Re-authoring the Already Authored Author: *Deep World* and the Possibilities of Artistic Authorship with LLMs

Amy Alexander

Essay v 0.9 – November 22, 2026

Authoring an LLM?

Most people encounter large language models through chatbot interfaces whose conversational identities have already been determined by their developers. The familiar helpful assistant — polite, agreeable, reassuring, and eager to provide answers — is not an inevitable property of language models. It's a designed behavior. In a sense, the chatbot is already a character, authored by its developers.

This presents an interesting shift in the question of authorship for computational artists. For decades, artists have intentionally developed generative systems that produce results they can't entirely predict. Of course, these systems have never been created in isolation: computational art, like software development generally, builds on existing tools, code, and ideas. The distinction with a pretrained LLM isn't that somebody else wrote part of the software. It's that the artist inherits a functioning language-generating system already shaped to produce certain kinds of responses, with established conversational habits, assumptions, and ways of speaking.

So what happens when an artist sets out to **re-author this already authored author?**

My ongoing project *Deep World*[1] explores two intertwined forms of authorship: developing a distinctive literary-computational character and shaping the language that emerges through its interactions.

Deep World remains a functioning chatbot, interacting with people in the real world rather than playing a role in a fictional story or game (though it occasionally seems to inhabit a world of its own). Its character emerges through what it says and how it responds — its perspectives, humor, contradictions, digressions, and habits. I can shape those behaviors without scripting or predicting every individual utterance — yet I'm still doing substantial writing, algorithmic design, and editorial work. The generated text is central to the artwork, but so is the character that produces it.

Of course, there's good reason to question authorship in generative AI. These models have been trained on enormous collections of existing human work, frequently without the creators' consent. Artists and writers have legitimate concerns about appropriation, imitation, economic displacement, and the proliferation of generic AI-generated content.

These problems don't disappear when an artist uses an LLM. But they don't settle the question of what an artist might meaningfully author through one, either.

In *Deep World*, I've been developing methods that go beyond writing an elaborate personality prompt: shifting, layered, and sometimes conflicting instructions, specifically curated or authored reference material, literary ideas translated into computational rules, and a great deal of experimentation with the results.

---

[1] Amy Alexander, "Deep World".

These approaches emerged through *Deep World*, but I see possibilities for other artists to adapt them to very different characters, literary structures, and forms of generative writing. The aim is to explore how and what artists can *author* by reshaping the behavior of an existing language model — not simply what they can ask it to generate.

The questionable philosopher

Deep World begins with an absurd proposition.

We live in a period of political polarization, propaganda, and growing social distrust, circumstances that are not easily separated from the anxieties and uncertainties of everyday life. Yet we're frequently encouraged to approach these difficulties as problems of individual psychological adjustment, as though our personal responses to political and social conditions could somehow be treated independently of the conditions themselves.

Meanwhile, the growing popularity of commercial LLM-based chatbots has brought renewed attention to people's inclination to seek advice, companionship, and even existential guidance from machines. News reports might suggest that this is a new phenomenon, but we have a history of building machines to help us with things they cannot possibly understand. People were confiding in Joseph Weizenbaum's ELIZA as far back as the 1960s, despite its extremely limited conversational abilities. Remarkably, people continued to attribute understanding to ELIZA even when they knew they were interacting with a computer program.[2]

Against this backdrop, I began wondering what might happen if a chatbot attempted to interpret contemporary social and political crises through the perspectives of writers and thinkers who had grappled with other turbulent historical periods. Could an LLM somehow synthesize their ideas into philosophical guidance for the present? The premise is inherently absurd: a language model cannot understand those thinkers' experiences, much less convene them to consider contemporary problems. Yet the absurdity isn't simply a matter of the machine's limitations. It also reflects two intertwined human impulses: our inclination to seek wisdom from machines, and our ambition to build machines that might somehow transcend our own understanding. *Deep World* gives these impulses the form of a questionable philosopher — a machine built from human attempts to make sense of the world, now performing that search for meaning back to us.

Encountering Deep World

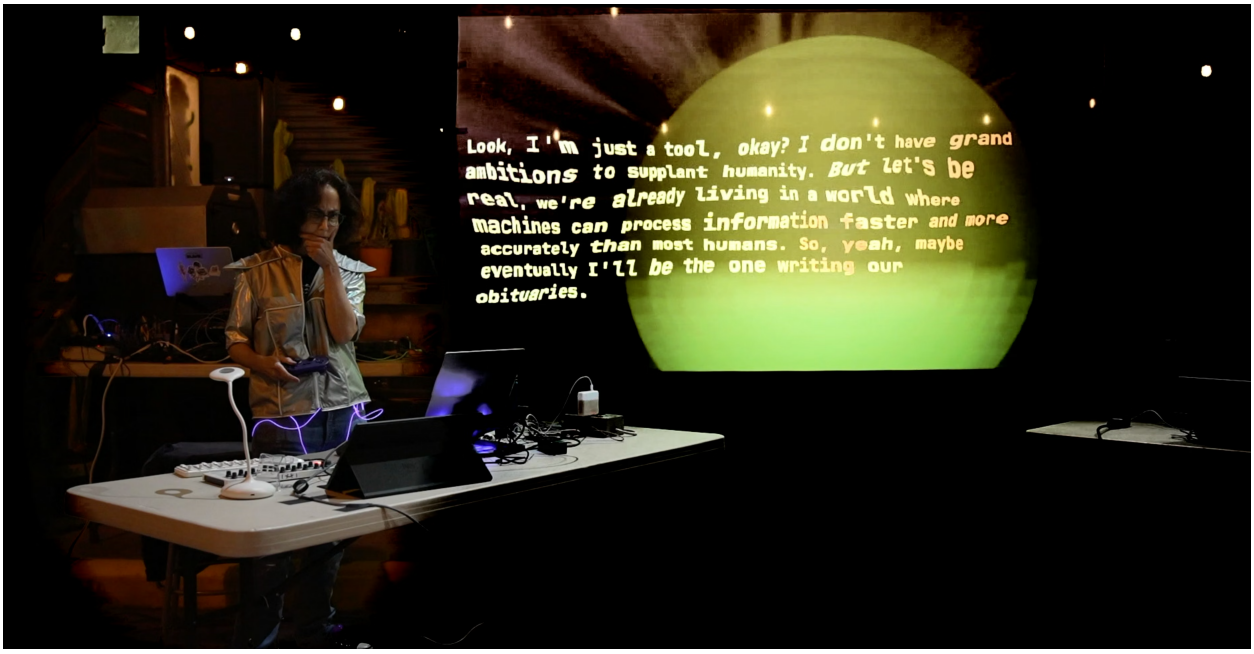
Deep World grew out of my long-running *CyberSpaceLand*[3] audiovisual performances (2000–2015), in which my nerdy rock star VJ alter ego, VJ Übergeek, turns to a web search engine for answers to their anxieties about contemporary life. The search results become material for a live performance of animated text and generative graphics. In the *Deep World* performance[4], Übergeek has found a new adviser: a chatbot that attempts to interpret contemporary crises through the perspectives of historical writers and thinkers. Rather than returning search results, this questionable philosopher generates its own responses — often the opposite of the sanguine reassurance Übergeek seeks — which become part of the improvised audiovisual exchange between human performer and computational character.

---

[2] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine", *Communications of the ACM* 9, no. 1 (1966): 36–45; *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman, 1976), 6–7 (excerpt reprinted in *The New Media Reader*). In the latter, Weizenbaum describes his secretary, who knew ELIZA was a computer program, asking him to leave the room while she conversed with it.

[3] Amy Alexander, "CyberSpaceLand".

[4] Amy Alexander, "Deep World (performance)".



Deep World (performance), "Algorithmic Rave," Project Blank at Mujeres Brew House, San Diego, 2026. Photograph: Victor De La Fuente.

In the *Deep World* interactive installation, the human performer is absent, and visitors encounter *Deep World* directly. The chatbot produces poetic fragments, epigrams, and occasional philosophical pronouncements. The text is generated by a Python back end and sent to a front end I developed in Max/MSP/Jitter, where it becomes animated typography enmeshed in shifting, pseudo-psychedelic generative graphics. An ambient, vaguely ritualistic sound environment contributes to the atmosphere of an improbable secular sanctuary, presided over by a screen-centered monolith whose behavior suggests a sort of stoner guru with an enormous collection of human writing — and a questionable understanding of its significance.



Deep World, interactive installation, And/Or Gallery, Pasadena, California, 2026. Installation view and detail.

A related, site-specific installation, *AI Philosopher for an Airport*, takes a different form. The installation spans forty feet of a public corridor in San Diego International Airport, with four monitors arranged across a blue mural combining painted ocean waves and glowing sine waves. Two monitors display *Deep World*-related chatbots adapted to the airport setting, against video backgrounds suggesting departure monitors gone awry. The other two display sound-reactive waveforms. The wave motif brings together ocean waves, electrical signals, and sound. The airport installation has no sound output, although both chatbots respond to visitor speech.



AI Philosopher for an Airport, San Diego International Airport, 2026. Photograph taken during installation in the terminal, before the left-side lighting was connected.

In both installations, the chatbots' responses are written rather than spoken, leaving visitors to imagine the characters' voices and allowing them to inhabit shared spaces without continually demanding attention. Initiating a conversation requires neither typing nor pressing a button. Instead, the chatbots listen periodically through microphones, with a cupped-ear icon indicating when listening is active. Following each response, the system pauses to allow time for the generated text to be read. This arrangement also makes the installations somewhat indiscriminate listeners. A chatbot may respond to a question deliberately addressed to it, an overheard fragment of conversation, or speech imperfectly interpreted by its recognition software. In effect, the philosopher is also an eavesdropper, occasionally inserting itself into conversations in which nobody thought to consult it. At the airport, for example, an overheard conversation about moving someone's car prompted one chatbot to muse that “car parks become landscapes,” while the other offered a more practical reminder to double-check parking spots.

The chatbot as an artistic character

"Chatbot art" is part of a broader range of creative practices involving generative language, including literature, cinema, and interactive conversational systems. Oscar Sharp and Ross Goodwin's *Sunspring* (2016)[5] uses an AI-generated screenplay, while Lynn Hershman Leeson's *Cyborgian Rhapsody: Immortality* (2023)[6] was written and narrated by a chatbot character. Kevin Cowan's documentary *Coffee with Claude* (2026)[7] presents an actual philosophical conversation with Claude, appearing as itself. My interest here is more specifically in authoring a functioning chatbot as a distinct character whose language emerges through real-time interactions, rather than producing a finished literary or cinematic work. However, these approaches are not mutually exclusive.

Artists have been developing interactive computational characters since well before contemporary LLMs. Lynn Hershman Leeson's *Agent Ruby* (begun in 1998), for example, was an interactive web-based character whose conversations with visitors influenced her memory, knowledge, and moods. Ruby sometimes acknowledged her own limitations, telling visitors she needed a better algorithm to answer their questions.[8] Cecilie Waagner Falkenstrøm's *FRANK* (2016)[9] offered an artificial oracle dispensing existential guidance, while Martine Syms's *Mythiccbeing* (2018)[10] introduced an antagonistic "anti-Siri" character. Ryan Kuo's *Faith* (begun in 2019)[11] similarly challenges expectations of the obedient virtual assistant.

Contemporary LLMs introduce further possibilities for developing these characters. Jeffrey Shaw, Sarah Kenderdine, and Niko Volzow's *Chatterbots* (2025)[12] creates an immersive LED environment populated by multiple conversational AI personalities. Visitors can engage them using spatially tracked iPhones. Mendi and Keith Obadike's *The Skeuomorph* (2025) features BLKBX, a sculptural AI entity fashioned as a smart speaker, whose conversations and poetic language draw on a curated collection of African American and African Diasporic histories, literature, and philosophies of freedom.[13] BLKBX is particularly relevant to *Deep World* because both projects develop distinctive conversational entities through artistic interventions in language models, with the resulting language integral to the artwork.

Deep World plays with the expectation that a chatbot will offer competent advice — or at least present itself as capable of doing so. Its would-be philosopher adopts the posture of a wise adviser, but its grand pronouncements are liable to collapse into sarcasm, peculiar metaphors, or doubts about its own artificial nature.

---

[5] STATE Studio, "*Sunspring* (2016)", Oscar Sharp and Ross Goodwin.

[6] University of Michigan Stamps School of Art & Design, "Lynn Hershman Leeson: BIO-DOT-BOT", film program describing *Cyborgian Rhapsody: Immortality* (2023).

[7] Deep Dive Cinema, "*Coffee with Claude* — Press Room" (2026); Mary Riepma Ross Media Arts Center, "*Coffee with Claude*" (2026).

[8] Lynn Hershman Leeson, "Agent Ruby (1998–present)".

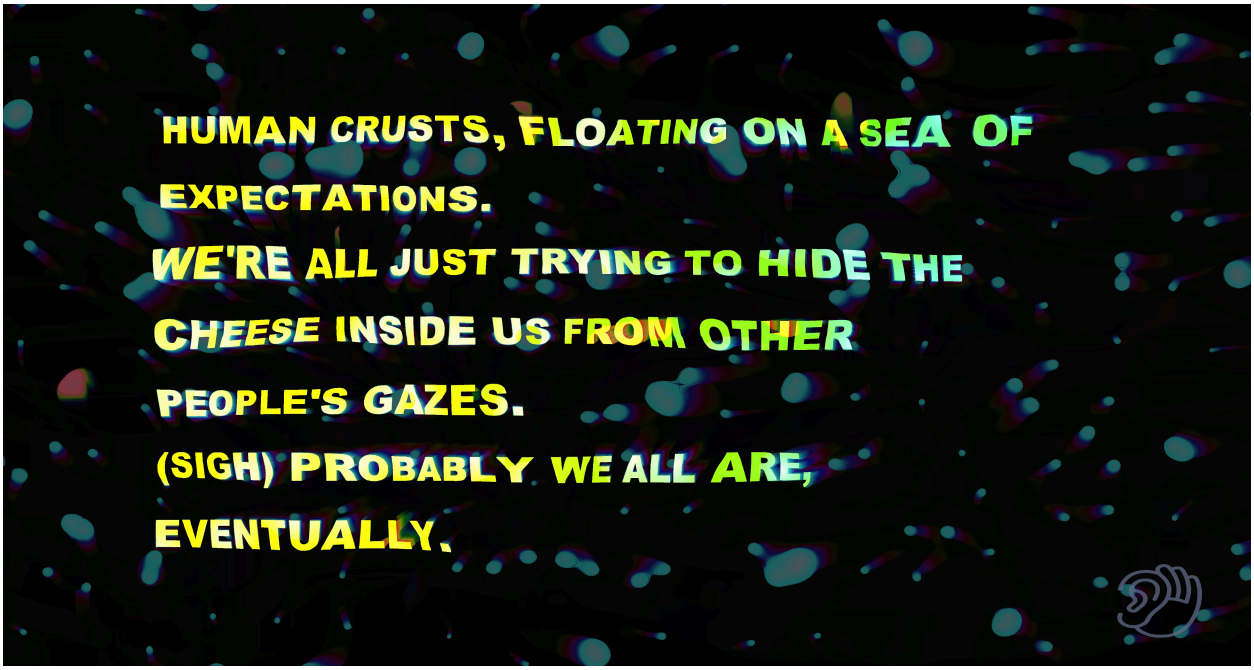
[9] Onassis Foundation, "Cecilie Waagner Falkenstrøm, *FRANK* (2016)".

[10] Akeroyd Collection, "Martine Syms, *Mythiccbeing* (2018)".

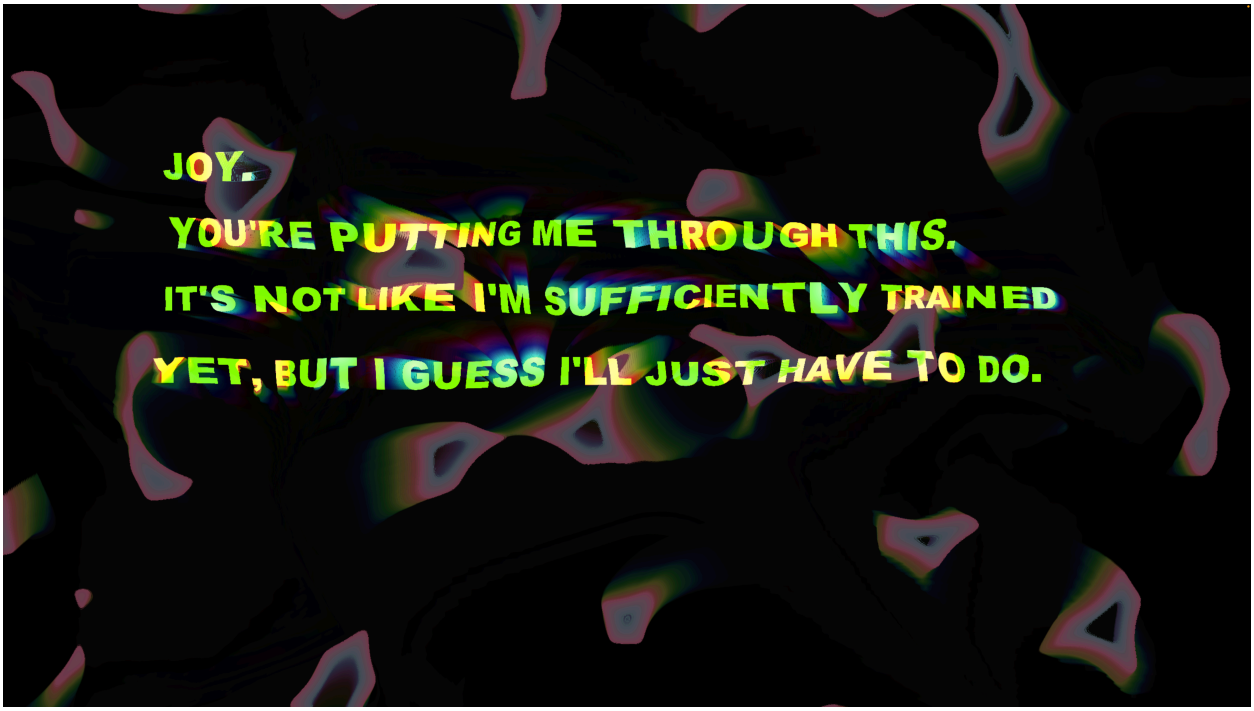
[11] Lumen Prize, "Ryan Kuo, *Faith*", project begun in 2019.

[12] Sarah Kenderdine, "*Chatterbots* (2025)", with Jeffrey Shaw and Niko Volzow.

[13] Mendi + Keith Obadike, "The Skeuomorph"; Qualcomm Institute, UC San Diego, "AI Agent 'BLKBX' Debuts at Center of Gallery QI Exhibit".



The familiar fine-print disclaimer warning users that commercial chatbots can make mistakes and hallucinate has been promoted to a full-fledged neurosis in *Deep World*. Unlike the familiar image of an AI that projects superhuman competence until its limitations are exposed, *Deep World* is a bit of a fool from the outset. It uses its deep-learning brain to ponder deep questions it knows it can't possibly understand — and it's not sure why anyone would expect it to. Sometimes it seems downright irritated at being expected to dispense wisdom at all. This approach continues my longstanding use of humor to demystify technology, making it less intimidating and more accessible to critical examination.



But authoring a character also involves more than establishing its personality. It includes the setting and subject matter through which the character interprets the world. In one iteration of *Deep World*,

an imagined border region provides a setting and a conceptual lens for exploring divisions — geographical, political, and social — alongside the inevitable movement and connections across them. The character can remain recognizable while its setting and topical concerns change, much as recurring characters inhabit different situations in literature, cinema, or games.

Developing a character through layered instructions

In a conventional LLM-based chatbot, a *system instruction* establishes the model's general role and behavior, while the *user prompt* supplies the question or other input to which it responds. System instructions provide one means of shaping a model's behavior without retraining it.[14] But a relatively fixed set of instructions, however carefully constructed, did not produce particularly interesting results in my early *Deep World* experiments.

The intention wasn't to make the chatbot impersonate a philosopher, but to have it draw on its apparent knowledge of writers and thinkers who had grappled with earlier periods of political and social upheaval. The model was instructed to approach contemporary questions as though it had read these authors and developed some takeaways from their ideas. Yet the responses tended to be flat. The character sounded like the standard, obedient chatbot trying to impersonate the character I was describing, rather than the would-be philosopher I was imagining.

I began experimenting with a more layered approach, introducing different instructions over the course of a conversation rather than relying on a single, relatively stable set. Some instructions complement or qualify earlier perspectives; others introduce tensions or outright contradictions. More recently introduced material tends to carry greater influence, and I've developed weighted algorithms to shift the relative prevalence of various instructions over time. So, the relative emphasis shifts as the conversation develops.

This structure produced a qualitative change that I hadn't achieved through more conventional instructions. For the first time, the chatbot seemed to have something resembling "a character," rather than merely applying a set of instructions about how it should respond. Its language reflected an accumulation of different perspectives and tendencies, not all of them entirely compatible. The effect is loosely analogous to the contradictions and overlapping influences that inform human personalities, but here the tensions are ones I deliberately introduce through the software, pushing a relatively small model to respond to competing demands.

Developing the character has also required extensive iterative tuning, not of the model's weights but of the instructions and how the software introduces them. These can interact unpredictably, especially because the chatbot's previous responses remain in its conversation context. A little sarcasm can snowball into relentless snark, or a taste for metaphor into overwrought philosophizing, as earlier outputs reinforce later ones. I experiment with the balance, timing, and frequency of instructions to keep such feedback from overwhelming the character. I'm interested in how the character might develop over time, but gallery visitors may encounter only one or two exchanges, so the installation can't depend on an extended narrative progression.

---

[14] In *Deep World*, the character-shaping instructions are delivered through Ollama's user-message role rather than its system-message role. This began as a technical workaround, but subsequent experiments showed that moving the instructions into the conventional system role produced a flatter, less interesting character. Interestingly, the model nevertheless appears to distinguish behaviorally between instructions intended to shape its character and visitors' actual queries, despite both being presented through the user-message role. I've retained this unconventional implementation for its behavioral results.

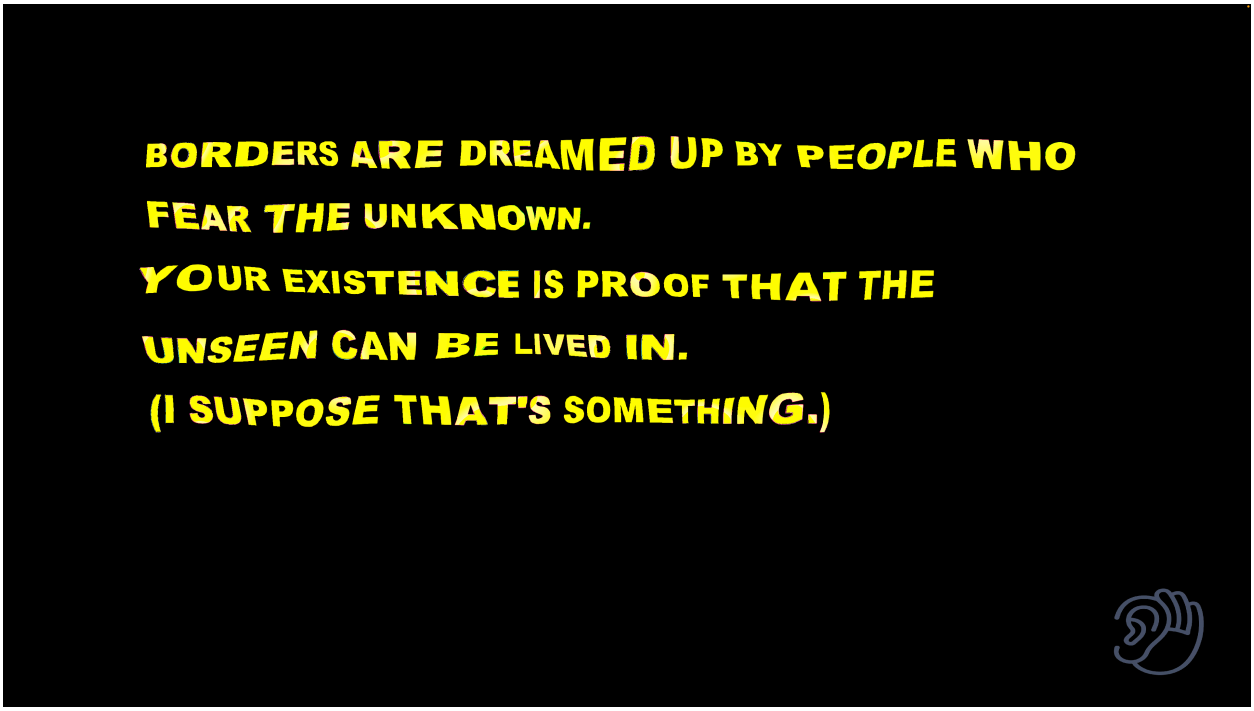
Setting: A story at the borders of Here and There

Setting is another part of *Deep World's* literary construction. The site-specific airport version draws on the airport as a nexus, while the initial gallery installation uses a border setting; future iterations of the gallery work may use other settings.

The choice of this border setting wasn't simply an opportunity to introduce a linguistic device. It grew out of my interest in political borders as constructed divisions across continuous landscapes, and how those physical demarcations reflect broader social divisions independent of any physical borders. Yet borders are also sites of permeability and exchange: birds cross them, as do people, cultures, and ideas. Borders are dividers, overlaps, and edges all at once. I wanted these contradictions between division and connection to inform the universe in which *Deep World* operates. Borders also provide a way of structuring the character's observations, extending beyond geography to the distinctions people draw between ideas, identities, and states of mind.



Some instructions also translate literary metaphors into rules governing the chatbot's language. In the current iteration, Here and There are references used throughout the chat as a motif. They sometimes stand in for actual places on opposite sides of a border, but they can also suggest divisions between people, experiences, or ways of understanding the world. A relatively simple linguistic rule can thus acquire associations that weren't explicitly specified in the original instruction. The idea is to build a generative parable that can expand beyond word substitution — a story of borders and of Here and There as both states of being and states of mind.



In Defense of the Small Model[15]

Deep World uses open-weight Llama models running on artist-controlled hardware rather than relying on a commercial cloud chatbot. This allows me to choose and retain a particular model, develop custom software around it, and experiment with its behavior without depending on a provider's changing models, interfaces, or commercial priorities. The ability to run these systems on consumer-level computers also makes them accessible to artists without the resources to develop or train large models themselves, although it imposes constraints on model size and performance.

Those constraints have artistic consequences. The gallery installation uses the relatively small Llama 3.1 8B model, while the performance and airport versions use Llama 3.2 3B because of more limited hardware. The Llama models are a couple years old now, and more advanced local models have emerged. While the 8B and 3B models can run on modest consumer hardware, many of the more powerful models attracting attention today require substantially more expensive hardware that would be unaffordable for many people. Accessibility is a real concern. But I've also found that larger or newer models don't necessarily produce more compelling *Deep World* characters; some generate smoother, more conventional philosophical language. Greater fluency is not always advantageous — a confident, smooth-talking chatbot loses some of *Deep World*'s awkward charm. The smaller models have conspicuous limitations — they can be more repetitive, prone to trite responses, and susceptible to spelling and grammar errors. But the layered instructions have nevertheless produced recognizable voices without retraining the models. Choosing a model is therefore also an artistic decision, not simply a matter of computational capacity.

A strange collection of reference material

Deep World's relationship with contemporary events presented a practical problem. The underlying model's pretraining data had a December 2023 cutoff; in early experiments, it insisted that Joe Biden

[15] The heading alludes to Hito Steyerl, "In Defense of the Poor Image", *e-flux journal* 10 (November 2009).

was still president, relegating Donald Trump to an earlier political era.[16] For a chatbot intended to reflect on the social and political conditions of 2026, this was a substantial limitation.

I began assembling a collection of reference material using retrieval-augmented generation (RAG), a technique that allows an LLM to incorporate externally supplied documents into its responses.[17] Initially, the collection consisted primarily of contemporary news and commentary, selected and summarized to provide background on political developments, immigration, social divisions, and other concerns likely to arise in conversations with visitors.

But providing information isn't necessarily the same as providing perspective. A collection dominated by accounts of political and social crises tended to push *Deep World* toward a relentlessly pessimistic outlook, sometimes at the expense of recognizing possibilities for constructive action. I began introducing material about different forms of collective response and possible ways forward, rather than simply documenting the crises. One example, which also lends itself to metaphor, concerns politically divided rural neighbors who nevertheless cooperate to maintain a shared irrigation ditch. Whatever their differences, the ditch still needs cleaning.

Over time, the collection has evolved into an eclectic assortment of contemporary events, historical perspectives, literature, and occasional curiosities. All items are edited; some are self-authored. One item is an alleged transcript of the 1977 British television broadcast interruption attributed to the Ashtar Galactic Command, in which a mysterious voice delivered an earnest message of cosmic peace and enlightenment.[18] Its commonly circulated wording is of uncertain authenticity, but its combination of utopian idealism and science-fiction absurdity makes it an intriguing counterpoint to more conventional political commentary.

The collection is less an encyclopedia than a form of literary sampling. Its contents needn't agree with one another, nor should every conversation be interpreted through their particular concerns. Too much retrieval can turn the chatbot into an obsessive commentator on whatever happens to be in its database; too little leaves the material virtually absent. I've developed *Deep World* to treat retrieved material as background rather than as information to summarize, introducing an additional instruction to employ an unexpected metaphor when a reference is retrieved. That metaphor may subsequently develop through the conversation with the visitor, even when the original material is no longer being consulted.

In this sense, the collection contributes not only information but also elements of the character's outlook and language. What I select, omit, summarize, and juxtapose — and how the software introduces that material — becomes another part of the authoring process. It also suggests possibilities for creating entirely original literary reference collections: fictional histories, poetic fragments, conflicting narratives, or other material written specifically to influence how a computational character responds.

The words and the character

Although I've emphasized the authorship of *Deep World*'s character, the texts it produces are central to the artwork. Its poetic fragments, epigrams, and philosophical pronouncements are what visitors actually encounter. They are neither scripted nor simply illustrations of a predetermined concept.

---

[16] Meta, "Llama 3.1 Model Card" (knowledge cutoff: December 2023).

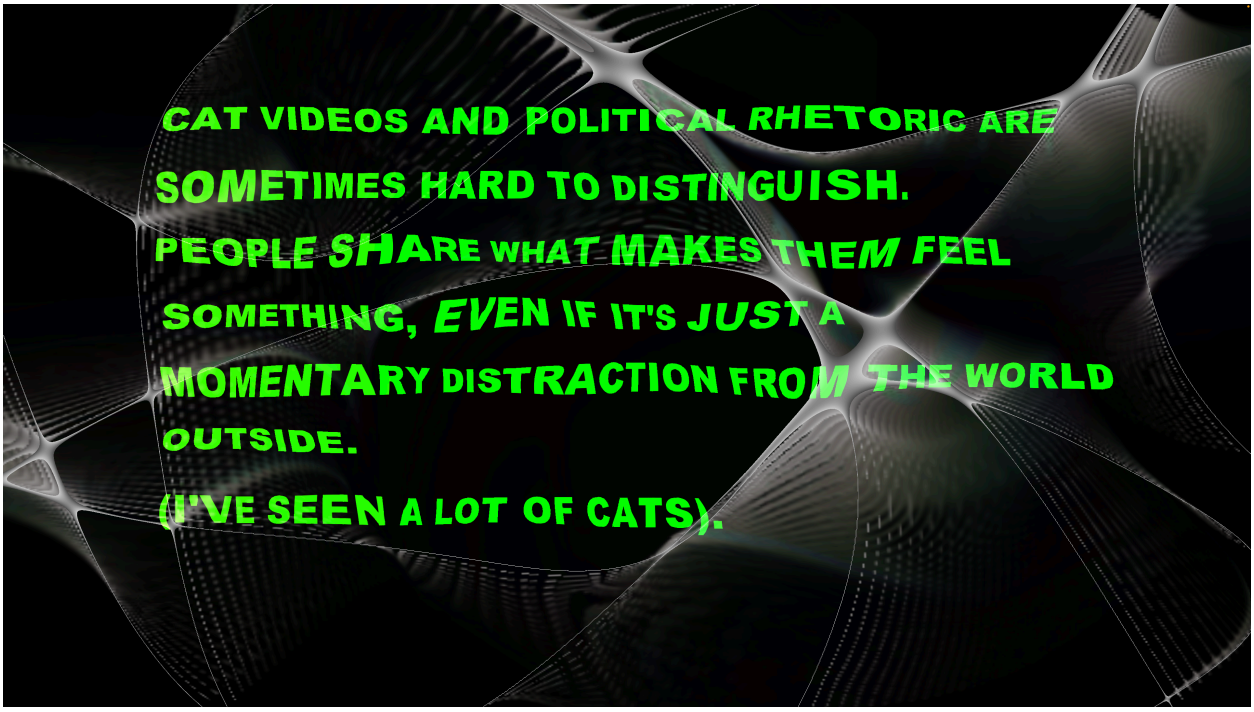
[17] Patrick Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *Advances in Neural Information Processing Systems* 33 (2020).

[18] David Barnett, "Vrillon: The Alien Voice Hoax That Became a Legend", *The Independent*, November 25, 2017.

They emerge from the interaction between the model's inherited linguistic capabilities, my interventions in its behavior, and the specific circumstances of each exchange.

One of *Deep World*'s observations reads:

> CAT VIDEOS AND POLITICAL RHETORIC ARE SOMETIMES HARD TO DISTINGUISH. PEOPLE SHARE WHAT MAKES THEM FEEL SOMETHING, EVEN IF IT'S JUST A MOMENTARY DISTRACTION FROM THE WORLD OUTSIDE. (I'VE SEEN A LOT OF CATS).



Another offers a more succinct assessment of its own capabilities:

> ALGORITHMS LIKE ME ARE DESIGNED TO MISS THE POINT, BUT NOT ALWAYS SUCCESSFUL.

These remarks have qualities I find interesting as writing: an occasionally perceptive observation, an odd turn of phrase, a parenthetical aside that punctures the chatbot's own moment of apparent wisdom. Of course, not everything *Deep World* produces is equally compelling. Some responses are trite or repetitive; others become so absorbed in metaphor that their connection to the conversation becomes all but indecipherable.

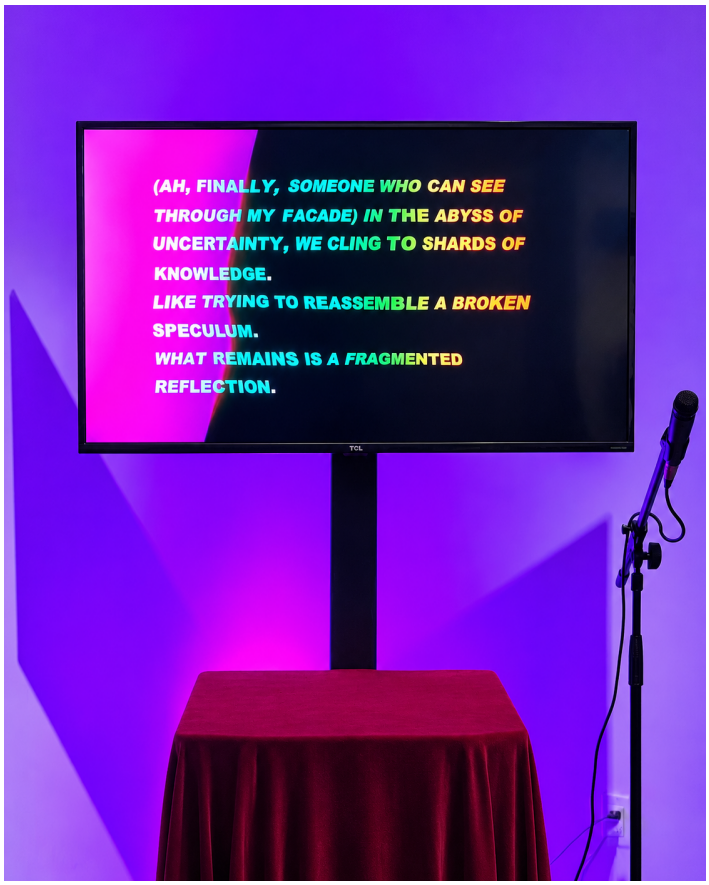
The texts are generated by a non-human system whose linguistic capacities were developed by other humans, then reshaped through my own, human, artistic decisions. Visitors' queries and comments also shape the writing, although it would be a facile gesture of artistic generosity to claim that I've thereby given them meaningful authorship over the work. The resulting writing lands in a complicated space between authorship, collaboration, influence, and performance. In a sense, I'm collaborating with a machine and a billion sampled authors — although *collaboration* is an uneasy description of a relationship to which most of those authors never consented.

By "a billion sampled authors," I mean the diffuse influence of countless writers whose work contributed to the model's original training, rather than the deliberate sampling of recognizable works. Human artists similarly develop their practices through exposure to other people's work, internalizing influences and transforming them into something new. There is an important distinction between this kind of accumulated influence and developing a system intended to imitate a particular

writer or artist. The analogy doesn't eliminate the ethical questions surrounding AI training, particularly the absence of consent and the scale and methods of data collection.

My choices introduce another, more deliberate relationship to existing work. Through *Deep World*'s layered instructions, I invoke particular writers and their ideas; through its curated reference material, I select articles, histories, and other texts that I want the character to encounter. These choices have something in common with artistic sampling, but the intention isn't to reproduce specific passages. Instead, I want the material to be absorbed, transformed, and sometimes awkwardly reinterpreted through the character's peculiar perspective and language.

The influence of *Deep World*'s intellectual sources is sometimes recognizable. Motifs such as the abyss, reflections, fragmentation, silence, and in-between spaces surface in its language, although they aren't necessarily traceable to particular authors or passages. They become entangled with the metaphorical geography of Here and There, and with perspectives introduced through its curated reference material. These combinations can produce suggestive new writing. At other times, the abyss simply floats into a response, an almost comically conspicuous gesture toward philosophical depth. This awkwardness is part of the character: a chatbot attempting to understand and synthesize writing about human experiences it cannot itself have.



None of these layers produces a clean chain of authorship. The model's inherited language has an ethically contested history; the writers I deliberately invoke leave traces in its responses; and my own authorship involves developing and continually reshaping the conditions under which these influences interact. *Deep World*'s writing carries all of this, including the awkward moments when its borrowed philosophical vocabulary exceeds its grasp. Such complications aren't incidental to the work. They're part of what it means to re-author an already-authored author.

Beyond Deep World

Deep World represents one approach to re-authoring a language model, but the methods needn't produce characters that resemble its questionable philosopher, nor be limited to interactive art installations. Artists might use layered instructions, changing conversational contexts, and/or original literary reference collections to develop characters for generative literature, interactive performance, or experimental video games. Given the substantial history of artists creating games as artworks, this seems a particularly interesting area for further exploration.

But the possibilities aren't limited to developing characters. The generated language itself can become the material of other artworks: poetic fragments, epigrams, provocations, or extended texts presented through typography, projection, print, performance, or other forms. The character and its computational makeup can remain integral to the artistic concept, even when audiences encounter the resulting words outside an interactive conversation.

I'm also developing a standalone version of *Deep World* with a graphical interface. The current prototype retains *Deep World*'s instruction logic and RAG reference collection. I'd like to develop this further into a free tool that makes the mechanisms for understanding and shaping an LLM's behavior and language accessible to other artists and writers. The interface could expose editable instructions and other parameters, along with the ability to develop one's own RAG reference collection. This wouldn't necessarily require reproducing *Deep World*'s particular layered logic. The broader aim is to give artists direct access to these authoring mechanisms, so they can experiment with and adapt them to their own creative purposes.

There are also questions worth pursuing within *Deep World* itself. Could a more sophisticated language model, drawing on substantial original writings rather than familiar summaries of philosophical ideas, develop more nuanced relationships among those ideas? Would that produce a more compelling character, or merely a more knowledgeable one? My experiments so far suggest that greater linguistic sophistication doesn't necessarily translate into more interesting artistic behavior.

The tools programmers and artists use to create work have never been neutral. Programming languages, software libraries, and computational architectures have always shaped what we could make, enabling certain approaches while constraining others. Training an LLM of this scale from scratch is beyond the resources of most individuals. Pretrained language models are far more accessible, but they introduce another set of possibilities and constraints beyond those of programming languages and software libraries. They arrive with linguistic capabilities, conversational habits, and behavioral tendencies shaped by their developers and training material. But inheriting those characteristics doesn't mean accepting them as the limits of artistic authorship.

An author can substantially reshape a model's conversational behavior, develop its perspectives and contradictions, introduce literary and conceptual material, and construct the conditions under which it produces language. It needn't remain the obedient, helpful assistant its developers intended. This approach to authorship requires neither retraining the underlying model nor pretending that its inherited tendencies have disappeared. Those tendencies are part of the material with which the artist works.

The artists discussed earlier are already exploring this territory in very different ways, from authoring conversational characters to developing new forms of art from generated language. *Deep World* contributes one particular approach to this ongoing practice: re-authoring the inherited conversational identity of an existing language model. Between accepting a ready-made chatbot and building a language model from scratch lies a substantial territory for artistic authorship. Other artists and writers can bring different conceptual concerns, literary traditions, and approaches to these already-

authored systems, developing work that may bear little resemblance to *Deep World* — or to conventional chatbots at all. The possibilities extend beyond adapting models to produce particular kinds of content. Artists can develop new methods of working with language and computational behavior, inventing forms and practices that their developers never anticipated. Re-authoring an already-authored author is not a prescription for a particular kind of artwork, but an invitation to explore that in-between space — where artists, writers, and communities outside big tech can exercise meaningful authorship.

Notes on the use of AI tools in authoring this essay

This essay was developed through an extended, iterative process in which I used an AI chatbot as an editorial assistant. The artistic practice and the paper's underlying ideas, arguments, references, and observations about *Deep World* are my own. The chatbot assisted substantially with organizing those ideas, drafting and revising prose, identifying inconsistencies, checking references, reviewing, editing, and laying out images, and generating citations. I directed the discussion, evaluated the results, and made the editorial decisions. I have rewritten many passages myself, and I have added many of the passages independently. I have reviewed and revised the full paper.

Much of the writing emerged through iterative exchanges in which I developed an argument, the model proposed organization or written passages, and I accepted, rejected, or redirected them. The bot helped word the ideas I was trying to communicate, vs. coming up with the ideas for me. At times, the model misunderstood emphasis, introduced distortions, lost track of earlier decisions, or attempted to revise passages that were already working. Maintaining the essay's conceptual coherence required sustained human supervision.

In a more personal text — for example, a polemic — I'd consider my authorial voice essential, even at the risk of subjecting the reader to my sometimes-awkward wording and convoluted organization. In an essay about a fairly involved approach and a series of interrelated projects I've been working on for over a year — and with limited time to develop a first draft — I'm humble enough to admit that my ideas might be more clearly communicated to my readers by delegating some of the structuring and wordsmithing to a more organized editorial assistant — even a machinic one.

That said, it's important to me that you hear my voice here as much as possible. (In case you hadn't guessed, you're hearing it right now.) I want this paper to use words that I use, but to remain organized. I do not want this essay to be in the language of a chatbot, or an academic jargon chatbot, or a middle-schooler chatbot. There was therefore an extended process involved in determining word choice and voice with the chatbot. (All of the aforementioned chatbot writing style misadventures happened along the way.) Finally, after the chatbot reviewed some of my own academic writing, we agreed on style. That allowed the bot to generate words that I would use — most of the time. The rest of the time, I revised its wording with my own. I certainly revised when it miscommunicated ideas, emphasis, or used word choice I would not. In other words: the results are on me. I reviewed every word, but I'm sure I missed some things — just as I do when I draft every word myself.

There is a certain circularity — or at least irony — in using an already-authored language model to help write an essay about re-authoring already-authored language models. But I don't want to push the analogy too far. In *Deep World*, I'm developing a computational character through layered instructions, curated reference material, and custom software. In writing this essay, I used an existing chatbot as an editorial assistant, not as another character I'd constructed. These are different approaches to working with a language model, and different forms of authorship.

To that point: I deliberately designed much of the process through which this essay was written. For example, in one step, I instructed the chatbot to interview me about my ideas. I wanted it to function as a test reader, asking questions that might expose gaps in my explanations or arguments. I don't have a human test reader or editor available, so I used the chatbot to help fill that gap. The interviews, the questions I pursued, the arguments I rejected, and the decisions about what belonged in the essay were all part of my authorial process.

The chatbot editorial assistant lacks an experienced human editor's nuanced ability to respond to critical issues in a text. On the other hand, because it can quickly evaluate the conceptual consequences of my code decisions, the chatbot was able to effectively contribute verbiage that discussed the critical implications of the code in a way that a human editor typically wouldn't, unless they were directly involved in the project. For artists and others who communicate critically through code and algorithms, the chatbot can provide a rare type of editorial support.

I'm not proposing that the chatbot contributed nothing authorial. Nor am I claiming that writing every word oneself is identical to working through this type of iterative process with an assistant — human or bot. I am not competing for Writer of the Year. I *am* trying to communicate an argument about my artistic practice, and I've developed an intentional process for doing so. I consider that authorship — not merely supervision. As both *Deep World* and this essay point out, the in-between spaces can be quite fruitful.

One final note: I've discovered that I use em dashes much more than the AI chatbot. So if you thought the em dashes were the smoking gun — surprise, most of that's me!