the camels"	Literal description	The phrase describes an action represented in the painting
Northern Lights	"a celestial explosion of colours"	Metaphorical expression with evaluative force	The visual effect is described through the image of an explosion
Northern Lights	"this fantastic landscape"	Evaluation; possible conventional figurative usage depending on context	"Fantastic" expresses positive appreciation; its use alone does not establish a distinct metaphor.
Houses	"an irregular square"	Primarily literal description	The expression identifies a visual form rather than clearly assessing artistic quality
Houses	"the extremely simple shapes create a memorable image"	Evaluation and description	"Memorable" explicitly assesses the effect, while the description of simple shapes provides a visual basis for the judgement
The Old and the New	"the most advanced museum of the 1920s"	Comparative evaluation	"Most advanced" presents a superlative judgement about the museum's status
The Old and the New	"a creative push"	Evaluative; metaphorical candidate	The phrase conveys creative energy and, may frame activity as a physical force

Table 1. *Representative expressions and their linguistic classification*

Expressions are quoted from the English-language catalogue entries. The categories represent contextual linguistic judgements, and some expressions may perform more than one function.

The examples suggest that the most important point of comparison is not simply whether ChatGPT identifies a relevant phrase. It is also necessary to determine whether the model distinguishes the phrase's evaluative function from its metaphorical or literal meaning. For instance, identifying "a celestial explosion of colours" as metaphorical would be incomplete without explaining how the expression describes the visual effect of the painting. Similarly, recognising "the most vivid and enduring symbols of Central Asia" as interpretive language does not establish that the phrase contains a lexical metaphor. A further issue concerns expressions that combine more than one function. "An endless quest for a new artistic language" is both evaluative and potentially metaphorical: it characterises artistic development positively while representing it as a search. A classification system that allows only one label may fail to capture this overlap. The analysis therefore treats category overlap as a feature to be examined, rather than automatically as an annotation error.

The present comparison remains qualitative. Although the examples identify clear distinctions that can be used to assess AI-generated analysis, a numerical agreement rate cannot be reported until every candidate expression in the defined corpus has been coded systematically under the same rules. The illustrative cases should therefore be understood as evidence of the analytical issues encountered, not as a complete measure of model performance.

Discussion

Evaluative Language as a Context-Dependent Category

The preliminary analysis indicates that evaluative language in museum catalogue descriptions cannot be identified solely through the presence of adjectives with positive or negative connotations. Evaluation may be expressed through intensifiers, superlative constructions, and statements about artistic influence or originality. For example, the phrase "a tremendous influence on his fellow artists in Central Asia" evaluates the significance of an artist's contribution, whereas "the most advanced museum of the 1920s" presents a comparative judgement. These expressions differ grammatically, but both communicate an assessment rather than merely report a visible feature (Museum of Art named after I.V. Savitsky, 2020).

This observation is consistent with the view that evaluation operates across a range of linguistic resources rather than through a fixed list of evaluative words. Martin and White (2005) describe evaluative meaning as a system through which language communicates attitudes, judgements, and assessments. Lazzeretti (2025) similarly examines evaluation as a feature of museum communication, including assessments related to quality, novelty, and significance. These perspectives support the treatment of evaluation as a contextual category in the present study. For AI-assisted analysis, the practical implication is that identifying an apparently evaluative word is only part of the task. A system must also establish what is being evaluated and how the expression functions in the sentence. An adjective describing a visual characteristic should not automatically be treated as an assessment of artistic quality. Conversely, a statement about an artist's influence may communicate evaluation even when it does not describe the visual appearance of a particular painting. The distinction is important when evaluating the usefulness of ChatGPT for linguistic analysis because classification must be supported by the function of the expression in its context.

Distinguishing Metaphor from Symbolism

A second issue concerns the relationship between metaphorical language and symbolic

interpretation. In the selected catalogue entries, "a celestial explosion of colours" provides a strong candidate for metaphorical analysis because the word "explosion" is used to describe a visual effect rather than a literal physical event. The phrase "an endless quest for a new artistic language" similarly presents artistic development through the conceptual framework of a search. In both cases, figurative wording contributes to the interpretation of the artwork or artistic process (Museum of Art named after I.V. Savitsky, 2020).

These examples can be examined in relation to conceptual metaphor theory, which explains how one domain of experience may structure the understanding of another (Lakoff & Johnson, 1980). However, not every interpretive statement in a museum catalogue should be classified as metaphorical. A text may explicitly describe an object as a symbol of a cultural or historical idea without using a metaphorical expression at the lexical level. The identification of metaphor therefore requires attention to the wording and its contextual meaning, rather than relying exclusively on the interpretive significance of the passage (Pragglejaz Group, 2007; Steen, 2007).

This distinction has direct implications for AI-based annotation. A model may recognise that a passage is symbolic or interpretive while still assigning an overly broad metaphor label. Alternatively, it may correctly identify a figurative expression but fail to explain how its contextual meaning differs from a more basic meaning. Such cases need to be distinguished from simple omissions because they concern the relationship between category assignment and linguistic justification.

At the same time, a borderline classification does not necessarily demonstrate that the model has made an unequivocal error. Metaphor identification can involve conventional expressions and context-sensitive decisions, and competing interpretations may remain plausible. Consequently, disagreements should be examined individually and explained with reference to the analytical criteria rather than treated as self-evident evidence of failure (Pragglejaz Group, 2007; Steen, 2007).

Implications for AI-Assisted Qualitative Analysis

The present analytical framework also raises a broader methodological question: how should the output of a large language model be evaluated when the task involves interpretation rather than simple information extraction? A fluent explanation may appear convincing even when its classification is insufficiently supported by the source text. For this reason, the analysis focuses on the relationship between the quoted expression, its assigned category, and the explanation offered for that decision.

Research on large language models in qualitative coding provides a relevant methodological context. Zhang et al. (2024), in their preprint on QualiGPT, explore the potential of language models as tools for qualitative coding and compare AI-supported approaches with manual coding. The present study addresses a narrower task within this broader area: the identification of evaluative language and metaphor in specialised museum descriptions. Its emphasis on individual expressions makes it possible to examine the reasoning behind classifications rather than relying solely on an overall impression of the model's performance.

Nevertheless, the usefulness of AI-assisted analysis cannot be established from illustrative examples alone. A systematic comparison requires a completed corpus, explicit annotation criteria, and a record of the decisions made by both the model and the human annotator. Where only one human annotator is involved, the resulting comparison should not be presented as inter-annotator agreement. Similarly, a single analysis by one model configuration cannot establish reproducibility across different sessions or versions. The methodological value of the present study therefore depends on transparency. Reporting the model, the

prompt, the selected texts, and the criteria for classification allows readers to understand the conditions under which the analysis was conducted. This approach does not eliminate interpretive uncertainty, but it makes the basis of the conclusions more visible and allows future research to repeat or extend the procedure.

Limitations and Directions for Further Research

Several limitations must be considered when interpreting the study. First, the corpus consists of a small purposive sample of English-language catalogue entries. The examples are useful for close analysis, but they do not represent all museum texts or all forms of evaluative and metaphorical language. Second, the study focuses on one AI system under a defined set of conditions. Its observations should not automatically be generalised to other models or versions. Third, the interpretation of evaluative and metaphorical expressions may involve borderline cases. Although explicit criteria can make annotation more consistent, they cannot guarantee that every expression has only one defensible classification. The use of manual annotation as a reference therefore requires a clear distinction between strong evidence of a classification error and a legitimate interpretive disagreement. A stronger design for future research could include an additional independent annotator, a documented reconciliation procedure, and repeated model runs to examine whether the same expressions receive consistent classifications. Finally, quantitative measures of agreement would be meaningful only after the full corpus had been coded and the units of comparison had been defined. The present study should therefore avoid reporting accuracy rates or percentages unless these have been calculated from a complete set of observations.Future work could extend the sample to additional catalogue entries and compare multiple model configurations while retaining the same annotation criteria.

Overall Interpretation

The central methodological implication is that AI-assisted analysis of museum discourse should be evaluated at the level of individual linguistic decisions. Identifying a potentially relevant expression, assigning it a category, and justifying that category are related but distinct tasks. The distinction is particularly important when evaluative language overlaps with metaphor or when symbolic interpretation is expressed through otherwise literal wording. By focusing on these decisions, the study connects research on evaluation in museum communication with established approaches to metaphor identification and current work on AI-assisted qualitative coding (Lazzeretti, 2025; Pragglejaz Group, 2007; Zhang et al., 2024). Its conclusions, however, must remain proportionate to the evidence obtained from the selected texts and the comparison procedure actually conducted.

Conclusion

This study examines evaluative language and metaphor in selected English-language descriptions from the I.V. Savitsky State Museum of Art of the Republic of Karakalpakstan, drawing on research into museum communication, metaphor identification, and AI-assisted qualitative analysis (Lazzeretti, 2025; Pragglejaz Group, 2007; Zhang et al., 2024). It highlights the need to distinguish evaluative expressions, metaphorical language, literal descriptions, and symbolic interpretations, whose identification depends on both wording and context (Lakoff & Johnson, 1980; Martin & White, 2005). The study also emphasises that AI-generated classifications require justification based on the source text rather than plausibility alone. However, the small purposive sample, use of a single AI system, and absence of independent annotation and repeated model runs limit the conclusions. A systematic comparison of ChatGPT's actual responses with manual annotation is necessary to establish areas of agreement, disagreement, and interpretive uncertainty. Further research could extend the

corpus and involve additional annotators and language models to assess the consistency of AI-assisted linguistic analysis in museum discourse.

References:

1. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
2. Cameron, L. (2003). *Metaphor in educational discourse*. Continuum.
3. Fairclough, N. (1992). *Discourse and social change*. Polity Press.
4. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
5. Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
6. Lazzeretti, C. (2025). *Evaluation in museum communication*. In D. Hocking (Ed.), The language of art and artists. 55–77. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-89745-0_3
7. Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan. https://doi.org/10.1057/9780230511910
8. Museum of Art named after I. V. Savitsky. (2020). *Collection highlights*.
9. Ntzani, D. (2015). Under the spell of metaphors: Investigating the effects of conduit and container metaphors on museum experience. *Curator: The Museum Journal, 58*(1), 59–76. https://doi.org/10.1111/cura.12098
10. Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol, 22*(1), 1–39. https://doi.org/10.1080/10926480709336752
11. Ravelli, L. J. (2006). *Museum texts: Communication frameworks*. Routledge.
12. Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE.
13. Semino, E. (2008). *Metaphor in discourse*. Cambridge University Press.
14. Steen, G. J. (2007). *Finding metaphor in grammar and usage: A methodological analysis of theory and research*. John Benjamins Publishing Company. https://doi.org/10.1075/celcr.10
15. Zhang, H., Wu, C., Xie, J., Rubino, F., Graver, S., Kim, C., Carroll, J. M., & Cai, J. (2024). *When qualitative research meets large language model: Exploring the potential of QualiGPT as a tool for qualitative coding* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2407.14925