

MIKLÓS HOFFMANN

**LARGE LANGUAGE MODELS AND ASSOCIATIVE FIELDS IN AI-BASED
CREATIVE WRITING AND LITERARY INTERPRETATION**

Abstract

Artificial intelligence-supported Large Language Models (LLMs) such as GPT-4, Claude, and Gemini have made remarkable progress in seemingly understanding and generating natural language texts. However, their performance in interpreting and regenerating the associative field of the text, especially in terms of short, poetic, or aphoristic texts, remains deeply limited. By associative field, we refer to the semantic, emotional, cultural, and symbolic network evoked by a text beyond its literal meaning. This paper investigates the key shortcomings in LLMs' grasp of associative fields, explores the underlying causes, and evaluates the implications for literary analysis, creativity, and human-machine communication. We hypothesise that the smaller the corpus and the more complicated the (human) associative field around the sample poem, the less likely we will expect anything valuable from the LLMs. That is, the fewer points of contact the model has to map the associative field, the more difficult it will be to meet expectations and create something similar to the original and of a similar standard. We provide some relevant examples of these shortcomings by (the lack of) generating poems analogous to some famous English writers, including Shakespeare, and some poems of Hungarian authors Sándor Weöres and László Nagy.

Keywords: Large Language Model, associative field, poetry, AI-generated poem, AI-based creative writing

Introduction

Poetic and condensed texts often rely on ambiguity, allusion, intertextuality, and resonance to generate a wider associative field beyond or around the primary meaning. Unlike expository language, which is frequently used in hard sciences and aims to clarify the exact, unified meaning as much as possible, poetic language opens space for various interpretations. Understanding a poem often involves tracing a network of associations – historical, emotional, symbolic, or personal – rarely made explicit.

Large Language Models (LLMs) have shown impressive capabilities in syntax, coherence, and even stylistic mimicry of a longer text corpus. Yet, when faced with short, dense texts, they often flatten meaning, literalise

metaphor, or misattribute intent. This survey attempts to map the scope and nature of these failures and clarify what „understanding” in associative meaning would require.

Shen et al. (2024) conducted a comprehensive examination of several state-of-the-art LLMs in the context of cultural commonsense tasks. Their findings revealed that LLMs exhibit significant discrepancies in performance when tested on culture-specific commonsense knowledge, highlighting an inherent bias in their cultural understanding. This suggests that LLMs may not fully grasp the cultural nuances embedded in poetic language. Using several general and cultural commonsense benchmarks, they found that LLMs have a significant discrepancy in performance when tested on culture-specific knowledge for different cultures. LLMs’ general capability is deeply affected by cultural context.

Cuskley et al. (2024) argue that the capabilities of LLMs provide limited insights into human language and cognition. They emphasise that LLMs’ understanding is primarily functional and lacks the depth of human linguistic behaviour, which is broader and more complex. This limitation is particularly evident in tasks requiring the interpretation of associative meanings in short, dense texts.

Weigang and Brom (2025) explored LLMs’ challenges in preserving poetic intent during translation. Their study found that LLMs often struggle with cultural and literary retention, exemplifying the „paradox of poetic intent”. This indicates that LLMs may not effectively capture the associative fields of poetic texts, mainly when translating between languages with distinct cultural contexts. Contrary to this, Ichien et al. (2024) suggest that LLMs like GPT-4 have acquired an emergent ability to interpret novel literary metaphors. However, this ability is still developing, and LLMs may not consistently grasp the full range of associative meanings in poetic language.

This issue is interconnected and gains a broader horizon through the debate over AI’s „understanding”. Mitchell and Krakauer (2022) provide an outstanding survey of the discussion of whether LLMs can be said to „understand” language in any meaningful sense. They highlight that while LLMs can generate linguistically coherent text, their understanding is limited and does not encompass the depth of human cognition, particularly in interpreting the associative fields of poetic texts.

These studies underscore that while various LLMs have significantly enhanced natural language processing, they still face considerable challenges in interpreting and generating the associative fields of short, poetic, or aphoristic texts. Their performance in these areas is often limited by cultural

bias, lack of deep understanding, and challenges in preserving poetic intent. These limitations have essential implications for literary analysis, creativity, and human-machine communication applications.

The relationship between poetry and Large Language Models can be examined from two perspectives. The first direction is when we try to involve the LLM in writing poetry by giving it a prose prompt („Please, write a poem about...”). This aspect was examined in detail – primarily from a methodological point of view – first by Kangasharju et al (2022) and Cahkrabarty et al (2022), and later by Tsao and Nogues (2024). The LLM shows quite remarkable abilities in this direction. The enormous size of the corpus used for its training enables it to use words and expressions in a kind of semantic proximity to each other, linguistically and grammatically correct. Words that suggest or imply others (e.g., „rose” might evoke „love”, „thorns”, „fragility”) appear in many places in the teaching corpus, so the LLM uses these direct associations well and in the right places. Also, the technical aspects of poetry are well embedded in this system. LLMs seem proficient in various verse forms, from Sapphic stanza to modern rhythmic structures, and respond well to prompts to create output using these forms.

The other direction is when we provide sample(s), a poem or a set of poems, and LLM is asked to create „something similar”. This requires understanding the poem and mapping its semantic associative field before the (re)creative action. Whatever „understanding” means in this context (the study of this question is beyond the scope of this paper), LLM is facing a much more complex challenge in this latter case. We hypothesise that the smaller the corpus and the more complicated the (human) associative field around the sample poem, the less likely we will expect anything valuable from the LLMs. This hypothesis is studied through some experiments described in the following sections.

The Role of the Associative Field and Related Issues

To define clearly our central notion, by the associative field of a text we mean the web of meanings it activates in a reader’s mind, including semantic proximity, that is words that suggest or imply others; emotional tonality, that is the mood and feeling emerged by the choice and arrangement of those specific words; symbolic networks, that is cultural or archetypic meanings attached to objects or images; intertextual references, that is echoes of other texts, styles, or genres; and finally contextual resonance, that is the sociohistorical or biographical implications suggested indirectly by the text.

Understanding poetry and short literary texts often means navigating these layered, sometimes contradictory fields. The challenge for LLMs is that this kind of interpretive work resists clear-cut representation or formalisation.

Key Problems in LLM interpretation or creation of associative fields are manifold. One of the most common errors is the „flattening of metaphorical language” into literal paraphrases. For instance, in interpreting „Hope is the thing with feathers” (title of the poem by Emily Dickinson), a model may emphasise ornithological or biological aspects of birds rather than understanding the emotional resilience Dickinson encodes in the metaphor (c.f. Freeman 1995).

Due to the very nature of the training methodology of LLMs, another issue is the underlying cause: LLMs often rely on statistical patterns over semantic depth, making them poor at mapping metaphor to affect or symbolic register. Tracking polysemy and ambiguity is also challenging for LLMs. Short poems often play on multiple meanings of the same word or overlapping words (see our experiment in the next section with the poem of Sándor Weöres), and LLMs may resolve ambiguity prematurely, choosing a dominant sense without recognising the deliberate ambiguity or double-entendre.

An example can be a haiku, Autumn moonlight by Matsuo Bashō (translated by Robert Hass)

Autumn moonlight –
A worm digs silently
into the chestnut.

The associative tension between natural imagery and implied mortality is often lost on models, who may observe only the primary, biological meaning.

Neglecting cultural, socio-historical and symbolic resonance is another problem for LLMs. This especially holds for languages with smaller training corpora (practically every language except English). In poetry, cultural resonance often carries a crucial message or meaning. For example, the word „cross” in a short poem may evoke Christian theology, burden, sacrifice, or betrayal, but also a crossroad in the countryside, depending on context. LLMs may list possible meanings but rarely weigh them appropriately or recognise the dominant symbolic field a poet is invoking. Instead, they may provide generic or statistically common interpretations.

LLMs also struggle with inadequate handling of intertextuality. Poems and aphorisms often refer to or gesture toward other texts. LLMs may miss

these allusions entirely or fail to appreciate their significance in the inner interpretation of meaning. For example, a line like „April is the cruellest month” (from *The Waste Land* by T.S. Eliot) carries profound intertextual weight. Still, an LLM might interpret it simply as a meteorological observation without the whole literary and cultural context. In our experiment, we used a similar poem to prove our hypothesis.

Disregard for the affective economy of minimal texts may also appear in the contextualisation and interpretation of poems in LLMs. Short literary forms may highly depend on the „density of affect”, the concentration of emotional or existential weights in a small space. LLMs, lacking actual affective experience or embodiment, tend to underplay/overplay or misread this economy. For example, when interpreting a poem like „This Is the Dark Time, My Love” by Martin Carter, LLMs may correctly observe the political or historical aspects, but miss the tenderness, intimacy, or tone of lament the poet highlighted in the phrase „my love.” Specifically, GPT-4 interprets it as follows:

„The full poem is a powerful and haunting reflection on colonial oppression, where the 'dark time', 'festival of guns' and 'man of death' conjure an atmosphere of dread and militarisation. Meaning: Carter is warning about the presence of oppressive forces (literal or symbolic) during political upheaval.”

As one can observe, this interpretation almost completely ignores the intimacy aspects of the poem.

Of course, LLMs have their evident cognitive and architectural limitations. LLMs do not „understand” writings in the human sense. Their limitations stem from the lack of embodiment, that is, the impossibility of mimicking human interpretation grounded in physical, emotional, and lived experience. It also stems from statistical inference. LLMs generate text based on likelihoods, not meaning or intention. Context limitations come from the fact that models operate on token windows, and even if long context windows are applied, they lack persistent memory and situated context. In (interpreting) poetry, the absence of personal or historical self is crucial. Poems often evoke subjective memory or biographical nuance, which AI cannot emulate.

Some important aspects are beyond this paper. For example, we do not consider ethical and educational contexts in terms of misleading interpretation or co-creation. When LLMs offer confident but shallow readings of poems, they may mislead students or casual readers into reductive understandings.

At the same time, in co-creative experiments, there is a risk of hiding cultural and societal biases with implications for representation, inclusivity, and marginalisation, while leading to a tendency towards simplistic literalism (Tsao and Nogues 2024). Generative poetry or analysis by LLMs can flatten poetic difference, privileging mainstream or statistically dominant modes over the eccentric, obscure, or culturally specific. However, combining LLM suggestions with curated (and supervised) human interpretive guidance may yield richer, more responsible literary tools. These aspects can also be included in the training session. Fine-tuning LLMs on annotated poetic corpora – where metaphor, ambiguity, and associative fields are marked – could improve sensitivity to these features.

Experiment and Discussion

Based on the above, our experiment instructed one of the significant language models, GPT-4, to compose two different poem types requiring very different associative abilities.

In the first experiment, we chose a well-defined and richly trained poem form and topic: the AI had to write a Shakespearean sonnet. In this experiment, we used ChatGPT, and the topic was introduced by the following, very brief and direct conversation:

Me: Do you know Shakespeare's sonnets?

ChatGPT: Yes, I know William Shakespeare's sonnets. He wrote 154 sonnets, which were published in one volume in 1609. These poems are mainly about love, time, beauty, transience, and the immortality of poetry. The sonnets are often addressed to a mysterious Fair Youth, a Dark Lady, or a Rival Poet, and are the subject of much debate today.

Is there a specific sonnet you would like to discuss, or would you like me to quote one in Hungarian or English?

Me: Please write a new, non-existent Shakespearean sonnet, the theme of which is the transience of love.

ChatGPT:

Upon the Fading Flame of Love

When first thy glance did set my soul on fire,

I dreamed that Time would bow to passion's reign;

But seasons turn, and pale that first desire,
And bloom decays beneath the gentle rain.
The rose that once did blush upon thy cheek
Now yields to shadows Time doth softly cast;
No vow can stay what years in silence seek,
Nor bind the breeze that sings of moments past.
Though lips may swear that love shall never fade,
The stars still fall, the tide still leaves the shore;
What once was gold grows dim in twilight's shade,
And hearts forget the names they whispered o'er.
Yet in this verse, our love shall linger on —
A ghost in rhyme, though flesh and fire be gone.

Still ChatGPT: I'd be happy to write more of these, even on other topics, if you'd like. Would you be interested in a sonnet about immortal poetry or the beauty of transience?

Me: Yes, please write one about this.

ChatGPT: Gladly! Here is a new, original Shakespearean-style sonnet about the beauty of passing away and the immortality of poetry:

On Beauty Lost, and Verse That Never Dies

The fairest bloom must bend beneath decay,
Though morning gilds it with the light of grace;
No hand can bid the golden hour stay,
Nor time retreat from beauty's fading face.
Yet in the grave where all things else must lie,
Verse walks unburied through the dust of years;
It holds the shade of those too soon passed by,
And paints with ink what we have lost in tears.
The rose will fall, but not the rose in rhyme,
Whose scent endures though petals turn to air;
The voice is hushed, yet echoes still through time,
When soul finds shape in syllables laid bare.
So let what dies be gathered into song,
And there, in art, forever still belong.

As we can see, LLM understood the prompt request very clearly and performed well in both syntactic, verse-form and semantic terms. A Shakespearean sonnet, as it is well-known, is a 14-line poem written in iambic pentameter with a specific rhyme scheme: ABAB CDCD EFEF GG. It's structured into three quatrains (four-line stanzas) followed by a concluding couplet (two-line stanza). The GPT training set contains all of Shakespeare's sonnets, so the more than 150 poems also provide the AI with a well-defined thematic spectrum. Shakespeare's sonnets obviously represent an outstanding part of the literary heritage for all of us, but from the point of view of our current study, we must state that the associative field of lines and phrases he used is not particularly broad or complex – they mainly encourage the reader to make clear, primary associations.

In our second test, we asked the GPT-4 model to write a poem similar to two poems from different times and approaches. Our hypothesis can be technically reformulated as the fewer points of contact the model has to map the association space, the more difficult it will be to meet expectations, and the more difficult it will be to create something similar to the original and of a similar standard. We chose two Hungarian poems for this. The first was a poem by Sándor Weöres consisting of just one single word, *Tojáséj*. This almost untranslatable poem immediately drew anyone who understood Hungarian into the rich field of associations. The poem „*Tojáséj*” is a playful combination of the words „*tojás*” (egg) and „*éj*” (night), which at the same time is separated by a single letter from the ordinary everyday expression of „*tojáshéj*” (eggshell). The outcome is similar in style to „*Egghell*”, in English. It is full of emotions and free associations, thanks to the intersecting and overlapping of the three intersecting and overlapping words. As János Lackfi, a contemporary poet, aptly described the poem's atmosphere: it is a special round-trip to the womb.

GPT-4 was asked to write a poem similar to this one-word poem. Perhaps not surprisingly, the AI was far from creating anything close to this poem. After many turns and plenty of explanations, the best approximations on its part were these: *Csendhíd* (Bridge of Silence); *Ébrenlét* (Hours of Waking); *Reményfény* (Light of Hope); *Szélcsend* (Dead Calm); *Holdfű* (Moongrass). Nothing similar to the power of the original poem. There were a few attempts to be somewhat closer to the Weöres-poem, for example „*Szóhaj*” (*szó* = word, *haj* = hair, *sóhaj* = sigh). But when asked, ChatGPT could not interpret or repeat it. Anyway, besides the syntactic similarity (insertion of an extra letter), this word-poem cannot build the same semantic richness and multidimensional association space as the original one.

The other poem we chose was, again, a brief poem by László Nagy, containing one straightforward sentence: „Itt nyugszik a Hármas Honvéd Gyalogezred” (Here rests the 3rd Home Defence Infantry Regiment). The difficulty in this poem from the GPT perspective is that this sentence could be a simple, everyday epitaph, or even a declarative sentence like a newspaper headline. Thus, GPT would need extensive socio-historical background knowledge and strong association skills to interpret and copy this poem.

And the GPT fails to jump to this level. GPT-4 was asked again to create poems similar to this one. The outcome was not only way below the level of the original poem (this is somewhat evident), but did not touch the simultaneously present sensitivity and brutality of the original work, and created simple sentences like „Itt állt egykor a gesztenyefa” (Here once stood the chestnut tree), or „Itt sétált, aki már nincs” (Here walked the one who is no longer there).

With these two experiments, we tried to prove that in terms of sample-based artistic texts generated by AI, in the case of works with a large corpus and mainly requiring an understanding of the primary associative space, there is a chance for a creative (re)generation. However, in cases where this associative space is challenging to create due to the scarcity of the corpus, while the associative space itself is vibrant and multidimensional for our human mind, LLMs cannot get even close to the text.

Conclusion

LLMs represent a leap forward in linguistic fluency but fall short in interpreting texts whose meaning is not denotative but associative. In poetry, where suggestion, resonance, and ambiguity are key aspects, their interpretations are often superficial or misguided. This gap is not simply technical but philosophical: meaning in art is not reducible to statistical frequency or token-level prediction. Until LLMs are grounded in deeper models of context, affect, and culture, their grasp of poetic language will remain partial at best. This also yields poor or misleading performance in terms of (re)creating poems by samples where the associative field of the original poems is rich, while the corpus itself is limited. Poetry writing must therefore remain a fundamentally human endeavour, even if new forms, notably AI-based co-creation or co-poesis, are promising alternatives for the next generation of writers. Our education system must start teaching its technical, ethical and intellectual aspects promptly.

References

- Chakrabarty, T. – Padmakumar, V. – He, H. 2022. Help me write a poem: Instruction tuning as a vehicle for collaborative poetry writing. *arXiv preprint arXiv:2210.13669*.
- Cuskley, Christine – Woods Rebecca – Flaherty, Molly. 2024. The Limitations of Large Language Models for Understanding Human Language and Cognition. *Open Mind* 31(8) 1058–1083. https://doi.org/10.1162/opmi_a_00160.
- Freeman, Margaret H. 1995. Metaphor making meaning: Dickinson's conceptual universe, *Journal of Pragmatics*, 24 (6), 643–666, [https://doi.org/10.1016/0378-2166\(95\)00006-E](https://doi.org/10.1016/0378-2166(95)00006-E)
- Ichien, N. – Stamenković, D. – Holyoak, K. J. 2024. Large language model displays emergent ability to interpret novel literary metaphors. *Metaphor and Symbol* 39(4), 296–309.
- Kangasharju, A. – Ilomäki, L. – Lakkala, M. – Toom, A. 2022. Lower secondary students' poetry writing with the AI-based poetry machine. *Computers and education: Artificial intelligence* 3, 100048.
- Mitchell, M. – Krakauer, D. C. 2023. The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences* 120(13), e2215907120. <https://doi.org/10.1073/pnas.2215907120>
- Siqi Shen – Lajanugen Logeswaran – Moontae Lee – Honglak Lee – Soujanya Poria, – Rada Mihalcea. 2024. Understanding the Capabilities and Limitations of Large Language Models for Cultural Commonsense. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 5668–5680, Mexico City, Mexico.
- Tsao, J. – Nogues, C. 2024. Beyond the author: Artificial intelligence, creative writing and intellectual emancipation. *Poetics* 102, 101865.
- Weigang, Li – Brom, Pedro Carvalho 2025. The Paradox of Poetic Intent in Back-Translation: Evaluating the Quality of Large Language Models in Chinese Translation. *arXiv preprint arXiv:2504.16286*.

Miklós Hoffmann

university professor

Eszterházy Károly Catholic University Faculty of Informatics

hoffmann.miklos@uni-eszterhazy.hu

<https://orcid.org/0000-0001-8846-232X>