




Amplifying the Anomaly: How Humans Choose Unproven Options and Large Language Models Avoid Them

Anthony K. Brandt 

Rice University

ABSTRACT

Both large language models (LLMs) and the human brain develop internal models of reality to make accurate predictions. Both typically prefer choices with the strongest track records. However, when faced with a creative challenge, LLMs remain committed to high-probability options while humans can opt for unproven ones. This paper delves into one way of making *unlikely events plausible*—“amplifying the anomaly.” The concept involves *extrapolating viable consequences from an unlikely proposition*. Rather than being treated as oddball events or “one-offs,” the anomaly permeates the creative work. Notably, novelty and appropriateness can be in tension with each other, with high utility coming at the cost of low novelty. Amplifying the anomaly aligns these competing demands. It enhances originality: the rarer a proposition and the more thoroughly it is worked out, the more unique and surprising the result. At the same time, the effectiveness and value of the unproven option also rises: thorough elaboration in the creative product establishes its fitness. Musical examples by Beethoven, Schubert, and contemporary composer Sky Macklay, along with creative products from other domains, illustrate this principle. Classic LLMs have several limitations that make it difficult to amplify the anomaly: they are steered toward norm-driven outcomes, make short-term decisions, and are not designed to self-evaluate. As a result, they have difficulty of developing unusual propositions in a non-obvious way without human guidance. Alternatives approaches, including adversarial networks and team AI, are briefly examined. Implications for the future of computational and human creativity are discussed.

ARTICLE HISTORY

Received December 1, 2023

Introduction

When large language models (LLMs) produce output that is novel and task appropriate (M. A. Runco & Jaeger, 2012), many feel that these computational systems meet the threshold of the standard definition of creativity (Miller, 2019; Orwig et al., 2024). Indeed, studies indicate that LLMs can match or exceed human performance on evaluations of creativity such as the Torrance Test (Guzik et al., 2023) and other measures (Hubert et al., 2024). However, others argue that these automated systems – lacking as they do agency, understanding, and judgment (Gilhooly, 2023; M. Runco, 2023) – are not on a par with human creativity. Still, it has been hard to be precise about the cost of these deficits, or how exactly human output differs from that of the machine. If agency, understanding, and judgment convey advantages, what are they? This paper is an attempt to address one facet of that question: how humans enhance originality.

Both classic LLMs like ChaptGPT and the human brain develop internal models of reality – an LLM from its training set, a brain from lived experience –

from which they predict which outcomes have the highest likelihood of success. Esling and Devis (2020) describe how this works in an LLM:

Training a given (AI) model usually relies on minimizing the *expectation* of a loss (error) term . . . This . . . means that the model is incentivized to perform correctly on the *most common* elements of knowledge, by trying to fit the *principal mode* of the distribution (where most examples are concentrated). (p. 7)

That is not that different from the brain’s go-to strategy. As Hiesinger (2021) writes:

When you learn something, you are prodding a system that tries to remain stable, and continues to make plausible predictions, based on the history of everything it ever learned so far (The brain) will align the new information as smoothly as possible with what is already there in the network. (p. 300)

But when it comes to the creative process, LLMs and brains diverge in a crucial way. LLMs are highly sighted: drawing on their vast databases, they remain committed to *mean* outcomes because those have the highest probability of success. Absent human supervision, an LLM is

an “infinite norm-generating machine” (Esling & Denis, p. 10). Meanwhile, under the right circumstances, human brains can step off paths of least resistance and steer our thoughts in unexpected directions. To produce surprising results, we are blind to the outcomes (Campbell, 1960) or, at best, semi-sighted (Dietrich, 2015).

Thus, while classic LLMs aim for high consensus outcomes, our ability to go off-road and select unlikely options confers a potential advantage. However, capitalizing on that advantage raises the risk of making something non-viable. After all, creativity requires more than inventing something unusual (Campbell, 1960; Dietrich, 2015; Mednick, 1962; Turner & Fauconnier, 2003). It also involves making *unlikely events plausible*. That is why most definitions such as the standard one include a stipulation that output is not only novel and surprising but also useful or appropriate (Amabile, 1988; Boden, 1998; M. A. Runco & Jaeger, 2012; Stein, 1974). The challenge is that the utility of non-obvious ideas is not always immediately apparent, especially when the work is avant-garde (Brandt, 2021).

So while human minds are able to grab an opportunity that classic LLMs are designed to avoid – selecting unproven options – the odds of failure increase. This paper delves into one way that human minds address this conundrum: “amplifying the anomaly.” The concept involves *extrapolating viable consequences from an unlikely proposition*. Rather than being treated as odd-ball events or “one-offs,” the unlikely proposition permeates the creative product. Crucially, novelty and appropriateness can be in tension with each other, with high utility coming at the cost of low novelty (Esling & Devis, 2020; Simonton, 2018). Amplifying the anomaly aligns these competing demands. It enhances originality: the rarer a proposition and the more thoroughly it is worked out, the more unique and surprising the result. At the same time, the effectiveness and value of the unproven option also rises: being thoroughly elaborated in the creative product establishes its viability. Amplifying the anomaly is one of the ways that human intelligence is reflected in creative products to an extent that computational systems cannot yet match.

To illustrate this concept, we will examine musical examples by Ludwig van Beethoven, Franz Schubert, and contemporary composer Sky Macklay. Why music? The arts, in general, are overt demonstrations of creativity that are often hard to view directly or document in other fields (Eagleman & Brandt, 2017). Music is a play of sonic patterns unconstrained by referential meaning (Meyer, 1956), making it a prime medium for observing general principles. Finally, notated music (as illustrated in the excerpts that

follow) involves *deliberate* creativity (Dietrich, 2015) – a process of extended elaboration for which amplifying the anomaly is particularly relevant. The analyses use as little musical terminology as possible, and no ability to read music is required. Necessary terms are explained in the footnotes. Written descriptions and graphic indications underscore relevant concepts. For those who wish to listen to the music, links to YouTube videos of representative recordings are listed in the footnotes, with timing cues indicated in the Figures.

The paper begins presents analyses of the three works, along with examples from other domains, unpacks some of the limitations and challenges of LLMs, and concludes with a discussion of the implications for generative AI and human creativity.

Beethoven: Symphony No. 8, IV

The last movement of Beethoven’s *Symphony No. 8* (van Beethoven, 1989) opens with a skittish, playful theme.¹ As it approaches its first cadence, the music arrives at a repeated figure affirming the home key of F-Major.² Had Beethoven wished, he could have continued directly to a restatement of the theme, as was typical. Instead, just as the theme winds down, Beethoven adds a jarring note outside the key (Figure 1).

This “jump scare” is clearly out of place – a low-probability event played very forcefully. Beethoven could have left it at that. But rather than treating it as a one-off, Beethoven continues to reference this incongruous event in various ways. In doing so, the composer takes a momentary aberration and magnifies its impact.

For instance, when Beethoven returns to the opening theme, he adds a loud jolt earlier in the phrase (Figure 2) – evoking the anomaly’s sudden dynamic change.

The first few measures shown in Figure 3 prepare a statement of a secondary melody in the home key. Except that isn’t what happens. Instead, Beethoven inserts a detour to D-flat Major – a key built on the anomalous note (Figure 3).³ A pitch that was originally out of place temporarily becomes the music’s resting point.

In the passage shown in Figure 4, hushed fragments of the main theme are interrupted by sudden jolts (Figure 4) – referencing the change in volume in a new way.

Later in the movement, Beethoven returns to the opening theme a third time. He could have returned to the theme in its original form. Instead, he *insists* on the anomaly over and over – a total of five times. This insistence knocks the music even further off-course to a distantly related key (f#-minor) (Figure 5).

Allegro vivace
♩ = 80

The musical score is presented in four systems. The first system begins with a piano (*pp*) dynamic and features a melodic line in the right hand and a bass line in the left hand, both containing triplets. The second system continues the melodic line with a sharp sign above the notes and triplets in the bass. The third system shows a change in dynamics to *più. p* and then *pp*, with the bass line becoming more active. The fourth system features a box around a specific note in the right hand, which is marked with a sharp sign and *ff*.

Figure 1. The opening of Beethoven: Symphony No. 8, IV [0:06]. The box indicates the anomalous note [0:22].

As the movement near its close, Beethoven offers one more surprise: the music is shaken awake by another startling chord (Figure 6) – this time adding first one and then a second note outside the key.

Thus, Beethoven's jarring note is not an isolated occurrence – it sprouts variants, each one unique. Beethoven's fringe option offers creative opportunities that would not exist without it. New experiences occur as a direct result of the initial novelty. In elaborating it, Beethoven shifts this movement's actuarial tables: events that might be out of place in other contexts become more anticipated (Meyer, 1956).

Crucially, Beethoven's movement does not offer a grab-bag of surprises: later ones are tethered to the earlier one – alluding to the original anomaly's vagrant pitch, unexpected loudness, and disruption of the phrase. Thanks to these repercussions, Beethoven turns the oddball event into one of his narrative's driving features – a fountain of novelty constantly refreshing

the music. The result is a customized through-line, exclusive to this work. Remove the anomaly and its follow-through and the music is totally acceptable – but more routine.

Beethoven's treatment of the anomaly makes it *intentional* and *relevant*: by insisting on it multiple times, he shows it is not an accident; and in having it repeatedly alter the course of the movement, he makes it worth paying attention to. In a communicative medium like music, this working out makes the outlier pitch contextually *useful* and *appropriate*: it adds to the drama and distinctiveness of the movement in an articulate way. Notably, it takes *time* for this to happen: when it first occurs, the vagrant pitch is a non-sequitur; it is only through its elaboration that its relevance is established. Beethoven is able to select an unproven option and, by developing it, establish its viability.

Current architecture sharply limits this creative strategy in LLMs. For starters, as discussed earlier, these



Figure 2. The first return of the theme [1:28], with a surprising outburst earlier in the phrase [1:34].

systems are norm-driven. Franceschelli and Musolesi (2023) write:

Even if prompted with the sentence “I wrote a new poem this morning;” they would complete it with what is most likely to follow such words, e.g. something close to what others have written in the past (Shanahan, 2022). It is a probabilistic process after all. The degree of dissimilarity would be small by design.” (p.4)

Indeed, Binz and Schulz (2023) found that GPT-3 continued to select higher probability outcomes even when lower outcome ones had greater odds of success.

Furthermore, LLM decision-making is very short-sighted. In making its next choice, an LLM can have considerable depth of field in its rearview mirror – the bigger the context it considers, the better the fit of its next choice (Bender et al., 2021). However, it has limited foresight: it makes these decisions one at a time and does not work out large-scale plans (Suri et al., 2024). If a system were to introduce an outlier event such as Beethoven’s vagrant pitch, its algorithm would look no farther ahead than the next step.

Third, LLMs are unable to evaluate their own production. Boden (1998) flagged this as a crucial limitation of early systems and it remains one of the most intractable problems in computer engineering (Franceschelli & Musolesi, 2023). As Esling and Devis (2020) remark, “One of the major flaws of AI applied to creativity

appears to lie within its inability to judge the creative structures that emerge from exploratory processes” (p. 8).

Given this architecture, autonomous LLMs function most effectively close to the mainstream, where exemplars are more robust. They are less well equipped for options at the margins and bespoke contexts – precisely what a forward-looking composer like Beethoven was striving for. It’s akin to the difference between a chain restaurant and a Michelin-star chef.

These shortcomings are reflected in a recent AI realization of Beethoven’s Tenth Symphony, created from the composer’s incomplete sketches. Because Beethoven had included voices for the first time in his Ninth Symphony, the programmers decided to incorporate a similar innovation in the Tenth – so they programmed the computational model to add an organ. However, unlike Beethoven’s elaborate treatment of voices in the Ninth, the generative system did not seem to recognize that it was dealing with something exceptional. It failed to amplify the anomaly in a compelling way, treating the organ too matter-of-factly: for instance, whereas Beethoven gives the choir a “front-stage fancy” entrance, the organ in the AI realization slips in unannounced as part of the orchestral texture; in later passages, there is nothing distinctive about the organ writing that could *only* have been written for that instrument.⁴ “To the

The figure shows a musical score for a piano piece. It is in G major and 3/4 time. The score consists of 26 measures. The first measure is marked *sfz*. Measures 7-11 are marked *p*. A box highlights measures 12-20, indicating Beethoven's insertion. The score includes various musical notations such as dynamics (*sfz*, *p*), articulation (accents), and phrasing slurs. The piece ends with a final chord in G major.

Figure 3. An interrupted move to the home key [3:04], with Beethoven's insertion indicated by the box [3:12–3:20].

extent the AI's dataset captures Beethoven's knowledge, it lacks crucial awareness of his imagination" (Brandt, 2023).

Schubert: String Quartet No. 15 in G-Major, I

In the Beethoven example, the anomaly is an accessory to the main theme. In Schubert's *String Quartet in*

G-Major (Schubert, 1973), it is the main theme itself. This anomaly leads to an even more unexpected result.⁵

Schubert's theme rests on a primary expressive contrast in Western tonality: the opposition between the Major and minor modes (Gagnon & Peretz, 2003; Horn & Huron, 2015). In traditional Western music, the Major mode is associated with positive feelings such as joy, triumph, and calm, whereas the minor mode is



Figure 4 shows a musical score in piano and bass staves. The piano part features dynamic markings *p*, *pp*, and *f*, along with triplet markings (3). The bass part also includes triplet markings. Two specific passages are highlighted with black boxes: one at measure 13 and another at measure 15.

Figure 4. This passage [3:47] incorporates sudden jolts [3:56, 3:59].

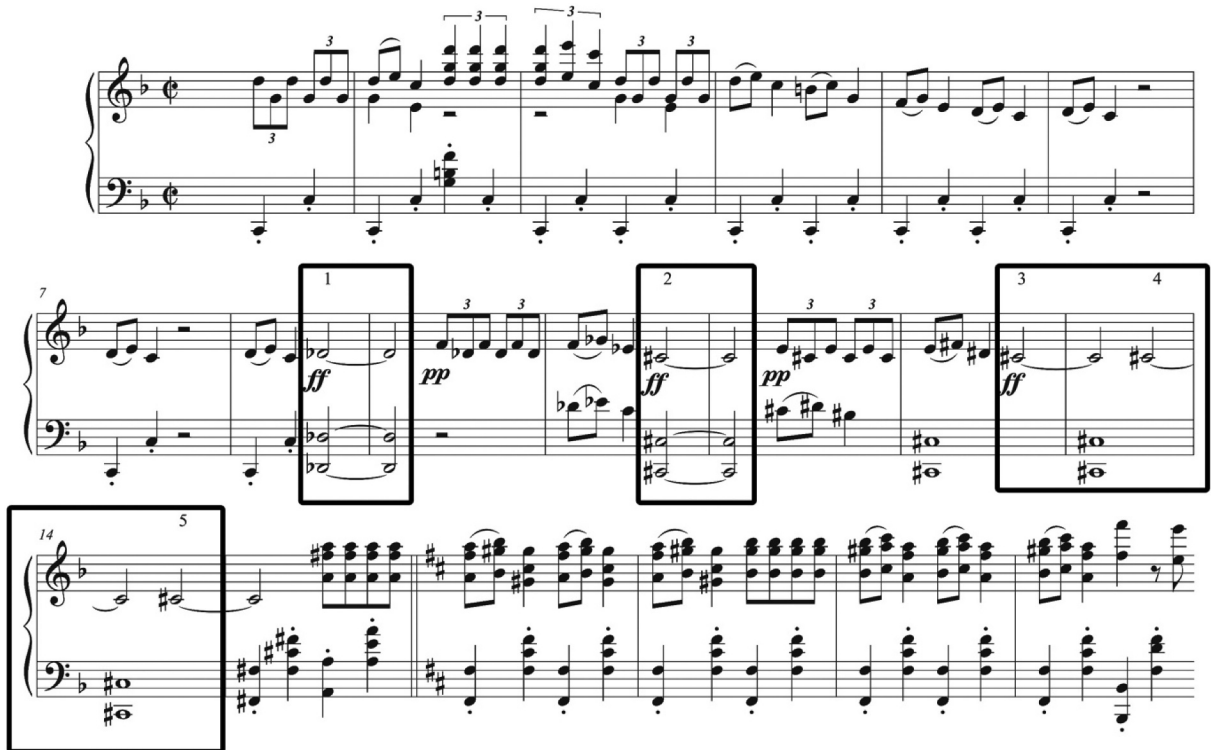


Figure 5 shows a musical score in piano and bass staves. The piano part features dynamic markings *ff* and *pp*, along with triplet markings (3). The bass part also includes triplet markings. Five specific passages are highlighted with black boxes and numbered 1 through 5, corresponding to the third statement of the theme.

Figure 5. The third statement of the theme [5:07] in which Beethoven repeats the anomalous note five times [5:13–5:18].

associated with negative ones, such as sadness and anger. For example, Mendelssohn’s well-known wedding march is in Major, and Chopin’s funeral march is in minor. While these associations are not culturally

universal (Jacoby et al., 2020), they can be virtually instantaneous in Western listeners, with processing as fast “as 200 ms in naïve listeners” (Bakker & Martin, 2015, p. 15).

Figure 6. The final arrival in the home key [5:40; Figure starts at 5:50] is disrupted by another startling chord [6:00].

In Schubert's day, changing modes was a well-established way to effect a mood change. Even funeral marches typically offered passages in Major as a relief from the pathos. These modal shifts can happen suddenly.

But nothing like the opening of Schubert's quartet had ever happened before. Typically, one mode would be firmly established before being supplanted by the other. Schubert's work begins with an immediate change of

Figure 7. The opening of Schubert's String Quartet in G-Major [0:15]. The change of modes is indicated [0:19].

mode within the *opening chord*: the strings play a Major chord that abruptly changes to minor (Figure 7). Before the music has found its footing, the mode flips.

As with Beethoven, Schubert does not treat this as a “one-off.” Instead, he integrates this modal uncertainty into the music that follows, which constantly oscillates between Major and minor, spending nearly equal time in both. Shifts happen both slowly and rapidly.

In the passage shown in Figure 8, as the music moves vigorously from key to key, the mode switches back and forth (Figure 8).

Schubert also employs a technique called *mixture of modes* in which chords that are usually exclusive to one mode are incorporated into the opposite one. This mixture happens at the level of the harmonic progressions as well as the larger key structure. In ways both small and large, minor and Major wrestle with each other.

None of this is too out of the ordinary in Schubert’s output, though here he elevates the frequency and intensity of the modal conflict: for instance, he switches from a full statement of the second theme in Major to a derived passage in minor *twice* in succession, giving it heightened emphasis.

At the opening’s return, however, Schubert does something truly unprecedented. At this pivotal moment, the mainstream solution would be to bring back the theme in its original form. Instead, the modes trade places: *minor* comes first, replaced by *Major*. Schubert reshapes the dynamic and rhythmic profile of the return to accentuate the change: this time, the modal flip is played softly by plucked strings (Figure 9).

Because of the workings of tonal progressions, such a scrambling of a theme’s harmony would generally render the music nonsensical. Schubert recognizes that his irregular opening affords him a one-of-a-kind opportunity: rather than “normalizing” his opening gambit, he *invents an outcome that can only exist in this alternate reality*. One novel idea generates another. Rather than reverting to routine, he does something unheard of.

In contrast, LLMs remain tethered to their databases. Even when considering counterfactuals, they adhere as closely as possible to established norms. As Franceschelli and Musolesi (2023) write:

The autoregressive nature of classic LLMs make them unlikely to generate surprising products (Bunescu and



Figure 8. A passage from Schubert’s quartet in which the mode switches rapidly back and forth [8:47–9:06].

The image shows a musical score for the recapitulation of Schubert's quartet. It consists of four staves: Violin I, Violin II, Viola, and Cello. The key signature is G major, and the time signature is 3/4. The score is divided into two sections by a vertical dashed line at 9:35. The first section is marked 'minor' and the second section is marked 'MAJOR'. Performance markings include 'pizz.' (pizzicato) and 'arco' (arco). Dynamics include 'pp' (pianissimo) and 'p' (piano). The Cello part has a 'p' marking with a hairpin symbol.

Figure 9. The recapitulation of Schubert's quartet with the modes reversed [9:31; the change in mode occurs at 9:35]. "Pizz" indicates plucked strings.

Uduehi, 2019), since they essentially trained to follow the current data distribution (Shanahan, 2022) (p. 4).

As an anecdotal example, consider the following exchange with ChatGPT-3.5. In 1972, the comedy troupe Monty Python created a comedic sketch titled "The Philosophers' Soccer Championship," which pitted ancient Greek thinkers such as Plato and Archimedes against their 19th German counterparts such as Hegel, Nietzsche and Kant (Davies & MacNaughton, 2020).

How would ChatGPT envision such an implausible contest? When queried, the LLM recognized that this was "purely a fictional and impossible scenario. However, if we were to entertain this imaginative idea for fun, we could speculate on what might happen" (for a full transcript, see [Appendix 1](#)).

ChatGPT imagined that the players' behavior would reflect their philosophies. Socrates would use "his method of questioning to outwit opponents," while Nietzsche would "bring a flair for the dramatic." It proposed that ancient Greeks would "prioritize fairness and justice . . . influenced by their ethical philosophies," whereas the Germans might "debate the very concept of rules and morality during the game."

The LLM concluded that

given the highly speculative and whimsical nature of this scenario, predicting the outcome is impossible. The game might be a chaotic blend of philosophical debates, Socratic irony, Kantian ethics, Hegelian dialectics, and Nietzschean will to power. The final score may not even be relevant.

There is a lot to admire in ChatGPT's response. It recognizes the hypothetical nature of the scenario and, drawing on its knowledge of Greek and German philosophers, comes up with plausible solutions. It knew

enough about games to say that, in this case, it may not matter who wins.

But Monty Python's skit goes a step further: in their match, *no one touches the ball*. Because LLMs are drawn to high-probability outcomes, their go-to approach is to "normalize" the irregular and fall back to the familiar as much as possible; their ingenuity is circumscribed. In contrast, Monty Python exaggerates its preposterous premise.

And we laugh – because although we've never seen anything like it, we understand Monty Python's satiric intentions. Under the right conditions, we can consider hypotheticals that have never existed before. "We don't have an airtight closed world; our world has porous borders that leak future" (Eagleman & Brandt, 2017, p. 250).

Prompted later for the most absurd possible outcome, ChatGPT offered the following: "As the players become more engrossed in their philosophical debates, the ball would likely be forgotten and the match would come to a standstill" ([Appendix 2](#)). At first, it might appear that GPT had broken through and generated an unproven thought. However, a subsequent query revealed that the Monty Python skit is part of its database; indeed, multiple references to the skit exist on the internet. The fact that the LLM did not initially return the fringe answer illustrates the degree to which the model aims for statistical norms as much as possible (Franceschelli & Musolesi, 2023).

Simonton's patent office definition of creativity requires that a creative product not only be novel, it must also be non-obvious (Simonton, 2012). Schubert and Monty Python clear that bar: Schubert did not have any examples in his database for restoring a theme with its harmonies flipped nor Monty Python for touchless

soccer; they extrapolate unexpected consequences from their anomalies. Yoked to its database, ChatGPT's responses are more straight-forward.

Sky Macklay: Many, Many cadences

Beethoven and Schubert created their works using the musical vernacular of their day: their novelties are aberrations within a communal framework. In the twentieth century, compositional practices began to diverge, eventually yielding a wide spectrum of rhetoric and aesthetics (Ross, 2007). As progressive as Beethoven and Schubert were, their music still considerably overlapped with their contemporaries. Not so for avant-garde modern composers, whose musical propositions are often quite singular. Each work establishes its own norms and exceptions, posing yet another level of creative challenge.

Sky Macklay's string quartet "Many, Many Cadences" (Macklay, 2014) makes an interesting case study. It takes stock harmonic progressions from Western classical music as its starting point but used in an irregular way.⁶

In the 17th and 18th centuries, aspiring musicians studied *partimenti*—catalogs of these harmonic formulas stitched together to improvise background music at social functions (Gjerdingen, 2007). Perhaps no feature of Western harmony is more often referenced than these sequences of chords.

In traditional music, these progressions were generally used in two ways: either different ones would be linked within a key to create longer phrases, or a single pattern would be repeated while *changing* keys. A composer would be unlikely to change *both* patterns and keys within the same phrase – it would be too bewildering.

Macklay takes an unexpected approach. Her quartet opens with a phrase consisting of a string of fifteen cadential patterns, with all but the first and last in different keys (Figure 10). This phrase then becomes the basis of a recurring harmonic cycle. The conventional rulebook has gone out the window: norms established across several centuries no longer apply. In its place, Macklay establishes a musical rhetoric in which events that would have been rare in earlier music become routine. Her work establishes its own probability tables.

To this singular scenario, Macklay adds small disturbances so that no two cycles are exactly alike: cadences oscillate between modes, there are substitute chords, and some cycles are abbreviated. These unsettle but do not overly distort the recurring cycle of cadences, which remains largely recognizable.

At the start of the seventh cycle, though, a more severe distortion occurs: the quartet begins to play *glissandi*—glides along the string – an effect atypical in the classical era that blurs the pitch (Figure 11).

At first, the *glissandi* are sporadic. But when the original cadential patterns return after a contrasting section, the *glissandi* begin to permeate the texture. Their speed and concentration create an effect similar to trying to see past windshield wipers during a downpour: it's hard to hear the underlying harmonies. Eventually, pitch is effaced altogether: Macklay just indicates the direction of the *glissandi* with approximate landing points (Figure 12).

The piece ends with the quartet percussively fingering the cadence patterns rather than bowing them. The cadential progressions are no longer audible, but the instrumental figures fall from high to low and change speed in ways reminiscent of the opening. Thus, "Many, Many Cadences" starts with precise pitch patterns that, overtaken by an accumulation of *glissandi*, become more indeterminate.

Macklay introduces an irregular sequence of cadences, and then radically transforms it, eventually abstracting its contour and rhythm. As anomalous as her opening patterns are to historical ones, so her final ones are to where she starts. Unprompted, an LLM would be unable to generate Macklay's cadential rhetoric since its database would have no precedent. Likewise, it would be unlikely to introduce *glissandi* since those have no established connection to cadential patterns. Finally, because LLMs make short-term decisions, it is hard to see how a model could regulate the timing and pacing of the *glissandi* that overtake the texture – that requires considering long-range outcomes in a way that LLMs are not designed to do. All of this underscores what a complex undertaking it is to produce something as idiosyncratic yet coherent as Macklay's quartet. Beethoven and Schubert were at least working within a shared language; Macklay is creating her own. The singularity of her approach poses a formidable challenge for a norms-based system.

Amplifying the anomaly in other domains

As in music, amplifying the anomaly is evident in creative products of other domains. In this poem by Cummings (1994), word spacing – from no separation to wider gaps – is modulated throughout the text. Note how capitalization is usually sparingly, underscoring the lack of punctuation, and individual words are set apart, heightening their emphasis. Cummings also creates a climactic compression, "onetwothreefourfive," evoking a skilled marksman. Thus, Cummings' anomalous syntax

frenzied, staccato notes
are non vibrato

Many Many Cadences

about ♩ = 120 Sky Macklay

The score consists of four staves: Violin I, Violin II, Viola, and Violoncello. Each staff contains musical notation with notes, rests, and slurs. Roman numeral analysis is provided below each staff. The first system of analysis is: C, Eb, A, eb, G, E, Db, bb, D. The second system is: B, Ab, F, Eb. The tempo is marked 'sempre f'.

Violin I
Violin II
Viola
Violoncello

sempre f

C Eb A eb G E Db bb D

sempre f

C Eb A eb G E Db bb D

sempre f

C Eb A eb G E Db bb D

sempre f

Cv I EbIV I A:vii^{o4} I eb:vii^{o4} i Gii V I Evi IV I Db:ii vii^{o3} I bbiv V i Dii V

sempre f

4

B Ab F Eb

B Ab F Eb

B Ab F Eb

I B:IV ii V I Ab:IV ii vii^{o7} I F:iii vi ii V I Eb:iii vi IV

Copyright 2014 by Sky Macklay (ASCAP), all rights reserved.

Figure 10. The opening measures of Many, Many Cadences [0:12–0:25]. Macklay includes an analysis of each cadence using Roman numerals from traditional music theory.

33

C *pp* — *f* Eb A eb G

C *pp* — *f* Eb A eb G

C *pp* — *f* Eb A eb G

pp — *f*

i VI iv vii⁷ i C I iii IV ii V I CV I Eb IV I Avii⁷ I eb V i Gii

Figure 11. The introduction of glissandi, notated by drawn-out lines between pitches [1:14].

permeates the text, creating persistent surprise and heightening the poem's meaning.

Buffalo Bill's
defunct
 who used to
 ride a watersmooth-silver
 stallion
and break onetwothreefourfive pigeonsjustlikethat
 Jesus
he was a handsome man
 and what i want to know is
how do you like your blueeyed boy
Mister Death
(p. 72)

This concept is found in technological innovation as well. In the 1970s and 80s, there was an epidemic of record piracy: at one point, retailers returned three million copies of Olivia Newton-John's hit album *Physical* (1981) for refunds – even though only two million copies had been printed. British inventor Kane Kramer initially set out to solve the problem by designing a print-on-demand machine, with the musical data transmitted digitally over phone lines and albums custom-made in store. But then he realized that carving an analog record was an unnecessary step: in its place, he designed the IXI, the first digital music player. That anomalous idea became a portal into reimagining the

record industry: in addition to a lightweight, portable machine, Kramer envisioned a world of unlimited supply in which warehouses would no longer be needed, music could be purchased anywhere, and the digital format meant you could read books on a music player – opportunities that were non-obvious to his contemporaries yet came to pass when his concept became the basis for the iPod (Eagleman & Brandt, 2017).

Scientific breakthroughs also involve amplifying the anomaly. Einstein's theory of relativity posited that the speed of light is constant to all observers. From that unlikely premise, Einstein extrapolated that absolute simultaneity did not exist, people moving at different speeds age at different rates, objects become infinitely massive as they approach the speed of light, and superluminal connections are possible – all hypothetical propositions that lay far beyond the databases of his day. By following through on the implications of his anomalous axiom, Einstein radically remade our understanding of reality (Greene, 1999).

Amplifying the anomaly can be an on-going progress, stretching out over a career. This is often the case in modern visual art, where a series of works can represent instantiations of a novel premise. In October 1952, Helen Frankenthaler mixed her paints with turpentine and applied them directly on an unprimed canvas, allowing the paint to soak in. "I

32

169

3 3 C C Eb A C Eb A C Eb A C Eb A

i VI iv vii⁷ i C I iii IV⁶ ii V I CV I EbIV I A: vii^o I iii IV⁶ I

171

G G G E G-V I Evi

Figure 12. The quartet texture dominated by glissandi [7:17]. The score uses indeterminate graphic notation, with only rare indications of precise pitch.

might have been very impatient to paint, and a combination of impatience, laziness, and innovation decided why not put it on straight?" (Nemerov 2021, p. 50). That experiment became the basis for her "soak stain" technique, explored over several decades, in

which the image resembles washes of watercolor: "the paint soaks into the canvas, retiring in the very act of appearing" (Nemerov 2021, p. 157). Artist Nick Cave created his first Soundsuit – wearable, whole-body sculptural costumes that completely conceal the

wearer's identity – in response to the Rodney King beatings. He continues to mine the possibilities of this unusual proposition, producing over 200 suits in a variety of shapes, materials and found objects (Lamm, 2017).

Amplifying the anomaly can also involve collective effort. When Steve Jobs took the stage in 2007 to launch the iPhone – “an iPod, a phone, and an internet communicator” (Jobs, 2007) – the prototypes were barely functional:

The software that ran the phone was full of bugs. The iPhone could play a section of a song or a video, but couldn't play an entire clip reliably without crashing. It worked fine if you sent an e-mail and then surfed the Web. But if you did those things in reverse, however, it might not. Hours of trial and error had helped the iPhone team develop what engineers called “the golden path,” a specific set of tasks in a specific way and order, that made the phone look as it worked. (Vogelstein, 2013, p. 4)

Thanks to the industriousness of countless software engineers, we now count on the iPhone to reliably pay for parking, track our fitness and driving habits, clip coupons, navigate, stream feature films, and record home movies – all extrapolations from Jobs' originally shaky proposition and non-obvious to prior generations accustomed to traditional phone handsets. Per Apple, there are now close to two million apps in its App store.

When amplifying the anomaly, the “strength” or “power” of an unproven idea is directly proportional to the quality and quantity of consequences it yields; it is this “working out” that establishes its bona fides (Sosa, 2019). This aligns with Simonton's formulation (Simonton, 2012) in which both ratings for novelty and utility both need to be high in order for a product to be optimally creative. Others argue for an inverted U-shaped curve (Diedrich et al., 2015), with novelty and usefulness as antagonists and the “sweet spot” somewhere in the middle. This, too, is implicated in many of the examples cited: Beethoven likely selected his vagrant note because it was outside *but not too far outside* the home key; the iPhone relied on skeuomorphs – digital icons such as bookshelves, trashcans and envelopes – to acclimate users to the analog features that it supplanted. In both these cases, novelty was held in check in order to be appropriate. This suggests that the novelty-usefulness axis may actually be a multi-dimensional plane: when it comes to unproven ideas, compromises along some axes allow for greater license along others.

Amplifying the anomaly is one of the ways humans leverage unlikely outcomes: if an anomaly can be sustained or propagated, its appeal and utility increase. By working out the implications of their most novel elements, artists enhance their originality, scientists uncover far-reaching consequences of their discoveries, and society is remade by new technologies.

Large language models and novelty

To ensure valid responses, classic LLMs are generally designed to *minimize* the anomaly. In such cases, these computational systems are working at cross-purposes with themselves: they are attempting to produce novel results by adhering to statistical norms. What would it require to address this? As Ganguli et al. (2022) have shown, even robust computational systems are capable of unpredictable behavior. Reliability increases if model size, training data and training time are scaled up in tandem, but even so “complex systems trained on complex data yield noisy or variable results” (p. 1749). So LLMs generate potential pools of low probability output that could be exploited. Just as the human brain offers a dopamine reward for prediction error (Diederer & Fletcher, 2021) – a key feature of learning – a model's programming could incorporate incentives for novelty.

Guiding an LLM toward more unusual output would require it to be assess its own production. Already, researchers have developed automated scoring methods for divergent thinking (Buczak et al., 2023; Organisciak et al., 2023), visual art (Patterson et al., 2023), and metaphor creativity (DiStefano et al., 2024). In many cases, results match those of human raters. Thus, it seems plausible that an LLM could eventually assess the novelty of its own work-in-progress. Accomplishing this would be a step toward computational self-awareness: the system would recognize when it had invented something rare.

However, indiscriminately spicing up output with novelty risks incoherence. That is where the strategy exemplified by Beethoven, Schubert and Macklay comes into play: *maximizing individual anomalies by developing them*. To that end, engineers would need to make an LLM capable of contingent choices: when it had generated an unproven option, the model would not automatically accept or toss it out. Instead, it would test its viability – for instance, by experimenting with how to elaborate it. If it could meaningfully build on a proposition, it would retain it; if not, it would abandon it. In this way, amplifying the anomaly would serve as a fitness function: unusual ideas with the greatest capacity for development would be favored. This strategy would put the system in line with Corazza's dynamic

definition of creativity, giving it a means to assess “potential effectiveness” (Corazza, 2016). Such speculative trials are more labor-intensive and prone to failure but would create a system more tolerant of ambiguity and capable of surprising outcomes. Supervised learning, in which human rankings fine tune the model, might be a way to accomplish this. Currently, though, engineers have other priorities – for instance, making sure text generation is tasteful: in those cases, “strict alignment to very careful and pre-designed human responses leads to the generation of text that might be considered *banal* (Hoel, 2022)” (Franceschelli & Musolesi, 2023, p. 5).

LLMs are dependent on human output for their knowledge base; but in other respects, their relationship to us is quite rudimentary. For instance, we take it for granted, but our brains are constantly updating. In contrast, many LLMs are frozen in time: if you ask ChatGPT3 “Who won the 2023 World Series?” it responds, “As of my last update in January 2022, I don’t have access to that information.” Adaptive models, which continue to take in new data, would enhance flexibility (Franceschelli & Musolesi, 2023) – especially if the dataset could incorporate the LLM’s on-going production. That would bring an LLM more in line with how we experience the world: memories are continually updated, and a creator is aware of his or her prior work.

While generating a novel idea can happen in a flash of insight (Kounios & Beeman, 2009), amplifying the anomaly depends on large-scale thinking. In the Beethoven, Schubert and Macklay works, surprises tend not to have happen in spasms but rather are distributed relatively equally across the span of the work. In addition, all three composers *escalate* the intensity of the original anomaly as their works progress: Beethoven’s maximum insistence on the vagrant pitch arrives near the end; Schubert concludes his movement with a culminating back-and-forth between Major and minor tonic chords; Macklay reaches her apotheosis in the movement’s percussive conclusion. Yet while each work trends toward greater intensity and distortion, that doesn’t happen in a straight line: there are steps forwards and back, in an intricate interplay with listener expectations (Meyer, 1956). In order to address this, an LLM would need to be programmed to consider the whole when making individual decisions and not only work forward but also back-chain (Brandt, 2023).

The shortcomings of these computational systems have not gone unnoticed. To increase the originality of the output, engineers have experimented with untethering neural nets from their datasets. For instance, in DeepDream, Mordvintsev created a program that interrupts feature extraction of an input such as

a photograph, externalizing the neural net’s hidden layers partway through (Al-Khazraji et al., 2023). Elgammal’s adversarial network incorporates a data-driven system trained on a vast corpus of visual art and an uninformed one designed to generate random images; the “generator” is rewarded when it creates an image the “discriminator” accepts as art but cannot pin down its style (Elgammal et al., 2017). Elwes’s generator is primed to produce an image that satisfies the discriminator; but once it reaches that threshold, the software decouples it, allowing it to explore the “latent space” on its own (Miller, 2019). While promising, these systems all have their limitations: for instance, as currently configured, they can produce “one-offs” but don’t have the awareness to build on their own production (Franceschelli & Musolesi, 2023); they are always starting over from scratch. Nor do they take the whole into account when making local decisions: while Beethoven, Schubert and Macklay’s output reflects careful judgments about the nature and pacing of their extrapolations, these systems are blunt tools. Developing human-like capabilities in these computational systems would give them the capacity for what Barron terms “controllable oddness” (1993): the ability to challenge convention without being overinclusive or unconstrained, and therefore capable of what Barron terms “refined complexity” (p. 183).

The “closed and orderly” worlds (González-Díaz & Palacios-Huerta, 2022, p. 4) of Chess and Go have yielded some of the most promising results. This game-based research has provided laboratories for computational creativity with explicit goals (i.e. winning) and well-defined and inviolate rules. AlphaZero taught itself chess through unsupervised reinforcement learning, rediscovering “ideas known in decades and centuries of human play” and creating “new ideas as well” (p. 5). To surpass these results – and address “blind spots” in computer play exploited by savvy humans– Zahavy et al. (2023) developed a novel approach: team AI. The authors observe:

Humans . . . seem capable of considering multiple good solutions when attempting to solve a new problem instance (Page, 2019), ruling out solutions that had worked before (Stokes, 2005), and accepting the notion of failure (Kasparov and King, 2000). This leads to a kind of creativity through diversity (of solutions) that is missing in current reinforcement learning (RL) agents. (p. 4)

The authors found that a heterogeneous group of RL systems, each incentivized to respond differently to the same game situations, outperformed a homogeneous one. The team approach makes the system more flexible and speculative in a controlled way. However, it remains

to be seen how readily these strategies can be adapted to creative tasks in which goals are more ambiguous and open-ended and rules malleable and capable of being overturned.

As engineers continue to explore the frontiers of computational creativity, a potential pitfall looms. We generally prize computers for their infallibility. Raising the stakes for originality introduces more risk – especially in low-consensus fields like the arts (Simonton, 2018). There’s no reason to think we will admire an algorithm’s avant-garde output more readily than a living artist – LLMs are likely to face the same crucible of competition and subjective appraisal. As composer Morton Feldman once put it: “In music, when you do something new, something original, you’re an amateur. Your imitators – these are the professionals” (Feldman & Zimmermann, 1985, p. 87). There is significant expense involved in training and maintaining these systems (Samsi et al., 2023) that will need to be weighed against possible benefits in low consensus fields. It may prove costly to produce a computational Beethoven or Macklay – with no guarantee of artistic success.

There is no doubt that LLMs are improving at breathtaking speed. What does these systems’ current progress mean for humans? Artists are already partnering with LLMs as co-creators, turning to them for idea generation and interactive design (Rafner et al., 2023; Liu et al., 2023). Ploin et al. (2022) write: “Technology-augmented artmaking may just be machine learning delivering on its full potential” (p. 74). So far, getting the most out of an LLM depends significantly on human input, which is why prompt engineering has emerged as a major new discipline (Meskó, 2023). However, this collaboration between humans and machine also has its vulnerabilities: Anderson et al. (2024) found that relying on LLMs as creative facilitators tended to homogenize the results. The authors conclude that LLMs may be best suited for the “rapid enumeration of relatively obvious solutions that users might otherwise fail, or take longer, to consider” (p. 7). Given an LLM’s exhaustive database, another potential use is road-testing a creative idea for uniqueness – the LLM would be like a “CarFax” for originality.

Much as the invention of photography has been credited with the genesis of abstract art (Moszynska, 2020), generative AI may lead to a flourishing of genre-defying, multimodal, boundary-breaking art. It’s possible that creators may push themselves as never before toward more bespoke output. Just as photography lessened the need for painters to visually reproduce reality, generative AI may lower human incentives to work in existing media and encourage adventurous work that is not easily assimilated into a database.

These systems may also prove to be potent research tools about human creativity. As Miller writes “artificial neural networks are designed to replicate the activities of the brain and cast light on how the brain works” (Miller, p. 107), and have already been used to model features of human cognition such as pattern recognition (Hinton et al., 2012), memory (Schmidhuber, 2015), attention (Brauwers & Frasinicar, 2021), and visual and auditory perception (Kanwisher et al., 2023). While even the most sophisticated computational systems can at best capture limited features of the machinery of human thought (Moro et al., 2023), LLM creativity may be instructive both in the ways it falls short and how it improves. Franceschelli and Musolesi (2023) write:

Transformational creativity is not achievable by means of the current LLM training solutions. *In theory*, other forms of training or fine-tuning might circumvent this limitation, allowing the model to forget the learned rules in order to forge others. (p. 5)

How engineers address these challenges may yield insights into the human creative process or at least offer transferable strategies. As much as we fear LLMs cannibalizing creative tasks, perhaps they will also help enlighten us on the black box of our own imaginations.

Conclusion

Despite LLMs’ impressive feats, human wetware still confers several distinct creative advantages. As their name implies, large language models require vast amounts of data: this means indiscriminately mixing the lofty with the mundane and the supreme with the ordinary (Bender et al., 2021). By virtue of their rarity, more original solutions will occupy the margins – especially the closer you get to the present, where there is less time for influence and imitation. In contrast, humans can pick and choose their role models and give greater weight to the exceptional. We can optimize our databases.

On top of that, one of humanity’s most remarkable traits is our comfort with mistakes. Tolerating errors is how we learn, adapt, and discover. In taking these risks, we “likely weigh perceived hazards against perceived benefits prior to acting” (Beghetto, 2021, p. 2). With LLMs still in their infancy, the reputational risks of failing far outweigh the potential benefits of originality, which is why software engineers work so hard to reduce hallucinations (Rawte et al., 2023) and reinforcement learning from human feedback (RLHF) is geared toward guaranteeing the most appropriate responses (Ouyang et al., 2022). We may be David to AI’s Goliath regarding

speed and fecundity, but we continue to hold the upper hand in risk-tolerance (M. Runco, 2023).

Beethoven, more than any other composer of his day, had an aversion to duplicating himself (Brandt, 2022). LLMs do not appear to frame creativity in this way. If you give ChatGPT the same prompt on different days, you get essentially the same response: the system does not care if it repeats itself. The fact that the answers are similar validates the LLM's decision-making – it means its choices were not arbitrary or careless. The drive to continue to speculate is an abiding human strength.

Finally, LLM creativity is closer to human improvisation, which is spontaneous, linear, and unconcerned with self-appraisal and even reproducibility (Dietrich, 2015). Because it can only think *forward*, it is unclear how an LLM would handle writing a whodunit, which requires devising a clever solution to a crime and then working backward to hide the clues. In contrast, deliberate human creativity depends upon reflection, filtering, revision, and weighing alternatives. To that end, it is unbound from chronology: a creator's mind can roam at will and contemplate different levels of scale. Beethoven, Schubert, and Macklay's skill in amplifying their anomalies depends on that cognitive flexibility, enabling them to test, time, and pace irregular events within an unfolding narrative.

Amplifying the anomaly is emblematic of the ways that agency, understanding, and judgment inform human creativity. Without those faculties, it would be difficult to ratify, elaborate, and integrate uncommon ideas – because, by definition, existing data about them is scarce. At a public forum in October 2023, Sam Altman, CEO of OpenAI, said, "Creativity has been easier for AI than people thought" (Jones, 2023). Maybe that's true for more routine production (Rafner et al., 2023). But as we probe the multi-faceted requirements of deliberate creativity, we find a more significant gap between human and machine. Sustained creativity with high marks for originality may be hard for humans, but it is even more challenging for algorithms. We shouldn't lose sight of that, lest we become overly dependent on systems that cannot yet rival our best work.

Notes

1. A videorecorded performance by Tafelmusik is available at <https://www.youtube.com/watch?v=RSqzJnLBtec>.
Timings of the musical examples are cued to this recording.
2. In Western tonal music, the key of a passage refers to the scale on which it is built. The "home" key refers to

the primary key of a movement: a movement generally begins in that key and *always* ends there.

3. In Western tempered tuning, C-sharp and D-flat are enharmonically equivalent – they are two ways of naming the same pitch.
4. Camille Saint-Saëns "Organ Symphony" offers a salient counter-example: for instance, in the second movement, the organ enters with a massive, fortissimo chord – something only an instrument of its vast dimensions can do.
5. A videorecorded performance by the Esme Quartet is available at https://www.youtube.com/watch?v=9One_hAjrI.
Timings of the musical examples are cued to this recording.
6. An audio recording with score by the Spektral Quartet is available at <https://www.youtube.com/watch?v=mrI39Nf7cj4>.
Timings of the musical examples are cued to this recording.

Acknowledgments

The author wishes to thank Dr. Lisa Serice (Johns Hopkins University) for her valuable feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Anthony K. Brandt  <http://orcid.org/0000-0001-6895-8221>

References

- Al-Khazraji, L. R., Abbas, A. R., & Jamil, A. S. (2023). A systematic review of deep dream. *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, 23(2), 192–209.
- Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior*, 10(1), 123–167.
- Anderson, B. R., Shah, J. H. & Kreminski, M. (2024). Homogenization Effects of Large Language Models on Human Creative Ideation. In *Creativity and Cognition (C&C '24)*, June 23–26, 2024, Chicago, IL, USA (p. 13). ACM. <https://doi.org/10.1145/3635636.3656204>.
- Bakker, D. R., & Martin, F. H. (2015). Musical chords and emotion: Major and minor triads are processed for emotion. *Cognitive, Affective, & Behavioral Neuroscience*, 15(1), 15–31. <https://doi.org/10.3758/s13415-014-0309-4>
- Barron, F. (1993). Controllable oddness as a resource in creativity. *Psychological Inquiry*, 4(3), 182–184. https://doi.org/10.1207/s15327965pli0403_3
- Beghetto, R. A. (2021). There is no creativity without uncertainty: Dubito Ergo Creo. *Journal of Creativity*, 31(100005), 100005. <https://doi.org/10.1016/j.jyoc.2021.100005>

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6), e2218523120. <https://doi.org/10.1073/pnas.2218523120>
- Boden, M. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, 103(1–2), 347–356. [https://doi.org/10.1016/S0004-3702\(98\)00055-1](https://doi.org/10.1016/S0004-3702(98)00055-1)
- Brandt, A. (2021). Defining creativity: A view from the arts. *Creativity Research Journal*, 32(2), 81–95. <https://doi.org/10.1080/10400419.2020.1855905>
- Brandt, A. (2022). Beethoven and divergent thinking. *Creativity Research Journal*, 1–18. <https://doi.org/10.1080/10400419.2022.2088131>
- Brandt, A. (2023). Beethoven's Ninth and AI's Tenth: A comparison of human and computational creativity. *Journal of Creativity*, 33(3), 100068. <https://doi.org/10.1016/j.yjoc.2023.100068>
- Brauwiers, G. & Frasincar, F. (2021). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3279–3298.
- Buczak, P., Huang, H., Forthmann, B., & Doebler, P. (2023). The machines take over: A comparison of various supervised learning approaches for automated scoring of divergent thinking tasks. *The Journal of Creative Behavior*, 57(1), 17–36. <https://doi.org/10.1002/jocb.559>
- Bunescu, R. C. & Uduehi, O. O. (2019). Learning to surprise: A composer-audience architecture. Proceedings of the 10th International Conference on Computational Creativity (ICCC'19) (pp. 41–48).
- Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, 67(6), 380–400. <https://doi.org/10.1037/h0040373>
- Corazza, G. E. (2016). Potential originality and effectiveness: The dynamic definition of creativity. *Creativity Research Journal*, 28(3), 258–267. <https://doi.org/10.1080/10400419.2016.1195627>
- Cummings, E. E. (1994). In Richard S. Kennedy (Ed.), *Selected poems. With introduction and commentary by Richard S. Kennedy* Liveright.
- Davies, J. H., & MacNaughton, I. (2020). *Monty Python's flying circus: The complete series* [DVD]. Universal Pictures Home Entertainment.
- Diederer, K. M., & Fletcher, P. C. (2021). Dopamine, prediction error and beyond. *The Neuroscientist*, 27(1), 30–46. <https://doi.org/10.1177/1073858420907591>
- Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, 9(1), 35. <https://doi.org/10.1037/a0038688>
- Dietrich, A. (2015). *How creativity happens in the brain*. Palgrave Macmillan.
- DiStefano, P. V., Patterson, J. D., & Beaty, R. E. (2024). Automatic scoring of metaphor creativity with large language models. *Creativity Research Journal*, 1–15. <https://doi.org/10.1080/10400419.2024.2326343>
- Eagleman, D., & Brandt, A. (2017). *The runaway species: How human creativity remakes the world*. Catapult.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks generating “art” by learning about styles and deviating from style norms. In A. Goel, A. Jordanous, & A. Pease (Eds.), *Proceedings of the 8th International Conference on Computational Creativity, ICCO*. Georgia Institute of Technology.
- Esling, P., & Devis, N. (2020). Creativity in the era of artificial intelligence. *arXiv preprint arXiv:2008.05959*. <https://doi.org/10.48550/arXiv.2008.05959>
- Feldman, M. (1985). *Essays* (W. Zimmermann Ed.). Beginner Press.
- Franceschelli, G., & Musolesi, M. (2023). On the creativity of large language models. *arXiv preprint arXiv:2304.00008*. <https://doi.org/10.48550/arXiv.2304.00008>
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies. *Cognition and Emotion*, 17(1), 25–40. <https://doi.org/10.1080/02699930302279>
- Ganguli, D., Hernandez, D., Lovitt, L., Askell, A., Bai, Y., Chen, A. & Clark, J. (2022, June). Predictability and surprise in large generative models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1747–1764). <https://doi.org/10.1145/3531146.3533229>
- Gilhooly, K. (2023). AI vs humans in the AUT: Simulations to LLMs. *Journal of Creativity*, 100071(1), 100071. <https://doi.org/10.1016/j.yjoc.2023.100071>
- Gjerdingen, R. O. (2007). Partimento, que me veux-tu? *Journal of Music Theory*, 51(1), 85–135. <https://doi.org/10.1215/00222909-2008-024>
- González-Díaz, J., & Palacios-Huerta, I. (2022). *AlphaZero Ideas* (SSRN Scholarly Paper 4140916). <https://doi.org/10.2139/ssrn.4140916>
- Greene, B. (1999). *The elegant universe*. W.W. Norton & Company.
- Guzik, E. E., Byrge, C., & Gilde, C. (2023). The originality of machines: AI takes the Torrance Test. *Journal of Creativity*, 33(3), 100065. <https://doi.org/10.1016/j.yjoc.2023.100065>
- Hiesinger, P. R. (2021). *The self-assembling brain: How neural networks grow smarter*. Princeton University Press.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. <https://doi.org/10.1109/MSP.2012.2205597>
- Hoel, E. (2022). *The banality of ChatGPT*. Retrieved on January 26, 2023. <https://erikhoel.substack.com/p/the-banality-of-chatgpt>
- Horn, K., & Huron, D. (2015). On the changing use of the major and minor modes 1750–1900. *Music Theory Online*, 21(1), 1–11. <https://doi.org/10.30535/mt0.21.1.4>
- Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), 3440. <https://doi.org/10.1038/s41598-024-53303-w>
- Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J. & Wald Fuhrmann, M. (2020).

- Cross-cultural work in music cognition: Challenges, insights, and recommendations. *Music Perception*, 37(3), 185–195. <https://doi.org/10.1525/MP.2020.37.3.185>
- Jobs, S. (2007). *Transcript – iPhone keynote 2007*. European Rhetoric. <http://www.european-rhetoric.com/analyses/ikeynote-analysis-iphone/transcript-2007/>
- Jones, R. (2023, October 18). *How Sam Altman got it wrong on a key part of AI: ‘Creativity has been easier for ai than people think* [Fortune]. <https://fortune.com/2023/10/18/openai-sam-altman-misjudged-ai-creativity-chatgpt/>
- Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask ‘why’ questions of minds and brains. *Trends in Neurosciences*, 46(3), 240–254. <https://doi.org/10.1016/j.tins.2022.12.008>
- Kasparov, G. K. & King, D. (2000). *Kasparov against the world: The story of the greatest online challenge*. Incorporated. Kasparov Chess Online.
- Kounios, J., & Beeman, M. (2009). The Aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, 18(4), 210–216. <https://doi.org/10.1111/j.1467-8721.2009.01638.x>
- Lamm, K. (2017). “The will to adorn”: Nick Cave’s soundsuits and the queer reframing of black masculinity. *Critical Arts*, 31(3), 35–52. <https://doi.org/10.1080/02560046.2017.1383494>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT- related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2), 100017. <https://doi.org/10.1016/j.metrad.2023.100017>
- Macklay, S. (2014). *Many, many cadences for string quartet*. Edition Peters.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. <https://doi.org/10.1037/h0048850>
- Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: Tutorial. *Journal of Medical Internet Research*, 25, e50638. <https://doi.org/10.2196/50638>
- Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.
- Miller, A. I. (2019). *The artist in the machine: The world of AI-powered creativity*. MIT Press.
- Moro, A., Greco, M., & Cappa, S. F. (2023). Large languages, impossible languages and human brains. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 167, 82–85. <https://doi.org/10.1016/j.cortex.2023.07.003>
- Moszynska, A. (2020). *Abstract art (world of art)*. Thames & Hudson.
- Nemerov, A. (2021). *Fierce Poise: Helen Frankenthaler and 1950s New York*. Penguin Books.
- Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, 49, 101356. <https://doi.org/10.1016/j.tsc.2023.101356>
- Orwig, W., Edenbaum, E. R., Greene, J. D., & Schacter, D. L. (2024). The language of creativity: Evidence from Humans and large language models. *The Journal of Creative Behavior*, 58(1), 128–136. <https://doi.org/10.1002/jocb.636>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P. & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
- Page, S. E. (2019). *The diversity bonus: How great teams pay off in the knowledge economy*. Princeton University Press.
- Patterson, J. D., Barbot, B., Lloyd-Cox, J., & Beaty, R. E. (2023). AuDrA: An automated drawing assessment platform for evaluating creativity. *Behavior Research Methods*, 1–18. <https://doi.org/10.3758/s13428-023-02258-3>
- Plain, A., Eynon, R., Hjorth, I., & Osborne, M. A. (2022). *AI and the arts: How machine learning is changing artistic work. Report from the creative algorithmic intelligence research project*. <https://www.oii.ox.ac.uk/news-events/reports/ai-the-arts/>
- Rafner, J. Beaty, R. E. Kaufman, J. C. Lubart, T. & Sherson, J. (2023). Creativity in the age of generative AI. *Nature Human Behaviour*, 7(11), 1836–1838.
- Rawte, V. Sheth, A. & Das, A. (2023). A survey of hallucination in large foundation models. arXiv preprint arXiv:2309.05922.
- Ross, A. (2007). *The rest is noise: Listening to the twentieth century*. Macmillan.
- Runco, M. (2023). AI can only produce artificial creativity. *Journal of Creativity*, 33(3), 100063. <https://doi.org/10.1016/j.yjoc.2023.100063>
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24(1), 92–96. <https://doi.org/10.1080/10400419.2012.650092>
- Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., & Gadepally, V. (2023). “From words to Watts: Benchmarking the energy costs of large language Model inference.” 2023 IEEE High Performance Extreme Computing Conference (HPEC), Boston, MA, USA (pp. 1–9). <https://doi.org/10.1109/HPEC58863.2023.10363447>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Schubert, F. (1973). *Complete chamber music for strings*. Dover Publications.
- Shanahan, M. (2022). Talking about large language models. arXiv:2212.03551 [cs.CL].
- Simonton, D. K. (2012). Taking the U.S. Patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal*, 24 (2–3), 97–106. <https://doi.org/10.1080/10400419.2012.676974>
- Simonton, D. K. (2018). Defining creativity: Don’t we also need to define what is not creative? *The Journal of Creative Behavior*, 52(1), 80–90. <https://doi.org/10.1002/jocb.137>
- Sosa, R. (2019). Accretion theory of ideation: Evaluation regimes for ideation stages. *Design Science*, 5, e23. <https://doi.org/10.1017/dsj.2019.22>
- Stein, M. I. (1974). *Stimulating creativity, Vol I*. Academic Press.
- Stokes, P. D. (2005). *Creativity from constraints: The psychology of breakthrough*. Springer Publishing Company.

- Suri, G., Slater, L. R., Ziaee, A., & Nguyen, M. (2024). Do large language models show decision heuristics similar to humans? A case study using GPT-3.5. *Journal of Experimental Psychology: General*, 153(4), 1066–1075. <https://doi.org/10.1037/xge0001547>
- Turner, M., & Fauconnier, G. (2003). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- van Beethoven, L. (1989). *Symphonies nos. 8 and 9 in full score*. Dover Publications.
- Vogelstein, F. (2013). And then Steve Said, “Let There Be an iPhone. *The New York Times Magazine*, 4.
- Zahavy, T., Veeriah, V., Hou, S., Waugh, K., Lai, M., Leurent, E. & Singh, S. (2023). Diversifying ai: Towards creative chess with alphazero. *arXiv preprint arXiv:2308.09175*.

Appendix 1: Transcript of the author's conversation

with ChatGPT-3.5 on November 9, 2023

USER QUERY:

What would happen if Ancient Greek philosophers played against 19th century German philosophers in the "Philosopher's Soccer Championship"?

CHAT GPT-5

The idea of Ancient Greek philosophers playing against 19th-century German philosophers in a "Philosopher's Soccer Championship" is a humorous and imaginative scenario, but it's important to note that this is purely a fictional and impossible scenario, as time travel and a soccer tournament involving historical figures are not possible in reality. However, if we were to entertain this imaginative idea for fun, we could speculate on what might happen.

1:Style of Play

- Ancient Greek Philosophers: The Ancient Greek philosophers may bring a more classical and strategic approach to the game. They might prioritize wisdom, intellect, and philosophical discussions over physical prowess. Socratic questioning could be their primary tactic.
- 19th-Century German Philosophers: The 19th-century German philosophers might approach the game with a more systematic and methodical approach, drawing from their philosophical traditions. They could employ tactics inspired by dialectics and Hegelian synthesis.

2:Team Composition

- Ancient Greek Team: Socrates might be the team's captain, using his method of questioning to outwit opponents. Plato and Aristotle could contribute their analytical skills to create well-structured strategies. Diogenes might be the goalkeeper, using his cynicism to distract the opposing team.
- German Team: Immanuel Kant could serve as the team's moral compass, while Friedrich Nietzsche might bring a flair for the dramatic. Georg Wilhelm Friedrich Hegel could provide a systematic approach to game tactics, and Arthur Schopenhauer could contribute his pessimistic outlook.

3:Refereeing and Fair Play

- Ancient Greek philosophers might prioritize fairness and justice in the game, influenced by their ethical philosophies. They could engage in discussions with the referee and opponents about the nature of fairness.
- 19th-century German philosophers might debate the very concept of rules and morality during the game, potentially leading to philosophical disputes and existential crises on the field.

4:Outcome

- Given the highly speculative and whimsical nature of this scenario, predicting the outcome is impossible. The game might be a chaotic blend of philosophical debates, Socratic irony, Kantian ethics, Hegelian dialectics, and Nietzschean will to power. The final score may not even be relevant.
- In reality, such a scenario is amusing to imagine but entirely fictional. Soccer is a physical sport that requires athleticism and skill, whereas philosophy is an intellectual pursuit. Mixing the two in a "Philosopher's Soccer Championship" would result in an absurd and comical spectacle rather than a competitive match.

Appendix 2: Follow-up query of ChatGPT-5 on the same day

USER QUERY:

If Ancient Greek philosophers were to play 19th century German philosophers in a soccer match, what is the most absurd thing that could happen?

CHATGPT-5

Imagining Ancient Greek philosophers playing a soccer match against 19th-century German philosophers is certainly an amusing and absurd concept. Here's one humorous scenario that could take place:

- Imagine the Ancient Greek philosophers, such as Socrates, Plato, and Aristotle, trying to engage in a deep philosophical discussion in the middle of the soccer field while the game is in progress. They might be so engrossed in their debate about the nature of reality, ethics, or the ideal form of government that they completely disregard the rules and objectives of the game. Meanwhile, the 19th-century German philosophers, like Immanuel Kant, Friedrich Nietzsche, and Georg

Wilhelm Friedrich Hegel, could be arguing passionately about their philosophical ideas, leading to a chaotic and heated exchange.

- As the players become more engrossed in their philosophical debates, the ball would likely be forgotten, and the match would come to a standstill. The absurdity of this scenario lies in the clash of philosophical discussions with the fast-paced nature of a soccer match, where the primary goal is to score goals, not engage in deep philosophical pondering.
- Of course, this is a playful and fictional scenario, and it's unlikely that such a match could ever take place in reality.