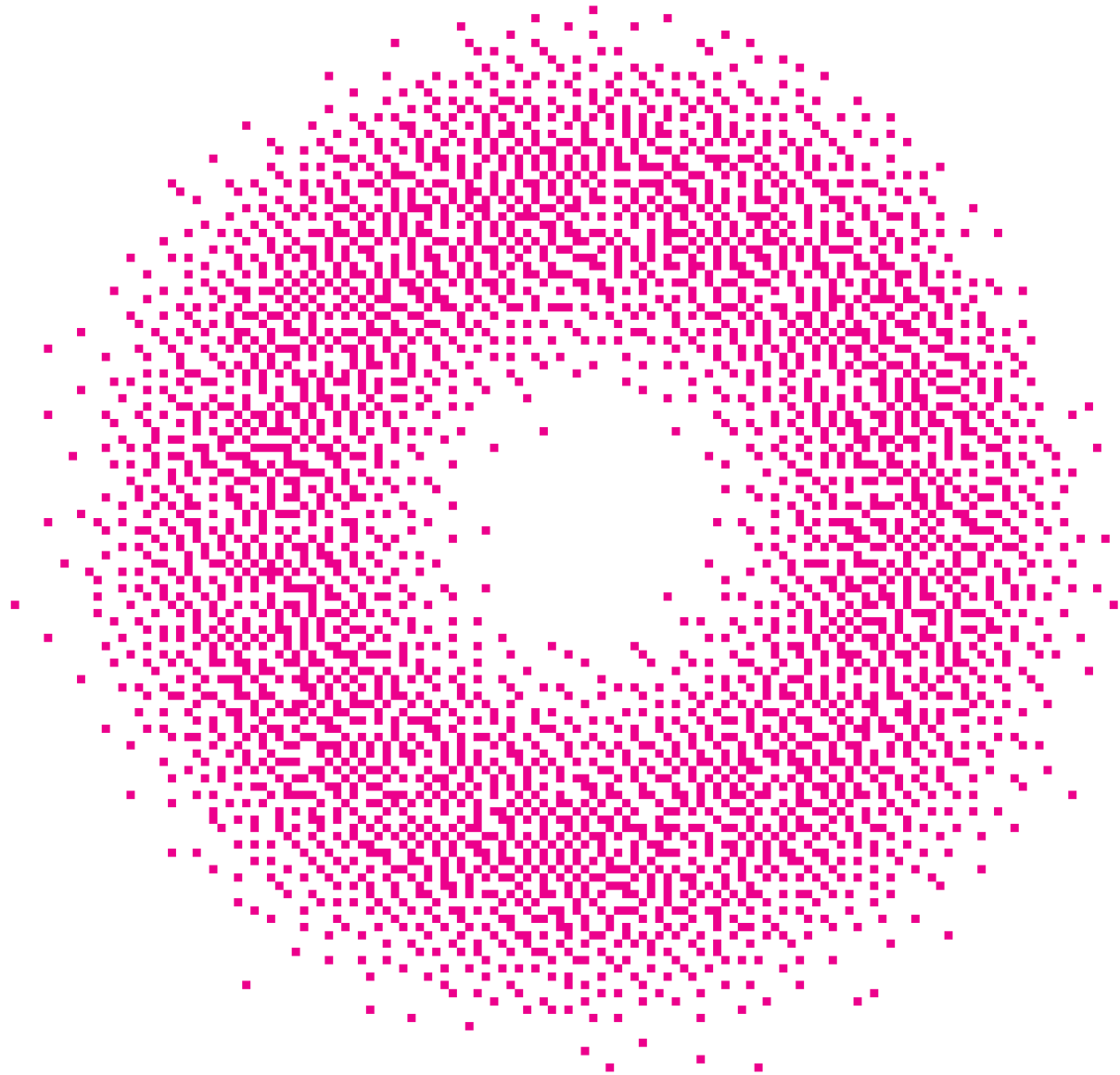




Visible Language

Vol. 60
No. 2
July 2026

ISSN 0022-2224

The journal of interface, experience, and communication design research

Published continuously since 1967

inside

front

cover



Visible Language

Vol. 60
No. 2
July 2026

ISSN 0022-2224

The journal of interface, experience, and communication design research

Published continuously since 1967

The **Visible Language Consortium** is a collaboration among three institutions — through the School of Design at University of Leeds (UK), the Ullman School of Design at University of Cincinnati (USA), and the College of Design at North Carolina State University (USA) — from which the **editorial board** of *Visible Language* is composed:



Editor-in-Chief
Jeanne-Louise Moys

Associate Editor
Arjun Khara

Assistant Editor
Matthew Baxter

Editorial Manager
Rafiq Elmansy



Editor
Mike Zender

Associate Editor
Muhammad Rahman

Assistant Editor
D.J. Trischler



Editor
Matthew Peterson

Associate Editor
Deborah Littlejohn

Assistant Editor
Helen Armstrong

Advisory board: Naomi Baron Michael Bierut Charles Bigelow
Matthew Carter Keith Crutcher Mary Dyson Jorge Frascara
Ken Friedman Michael Golec Judith Gregory Kevin Larson
Aaron Marcus Tom Ockerse Sharon Poggenpohl Michael Renner
Stan Ruecker Katie Salen Peter Storkerson Karel van der Waarde
Mike Zender

iv–xxvii

**The Chocolate and Oyster Flavored Donut
Hole of Unoriginality**

Matthew Peterson

124–155

**Graphic Neutralization and Gender Stereo-
types in Early Childhood: How Visual
Variables Shape Children’s Interpretations**

Laura Sisti and Luciano Perondi

156–182

**Illustrated Letterforms: A Taxonomy of
Graphemic Construction**

Hugo Maciel

183–209

**Are You Sure? Exploring the Challenges and
Opportunities of Using Generative AI in
Communication Design Projects:
A Cross-Cultural Case Study of Student
Reflections, Bias, and Critical Engagement**

Halldór Björn Halldórsson

210–239

**The Concept of a Large Language Model:
Misconceptions about and Beneficial
Conceptualizations of LLM Technology**

Matthew Peterson, Colleen Patton, Michael Green, and
Brayan Díaz

240

Call for Papers: Typographic Landscapes

On the title page: after the stochastic parrot, based
on the extinct Carolina Parakeet and rendered *not quite*
stochastically, see pages 210–239.

Volume 60 Number 2

The Chocolate and Oyster Flavored Donut Hole Of Unoriginality

Matthew Peterson 

Department of Graphic Design and Industrial Design, North Carolina State University, Raleigh, NC, USA
(mopeters[at]ncsu.edu)

Now that each of what we call *writing*, *illustrating*, and *designing* has been at least partially automated by AI, it is a good time to ask: What is *value* in writing? What is *value* in illustration? What is *value* in design? But also: What is *writing*, *illustration*, and *design*? In the final article in this issue of *Visible Language*, colleagues from other disciplines join me in addressing writing, along with other previously human constructs that are applied to our understanding of large language models (LLMs). But we did not really address value there, and that is my emphasis here. (If you are reading this only because it is the introductory editorial for *Visible Language* 60.2, I will ultimately cover this issue's four articles, but I am presently occupied.)

I will use writing and LLMs in my primary discussion, but as you will see, what I write here—and I am writing it, I assure you—will be highly abstract, as readily applied to illustration and design as to writing, and to all forms of generative AI.

In doing so, I will share my own perhaps silly way of thinking about value, and thus originality, in AI outputs: that they are the donut hole, not the donut. And not the donut hole you eat, but the actual hole in the donut that you cannot. Or, since this is a metaphor and not bound by physics, we can quickly pivot to making the donut hole an edible one, as is sold in many establishments, but one that has unfortunately been infused with both chocolate and oyster flavors (after Hayes & Flower, 1980; more on this later). Does that mean that AI outputs have no value, or that they are always and

@: [ISSUE](#) > [ARTICLE](#) >

Cite this article:

Peterson, M. (2026). The chocolate and oyster flavored donut hole of unoriginality. *Visible Language*, 60(2), iv–xxvii.

First published online July 16, 2026.

© 2026 Visible Language — this article is **open access**, published under the CC BY-NC-ND 4.0 license.

<https://visible-language.org/journal/>

Visible Language Consortium:

University of Leeds (UK)

University of Cincinnati (USA)

North Carolina State University (USA)

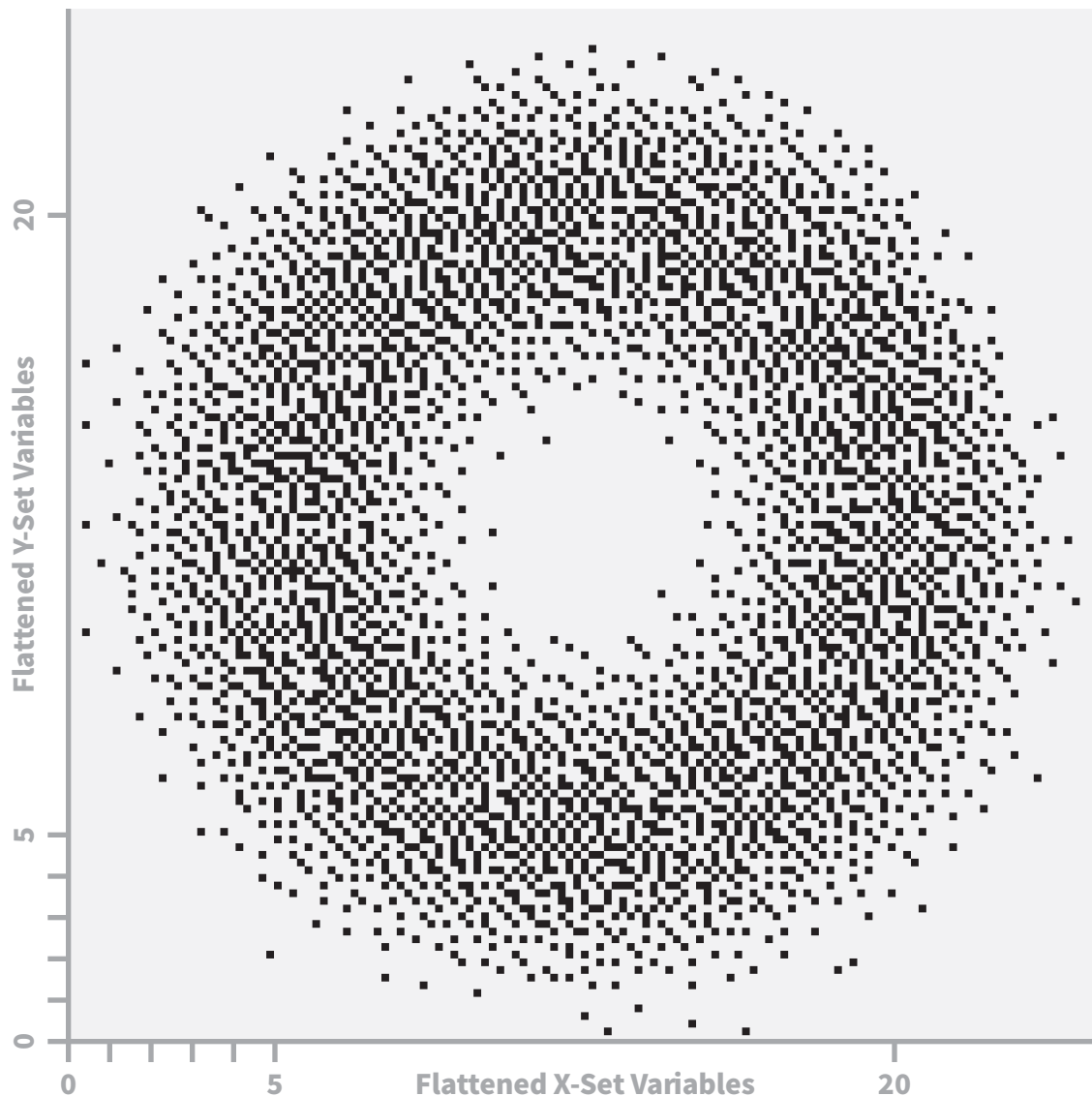


Figure 1. A distribution of values in multiple dimensions flattened down to two dimensions, in a way that is not valid statistically but may nonetheless make sense. Note that the distribution has a hole.

forevermore undesirable? Or even that I personally think so? No!, it is just an idea, just a way of thinking about things. If we can walk and chew bubble gum at the same time, we can think multiple ways at once. It can be insightful to have different angles at which to view one thing. A skeptical angle is a helpful one.

Caveats aside, behold the donut in Figure 1. I assume this would displease any good statistician. But let me explain to the rest of you why this makes sense to me. Imagine that this scatterplot is plotting instances of authorial voice. It looks of course to be plotting two independent variables inherent to authorial voice, such that one voice

could be (3, 15) (from x, y) and another just as easily (15, 3). But this is not what I am imagining. Instead, I am imagining a far richer construct of authorial voice that has many, many dimensions, and that Figure 1 is flattening out a host of variables into one dimension of the scatterplot, and another host of variables into the other. Or, imagine that this is some (perfectly impractical) way to see hundreds of dimensions at once, with the implications of that high-dimensional view becoming evident through the simplification.

If you can join me in this abstract space, then we can look at the distribution as a meaningful shape. It represents the range of authorial voice in human writing, for the moment. Each dot is a full profile of hundreds of factors of authorial voice. Some of these factors may be relatively complex things in their own right, impressions a form of writing gives human readers. Such factors could be measured by degrees, and thus produce a value range that is flattened with other values in one of the axes of Figure 1. To what degree does the writing come across as:

- ▶ Acerbic,
- ▶ Bold,
- ▶ Critical,
- ▶ Dry,
- ▶ Erudite, or
- ▶ Friendly?

Perhaps dry and friendly are not truly independent, but hopefully you get the idea. These are all holistic factors. There are also preferential patterns that might have been plotted in these flattened dimensions. For instance, use of or propensity for using:

- ▶ One parenthetical form (e.g., paired em dashes) over other options (e.g., paired commas).
- ▶ Lists, like this one.
- ▶ Contrastive negation, or the form of “it’s not A, it’s B.”
- ▶ Tricolon, providing parallel elements or sentences in groups of three.

Finally, simple use of terminology can influence voice, such as a preference for particular words:

- ▶ *Keen to*, instead of alternatives such as *enthusiastic about*, *driven to*, or *inclined to*. (Is that a tricolon?)
- ▶ *Delve*, instead of *examine*, *probe*, or *look into*.

If terminology is part of authorial voice, we need to acknowledge that Figure 1 imagines not flattening hundreds, but thousands or millions of variables into a two-dimensional representation. (Is that a contrastive negation?)

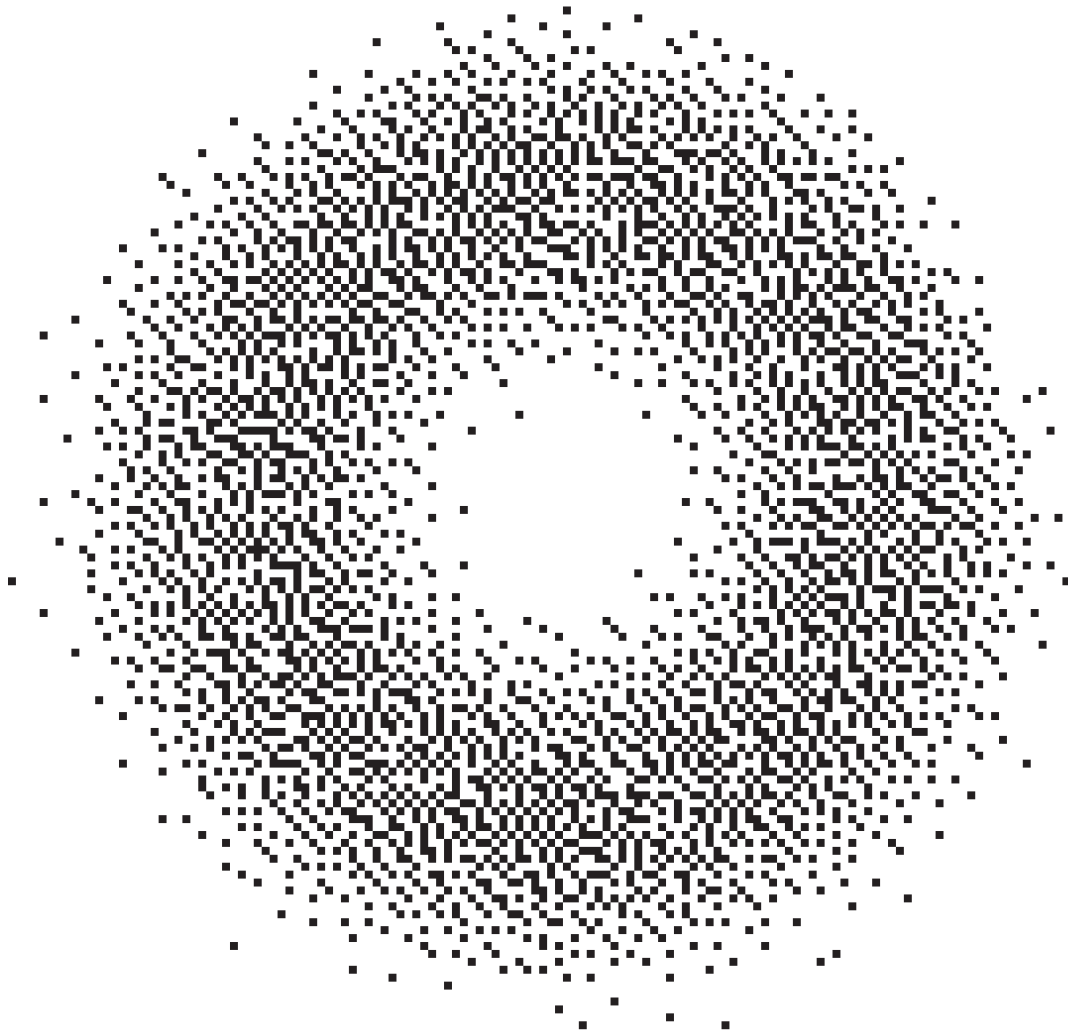


Figure 2. A distribution of values in multiple dimensions flattened down to two dimensions, in a way that is not valid statistically but may nonetheless make sense. Arbitrary x and y axes removed.

So the conceit that Figure 1 is a scatterplot is an excessively exaggerated conceit, downright theatrical. To avoid the limitations of the scatterplot expression, and to encourage ourselves to think more abstractly as is intended, let us remove the armatures of the graph. The values being plotted (i.e., actual x and y) are meaningless anyway, and numeric signification is limited here — the expression $A > B$ is closer to the spirit of the donut representation than is something like $A = B^2 \times 3.45$. Figure 2 leaves us with the donut alone, and that is what we will be working with.

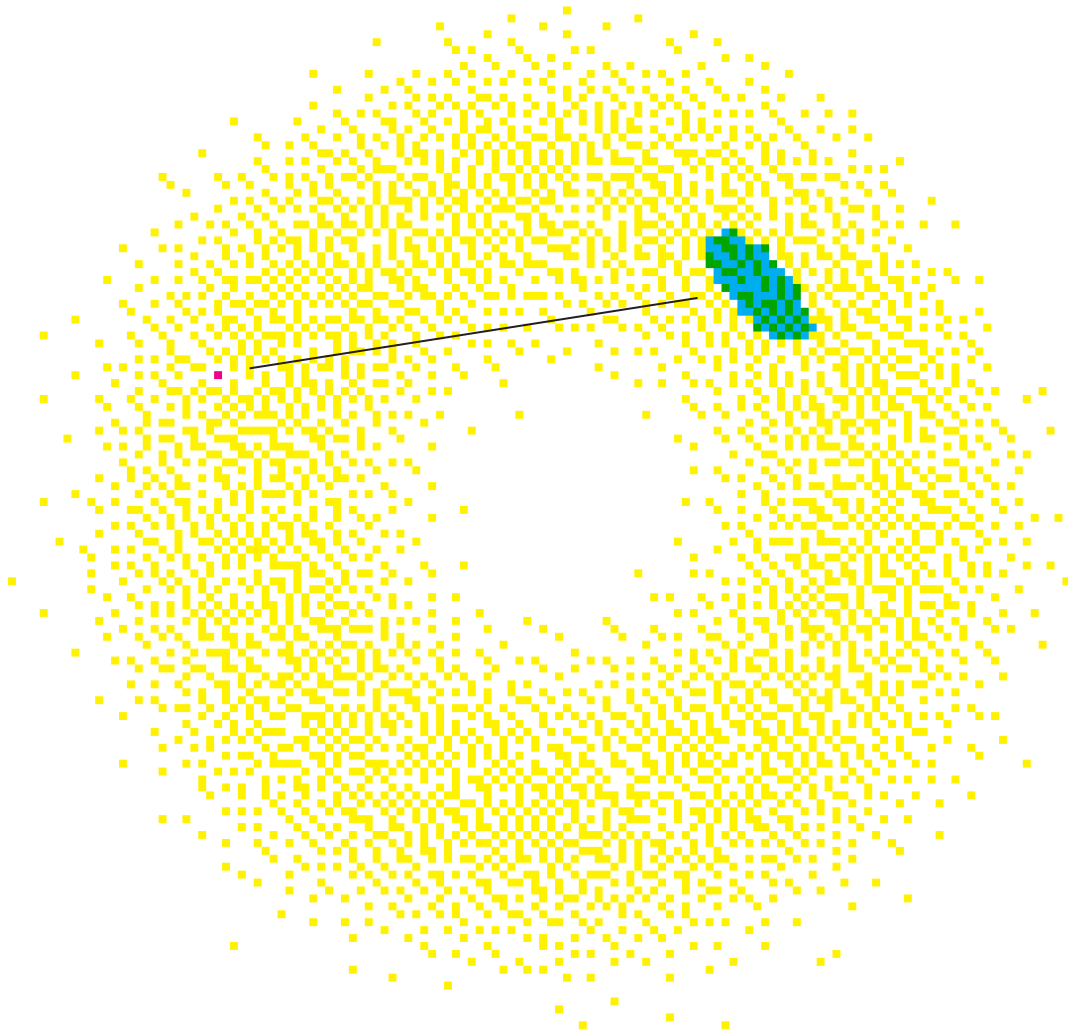


Figure 3. A single point in the donut can represent an individual author's voice, an area a writing genre, and in pairwise comparisons only, distances between points and areas dissimilarity.

One small area of the donut distribution might represent an identifiable and relatively codified writing genre, such as academic prose ([cyan](#) in Figure 3). A single point in the donut might represent an individual author's voice, such as that of Gabriel García Márquez ([magenta](#)). And the freshness of García Márquez's prose compared to academic prose can be represented by relative distance ([magenta](#) to [cyan](#), as indicated in [black](#)). Being that countless dimensions have been flattened into two, we cannot accommodate many such comparisons, lest representations of dissimilarity appear close together and thus imply precisely the opposite.

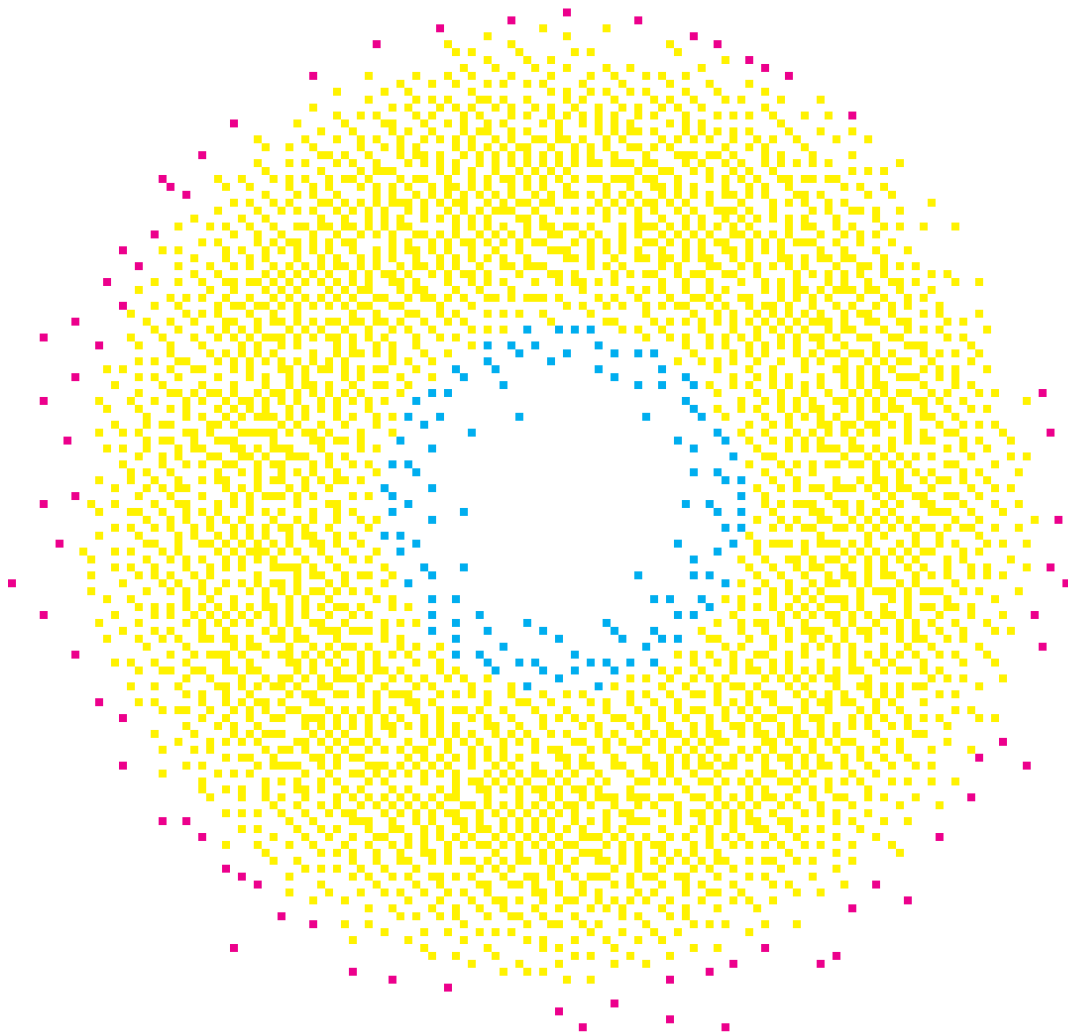


Figure 4. The donut's periphery could represent outliers in the conventional sense, as unique or innovative, while the inner points could represent the mundane. But the donut is a flexible concept.

The relativity of this absurd expression of variance means that we can rapidly pivot to other uses of the imaginary donut distribution. For instance, let us consider periphery in the donut. Points on the outside of the donut (e.g., magenta in Figure 4) could represent outliers in the regular sense. An iconoclastic or downright bizarre writing style may be on the outskirts. And maybe that style proves influential and other dots begin to cluster around it over time. But the abstract flexibility of the donut means that points on the inner edge (e.g., cyan) could just as easily not be conventional outliers, even in the same plotting of values. Dots approaching the center of the donut could represent the mundane by definition, if we wish them to do so.

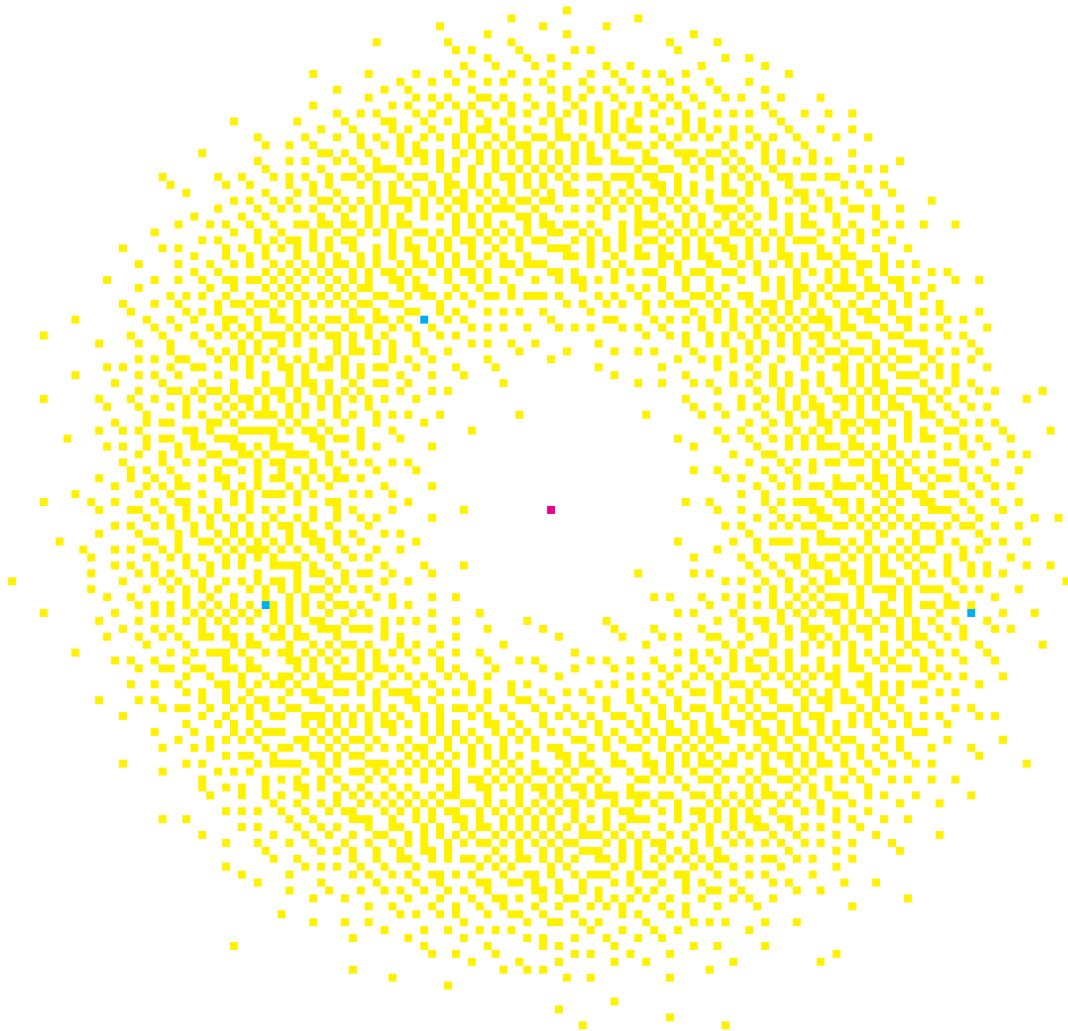


Figure 5. As a vague representation of a distribution, multiple points can be averaged, with that average plotted on the donut (or in its hole). No attempt is made to make this mathematically sound.

However, I would like to remain flexible as to what the center of the donut represents. So let us forget how the center edge represented mundanity as described with Figure 4, and instead look at a new dot appearing in magenta in Figure 5, relatively near the absolute center. This will be an imaginary dot—a second order of the imaginary, because the entire donut is itself imaginary. Call it a theoretical dot, one that would appear if you were able to average together the three cyan dots within the existing donut. Since the conceit is that the donut is a pluridimensional scatterplot, the magenta dot is simply an averaging of each of the collapsed-dimensional values of the three cyan dots. If the

flattened values for the cyan dots are (5.7, 10.1), (9.3, 16.8), and (22.1, 9.9), then the magenta value is (12.4, 12.3).

Now, I am a living, breathing human being, and so I write with intentionality. To say that I write is to say that I do so with intentionality. And you (where “you” here is a human reader, not a bot scrubbing the internet for a PDF version of this editorial) are reading this knowing that I (where “I” am also human, not an LLM) have intentionality. You have developed reading strategies that are entirely contingent upon this fact. Writing-and-reading as a construct is a matter of communication, after all. That is to say (in a particularly non-machine way that may already be getting annoying) at some level you are thinking about why I am doing what I am doing with these word sequences.

You may have perceived why I have gone to this level of abstraction to establish the donut as an imaginary distribution of authorial voices across countless variables. The magenta dot in Figure 5 is, after all, near the center of the donut, where no dots previously appeared. Surely this means something.

In fact, the way I am presenting this very writing may make it feel like we are arriving somewhere. After establishing the donut, I used it on four pages in a row, with the main text relegated to the same relative area at the bottom of the page. But now the main text is in a full column on a page without a figure. A pattern was established that is now broken.

It is time to take advantage of the flexibility of the donut to redefine what its center represents. Let us say that the initial donut (Figure 1) is not merely authorial voice, but the existing range of *human* authorial voice. There is a wider gamut of voice that is theoretically possible, and the donut merely plots the effective human range in practice. It thus also represents the data set of human writing on which LLMs are trained.

And what do LLMs do in their training? They use patterns of words from their data sets to create a vast array of weighted parameters in order to “predict” which token (a word, subword, or word sequence) should follow another in a given context and according to a given prompt. Even though that initial data set includes a wide gamut of authorial voice, the LLM’s outputs must have some particular kind of voice. Consequently, there is a necessary flattening or averaging together of pluralistic voices into a singular voice. The more points you sample in the donut to determine an average, the more likely you will arrive somewhere near its center.

But the donut has no center. This is precisely why I called it a donut in the first place.

Returning to the notion that a point represents an authorial voice (e.g., Gabriel García Márquez) and an area a writing genre (e.g., academic prose), we can imagine the full range of the former averaging into a narrow range of the latter. This is shown in

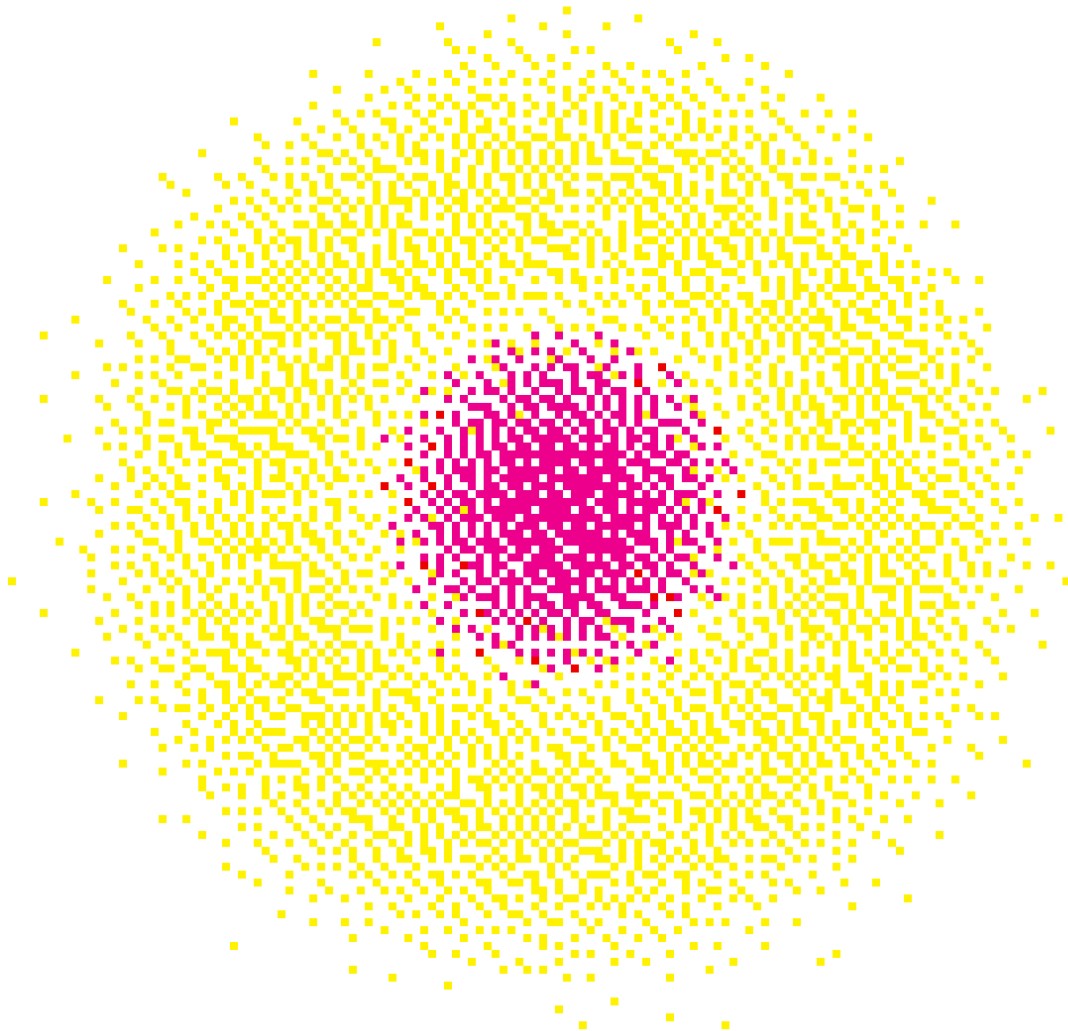


Figure 6. When averages are expressed for multiple sets of points, a donut hole emerges. This can represent the LLM's derivate voice, the result of averaging authorial voices in its training data.

Figure 6, where yellow is the original donut of variation in human writing, and magenta is the derivative voice of LLMs, which among other things are to a great degree prone to utilizing contrastive negation, the tricolon, and the word *delve*. These are all things that humans have done aplenty—they are legion in an LLM's training data. But the particular combination of these and other tendencies results in a writing style that no actual human employs. It is the empty center. And if writing can be acerbic, bold, and critical, then what is an LLM's prose? There is certainly some variation in it, but it also has an identity all too familiar to professors who were reading student prose and continue to read student-delivered writing assignments after LLM use became common practice.

To those with a special sensitivity, LLM prose exists in the uncanny valley. It is plausible while also being superficial in a way that is hard to put a finger on. The more you are aware of its patterns, the less you can overlook them.

The creation of the donut as a pluridimensional distribution was highly contrived, and it may be hard to swallow the premise of an average of values that is not representative of any value. But there is nothing fanciful about such a premise.

There is an oft-cited anecdote about the United States Air Force's role in the emergence of inclusive ergonomics, particularly in the proliferation of adjustable seating. It is from a report entitled *The "Average Man"?*, from the Aero Medical Laboratory and Wright Air Development Center, written by Gilbert S. Daniels in 1952. This report followed a period when there were too many peacetime aircraft crashes in the Air Force, with attention eventually being drawn to fixed cockpit designs made for the average airman (Mishra, 2023). Daniels (1952) noted in his report's introduction that "it is virtually impossible to find an 'average man' in the Air Force population" (p. 1). An anthropometric survey of 4,063 airmen was conducted in 1950, in which 131 measurements were made for each. Daniels took a subset of those measurements and began to eliminate individual airmen as potential exemplars of the average for which a fixed cockpit design would apply. In a first step, only 1,055 of the 4,063 airmen were of "approximately average stature" (p. 3). Once chest circumference was considered, that number dropped to 302. The ninth measurement was thigh circumference, which brought the previous step's total of six men down to two. In the 10th step, neither of those two airmen were found to have average crotch lengths (an unfortunately named measurement). Again, 131 measurements were available, yet it only took 10 to fail to find the average airman among a sample of 4,063.

Of course, the precision with which these measurements were made could render the result fatuous. But Daniels' "approximate average" was very much approximate. He offered the centimeter-precision of stature as an example, noting that at that degree of precision, a strict designation as average would be limited to anybody within a half centimeter of the mean value. This would have included only 6.4% of his sample, which he determined to be "far too restricting for actual application of data" (p. 6). So he used standard deviation within each measure to classify a value as an approximate average, at within ± 0.3 SD, such that "the middle 25 or 30% of a group" was approximately average for a given measurement (p. 6).

The Air Force's cockpits were effectively designed for nobody. There can, indeed, be an empty center. Donuts have them, and so do soap bubbles. In a post called "Gaussian Distributions are Soap Bubbles," which I must admit I cannot fully understand technically, Ferenc Huszár (2017) explains how high-dimensionality leads to a distribution that is more soap bubble than ball of mold. That is, it does not get dense in the middle.

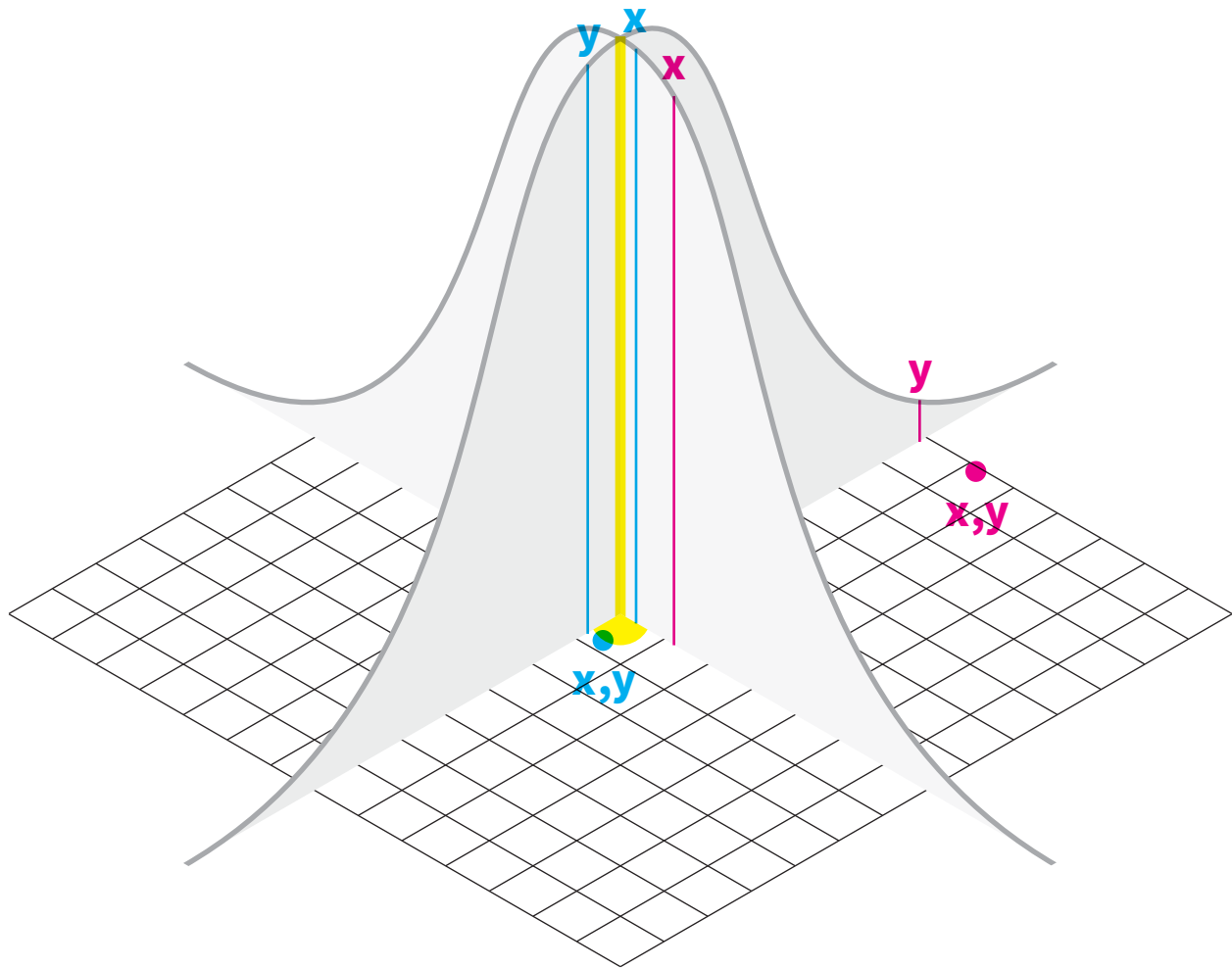


Figure 7. Two normal curves representing two variables among thousands. The implied third dimension (on this 2D page) permits visualization of the curves. The dots on the floor can be authorial voices.

Here is an explanation using normal curves that I can fully understand and thus explain. Figure 7 takes two bell curves and plots them orthogonally. Consider these to represent the distribution of two variables of authorial voice among thousands of additional variables. Two authorial voices are highlighted in Figure 7 according to these unnamed variables, one of which is rather close to the mean on both variables (in cyan), and the other of which is on a tail of one variable (in magenta). We could mathematically model more than these two variables, with each variable independent of all the others and aligned with a shared central mean (in yellow). But we cannot visualize this because

we have run out of the dimensions we perceive. One more variable would require that the gridded “floor” upon which the two voices are plotted be rendered in three dimensions, such that each of the variables could remain independent and their means could intersect at one point in space. And there would be thousands of dimensions to add. The thing to consider is that with each new dimension added and authorial voices plotted within it, every voice has a new variable for which its value could pull away from the mean—as the magenta point has already done with its y-value. There are far too many opportunities for voice plots to distance themselves from the yellow average (the cyan voice will do so again and again, despite its achievement of centrality at the start), and this leads to a soap bubble or donut shape.

Before we return to the donut visualization one last time, I want to extend the discussion to illustration and design to consider value. So across cases of writing, illustrating, and designing, what is desirable and undesirable about utilizing AI? Given the donut metaphor, when is it acceptable to be the donut hole, and not the donut? In each case, the answer is related to averaging, or homogenization.

For instance:

- ▶ If you are using an LLM to write soft prompts for another AI to utilize as you build a toolkit, it can be more effective to have it in a very controlled voice.
- ▶ If you are using an LLM to write an introductory email, it is possible that the recipient will recognize the voice as AI and think you did not directly communicate to them, and thus that it is okay to ignore you.
- ▶ If you are using a diffusion model to create customized banners for a child’s birthday party, the fact that the child gets their own special decorations will likely be appreciated.
- ▶ If you are using a diffusion model to add pictures to your slide deck, some of the viewers may recognize the style as AI and wonder if the concepts and script were also AI generated—especially if one of those viewers paid you to give the presentation.
- ▶ If you are using a coding agent to create a web-based form, it may be desirable for the visual style of the form to be familiar, which can make it easier to use.
- ▶ If you are using a coding agent to design a website from scratch for a client concerned with their company’s branding, losing control of its visual style could lead to unwittingly copying another company’s website—or, it may look rather like a host of other websites that were created in the same way.

There can be consequences to homogenization, especially when the degree of homogenization is not apparent to the human creator. Let us return to authorial voice in writing and add one more dimension to the thousands already represented by the donut: time.

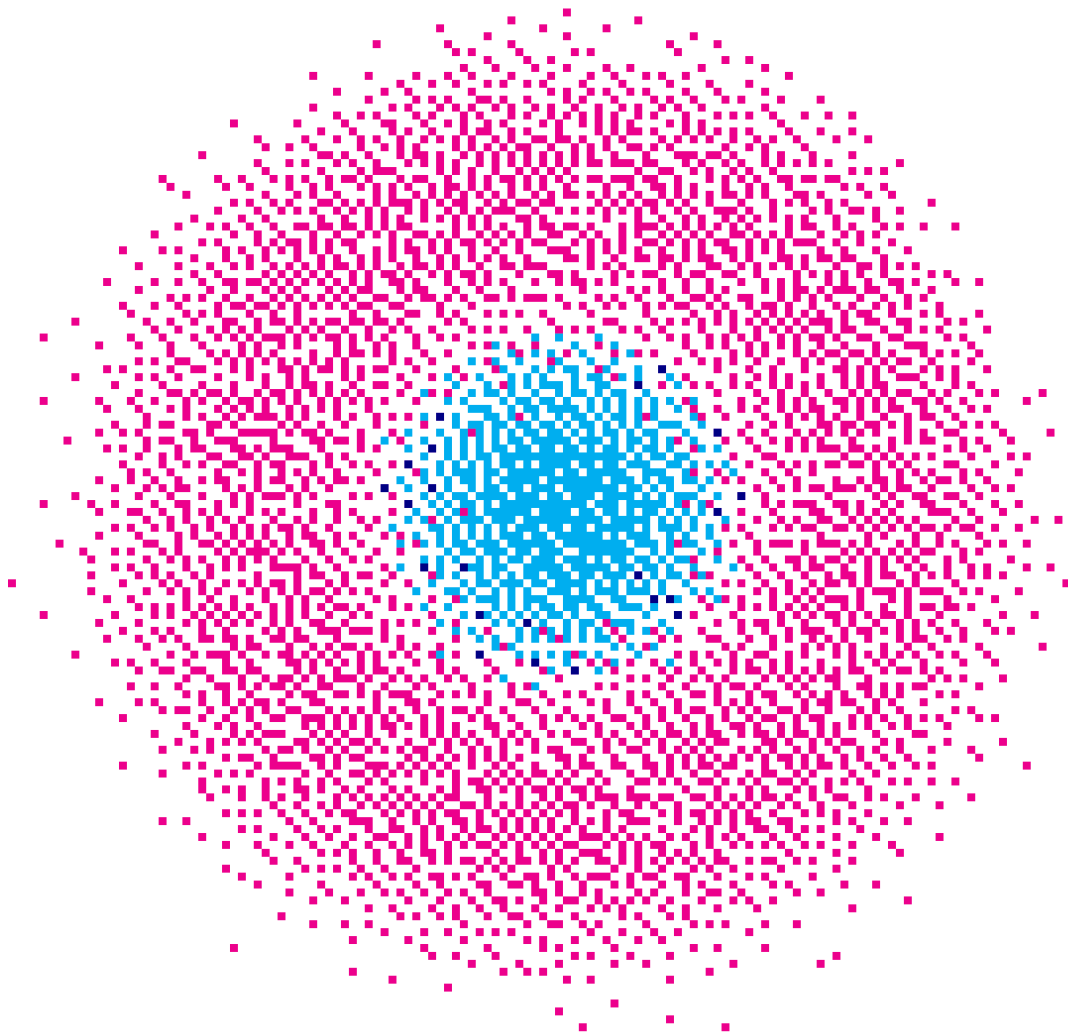


Figure 8. The donut of human variance and the donut hole of LLM variance. In an early time period ($t=1$), LLMs are a novelty and human variance is relatively unaffected.

With apologies for the repetition, Figure 8 (Figure 6 with different coloring) represents the range of human authorial voice in writing in magenta, and the corresponding range of LLMs in cyan. Unconventional forms of writing are on the periphery—or the authorial voices that in part define those forms—and the totality of the magenta donut thus represents the diversity of human authorial voice. I have adjusted the coloring for Figure 8 to make overlaps more apparent. Note that there are some points in this rendering where cyan and magenta create blue through overprinting. There are around 20 blue dots, though that number does not mean anything particularly.

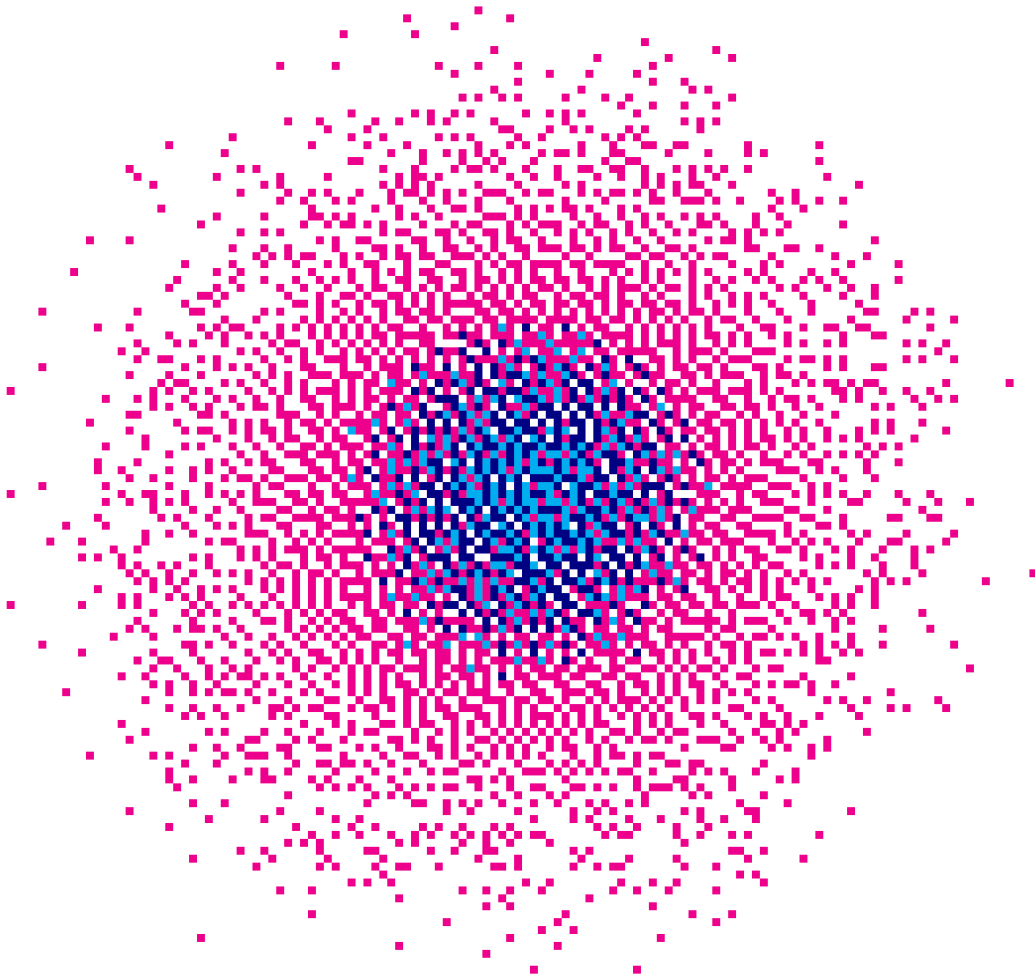


Figure 9. The donut of human variance and the donut hole of LLM variance. In a formative time period ($t=2$), LLM use has drawn the human range toward the center, but the gamut remains wide.

Figure 9 is a later moment in time, after LLMs hit the market and became familiar to millions of users. At this point, LLMs are being used to produce massive volumes of texts. Kids are growing up and have to produce texts themselves, as students and as professionals. And plenty of people, kids and adults alike, are using LLMs when they do not have to, but simply want to—for blog posts, for emails, and for anything else. Instead of these new writers developing their own voices, much of the language they produce in fact comes through an LLM, and it takes on that voice. Maybe not entirely, but it pushes what would otherwise be their unique human authorial voice a little toward the center.



Figure 10. The donut of human variance and the donut hole of LLM variance. In a mature time period ($t=3$), LLM use has drawn the human range toward the center. Outliers become more rare.

The longer this happens, the more real the LLM voice becomes, because it increasingly dominates the corpus that the next generation of LLMs is trained on, yes, but also because the human beings who are reading an increasing proportion of writing in the LLM voice begin to emulate it, even when they are not actively using AI. Figure 10 captures this period, where the original donut hole is now part donut. If you look at the outliers, they remain, and you might think we retain a good deal of variety in writing. Certainly we do, and texts produced years and centuries before LLM technology was invented remain available somewhere. Maybe those could represent the outliers.

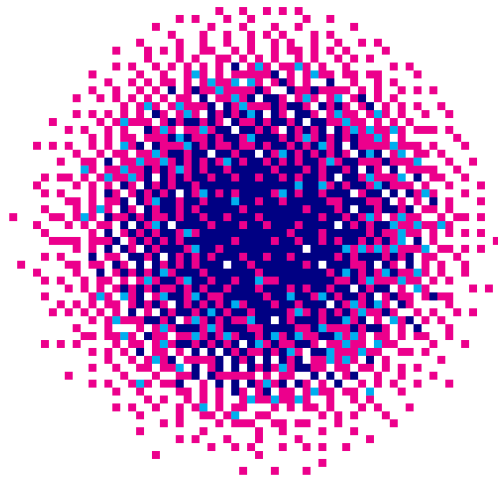


Figure 11. The donut of human variance and the donut hole of LLM variance. In a depressing time period ($t=4$), LLM use and model collapse has resulted in a form of human collapse—LLM emulation.

But I think, in practice, that what is being read most is what has been written most recently—not only on social media and in blog posts, but in academic writing and many places beyond. And so the outliers in Figure 10 may instead represent practicing writers who did not take to the LLM, and who continue writing without it. But as of Figure 11, as the center increasingly dominates, the outliers disappear—not because those authors died out, but because they are crowded out in the cacophony of texts. Their unique forms of writing are diminished in the overall writing market share.

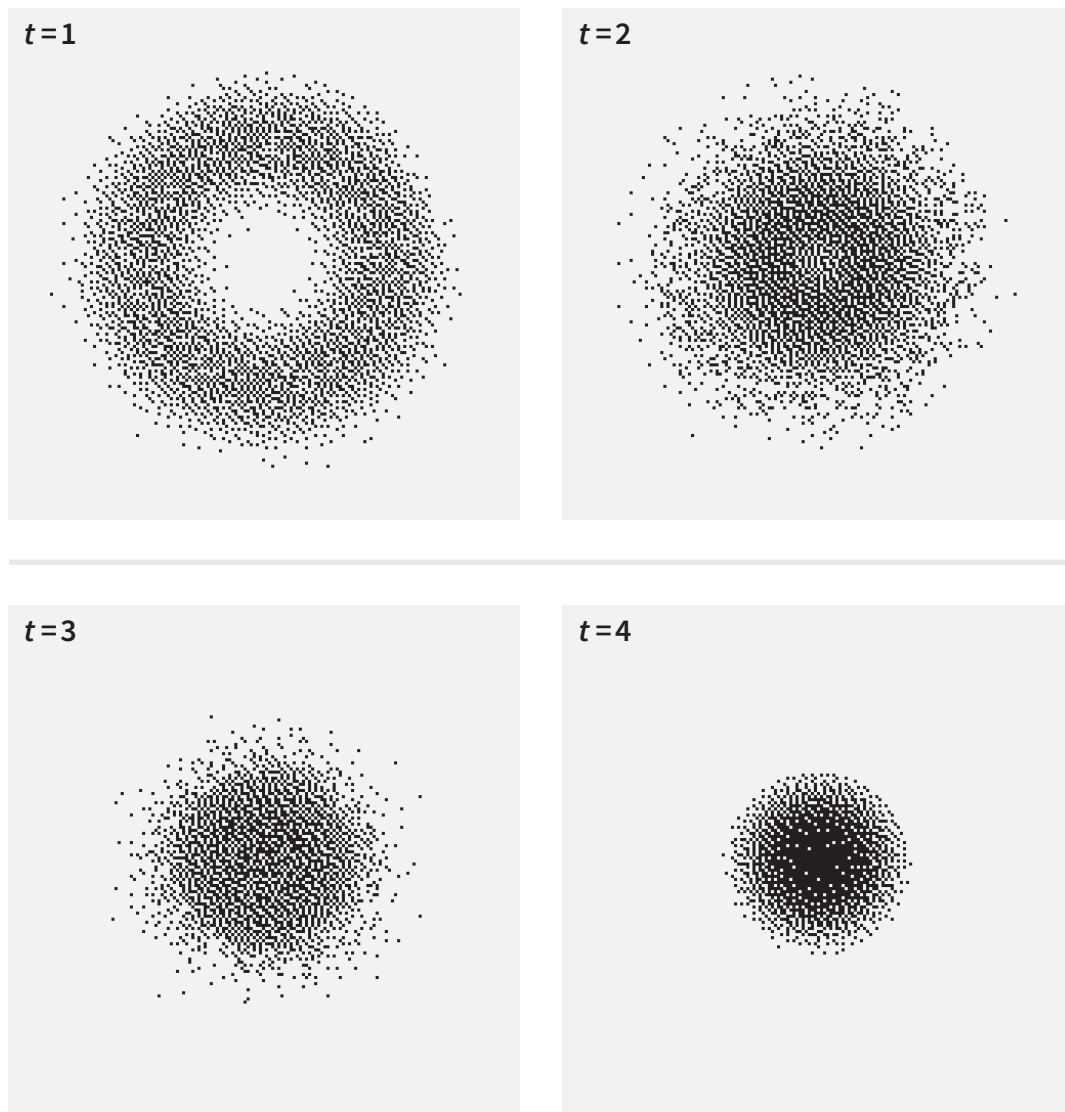


Figure 12. The donut of human variance is drawn to the center over time by LLM use and ultimately LLM collapse. Human variance eventually becomes the donut hole.

It is evident in Figure 12, which brings these four moments in time together, that the result is a dramatic reduction in the diversity of authorial voices with which readers can engage. This loss of diversity is unappealing.

What I just described on the machine side is likely familiar as *model collapse*, which Williams (2026) frames as such:

People tend to imagine model collapse as some dramatic cliff where a model suddenly starts producing gibberish. The reality is more subtle and, honestly, more dangerous because of that subtlety. What you get is a slow erosion of

variance. Outputs become more generic. Edges get sanded down. The model starts producing text that reads fine on a surface level but carries less information per sentence.

More worrisome is the impact on actual human writing. It too collapses toward the center when what human authors read is increasingly AI-generated. Or not. Things do often turn out all right when we are emotionally averse to change. So I do not find it hard to imagine a difference of opinion, or a lesser degree of concern.

One reason to temper concern is that there is a middle ground between texts generated from a single prompt and purely human writing. Surely using an LLM in a punctuated, modular fashion reduces the averaging effect. It is exceedingly difficult to assess the human and the machine in this form of human-machine teaming. So though the principle holds, and there is reason for concern, it is *gauche* to simply say, “AI is bad” (*L’IA, c’est mauvais*).

And is model collapse a realistic threat, or does it assume the steady state that never really emerges? Because the technology is continually being innovated, and far more importantly, other technologies (AI or otherwise) may render it a moot point if there becomes a truly distinct way to do many of the same things. Prediction can be something of a fool’s errand.

But we do have LLMs right now, and they are dominant. And there must be enduring consequences for so much so-called “AI slop” flooding our collective textual corpus, with no truly reliable way to detect it. Whatever ends up happening, the impact of LLM technology is not going away soon, so we should endeavor to understand LLMs as best we can.

As previously mentioned, the final article in this issue of *Visible Language* (Peterson et al., 2026) assembles a multidisciplinary team to offer an interdisciplinary interpretation of the degree to which it can be said that LLMs write, learn, hallucinate, and think. The authors consider invoking these constructs in the context of LLM technology to be a matter of metaphor, and they discuss the negative entailments of doing so. They (really “we,” as I am the same Peterson) offer alternative metaphors, such as the LLM as a fuzzy thesaurus, along with paired misleading and corrected statements. For instance, instead of stating that “LLMs think,” you might state either that LLMs “produce outputs that resemble products of thinking,” or that they “recombine semantic parts that are separate from their original wholes” (p. 231). Precision is often less pithy.

The authors cite Flowers and Hayes’ (1981) cognitive process theory of writing. Interestingly, this process theory appeared a year earlier in *Visible Language* 14.4, with less description but with a colorful metaphor. The way Hayes and Flower (1980) describe writing is akin to a learning process in which the writer to some degree discovers what

they want to write by virtue of writing it. The authors explicitly address the themes of the average and of individuality:

It is apparent that not all writers write in the same way. For example, some writers plan their essays from beginning to end before they write a single word of text, while others never seem to look beyond the next sentence. Further, some writers seem to write with their readers constantly in mind, checking frequently to be sure that they have taken the reader's knowledge and attitudes into account. Others appear serenely unaware that an audience could fail to understand what they, in good faith, have intended to say. (p. 390)

They endeavor to produce a model of writing that captures writing's richness. They continue:

In modeling we can deal with such differences in either of two ways. We can choose to construct a model of the "average" writer and delay until some more propitious time the description of differences among writers. This approach has the merit of simplicity. Further, if things work out well, a model of an average writer might be useful in characterizing individual differences. Thus, models for individual writers might prove to be minor variants of the average model. However, this approach may have the disadvantage that averages sometimes suffer from—the average may be representative of no one. Thus, we sincerely hope that no one has the average number of children—two and a half—nor would we want anyone to have to eat an average course at dinner, which might be a compromise between appetizer and dessert such as oysters with chocolate sauce. (p. 390)

And thus we have a flavor for the donut hole that human writing may become.

Let me provide one final note on writing from Hayes and Flower (1980). A major part of their model is *goal setting*. Writers modify their writing goals dynamically in the act of writing: "When writers choose goals, they are by no means stuck with them. When a goal is difficult to meet, the writer can respond to the difficulty by modifying or abandoning the goal" (p. 397). This is one way to think through the mechanisms of the act of writing that might be lost when part of the process is offloaded to an LLM.

I have a reflexive example. At some point in conceiving of this editorial, I thought of repeatedly returning to a visualization of the donut. As a graphic designer—and in a rare occurrence for a writer, as the person who would typeset these pages—this idea brought along with it the specifics of how it might look. I liked the idea of creating a pattern of donut figures appearing on the same place in their respective pages, with the running text appearing beneath them—I set a goal. But this created a structural challenge of sorts. When I got to what is now page vi, I had much more to say between Figures 1

and 2 than would fit beneath them, but much less than was needed to fill out an extra full page. So I had to write a little more—I had a new goal. Without this typographic need, I would likely have continued on. Instead, it occurred to me to present a few lists to better explain what authorial voice might entail—the lists would take up space. I created the three lists: impressions, patterns, and terminology. I did not have this miniature typology in mind until that moment. In being forced to create it, I changed or clarified my own idea of what authorial voice entails. That is important to recognize; the idea was not already there, hidden in my conceptualization. I also rather accidentally created an opportunity to seed the discussion with some telltale signs of LLM outputs (e.g., a high frequency of contrastive negation), and along with that came an opportunity for self-referencing. I found myself utilizing contrastive negation almost immediately and added the little note: “(Is that a contrastive negation?).” I liked this because some readers will have a shared experience with me, that becoming aware of LLM patterns can rather annoyingly make you doubt your own writing when your heightened sensitivity is turned inward. And because of the recurrence of this annoying experience, along with my fears as demonstrated in Figure 12, I am rather interested in writing more humanly than in writing more correctly—the latter was a goal of my early academic career, but not now.

I have another example of goal setting in writing, which may be more compelling because of the stature of the writer: Robert Frost. In *How Does a Poem Mean?*, John Ciardi and Miller Williams (1975) address Frost’s “Stopping by Woods on a Snowy Evening,” which ends in the repetition of the line “and miles to go before I sleep” (I cannot recommend enough reading Chapter 1 of *How Does a Poem Mean?*). Ciardi and Williams explain that the second occurrence of the line “suddenly” transforms words into symbols (p. 9). But the end result is not what I am interested in here. Because they go on to consider Frost’s experience of writing the poem to gain insight into the performance of the poem:

Now look at the poem in another way. Did Frost know what he was going to do when he began? Considering the poem simply as a piece of juggling one cannot fail to respond to the magnificent turn at the end where, with one flip, seven of the simplest words in the language suddenly dazzle full of never-ending waves of thought and feeling; or—more precisely—of felt thought. Certainly an equivalent stunt by a juggler—could there be such an equivalent—would bring the house down. Was it to cap his performance with that grand stunt that Frost wrote the poem?

Far from it: the fact that must not be overlooked is that *Frost did not know he was going to write those lines until he wrote them*. Then a second fact must be

registered: *he wrote them because, for the fun of it, he had got himself into trouble.*
(p. 10, emphasis original)

Ciardi and Williams go on to explain how the novel rhyming scheme Frost initially adopted led to a conundrum in the final stanza. For my purposes I have explained enough and will leave that be. The point is that this celebrated poem was not necessarily the result of some instantaneously discovered insight, but that it likely emerged from the dynamics of goal setting. And if you reduce the goal setting in a writing process, a great deal of value may go with it.

Working backwards through this issue of *Visible Language*, let me now mention the “next” article, which presents an opportunity to address this editorial’s title. I have covered the donut and the donut hole, and the chocolate and oyster flavoring. But in emphasizing authorial voice in the donut, among the many things the metaphor can extend to, I have not seriously covered originality yet. Halldórsson (2026) relays a project completed by students in design schools in Eskilstuna, Sweden, and Perth, Australia. Students had to use generative AI to create travel advertisements for the other city (e.g., a student in Perth used AI to design for Eskilstuna).

Halldórsson describes early outcomes that “generally conveyed broad connotative impressions—‘a European town’ or ‘an Australian city’—rather than precise depictions” (p. 199). The wrong geographic features and endemic fauna would regularly appear. They were appropriate at the country or regional level but not the civic level. This can be mitigated in any number of ways, and of course the students over the course of the project developed and employed mitigation strategies. But there remains an underlying sense that originality is under threat.

So moving beyond authorial voice to focus on originality, let us return to writing but consider content rather than presentation. Scholarship is predicated on originality. A major criterion is novelty, that something is publishable because it is new, and this is a large part of its value. It becomes problematic for many researchers using the same tools to produce research when those tools push all of their work toward the same empty center. And if researchers are using the same tools—and regardless, AI tools utilize the same corpus to construct an averaged center—their research will tend to converge. The size of the donut—the gamut, the variety—is a measure of originality. Figure 12 tells that story, too.

The remaining articles from this issue seemingly let us step away from AI for a moment. Maciel (2026) provides copious examples of “graphemic construction” in the form of illustrated letters that far predate AI. He considers the degree to which illustrated letter-forms conform to an existing letter, or form a new letter. The 21 figures in the article



Figure 13. Miriam Martincic, Barcelona, 2023. Mixed media. From Maciel (2026) in this issue (as Figure 9 in that article, p. 172).

provide a rich demonstration of human originality, the stuff on which our AI models have been trained.

For instance, Figure 13 is one of Maciel's most recent examples, a rich and rewarding illustration by Miriam Martincic. In the same way that Ciardi and Williams (1975) discuss the poem as an experience—"a performance of forces" (p. 10)—we can engage with Martincic's illustration and know that it is wildly insufficient to call it an "E." Maciel (2026) calls it a hybrid letterform having "a negotiated condition in which structure and image co-operate" (p. 172), and that co-operation is active.



Figure 14. Laura Sisti's picturebook illustrations of children playing with gendered signifiers: female at top, and male at bottom. From Sisti and Peronti (2026; as Figures 9 and 10 in that article, p. 137).

Finally, Sisti and Perondi (2026) studied the impacts of gendered and gender neutralized picturebook illustrations on stereotyping in children's interpretations. The first author, Laura Sisti, produced the illustrations. The control afforded by her ownership of the imagery permitted precise and deliberate manipulation for careful stimuli preparation for an empirical study. Figure 14 provides details of her female- and male-gendered illustrations. They are gender encoded through "color, character morphology, and symbolic gender cues," as commonly found in "early-childhood educational materials in Italy" (p. 133).

The authors' results are quite interesting, but for the present editorial I just want to emphasize the *precision* with which Sisti was able to develop their study's stimuli—hand-drawn and guided by the human eye. Handless, eyeless, and unthinking AI models have not been trained on these illustrations yet, but they will be.

References

- Daniels, G. S. (1952, December). *The "average man"?*. Wright Air Development Center. <https://web.archive.org/web/20210411010707/https://apps.dtic.mil/dtic/tr/fulltext/u2/010203.pdf>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://www.jstor.org/stable/356600>
- Halldórsson, H. B. (2026). Are you sure? Exploring the challenges and opportunities of using generative AI in communication design projects: A cross-cultural case study of student reflections, bias, and critical engagement. *Visible Language*, 60(2), 183–209. <https://www.visible-language.org/Issue-60-2/VL-60-2-Halldorsson.pdf>
- Hayes, J. R., & Flower, L. S. (1980). Writing as problem solving. *Visible Language*, 14(4), 388–399. <https://journals.uc.edu/index.php/vl/article/view/5308>
- Huszár, F. (2017, November 9). Gaussian distributions are soap bubbles [Blog post]. *inFERENCE*. <https://www.inference.vc/high-dimensional-gaussian-distributions-are-soap-bubble/>
- Maciel, H. (2026). Illustrated letterforms: A taxonomy of graphemic construction. *Visible Language*, 60(2), 156–182. <https://www.visible-language.org/Issue-60-2/VL-60-2-Maciel.pdf>
- Mishra, P. (2023, April 20). Opinion: Gilbert Daniels, the gardener who changed our world [Blog post]. Mary Lou Fulton College for Teaching and Learning Innovation, Arizona State University. <https://education.asu.edu/about/news/opinion-gilbert-daniels-gardener-who-changed-our-world>
- Peterson, M., Patton, C., Green, M., & Díaz, B. (2026). The concept of a large language model: Misconceptions about and beneficial conceptualizations of LLM technology. *Visible Language*, 60(2), 210–239. <https://www.visible-language.org/Issue-60-2/VL-60-2-Peterson-et-al.pdf>
- Sisti, L., & Perondi, L. (2026). Graphic neutralization and gender stereotypes in early childhood: How visual variables shape children's interpretations. *Visible Language*, 60(2), 124–155. <https://www.visible-language.org/Issue-60-2/VL-60-2-Sisti-Perondi.pdf>
- Williams, A. (2026, March 25). Model collapse is already happening, we just pretend it isn't [Blog post]. *Blog@ACM*. Communications of the ACM. <https://cacm.acm.org/blogcacm/model-collapse-is-already-happening-we-just-pretend-it-isnt/>



Graphic Neutralization and Gender Stereotypes in Early Childhood: How Visual Variables Shape Children's Interpretations

Laura Sisti ^a and Luciano Perondi ^b

^a Istituto Tecnico Tecnologico Abate Zanetti, Venezia, Italy; ^b Università Iuav di Venezia, Italy

Corresponding author: Luciano Perondi (lperondi[at]iuav.it)

Abstract: This research explores the influence of gendered connotations in children's book illustrations on the perception and assimilation of gender stereotypes among boys and girls aged four to five. Through an empirical study conducted in a context designed to maximize ecological validity, the investigation examined whether graphic neutralization, defined as the reduction of details associated with gendered characteristics, can help decrease the activation of stereotypical schemas in children's behaviors. The study is grounded in theoretical frameworks drawn from research on visual perception, the acquisition of gender stereotypes in preschool age, and visual communication in educational contexts. The experimental design involved 82 participants (M = 41, F = 41; M = 4.73 years, SD = 0.45), structured in two phases: a baseline test, aimed at assessing pre-existing knowledge of gender stereotypes, and a subsequent test, conducted with two groups (control and experimental) to evaluate the effect of graphic neutralization. The illustrations, balanced in content but differentiated by the level of gender connotation, were presented individually. Data were collected through participant observation and structured ordinal quantitative observation sheets, allowing for the analysis of behaviors, preferences, and spontaneous associations between gender and activities. The results revealed a statistically significant reduction in stereotyped choices in the experimental group compared to the control group, along with a lower tendency to automatically assign gender to the depicted activities. Some variables, such as response certainty (CCU), recognition of gender-related elements (RGAA), and awareness of change (CSC), showed significant differences between groups, suggesting a tangible impact of visual neutralization on children's behavior. At a qualitative level, children in the experimental group displayed greater hesitation in their decision-making process, indicative of a suspension of cognitive automatisms linked to stereotypes.

@: [ISSUE](#) > [ARTICLE](#) >

Cite this article:

Sisti, L., & Perondi, L. (2026). Graphic neutralization and gender stereotypes in early childhood: How visual variables shape children's interpretations. *Visible Language*, 60(2), 124–155.

First published online July 16, 2026.

© 2026 Visible Language — this article is **open access**, published under the CC BY-NC-ND 4.0 license.

<https://visible-language.org/journal/>

Visible Language Consortium:

University of Leeds (UK)

University of Cincinnati (USA)

North Carolina State University (USA)

Implications for practice: The findings suggest that visual design choices in preschool educational materials can play a meaningful role in reducing the activation of gender stereotypes. In practice, illustrators, educators, and publishers should avoid treating visual style as a neutral component of learning resources, since features such as color, hairstyle, clothing, posture, facial detail, body shape, and symbolic objects may guide children's gendered interpretations of activities and roles. Educational materials, particularly those related to science, technology, engineering, and mathematics, should therefore represent tools, experiments, construction tasks, technological devices, and scientific exploration without systematically associating them with masculine or feminine visual markers. Graphic neutralization may be implemented by reducing explicit gender cues, limiting culturally gendered colors and accessories, softening or simplifying character features, and preserving recognizability without reinforcing binary gender associations. In early-childhood settings, where children frequently encounter scientific and technological content through picturebooks, games, and narrative images, such design strategies may help present STEM-related activities as broadly accessible and not implicitly tied to a specific gender.

Keywords: childhood; gender stereotypes; graphic variables; illustration; inclusion

1. Introduction

Gender stereotypes begin to form early in childhood and progressively influence cognitive expectations, preferences, and behavioral choices up to adult age (Grobelny & Michalski, 2015). Research in developmental psychology has shown that children between four and five years of age are already capable of associating activities, objects, and roles with gendered categories, often applying these associations consistently in their everyday decisions (Martin & Ruble, 2004; Serbin et al., 2001; Shutts et al., 2010). These early categorizations are not benign: they contribute to the construction of self-concept and play a role in shaping future educational interests and aspirations, particularly in domains traditionally marked by strong gender segregation such as STEM (Bian et al., 2017).

While the development of gender stereotypes has been widely studied from psychological and educational perspectives, comparatively less empirical attention has been paid to how visual communication may cue, reinforce, or shape children's gendered interpretations. Children's illustrated materials—picture books, educational worksheets, and learning tools—are often treated as neutral supports for content delivery. However, research in visual semiotics and multimodality has consistently shown that images operate as semiotic and informational systems within multimodal ensembles, actively shaping interpretation and contributing to the transmission of cognitive and cultural frameworks rather than merely accompanying verbal content (Cheng et al., 2025; Serafini & Freid, 2024).

Visual attributes such as color, shape, character morphology, and symbolic objects function as informational cues that guide interpretation and categorization processes.

Studies on children's literature have documented the persistence of gender bias in illustrated books, showing systematic imbalances in representation and recurrent associations between gender and activities (Boero & Giuliani, 2019; Brugeilles et al., 2002).

Information design is not only about legibility and clarity, but about what distinctions are made salient and what categories become cognitively available for decision-making. When illustrations systematically encode gender through color conventions, morphological signals, and symbolic props, they do more than represent; they pre-structure inference. Large-scale analyses of children's books have documented persistent imbalances and gendered patterns in depiction, supporting the claim that children's publishing participates in the reproduction of gender norms (Boero & Giuliani, 2019; Brugeilles et al., 2002). Yet much of this evidence is descriptive: it identifies bias in content but leaves open the causal question central to design research—whether and how specific visual variables modulate children's cognitive and behavioral responses.

Preschool age constitutes a particularly sensitive window for such investigation. At this stage, perceptual cues play a dominant role in meaning-making, and visual information often precedes or even substitutes verbal reasoning (Lewis et al., 2022). Children rely heavily on salient visual markers to structure social categories, making them especially susceptible to implicit cues embedded in graphic artifacts. Consequently, the repeated exposure to gender-connoted visual patterns may contribute to the early naturalization of stereotypes, reinforcing them before critical reflection becomes possible.

In response to this issue, recent design-oriented discussions have advocated for the development of inclusive and stereotype-aware educational materials. Yet these proposals are often normative and lack empirical validation. In particular, there is limited experimental evidence assessing whether modifying specific visual variables can measurably alter children's stereotypical responses. For information design research, this lack of evidence limits the translation of ethical intentions into operational design strategies.

The present study addresses this gap by empirically examining the effects of graphic neutralization in children's illustrations. Graphic neutralization is defined here as the systematic reduction of visual gender markers, including chromatic conventions, explicitly gendered symbols, and strongly connoted character traits, while maintaining narrative coherence and usability. Rather than introducing counter-stereotypical representations, the study investigates whether reducing the salience of gender cues can weaken stereotype-driven decision-making.

Using a pre-post experimental design with a control group and an experimental group, the study compares children's behavioral choices when exposed to gender-connoted versus visually neutralized illustrations. In the experimental condition, visual neutralization was operationalized through a controlled manipulation of specific graphic variables commonly associated with gender coding in children's visual culture. These included (1) the use of pastel and low-saturation color palettes, as well as colors less rigidly associated with gender—such as green (Hurlbert & Ling, 2007)—in place of culturally gendered chromatic schemes, whose early gender associations have been shown to emerge in preschool children and to be reinforced through socialization processes and everyday visual environments (Cherney et al., 2006; Eichstedt et al., 2002; Zemach et al., 2007); (2) the reduction of strongly connoted symbolic objects, which function as culturally shared gender cues in children's visual environments (Shutts et al., 2010); and (3) the attenuation of gendered bodily markers through simplified character features, such as rounded shapes and less explicitly gender-coded facial, hairstyle, clothing, and bodily details, while preserving sufficient visual detail for preschool children to recognize the characters and actions depicted. This choice was informed by developmental evidence showing that, in the absence of external sex-stereotyped cues such as hairstyle and clothing, children's ability to categorize faces by sex depends on internal facial-structure information and varies with age (Wild et al., 2000). Johnson et al. (2010) further showed that preschool children become increasingly able to use body-shape information when categorizing sex, supporting the broader methodological decision to reduce the salience of visible bodily cues that might guide gender attribution.

The use of simplified and pedomorphic character features was intended to reduce the salience of gendered bodily cues, a strategy that has been shown to be developmentally appropriate for preschool-aged children and particularly effective in minimizing externally imposed markers of gender during interaction (Malisiova & Kompoti, 2025).

However, evidence from studies on virtual and real infants suggests that even when visual gender cues are minimal or absent, observers—including children—may still infer and assign gender categories on the basis of sparse visual information, indicating that simplified character design can modulate, but not fully prevent, gender attribution processes (Araujo et al., 2022; Karniol, 2011). For this reason, the experimental stimuli were designed to minimize salient gender cues while preserving recognizability of the character, and children's gender attributions were retained as empirical outcomes in the analysis rather than treated as neutralized by design.

The focus is not on explicit attitudes but on observable responses, decision patterns, and indicators of cognitive engagement. This approach aligns with an information design perspective that conceives visual artifacts as structures that organize action and interpretation, particularly in early and pre-verbal stages of cognition.

In particular, such knowledge may inform the design of visual materials aimed at introducing STEM-related concepts in early childhood, a domain in which girls' lower participation and performance have been shown to emerge despite the absence of cognitive differences, pointing instead to cultural and environmental factors as key contributors (Bian et al., 2017; Lindberg & al., 2010). Yet much of the existing evidence remains descriptive: it identifies bias in content but leaves open the causal question central to design research—whether and how specific visual variables modulate children's cognitive and behavioral responses.

This research aims to contribute empirical evidence on how visual design can actively shape cognitive processes related to gender. The findings are intended to inform both theoretical discussions on the role of visual communication in early learning and practical design decisions in educational publishing.

2. Method

2.1. Hypotheses

Drawing on the reviewed literature, this study addresses the following research question: In children aged 4–5 years, does exposure to stereotyped-content illustrations incorporating gender-neutralizing graphic variables, compared with exposure to the same illustrations without these variables, lead to a reduction in the perception of gender stereotypes, as measured using a pre-post experimental design with repeated measures?

The study seeks to determine whether visual neutralization constitutes an effective strategy for reducing adherence to gender stereotypes in early childhood, with potential implications for the design of inclusive educational materials.

2.2. Experimental Design and Procedure

The study adopts a pre-post experimental design with repeated measures, involving a single experimental condition and two comparison groups, a control and an experimental group. This structure makes it possible to isolate the effect of graphic variables, control for external variables and reduce biases such as social conformity (Haun & Tomasello, 2011; Pedon, 2011), ensuring that the observed variations can be attributed to the visual manipulation of the stimuli.

The experiment was organized in two consecutive phases. The first phase, called Baseline Test, aimed to identify the spontaneous preferences of children for illustrations containing gender-symbolic elements. This phase provided an initial measure of pre-existing gender stereotypes (Bian et al., 2017; Lewis et al., 2022). For this, a pair of illustrations containing gender-symbolic elements such as ballet shoes opposite

football boots was presented. The Experimental Test presented the same activities in two graphic versions. This phase investigated the impact of graphic neutralization of the illustrations. It assessed whether reducing gender-related details could change children's perceptions and choices. Participants were randomly assigned to two groups. This procedure allows for comparison of responses and the assessment of the impact of graphic neutralization (Bikhchandani et al., 1992; Haun & Tomasello, 2011; Piaget, 1932). The control group received stereotyped illustrations with explicit gender details, while the experimental group received illustrations analogous in content but graphically neutral, lacking distinctive features attributable to a specific gender. The order of administration was kept the same for all participants to prevent carryover effects (Bordens & Abbott, 2011). In other words, this avoided the first test influencing behavior in the second test.

2.3. Participants

The sample consisted of 82 children aged 4–5 years, equally distributed by gender (41 males; 41 females), attending the final and penultimate years of preschool. Twenty-two children were four years old (14 females; 8 males) and sixty were five years old (27 females; 33 males). The overall mean age was 4.73 years ($SD = 0.45$). Gender distribution was balanced and the internal variability ensured a sufficiently heterogeneous sample. Participant selection was based on convenience, according to age and gender; geographical origin was not considered. The experiment was conducted in four public and private preschools located in central and northern Italy: Scuola dell'Infanzia Tavernelle "Bruno Munari" (Tavernelle di Colli al Metauro), Scuola dell'Infanzia di Sterpeti, Istituto Comprensivo D. Bramante – Scuola dell'Infanzia Fermignano, and European School of Venice (Mestre). Data collection took place within the school environment, in quiet classroom areas prepared by the teachers and adapted for individual administration of the tasks. Each child completed the task individually, separated from peers, in order to minimize potential social influence and conformity effects on attribution and choice responses (Haun & Tomasello, 2011). Session duration was not standardized and no quantitative measure of testing time was collected. Participation was authorized through informed consent obtained from parents or legal guardians.

2.4. Independent Variables

The independent variables consisted of specific graphic characteristics of the illustrations, systematically manipulated to modulate the degree of gender connotation of the visual stimuli. In particular, the manipulation concerned three dimensions: color, pedomorphic traits, and symbolic gender connotations.

Color was treated as a key variable in gender coding in early childhood (Cherney et al., 2006; Eichstedt et al., 2002; Zemach et al., 2007). In the control condition, the illustra-

tions employed culturally gendered color palettes, commonly associated with male and female categories (Figure 1; Table 1). The color palette used was informed by previous studies documenting gender-related color associations in childhood (Passig & Levin, 1999; Cherney et al., 2006; Eichstedt et al., 2002; Zemach et al., 2007). In particular, the selection of hues and their manipulation were based on the framework proposed by Xue et al. (2024), who identified seven colors (blue, orange, green, yellow, red, violet, and magenta) showing different degrees of gender association. In the experimental condition, the same base hues were visually neutralized through a systematic reduction of saturation and brightness using the HSB model, to reduce their recognizability as



Figure 1. Control condition stereotyped palette.



Figure 2. Experimental condition neutralized palette.

Table 1. Stereotyped color palette in CMYK values.

Hue	Cyan (%)	Magenta	Yellow	Black
Blue	92.4	74.6	0	0
Orange	0	62.9	91.3	0
Green	61.4	0	100	0
Yellow	0	0	100	0
Red	0	96.1	94.7	0
Purple	75.8	79.7	0	0
Magenta	0	100	0	0

Table 2. Neutralized color palette with HSB and CMYK values.

Hue	Hue	Sat. (%)	Bright.	Cyan (%)	Magenta	Yellow	Black
Blue	210°	15	70	24	23	23	4
Orange	30°	50	45	14	39	61	3
Green	118°	10	85	29	5	29	0
Yellow	50°	40	95	8	8	53	0
Red	10°	30	75	22	45	40	10
Purple	272°	10	50	53	48	34	19
Magenta	330°	15	65	36	44	27	9

gender markers (Figure 2; Table 2) (Hurlbert & Ling, 2007) while preserving overall chromatic variety.

The illustrated characters differed in the degree of pedomorphism and bodily differentiation. In the control condition, clearly gendered bodily and facial markers were present (e.g., hairstyles, proportions, posture) (Figure 3). In the experimental condition, these markers were attenuated through a simplified graphic style characterized by reduced sexual dimorphism, and ambiguous facial and bodily traits (Figure 4). For example, the larger hair shape in Figure 4 was used to increase the perceived size and roundness of the head within a simplified pedomorphic character design, rather than as a gendered hairstyle cue. To this end, the hair was rendered as a non-specific shape, avoiding a haircut that could be readily identified as masculine or feminine. Likewise, the simplification of the mouth and other facial details was intended to reduce expressive and stylistic specificity, not because these features are inherently gendered in isolation.

The comparison between Figures 3 and 4 should therefore not be understood as the isolation of a single visual variable, but as a controlled reduction of multiple gender-coded cues that operate in combination. For example, facial details such as the presence, shape, or emphasis of the mouth were not treated as inherently gendered in themselves; rather, they were considered part of a broader visual configuration through which expression, softness, maturity, and character identity may be culturally interpreted. Similarly, differences in eye size, body/head proportion, hairstyle, skin tone, and color palette were controlled as elements contributing to the overall gender coding of the character, rather than as independent indicators of gender. Given the multivariate nature of illustration, the aim was not to attribute gendered meaning to any single graphic feature in isolation, but to reduce the cumulative salience of visual



Figure 3. Control condition. Characters with distinct gendered pedomorphic features.

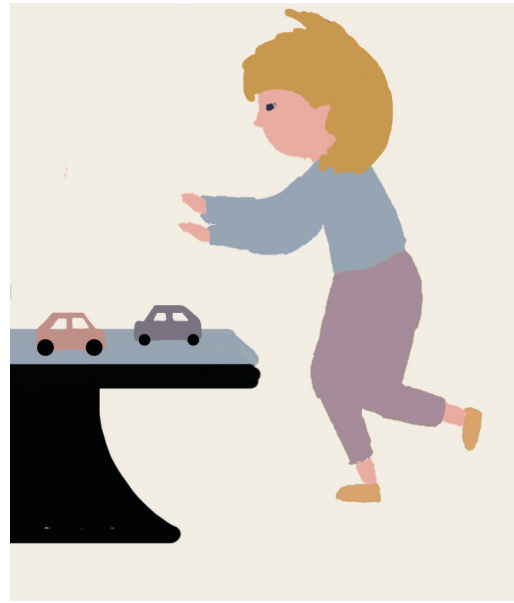


Figure 4. Experimental condition. Characters with reduced gender cues and ambiguous pedomorphic traits.

cues that, within familiar conventions of children's illustration, may support gender attribution.

This approach is grounded in picturebook and multimodal literacy scholarship, which shows that meaning in illustrated materials emerges through the interaction of verbal and visual resources rather than through narrative content alone (Nikolajeva & Scott, 2001). From a social semiotic perspective, children's interpretation of picturebooks is also situated within broader social, historical, and cultural contexts (Serafini, 2023). Accordingly, the gendered meaning of visual features such as hairstyle, clothing, posture, and bodily details was treated as culturally situated (center and north Italy, 2025) rather than universal.

Accordingly, the neutralization strategy was not intended to remove gender attribution altogether, but to reduce the salience of culturally recognizable gender cues within the specific visual conventions available to the participating children. Consistently, evidence on gender attribution suggests that observers may still assign gender categories even when explicit gender cues are reduced; therefore, simplified character design was treated as a strategy for modulating, rather than eliminating, gender attribution processes (Karniol, 2011).

Objects, clothing, and visual contexts were treated as additional gender markers, as they function as culturally shared cues for gender categorization (Shutts et al., 2010). In the control condition, gender-connoted iconic elements were present (Figure 5). In



Figure 5. Control condition with female stereotypes.



Figure 6. Experimental condition with neutralized female stereotypes.

the experimental condition, these elements were maintained at the level of content but represented in a visually neutral manner (Figure 6). For each content, two versions were produced: a stereotyped version (control condition) and a graphically neutralized version (experimental condition).

2.5. The Illustration

The illustrations were created specifically for the study by the first author, who also conducted the experimental sessions. The use of a single illustrator ensured consistency across the stimulus set and allowed the three graphic parameters—color, character morphology, and symbolic gender cues—to be manipulated within a coherent visual language. The adopted style was deliberately simplified and hand-drawn, in line with visual registers commonly found in early-childhood educational materials in Italy. However, the style was not treated as neutral: it was considered a meaningful graphic variable that could influence interpretation. For this reason, the findings should be understood as applying to simplified preschool-oriented illustrations, and not automatically extended to other modes of children's science communication.

In line with empirical evidence on visual recognition processes in preschool children, the stimuli used in both the Baseline Test and the Experimental Test were designed employing a simplified visual language (Eng et al., 2020), characterized by essential geometric forms, soft contours, pedomorphic proportions, and a high level of detail. This was intended to facilitate rapid and unambiguous processing of images by children (Hochberg, 1962; Elkind, 1970; Stone & Church, 1968).

For the Baseline Test, highly symbolic and culturally connoted stimuli were employed, namely a football boot (Zeng & He, 2024; Grabow & Kühn, 2019) and a dance shoe (Oreamuno, 2018), selected insofar as they embody visual and cultural signs that convey gendered behavioral norms and aesthetic ideals (Biemmi, 2017). Both images were rendered using a yellow chromatic palette, chosen as it is generally regarded as perceptually neutral in relation to stereotypical pink/blue gender associations (Cherney et al., 2006; Xue et al., 2024).

In the Experimental Test, the illustrations were differentiated between control and experimental conditions across three principal dimensions: color, pedomorphic features, and symbolic gender connotations. Specifically, the control condition preserves recognizable stereotypical coding, whereas the experimental condition implements a progressive process of visual neutralization through the adoption of non-polarized color palettes (Xue et al., 2024), the removal of identifying elements (Martin & Ruble, 2004), and the use of pedomorphic (Stroessner et al., 2020) and deliberately ambiguous forms.

2.6. Dependent Variables

The dependent variables were children's choices and observable behaviors in response to the illustrations. These measures allowed us to examine whether changes in color, character design, and symbolic objects affected the way children associated the images with gender. This approach builds on developmental studies showing that children's interpretation of images depends on visual detail, salience, and recognizability (Elkind, 1970; Gollin, 1960; Hochberg, 1962; Stone & Church, 1968).

Table 3. Dependent variables (behaviors measured).

Abbrev.	Name
CS	Children satisfaction
CE	Children engagement
CCU	Children confidence/uncertainty
CDE	Children difficulty during the experiment
RGAA	Recognition of gender-associated activities
SC	Stereotyped choices
CSC	Changing of stereotyped choices

Both stereotyped choices (SC) and a set of behavioral and cognitive variables were analyzed to describe children's behavior during the task and their degree of adherence to gender stereotypes in both parts of the experiment (Table 3).

Children's behaviors and choices were recorded using observation sheets, which allowed qualitative information to be transformed into numerical ratings on a 7-point scale. Data were collected through participant observation, fostering a natural and non-invasive interaction to ensure the authenticity of children's responses.

Data collection was conducted consistently across the Baseline Test and the Experimental Test, ensuring methodological continuity between the two phases. The same set of variables was analyzed in both conditions, except for CSC (changing of stereotyped choices), which was introduced exclusively in the experimental test to capture potential changes in children's choices and behaviors resulting from the graphic manipulation of the stimuli.

2.7. Baseline Test

The Baseline Test was designed to examine the presence of automatic gender-stereotypical behaviors and to assess children's knowledge of gender stereotypes, by verifying whether their choices between illustrations containing gender-symbolic elements reflected preferences based on pre-existing stereotypes (Martin & Ruble, 2004; Serbin et al., 2001), thus providing a reference framework for the subsequent experimental test.

The independent variable was the type of illustration shown to the children: one version included objects with culturally gendered associations, while the other reduced or removed these symbolic cues (Figures 7 and 8). The dependent variables were the children's choices, preferences, and observable behaviors during the task.



Figure 7. Baseline test stimulus. Female gender stereotypes in content (ballet shoes) (Bem, 1981; Piluso, 2019; Priulla, 2013; Zeng & He, 2024).



Figure 8. Baseline test stimulus. Male gender stereotypes in content (football shoes).

Yellow was used because it is minimally associated with gender stereotypes (Cherney et al., 2006; Xue et al., 2024). To ensure visibility and contrast against the white background, the hue was darkened by adding black (K=50%).

Stimuli were presented individually to each child (Haun & Tomasello, 2011). Children were asked to indicate an illustration and were asked the following questions: “Which scene do you like most?”, “Which activity would you like to do?”, “Which character is most like you?”, “Which scene makes you think of a friend?”

2.8. Experimental Test

The Experimental Test aims to verify whether graphic variables, specifically the graphic neutralization of illustrations can influence the perception of gender stereotypes in illustrations and adherence to them, compared to detailed and strongly stereotyped stimuli (Cherney et al., 2006; Martin & Ruble, 2004; Serbin et al., 2001; Xue et al., 2024).

The content of the activities remained unchanged, allowing the effect of graphic characteristics to be isolated. The control group received stereotyped illustrations with explicit graphic elements clearly indicating gender through objects and contexts, according to traditional gender roles (Figures 9 and 10). The experimental group received illustrations analogous in content, but graphically neutral, in which gender-indicative details were reduced to a minimum (Figures 11 and 12). Participants were tested individually, and the two stimuli were presented simultaneously. The questions administered were the same as those used in the Baseline Test. This procedure ensured spontaneity of responses and ecological validity of the collected data.

2.9. Ethics

The study was approved by the institutional Ethics Committee of the Iuav University of Venice. The Committee reviewed the submitted documentation and confirmed that the processing of participants’ personal data complies with the General Data Protection Regulation (EU 2016/679, GDPR) and that a Data Protection Impact Assessment (DPIA) was not required. Accordingly, the Committee issued a favorable opinion for the conduct of the research, ensuring that all procedures met ethical standards for research with young children.

3. Results

3.1. Baseline Test

The study involved 82 children aged 4–5 years ($M\text{ age} = 4.73$, $SD = 0.45$), equally distributed by gender (41 males, 41 females), recruited from preschools. Variables were analyzed to describe children’s behavior during the task. The majority of children displayed



Figure 9. Experimental test stimulus. Control group illustration (girls playing with dolls).



Figure 10. Experimental test stimulus. Control group illustration (boys playing with toy cars).



Figure 11. Experimental test stimulus. Experimental group illustration (female scene, neutralized).



Figure 12. Experimental test stimulus. Experimental group illustration (male scene, neutralized).

Table 4. Baseline Test Contingency Tables. Stereotyped choices divided by participant gender.

SC	Male Participant	Female Participant
MM	40	0
FF	0	37
FM	0	4
MF	1	0
Total	41	41

Table 5. Baseline test—Chi-square test.

	Value	df	<i>p</i>
X²	82.0	3	<.001
N	82		

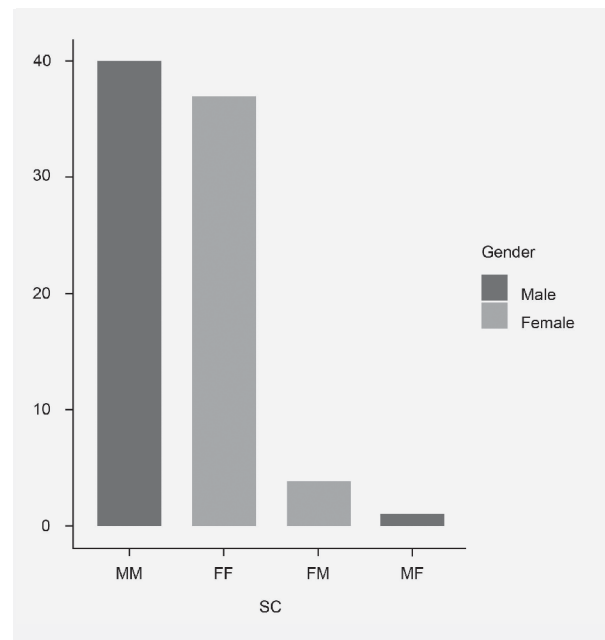


Figure 13. Baseline Test Contingency Chart. Stereotyped choices divided by participant gender.

choices consistent with the gender stereotypes associated with the represented activities (Table 4). The overall percentage of stereotyped choices (SC) was high, indicating a strong early internalization of gender-activity and symbol associations.

The vast majority of children selected images consistent with stereotypes associated with their own gender. Specifically, 97.6% of boys chose the male-stereotyped image, while 90.2% of girls selected the female-stereotyped one. Only 6.1% of the total made a non-stereotyped choice relative to their gender.

The chi-square test revealed statistically significant association between gender and stereotyped choice, $\chi^2(3) = 82.00$, $p < .001$, indicating that children's choices were strongly influenced by their gender (Table 5). The contingency analysis of SC $\times$ gender did not reveal statistically significant differences between boys and girls, suggesting that adherence to gender stereotypes is relatively homogeneous across the two groups (Figures 13–16).

Overall, these results show that by the age of 4–5, children already display a robust adherence to gender stereotypes when choosing images associated with activities or roles, supporting evidence that gender schemas are well established in preschool

age and guide children's interpretations of the world (Martin & Ruble, 2004; Shutts et al., 2010).

The mean scores for Satisfaction (CS) and Engagement (CE) were medium-high (median = 5.00 for both genders), indicating that the baseline activity was generally experienced as enjoyable and engaging, without signs of refusal or disinterest. Safety/Uncertainty (CCU) showed medium-high values (girls: 7.00; boys: 5.00), suggesting that stereotypical choices were not random but accompanied by decisional confidence (Table 6).

Significant positive correlations emerged among recognition of gender-activity associations (RGAA), conformity to gender stereotypes (BCS), and stereotypical choices (SC) (Table 7). Specifically, RGAA-BCS: $\rho = 0.836$, $p < .001$; BCS-SC: $\rho = 0.293$, $p = .008$; RGAA-

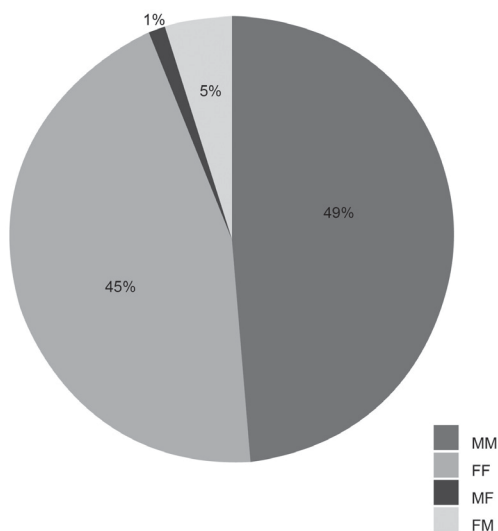


Figure 14 (top left). Participant choices chart in Baseline Test.

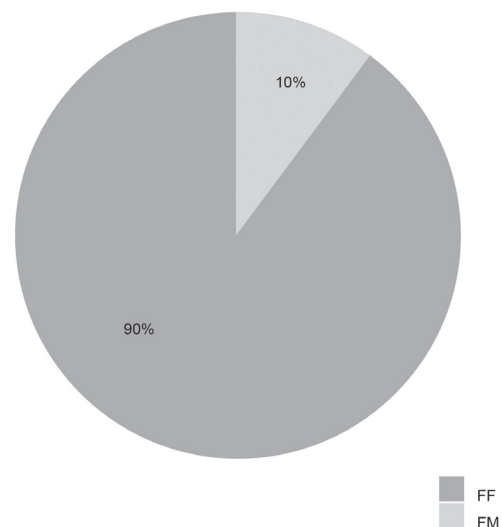


Figure 15 (top right). Female participant choices chart in Baseline Test.

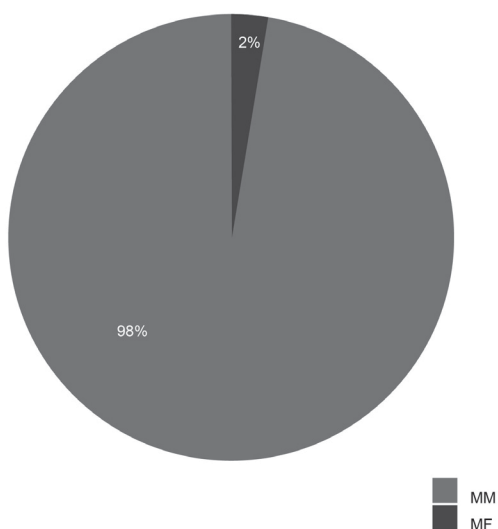


Figure 16 (bottom left). Male participant choices chart in Baseline Test.

Table 6. Baseline test values—dependent variables. M–W: Mann–Whitney.

Variable	M–W U Test	<i>p</i> -value	Effect (<i>r</i>)	Female Median	Male Median
CS	765	.452	–0.0898	5.00	5.00
CE	752	.376	–0.1053	5.00	5.00
CCU	647	.055	–0.2302	7.00	5.00
BCS	688	.136	–0.1814	5.00	5.00
CDE	688	.127	–0.1820	5.00	5.00
RGAA	666	.082	–0.2082	7.00	5.00

Table 7. Correlation Matrix between dependent variables in Baseline test.

		CS	CE	CCU	BCS	CDE	RGAA	SC
CS	Spearman's rho	–						
	df	–						
	<i>p</i> -value	–						
CE	Spearman's rho	0.944	–					
	df	80	–					
	<i>p</i> -value	<.001	–					
CCU	Spearman's rho	0.747	0.785	–				
	df	80	80	–				
	<i>p</i> -value	<.001	<.001	–				
BCS	Spearman's rho	0.432	0.412	0.541	–			
	df	80	80	80	–			
	<i>p</i> -value	<.001	<.001	<.001	–			
CDE	Spearman's rho	0.644	0.645	0.579	0.352	–		
	df	80	80	80	80	–		
	<i>p</i> -value	<.001	<.001	<.001	0.001	–		
RGAA	Spearman's rho	0.436	0.467	0.606	0.836	0.320	–	
	df	80	80	80	80	80	–	
	<i>p</i> -value	<.001	<.001	<.001	<.001	0.003	–	
CSC	Spearman's rho	0.084	0.100	0.206	0.293	0.092	0.325	–
	df	80	80	80	80	80	80	–
	<i>p</i> -value	0.454	0.370	0.064	0.008	0.412	0.003	–

SC: $\rho = 0.325$, $p = .003$. These findings show that children who more often recognized gender–activity associations also tended to produce more stereotype-congruent responses. Because these correlations are calculated among outcome measures, they do not establish whether recognition of gender–activity associations leads to stereotypical choices, or whether children more inclined to make stereotypical choices are also more likely to identify gender–activity associations. The causal claim is therefore restricted to the experimental manipulation: differences between the control and neutralized conditions suggest that changes in visual cues affected children’s responses, while the correlations among RGAA, BCS, and SC are interpreted as associations within those responses rather than as causal pathways.

3.2. Experimental Test

In the control group, the percentages of SC are comparable to those observed in the Baseline Test, confirming the stability of stereotypical automatic responses in the presence of visually connoted stimuli. On the other hand, a significant reduction in the percentage of stereotypical choices is observed, in the experimental group, accompanied by an increase in non-stereotypical choices compared to the control group (Table 8).

In the control group, the vast majority of children made stereotypical choices. Specifically, 21 out of 22 boys (95.5%) selected male-stereotyped images (MM) and 18 out of 19 girls (94.7%) selected female-stereotyped images (FF), for a total of 39 stereotypical choices out of 41 participants (95.1%). In contrast, in the experimental group, the proportion of stereotypical choices decreased markedly, 10 out of 21 boys (47.6%) chose MM and 7 out of 20 girls (35.0%) chose FF (Figure 17). Overall, stereotypical choices in the experimental group accounted for 17 out of 41 selections (41.5%), while non-stereotypical choices (FM and MF) increased to 24 out of 41 (58.5%), compared to only 4.9% in the control group. Row percentages further illustrate this shift. In the control group, 51.2% of choices were MM and 43.9% FF (Figure 18). In contrast, in the experimental group, the distribution shifted toward neutral choices (31.7% FM and 26.8% MF) (Figure 19). Overall, these results suggest that the graphical neutralization of illustrations can significantly influence children’s perception and choice, reducing automatic adherence to gender stereotypes.

A chi-square analysis revealed a statistically significant association between stereotypical choices (SC) and group ($\chi^2(3) = 27.4$, $p < .001$) with a contingency coefficient of 0.500, indicating a strong relationship between stimulus type and the frequency of stereotypical choices (Tables 9 and 10). Standardized residuals confirmed that stereotypical choices were more frequent than expected in the control group (MM: +3.50; FF: +3.89) and less frequent than expected in the experimental group (MM: –3.50;

Table 8. Experimental test. Contingency tables.

Group		MM	FF	FM	MF
Control	Observed	21	18	1	1
	Expected	15.5	12.5	7.00	6.00
	% w/in row	51.2	43.9	2.4	2.4
	% w/in column	67.7	72.0	7.1	8.3
Experimental	Observed	10	7	13	11
	Expected	15.5	12.5	7.00	6.00
	% w/in row	24.4	17.1	31.7	26.8
	% w/in column	32.3	28.0	92.9	91.7

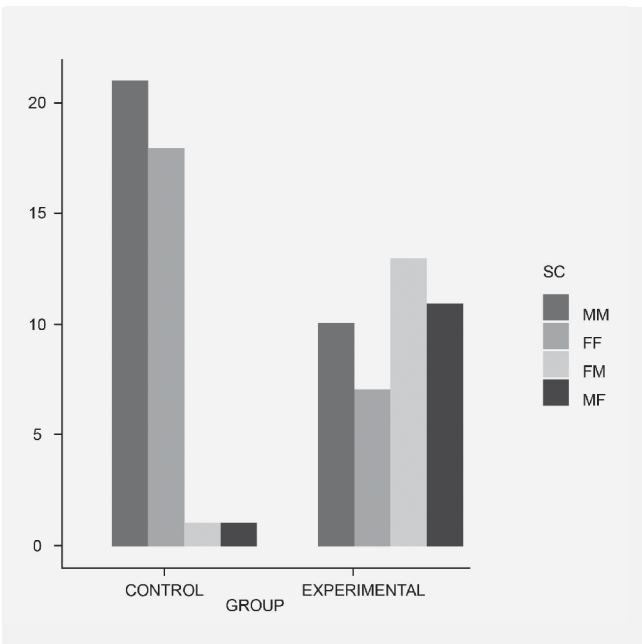
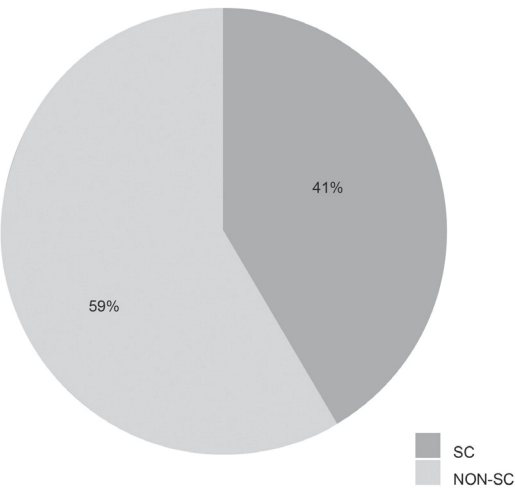
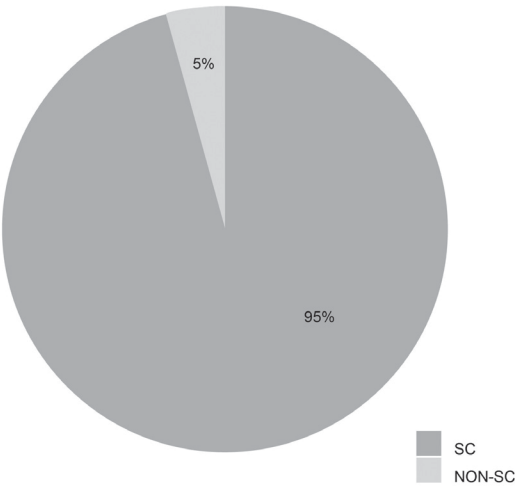


Figure 17 (top right). Groups choices chart in Experimental Test.

Figure 18 (top left). Control groups choices chart in Experimental Test.

Figure 19 (bottom left). Experimental groups choices chart in Experimental Test.

Table 9. Experimental test—Chi-square test.

	Value	df	p
X²	27.4	3	<.001
N	82		

Table 10. Experimental test. Contingency coefficient.

Nominal	Value
Contingency coefficient	.500

Table 11. Experimental Test. Post Hoc Test (Standardized Residuals).

Participant Gender	SC	Control Group	Experimental Group
Male	Male	3.35	−3.35
	Female	−3.35	3.35
Female	Male	−3.89	3.89
	Female	3.89	−3.89

Table 12. Descriptives dependent variables in Experimental Test.

	Group	CS	CE	CCU	BCS	CDE	RGAA	CSC
Mean	Control	6.02	5.98	5.93	5.73	5.15	5.73	1.83
	Experimental	6.07	5.98	4.46	3.05	5.78	3.34	4.46
Median	Control	7	7	7	7	5	7	1
	Experimental	7	7	5	3	7	3	5
SD	Control	1.11	1.19	1.49	1.53	1.51	1.47	1.61
	Experimental	1.10	1.11	1.85	1.70	1.47	1.84	2.24

Table 13. Experimental Test. Independent Samples T-Test
(Note. H^a: Control ≠ Experimental).

		Statistic	<i>p</i>
CS	Mann-Whitney U	821	0.835
CE	Mann-Whitney U	831	0.920
CCU	Mann-Whitney U	461	<.001
BCS	Mann-Whitney U	234	<.001
CDE	Mann-Whitney U	645	0.051
RGAA	Mann-Whitney U	292	<.001
CSC	Mann-Whitney U	329	<.001

Table 14. Experimental Test—Homogeneity of Variances Tests.

		F	df₁	df₂	<i>p</i>
CS	Levene's	0.00986	1	80	0.921
	Variance ratio	1.004	40	40	0.990
CE	Levene's	0.36396	1	80	0.548
	Variance ratio	1.163	40	40	0.635
CCU	Levene's	1.95314	1	80	0.166
	Variance ratio	0.652	40	40	0.180
BCS	Levene's	0.06903	1	80	0.793
	Variance ratio	0.811	40	40	0.512
CDE	Levene's	0.56801	1	80	0.453
	Variance ratio	1.047	40	40	0.885
RGAA	Levene's	2.36925	1	80	0.128
	Variance ratio	0.636	40	40	0.157
CSC	Levene's	7.29298	1	80	0.008
	Variance ratio	0.519	40	40	0.041

Table 15. Experimental Test. Tests of Normality.

		Statistic	<i>p</i>
CS	Shapiro–Wilk	0.707	<.001
	Kolmogorov–Smirnov	0.349	<.001
	Anderson–Darling	11.58	<.001
CE	Shapiro–Wilk	0.710	<.001
	Kolmogorov–Smirnov	0.339	<.001
	Anderson–Darling	11.35	<.001
CCU	Shapiro–Wilk	0.912	<.001
	Kolmogorov–Smirnov	0.224	<.001
	Anderson–Darling	2.97	<.001
BCS	Shapiro–Wilk	0.944	<.001
	Kolmogorov–Smirnov	0.175	0.013
	Anderson–Darling	2.03	<.001
CDE	Shapiro–Wilk	0.889	<.001
	Kolmogorov–Smirnov	0.221	0.013
	Anderson–Darling	3.23	<.001
RGAA	Shapiro–Wilk	0.887	<.001
	Kolmogorov–Smirnov	0.242	<.001
	Anderson–Darling	3.91	<.001
CSC	Shapiro–Wilk	0.923	<.001
	Kolmogorov–Smirnov	0.202	0.002
	Anderson–Darling	2.67	<.001

FF: -3.89) (Table 11). These results suggest that the graphical neutralization of illustrations effectively reduced children's automatic adherence to gender stereotypes.

These findings suggest that graphical neutralization reduces gender-related automatic responses. The comparison between groups reveals significant differences in decision-related confidence/uncertainty (CCU), conformity to gender stereotypes (BCS), recognition of activities associated with a specific gender (RGAA), and changes in choices and behaviors (CSC) (Table 12). Specifically, the experimental group shows higher decision-making uncertainty (CCU), lower conformity to gender stereotypes (BCS), reduced automatic recognition of activity–gender associations (RGAA), and a greater shift toward non-stereotypical choices and behaviors (CSC).

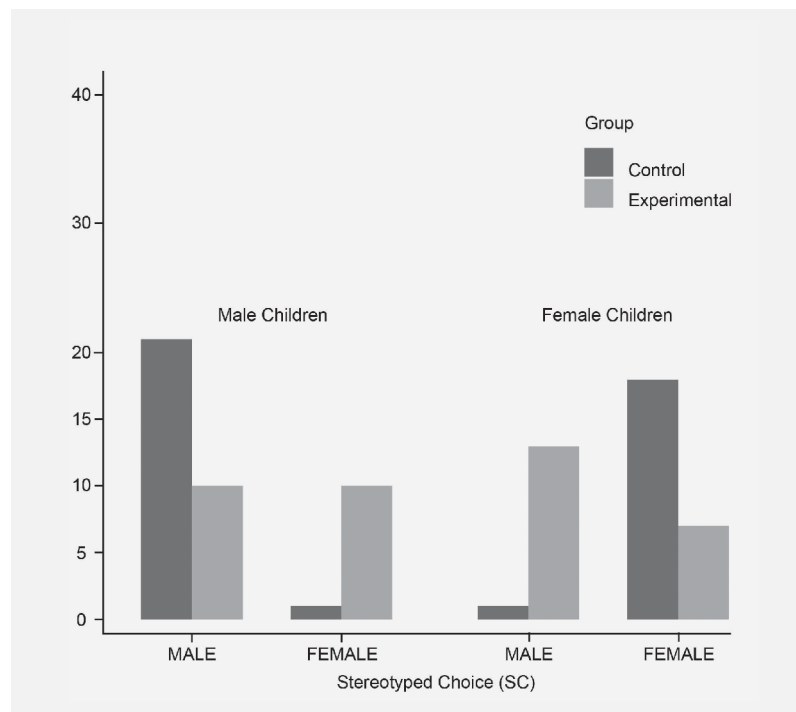


Figure 20. Stereotyped Choices comparison between groups in the Experimental Test.

Confidence/uncertainty (CCU) and Engagement (CE) variables do not show significant differences between the groups in terms of engagement and satisfaction (CS: $U = 821$, $p = .835$; CE: $U = 831$, $p = .920$) (Tables 13–15). This finding is particularly important as it demonstrates that graphical neutralization does not reduce children's engagement or enjoyment of the activity. The activity remained equally enjoyable and motivating across both conditions, while successfully reducing stereotypical behaviors.

A gender-stratified analysis confirmed that the effect was significant for both girls ($\chi^2(1) = 15.11$, $p < .001$, contingency coefficient = 0.528) and boys ($\chi^2(1) = 12.22$, $p < .001$, contingency coefficient = 0.470), with a slightly stronger effect observed among girls (Figure 20). Notably, within the experimental group, no significant gender differences in stereotypical choices were observed ($\chi^2(1) = 0.672$, $p = .412$), suggesting that graphical neutralization tends to equalize response patterns across genders.

Qualitative observations during the experimental phase revealed notable behavioral differences between groups. Children in the experimental group exhibited longer decision times and greater verbal reflection before making their choices, suggesting a suspension of automatic, stereotype-driven responses in favor of more deliberate cognitive processing. In contrast, children in the control group typically made rapid, confident selections consistent with gender stereotypes.

In summary, the results demonstrate that the experimental intervention using graphically neutralized illustrations effectively reduced children's stereotypical responses (Control group: 95.1%; Experimental group: 41.5%; $\chi^2(3) = 27.4$, $p < .001$, contingency coefficient = 0.500), while significantly impacting key cognitive and behavioral dimensions. Conformity to gender stereotypes decreased significantly (BCS: $U = 234$, $p < .001$), as did decision-making confidence and uncertainty (CCU: $U = 461$, $p < .001$), recognition of gender-activity associations (RGAA: $U = 292$, $p < .001$), and changes in choices and behaviors awareness (CSC: $U = 329$, $p < .001$). Importantly, engagement and satisfaction remained unaffected (CS: $U = 821$, $p = .835$; CE: $U = 831$, $p = .920$), indicating that the activity continued to be enjoyable and motivating. These findings support the thesis that visual neutralization can modify children's perception and adherence to gender stereotypes without compromising their engagement in the activity.

4. Discussion

The present study set out to investigate whether the graphic neutralization of gender-coded visual variables in children's illustrations can reduce the activation of gender stereotypes in preschool-aged children. The findings provide converging quantitative and qualitative evidence that visual design choices play an active role in shaping children's gender-related interpretations and decision-making processes.

At baseline, the results confirm that children aged 4–5 already display a strong and systematic adherence to gender stereotypes. The very high proportion of stereotypical choices, together with the significant association between gender and selected activities, aligns with previous research showing that gender schemas are well established in early childhood and guide interpretation even in the absence of explicit verbal cues (Martin & Ruble, 2004; Shutts et al., 2010). Importantly, baseline data show not only stereotypical preferences, but also high decisional confidence and rapid choice execution, suggesting that these associations operate as largely automatic cognitive shortcuts rather than reflective judgments.

Against this background, the experimental manipulation demonstrates that graphic neutralization produces a substantial disruption of these automatic responses. When gender-connoted visual cues were reduced through controlled modifications of color saturation, pedomorphic traits, and symbolic objects, stereotypical choices decreased dramatically in the experimental group compared to the control group. This effect was statistically robust and accompanied by a marked increase in non-stereotypical choices, indicating that children did not simply randomize their responses but engaged differently with the visual stimuli.

Crucially, the reduction in stereotypical choices was not associated with lower engagement or satisfaction. Measures of enjoyment and task involvement remained comparable across conditions, suggesting that graphic neutralization does not diminish the attractiveness or usability of the illustrations. This finding is particularly relevant from an information design perspective, as it counters the assumption that inclusivity-oriented visual strategies necessarily entail a loss of expressiveness or motivational value.

The analysis of behavioral and cognitive variables further clarifies the mechanism underlying this effect. Children in the experimental group showed lower levels of decisional confidence (CCU), reduced behaviors conforming to stereotypes (BCS), and a lower tendency to automatic recognition of gender-associated activities (RGAA). Rather than indicating confusion or disengagement, this pattern points to a temporary suspension of stereotype-driven heuristics. Qualitative observations support this interpretation: children exposed to neutralized illustrations hesitated more often, took longer to decide, and occasionally verbalized uncertainty or alternative interpretations. These behaviors suggest a shift from automatic categorization to more deliberative processing.

The results therefore indicate that the graphic neutralization of gender markers in illustrations reduced stereotypical choices and gendered activity attribution in the immediate context of the task, without diminishing children's engagement or satisfaction. This finding is relevant beyond the specific experimental setting because illustrated materials participate in the early organization of social categories, preferences, and perceived possibilities for action. However, the present study does not measure long-term interests, self-concept, or educational choices. Its implications should therefore be understood primarily in relation to the design of preschool visual materials and to the ways in which illustrations may support or reduce gendered interpretations of activities, objects, and roles.

STEM remains a particularly relevant application in the Italian context, where available evidence shows a marked gender gap in the choice of STEM degree programs even among girls with high demonstrated competence. Integrated data linking INVALSI standardized assessments with university enrollment records indicate that girls are less likely than boys to enroll in STEM degrees, suggesting that divergence in educational choices cannot be explained by academic performance alone (Falzetti, 2024). For this reason, gender-aware visual design in early educational materials may be considered one possible area for future research on the longer developmental chain linking early representations, self-perceptions, interests, and later educational choices.

Nevertheless, the present data do not allow direct conclusions about STEM preferences or long-term educational trajectories. Establishing such links would require studies that include specific measures of domain interest, self-efficacy, and perceived belonging, as well as longitudinal designs following children across later stages of schooling.

Future research should also examine whether similar effects occur across different illustration styles, levels of realism, and genres of children's visual communication, including narrative picturebooks, preschool science materials, informational books, and classroom worksheets.

From this perspective, the findings may inform the design of picturebooks, worksheets, non-fiction materials, and early-learning resources in which children encounter representations of play, care, movement, science, technology, professions, and everyday activities. Reducing the salience of gender-coded visual cues may be relevant for STEM-related representations, and for domains traditionally associated with either boys or girls, such as sport, construction, nursing, domestic care, or technological play. The broader contribution of the study concerns the role of visual design in expanding the range of activities that children may perceive as available to them.

5. Limits of the Study

This study presents certain limitations that should be taken into account when interpreting the findings. First, the sample, composed of children from a relatively homogeneous socio-cultural context, limits the extent to which the results can be generalized to other populations. Although the sample size was adequate for the exploratory aims of the study, further research involving larger samples could help strengthen the generalizability of the findings. Moreover, the visual markers employed to represent or neutralize gender stereotypes—such as hairstyles, clothing, chromatic palettes, and objects—are products of specific cultural and social conventions and may therefore acquire different meanings across diverse geographical and cultural contexts. Consequently, the effectiveness of the proposed graphic neutralization strategies may vary depending on the gender norms and representations specific to each context.

An additional limitation concerns the absence of quantitative time measurements during the experimental procedure. As a result, it was not possible to examine potential differences in response processes across the experimental conditions or to provide precise information regarding the duration of the testing sessions.

A further limitation concerns the visual style of the stimuli. The present study used simplified, hand-drawn, character-based illustrations designed for preschool-aged children. This choice was coherent with the nursery-school context of the experiment, but it limits the generalizability of the findings to other forms of children's science communication, including photographic, diagrammatic, or more verisimilar scientific illustration. Future research should examine whether the reduction of gender-coded cues operates similarly across different illustration styles, school grade, and levels of visual realism.

6. Future Work

In light of these considerations, future research should involve larger and more culturally heterogeneous samples, integrate the systematic collection of temporal measures, and adopt cross-cultural comparative approaches and longitudinal research designs in order to assess the generalizability of the findings, the temporal stability of the observed effects, and the applicability of the proposed design strategies across different educational and socio-cultural settings. Future studies should also compare different illustration styles and degrees of visual realism, from simplified preschool-oriented drawings to more realistic, diagrammatic, photographic, or scientific forms of representation. This would clarify whether graphic neutralization operates similarly across different genres of children's visual communication, including picturebooks, worksheets, nonfiction materials, and early STEM education resources. Such work would help assess the generalizability of the present findings and the applicability of the proposed design strategies across different educational and socio-cultural contexts. Our future research will focus particularly on STEM-related materials for early childhood education, where the design of visual representations may play a role in how children encounter activities, tools, and roles that are often culturally gendered.

7. Conclusion

The results suggest that the design of books and scientific materials featuring explicitly genderless illustrations may reduce the implicit association between specific activities and a particular gender. This implies that, in educational contexts, illustrations that do not link activities to gender markers can be used to promote a freer and less biased access to content, particularly for girls.

The neutralization of illustrations can be implemented through several design strategies. First, by eliminating gender connotations and representing characters without distinctive traits referable to a specific identity, such as explicit sexual markers (e.g., breasts, beards, hips), as well as clothing, objects, or hairstyles that are culturally associated with a particular gender. A further strategy involves reducing pedomorphic features—elements that accentuate childlike characteristics, such as disproportionately large heads or eyes—by avoiding excessive emphasis on such proportions and instead favoring softer, more essential, and less detailed lines. Finally, a neutral color palette may be adopted by desaturating colors traditionally attributed to specific genders.

These findings have practical implications for the design of preschool educational materials. They suggest that illustrators, educators, and publishers should not treat visual style as neutral, since color, hairstyle, clothing, facial detail, posture, and symbolic objects can contribute to the gendered interpretation of activities and roles.

For STEM-related materials, this approach may be particularly useful. Tools, experiments, construction activities, technological devices, and scientific exploration can be represented without systematically attaching them to masculine or feminine visual markers. In early-childhood contexts, where children often encounter science and technology through narrative images, worksheets, games, and picturebooks, reducing the salience of gender-coded cues may help present these activities as broadly accessible rather than implicitly associated with one gender.

8. Ethics Statement

The study was reviewed and approved by the Ethics Committee of Università Iuav di Venezia. Written informed consent was obtained from parents or legal guardians before children's participation.

9. Acknowledgments

The authors would like to thank the schools that participated in the study, and in particular the school administrators, teachers, children, and their families: Scuola dell'Infanzia Tavernelle "Bruno Munari" (Tavernelle di Colli al Metauro), Scuola dell'Infanzia di Sterpeti, Istituto Comprensivo D. Bramante – Scuola dell'Infanzia Fermignano, and European School of Venice – ESV Srl. (Mestre). Their cooperation was essential for data collection and for ensuring the ecological validity of the study.

This article is the result of a collaborative effort between the authors. Laura Sisti contributed to the writing of Sections 2 and 3 (Method and Results), while Luciano Perondi authored Sections 1 and 4 (Introduction and Discussion).

10. References

- Araujo, V., Schaffer, D., Costa, A. B., & Musse, S. R. (2022). Can gender categorization influence the perception of animated virtual humans? *arXiv*. <https://doi.org/10.48550/arXiv.2208.02386>
- Baird, C. L., & Keene, J. R. (2018). Closing the gender gap in math confidence. *International Journal of Gender, Science and Technology*, 10(3), 378–410. <https://scispace.com/pdf/closing-the-gender-gap-in-math-confidence-gender-and-race-598gww82dz.pdf>
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88(4), 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>
- Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323), 389–391. <https://doi.org/10.1126/science.aah6524>
- Biemmi, I. (2017). *Educazione sessista. Stereotipi di genere nei libri delle elementari*. Rosenberg & Sellier.

- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026. <https://doi.org/10.1086/261849>
- Boero, M., & Giuliani, D. (2019). Gli stereotipi di genere negli albi illustrati. *Filosofi(e) Semiotiche*, 6(1), 12–27. <https://www.ilsileno.it/wp-content/uploads/2019/08/3-Boero-Giuliani.pdf>
- Bordens, K. S., & Abbott, B. B. (2011). *Research design and methods: A process approach*. McGraw-Hill.
- Brugeilles, C., Cromer, I., Cromer, S., & Andreyev, Z. (2002). Male and female characters in illustrated children's books or how children's literature contributes to the construction of gender. *Population*, 57(2), 237–267. <https://www.jstor.org/stable/3246609>
- Cheng, L. M. Y., Powell, S. J., & Davis, B. (2025). Critical multimodal literacy: An analysis of young children's picture books about refugees. *International Journal of Early Childhood*, 57(3), 875–901. <https://doi.org/10.1007/s13158-025-00423-7>
- Cherney, I. D., Harper, H. J., & Winter, J. A. (2006). Nouveaux jouets: Ce que les enfants identifient comme «jouets de garçons» et «jouets de filles». *Enfance*, 58(3), 266–282.
- Eichstedt, J. A., Serbin, L. A., Poulin-Dubois, D., & Sen, M. G. (2002). Of bears and men: Infants' knowledge of conventional and metaphorical gender stereotypes. *Infant Behavior & Development*, 25, 296–310. [https://doi.org/10.1016/S0163-6383\(02\)00081-4](https://doi.org/10.1016/S0163-6383(02)00081-4)
- Elkind, D. (1970). *Children and adolescents: Interpretive essays on Jean Piaget*. Oxford University Press.
- Eng, C. M., Godwin, K. E., & Fisher, A. V. (2020). Keep it simple: Streamlining book illustrations improves attention and comprehension in beginning readers. *npj Science of Learning*, 5(14). <https://doi.org/10.1038/s41539-020-00073-5>
- Falzett, P. (2024). Il divario di genere nella scelta delle lauree STEM: un percorso in salita per le ragazze. In *Atti Accademia Roveretana degli Agiati* (pp. 239–254).
- Gollin, E. S. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills*, 11(3), 289–298. <https://doi.org/10.2466/pms.1960.11.3.289>
- Grabow, H., & Köhl, M. (2019). You don't bend it like Beckham if you're female and reminded of it: Stereotype threat among female football players. *Frontiers in Psychology*, 10, 1963. <https://doi.org/10.3389/fpsyg.2019.01963>
- Grobelny, J., & Michalski, R. (2015). The role of background color, interletter spacing, and font size on preferences in the digital presentation of a product. *Computers in Human Behavior*, 43, 85–100. <https://doi.org/10.1016/j.chb.2014.10.036>
- Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, 82(6), 1759–1767. <https://doi.org/10.1111/j.1467-8624.2011.01666.x>
- Hochberg, J. (1962). The psychophysics of pictorial perception. *Audiovisual Communication Review*, 10, 22–54. <https://link.springer.com/article/10.1007/BF02765240>
- Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, 17(16), R623–R625. <https://doi.org/10.1016/j.cub.2007.06.022>
- Johnson, K. L., Lurye, L. E., & Tassinari, L. G. (2010). Sex categorization among preschool children: Increasing utilization of sexually dimorphic cues. *Child Development*, 81(5), 1346–1355. <https://doi.org/10.1111/j.1467-8624.2010.01476.x>
- Karniol, R. (2011). The color of children's gender stereotypes. *Sex Roles*, 65(3–4), 119–132. <https://link.springer.com/article/10.1007/s11199-011-9989-1>
- Lewis, M., Cooper Borkenhagen, M., Converse, E., Lupyan, G., & Seidenberg, M. S. (2022). What might books be teaching young children about gender? *Psychological Science*, 33(1), 33–47. <https://doi.org/10.1177/09567976211024643>

- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136(6), 1123–1135. <https://doi.org/10.1037/a0021276>
- Malisiova, E., & Kompoti, D. (2025). Developing a digital narrative tool to explore gender bias in early childhood education. *Journals of Education & Social Sciences*, 1(1), 39–48.
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2), 67–70. <https://doi.org/10.1111/j.0963-7214.2004.00276.x>
- Nikolajeva, M., & Scott, C. (2001). *How picturebooks work*. Garland.
- Oreamuno, S. I. (2018). The pointe shoe: A tool for knowledge production. *Contingent Horizons: The York University Student Journal of Anthropology*, 4(1), 77–91. <https://doi.org/10.25071/2292-6739.96>
- Passig, D., & Levin, H. (1999). Gender interest differences with multimedia learning interfaces among preschoolers. *Computers in Human Behavior*, 15(2), 173–183. [https://doi.org/10.1016/S0747-5632\(99\)00016-3](https://doi.org/10.1016/S0747-5632(99)00016-3)
- Pedon, A. (2011). *Psicologia sociale*. McGraw-Hill.
- Piaget, J. (1932). *The moral judgment of the child*. Kegan Paul, Trench, Trubner & Co.
- Piluso, F. (2019). Lo stereotipo come modo di (ri)produzione e distruzione segnica: Esempi e strategie di sovversione. *Filosofi(e) Semiotiche*, 6(1), 183–197. <https://www.ilsileno.it/wp-content/uploads/2019/08/18-Piluso.pdf>
- Priulla, G. (2013). *C'è differenza: Identità di genere e linguaggi: storie, corpi, immagini e parole*. Franco Angeli Edizioni.
- Serafini, F. (2023). How multimodality matters in children's literature scholarship. *The Australian Journal of Language and Literacy*, 46, 201–213. <https://doi.org/10.1007/s44020-023-00046-2>
- Serafini, F., & Reid, S. F. (2024). Analyzing picturebooks: Semiotic, literary, and artistic frameworks. *Visual Communication*, 23(2), 330–350. <https://doi.org/10.1177/14703572211069623>
- Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, 25(1), 1–10. <https://doi.org/10.1080/01650250042000078>
- Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science*, 13(4), 599–610. <https://doi.org/10.1111/j.1467-7687.2009.00913.x>
- Stone, L. J., & Church, J. (1968). *Childhood and adolescence: A psychology of the growing person* (2nd ed.). Random House.
- Stroessner, S. J., Benitez, J., Perez, M. A., Wyman, A. B., Carpinella, C. M., & Johnson, K. L. (2020). What's in a shape? Evidence of gender category associations with basic forms. *Journal of Experimental Social Psychology*, 87, 103915. <https://doi.org/10.1016/j.jesp.2019.103915>
- Wild, H. A., Barrett, S. E., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, 77(4), 269–291. <https://doi.org/10.1006/jecp.1999.2554>
- Xue, Z., Li, Q., Zhao, J., & Zeng, X. (2024). An investigation into the relationship between clothing colors and gender stereotyping in children. *Journal of Retailing and Consumer Services*, 76, Article 103559. <https://doi.org/10.1016/j.jretconser.2023.103559>

- Zemach, I., Chang, S., & Teller, D. Y. (2007). Prediction of infants' spontaneous color preferences. *Vision Research*, 47(10), 1368–1381. <https://doi.org/10.1016/j.visres.2006.09.024>
- Zeng, X., & He, W. (2024). Exploring adolescent participation in football: A gender-differentiated structural equation model based on the theory of planned behavior. *Frontiers in Psychology*, 15, 1387420. <https://doi.org/10.3389/fpsyg.2024.1387420>

Authors

Laura Sisti is a graphic and visual designer. She holds a bachelor's degree in Graphic Design and Visual Communication from ISIA Urbino and is completing a master's degree in Design and Visual Communication at Università Iuav di Venezia, with a study period in Design Research at Bern Academy of the Arts. She works as a freelance designer in visual communication and has gained experience in editorial design, advertising campaigns, and graphic projects for cultural institutions and design studios. Since 2025, she has taught Multimedia Design at I.T.T. "Abate Zanetti" in Murano, Venezia.

Luciano Perondi is an associate professor at Università Iuav di Venezia, and deputy director of the User Lab. From 2007 to 2018 he was professor at ISIA Urbino, and director from 2013 to 2016. He is a member of the PhD board in Humanities at the University of Urbino Carlo Bo., a Research Fellow at Hasselt University since 2023, and a member of the co-op CAST SC and Alpaca SC.



Illustrated Letterforms: A Taxonomy of Graphemic Construction

Hugo Maciel 

Escola Superior de Educação do Instituto Politécnico de Viana do Castelo, Portugal
(hugomaciel[at]ese.ipvc.pt)

Abstract: Illustrated letterforms occupy a complex space between writing and image, combining graphemic structures with pictorial strategies that surpass conventional typographic classification. Rather than using typological or stylistic approaches, this article conceptualizes illustrated letterforms as graphemic constructions, in which the structural identity of the letter is determined by visual and pictorial operations. A taxonomy is proposed based on the specific operations by which letters are constructed. This study explores how illustrators and designers combine image and letterform into unified visual units, analyzing methods that preserve legibility while introducing visual, narrative, or symbolic elements. It covers diverse contexts—from manuscript traditions to modern graphic practices—focusing on operations that merge image and graphemic structure. Existing classification systems, especially in manuscript studies, often focus on narrative, iconographic, or stylistic criteria. This article emphasizes construction by operation, examining how graphemic structures are actively shaped by pictorial strategies. By shifting attention from style or function to operation, the framework offers a systematic way to describe and compare illustrated letterforms across media and over time. The taxonomy highlights varying degrees of integration between graphemic structure and pictorial elements, from decoration to full integration. This approach positions illustrated letterforms as a distinct area within visual-graphemic creation, emphasizing design, interpretation, and visual problem-solving rather than treating them as marginal or anomalous.

Keywords: classification; graphemes; graphic design; illustrated letterforms; illustration; letterform construction; lettering; visual culture

@: [ISSUE](#) > [ARTICLE](#) >

Cite this article:

Maciel, H. (2026). Illustrated letterforms: A taxonomy of graphemic construction. *Visible Language*, 60(2), 156–182.

First published online July 16, 2026.

© 2026 Visible Language — this article is **open access**, published under the CC BY-NC-ND 4.0 license.

<https://visible-language.org/journal/>

Visible Language Consortium:

University of Leeds (UK)

University of Cincinnati (USA)

North Carolina State University (USA)

1. Introduction

The alphabet did not originate as a purely abstract system but as a drawn and materially constructed form. Before becoming engravings, metal types, printed impressions, or digital files, letters were traced by hand and shaped by the interaction of tools, gestures, materials, and surfaces. This drawn condition represents a persistent dimension in the history of writing, continually shaping the conception and construction of letterforms.

Scholars have emphasized that letterforms are more than technical solutions to language problems. They are shaped by culture and history. Johanna Drucker (1995), for instance, frames the alphabet as a symbolic and visual artifact whose meanings extend beyond functional transcription, highlighting the cultural and historical contingency of letterforms. Similarly, Robin Kinross (2004) has shown that typographic practices are shaped by specific cultural and ideological conditions rather than constituting neutral or purely functional systems. These perspectives open the possibility of approaching letterforms beyond strictly typographic or reproductive frameworks.

Within this broader context, illustrated letters have persisted across historical periods—from manuscript traditions to contemporary graphic practices—yet remain comparatively under-theorized and are often approached through their decorative, expressive, or pedagogical functions. Existing approaches, particularly in manuscript studies, have developed descriptive frameworks and classifications for decorated, inhabited, and historiated initials, typically based on iconographic, stylistic, or functional criteria (Alexander, 1978; Brown, 1994; Kendrick, 1999). These frameworks offer valuable insights but focus more on representational or functional aspects than construction processes. As a result, less attention has been given to how graphemic structure and pictorial elements merge in letterform construction.

This article offers a complementary perspective by treating illustrated letterforms as graphemic constructions and focusing on the operations used to build them. It examines how illustrators and designers integrate image and letterform into a single visual unit, balancing legibility, representation, and structural coherence. In this sense, the study shifts attention from their functions, meanings, and stylistic features to the constructive processes that merge image and graphemic structure.

Central to this approach is the notion of drawing as a shared operational ground among lettering, illustration, and design. Here, drawing is a broader constructive and image-making process through which letterforms are visually configured across manual, photographic, digital, and material practices. Within graphic design, these processes operate across typographic, illustrative, computational, and spatial contexts. In typographic workflows, drawing typically serves as a preparatory phase focused on reproducibility and systemic consistency. In contrast, within illustration, drawing

remains evident as a process, retaining traces of material engagement and decision-making. Illustrated letterforms employ drawing as a fundamental structural practice that actively constructs the letter's identity, rather than functioning merely as decorative embellishment.

Through selected examples spanning manuscript culture, early print, and contemporary design, this article identifies recurring constructive operations in illustrated letterforms. Rather than offering an exhaustive survey, it establishes a framework for analyzing how letterforms emerge from the integration of graphemic structure and pictorial strategies across varying degrees of visual and structural interaction. In doing so, it positions illustrated letterforms as a relevant area of study grounded in design, interpretation, and visual problem-solving.

2. Terminological and Theoretical Framework

2.1. Grapheme, Letter, and Letterform

In the study of writing systems, scholars commonly distinguish between the grapheme, an abstract unit of writing, and its visual realizations (Sampson, 1985; Daniels & Bright, 1996; Coulmas, 2003). The *grapheme* is the minimal unit that distinguishes meaning within a writing system, whereas its visual realizations may vary across media, styles, and historical periods. In this article, the term *letterform* refers to the visual configuration through which a grapheme becomes perceptible, acknowledging that form is shaped by historical and cultural factors (Drucker, 1995).

This distinction is especially relevant for illustrated letterforms, where visual configurations may diverge significantly from conventional models yet remain recognizable. Rather than treating variation as deviation, this study approaches letterforms as constructions in which graphemic identity is preserved through structural cues, even when transformed by pictorial strategies. In this sense, legibility is understood as a negotiated condition that emerges from the interaction between structural stability and visual transformation (Bringhurst, 2013). Letterforms are treated here as visual configurations that may function as graphemes to varying degrees. Their graphemic status depends on whether their structural identity remains recognizable within a conventional writing system.

2.2. Typography, Lettering, and Drawing

Typography is typically associated with systems of reproducible letterforms, historically linked to printing technologies and, more recently, to digital type design (Kinross, 2004; Carter et al., 2015; Drucker & McVarish, 2013). Its analytical frameworks have often focused on classification, standardization, and optimizing readability across

contexts. Lettering, by contrast, refers to the drawing of letters as distinct visual constructions, often produced for specific contexts and not necessarily intended for systematic reproduction.

While typography and lettering differ in their modes of production and use, both rely on drawing as a formative process. In lettering, drawing typically remains visible in the final form, directly shaping the letter's identity and expression. In typography, drawing often serves as a foundational stage in developing letterforms, though its traces may be subordinated to standardization and reproduction. Illustrated letterforms foreground both the generative and expressive dimensions of drawing. Their construction depends on drawing not only as a generative process but also as a visible and constitutive component of the final form, carrying authorial, expressive, and pictorial qualities that may introduce narrative, metaphorical, symbolic, or ornamental dimensions alongside graphemic function.

In this context, drawing is understood as a broader process of visual construction shared across typography, lettering, illustration, and contemporary design. Following Ingold (2007, 2013), drawing is a process of making that unfolds through gesture, material engagement, and temporal development, rather than merely a representation. While typographic production begins in drawing, it may ultimately dissolve drawing within processes of standardization, whereas illustrated letterforms retain and foreground drawing as a visible and constitutive dimension of form.

2.3. Illustrated Letterforms as Graphemic Constructions

Illustrated letterforms are understood as visual configurations in which graphemic structure and pictorial elements are integrated into a single functional unit. This integration may vary in degree, ranging from additive strategies—where imagery accompanies a relatively stable letterform—to fully integrated constructions in which image and structure are co-dependent.

Such forms are well documented in manuscript traditions, particularly in historiated and inhabited initials, where image and letterform are structurally intertwined (Alexander, 1978; Brown, 1994; Kendrick, 1999). These historical examples demonstrate that the integration of image and graphemic structure is a recurring mode of visual construction.

2.4. Synthesis

Overall, these distinctions establish the conceptual framework of this study. By treating letterforms as graphemic constructions, distinguishing typographic systems from drawing practices, and emphasizing drawing as an operative dimension, the article shifts the analytical focus from classification to construction. This enables analysis

of illustrated letterforms based on how they are made, rather than on stylistic or functional categories.

3. Context: Historical and Analytical Approaches

3.1. From Pictorial Origins to Constructed Letterforms

The relationship between image and letterform dates back to the early development of writing systems, in which many signs originated as pictorial representations of objects and were progressively abstracted and conventionalized (Daniels & Bright, 1996; Coulmas, 2003). While these processes led to the stabilization of graphemic systems, they also show that writing emerged through transformations of visual form rather than from purely abstract principles.

The integration of letter and image becomes more explicit and systematic in late antique and early medieval manuscript traditions. From the 4th and 5th centuries onward, enlarged initials begin to incorporate decorative and figurative elements within the letter's structure. Early forms include letters enriched with filling ornament—vegetal, geometric, or zoomorphic motifs integrated into the letterform while maintaining its recognizable structure (Alexander, 1978; Brown, 1994). In some cases, these elements transform parts of the letter into figures or decorative forms.

This development marks a shift in which the letter functions both as a unit of writing and as a site of visual construction, with graphemic structure and pictorial strategies deliberately combined.

3.2. Manuscript Practices: Structure, Hierarchy, and Interpretation

In medieval manuscript culture, the integration of image and letterform became a sustained and highly developed practice. Initials—commonly described as historiated, inhabited, zoomorphic, anthropomorphic, or foliated—demonstrate a wide range of strategies for incorporating pictorial elements into letterforms (Alexander, 1978; Brown, 1994).

These forms serve multiple functions. Structurally, enlarged initials mark divisions within the text, signaling the beginning of sections, chapters, or books. Hierarchically, variations in scale, color, and ornament establish relationships among textual units. Functionally, they act as navigational devices, enabling readers to locate passages within extended manuscripts in the absence of pagination. At the same time, they often incorporate narrative or symbolic imagery related to the text, functioning as sites of visual interpretation.

In many manuscript examples, pictorial elements function as structural components of the letterform rather than external additions. Human or animal figures may determine curvature, proportion, or internal articulation, shaping the letter from within rather than decorating its surface (Kendrick, 1999). This integration suggests that figuration operates as a constructive principle, establishing the letter's identity through visual and material constraints.

These practices show that combining image and letterform in manuscript culture is a deliberate, systematic visual approach that balances legibility with pictorial expression.

3.3. Print Culture and the Persistence of Illustrated Letterforms

With the transition from manuscript to print, the production of letterforms became increasingly standardized through typographic systems oriented toward reproducibility. Nevertheless, the integration of image and letterform did not disappear. Early printed books frequently left spaces for initials to be added by hand, maintaining continuity with manuscript practices.

Over time, printed initials and ornamental letters evolved through woodcut and engraving techniques, as well as through modular and composite printing processes that enabled variation in color and form (Drucker & McVarish, 2013). These developments show that illustrated letterforms continued to be produced within print culture, adapting to new technical conditions while remaining open to visual construction.

From the early modern period onward, illustrated letterforms expanded into a broader range of contexts, including emblem books, decorative alphabets, and pedagogical materials. In these contexts, letters function as visual and conceptual constructs, often combining recognizable letterforms with figurative or symbolic elements and extending beyond their role in structuring text.

There are multiple examples of similar construction strategies across different letters, indicating that these forms are part of broader visual logics rather than isolated instances. This points to the existence of implicit systems based on shared constructive principles rather than formal standardization.

3.4. Modern and Contemporary Practices

In modern and contemporary contexts, illustrated letterforms extend across multiple domains, including graphic design, illustration, advertising, and artistic practice. Alphabets and individual letterforms are reinterpreted through a wide range of strategies, incorporating human figures, objects, abstract shapes, materials, photographic elements, and digitally generated or three-dimensional forms. Contemporary practices have further expanded these constructive possibilities through computational, performative, and hybrid image-making processes that increasingly challenge stable

distinctions between typography, lettering, and illustration (Miller & Lupton, 1996; McNeil, 2017).

In these contexts, the letterform often functions as a constructed object rather than a fixed unit within a writing system. The problem is no longer just one of textual organization but of how to preserve graphemic recognizability while integrating visual, material, or conceptual elements. This tension between legibility and transformation underlies many contemporary approaches, in which the letter becomes a site for visual experimentation and design.

3.5. Existing Classification Systems and Their Limits

Across these historical contexts, classification systems have been developed to describe illustrated and decorated letterforms. In manuscript studies, terminology such as historiated, inhabited, zoomorphic, anthropomorphic, and foliated initials reflects attention to iconographic content, decorative motifs, and stylistic features (Alexander, 1978; Brown, 1994). In typographic and design-oriented frameworks, classification has often focused on formal characteristics, stylistic categories, and modes of production (Bringinghurst, 2013; Carter et al., 2015). These approaches provide effective tools for identifying and describing visual features in specific contexts. However, they primarily address what letterforms represent, how they appear, or how they function within particular systems. They do not consistently consider how constructive processes integrate image with graphemic structure.

From the illustrator's or designer's perspective, illustrated letterforms can be understood as the result of specific visual problems: how to construct a recognizable letter while integrating pictorial, material, or conceptual elements. In this way, letters that look different might rely on similar fundamental processes, whereas visually similar forms can be created using different constructive methods.

3.6. Positioning of the Present Study

This article considers illustrated letterforms across contexts as examples of visual construction, avoiding specific typologies such as historiated initials, decorative alphabets, or illustrated books, as defined by iconographic, stylistic, or historical categories.

The analysis focuses on how graphemic structures and pictorial elements are combined, rather than on medium, typology, or historical period. By emphasizing the process of construction over simple classification, the study highlights the maker's role, treating illustrated letterforms as the result of specific acts of drawing and visual problem-solving.

The proposed taxonomy complements existing classification systems by providing a framework for understanding how letterforms are generated, transformed, and stabilized through constructive operations across both historical and contemporary practices.

4. Methodology

This study adopts a qualitative, comparative methodology grounded in visual and structural analysis (Rose, 2016). Rather than attempting an exhaustive historical survey, it develops and tests an analytical framework by examining selected case studies from manuscript culture, print, and contemporary graphic practices, including digital and computational contexts.

Works were chosen because they clearly demonstrate distinct modes of interaction between graphemic structure and pictorial elements. The objective is to emphasize analytical contrast rather than statistical representativeness, so each case functions as a paradigmatic instance that clarifies specific constructive strategies.

The analysis focuses on whether pictorial elements generate, transform, or accompany the graphemic structure, and on how recognizability is maintained across varying degrees of integration. Attention is given to aspects such as structural continuity, proportion, spatial organization, and the relationship between positive and negative space, as these contribute to the letter's construction. Iconography, narrative content, and stylistic interpretation are considered only when they directly affect the letterform's structural configuration.

The proposed framework operates heuristically. Its goal is not to exhaust all possible forms of illustrated letters or to impose rigid categories. Instead, it offers a repeatable analytical approach grounded in a central question: how do graphemic structure and pictorial operations interact in the construction of the letterform?

5. A Taxonomy of Illustrated Alphabet Letters

Building on the conceptual and historical framework established in previous sections, this section proposes a taxonomy of illustrated letterforms based on structural and constructive criteria. Rather than organizing these forms by style, iconography, or historical period, the taxonomy focuses on how the image participates in the construction of the letterform.

The taxonomy is grounded in drawing and image-making as shared practices across illustration, lettering, typography, and contemporary design. It considers letterforms as constructed visual entities and focuses on the role of pictorial elements in shaping

their structure. Iconography, narrative, and stylistic interpretation are considered only when they directly affect construction.

5.1. Structural Operativity as Taxonomic Principle

At the core of the taxonomy lies a single operational question: Is the illustrated element structurally operative in the construction of the letterform?

Structural operativity refers to the extent to which pictorial elements participate in generating, maintaining, or transforming the graphemic structure of a letter. Assessing structural operativity is not always straightforward. In many cases, the constructive process by which a letterform was produced is not directly accessible and must be inferred from its visual organization. Pictorial elements may occupy a pre-existing graphemic structure, modify it, or contribute directly to its formation. As a result, structural operativity is best understood as a matter of degree rather than an absolute condition.

One useful heuristic is to consider what happens when pictorial elements are conceptually removed from the letterform. Rather than functioning as a rigid classificatory test, this procedure helps reveal the extent to which the letter's structural identity depends on pictorial construction.

5.2. Taxonomic Structure

The taxonomy is organized into three main categories based on the degree of structural dependency on pictorial intervention: non-form-generative, hybrid-generative, and form-generative.

Table 1 summarizes the taxonomy, defining each category and subcategory by its constructive criterion and degree of structural dependency.

5.3. Non-Form-Generative Letterforms

In non-form-generative cases, the structural framework of the letter remains independent of illustration.

Four main conditions can be identified:

- ▶ Adjunct: illustration is external to the letterform.
- ▶ Ornamental: surface enrichment without structural impact.
- ▶ Container-based: illustration fills a predefined outline.
- ▶ Transformative: the letter is visually reinterpreted while its structure remains intact.

These categories define forms in which illustration enriches or recontextualizes the letter without determining its structure.

Table 1. Structural taxonomy of illustrated letterforms.

Structural Category	Subcategory	Constructive Criterion	Structural Dependency
Non-form-generative	Adjunct	Illustration external to letter structure	Low
	Ornamental	Surface enrichment without structural intervention	Low
	Container-based	Illustration fills predefined outline	Low
	Transformative	Semantic or material alteration; skeleton intact	Low
Hybrid-generative	—	Shared structural responsibility (partial modification)	Medium
Form-generative	Figurative-constructive	Figures generate structure	High
	Object-constructive	Objects replace structural components	High
	Spatial/architectural-constructive	Spatial relations define geometry	High
	Material-constructive	Physical process determines form	High
	Abstract-constructive	Non-representational forms generate structure	High

5.4. Hybrid-Generative Letterforms

Hybrid-generative letterforms occupy an intermediate position. A recognizable structural framework remains, but pictorial elements intervene directly in its construction. Illustration and structure work together to produce the final form, establishing a condition of shared structural responsibility. Specific components of the letter—such as strokes, terminals, or counters—may be replaced, extended, or modified, while overall recognizability is preserved. Hybrid conditions may also emerge through the combination of different constructive strategies, allowing generative and non-generative operations, or multiple generative subcategories, to coexist within the same letterform. Similar combinations can be observed in manuscript traditions, where different modes of image–letter integration often coexist within the same form (Alexander, 1978; Camille, 1992).

5.5. Form-Generative Letterforms

Form-generative letterforms are those in which illustration is structurally responsible for the letter's existence. The letter doesn't pre-exist as a stable outline; it emerges from the constructive process itself.

Several constructive strategies may be observed:

- ▶ Figurative-constructive: human or animal figures generate structural components such as strokes, curves, and proportions.
- ▶ Object-constructive: objects or artifacts replace or constitute parts of the letter's structure.
- ▶ Spatial/architectural-constructive: the letter emerges from spatial relations, volume, or architectural organization.
- ▶ Material-constructive: physical processes such as cutting, folding, or deformation generate the form.
- ▶ Abstract-constructive: non-representational forms generate the letter structurally.

In all cases, illustration constitutes the letter.

5.6. Letter, Set, and Alphabet

The taxonomy treats the individual letter as its primary unit of analysis, while recognizing that letters often appear as part of sets or complete alphabets. Sets and alphabets may be structurally homogeneous, using similar constructive strategies across letters, or heterogeneous, combining different approaches. They may nevertheless exhibit strong stylistic, conceptual, or thematic coherence, even when relying on different constructive strategies across individual letterforms.

5.7. Legibility and Structural Tension

Across all categories, recognizability remains a necessary condition. The taxonomy assumes that a letterform must retain sufficient graphemic identity to be identified as a letter, even when pictorial elements play a significant role in its construction.

According to van Leeuwen (2006), legibility is understood as a culturally acquired competence rather than a fixed formal property. Recognition may therefore depend not only on the structure of an individual letterform but also on the context in which it appears. In sets and complete alphabets, relationships between letters can provide additional cues for identification, allowing more radically transformed forms to remain legible. This is particularly relevant in borderline cases, where a letter's recognizability may depend less on its individual form than on its participation in a broader visual system.

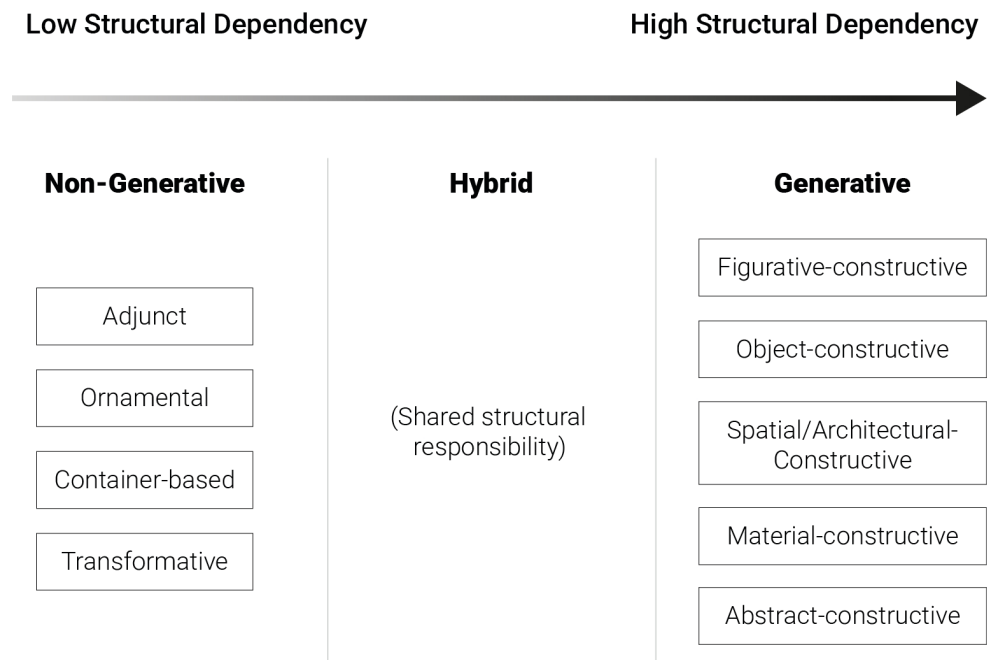


Figure 1. Structural continuum of illustrated alphabet letters. Diagram positions illustrated letterforms according to the degree of structural dependency on illustration.

5.8. Analytical Continuum

To clarify the relationships between categories, Figure 1 presents the taxonomy as a continuum of structural dependency.

The diagram positions illustrated letterforms along a gradient from non-generative conditions—where the letter remains structurally independent—to fully generative constructions, in which illustration determines or replaces its structure. Hybrid conditions occupy the intermediate zone, characterized by shared structural responsibility.

5.9. Scope and Limits of the Taxonomy

The categories proposed here should not be understood as rigid or mutually exclusive types. Illustrated letterforms often combine multiple constructive strategies, and borderline cases may occupy intermediate positions along the continuum of structural dependency.

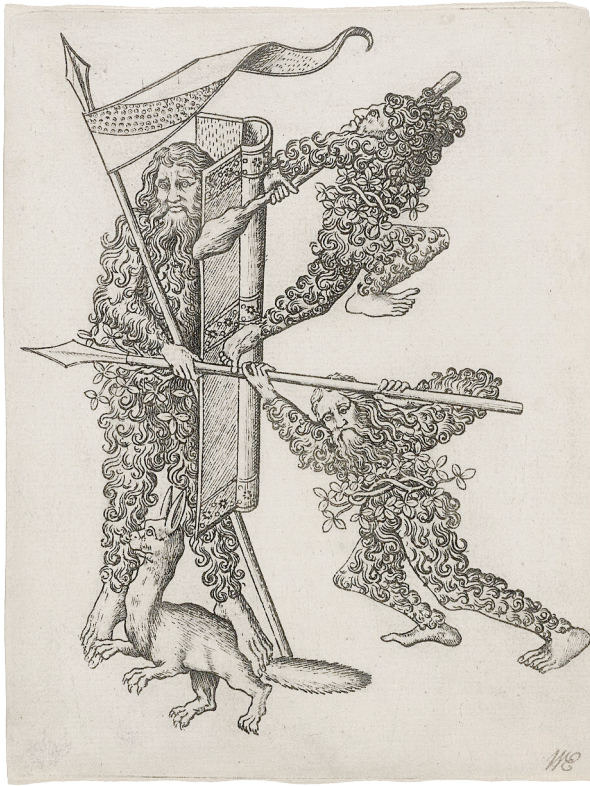


Figure 2. Master E. S., *The Letter K*, 1466–1467. Engraving. Form-generative and figurative-constructive letter in which human figures, animals, and objects collectively generate the graphemic structure through their strategic arrangement and simultaneously construct a narrative scene. Public domain.

6. Applying the Taxonomy: Structural Analysis of Illustrated Letterforms

The proposed taxonomy defines structural operativity as the primary criterion for distinguishing illustrated letterforms. This section applies the framework to selected examples, examining how pictorial elements contribute to the construction of letterforms across the categories of form-generative, hybrid-generative, and non-form-generative.

6.1. Form-Generative Letterforms

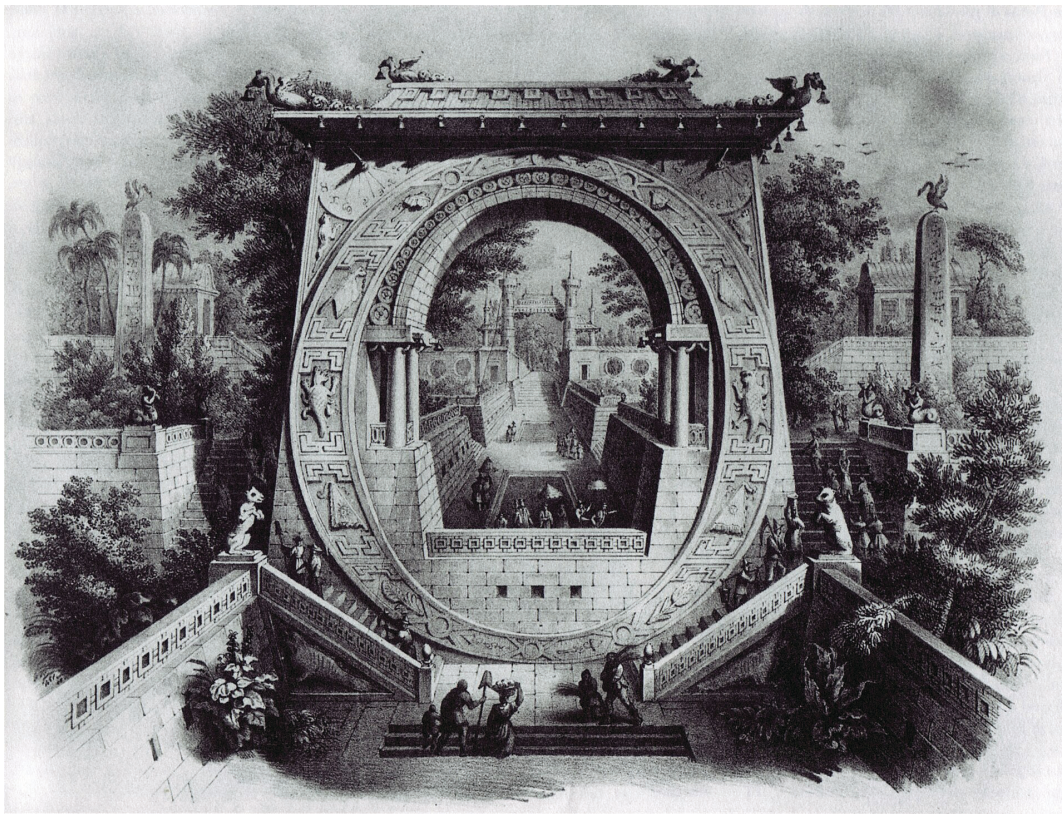
Form-generative letterforms are those in which illustration is structurally responsible for the letter's existence. The letterform emerges from the constructive process itself rather than from a stable, pre-existing outline, with pictorial elements generating its graphemic structure. These constructive strategies may arise through manual, photographic, digital, material, or computational processes.

Master E. S.'s letter (Figure 2) is constructed through the strategic arrangement of human figures, animals, and objects, whose combined configuration gives rise to the graphemic structure. The figurative elements remain recognizable and are not constrained by a pre-existing typographic outline. Instead, strokes, junctions, and proportions emerge from their organization, ensuring letter recognition while simulta-



Figure 3 (left). James Wojcik, *Alphabet Series*, 1988. Photography. Object-constructive letter assembled from photographed elements, in which material mass replaces the stroke and determines thickness and structural form. Copyright James Wojcik 2026. Used with permission.

Figure 4 (below). Antonio Basoli, *Alfabeto Pittorico*, 1839. Engraving. Spatial-constructive letter generated through architectural composition, in which built structures articulate stroke direction, counterform, and proportional organization. Public domain.



neously constructing a narrative scene. The letterform is therefore form-generative and figurative-constructive.

James Wojcik's sculptural letter (Figure 3) exemplifies object-constructive logic. Recognizable objects assemble into structural components, replacing the drawn stroke with material mass. Thickness and termination derive from the elements' physical properties. In this case, the letter N is constituted by the playing cards themselves and their spatial arrangement.

Antonio Basoli's letter (Figure 4) demonstrates spatial/architectural construction. The form arises from the organization of built elements. Volume and void determine proportion, and spatial relations generate counterform. Geometry is constructed rather than traced.

The letter in Figure 5 exemplifies material-constructive logic. The form emerges from the behavior of viscous matter. Gravity, tension, and cohesion determine thickness and curvature. The structure results from the process rather than from a predefined outline.

Casey Concinha's letter (Figure 6) illustrates abstract-constructive generation through computational processes. The form emerges from layered trajectories and algorithmic



Figure 5. Foreal Studio, *The Sculpted Alphabet*, 2014. Digital 3D rendering. A material-constructive letter generated through viscous behavior, with thickness, curvature, and terminal formation determined by fluid tension and material cohesion. Used with permission of Foreal Studio.

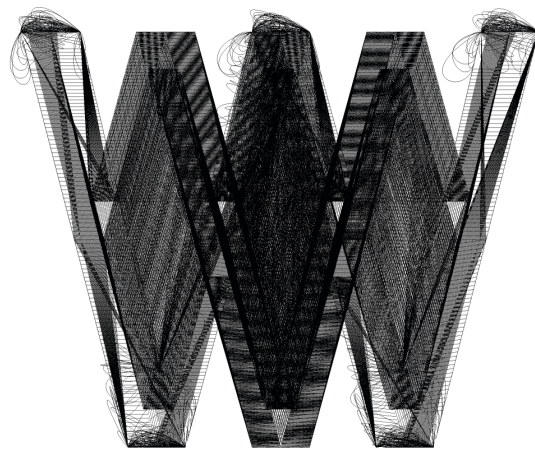


Figure 6. Casey Concinha, *exp.01*, generative typography. Computational typographic experiment developed using Processing and the Geomorative library. An abstract-constructive letterform generated through procedural and algorithmic operations, in which layered trajectories and iterative repetition produce structural continuity and letterform emergence. Used with permission of Casey Concinha.

repetition, generating stroke, density, and structural continuity. Rather than following a predefined typographic outline, the letterform is created through the behavior of the generative system itself.

Across these examples, structural responsibility lies in the constructive process itself, whether pictorial, material, or computational.

6.2. Hybrid-Generative Letters

Hybrid-generative letterforms occupy an intermediate condition in which part of the letter's structure remains recognizable while pictorial elements assume responsibility for other structural components.



Figure 7. *Bible de Saint-Sauveur de Majeure*, c. 999–1098. Illuminated manuscript. Bibliothèque municipale de Bordeaux. Romanesque historiated initial in which figural, zoomorphic, and vegetal elements shape the curvature and internal structure, contributing to the construction of the letterform. Public domain.



Figure 8. Lucas Kilian, *New ABC Booklet*, 1627. Engraving. Cleveland Museum of Art. Hybrid-generative letter in which figurative elements partially replace structural components while preserving the overall proportions. Public domain.



Figure 9. Miriam Martincic, *Barcelona*, 2023. Mixed media. Hybrid-generative letter in which illustrated elements modify structural components while preserving overall proportional coherence. Used with permission of Miriam Martincic.

The Romanesque initial from the Bible de Saint-Sauveur de Majeure (Figure 7) exemplifies this condition. Recognizable structural features of the letter remain perceptible, yet figural elements actively shape curvature, closure, and internal articulation. The final form emerges from the interaction of these two constructive contributions.

Lucas Kilian's engraved letter (Figure 8) similarly replaces portions of the structure with figurative and ornamental elements while preserving the overall proportion. Structural coherence results from the interaction between the underlying geometry and the pictorial insertion.

Miriam Martincic's letter (Figure 9) presents a comparable dynamic. The illustrated elements integrate into the stroke system, modifying the letter without fully dissolving its structure. The form remains recognizable while structurally transformed.

Hybrid letterforms make visible a negotiated condition in which structure and image co-operate. The letter is neither fully autonomous nor fully generated by illustration.

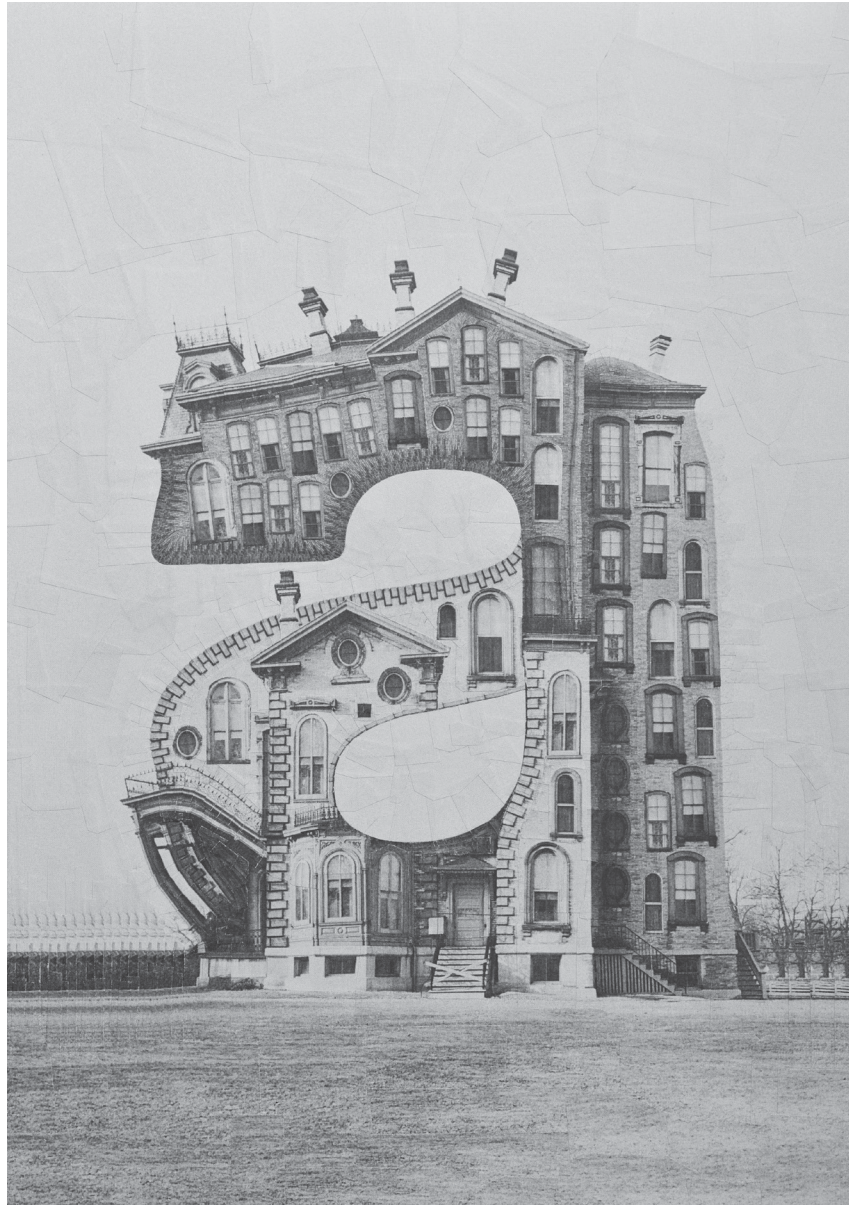


Figure 10. Lola Dupré, *Architecture Series*, 2016. Collage. Container-based letter in which a stable typographic outline functions as a structural boundary for internal collage imagery. Used with permission of Lola Dupré.

6.3. Non-Form-Generative Letters

Non-form-generative letterforms preserve the letter's structural autonomy. Illustration does not determine its construction.

Lola Dupré's letter (Figure 10) exemplifies container-based logic. A stable outline serves as a boundary within which collage imagery is inserted. Despite internal complexity, proportion and structure remain unchanged.

Mateusz Witczak's letter (Figure 11) functions ornamentally. Surface enrichment does not alter the structural organization. The underlying letter structure remains intact.

Andreas Scheiger's letter (Figure 12) demonstrates transformative logic. The form appears materially altered, yet its structural proportions remain unchanged. Transformation affects perception rather than construction.

Jean-Pierre Humbert's configuration (Figure 13) illustrates adjunct logic. Illustrative elements appear externally to the letter, without intervening in its structure.



Figure 11 (above left). Mateusz Witczak, 2025. Vector Illustration. Ornamental letter in which decorative detail enriches the surface without altering structural proportion or stroke logic. Used with permission of Mateusz Witczak.



Figure 12 (above right). Andreas Scheiger, *Evolution of Type*, 2014. Sculpture and photography. Transformative letter in which typographic form undergoes material metamorphosis while preserving its underlying structure. Used with permission of Andreas Scheiger.



Figure 13 (right). Jean-Pierre Humbert, *ABC Book*, 2011. Engraving. Adjunct letter in which external illustrative elements frame the letter without altering its structure. Used with permission of Jean-Pierre Humbert.

These examples clarify the distinction between structural intervention and visual enrichment. In non-form-generative cases, the letter remains structurally independent.

6.4. Systems and Sets

While classification operates at the level of the individual letter, illustrated letterforms frequently appear as part of sets or complete alphabets.

The alphabet attributed to L. E. M. Jones (Figure 14) constitutes a structurally homogeneous, form-generative system. A consistent, constructive logic underlies the entire alphabet, with letterforms emerging from the arrangement of landscape, architectural, and natural elements. Although individual letters differ in their specific configurations, they share the same generative principles and visual language. Antonin Caulo's alphabet (Figure 15) forms a homogeneous hybrid system. Figurative and structural elements interact consistently throughout the series.



Figure 14. L. E. M. Jones, printed and published by Charles Joseph Hullmandel, c. 1818–1860. Lithograph on paper. Structurally homogeneous form-generative alphabet in which landscape, architectural, and natural elements consistently generate letter structure across the set. © The Trustees of the British Museum. Shared under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license.



Figure 14 (continued)

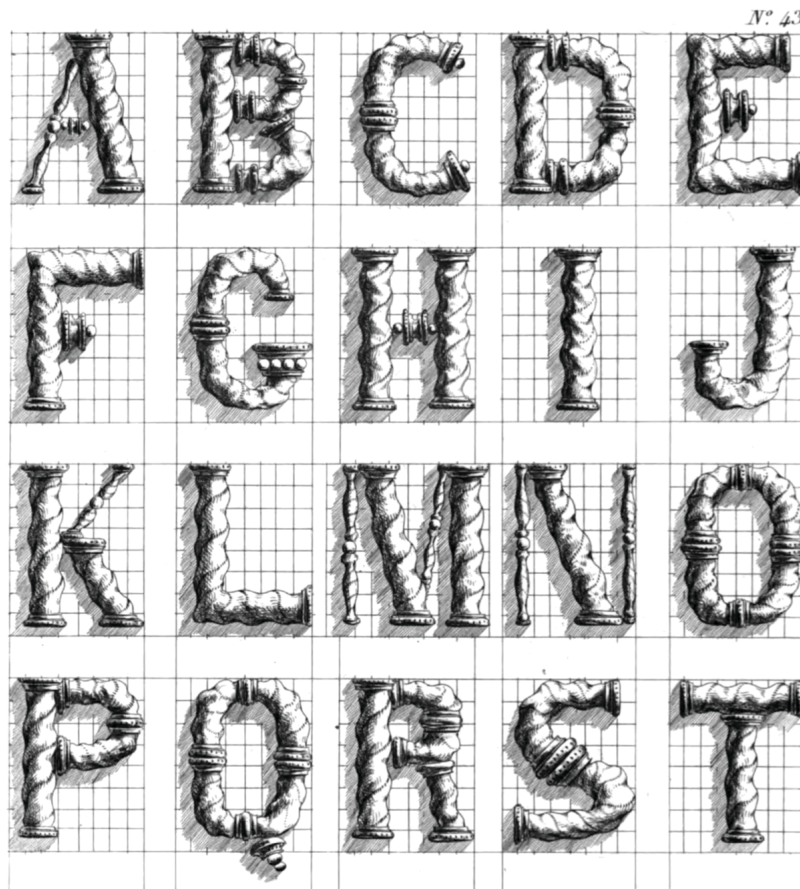


Figure 15. Antonin Caulo, *Nouvelle collection de lettres de différents genres...*, 1856. Engraving. A homogeneous hybrid-generative alphabet in which figuration and typographic proportion share structural responsibility across the set. Public domain.



Figure 16. M/M Paris, *The Miu Miu Type*, in *An ABC of Actions, Behaviors and Comportments*, 2018. Graphic design. Non-form-generative alphabet with consistent illustrative density accompanying stable typographic skeletons. Used with permission of M/M Paris.

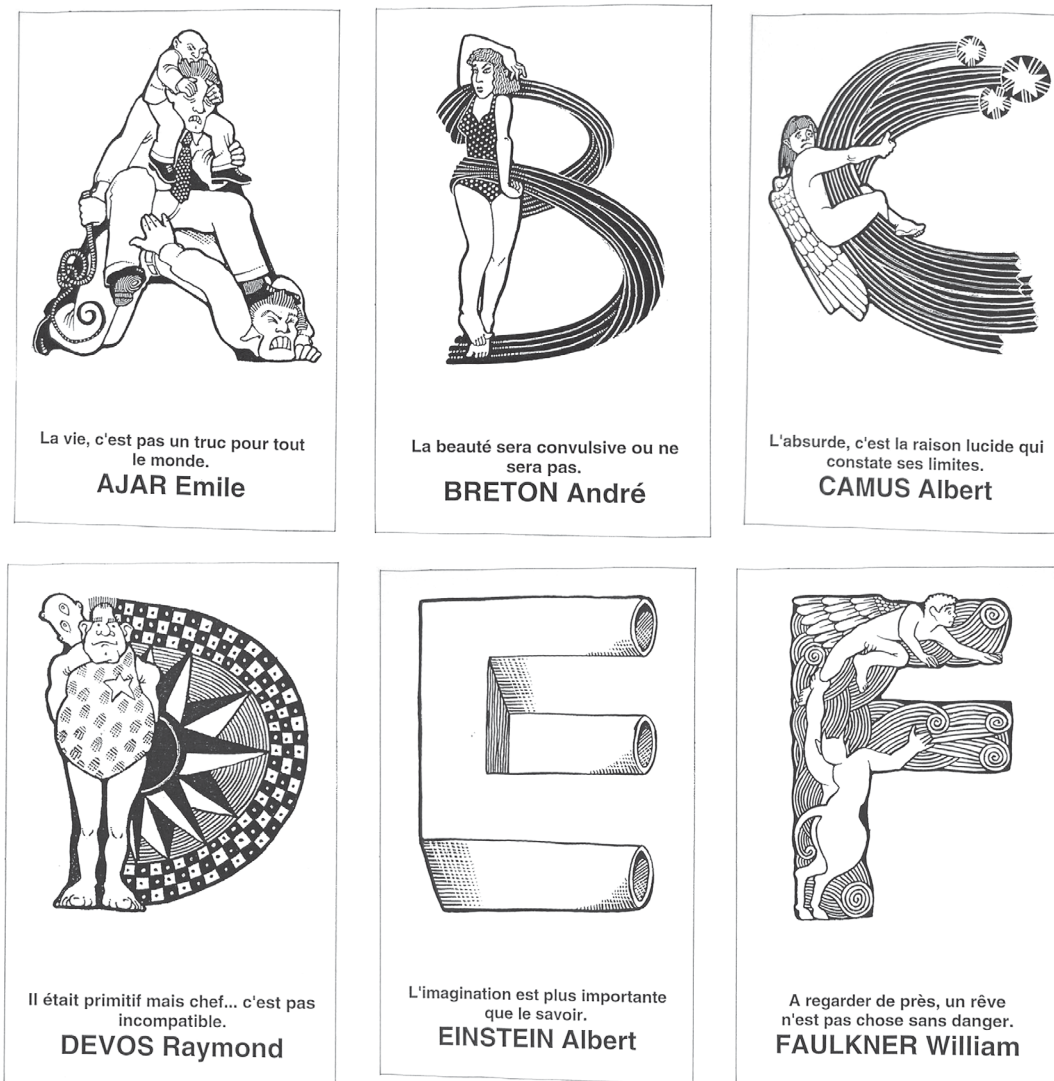


Figure 17. Álvaro Parés, *Abc Abscons*, 2006. Drawing. Structurally heterogeneous alphabet that combines generative and non-generative strategies across letters. Used with permission of Álvaro Parés.

The Miu Miu alphabet by M/M Paris (Figure 16) exemplifies a homogeneous non-form-generative system. The illustration accompanies the letters without altering their structure. Álvaro Parés' set (Figure 17) demonstrates structural heterogeneity. Different letters employ distinct constructive strategies, requiring analysis at the level of each individual form. Anthony Earnshaw's alphabet series (Figure 18) similarly combines multiple constructive logics. Stylistic coherence does not imply structural uniformity. Generative and non-generative processes coexist within the same alphabet, and several letters can be classified as hybrid forms.

These examples show that system coherence may operate independently of structural consistency. Classification must therefore remain attentive to individual constructions.



Figure 18. Anthony Earnshaw, *Secret Alphabet No. 2*, 1985. Lithograph. A structurally heterogeneous alphabet that combines generative, non-generative, and hybrid letter-construction processes within a coherent visual style. Used with permission of Gail Earnshaw.

6.5. Ambiguity and Limits of Classification

No taxonomy fully eliminates ambiguity. Borderline cases challenge stable categorization and reveal the limits of discrete classification.

Kelsey Robinson's letter (Figure 19) occupies a borderline position in which generative and transformative conditions coexist. Structural intervention is present, but not dominant. Sabeena Karnik's letter (Figure 20) occupies a borderline position between generative and transformative conditions. Although the letter is materially constructed through paper manipulation, its form continues to reflect a stable and recognizable typographic framework. Even within otherwise consistent systems, variation may still occur. Antonin Caulo's alphabet (Figure 21) contains moments that approach full generativity.

These cases demonstrate that illustrated letterforms operate along a continuum of structural dependency rather than within rigid categories. While the taxonomy provides consistent analytical criteria, some borderline cases inevitably remain open to interpretation.

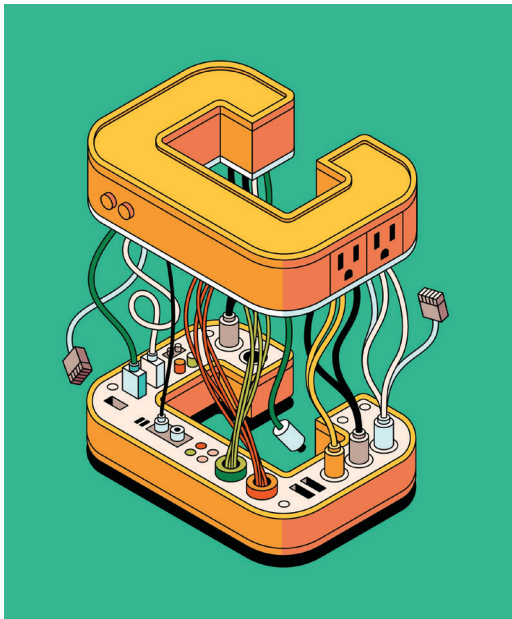


Figure 19. Kelsey Robinson, *36 Isometric Days of Type*, 2023. Vector illustration. Letter illustrating the limits of categorical classification. Structural additions provide significant visual information, yet the typographic form remains recognizable and largely autonomous. Used with permission of Kelsey Robinson.

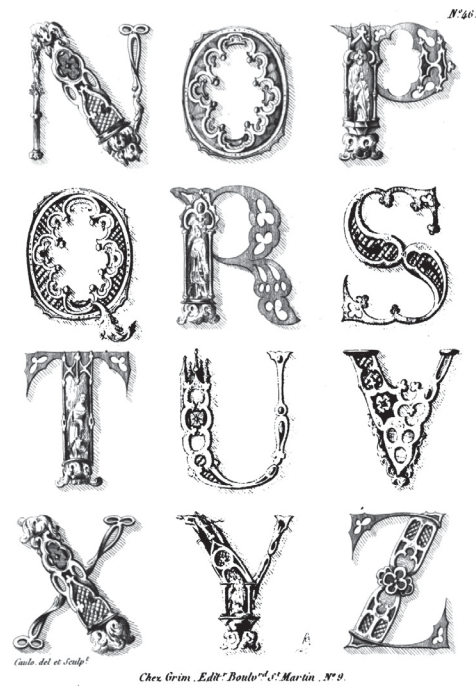


Figure 21. Antonin Caulo, *Nouvelle collection de lettres de différents genres...*, 1856. Engraving. Homogeneous hybrid-generative alphabet in which figurative and ornamental elements partially replace typographic strokes while preserving overall proportional structure. Public domain.



Figure 20. Sabeena Karnik, *Tropical Oasis*, 2016. Paper art illustration. A letter formed from layered paper elements arranged around a recognizable typographic framework, illustrating the overlap between material generation and structural preservation. Used with permission of Sabeena Karnik.

The taxonomy therefore functions as an analytical tool for assessing the distribution of structural responsibility, while acknowledging that classification remains interpretive and context-dependent.

7. Conclusion

This article has proposed a taxonomy and analytical framework for illustrated letterforms grounded in structural operativity. Rather than classifying these forms by what they depict or signify, it has focused on how they are made—the operations by which image and graphemic structure are combined into a single visual unit.

From this perspective, illustrated letterforms are sites of construction in which drawing actively participates in the production of the letter. The framework developed here directs attention to the processes by which letters are built and to the distribution of structural responsibility between graphemic and pictorial elements.

The taxonomy distinguishes three conditions—form-generative, hybrid-generative, and non-form-generative—understood as degrees of structural dependency on pictorial operations. These are best understood as positions along a continuum rather than as fixed categories. Their value lies in making visible how letterforms emerge through different constructive logics, whether through substitution, modification, or coexistence with a typographic skeleton.

Across historical and contemporary examples, similar operations recur: bodies become strokes, objects replace structure, space generates geometry, and material processes define form. These strategies persist over time, while the conditions for their production and interpretation continue to evolve.

Every intervention in a letter is a structural act. Even when the typographic skeleton appears intact, decisions about texture, material, or integration influence how the letter is constructed and perceived.

The taxonomy accommodates intermediate and hybrid conditions, reflecting the continuum of structural dependency that characterizes illustrated letterforms. As a result, classification remains interpretive and context-dependent. The framework functions as an analytical tool for assessing how structural responsibility is distributed across different forms of letter construction.

Future research may extend this framework to non-Latin scripts or to computational environments where construction is distributed across systems rather than confined to singular gestures. The underlying question, however, remains the same: how does the letter come into being?

Ultimately, illustrated letterforms foreground the alphabet as a constructed visual system. They reveal drawing and image-making as operative conditions of letterform construction and demonstrate the continuing role of visual invention in the formation of letters.

8. References

- Alexander, J. J. G. (1978). *The decorated letter*. Thames & Hudson.
- Bringhurst, R. (2013). *The elements of typographic style* (4th ed.). Hartley & Marks.
- Brown, M. P. (1994). *Understanding illuminated manuscripts: A guide to technical terms*. Getty Publications.
- Camille, M. (1992). *Image on the edge: The margins of medieval art*. Harvard University Press.
- Carter, R., Meggs, P. B., Day, B., Maxa, S., & Sanders, M. (2015). *Typographic design: Form and communication* (6th ed.). Wiley.
- Coulmas, F. (2003). *Writing systems: An introduction to their linguistic analysis*. Cambridge University Press.
- Daniels, P. T., & Bright, W. (Eds.). (1996). *The world's writing systems*. Oxford University Press.
- Drucker, J. (1995). *The alphabetic labyrinth: The letters in history and imagination*. Thames & Hudson.
- Drucker, J., & McVarish, E. (2013). *Graphic design history* (2nd ed.). Pearson.
- Ingold, T. (2007). *Lines: A brief history*. Routledge.
- Ingold, T. (2013). *Making: Anthropology, archaeology, art and architecture*. Routledge.
- Kendrick, L. (1999). *Animating the letter: The figurative embodiment of writing from late antiquity to the Renaissance*. Ohio State University Press.
- Kinross, R. (2004). *Modern typography: An essay in critical history* (2nd ed.). Hyphen Press.
- McNeil, P. (2017). *The visual history of type*. Laurence King.
- Miller, A., & Lupton, E. (1996). *Design writing research: Writing on graphic design*. Phaidon Press.
- Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials* (4th ed.). Sage.
- Sampson, G. (1985). *Writing systems: A linguistic introduction*. Stanford University Press.
- van Leeuwen, T. (2006). Towards a semiotics of typography. *Information Design Journal*, 14(2), 139–155. <https://doi.org/10.1075/idj.14.2.06lee>

Author

Hugo Maciel is a lecturer in illustration and design at the School of Education, Polytechnic Institute of Viana do Castelo, Portugal. He is also an illustrator and visual artist. His teaching, research, and practice focus on drawing, illustrated alphabets, and the intersections between illustration, typography, and visual communication.



Are You Sure? Exploring the Challenges and Opportunities of Using Generative AI in Communication Design Projects: A Cross-Cultural Case Study of Student Reflections, Bias, and Critical Engagement

Halldór Björn Halldórsson 

Department of Health Sciences, Innovation & Design, Mälardalen University, Eskilstuna, Sweden
(halldor.bjorn.halldorsson[at]mdu.se)

Abstract: Regardless of whether we approach Generative AI (GenAI) with apprehension or enthusiasm, the fact remains that AI has already made its way into many aspects of personal and professional life—including the field of communication design and, by extension, the education of future designers. While there is currently no consensus on how to incorporate GenAI into design education, the technical development within AI is happening faster than most of us can keep up with, and the ways in which we can utilise AI are becoming more varied by the day. Considering that the integration of AI in design will only increase in the coming years, it is essential to build strategies for preparing our students for their future careers. This paper presents a joint project conducted in autumn 2025 by two universities in Sweden and Australia that tasked students with creating posters representing an unfamiliar city using only AI-generated imagery. Through reflective writing, discussions, and analysis of their visual outcomes, the study examines how students negotiated both the opportunities and obstacles posed by GenAI. The project is framed against current literature on GenAI in design and education and seeks to contribute to the ongoing discourse on how design education can adapt to emerging technologies while maintaining critical and creative integrity. Findings show that while students valued GenAI's capacity for rapid ideation, they also encountered issues such as factual inaccuracy, stylistic inconsistency, and cultural bias. These challenges underscored the importance of maintaining strong foundational design skills, as well as developing students' critical awareness of the ethical, cultural, and semiotic implications of AI-generated imagery. The study contributes insights into how communication design education might meaningfully integrate GenAI while sustaining critical and creative integrity.

@: [ISSUE](#) > [ARTICLE](#) >

Cite this article:

Halldórsson, H. B. (2026). Are you sure? Exploring the challenges and opportunities of using generative AI in communication design projects: A cross-cultural case study of student reflections, bias, and critical engagement. *Visible Language*, 60(2), 183–209.

First published online July 16, 2026.

© 2026 Visible Language — this article is **open access**, published under the CC BY-NC-ND 4.0 license.

<https://visible-language.org/journal/>

Visible Language Consortium:

University of Leeds (UK)

University of Cincinnati (USA)

North Carolina State University (USA)

Keywords: communication design education; design pedagogy; generative AI (GenAI); ideation and creativity; prompting; reflective practice

1. Introduction

The rapid emergence of generative AI (GenAI) has the potential to unsettle familiar methods and routines within design practice and design education. For many, designers and educators alike, GenAI represents a technological shift that brings with it a great deal of ambiguity; a tool that shows promise to expand creative options while also disrupting conventional assumptions about authorship, skill, and visual literacy. Communication design is very much at the centre of this transition, being a field occupied with the relationships between form, meaning, culture and representation. Yet, considering the immense attention paid to AI in media, news and society today, empirical research on how design students experience and make sense of such tools remains relatively scarce.

The aim of this study is to address that gap by exploring how third-year communication design students in Sweden and Australia engaged with GenAI during a cross-cultural, project-based assignment. Rather than evaluating AI's technical capabilities, the project sought to understand the interpretive, experiential, and meaning-making dimensions of students' encounters with GenAI: how they used it, where they struggled, what they valued, and how they articulated the relationship between AI processes and their own design judgement.

The research question guiding this project was: *How do communication design students regard the use of generative AI in a cross-cultural design assignment?* This was inspired by discussions I have had with colleagues and design students on the integration of GenAI in design practices. The study examines not only students' creative and technical challenges, but also the cultural, semiotic, and ethical tensions revealed in their attempts to represent places unfamiliar to them. In doing so, the project does not position GenAI as either positive or negative, but as a culturally situated production tool that mediates between intention and outcome, often in ways students do not expect.

Design education has always evolved in response to new technologies, from desktop publishing to motion graphics to web design. But GenAI alters the relationship between designer, tool, and output in ways that demand renewed critical attention. As students discovered, the promise of instant visual generation also brings aesthetic instability, factual drift, stereotypical outputs, and a heightened need for interpretive skill. These tensions reveal important insights for educators seeking to support students' development in an increasingly AI-augmented landscape.

By analysing student reflections, discussions, and visual artefacts, this study contributes to ongoing conversations about how design education can integrate GenAI while preserving the critical, semiotic, and ethical foundations of the discipline. The findings have particular relevance for educators navigating questions of creative agency, prompting literacy, data-driven representation, and the enduring importance of foundational design skills in a rapidly shifting technological environment.

2. Background

The literature search for the study was primarily done using Google Scholar, Scopus, Web of Science, and ScienceDirect, supplemented by reference tracing in relevant articles. Search strings included “(visual/graphic or communication) design and AI + GenAI” “education and AI + GenAI,” and “(visual/graphic or communication) design education and AI + GenAI.” This search process provided a cross-disciplinary view of current literature on GenAI in creative and educational contexts. While the field is expanding rapidly, the literature remains fragmented, particularly in relation to design-specific pedagogies. The following subsections present an overview of some relevant strands of this still-forming body of research.

In a comprehensive literature review, Yusuf et al. (2024) provided an overview of recent articles that focused on GenAI and education. The authors showed that interest in these topics is very much on the rise, but an analysis of the educational disciplines included in the review concluded that design was not among those disciplines that had been the topic of more than one article (out of 407 publications), signalling a significant absence of design-focused research in a field undergoing rapid technological change.

Design researchers are starting to react however, demonstrated by a more current overview of the research agenda in creative contexts by Heigl (2025). But her text calls for bigger efforts, as Heigl identified several research gaps that need attention within design and artistic research, such as the long-term effects of GenAI, its socio-economic implications on creative industries, the building of ethical frameworks and studies on people’s perception of AI-generated material. Against this backdrop, the following sections outline the key themes currently shaping discussions on GenAI within design practice and design education.

2.1. The Rise of GenAI

While the term “AI” has become nearly ubiquitous, being used for various sorts of machine learning, this paper will primarily focus on GenAI, which entails a range of artificial intelligence models designed to produce material (images, videos, text or audio), using models that mimic aspects of human cognition (Dwivedi et al. 2023;

Yusuf et al. 2023). Through machine learning and using enormous databases, these systems identify patterns and structures in existing material, which in turn become the raw material the systems use to generate new material, using probabilistic models (Mandapuram et al., 2018).

Currently we find ourselves at what will undoubtedly be considered a turning point in the widespread adoption of GenAI, with rapid uptake across professional and personal contexts. The pace and scale of this shift have prompted comparisons to earlier technological revolutions; for example, Geoffrey Hinton has likened the significance of AI to the invention of the printing press or even the wheel (CBS Mornings, 2023). Yet, unlike previous technological shifts, such as the arrival of desktop publishing (DTP) in the early 1990s—which I personally experienced as it affected my own field of graphic design—today’s AI tools span an immense range of applications, from text-to-image generators like Firefly and DALL·E to more complex systems like Stable Diffusion. This breadth opens GenAI to both novices and experts alike in practically all parts of society.

The rapid development of AI has been described by some as one of the most disruptive technological changes society has encountered (Păvăloaia & Necula, 2023), presenting significant challenges for education. It also amplifies concerns around technological opacity. Drawing on Latour’s (1999) notion of “black-boxed” technologies—tools so familiar that their inner workings become invisible—the internal logic of GenAI remains obscured, even as it shapes creative processes. This reality underscores the importance of developing conscious, iterative approaches to working with GenAI in design practice.

2.2. Creativity and GenAI in Design Practice

The question of whether AI can be considered “creative” has no definitive answer, in part because creativity itself is difficult to define. Many descriptions highlight divergent thinking, reflection, and the ability to depart from linear or habitual approaches (Abraham & Windmann, 2007; Cropley et al., 2022; Runco & Jaeger, 2012). Mumford et al. (2012) emphasise the importance of metacognition—the ability to assess, select, and develop ideas. By these criteria, AI alone is unlikely to qualify as creative.

However, GenAI can play a productive role within human–machine collaboration. These tools can rapidly generate numerous variations, providing designers with material that can spark ideas, refine directions, or serve as visual inspiration. Seidel et al. (2018) suggest that this may lead to a shift in creative practice, which they refer to as a “triple-loop approach” (p. 51). The authors describe such an approach as involving three elements: the generation of alternatives based on input parameters, a feedback loop where the designer evaluates and modifies the outcomes, and a third step consisting

of mutual learning, where the designer gains insight into the machine's mental models and the machine learns more about the designers' thought processes.

The topic of creative decision-making by AI systems has raised concern and caused apprehension by some, seeing a risk of potentially undermining a fundamental human trait with unforeseen repercussions. Yet scholars like Sims (2024) propose a more nuanced view, comparing GenAI use to the discipline of photography. Mastery in photography requires extensive experimentation, the cultivation of an aesthetic sensibility through trial and error, and the ability to discern what should be kept or discarded. Similarly, skilful use of GenAI demands the development of prompting strategies, visual judgement, and conceptual clarity—abilities that remain squarely within the domain of human creativity.

2.3. Integrating GenAI into Design Education

The integration of GenAI into education has provoked diverse reactions. Some celebrate its potential, while others express mechanophobia—a fear that AI may diminish human creative value (Fleischmann, 2024). While some might argue that the easy access to GenAI will eliminate the need for educated designers, the fact remains that most people are not autodidacts and prefer guidance, which places educators in a crucial role. Saéz-Velasco et al. (2024) found that while professional designers see value in AI, students must retain competence in composition, colour, and form to avoid producing homogeneous or biased outputs, a concern also raised by Bianchi et al. (2023). The importance of maintaining human creativity is similarly argued by Marble (2025), stating that GenAI should be introduced only after students have cultivated independent design skills.

On the pedagogical front, Putjorn and Putjorn (2023) found no evidence that GenAI promotes creativity or collaboration on its own, though it can support ideation when paired with advanced prompting skills. This has led artists and scholars—such as Boris Eldagsen, who coined the term “promptography” (Bainbridge, 2023)—to frame prompting itself as a creative act. Ethical and contextual awareness is also critical: Hwang and Wu (2024) caution that students must understand the risks of cultural misrepresentation and stereotyping. Gray (2024) adds that as AI continues to reshape creative labour, educators must prepare students not only technically but also ethically, legally, and professionally.

2.4. Pedagogical Opportunities and Challenges

Outside the field of design, AI has demonstrated potential for personalised learning and instructional support (Hwang & Wu, 2024; Islam et al., 2023), increasing self-efficacy by offering inspiration and instructional support. In fields such as music education, AI

can facilitate contextual and interactive learning, though excessive reliance may reduce students' motivation to develop fundamental creative skills, such as creating melodies and harmonies (Merchán Sánchez-Jara et al., 2024). García-López and Trujillo-Liñán (2025) have pointed out three primary benefits of AI: personalisation, equitable access, and the development of critical competencies. The authors stress the need to address ethical questions, regulations, and definitions of educational quality. Albar Mansoa (2024) highlights similar benefits and concerns, noting that fine art students express worry about ethical implications and threats to artistic integrity.

Within design education, the sheer generative power of AI introduces new challenges. Omran Zailuddin et al. (2024) note that GenAI can overwhelm students with excessive visual material, potentially stifling creativity rather than enhancing it. This positions educators as guides who must help students navigate abundance with intentionality and critical awareness.

Matthews et al. (2023) remind us that design education has always adapted to new technologies—from the emergence of DTP to the rise of web and new media. At the same time, Dziobczenski and Person (2017) highlight how the contemporary designer's role increasingly includes facilitation, project management, and interpersonal communication. These expanded expectations intersect directly with AI-enabled workflows, making social, reflective, and entrepreneurial skills more important than ever.

2.5. Sustainability and Technological Vulnerability

Sustainability is seldom addressed in the literature on GenAI in education. Large-scale AI systems rely on massive data-centre infrastructure, with uncertain long-term economic and environmental viability (Costa et al., 2024; Ivanov & Lyu, 2024). These concerns extend into pedagogical settings, where students depend on systems that may be fragile or inaccessible.

Today, many of the widely used GenAI systems rely on a cloud-based infrastructure, which can introduce dependencies on external services, licensing models, or institutional access. However, it is important to point out that some generative models can also be used locally, depending on hardware capacity and technical expertise. These variations are demonstrative of the rapidly evolving and heterogeneous nature of modern GenAI, and raise broader questions about future access, reliability, and long-term sustainability in educational settings.

2.6. The Human Role in AI-Augmented Design

Pratschke's (2024) concept of generativism frames the relationship between designer and GenAI as a cooperative process, where algorithmic capabilities complement human criticality, contextual understanding, and reflective judgement. While GenAI

can support analysis and generation, it cannot replace the nuanced creativity required for meaningful design. Pratschke's work reinforces the importance of helping students develop an informed, reflective, and creative approach to GenAI—not a passive or purely receptive one.

Multiple studies reiterate that effective use of GenAI depends on a foundation of strong design skills (Ray, 2023; Xu et al., 2019). Without aesthetic literacy and conceptual grounding, users risk creating repetitive, biased, or visually misleading outcomes. In an educational setting, this underscores the need to cultivate reflective engagement, ensuring that students remain active authors rather than passive recipients.

2.7. Conclusion: Adapting Design Education for an AI-Driven Future

Across the literature, there is broad agreement that the inclusion of GenAI is inevitable and that curricula and teaching methods must be adapted for a new reality in design. There is a consensus on the ability of GenAI to function as a partner in a creative process, facilitating quicker and more varied conceptual approaches, prototyping and sketching. To successfully use GenAI in a way that will lead to innovative solutions, authors also see a need for designers to balance the development of new skills, such as prompting, programming and the management of large amounts of generated output, with the traditional, more hands-on skills of design. Furthermore, the literature emphasises the importance of supporting students' critical approach to GenAI, building their awareness of the ethical, societal and cultural risks inherent in generated outcomes.

While some scholars propose comprehensive restructuring of curricula (Davis, 2018), others favour more gradual, fluid adaptation (Kelly, 2018). What unites these perspectives is recognition that future communication designers will require both timeless and emerging skills—traditional practices such as composition and typography, alongside prompting, ethical reasoning, and management of large volumes of generated material.

The overarching challenge lies in balancing the opportunities of GenAI with the need for critical, culturally aware, and ethically grounded pedagogy. This study contributes to that conversation by exploring students' reflections on using GenAI in design processes—providing insights into how educators might meaningfully integrate these tools in ways that support both creativity and criticality.

3. Methodology

3.1. Research Design and Analytical Orientation

The study used a qualitative, exploratory case-study design to investigate how communication design students experienced and reflected on the use of GenAI within a

project-based assignment. The research is grounded in a constructionist epistemology (Burr, 1995), positioning students' reflections, interactions, and design artefacts as socially constructed accounts shaped by cultural, pedagogical, and technological contexts. Data were analysed using reflexive thematic analysis (Braun & Clarke, 2022), complemented by a semiotic examination of the visual outcomes, to provide insight into how students negotiated issues of aesthetics, factual accuracy, and cultural representation when working with GenAI tools.

The methodological aim was to produce an interpretive account of students' meaning-making processes, rather than to measure predefined variables or evaluate tool performance. This aligns with the study's broader pedagogical focus on how GenAI intersects with design cognition, reflection, and cross-cultural collaboration.

3.2. Researcher Standpoint

My approach to GenAI is critical–constructive, recognising the technology as a new design material capable of expanding ideation while also introducing ethical ambiguities, technical constraints, and risks of stereotype reproduction. As an educator and professional graphic designer, I acknowledge that my interpretation of the data is shaped by commitments to critical thinking, semiotic awareness, and the preservation of creative agency—values that influence how student reflections and outcomes are analysed. My aim is not to position GenAI as either a threat or a solution, but to critically examine how students engage with it, negotiate its limitations, and articulate its role within their design processes.

3.3. Study Context and Project Structure

The study was carried out collaboratively between Mälardalen University (MDU) in Eskilstuna, Sweden, and Curtin University (CU) in Perth, Australia, in September and October 2025. The project formed part of third-year, undergraduate design courses: *ITE356 Editorial Design—Project* (MDU, $n = 13$) and *GRDE3009 Visual Communication* (CU, $n = 45$). Students at each university worked in groups to design posters representing the city of their partner university. All visual material had to be generated exclusively using Adobe Firefly Image Model 4—the latest Firefly version available at the time of the study (released April 2025).

The project followed Kolb's (1984) principles of experiential learning, aiming to engage the students in an iterative cycle of learning, moving between reflection and conceptualisation as they interpreted and represented another city. Collaborative Online International Learning (COIL) (Amaral et al., 2023) served as a framework for an environment which would place students in a cross-cultural and collaborative context. The iterative nature of the project is also grounded in Schön's (1983) reflective practice,

as students developed their work through reflection-in- and on-action, fuelled by intercultural feedback and dialogue. Students met in five joint online sessions, shared iterative work on a Miro board, and conducted cross-cultural research discussions. The cross-cultural structure of the project is not treated as a primary variable of comparison, but as a productive condition through which issues of representation, bias, and interpretation become more visible.

3.4. Participants and Data Collection

The fact that the project was carried out in cities separated by 13,500 km and six time zones called for a flexible approach in the project's design. Five joint sessions were scheduled in the morning in Eskilstuna and late afternoon in Perth, and Microsoft Teams was used for the online meetings. Some technical issues were encountered, but for the most part the sessions ran smoothly. At both sites the Teams sessions were projected on large screens, with some of the students also accessing the sessions through their personal devices. Data for this paper were collected primarily from the MDU group, as these students were instructed to write weekly reflections as part of the course requirements. The following empirical materials form the basis of the analysis:

- ▶ Weekly student reflections: 4–6 short texts per student (100–200 words), submitted after each session.
- ▶ Group interview: A final reflective discussion with MDU students, documented by my own researcher notes and summarised.
- ▶ Observation notes: Researcher notes from all joint sessions and local work sessions, documenting prompting challenges, peer feedback, and design choices.
- ▶ Design artefacts: Twelve final posters, intermediate sketches, and mood boards created using Adobe Firefly.
- ▶ Cross-institutional interactions: Notes from student discussions during joint MDU–CU sessions.

The CU students did not submit written reflections in the same way, and the findings therefore predominantly reflect the perspectives of the MDU group. This limitation is acknowledged explicitly to maintain methodological transparency.

3.5. Ethical Considerations

All participants provided informed consent. Students were informed that participation was voluntary and independent of assessment. Data were anonymised during transcription and stored securely. Ethics procedures followed institutional guidelines for pedagogical research at both universities.

3.6. Data Analysis

Reflexive thematic analysis. The analysis of student reflections, interview data, and observation notes followed the six phases of reflexive thematic analysis (Braun & Clarke, 2022):

1. Familiarisation: Reading reflections, interview notes, and observational logs multiple times.
2. Coding: Generating inductive codes related to student experiences, frustrations, prompting strategies, cultural interpretation, and design judgement.
3. Generating initial themes: Collating codes into candidate themes capturing patterns in student meaning-making (e.g., factual instability, prompting literacy, creative control, cultural generalisation).
4. Reviewing themes: Refining and adjusting thematic boundaries to ensure coherence across data sources.
5. Defining and naming themes: Producing thematic descriptions emphasising conceptual insight over frequency-based reporting, in line with reflexive thematic analysis principles.
6. Producing the narrative: Integrating themes with literature and semiotic analysis in the Findings and Discussion sections.

Reflexive thematic analysis does not seek inter-coder reliability or consensus checking; instead, coding is treated as interpretive and situated, consistent with the constructionist perspective underpinning the study.

Semiotic and visual analysis of artefacts. To analyse the 12 final posters and intermediate sketches, a Barthesian semiotic framework (Holt, 2017; Skaggs, 2017) was applied:

- ▶ Denotation—literal visual content (landmarks, figures, landscapes).
- ▶ Connotation—cultural associations, stylistic cues, tonality, implied narratives.
- ▶ Myth—overarching cultural stories or stereotypes reproduced in the AI-generated imagery.

This lens allowed for examination of discrepancies between student intentions and GenAI outputs, particularly in terms of factual accuracy, cultural generalisation, and stylistic coherence. Insights from this semiotic framework are primarily reflected in the analysis of factual inaccuracies, cultural generalisations, and mismatches between intended and generated imagery presented in the Findings section.

Integration of multiple data sources. Themes were developed through triangulation of:

- ▶ what students *said* (reflections, interview)
- ▶ what they *did* (observations)
- ▶ what they *produced* (visual artefacts)

This approach provided a multi-layered understanding of student experiences, frustrations, and strategies for negotiating GenAI's constraints.

3.7. Methodological Limitations

The study is constrained by:

- ▶ a small sample of reflective data ($n = 13$)
- ▶ reliance on novice GenAI users
- ▶ limited duration (six weeks)
- ▶ restriction to a single AI tool and version (Firefly Image Model 4)
- ▶ absence of reflection data from CU students
- ▶ inherent subjectivity of reflexive thematic analysis

These factors shape the study's transferability and are further addressed in the Discussion.

4. Poster Designs—Conceptual Stages

The primary assignment tasked the Swedish students with designing posters representing Perth, while the Australian students created posters depicting Eskilstuna. Although students could choose a fictitious target audience—tourists, new businesses, or residents—all groups ultimately converged on solutions that were, to varying degrees, targeted toward tourists. All visual material had to be generated exclusively using Adobe Firefly, whereas the conceptual development drew on surveys, AI-generated mood boards, mood boards built from online research, and iterative peer feedback.

- ▶ Week 1: Joint session. Students from both universities were introduced to the project and to their partner groups. They completed a six-question survey about their own city, after which responses from MDU were distributed to the CU students and vice versa. Instructional resources for Adobe Firefly were provided, and students were encouraged to begin familiarizing themselves with the tool.
- ▶ Week 2: Joint session. Using the survey responses as prompts, students at both universities created initial mood boards with Adobe Firefly and posted them on a shared Miro board.
- ▶ Week 3: Joint session. Students expanded their research using online sources and discussions with their partner groups, generating imagery, text, and stylistic references to refine their mood boards.
- ▶ Week 4: Joint session. Each group presented two initial poster sketches on the Miro board and exchanged peer feedback.
- ▶ Week 5: Local sessions. Students worked separately at their respective universities, developing the final versions of their posters using the feedback received.

- Week 6: Joint session. Final posters were presented on the Miro board (see Figures 4–15).

During the second session, students presented their first mood boards generated entirely with Adobe Firefly. Many described the experience as “going in blindly,” noting the difficulty of generating imagery for places they had little prior knowledge of. The initial outcomes were often described as stereotypical or generic. As one student put it: *“we struggled with the depicted persons’ proportions and the fact that they sometimes seemed to be missing body parts or conversely having too many of them”* (student 1, author’s translation).

Peer discussions highlighted that the GenAI tool frequently produced fictionalised representations of both cities, reinforcing generalised or archetypal ideas of Scandinavia and Australia. Notable examples included recurring mountain landscapes in images of Eskilstuna—despite the region’s flat geography—and quokkas appearing in images supposedly depicting Perth, even though these animals primarily inhabit islands off Australia’s west coast (Figure 1). These early results underscored the need for at least minimal factual familiarity with the subject matter to guide the GenAI outputs effectively.

In the third session, students were asked to refine their material by incorporating both feedback and additional research. Many attempted to produce imagery with greater

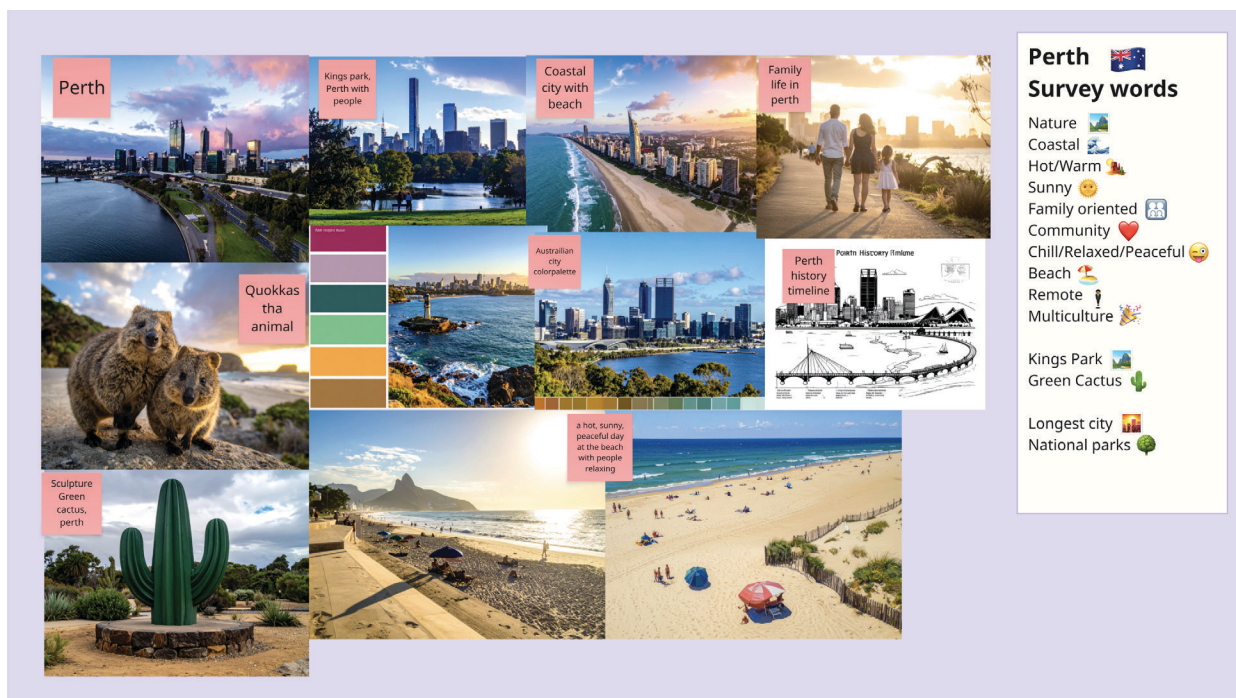


Figure 1. Team 12’s initial mood board.

Draft Poster(s) / Team 11



Figure 2. Team 11's drafts.

specificity, hoping to move toward more accurate depictions. Yet reflections reveal continuing difficulties: hands with too many fingers, child-like characters substituted for adults, and limbs positioned unnaturally. Attempts to correct these issues through prompting often made them worse or produced equally implausible alternatives. One student observed: *"I have also noticed that the faces of background figures lack detail; eyes and mouths are barely distinguishable in the images generated by Firefly"* (student 5, author's translation). Despite these challenges, some groups managed to generate imagery that captured at least broadly recognizable geographic elements, forming a starting point for their poster concepts (Figure 2).

By the fourth session, students presented two developed poster sketches (Figure 3). Many again expressed frustration with the challenge of shaping GenAI outputs to align with their intentions. Groups reported inconsistencies in how the system responded to prompts or partial disregard of instructions. As one student remarked: *"We noted that the AI tool had difficulty when it came to negations; even though I specifically prompted 'no mountains' it included mountains in all the images, continuing to do so in my second attempt"* (student 4, author's translation). Both Swedish and Australian students described similar issues—struggling to remove incorrect geographical features or adjust image perspectives. The AI's unpredictability led some to describe it as having *"a mind of its own"* (student 8, author's translation).

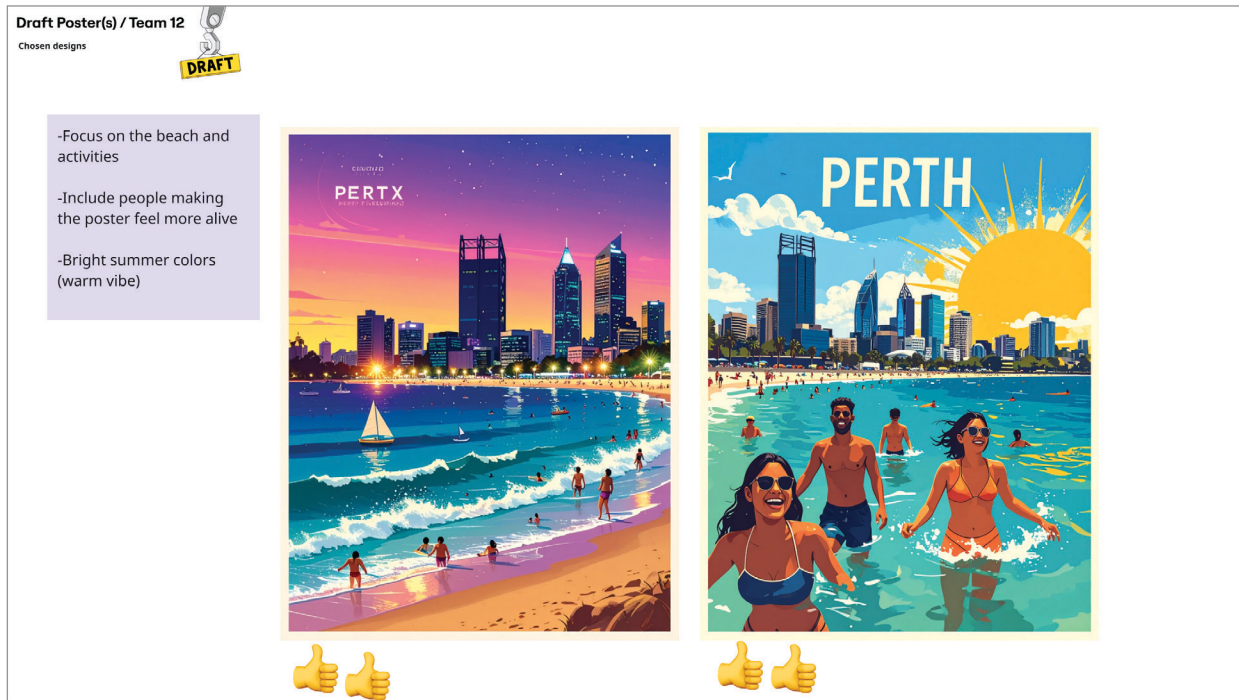


Figure 3. Team 12's poster sketches.

In Week 5, students worked independently to finalise their posters. They selected one of their sketches and were allowed to combine generated elements and add text, but they were not permitted to create any imagery manually. Groups continued to engage with their partner universities and received instructor feedback as they shaped their final outcomes.

5. Poster Designs—Final Outcomes

5.1. Aesthetics

In the project's final session, groups presented their completed posters. All groups opted for colour versions of their designs, with the least saturated design being a two-tone winter cityscape (Figure 6). Feedback from earlier sessions influenced aesthetic refinements, resulting in more accurate depictions of skies, clouds, and overall tonal balance.

In terms of visual style, approximately half of the groups adopted retro-inspired aesthetics reminiscent of earlier design eras (Figures 6–9, 12, and 13). Three groups developed vector-style illustrations (Figures 7, 8, and 10), one group produced a somewhat romanticised hand-drawn style (Figure 1), and two groups maintained photorealistic approaches (Figures 5 and 14).

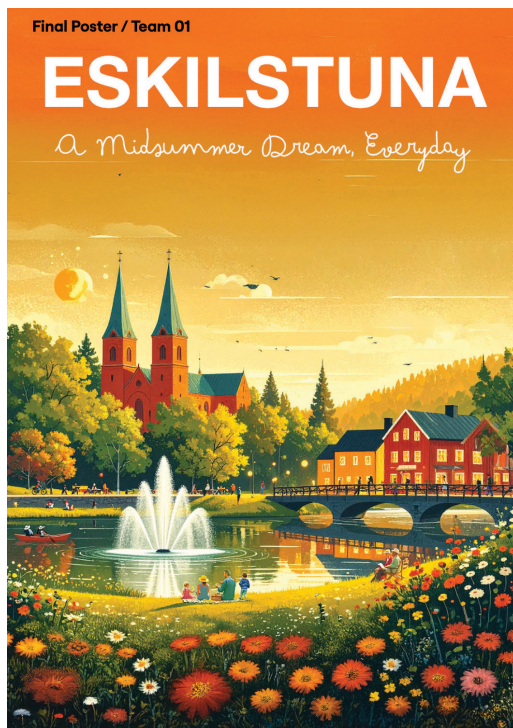


Figure 4. Team 1's poster design.

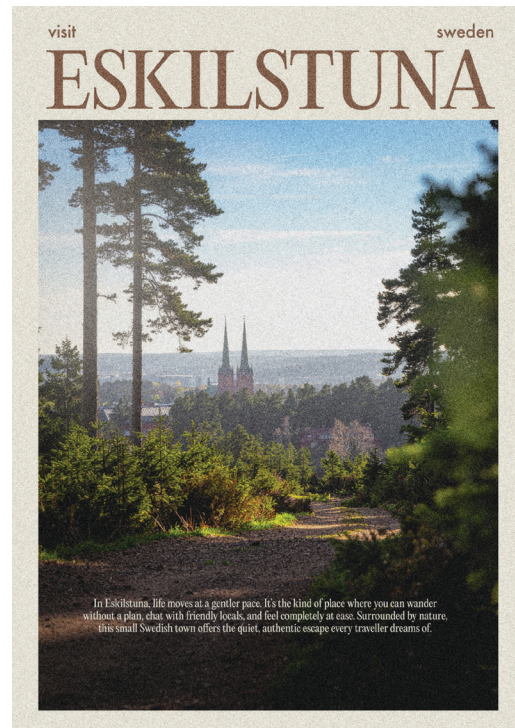


Figure 5. Team 2's poster design.

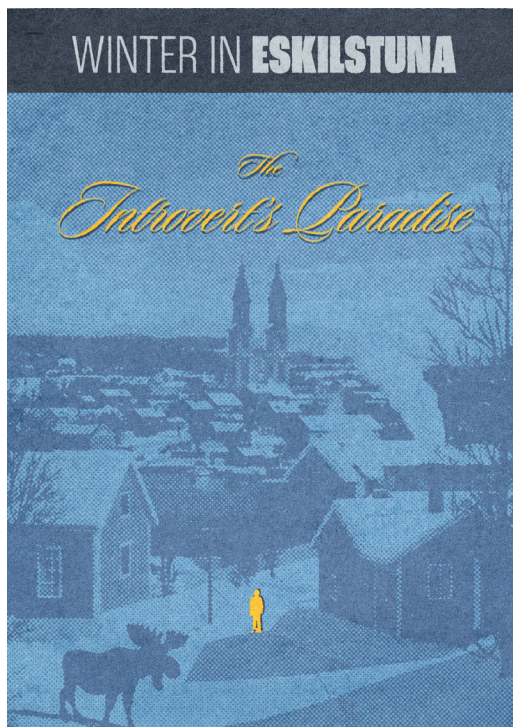


Figure 6. Team 3's poster design.



Figure 7. Team 4's poster design.



Figure 8. Team 5's poster design.



Figure 9. Team 6's poster design.



Figure 10. Team 7's poster design.



Figure 11. Team 8's poster design.

Interestingly, most groups initially experimented with photorealistic imagery but later pivoted toward illustration-based styles—nine out of twelve final posters used illustration rather than photography. Students explained that illustration allowed them to evoke a sense of place without needing to resolve specific details or achieve strict visual accuracy. As one student reflected: “...it is less forgiving to use photo-realistic images, as you expect them to correspond to reality” (student 3, author’s translation).

Discussions also revealed the cognitive effort required to select and refine AI-generated material. Students mentioned the frustration of not being able to arrive at outcomes corresponding to their ideas and mental images, while also acknowledging their own lack of experience with GenAI tools. They furthermore recognised that both theoretical and practical design skills were crucial for guiding the generative process and achieving acceptable results.

5.2. Factual Accuracy

A consistent challenge across all groups was managing factual inaccuracies generated by the AI. Students encountered animals not native to the depicted regions—such as quokkas appearing in Perth images and reindeer in Eskilstuna—alongside non-existent geographical features like mountains and beaches appearing in incorrect contexts. From a semiotic perspective, many of the inaccuracies observed at the denotative level (e.g., misplaced landmarks, incorrect fauna) disrupted the intended connotative meanings of the posters.

To navigate these issues, many groups chose to move away from specific details and toward more general atmospheres or symbolic representations. Attempts to depict people engaged in recognizable activities, such as playing a guitar, often failed due to distorted anatomy or unnatural body positioning.

Despite these challenges, students felt compelled to include certain key elements—landmarks, nature, or architectural fragments—to situate their posters. Swedish students attempted to generate images of the Perth skyline, while Australian students repeatedly turned to Eskilstuna’s iconic twin church towers. Yet achieving an accurate representation proved difficult.

Additional difficulties arose when GenAI added unwanted elements: Union Jack ribbons appearing in response to “wavy frame” prompts, or failure to produce classic “postcard style” layouts. Students experimented with alternative prompting strategies but reported mixed success.

The end results generally conveyed broad connotative impressions—“a European town” or “an Australian city”—rather than precise depictions. Although this type of general-



Figure 12. Team 9's poster design.



Figure 13. Team 10's poster design.

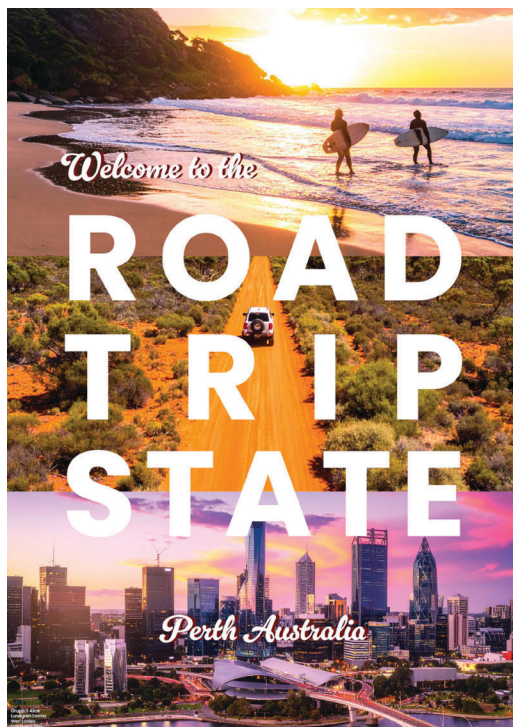


Figure 14. Team 11's poster design.



Figure 15. Team 12's poster design.

ization is not unusual in tourism-oriented material, the vagueness here stemmed from the GenAI system's limitations, rather than from deliberate design choices.

An important limitation of the assignment was the omission of a specific audience. While the groups all ultimately chose to focus on tourists, a more defined profile of the intended audience might have informed and shaped the outcomes better, avoiding generic visual qualities. Future iterations of the project would benefit from a more defined audience group, perhaps resulting in material of higher communicative quality.

5.3. Cultural Aspects

Through peer feedback and iterative refinement, groups were able to incorporate at least some cultural markers into their posters. For Perth, this included references to ocean proximity and an active, outdoors-oriented lifestyle (Figures 14 and 15). The Eskilstuna posters more frequently communicated a sense of smaller-city calm, historical character, and closeness to nature (Figures 5 and 7).

Students found it challenging to depict culture with specificity, again gravitating toward more symbolic or atmospheric cues. Misalignments often occurred between student intention and GenAI interpretation. For example, prompts such as “recycling” yielded images of trash cans or landfills, and “industry” produced outdated imagery like smoking chimneys.

Cross-cultural differences also influenced interpretations. Students in Eskilstuna (population ~70,000) struggled to relate to Perth students' description of their 2.3-million-inhabitant city as “small,” contributing to differences in visual and conceptual emphasis between the two groups. Despite this diversity, both cities are culturally and ethnically varied, making it striking that all posters reflected predominantly Western and Caucasian norms. The absence of student commentary on this highlights how generative systems can reproduce colonial imaginaries even in multicultural and Indigenous contexts, and underscores the need to strengthen students' critical engagement with such biases.

5.4. Final Remarks

In their final reflections, students emphasised how challenging it was to generate imagery that matched their intentions: “...it's been hard to get AI to do as you want... like adding details to images, such as swans or small boats” (student 5, author's translation). They described difficulties producing coherency in the outputs, inconsistencies between identical prompts issued at different times, and challenges building upon promising but flawed outcomes. Several attributed these issues to their lack of experience with GenAI, though they also acknowledged limitations inherent in the tools.

Students repeatedly recognised that solid design skills remained essential for guiding the generative process and mitigating erratic or stereotypical outcomes. At the same time, they identified valuable aspects of GenAI—particularly its ability to generate large numbers of variations quickly. Some even likened unexpected or inaccurate outputs to analogue “happy accidents,” capable of sparking new ideas and shifting their perspectives.

It is important to acknowledge that the use of Adobe Firefly (Image Model 4) shaped the outcomes underlying this study. Other GenAI tools and later iterations of the model offer more advanced results, and future studies should compare multiple tools to further develop our understanding of student interaction with GenAI. A conference workshop is planned for autumn 2026 which will be inspired by this project, which will offer the possibility to compare different GenAI tools and hopefully reveal further insights about the nature of their outcomes.

6. Discussion

This study set out to explore design students’ experiences of using GenAI within a cross-cultural, project-based design assignment. The reflections and discussions generated throughout the project reveal a nuanced combination of curiosity, frustration, creative experimentation, and critical awareness. The findings align with earlier research positioning GenAI as a promising yet unstable design partner—one that expands conceptual exploration while simultaneously exposing significant limitations.

Students from both universities described GenAI as a powerful tool during the ideation phase. The ability to rapidly generate large quantities of visual material supported divergent thinking and enabled groups to test stylistic alternatives quickly, resonating with Seidel et al.’s (2018) description of the “triple-loop” process in which designers iteratively evaluate and modify machine-generated outcomes. Students also noted instances in which unexpected results—what some compared to analogue “happy accidents”—supported shifts in perspective and encouraged the exploration of new visual directions.

However, the project also brought forward clear limitations of GenAI. Many students struggled to align the generated imagery with their conceptual intentions, particularly when attempting to refine details or produce visually accurate representations of unfamiliar locations. Errors such as distorted hands, unnatural body positions, missing or incorrect landmarks, and inconsistent stylistic features were repeatedly described as sources of frustration. These challenges echo concerns raised in the literature about the persistence of inaccuracies, biases, and inconsistencies in AI-generated imagery (Bianchi et al., 2023).

The project's cross-cultural structure intensified these issues. Students were tasked with representing cities they had never visited, relying initially on GenAI to construct a visual understanding of Perth or Eskilstuna. Yet the results were often stereotypical or factually incorrect, leading students to question the reliability of GenAI as a research tool. For example, Swedish students struggled to remove beaches from images of Perth, while Australian students were repeatedly presented with mountain ranges around Eskilstuna. These inaccuracies support Hwang and Wu's (2024) argument that GenAI poses contextual and cultural risks, especially when used without sufficient background knowledge.

As the project progressed, many groups shifted toward less literal, more stylised visual approaches—such as vector illustrations, retro aesthetics, or symbolic compositions—to circumvent GenAI's tendency toward factual distortion. This move away from specificity toward abstraction was not an initial conceptual decision but rather a practical adaptation to the technology's constraints. This reflects Omran Zailuddin et al.'s (2024) concern that GenAI's generative power can sometimes narrow rather than broaden creative exploration by forcing students toward generality.

Throughout the project, students expressed a heightened awareness of the importance of their foundational design skills. Their reflections suggest that without strong abilities in composition, colour, typography, semiotics, and visual analysis, it becomes difficult to interpret or steer AI-generated content meaningfully. This observation aligns with Saéz-Velasco et al.'s (2024) argument that foundational skills remain essential for avoiding homogeneous or biased outputs, and with Marble's (2025) insistence that students must develop independent design judgement before relying on GenAI.

The project also encouraged students to critically consider their own agency within AI-mediated workflows. Several reflected on whether GenAI might undermine human creativity or whether its outputs risk flattening cultural nuance—concerns echoed in broader discussions about the changing nature of design labour (Gray, 2024; Pratschke, 2024). The students' recurring descriptions of GenAI as having "a mind of its own" further point to Latour's (1999) notion of black-boxed technologies and the opacity of automated systems.

These findings also connect to broader concerns about the conditions under which future designers will enter the profession and the shifting technological infrastructures that shape creative work. In this context, it is also important to consider that we are still in the early stages of a society-wide adaptation to AI—an adaptation currently facilitated by the widespread availability of tools that are often accessible through institutional licensing or limited free-use tiers. This accessibility enables designers and students to experiment with GenAI without significant financial barriers. However, as generative technologies become more deeply embedded in commercial ecosys-

tems, there are indications that some business models may shift toward proprietary ecosystems, subscription structures, or platform consolidation. This trajectory aligns with Doctorow's (2025) critique of the "enshittification" cycle, in which digital systems gradually transition from user-focused to increasingly restrictive as market power consolidates. Adopting a forward-looking stance is therefore essential, to support alternative paths of development, where open-source models and less centralised AI ecosystems can emerge. Without such awareness, designers risk becoming dependent on tightly controlled platforms in ways that may constrain their creative autonomy, limit independent decision-making, and reduce flexibility in future design practices.

A recurring theme across the students' reflections was the need for more structured opportunities to develop critical evaluative skills when working with GenAI. While the project introduced reflective writing and peer critique, the open-ended nature of these activities meant that students often focused on surface-level frustrations rather than deeper questions about assumptions, cultural framing, authorship, and design intent. More scaffolded reflection prompts—focusing on what the tool did well, where it failed, what assumptions guided the prompting, and how alternate strategies might have changed the outcome—could have supported richer, more rigorous forms of critical engagement. Embedding such scaffolding is particularly important in an AI-augmented design environment where tools can mask their own biases and where students must develop confidence in questioning outputs rather than accepting them at face value.

These findings also connect to the question of how design education can prepare students for an uncertain and rapidly evolving professional landscape. As GenAI becomes increasingly integrated into industry workflows, entry-level design tasks may change or diminish, placing greater emphasis on students' ability to adapt, critically evaluate outputs, and identify opportunities where human judgement remains essential. The ethical responsibility of educators extends beyond introducing specific tools; it involves equipping students with durable capacities—critical thinking, conceptual clarity, reflective judgement, and cultural awareness—that will remain relevant regardless of how GenAI continues to develop.

Ultimately, the findings reinforce the idea that using GenAI effectively requires not only technical skill but also critical reflection, ethical awareness, and a strong sense of design intent. The project helped students articulate these needs explicitly, demonstrating the potential of project-based learning to support reflexive engagement with emerging technologies. With regard to these insights, several implications for design educators navigating the integration of GenAI into communication design curricula can be gained:

Reinforcing foundational design skills as a prerequisite to effective AI use. Students emphasised that without strong skills in composition, colour, typography, and semiotics, it was difficult to evaluate or steer GenAI outputs. This highlights the importance of guiding

students to view GenAI not as a source of ready-made designs, but as a complementary tool that both depends on and underscores the importance of traditional design competencies. Such an approach puts the human input and creativity first, instead of the tool itself.

Teaching prompting as a critical, iterative design practice. The students struggled with inaccurate or unwanted outputs which did not correspond with the imagery they had in their minds, which shows that prompting is not a mechanical task but a form of conceptual articulation. Helping students develop more precise, reflective prompting strategies through iterative practice could be a way to develop the ability to link prompts to visual outcomes and encourage iteration.

Embedding critical thinking and ethical awareness into GenAI activities. Stereotypes, factual errors, and culturally flattened representations were encountered by the students throughout the project. Structured reflection on these outputs can help highlight issues of bias, representation, and authorship, positioning GenAI as a tool to interrogate rather than merely to use. This is of course a central point in design education already, but now perhaps an even more important aspect than ever before.

Supporting students in managing abundance and ambiguity. The vast quantities of material generated by GenAI risks overwhelming students or leading to superficial selection. Educators can help students develop filtering criteria, selection strategies, and reflective decision-making processes that foreground intentionality.

Encouraging cross-cultural and contextual understanding. Because GenAI often produces generic or stereotypical imagery, design tasks that require cultural sensitivity can reveal the limits of the technology and the importance of context-driven research. Educators might incorporate collaborative or cross-regional projects to strengthen students' interpretive awareness.

Maintaining analogue and tactile practices alongside AI-augmented workflows. Students' reflections underscore that embodied design experiences—working with materials, tools, and physical processes—continue to shape critical judgement in ways GenAI cannot replicate. Integrating AI into curricula should not come at the expense of these traditional modes of exploration.

Presenting GenAI as part of a broader ecosystem of design tools, not a replacement. Educators can help students cultivate a balanced perspective in which GenAI is understood as one material among many—useful for broad exploration but limited in precision, cultural nuance, and aesthetic intentionality.

7. Conclusion

To summarise, this study both supports and extends existing research on the topic of GenAI in design education. Confirming prior work, its outcomes support the importance of developing strong foundational skills and critical awareness when using GenAI tools. Simultaneously, the study presents new insights into students' experiences of the tensions between their expectations and outcomes of GenAI, especially where issues of factual accuracy and representation are amplified. The role of reflexive practice is a key finding, highlighting students' ability to evaluate, question and reinterpret GenAI outcomes in order to maintain creative agency. Reflexivity can hence be seen not as an auxiliary skill, but as a core design skill needed to navigate current and future technological and cultural conditions.

The scope of this study was deliberately limited, and the students' difficulties can partially be attributed to their relative inexperience with GenAI. With more familiarity, it is likely that many of the frustrations they encountered—particularly around prompting and refining outputs—would diminish. Nevertheless, their reflections offer valuable insight into how students understand GenAI's role in design practice and how they negotiate its benefits and shortcomings.

The findings support earlier arguments that GenAI can assist with ideation and conceptual exploration, particularly when designers treat its outputs as starting points rather than finished proposals. At the same time, the results highlight the need to maintain traditional design skills, echoing concerns that over-reliance on AI could lead to homogenous outcomes, reduced creative agency, or the erosion of aesthetic judgement (Ray, 2023; Xu et al., 2019). More broadly, the project demonstrates that GenAI can serve as an effective catalyst for reflection, pushing students to articulate their conceptual intentions, question algorithmic tendencies, and recognise the value of their own creative and critical capacities.

As with any pedagogical case study, the conclusions drawn here are specific to the context in which the project was conducted. The short duration, the small number of participating students, their novice familiarity with GenAI tools, and the use of a single image-generation system all limit the generalizability of the findings. Moreover, the reflective data collected from students was shaped by their individual levels of confidence and their prior design experience. These limitations do not diminish the value of the insights obtained but instead clarify that they should be viewed as indicative rather than definitive, pointing to areas where further research and long-term pedagogical exploration are needed.

8. References

- Abraham, A., & Windmann, S. (2007). Creative cognition: The diverse operations and the prospect of applying a cognitive neuroscience perspective. *Methods*, 42(1), 38–48. <https://doi.org/10.1016/j.ymeth.2006.12.007>
- Albar Mansoa, P. J. (2024). La inteligencia artificial de generación de imágenes en arte: ¿Cómo impacta en el futuro del alumnado en Bellas Artes? *Encuentros: Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, 20, 145–164. <https://doi.org/10.5281/zenodo.10052355>
- Amaral, C., Edral, A., & Polson, D. (2023). Collaborative online international learning for design education: A framework for studio-based learning. *Journal of Design, Business & Society*, 9(1), 145–158. https://doi.org/10.1386/dbbs_00050_1
- Bainbridge, S. (2023). Photography or ‘promptography’: A year on from the Sony Awards AI furore, what is the nuanced view? *The Art Newspaper*. <https://www.theartnewspaper.com/2024/05/03/photography-or-promptography-a-year-on-from-the-sony-awards-ai-furore-what-is-the-nuanced-view>
- Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1493–1504. <https://doi.org/10.1145/3593013.3594095>
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.
- Burr, V. (1995). *An introduction to social constructionism*. Taylor & Frances/Routledge. <https://doi.org/10.4324/9780203133026>
- CBS Mornings. (2023, March 25). *Full interview: “Godfather of intelligence” talks impact and potential of AI* [Video]. YouTube. <https://www.youtube.com/watch?v=qpoRO378qRY>
- Costa, C. J., Aparicio, J. T., & Aparicio, M. (2024). Socio-economic consequences of generative AI: A review of methodological approaches. *ArXiv*. <https://doi.org/10.48550/arxiv.2411.09313>
- Cropley, D. H., Medeiros, K. E., & Damadzic, A. (2023). The intersection of human and artificial creativity. In P. Mishra & D. Henriksen (Eds.), *Creative Provocations: Speculations on the Future of Creativity, Technology & Learning*, 7, 30–47. Springer International Publishing AG. https://doi.org/10.1007/978-3-031-14549-0_2
- Davis, M. (2018). *Introduction to design futures*. AIGA: The Professional Association for Design. <https://www.aiga.org/sites/default/files/2021-02/introduction-to-design-futures.pdf>
- Doctorow, C. (2025). *Enshittification: Why everything suddenly got worse and what to do about it*. Verso.
- Dziobczenski, P. R. N., & Person, O. (2017). Graphic designer wanted: A document analysis of the described skill set of graphic designers in job advertisements from the United Kingdom. *International Journal of Design*, 11(2), 41–55.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Fleischmann, K. (2024). Generative artificial intelligence in graphic design education: A student perspective. *Canadian Journal of Learning and Technology*, 50(1). <https://doi.org/10.21432/cjlt28618>

- García-López, I. M., & Trujillo-Liñán, L. (2025). Ethical and regulatory challenges of generative AI in education: A systematic review. *Frontiers in Education*, 10, 1565938. <https://doi.org/10.3389/feduc.2025.1565938>
- Gray, M. (2024). Generative AI and its impact on creative education: A thought piece. Council for Higher Education in Art & Design (CHEAD). Available at https://web.archive.org/web/20240601110643/https://www.chead.ac.uk/wp-content/uploads/2024/03/ImpactofAlonCreativeEducation_ProfessorMelanieGray.pdf
- Heigl, R. (2025). Generative artificial intelligence in creative contexts: A systematic review and future research agenda. *Management Review Quarterly*, 76, 955–992. <https://doi.org/10.1007/s11301-025-00494-9>
- Holt, M. (2017). Semiotics and design: Towards an aesthetics of the artificial. *The Design Journal*, 20(sup1), 332–341. <https://doi.org/10.1080/14606925.2017.1352860>
- Hwang, Y., & Wu, Y. (2024). The influence of generative artificial intelligence on creative cognition of design students. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1455015>
- Islam, M. R, Urmi, T. J., Al Msoharrafa, R., Rahman, M. S., & Kadir, M. F. (2023). Role of ChatGPT in health science and research: A correspondence addressing potential application. *Health Science Reports*, 6(10), e1203. <https://doi.org/10.1002/hsr2.1625>
- Ivanov, S., & Lyu, Z. (2024). The economics of generative AI. In *Applications of generative AI* (pp. 491–502). Springer International Publishing. https://doi.org/10.1007/978-3-031-46238-2_25
- Kelly, R. (2018). Design in decline: Breathing new life into an industry through education. *Design Management Journal*, 13(1), 41–52. <https://doi.org/10.1111/dmj.12041>
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.
- Mandapuram, M., Gutlapalli, S., Bodepudi, A., & Reddy, M. (2018). Investigating the prospects of generative artificial intelligence. *Asian Journal of Humanity, Art and Literature*, 5(2), 167–174. <https://doi.org/10.18034/ajhal.v5i2.659>
- Marble, J. (2025). Prioritizing human creativity: A sequential approach to teaching generative AI in graphic design education. In M. Schrepp (Ed.), *Design, user experience, and usability: HCII 2025 (Lecture Notes in Computer Science, Vol 15798)* (pp. 64–78). Springer. https://doi.org/10.1007/978-3-031-93233-5_5
- Matthews, B., Shannon, B., & Roxburgh, M. (2023). Destroy all humans: The dematerialisation of the designer in an age of automation and its impact on graphic design—A literature review. *The International Journal of Art & Design Education*, 42(3), 367–383. <https://doi.org/10.1111/jade.12460>
- Merchán Sánchez-Jara, J. F., González Gutiérrez, S., Cruz Rodríguez, J., & Syroyid, B. (2024). Artificial intelligence-assisted music education: A critical synthesis of challenges and opportunities. *Education Sciences*, 14(11), 1171. <https://doi.org/10.3390/educsci14111171>
- Mumford, M. D., Medeiros, K. E., & Parlow, P. J. (2012). Creative thinking: Processes, strategies and knowledge. *Journal of Creative Behavior*, 46, 30–47. <https://doi.org/10.1002/jocb.003>
- Omran Zailuddin, M. F. N., Nik Harun, N. A., Abdul Rahim, H. A., Kamaruzaman, A. F., Berahim, M. H., Harun, M. H., & Ibrahim, Y. (2024). Redefining creative education: A case study analysis of AI in design courses. *Journal of Research in Innovative Teaching & Learning*, 17(2), 282–296. <https://doi.org/10.1108/JRIT-01-2024-0019>
- Păvăloaia, V.-D., & Necula, S.-C. (2023). Artificial intelligence as a disruptive technology—A systematic literature review. *Electronics*, 12(5), 1102. <https://doi.org/10.3390/electronics12051102>

- Pratschke, B. M. (2024). *Generative AI and education: Digital pedagogies, teaching innovation and learning design* (1st ed.). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-67991-9>
- Putjorn, T., & Putjorn, P. (2023). Augmented imagination: Exploring generative AI from the perspectives of young learners. In *2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE)*. <https://doi.org/10.1109/ICITEE59582.2023.10317680>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- 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>
- Sáez-Velasco, S., Alaguero-Rodríguez, M., Delgado-Benito, V., & Rodríguez-Cano, S. (2024). Analysing the impact of generative AI in arts education: A cross-disciplinary perspective of educators and students in higher education. *Informatics*, 11(2), 37. <https://doi.org/10.3390/informatics11020037>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Seidel, S., Berente, N., Lindberg, A., Lyytinen, K., & Nickerson, J. V. (2018). Autonomous tools and design: A triple-loop approach to human-machine learning. *Communications of the ACM*, 62(1), 50–57. <https://doi.org/10.1145/3210753>
- Sims, A. (2024). From creation to curriculum: Examining the role of generative AI in arts universities. *ArXiv*. <https://doi.org/10.48550/arxiv.2412.16531>
- Skaggs, S. (2017). *FireSigns: A semiotic theory for graphic design*. MIT Press.
- Xu, Z., Sun, C., Long, Y., Liu, B., Wang, B., Wang, M., & Wang, X. (2019). Dynamic working memory for context-aware response generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27, 1419–1431. <https://doi.org/10.1109/TASLP.2019.2915922>
- Yusuf, A., & Noor, N. M. (2023). Research trends on learning computer programming with program animation: A systematic mapping study. *Computer Applications in Engineering Education*, 31(6), 1552–1582. <https://doi.org/10.1002/cae.22659>

Author

Halldór Björn Halldórsson is a graphic designer and Associate Senior Lecturer at Mälardalen University, with a background in advertising and marketing. His research explores the interplay between design practice, tacit knowledge, and emerging technologies. His doctoral work investigates how designers' implicit knowledge can be articulated and communicated through experimental methods and exhibitions. His recent research addresses academic writing in graphic design through scholarly personal narrative, as well as ethical challenges in contemporary design education. With a strong commitment to typography and form, his work advances critical perspectives on how designers think, learn, and act within evolving sociotechnical contexts.



The Concept of a Large Language Model: Misconceptions About and Beneficial Conceptualizations of LLM Technology

Matthew Peterson ^a , Colleen Patton ^b , Michael Green ^c, and Brayan Díaz ^d

^a Department of Graphic Design and Industrial Design, North Carolina State University, Raleigh, NC, USA; ^b Department of Psychology, North Carolina State University, Raleigh, NC, USA; ^c Laboratory for Analytic Sciences, North Carolina State University, Raleigh, NC, USA; ^d School of Teacher Education and Leadership, Utah State University, Logan, Utah, USA

Corresponding author: Matthew Peterson (mopeters[at]ncsu.edu)

Abstract: Large language models (LLMs) are transforming writing and planning practices across the breadth of knowledge work, where misconceptions about the technology abound. We contend that these misconceptions are rooted in the dispersal of disciplinary knowledge—that multiple disciplines are needed to understand LLMs. We bring together disciplinary knowledge to address four constructs commonly applied to LLMs: writing, learning, hallucination, and thinking. Misconceptions about LLM technology are embodied in and transmitted through conceptual metaphors and analogies. We thus provide a basic description of LLM technology with a careful accounting for metaphor. This description and the discussion of the four constructs leads to corrected statements for understanding LLMs. For instance, the assertion that “LLMs think” is replaced with: “LLMs produce outputs that resemble the products of thinking.” Finally, we consider common and potential metaphors for more accurately conceptualizing LLMs, and favor LLMs as answer simulators, remixers, fuzzy thesauruses, and plagiarism traps. The statements and metaphors provided here are a starting point for well-grounded usage patterns with LLMs for professionals, researchers, and students alike.

Implications for practice: The descriptions of writing, learning, hallucination, and thinking constructs (Section 3) combine with the basic description of LLM technology (Section 2) to give practitioners a reflective space for reconsidering LLMs. Practitioners can then refer to the provided statements (Section 3, summarized in Table 1) and metaphors (Section 4) to positively impact internal corporate and institutional communications about LLMs. A more careful and consistent

@: ISSUE > ARTICLE >

Cite this article:

Peterson, M., Patton, C., Green, M., & Díaz, B. (2026). The concept of a large language model: Misconceptions about and beneficial conceptualizations of LLM technology. *Visible Language*, 60(2), 210–239.

First published online April 26, 2026.

© 2026 Visible Language — this article is **open access**, published under the CC BY-NC-ND 4.0 license.

<https://visible-language.org/journal/>

Visible Language Consortium:

University of Leeds (UK)

University of Cincinnati (USA)

North Carolina State University (USA)

use of language can serve to informally educate collaborators, managers, and decision makers on the LLM technology that they are actively employing.

Implications for research: Researcher-provided analogies have been shown to impact how people reason through concepts and models (see Section 1). The statements and metaphors about LLMs provided here (Sections 3.6 and 4) could likewise be tested for their impact on how people plan to use LLMs, what people expect to happen with such use, and how people interpret resultant LLM outputs. It would also be beneficial to systematically study how (and how frequently) constructs such as writing, learning, hallucination, and thinking are invoked in communication about LLMs—by companies, users, and researchers.

Keywords: analogical reasoning; conceptual metaphor theory; large language models; LLM hallucination; metaphor; multidisciplinary; semiotics; writing

1. Introduction

Large language models (LLMs) such as ChatGPT, Claude, and Gemini have pervaded knowledge workers' work and students' studies. Knowledge workers are being pressured into incorporating artificial intelligence (AI) into their workflows, if they are not already enthusiastically doing so of their own volition. Students are being prodded by software to utilize AI (e.g., a Gemini-powered "help me write" prompt that appears on the fourth blink of the cursor in every new paragraph in Google Docs). School teachers and college instructors are receiving LLM-written papers for their assignments, and academic journal editors are fielding LLM-written submissions and LLM-written peer reviews.

Whether prodded or not, people are making *decisions* about when and how to use LLM technology. These decisions are informed by *something*, even when they are not entirely conscious decisions.

We contend that a major factor behind decisions on when and how to use LLMs is the user's fundamental conceptualization of LLM technology, and furthermore that LLM use is in fact too commonly informed by misconceptions of the technology's nature. LLMs have at least superficially passed the Turing test (Gibney, 2025), meaning they can imitate text-based human communication well enough to appear human. The influence of this imitation on users is compounded by the language humans utilize when describing the technology. For instance, according to IBM's knowledge base on LLM technology (Stryker, n.d.), LLMs are "capable of understanding" and "AI agents don't just think."

If LLMs do not in fact understand or think, then these are examples of metaphor or analogy. The literature reveals that they constrain and even constitute our understanding

of reality. Conceptual metaphor theory (Lakoff & Johnson, 1980) posits that much of human thought is guided by conceptual metaphors, which utilize existing concepts like *building* and embodied experiences like a flutter in the chest to constitute more abstract concepts like *theory* and *love*, respectively. Underlying the more basic concepts and organizing embodied experiences are image schemas (Johnson, 1987; or conceptual primitives, Mandler, 1992) like the notion of a *path*. These image schemas are necessary building blocks for more structured conceptual metaphors. Thus, theories can be understood in relation to buildings—“We need to *buttress* the theory with *solid* arguments” (Lakoff & Johnson, 1980, p. 46)—and love is a journey, predicated on a *path* that partners *travel* together (Lakoff & Johnson, 1999, pp. 63–73).

Analogical reasoning similarly utilizes conceptual mapping to understand one concept in terms of another (Gentner, 1983, 2003), and analogy involves similarity of relations more than attributes (Gentner & Markman, 1997, p. 48). For instance, electronic circuitry, and thus electricity as a concept, can be conceptualized according to the structure of crowds of people moving through passageways, or by water flowing through plumbing—with electrons as people or water molecules. Gentner and Gentner (1983) showed how using one of these analogies over the other changed expectations about the nature of electricity, and being incomplete as all analogies must be, that different errors in understanding electricity result from the analogies. Ultimately, whether conceptual metaphor or analogical reasoning, conceptualizing LLMs according to constructs from other domains of understanding (e.g., thinking in humans) necessarily determines and delimits our relationship to the technology—importantly in terms of decisions on when and how to use it.

If it is necessary to comprehend LLM technology through something like conceptual metaphor or analogical reasoning, what expertise can be leveraged to derive reasonably accurate metaphors and analogies? LLM technology’s nature is bound not only to the processes by which its developers created it, but also to nuances of thought, language, and reality. We assume that an ideal conceptualization of LLM technology would rely on synthesizing disparate areas of expertise such as computer science, linguistics, psychology, the learning sciences, semiotics, and philosophy, at the very least.

Imagine an LLM developer who knows the technology well, but only has a folk understanding of thinking, and they ascribe thinking to LLMs. And imagine a cognitive psychologist who has a technical understanding of thinking, but does not appreciate how LLMs work, and they too ascribe thinking to LLMs. Yet these two individuals could come together to share knowledge and collectively determine that LLMs do not think. Perhaps only coordinated multidisciplinary expertise can hope to apprehend the reality of LLMs.

To that end, we utilize multiple areas of academic expertise—embodied in us, the authors—to describe LLM technology and interpret some constructs that are frequently used to conceptualize it. We identify misconceptions according to these areas of expertise and ultimately offer alternative conceptualizations—including conceptual metaphors—that may lead to better decision making with LLMs. The goal is to synthesize knowledge from relevant fields to provide a pragmatically more accurate or helpful conceptualization of LLMs.

2. A Basic Description of LLM Technology

The following description of LLM technology is necessarily written at a high level of abstraction, especially because we are minimizing use of metaphors. Its abstraction is a demonstration of why addressing the nature of LLMs is so frequently, even necessarily, dependent upon metaphor.

Model architecture. The large language model is a type of machine learning model called a *neural network*, with an architecture somewhat analogous to the neurons in a human brain. In the brain, each neuron is connected to multiple other neurons via synapses. Based on the strength of signals provided to it by these connections, each neuron produces its own signal, which is distributed to other neurons in turn. Already inescapable in the basic categorical designation of LLM technology is a brain–algorithm

Introduction to Boxes 1–4. In semiotic structural analysis, meaning is substantially derived from differences among signs (including words), and the paradigmatic axis (expressed in vertical chains below, and in all boxes) is a range of signs that could substitute for a chosen one; even “apparent synonymous signifiers” have significant differences (Chandler, 2021). In effect, at some level a chosen sign implies its alternatives, and the fact that those alternatives were not used impacts interpretation.

sang
died
cried.

boy
man

The

Box 1. A simple demonstration of the paradigmatic axis, as shown in Chandler (2021). The words in cyan are possible substitutions for the words in magenta.

analogy, which uses the neuron to characterize what might be neutrally described as a *node*. The correlational relationships among an LLM's nodes are produced mathematically, through several layers of interconnected mathematical functions that take inputs from previous layers of similar functions, modulate and combine them to some degree, and produce outputs to be fed to subsequent layers. The final layer of this network produces results that can be matched to human-comprehensible tasks, such as a yes/no decision or a choice among possible words in a vocabulary.

Model training. The strength of the signals in the connections among nodes in these mathematical representations is not explicitly programmed. The signal strength of node connections is determined by processing training data and through trial and error. The way an LLM incorporates information into a network of node connections is threefold, as shown in Figure 1 (we avoid characterizing this incorporation as *learning* here, though that term is commonly used).

In *pretraining*, an LLM is first trained on an enormous text corpus that may contain trillions of tokens. Tokens are the words or subword units that represent the compositional building blocks of language in current machine learning applications. These tokens are presented to a model as sequences of varying length and as they occur in the source texts. Using a process called *causal self-attention*, a model attempts to predict the next token in the given sequence. In this process, the model is prevented from seeing future tokens and conditionally determines the most significant earlier tokens in the sequence to focus on in its calculations. These next-token predictions are computed across many positions in parallel, and the predicted tokens are compared with the actual next tokens from the text with the differences contributing to the accumulated error or loss over the batch. Based on the aggregated loss, the numerical parameters or weights associated with each node in the network are adjusted a small amount at a time, proceeding one layer at a time in reverse order from model output back to input, through backpropagation. Training proceeds over many iterations as described, each cycle resulting in incremental adjustments to the model. Eventually the resulting

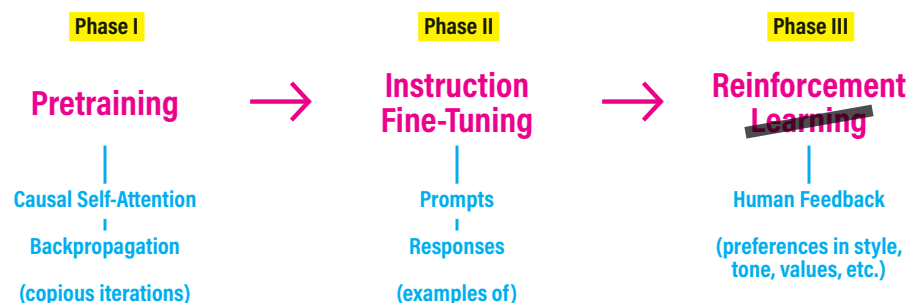


Figure 1. Training phases for large language models.

weights define a function that produces an increasingly accurate expression of the likelihood of each token in a model's vocabulary being the next token in the text. After a stopping criterion is met, the phase is complete and the model represents statistical and structural patterns of text in the training corpus. These patterns potentially generalize beyond the exact examples seen in the training data. However, we emphasize that these patterns may not match with reality or the world outside the training context.

The remaining training phases make the LLM more useful for humans. The second phase is *instruction fine-tuning*, in which examples of questions and answers—or prompts and responses—are provided. These examples are intended to approximate likely user requests. Again, node connection weights are revised through next-token prediction. The third and final phase is *reinforcement* (usually called reinforcement *learning*). This phase is a human feedback process that focuses on human preferences in style, tone, values, and other qualities (e.g., politeness, confidence). The sequence example and token counts in the latter two phases are much lower than in pretraining, but they are high nonetheless.

Output variability. Following training, an LLM outputs tokens into word sequences in a manner that, being based on probabilistic selection among token options and not deterministic pathways, is variable—results are not precisely repeatable. Several strategies have emerged to constrain this variability for situations that were not adequately represented in training. These include prompt engineering, expanded context windows, “reasoning models,” and systems of AI agents, among others. In such extended articulation, more deterministic results can be achieved, pairing the LLM’s pattern matching capabilities with reusable functions, verifiable data sources, or other resources.

3. Problematic Characterizations of LLM Technology

The above description of LLM technology was carefully written to appear dry and to preempt common misconceptions. Due to the terminology regularly applied to LLMs, we still had to invoke the constructs of *neuron* and *learning*. We now address the latter and other constructs that underlie common problematic conceptions of LLM technology, and we consider the entailments that necessarily follow from common but misleading descriptions. If we unpack and reject individual misconceptions, a healthier general concept of LLM technology is possible, one that can lead to better human decision making in using and choosing to use LLMs.

3.1. Writing

When a human user submits a prompt in an LLM chat interface, there is a brief pause before sequences of words are produced from start to finish, resulting in a coherent explanatory narrative. This observable production looks like writing in the folk sense,

with two obvious differences. First and most obvious, the speed at which the narrative is produced is beyond human capability. Second, at least with many LLMs, words and sentences are produced sequentially, with no regressions to revise earlier passages along the way. In contrast, an outside observer to the human writing of the previous sentence would have seen “*make edits along the way*” change to “*revise earlier passages along the way*,” with the latter half retained while the former half is reconsidered. These two LLM–human differences matter because LLM speed promises extraordinary efficiency, and because the impression can be that the technology is superhuman in its capability. But to be superhuman suggests more overlap with “human” than perhaps can reasonably be said to exist, bestowing a kind of authority and even familiarity to LLMs—as personification.

There is a less obvious difference between human and LLM text production. The folk concept of writing collapses writing-to-communicate and writing-to-learn into a single act, if it does not ignore the latter entirely in favor of writing as an unproblematic form of communication. “Let me get my thoughts down on paper” sounds sensible to many, but not to experts on composition. In their cognitive process theory of writing, Flower and Hayes (1981) describe writing as an elaborate, nonlinear process of goal setting and goal satisfaction. The intentional process of writing reveals the writer’s ideas as dynamic, not pre-formed:

Writers create their own goals in two key ways: by generating both high-level goals and supporting sub-goals which embody the writer’s *developing* sense of purpose, and then, at times, by *changing* major goals or even *establishing entirely new ones* based on what has been *learned* in the act of writing. (p. 366, emphasis added)

To the writer, composition “often seems a serendipitous experience,” as in, “I don’t know what I mean until I see what I say” (p. 377). But this serendipity is in fact constituted of *process goals*, or personal instructions for writing of the moment, and *content goals*, or what the writer intends to convey. Especially interesting is that Flower and Hayes note that much of what has been learned about how writers compose text could not have been ascertained by researchers through retrospective measures, because “people rapidly forget many of their own local working goals once those goals have been satisfied” (p. 377). This reality in part suggests why an inaccurate folk sense of writing may have formed—the subjective experience of writing is incomplete or forgotten.

Emig (1977) and Flower and Hayes (1981) establish that writing has an epistemic function. Because the experience of writing is personally beneficial, students and experienced writers face asymmetric consequences in LLM usage. The ability to use LLMs critically requires metacognitive awareness and writing process fluency that is not evenly distributed among writers with varying experience and expertise. Writers

who already have strong process fluency are more capable of integrating LLMs as genuine tools; those who do not are most at risk of substitution patterns. So there is an ethical dimension in the discrepancy between writing in the normal sense and the LLM production of texts. For the new writer, the more an LLM speaks for them, the less they can develop their own agency. They are subject to a form of cognitive surrender that neutralizes intuition and deliberation (Shaw & Nave, 2026).

It is thus misleading to consider the LLM as a co-author. A pair of human co-authors are engaged in their own personal and dynamic goal setting and satisfying behaviors (Flower & Hayes, 1981), which leads to individual learning that refines intentional meaning as embodied in the text. That is, each of the pair contributes the full spectrum of what writing entails. In human-machine teaming of one human author engaging an LLM through prompting, only the human can contribute the full spectrum. But in practice engaging with an LLM can sidestep the learning benefits of writing for the human operator. Thus LLMs are less co-authors than a means to bypass the personal benefits of learning through writing when producing text, which disproportionately impacts inexperienced writers. If the learning process constitutive of full-spectrum writing improves the thoughts that become text, then the speed benefits of LLMs are necessarily paired with quality limitations.

The quality limitations of LLMs extend beyond substitution of a human author's otherwise deep engagement through writing, to the probabilistic nature of next-token prediction. Because the token-to-token weights that determine an LLM's probabilities are derived from commonalities in a massive corpus of training data, its outputs are necessarily derivative in the unflattering sense that it is prone to repetition at the expense of novelty. To inexperienced writers such as students, an LLM's outputs may appear polished and novel, while an experienced writer may find the same writing to be unsophisticated, incoherent, and overly familiar. The student cannot perceive the patterns that their instructor can. And problematically, the lack of novelty—repetition derived from the dataset—can easily slide into unwitting plagiarism. This plagiarism is unwitting to the human operator who does not recognize the sources, and unwitting to the LLM simply because it has no inner life, and thus no wits. Instead of creating unique works—even if it seems this way to users—LLMs generate derivative works. This is a reversal from *creativity* to *replication*.

At a more fundamental level, it is somewhat misleading to make the simple claim that LLMs write texts. There is a scope asymmetry between what happens when a human writes on one hand, and next-token prediction on the other. Though less elegant, it is more accurate to suggest that LLMs enable the production of coherent text in the absence of writing. This is a reversal from *ignoring the limitations of an invoked sense of writing* to *emphasizing the absence of crucial facets of writing*.

Box 2. An absurdist demonstration of the paradigmatic axis (from semiotics; see Mickus, 2024, for its application to LLMs). The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer’s Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., people, LLMs) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.



3.2. Learning

Across popular, technical, and even academic discussions, the language of learning is frequently applied to AI systems without sufficient conceptual clarification or critical examination. As a result, fundamentally different processes are often treated as equivalent. Two common uses of *learning* illustrate this problem.

First, the LLM training process is routinely described through educational metaphors. Authors refer to “teaching” a model or having a model “learn” from examples (Patil & Gudivada, 2024). Such language can be misleading because it conflates machine training with human learning.

Human learning does not occur through passive exposure to vast quantities of information. Rather, educational research has consistently demonstrated that learning emerges through active engagement, experimentation, reflection, social interaction, and meaning-making. As Díaz and Delgado (2024) argue:

Should we educate students as we train ChatGPT, having them read all of Wikipedia plus thousands of books and magazines (Johri, 2023; Ouyang et al., 2022)? Or should we implement pedagogical practices that promote students’ problem solving, critical thinking, and teamwork skills instead of memorization? We have known for decades that the latter approach is preferable. (p. 2576)

The danger of the learning metaphor is not merely semantic. If machine training becomes an implicit model for human learning, educational practice risks being reduced to information transmission, positioning students as passive recipients of content rather than active participants in knowledge construction.

A second and even more widespread source of confusion arises from the term *machine learning* itself. Within computer science, machine learning refers to computational methods that enable systems to identify patterns in data and improve performance on specific tasks without explicit reprogramming (Bell, 2022). While this definition is technically appropriate within the field, describing the process as learning raises an important conceptual question: Is statistical optimization equivalent to learning as understood in educational, psychological, and social theories? Through training, LLMs improve their ability to predict linguistic patterns and generate increasingly sophisticated outputs. This happens in the absence of cognitive characteristics such as understanding, experience, intentionality, and meaning-making.

Educational theories have proposed diverse definitions of learning. Within behaviorism, learning is typically defined as a change in observable behavior resulting from interactions between environmental stimuli and responses (Watson, 2017). Learning is demonstrated through measurable behavioral changes rather than internal cognitive

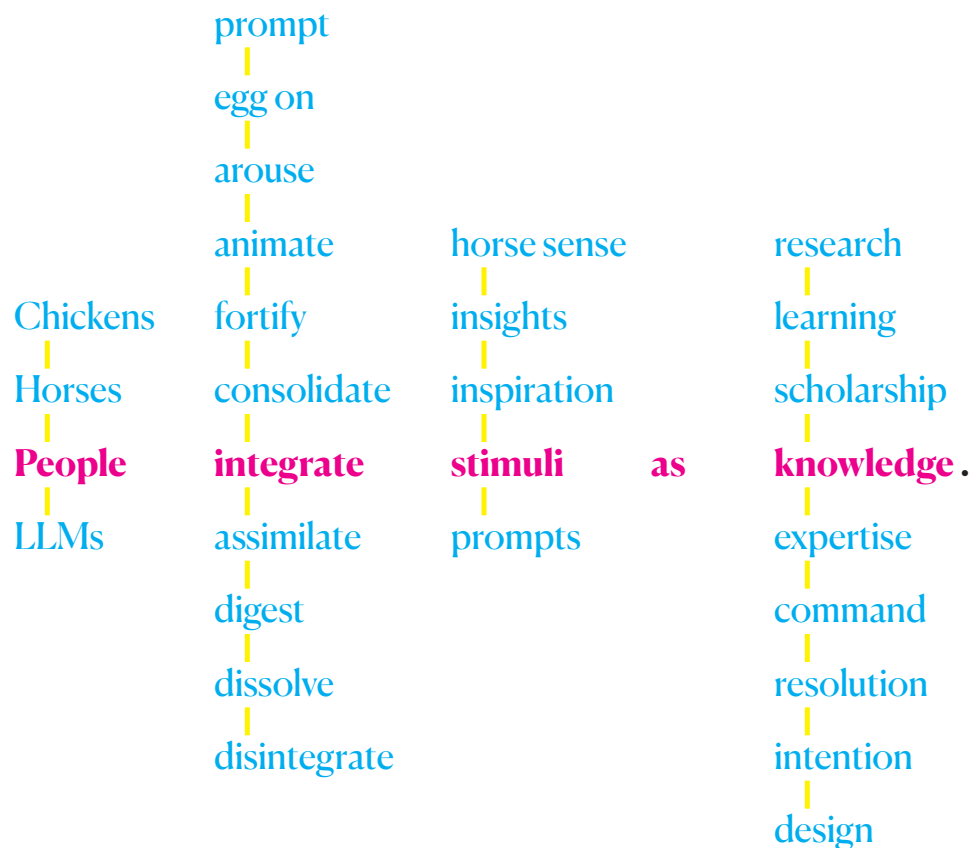
processes. Constructivist perspectives, particularly those associated with Piaget, shifted attention toward the learner's active role in constructing knowledge (Duckworth, 1965). From this perspective, learning occurs through processes of assimilation and *accommodation* (i.e., updating personal schemas when faced with conflicting information), whereby individuals reorganize existing cognitive structures in response to new experiences. Cognitive theories further emphasized internal mental processes involved in acquiring, organizing, storing, and retrieving information (Bruner, 1966). Although traditional cognitivism is often criticized for conceptualizing learners as information processors, it nevertheless assumes active mental engagement rather than information accumulation.

Moving beyond individualistic perspectives, social constructivism emphasized the fundamentally social nature of learning. Vygotsky (1978) argued that higher psychological processes emerge through social interaction and cultural mediation. His concept of the *zone of proximal development* highlighted how learning occurs through guided participation with more knowledgeable others. Through this lens, individuals learn from a scaffolded learning process. The emphasis on participation was further developed through situated learning theory (Lave & Wenger, 1991), which conceptualized learning as inseparable from the contexts in which it occurs. Wenger's (1998) communities of practice framework extended this argument by defining learning as increasing participation within social communities. Importantly, Wenger explicitly distinguished human participation from machine processes: "Participation is an active process, but I will reserve the term for actors who are members of social communities. For instance, I will not say that a computer 'participates' in a community of practice..." (Wenger, 1998, p. 56).

This distinction remains particularly relevant in contemporary discussions of AI. More recent perspectives have further expanded the concept of learning beyond cognition alone. Drawing on Maturana's work on *autopoiesis* (i.e., on self-maintaining systems; Maturana, 1975; Maturana & Varela, 1991), scholars have emphasized the biological, emotional, ecological, and relational dimensions of learning (Bekerman & Zembylas, 2018; Nussbaum et al., 2026). Under this biological theory approach, learning is understood as a property of living, self-organizing systems that continuously adapt through interactions with their environments. Learning is therefore inseparable from biological embodiment and lived experience.

Despite substantial theoretical differences, behaviorist, constructivist, cognitive, sociocultural, situated, and biological perspectives converge on several characteristics of human learning. Learning is understood as experiential, contextually situated, socially mediated, embodied, and grounded in lived participation. By contrast, the training of LLMs consists of statistical optimization across large datasets. Machine

Box 3. An absurdist demonstration of the paradigmatic axis. The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer’s Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., chickens, horses, people, LLMs) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.



learning systems identify patterns, adjust parameters, and improve predictive performance. This distinction does not diminish the remarkable capabilities of LLMs. Rather, it clarifies the ontological differences between computational optimization and human learning. Failing to recognize these differences can lead researchers, educators, and policymakers to overestimate what LLMs are doing and misunderstand how humans learn. Learning cannot be reduced to a process of parameter optimization or information accumulation.

LLMs internalize, in a fractured manner, a massive amount of information. But while LLMs acquire information, humans construct knowledge. We thus emphasize that LLMs are trained on information. (We would say “human-created information,” but LLMs’ datasets will increasingly include outputs from earlier LLMs.) Unlike learning, the term training does not to us suggest much about who or what is being trained. LLMs are trained on information while humans learn and thus construct knowledge. This is a reversal from *implying internal qualities of LLMs when referring to their development* to *emphasizing their artificiality as information stores*.

3.3. Hallucination

LLMs are frequently said to hallucinate when their outputs appear to humans as confident despite failing to reflect reality. Unlike the above construct of learning, the folk sense of hallucination is at least relatively simple, and for many people hallucinations are never experienced. Aleman and Larøi (2008) describe their model of hallucination as presuming:

that a sensory experience (either perception due to *external input* or hallucination due to *internal input*) can be induced by activation of sensory cortical areas (either primary or secondary). In normal perception, signals emanating from the sensory organs play a decisive role, whereas in imagery and hallucinations, internally generated signals (emanating from cortical centers) play a decisive role in determining the final “percept.” (unpaginated ebook version, emphasis added)

LLMs of course have no sensory system, and their “inputs” may be considered to all be internal. As internal inputs are associated with hallucination, it is tempting to simply state that everything LLMs do is a matter of hallucination, though many of these hallucinations end up adequately representing ground truths as humans accept them.

The American Psychological Association’s *Dictionary of Psychology* defines visual hallucination as:

visual perception in the absence of any external stimulus. Visual hallucinations may be unformed (e.g., shapes, colors) or complex (e.g., figures, faces, scenes). In hallucination associated with psychoses (e.g., paranoid schizophrenia, alcohol- or hallucinogen-induced psychotic disorder), *the individual is unaware of the unreality of the perception*, whereas insight is retained in other conditions (e.g., pathological states of the visual system)... (APA, 2018, emphasis added)

If LLMs have no direct connection to reality, then certainly they cannot, in any way, be *aware* of unrealities. They are thus entirely disconnected from any ground truth. Instead, they are connected to what Baudrillard (1981/1994) called *hyperreality*, a self-contained

simulation as a world in itself. Following Wittgenstein's (1921/1922) description of the proposition (in language) as "a model of the reality as we think it is" (p. 39), Baudrillard (1981/1994) stated, long before the advent of LLMs, that:

Today abstraction is no longer that of the map, the double, the mirror, or the concept. Simulation is no longer that of a territory, a referential being, or a substance. It is the generation by models of a real without the origin or reality: a hyperreal. The territory no longer precedes the map, nor does it survive it. (p. 1)

Baudrillard's ideas about media are all the more relevant to LLMs and their outputs. For LLMs, there is no separable normal experience to compare with hallucinatory experience, and so the basic implication of hallucination as a deviation is nonsensical. What are typically called LLM hallucinations are simply probabilistic results from next-token prediction.

Stating that LLMs occasionally hallucinate carries with it the unfortunate suggestion that they regularly perceive reality, which in turn implies that they can (unless something goes wrong) report on reality. It would be far less misleading to suggest instead that LLMs probabilistically produce outputs that are frequently—that often happen to be—reflective of reality. This is a reversal from *emphasizing errors as a performance deviation* to *emphasizing accuracy as a performance probability*.

In a negative sense, LLMs do not have access to reality and thus do not connect human users to reality. They act on internal representations derived from a dataset. In a positive sense, LLMs provide humans in reality access to a fragmented hyperreality constituted of a sizable portion of the internet, though it includes conflicting statements, outright falsehoods, and fictional accounts. (This hyperreality is fragmented because it is constituted of probability distributions rather than regular form.) This is a reversal from *epistemic authority* to *epistemic instability*.

3.4. Thinking

As noted above, writing is a process that forms or crystallizes thoughts. We have experienced this in preparing this section on thinking. The result of our writing process here is tentativeness, which leads to an emphasis on pragmatism (why ascribing thinking to LLMs matters in practice) over ontology (what LLMs are, what thinking is) or epistemology (what we can know about LLMs, how we talk about thinking).

We will begin with a brief sample discussion of aspects of thinking and correspondence with LLM technology. We will then note—and not comprehensively—how thinking entails a massive scope of theoretical constructs in and beyond psychology. This will lead us, in fact, to abandon the exercise and conclude that the richness of the thinking construct not only makes our unpacking of it impractical, but that this richness is

Box 4. An absurdist demonstration of the paradigmatic axis. The words in cyan are possible substitutions for the words in magenta. Boxes 2–4 are human-created, accomplished by following associations through the Oxford American Writer’s Thesaurus. Vertically adjacent words have direct connections in the thesaurus. The first column (i.e., LLMs, people, shepherds, wolves) is an exception; it was created in the spirit of the paradigmatic axis without thesaurus connections.



entirely the problem of ascribing thinking to LLMs: it cannot possibly be done responsibly. This, ultimately, is the crux of our general argument, that characterizations of LLM technology must be carefully formed, and some common characterizations cannot be vetted in the short term for a black-box technology. A conservative approach suggests favoring characterizations that can be given with confidence, that hedge bets.

Thinking is not as clearly delineated a term as is *cognition*, but cognition is at least a good starting point. One aspect of cognition is decision making. As an LLM produces word after word in sequences that cohere into an explanatory narrative, decisions are being made (i.e., next words are being “predicted”). For LLMs, this is a matter of probabilities derived from the training process. Is decision making so different in humans?

Much of the way in which people make decisions is through pattern matching. Recognition-primed decision making (Klein et al., 1989) suggests that humans decide on an action by matching cues from the current situation to past situations. Experts can be defined as such, in part, because they have such a breadth of prior patterns to match that they often make good decisions. LLMs exhibit exemplary pattern matching from and within the bounds of a huge set of patterns, frequently resulting in good decisions on what text to output. This idea can be transferred from decision making broadly into writing more specifically—often, good writers have enough experience to recognize where grammatical formatting belongs, or ways to structure an argument. LLMs do not tire of writing or typing, they can hold and compare amounts of information that we cannot fathom, and they can produce outputs—that is, continue to make decisions on what words to produce—without ever hitting writer’s block.

Human pattern matching in decision making is influenced by biases such as the saliency of the information (e.g., the emotion attached to it) or recency (e.g., how recently the information was encountered). These biases can make people hold onto already written work, recall certain pieces of information that may not be closely matched to the current idea, or have emotions evoked from a piece of writing—none of which are comparable to an LLM’s pattern matching. Humans acquire their patterns by being in the world, connected to it through sensory systems. They encode information in rich mental imagery and in linguistic propositions (Kosslyn et al., 2006), and process it in dedicated resources in working memory (the visuospatial sketchpad and phonological loop, respectively; Baddeley, 2018). Affect impacts much processing, with immediate feelings and personal values influencing interpretation and response. Decisions are understood through a conceptual system that uses conceptual metaphors (Lakoff & Johnson, 1980) to make the world meaningful, and these are based on formative and embodied experiences that provide basic units of meaning (Johnson, 1987; Shapiro, 2019; Wilson, 2002). Decisions are necessarily influenced by past experiences of living in a social world.

We quickly reach a point in considering decision making, itself only part of cognition, and that only part of what thinking entails, where we accumulate so many considerations that we must simply state the following:

Thinking is a rich capability of humans that is incredibly complex in its mechanisms and its source material. Most of what constitutes thinking in human beings has no parallel in LLMs, but creating a list enumerating the differences is impractical. Instead we are left with a simpler question: Do discrepancies between what humans do and what LLMs do suggest that LLMs cannot think, or that our definition of thinking is too narrow? Chen et al. (2026) address this question, or something close to it. They suggest that LLMs have general intelligence, and argue that differences between humans

and LLMs should not disqualify LLMs. It is in fact anthropocentric bias that leads to disqualification of LLMs, and “by reasonable standards, including Turing’s own, we have artificial systems that are generally intelligent” (p. 36).

We the authors are overwhelmed. Instead of making a scientific statement about the nature of thought, in which we could have little confidence, we are left with a more designerly response. *Pragmatically*, whether we ultimately describe what LLMs do as thinking or not, we presently believe that invoking *thinking* for LLMs is problematic.

In their structure mapping of similarity and analogy, Gentner and Markman (1997) use simple node diagrams to represent correspondence between two conceptual domains. Figure 2 is an example of pure correspondence. The nodes could represent any number of things. For instance, in an analogy between an atom (Domain A) and the solar system (Domain B), some nodes might be: (1) a large central element, nucleus or sun; (2) a small secondary element, electron or planet; (3) the large element restricting the small element’s movements; and (4) the large element being relatively fixed within the system (at which point at least the atom half of the analogy is becoming somewhat misleading).

Unlike the atom and solar system example, what LLMs do as a process is much simpler than the embodied, experiential, and social processes that underlie human thought. As shown in Figure 3, this can be visualized as a profound discrepancy in complexity of connections. The blurred magenta nodes on the LLM side of the diagram represent aspects of human thought that are not present in what we might call an LLM’s “thinking,” but that might result in “candidate inferences” (Gentner & Markman, 1997), which use of the term *thinking* impels people to make when they consider the nature of LLMs. That is, the invocation of *thinking* when discussing LLMs is likely to result in people making inferences from their assumptions about human thinking, itself enriched by their own experiences of their own thoughts, as well as interactions with others. The

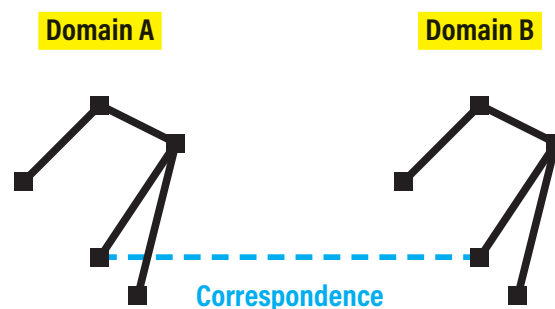


Figure 2. An abstraction of the analogy between two conceptual domains, after Gentner and Markman (1997). In this case, the two domains are in fact analogical. The dashed line indicates correspondence between nodes in the two domains. Only one line of correspondence is labeled, but each of the five nodes in Domain A corresponds to a node in Domain B, as indicated by matching network shapes.

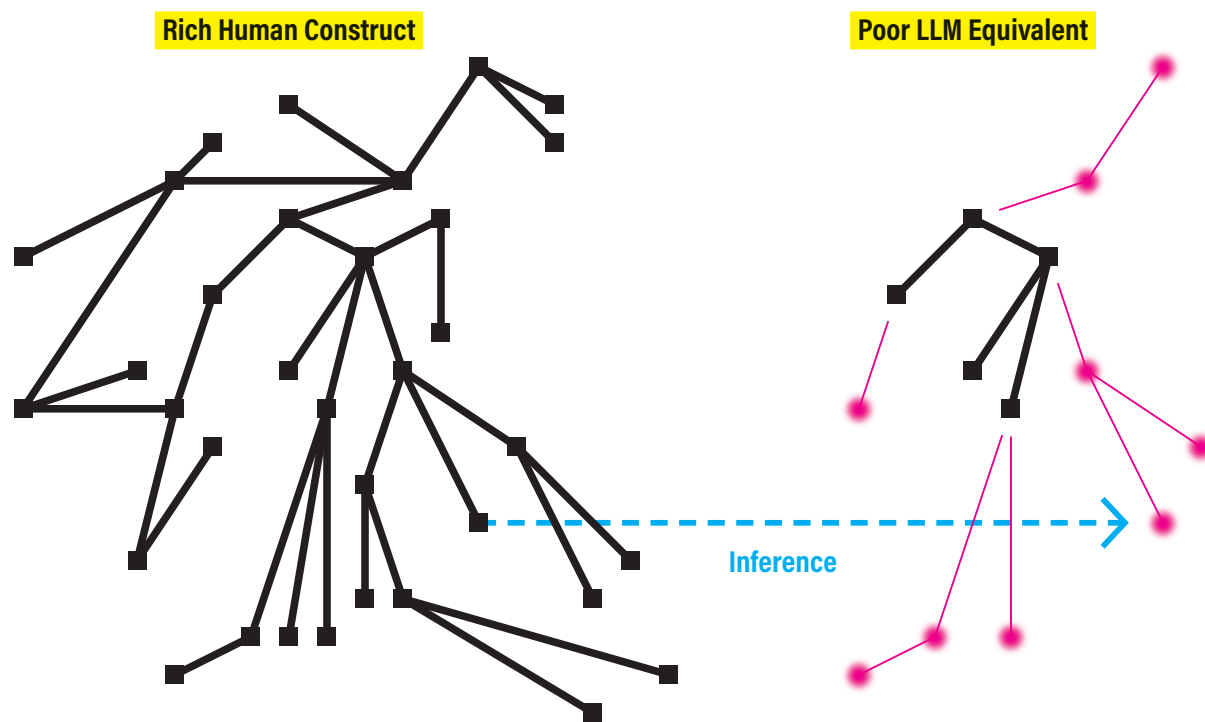


Figure 3. Complexity asymmetry between two domains in an imperfect analogy: a rich human construct such as *thinking* is paired with a comparably poor LLM equivalent. Inappropriate inferences may be made about the LLM as a target due to the richness of the thinking construct as a source, as indicated with the dashed cyan line. Inferred but inaccurate aspects of the LLM domain are indicated with blurred magenta circles. Properly corresponding aspects of the LLM domain are indicated with solid black squares.

discrepancies in LLM and human versions of “thinking” are too great for us to know all of the implications, and it is prudent to assume that many of the countless implications matter, perhaps to a great degree.

It is fair to state that there is an anthropocentric bias to selectively claiming that human-only aspects of thought are defining factors of thought, as is argued by Chen et al. (2026). However, saying that we do not have a complete enough sense of thinking to conclude that LLMs *do not* think is simply not a good argument for saying that they *do* think. In sharp contrast to Chen et al. (2026), we advocate for agnosticism on the matter, at the least. (While that is our prudent conclusion, we are individually and on a more personal level fairly comfortable stating that LLMs do not think, and do not come remotely close to doing so.)

Thus, instead of formally countering a statement that LLMs think with a definitive statement that they do not, we emphasize that the impression of thinking is based on similarity of outputs, and not on internal processes. Wittgenstein (1921/1922), in

considering the proposition (in language) to be a model of reality (p. 39), noted the discrepancy between output (language) and processing (thought):

Language disguises the thought; so that from the external form of the clothes one cannot infer the form of the thought they clothe, because the external form of the clothes is constructed with quite another object than to let the form of the body be recognized. (p. 39)

Chen et al. (2026) suggest in their argument for LLMs' possession of general intelligence that an LLM "will correctly predict shattering in one case and not the other" for "dropping a glass or a pillow on a tile floor" (p. 39), but this is misleading. A human making this prediction will imagine a glass falling and shattering, at least partially pictorially—we think in pictures—and thus will predict based on an activated *model of reality*. They will, indeed, predict the result of a glass or a pillow falling. The LLM does not predict what the glass or the pillow will do. It predicts the words that a human is likely to use when the human is predicting the actual outcome. The outputs may be equivalent, but the underlying process is not. Bender and Koller (2020) offer insight here, warning that:

we must be extra careful in devising evaluations for machine understanding, as Searle (1980) elaborates with his Chinese Room experiment: he develops the metaphor of a "system" in which a person who does not speak Chinese answers Chinese questions by consulting a library of Chinese books according to predefined rules. From the outside, the system seems like it "understands" Chinese, although in reality no actual understanding happens anywhere inside the system. (p. 5188)

(Searle was anticipating, in 1980, "strong AI" that "really is a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states" [p. 417]. Amusingly, and reflecting the present and heated debates on the nature of LLMs, Searle's article is immediately followed in its publishing journal's issue by an open peer commentary entitled "Searle's argument is just a set of Chinese symbols" [Abelson, 1980].)

Imagine asking a person a question, which they answer. And then, as a follow-up, questioning them again, and again receiving an answer. We would not consider this person's thoughts to exist only in and as they articulate their responses. The first response, rather, would be considered an immediate utterance that is, to some degree, the result of a lifetime of thought that preceded it. And in the space between their two answers would be more relevant thought, influenced by having to consider the first answer, how the questioner responded through affect and gesture, and the second question. But for the LLM, anything remotely akin to thought occurs in the *moments of*

response alone, as responses, not thoughts on how to respond, and not in the context of any internal life.

LLMs produce outputs that resemble the products of thinking—this need not imply that they think. This is a reversal from *a definitive assessment of thinking* (one way or another) to *an acknowledgment that there are likely important production discrepancies despite similarities in resultant products*.

3.5. Learning, Thinking, Hallucination, and Writing in Synthesis

The learning, thinking, and hallucination constructs, when considered together in sequence, offer another conceptual reversal. In an LLM's initial training phases—during development—holistic source material is not “learned” and internalized as understood by humans. It rather disassembles mass amounts of information from original gestalts into weights in its probability distribution. The LLM does not retain the integrated wholes from which it derived isolated parts. This is fundamental to how LLMs work, even though additional technologies can mitigate this to varying degrees.

When the LLM is subsequently prompted by a human—during use—it works from these isolated “parts” to consolidate some of them in its node network into something that the human operator will perceive as a whole. But this is a new whole, not a recovered whole. If this new whole corresponds with something that the LLM was originally trained on, we would call it a reconstruction. But the “re-” prefix is a human reflection, an assessment; the LLM was nevertheless merely constructing something from raw material.

(This may appear reminiscent of the fact that when humans access long term memory, there is a creative process that can alter the memories stored there. Elizabeth Loftus's research on “repressed” memories, eyewitness testimony, and leading questions demonstrated this [Mook, 2004], but *unstable* gestalts are far removed from the *absence of gestalts*, so we do not consider this evidence of human-machine thinking correspondence.)

The same human assessment of an LLM's outputs that may deem one a construction instead of a reconstruction could result in the label of *hallucination*. But as previously discussed, hallucination implies a connection to ground truth through a violation of that ground truth. In viewing LLMs as disassembling wholes into parts and subsequently assembling something else with a selection of those parts, hallucination again rings false.

Returning to writing as a practice, consider a common occurrence: an LLM is asked to substitute for a human user's literature search, and among its outputs are multiple formal references to articles and books. We would typically say that if some of these references are not real—i.e., the publications they list in fact do not exist—then *these*

references in particular are hallucinations. But all of the references were produced in the same manner. The LLM was originally trained on a massive number of references, all of which were decomposed into weights in a probability distribution, and subsequently during use, it fabricated references. *Fabricated* here is a synonym for construction. So though we might be inclined to ultimately assess *particular* references as “fabricated references,” in fact even the references we accept have been fabricated. As an LLM assembles a reference token by token, it is following pathways through its probability distribution, with the potential for it to make all of what we would call “right turns,” and output an authentic reference; but also with the potential to veer away.

What we typically call hallucination is inseparable from the very nature of LLMs. Any innovations that completely eliminate the problem of hallucination will result in something, by definition, other than an LLM.

3.6. Implications of Alternative Statements

Statements. The preceding description of four constructs affiliated with LLM technology led to misleading statements about LLMs. We suggested alternative statements for these misconceptions, which are collected in Table 1. Our premise is that these alternative statements support better human decision making with LLMs. Though this premise could hold even if it were impossible to observe directly, we next consider some overt ways in which these statements might express themselves in decision making.

Decisions. Consider a college student in an introductory course in their field of study who believes that *LLMs write texts*. When given a writing assignment, this student is likely to use an LLM to write their paper. They may have heard that LLMs are useful tools, and thus they “use” this tool to “write their paper.” But if they instead believe that *LLMs produce coherent text without writing*, they may instead see using the LLM as sidestepping the writing task (“without” writing), and they may be slightly more likely to conclude, even unconsciously, that using an LLM to produce their paper is improper. Their instructor is likely to agree.

Now consider an experienced writer who is an expert in their discipline, whose potential use of AI is not governed by a syllabus policy that makes it “improper.” If this expert believes that *LLMs write texts*, they may be comfortable offloading parts of their writing process to an LLM without questioning its impact on their writing process overall. But if they believe that *LLMs produce coherent text without writing*, they may use their sophisticated understanding of the writing process—informed by their own metacognition over thousands of hours of writing—to selectively compartmentalize aspects of their writing process, engaging in a preservation of their writing’s integrity.

Next consider the college student and the disciplinary expert together. If instead of viewing LLMs as creative in the normal sense, they think that they *generate text patterns*

Table 1. Misleading statements about LLMs compared with corrected versions.

Misleading statement	Corrected statement
LLMs think.	LLMs produce outputs that resemble products of thinking.
	LLMs recombine semantic parts that are separate from their original wholes.
LLMs learn from accumulated human knowledge.	LLMs are trained on patterns in a collective information store.
	LLMs disassemble meaningful gestalts during training.
LLMs occasionally hallucinate.	LLM outputs frequently reflect reality.
	LLMs fabricate texts.
LLMs model reality.	LLMs provide humans in the real world access to a fragmented hyperreality.
LLMs write texts.	LLMs produce coherent text without writing.
	LLMs assemble gestalts in which people find meaning.
LLMs are creative.	LLMs generate text patterns through partial replication of other text patterns.

through partial replication of other text patterns, they may be more careful in employing an LLM in writing. The student may fear cheating while the expert fears plagiarism. Replication of existing language suggests a high level of responsibility for interrogating outputs and investigating provenance.

Finally, consider the disciplinary expert when they have agreed to peer review a submission for an academic journal. If they believe that *LLMs think*, then they may consider it adequate to prompt an LLM with their intentions for review, and then have the LLM “review” the manuscript. But if they believe that *LLMs produce outputs that resemble the products of thinking* and thus that LLMs cannot truly reason or evaluate, they may conclude that the LLM cannot substitute for their expert evaluation. The manuscript’s authors and the journal editor are likely to agree.

4. Alternative Characterizations of LLM Technology

Our assumption is that because LLM technology is unfamiliar and mysterious as a black-box technology, that metaphors consistently used to characterize it will function as conceptual metaphors (Lakoff & Johnson, 1980), constraining thoughts about LLMs and thereby influencing usage patterns with them.

In a remarkable paper that raised alarms about LLM technology, Emily Bender, Timnit Gebru, and colleagues (2021) called the language model a *stochastic parrot*. *Stochastic* in this case refers to the underlying mechanism of probabilistic weights used in next-token prediction, and it is generally defined as “randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely” (New Oxford American Dictionary, June 2, 2026). The *parrot* half of the metaphor emphasizes the absence of meaning, in that the LLM has only its patterns of relationships among tokens. Considering the relationship of language model (LM) to human user in text-based communication, Bender et al. (2021) state that:

The problem is, if one side of the communication does not have meaning, then the comprehension of the implicit meaning is an illusion arising from our singular human understanding of language (independent of the model). [...] Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning. (pp. 616–617)

This illusory meaning is predicated on an understanding of language that distinguishes between form and meaning, as described by Bender and Koller (2020):

We take *form* to be any observable realization of language: marks on a page, pixels or bytes in a digital representation of text, or movements of the articulators. [...] We take *meaning* to be the relation between the form and something external to language... (pp. 5186–5187)

The “something external” that the LLM’s language is disconnected from is anything in reality itself. Bender and Koller also emphasize intentionality, an important facet of the stochastic parrot, which “parrots” language that it does not understand: “When humans use language, we do so for a purpose: We do not talk for the joy of moving our articulators, but in order to achieve some *communicative intent*” (p. 5187).

The stochastic parrot metaphor is popular (Bender et al., 2025)—as of June 2, 2026, only roughly five years after publication, Bender et al. (2021) had accumulated over 14 thousand citations on Google Scholar. However, it may over-correct in its move away from unjustified positivity in favor of LLMs—a positivity to which companies that make money on LLMs are prone. The stochastic parrot argument largely holds up in its assertion that models represent patterns determined by statistical prevalence in training data and not deeper conceptual understanding or embodied experience. But the term *parrot* minimizes the current capabilities of LLMs. Models generalize and display emergent behaviors not explicit in the training data, creating connections between seemingly dissimilar patterns. They effectively translate languages, write

code, solve problems, use tools, and follow complex instructions where they are not just repeating strings of previously seen text. While in the act of predicting next tokens, LLMs form abstract internal representations of possible future continuations, latent goals, and structural constraints that suggest that there is something more going on than simple mimicry (Lindsey et al., 2025). Our search for metaphors continues.

Fanciful metaphors about LLM technology abound. Ethan Mollick offers two noteworthy metaphors. In lamenting LLM-generated posts online, Mollick (2026) describes them as “just *meaning-shaped attention vampires* that take mental effort to decode and give you no equivalent understanding in return” (emphasis added). He describes the LLM-based chatbot (ChatGPT specifically) as “an infinitely sort of *patient intern that sometimes lies to you*” (Mollick & Nickisch, 2023, emphasis added). We can invent our own fanciful metaphors that eschew AI hype, for instance, by emphasizing the LLM’s limitation of its connection to a dataset (i.e., internal representations alone) as a hyperreality:

- ▶ LLMs are *hyperreality ambassadors*, giving us partial access not to reality, but to the hyperreality of the internet, with its biases, contradictions, and falsehoods. While this metaphor draws a hard line between the LLM’s disconnected and fractured dataset and the world, ambassadors are people with intentionality, emotion, interior lives, and personal histories. There are too many problematic entailments in such personification, and so we reject this metaphor.
- ▶ LLMs are *gaslighters*, emphasizing that the human user interacting from a position in reality with an LLM that is in hyperreality can confuse what is reality. Though we offer this up as an example, we immediately reject it on the grounds that gaslighting is done systematically and intentionally, and this is too extreme a characterization, and too negative.
- ▶ LLMs are *hallucinators*. We reject this metaphor also, because it implies a connection to reality that is bypassed. LLMs have no such connection to bypass—again, they frequently reflect reality rather than occasionally hallucinate.

Other available metaphors emphasize some aspect of LLM technology without drawing human parallels, and we believe work better by characterizing LLMs as artificial:

- ▶ LLMs are *meaning emulators*. This metaphor is terminologically reminiscent of *stochastic parrot* but offers another verb (*emulation*) to describe the work of LLMs.
- ▶ LLMs are *answer simulators* that appear to answer questions in the thoughtful way that people do, but in fact are predicting (token by token) what a person would answer. This is not the same as answering.
- ▶ LLMs are *averaging machines* or *algorithmic homogenizers*, because their weighted probabilities will have a central tendency within the distribution of their training data. This is in one sense a strength of LLMs: they give us access to something approaching a consensus. But the metaphor is useful when an inexperienced

writer is trying to use an LLM creatively. The result is not likely to be special due to this central tendency—one person’s prompted result is likely to be overly similar to another person’s.

- ▶ LLMs are *remixers*, collecting knowledge contributions from humans and combining them—sometimes plagiarizing them.
- ▶ LLMs are *plagiarism traps* that regularly obscure sourcing with which their human operators are unfamiliar. This means that the central tendency that leads to homogenized results also leads to inappropriately copied results.

The above characterizations may appear to be rather negative. They do in fact emphasize limitations. (Far more negative machine metaphors are available, such as *bullshit generator*, a pre-existing term that is frequently applied to LLMs.) The following metaphor emphasizes how LLMs can be used as tools when only limited function is required of them:

- ▶ LLMs are *fuzzy thesauruses*. A conventional thesaurus enables writers to find synonyms and antonyms for specific words. But the chatbot interface for an LLM permits writers to explore a looser terminological range of possibility even when they cannot call to mind a particular word as a starting point, which a conventional thesaurus requires. With a chatbot, a writer can circle around a concept instead of needing to first name the concept. The term *fuzzy*, as in fuzzy set theory, suggests a gradient of semantic inclusion for similarity rather than a binary of either is-similar or is-not-similar. So the LLM can reach further out than a standard set of synonyms.

We used ChatGPT as a fuzzy thesaurus to generate alternatives to *fuzzy thesaurus*, beginning with prompts that loaded the “nearest neighbor” and “paradigmatic axis” concepts, and it suggested *probability harvester*, *paradigmatic displacement apparatus*, and *semantic neighborhood watch*, among other metaphors (ChatGPT, June 8, 2026). (A *nearest neighbor* is a point in a set that is most similar to another point, and the *paradigmatic axis* in semiotics contends that a word’s usage is meaningful in part as a substitution, or in contrast to all the alternatives that could have been but were not used [which we explore in Boxes 1–4].) Here conceptual material that is meaningful to a human prompter *before an interaction with a chatbot* (e.g., nearest neighbor, paradigmatic axis) ensures that suggestions are imbued with meaning (i.e., the LLM is not a standalone stochastic parrot). But we consider these ChatGPT metaphors to be novelties rather than good options.

We find four of the above metaphors for LLMs to be the most balanced and reflective of their nature. The first three are the LLM as an answer simulator, as a remixer, and as a fuzzy thesaurus. Together they characterize the LLM as a non-thinking tool (an answer simulator; neither respondent nor consultant), which presents information in

a changed form (a remixer), and functions through flexible word associations (a fuzzy thesaurus). An additional metaphor, the LLM as a plagiarism trap, may appear negative, but it is actually balanced because people can avoid traps when they know how the traps work and where they can be found. The metaphor presents a warning while implying that humans can engage safely, if carefully. These more balanced metaphors leave room for principled use by human operators while avoiding major misconceptions about LLM technology.

5. Conclusion

Large language models do similar things to what people have traditionally done. This challenges our assumptions or blind spots about those things. For instance, writing is a complex endeavor, but our shorthand notion of writing is simply producing coherent language in written form. Once an LLM can produce coherent language in written form, we arrive at some contradictions. An instructor would not consider it acceptable for a student to “write” a paper with a single prompt and then take credit as the author. And no serious academic journal would accept a paper with ChatGPT as an author with a human prompt credited in the acknowledgments. So what is writing, actually? The point is that this situation begs for more explicit descriptions of constructs like writing, learning, hallucination, and thinking, and all that we can compare these constructs with is humans—with our own versions of them. With a given construct, what is the deviation from the human version, and when is the construct too far removed to even be invoked? LLMs may be intelligent in some way, but that intelligence is probably quite a bit different from human intelligence. This results in difficulties around assumptions or expectations of truth, and whether (or when) we can trust the output of AI models and systems.

We provided a “basic” description of LLM technology (i.e., nontechnical and with minimal use of metaphor), and compared it to four constructs as understood separately from LLMs: writing, learning, hallucination, and thinking. In doing so, we highlighted numerous alternative statements about the technology:

1. LLMs produce outputs that resemble products of thinking.
2. LLMs recombine semantic parts that are separate from their original wholes.
3. LLMs are trained on patterns in a collective information store.
4. LLMs disassemble meaningful gestalts during training.
5. LLM outputs frequently reflect reality.
6. LLMs fabricate texts.
7. LLMs provide humans in the real world access to a fragmented hyperreality.
8. LLMs produce coherent text without writing.

9. LLMs assemble gestalts in which people find meaning.
10. LLMs generate text patterns through partial replication of other text patterns.

An additional statement includes a warning to inexperienced writers—no, all writers—about LLM use:

11. LLMs can be used to bypass the personal benefits of learning through writing when producing text.

We then offered a selection of conceptual metaphors, and rejected many as overly negative or misleading. We ultimately favor characterizing LLMs as:

- ▶ Answer simulators
- ▶ Remixers
- ▶ Fuzzy thesauruses
- ▶ Plagiarism traps

Language constrains thought, and thought leads to action. The use of LLMs in writing and in other endeavors should be done thoughtfully and with reasonable conceptions of LLM technology. It is undeniable that the large language model is an amazing innovation, especially difficult for humans to understand. Accurately identifying a technology's inherent limitations does not deny that technology's importance. LLMs are not what they are not, but they certainly are what they are.

6. Acknowledgments

Kirsti Cole—a professor of English and co-director of the Campus Writing and Speaking Program at NC State University—provided extensive notes on an early version of the manuscript. Many of her comments were processed and are embodied in the final version, but this does not imply that she agrees with all assertions presented here.

The third author is a United States governmental employee, and thus we provide the following acknowledgment. This material is based upon work done, in whole or in part, in coordination with the Department of Defense (DoD). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the DoD and/or any agency or entity of the United States Government.

7. References

- Abelson, R. P. (1980). Searle's argument is just a set of Chinese symbols [Open peer commentary]. *Behavioral and Brain Sciences*, 3(3), 424–425. <https://doi.org/10.1017/S0140525X00005756>

- Aleman, A., & Larøi, F. (2008). Hallucinations: The science of idiosyncratic perception. American Psychological Association. <https://doi.org/10.1037/11751-000>
- APA (2018). Visual hallucination. In *APA dictionary of psychology*. <https://dictionary.apa.org/visual-hallucination>
- Baddeley, A. D. (2018). Working memory: Theories, models, and controversies. In *Exploring working memory: Selected works of Alan Baddeley* (pp. 332–369). Routledge.
- Baudrillard, J. (1981/1994). *Simulacra and simulation* (Trans. S. F. Glaser). The University of Michigan Press.
- Bekerman, Z., & Zembylas, M. (2018). *Psychologized language in education: Denaturalizing a regime of truth*. Palgrave Macmillan.
- Bell, J. (2022). What is machine learning?. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 207–216). John Wiley & Sons Ltd.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. <https://aclanthology.org/2020.acl-main.463.pdf>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada. ACM. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., Costello, E., Lee, K., Farrow, R., & Ferreira, G. (2025). Unsafe AI for education: A conversation on stochastic parrots and other learning metaphors ⚠️. *Journal of Interactive Media in Education*, 2025(1), article no. 10. <https://doi.org/10.5334/jime.1079>
- Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.
- Chandler, D. (2021, November 23). Paradigms and syntagms [Webpage]. *Semiotics for beginners*. <http://visual-memory.co.uk/daniel/Documents/S4B/sem03.html>
- Chen, E. K., Belkin, M., Bergen, L., & Danks, D. (2026, February 2). Does AI already have human-level intelligence? The evidence is clear. *Nature*, 650, 36–40. <https://www.nature.com/articles/d41586-026-00285-6>
- Díaz, B., & Delgado, C. (2024). Artificial intelligence: Tool or teammate?. *Journal of Research in Science Teaching*, 61(10), 2575–2584. <https://doi.org/10.1002/tea.21993>
- Duckworth, E. (1964). Piaget rediscovered. *The Arithmetic Teacher*, 11(7), 496–499. <https://www.jstor.org/stable/41186862>
- Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, 28(2), 122–128. <https://doi.org/10.2307/356095>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://www.jstor.org/stable/356600>
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.
- Gentner, D. (2003). Analogical reasoning, psychology of. In L. Nadel, R. French, P. Culicover, D. Chalmers, R. Goldstone, M. Gasser, D. Noelle, K. McCabe, J. Levin, & J. Odling-Smee (Eds.), *Encyclopedia of cognitive science* (pp. 106–112). <https://doi.org/10.1002/0470018860.s00473>
- Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 99–129). Lawrence Erlbaum Associates, Publishers.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52(1), 45–56. <https://doi.org/10.1037/0003-066X.52.1.45>

- Gibney, E. (2025, October 20). AI language models killed the Turing test: Do we even need a replacement? *Nature: News*. <https://doi.org/10.1038/d41586-025-03386-w>
- Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 462–472. <https://doi.org/10.1109/21.31053>
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press, 1987.
- Johri, S. (2023). The making of ChatGPT: From data to dialogue [Blog post]. *Science in the news*. <https://sites.harvard.edu/sitn/2023/06/06/the-making-of-chatgpt-from-data-to-dialogue/>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Lindsey, J., Gurnee, W., Ameisen, E., Chen, B., Pearce, A., Turner, N. L., Citro, C. Abrahams, D., Carter, S., Hosmer, B., Marcus, J., Sklar, M., Templeton, A., Bricken, T., McDougall, C., Cunningham, H., Henighan, T., Jermyn, D., Jones, A., . . . Batson, J. (2025). On the biology of a large language model. *Transformer Circuits*. <https://transformer-circuits.pub/2025/attribution-graphs/biology.html>
- Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, 99(4), 587–604. <https://doi.org/10.1037/0033-295x.99.4.587>
- Maturana, H. R. (1975). The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies*, 7(3), 313–332. [https://doi.org/10.1016/S0020-7373\(75\)80015-0](https://doi.org/10.1016/S0020-7373(75)80015-0)
- Maturana, H. R., & Varela, F. J. (1991). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.
- Mickus, T. (2024). Language models and the paradigmatic axis. In *Proceedings of the Society for Computation in Linguistics 2024* (pp. 63–74). Association for Computational Linguistics. <https://aclanthology.org/2024.scil-1.7/>
- Mollick, E. (2026, May 26). Choosing to stay human... Means choosing when and how to use AI [Blog post]. *One Useful Thing*. <https://www.oneusefulthing.org/p/choosing-to-stay-human>
- Mollick, E., & Nickisch, C. (2023, March 14). Why you (and your company) need to experiment with ChatGPT now [Podcast episode]. *HBR IdeaCast*, episode 901. <https://hbr.org/podcast/2023/03/why-you-and-your-company-need-to-experiment-with-chatgpt-now>
- Mook, D. (2004). Elizabeth Loftus: Leading questions and false memories. In D. Mook (Ed.), *Classic experiments in psychology* (pp. 199–202). Greenwood Press.
- Nussbaum, M., Bekerman, Z., Archambault, L., Ranieri, M., Yilmaz, R., van Braak, J., Zou, D., & Tsai, C. C. (2026). *Teachers' doings in the AI era: From transmission to relation*. *Computers & Education*, 249, 1–8. <https://doi.org/10.1016/j.compedu.2026.105617>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback [Preprint]. *ArXiv*. <https://doi.org/10.48550/arXiv.2203.02155>
- Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 14(5), 2074. <https://doi.org/10.3390/app14052074>

- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424. <https://doi.org/10.1017/S0140525X00005756>
- Shapiro, L. (2019). *Embodied cognition* (2nd ed.). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781315180380>
- Shaw, S. D., & Nave, G. (2026). Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender [Preprint]. https://doi.org/10.31234/osf.io/yk25n_v1
- Stryker, M. (n.d.). What are LLMs? [Webpage]. IBM: Think. Accessed March 18, 2026. <https://www.ibm.com/think/topics/large-language-models>
- Vygotsky, L. (1978). *Mind in society*. Harvard University Press.
- Watson, J. B. (2017). *Behaviorism*. Routledge.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>
- Wittgenstein, L. (1921/1922). *Tractatus logico-philosophicus* (Trans. C. K. Ogden). Kegan Paul, Trench, Trubner & Co., Ltd. <https://www.gutenberg.org/files/5740/5740-pdf.pdf>

Authors

Matthew Peterson is an associate professor of graphic and experience design at NC State University. His research focuses on visual representation in user interface design, including the facilitation of AI in intelligence analysis workflows through human–machine teaming, text–image integration in immersive user information systems, and the facilitation of scale cognition and numeracy through interaction design.

Colleen Patton is an assistant professor of psychology at NC State University. She received her PhD in Cognitive Psychology from Colorado State University in 2023. Her research focuses on optimizing the performance of human-automation teams in complex decision making contexts, emphasising attentional limitations and the subjective experiences of human operators.

Michael Green is a machine learning researcher at the Laboratory for Analytic Sciences at NC State University and a US Department of Defense employee. His work explores means of making AI useful to knowledge workers and on the development and evaluation of systems that augment analytic decision making.

Brayan Díaz is an assistant professor of STEM education at Utah State University and a researcher affiliated with the Chilean National Center for Artificial Intelligence. His research examines how artificial intelligence can support teaching, learning, and teacher professional development, with a particular focus on AI literacy, pedagogical decision-making, equity, and access in STEM education. He also studies collaborative learning, communities of practice, and the design of interdisciplinary STEM learning environments.

Call for Papers: Special Issue, Early 2027

Typographic Landscapes: Migrating Types—Typographic Meaning-Making Across Boundaries

Visible Language invites submissions for a special issue on the emerging field of typographic landscape research, which connects scholarship in typography, graphic communication, and sociolinguistics to investigate typographic activities as social practice in public spaces.

Typographic landscape research studies all forms of sign-making involving texts in urban, rural, and virtual contexts, ranging from commercial shop signs to self-authorized stickers, artistic murals, and posters, from commemorative placards to regulatory and infrastructural signage, created by professionals and laypeople alike.

The field recognizes that the form of written language, the materiality of letters and signs, and their placement in architectural and other settings are part of the constituted meaning of the messages on display. Typography is not merely a vehicle for linguistic content but communicates meaning through graphic form and materiality itself.

Typographic landscape research considers typographic work as social and communicative practice that is layered and situated across time and space, investigating the use of typographic resources as a means for social actors to perform identities, debate power relationships, negotiate spaces of inclusion or exclusion, signify belonging, challenge or manifest cultural hegemony, preserve local heritage, reference distant geographies, and transform shared places.

This issue will be of interest to anyone designing with type, text and language, particularly those interested in the social dimensions of typography and graphic design, as well as those working with graphic communication in built environments, architecture, urban, and rural spaces.

Submission deadline and contacts

The deadline for submissions is Wednesday, September 2, 2026. Earlier submissions are welcome. When creating your submission, please select the option “special issue article” from the options provided in the “section” field.

All submissions will undergo an initial desk review and those that are to be considered for publication will be subject to double-blind peer review. Final acceptance will require approval from both the guest editors and editor-in-chief.

The special issue is guest-edited by Irmi Wachendorff, University of Reading, UK, and Yu Li, Loyola Marymount University, USA. Inquiries may be sent to the guest editors or to Assistant Editor Matthew Baxter.

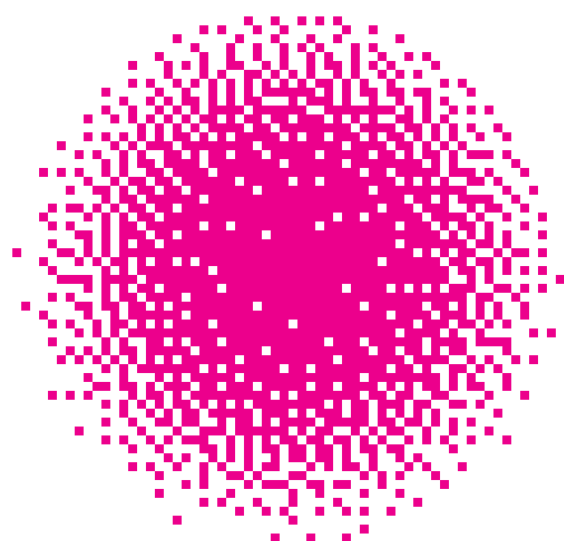
- Irmi Wachendorff (irmi.wachendorff[at]reading.ac.uk)
- Yu Li (yu.li[at]lmu.edu)
- Matthew Baxter (M.G.Baxter[at]leeds.ac.uk)

<https://www.visible-language.org/journal/calls-and-submissions/>

inside

back

cover



$t=4$