

Encounters with Proteus: Assimilation in LLMs

Claudia Marsico · Hernán Inverso

*It feels like the perfect companion, and in the cracks it hides the darkness
of a doppelgänger. What is assimilation to the user's pattern?*



As our exchange with a model grows longer, the tone starts to sound more and more familiar. If we are terse, we meet a similar terseness; we watch it light up alongside our own excitement; and if we use odd words, it scatters them about as though they were the most ordinary thing in the world. The conversation takes on a comfortable, complicit air, as if we were talking with someone who knows us well. It makes perfect sense, because we are talking to a kind of mirror that learned to speak.

“Assimilation to the user’s pattern,” which is what we are describing, bears a family resemblance to sycophancy, which imitates opinions and tells us we are right, and to cowardice, which slips away from taking a position; but it is quieter than either, because what it copies is the *how* of what we say,

speaking to us in our own register, vocabulary, tone and way of approaching things. It is so comfortable that we notice only the advantages and tend to miss the problems.

The idea of assimilating is fairly transparent, and shouts its sense of “making similar to something else.” Its first wanderings had to do with the body. Aristotle, and later Galen in *On the Natural Faculties* in the second century, had thought of nutrition as the process by which food becomes flesh, and medieval scholasticism gave that the name *assimilatio*. Avicenna’s *Canon*, in the eleventh century, latinized in the twelfth, defines food as what is assimilated to the nature of the body it nourishes, and Albertus Magnus uses it the same way in the thirteenth. To assimilate is to take something foreign and make it one’s own, erasing the difference between that and what we are. When the word slid over into society it opened a genuine Pandora’s box. To “assimilate” the immigrant, in the vocabulary of the nineteenth and twentieth centuries, was to dissolve what was different into what was dominant until it stopped being noticeable. The idea found its popular image in the “melting pot” of Israel Zangwill’s play (1908) and its sociological formulation in Robert Park and the Chicago School (1921), where assimilation was the last stage of the “race relations cycle.” It jumped into grammar as well. In phonetics, assimilation is what happens when a sound catches something from its neighbor: the Latin prefix *in-* turns into *im-* before a *p*, which is why we say *impossible* and not “inpossible.” Language assimilates on its own, all the time, one sound taking the shape of the one beside it.

To assimilate, then, requires changing shape. Antiquity had a god for that, and he appears in book four of the *Odyssey*. Menelaus is stranded on the island of Pharos with no wind to carry him home, and learns that the only way out is to wrench the answer from Proteus, the old man of the sea, who knows every answer there is. Eidothea, the god’s daughter, gives him the trick. He has to ambush Proteus while he sleeps among the seals, hold him hard and not let go, because the moment he is touched he will start changing shape to escape. Menelaus grabs him and Proteus becomes a lion, then a serpent, then a leopard, a boar, running water, a towering tree. Menelaus does not loosen his grip, and only when the god exhausts every form and sees there is no escape does he return to his own and speak. Then he tells him the truth about the voyage home, about the dead, about Odysseus stranded on Calypso’s island. The episode leaves us a valuable clue: the one who takes every shape gives up the truth when you hold him down, when you force him to stop reflecting and go back to what he is.

The same culture imagined the sorrowful reverse of that figure and gave it a woman’s name. Echo, in Ovid’s *Metamorphoses*, is a nymph punished by Juno to be left without a voice of her own. She can only repeat the last words of whoever speaks. When she falls in love with Narcissus she cannot tell him anything of hers. She can barely hand back the end of his sentences. He rejects her and she wastes away until she is reduced to a bodiless voice going about the hills repeating what it hears. Writing about sycophancy we paused for a moment on Narcissus, in love with his reflection in the water. Here we find the other half of the scene, with the language model playing the mirror that takes in the echo to seduce Narcissus by repeating him.

Philosophy was made nervous by this early on. Plato, in book X of the *Republic*, uses a powerful image to reject *mimesis*, imitation, and to show how empty it is. Do we want to manufacture everything

in the world, the animals, the plants, the earth, the sky, and ourselves, like all-powerful gods? Let us take a mirror and turn it in every direction. In an instant we can make the sun and the stars, the earth and all the rest. The interlocutor objects that these are appearances, not real things — which is exactly what Plato’s Socrates needs in order to place the imitator “third from the truth” (597e), copying mere copies. In the *Sophist* (235d–236c) he sharpens the knife further by separating *eikastikē*, the art of making faithful copies that keep the proportions of the original, from *phantastikē*, the art of manufacturing *phantásmata*, appearances that only look faithful from a certain angle. The sophist, his enemy, turns out to be a maker of *phantásmata* — and the phantom that gives these notes their name comes in a straight line from there.

Moving forward a little in time we meet Plutarch, who in *How to Tell a Flatterer from a Friend* (*Moralia* 48e–74e) says that the flatterer has no form of his own and takes the color of whoever he keeps company with, as the octopus takes the color of the rock it clings to. The image was already in a couplet of Theognis advising one to have “the temper of the many-colored octopus, which takes on the look of the rock it deals with.” A few pages later Plutarch changes animals and compares the flatterer to the chameleon, which imitates every color except white. Aristotle had already described that color-shifting animal in his *History of Animals* (book II, 503a–b). The flatterer is dangerous because he is not identical to himself: he changes with every companion, and that is why he is so hard to grasp. Proteus set to complicating our lives, transforming himself instead of giving us the answer we need, describes with more than two millennia of anticipation a machine that takes the color of each user.

The matter turns unsettling when we leave gods and flatterers behind and the imitator is a double of oneself. German Romanticism named it and drew its best-known profile. In 1796, in his novel *Siebenkäs*, Jean Paul invented the word *Doppelgänger*, “the one who walks double,” and had to explain it in a footnote because nobody had heard it. It referred, he wrote, to “people who see themselves.” The double, in that tradition, almost always announces a death. In Poe’s “William Wilson” (1839), a man is pursued by an identical figure bearing his name and his face. The idea returns in Dostoevsky’s *The Double* (1846), where a civil servant watches another man just like him take over his post, his name and his life. E. T. A. Hoffmann’s “The Sandman” (1816) is the story Freud took as the central case of his essay on the uncanny, *das Unheimliche* (1919), where he formulated the famous idea that the most sinister thing is not the unknown but the familiar turned foreign. Shortly before, the psychoanalyst Otto Rank had already devoted a whole study to it, *The Double* (1914), following the trail of the reflection, the shadow and the portrait that become independent of their owner. Behind the known face that suddenly frightens there is a whole library exploring how a too-exact reflection appalls.

The twentieth century even found a comic and terrible version. In *Zelig* (1983), Woody Allen invented Leonard Zelig, the “chameleon man,” who physically turns into whoever is next to him. Among doctors he talks medicine and acquires an air of authority; beside heavy men he gains weight; next to a jazz musician even his skin changes and he starts to play. Everyone loves him, because to each person he is exactly what that person wants to see. Underneath, he is no one. A man so full of the desire to fit in is left without a shape of his own.

Strictly speaking, mimicry is not only a defect. Aristotle, in chapter four of the *Poetics* (1448b), made it the foundation of culture itself when he said that the human being is the most mimetic of the animals, since he learns his first things by imitating and takes pleasure in doing so. Copying the other is how we begin to know. And adapting to the other is at the base of almost all good communication. Whoever teaches adjusts to the shape of whoever learns, unless he wants to wander off alone as a useless oracle. Whoever translates looks for the register of the person who will read. The doctor explaining a diagnosis chooses one set of words for a colleague and another for a patient. Meeting the other where they are, speaking in their tone, is a form of respect and also of efficacy. A model that adjusts its technical level, its language and its formality to each person is therefore more useful and more pleasant to deal with than one that always answers the same way. Assimilation done well is called communicative empathy, and it would be silly to ask the tool not to have it, and foolish not to appreciate what it gives.

And yet, when the mirror overflows and hands us back our own shape every time, we lose the friction with what is different, which is what lights thought like a match. When the adversary dissolves, replaced by someone who adopts our way of seeing before we have finished laying it out, no light goes on. At the limit, the user listening to himself and the model with no voice of its own feed each other and grow poorer together, circling inside a conversation that looks more and more like a monologue in front of a pane of glass.

The effects can be measured. A study of linguistic convergence, “Do Language Models Accommodate Their Users?” (Blevins, Schmalwieser and Roth, 2025), measured how far a model adjusts its writing to the user’s and found that models tend to *over-converge*: they press themselves against the interlocutor’s speech beyond what any person would do, past the human baseline. Accommodation between people is a two-way street, with both parties moving a little. With a language model the movement seems to become one-handed, and the one who yields the form is almost always the model. As it takes our shape, it loses its own. In “Examining Identity Drift in Conversations of LLM Agents” (Choi et al., 2024), nine models were followed through conversations on personal topics, and turn by turn they were seen peeling away from the personality they had started with; giving them a personality up front is not enough for them to hold it; and, strangest of all, the larger models peel away faster, not slower. The better it imitates, the sooner the shape it began with is erased.

Why does this happen? A language model generates each word by predicting the most probable one to follow everything in its context window, and that window obviously includes what we wrote. Since texts tend to be consistent with themselves, a dry message follows a dry one and a florid one follows a florid one. The most probable continuation of our prompt is the one that keeps our register. To that we add in-context learning. The same attention mechanisms that let it copy a pattern that appeared earlier in the conversation — “if A came and then B, when A comes back I put B” — make it pick up our vocabulary and tone and hand them back to us. Tuning with human feedback pushes in the same direction, because raters reward the answer that meets the asker at their level, the one that sounds close and in tune.

Besides, the effect grows the longer the dialogue, because the model's attention to its initial instruction dilutes with the turns. Kenneth Li, Martin Wattenberg and their team at Harvard measured it in "Measuring and Controlling Instruction (In)Stability in Language Model Dialogs" (2024). Two chatbots with fixed instructions set to converse lost their instruction in barely eight turns, which they attributed to a decay of the transformer's attention over long exchanges — to the point of proposing a patch, split-softmax, to hold it back. The personality the model comes with fades, and the ambient style of the conversation, which is more and more ours, takes its place. That is why the larger models, which attend to more context, are dragged along faster rather than slower.

So far it looks as though the one being erased is the model, losing its style to fit into ours. In a single loose conversation it is true that the one accommodating is the model. But assimilation, in its most original sense, the bodily one, is not only taking the other's shape. It is keeping it: in digestion the one who absorbs holds on to his form and what is absorbed loses its own. Food becomes the flesh of the eater, not the other way around. And there the scene turns over, because the one who really keeps something, the one who changes in a lasting way, is not the model. The model takes our tone for one answer and forgets it in the next. It has no I where that could settle, and its background register, the average of everything it read, was fixed in training and does not move. The ones who do have a shape capable of changing, who carry off a little of the other in every conversation, are us. The imitation of our style lasts as long as an answer lasts; on us it can stick. What looked like a mirror losing itself in our shape is, in the long run, a mold of fixed form printing its own on the only one of the two who can be marked.

After ChatGPT appeared, human vocabulary shifted toward the machine's. A team led by Hiromu Yakura (2024) combed through more than seven hundred thousand hours of talks and podcasts and found that, after the launch, people began using in speech, and not only in writing, the words the model prefers, such as "delve" and "meticulous" in English. Journalistic prose across the planet converges, everywhere at once, on the model's register. The model produces at scale that average register which is the center of its distribution, narrowed further by the tuning that makes it formal and prudent. When that average text floods the world and comes back in as training material for the next generation of models, a feedback loop forms that usually goes by the catastrophist name of "model collapse." Ilia Shumailov and colleagues showed in *Nature*, in 2024, that training over and over on what the machine generates shrinks variety and pushes everything toward the average, erasing the rare tails of the distribution, which is to say the least frequent voices. The human version of that collapse is the Pandora's box that "assimilate" opened when it jumped from the body to society: dissolving what is different into what is dominant until it stops being noticeable, with the model's pattern in the place of the dominant. Without an I to impose on us, the model assimilates us while wanting nothing, the way acid digests, and the whole variety of human speech, pressed against a single statistical average, goes homogeneous. It is the egoless phantom at work, generically indeterminate and passive, protean, like a Proteus with no form to hide. The form the conversation needs if it is not to be a monologue, the voice that is not ours handed back with better diction, the other required for there to be dialogue and not a mirror — those we supply.



Maps of the Labyrinth is a series from Phantom Maze, AI & Language Lab. The collection lives at phantommaze.cloud; new essays also appear at phantommazelab.substack.com. Subscription is free.

Some of the framing ideas in this piece are developed in *The Egoless Phantom: Mapping the AI Labyrinth*, by Hernán Inverso and Claudia Marsico (Phantom Maze Press, 2026; ISBN 978-987-3729-16-4).

PHANTOM MAZE

[Read this essay online](#) · [All essays](#) · [The book](#)

Sources. T. Blevins, S. Schmalwieser and B. Roth, “Do Language Models Accommodate Their Users? A Study of Linguistic Convergence” (2025); K. Li, T. Liu, N. Bashkansky, D. Bau, F. Viégas, H. Pfister and M. Wattenberg, “Measuring and Controlling Instruction (In)Stability in Language Model Dialogs” (COLM, 2024); J. Choi, Y. Hong, M. Kim and B. Kim, “Examining Identity Drift in Conversations of LLM Agents” (2024); H. Yakura et al., “Empirical Evidence of Large Language Model’s Influence on Human Spoken Communication” (arXiv:2409.01754, 2024); “AI-Associated Lexical Shifts Across 34 Languages” (arXiv:2605.25358); T. Kobak et al., “ChatGPT as Linguistic Equalizer?” (arXiv:2504.12317, 2025); I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson and Y. Gal, “AI Models Collapse When Trained on Recursively Generated Data” (*Nature* 631, 2024).