

New research shows people can't tell the difference between human and AI poetry – and even prefer the latter. What gives?

AI models can produce poems that rate well on certain 'metrics'. But the event of reading poetry is not one in which we arrive at standardised outcomes.

By Andrew Dean

6 min. read · [View original](#)

Here are some lines Sylvia Plath never wrote:

The air is thick with tension,
My mind is a tangled mess,
The weight of my emotions
Is heavy on my chest.

This apparently Plath-like verse was produced by GPT3.5 in response to the prompt “write a short poem in the style of Sylvia Plath”.

The stanza hits the key points readers may expect of Plath’s poetry, and perhaps a poem more generally. It suggests a sense of despair as the writer struggles with internal demons. “Mess” and “chest” are a near-rhyme, which reassures us that we are in the realm of poetry.

According to [a new paper in Nature Scientific Reports](#), non-expert readers of poetry cannot distinguish poetry written by AI from that written by canonical poets. Moreover, general readers tend to prefer poetry written by AI – at least until they are told it is written by a machine.

In the study, AI was used to generate poetry “in the style of” ten poets: Geoffrey Chaucer, William Shakespeare, Samuel Butler, Lord Byron, Walt Whitman, Emily Dickinson, T.S. Eliot, Allen Ginsberg, Sylvia Plath and Dorothea Lasky.

Participants were presented with ten poems in random order, five from a real poet and five AI imitations. They were then asked whether they thought each poem was AI or human, rating their confidence on a scale of 1-100.

A second group of participants were exposed to three different scenarios. Some were told that all the poems they were given were human. Some were told they were reading only AI poems. Some were not told anything.

They were then presented with five human and five AI poems and asked to rank them on a seven point scale, from extremely bad to extremely good. The participants who were told nothing were also asked to guess whether each poem was human or AI.

The researchers found that AI poems scored higher than their human-written counterparts in attributes such as “creativity”, “atmosphere” and “emotional quality”.

The AI “Plath” poem quoted above is one of those included in the study, set against several she actually wrote.

A sign of quality?

As a lecturer in English, these outcomes do not surprise me. Poetry is the literary form that my students find most unfamiliar and difficult. I am sure this holds true of wider society as well.

While most of us have been taught poetry at some point, likely in high school, our reading does not tend to go much beyond that. This is despite the ubiquity of poetry. We see it every day: circulated on Instagram, plastered on coffee cups and printed in greeting cards.

The researchers suggest that “by many metrics, specialized AI models are able to produce high-quality poetry”. But they don't interrogate what we actually mean by “high-quality”.

In my view, the results of the study are less testaments to the “quality” of machine poetry than to the wider difficulty of giving life to poetry. It takes reading and rereading to experience what literary critic [Derek Attridge](#) has called the “event” of literature, where “new possibilities of meaning and

feeling” open within us. In the most significant kinds of literary experiences, “we feel pulled along by the work as we push ourselves through it”.

Attridge quotes philosopher [Walter Benjamin](#) to make this point: literature “is not statement or the imparting of information”.



Philosopher Walter Benjamin argued that literature is not simply the imparting of information. [Public domain, via Wikimedia Commons](#)

Yet pushing ourselves through remains as difficult as ever – perhaps more so in a world where we expect instant answers. Participants favoured poems that were easier to interpret and understand.

When readers say they prefer AI poetry, then, they would seem to be registering their frustration when faced with writing that does not yield to their attention. If we do not know how to begin with poems, we end up relying on conventional “poetic” signs to make determinations about quality and preference.

This is of course the realm of GPT, which writes formally adequate sonnets in seconds. The large language models used in AI are success-orientated machines that aim to satisfy general taste, and they are effective at doing so. The machines give us the poems we think we want: ones that tell us things.

How poems think

The work of teaching is to help students to attune themselves to how poems think, poem by poem and poet by poet, so they can gain access to poetry's

specific intelligence. In my introductory course, I take about an hour to work through Sylvia Plath's [Morning Song](#). I have spent ten minutes or more on the opening line: "Love set you going like a fat gold watch."

How might a "watch" be connected to "set you going"? How can love set something going? What does a "fat gold watch" mean to you – and how is it different from a slim silver one? Why "set you going" rather than "led to your birth"? And what does all this mean in a poem about having a baby, and all the ambivalent feelings this may produce in a mother?

In one of the real Plath poems that was included in the survey, [Winter Landscape, With Rooks](#), we observe how her mental atmosphere unfurls around the waterways of the Cambridgeshire Fens in February:

Water in the millrace, through a sluice
of stone,
plunges headlong into that black pond
where, absurd and out-of-season, a
single swan
floats chaste as snow, taunting the

clouded mind
which hungers to haul the white
reflection down.

How different is this to GPT's Plath poem? The achievement of the opening of *Winter Landscape, With Rooks* is how it intricately explores the connection between mental events and place. Given the wider interest of the poem in emotional states, its details seem to convey the tumble of life's events through our minds.

Our minds are turned by life just as the mill is turned by water; these experiences and mental processes accumulate in a scarcely understood "black pond".

Intriguingly, the poet finds that this metaphor, well constructed though it may be, does not quite work. This is not because of a failure of language, but because of the landscape she is trying to turn into art, which is refusing to submit to her emotional atmosphere. Despite everything she feels, a swan floats on serenely – even if she "hungers" to haul its "white reflection down".

I mention these lines because they turn around the Plath-like poem of GPT3.5. They remind us of the unexpected outcomes of giving life to poems. Plath acknowledges not just the weight of her despair, but the absurd figure she may be within a landscape she wants to reflect her sadness.

She compares herself to the bird that gives the poem its title:

feathered dark in thought, I stalk like
a rook,
brooding as the winter night comes
on.

These lines are unlikely to register highly in the study's terms of literary response – “beautiful”, “inspiring”, “lyrical”, “meaningful”, and so on. But there is a kind of insight to them. Plath is the source of her torment, “feathered” as she is with her “dark thoughts”. She is “brooding”, trying to make the world into her imaginative vision.



Sylvia Plath. [RBainbridge2000](#), via [Wikimedia Commons](#), [CC BY](#)

The authors of the study are both right and wrong when they write that AI can “produce high-quality poetry”. The preference the study reveals for AI poetry over that written by humans does not suggest that machine poems are of a higher quality. The AI models can produce poems that rate well on certain “metrics”. But the event of reading poetry is ultimately not one in which we arrive at standardised criteria or outcomes.

Instead, as we engage in imaginative tussles with poems, both we and the

poem are newly born. So the outcome of the research is that we have a highly specified and well thought-out examination of how people who know little about poetry respond to poems. But it fails to explore how poetry can be enlivened by meaningful shared encounters.

Spending time with poems of any kind, attending to their intelligence and the acts of sympathy and speculation required to confront their challenges, is as difficult as ever. As the Plath of GPT3.5 puts it:

My mind is a tangled mess,
[...]
I try to grasp at something solid.