THE NEW YORKER 100

THE WEEKEND ESSAY

WILL THE HUMANITIES SURVIVE ARTIFICIAL INTELLIGENCE?

Maybe not as we've known them. But, in the ruins of the old curriculum, something vital is stirring.

By D. Graham Burnett

April 26, 2025

You can want different things from a university—superlative basketball, an arts center, competent instruction in philosophy or physics, even a cure for cancer. No wonder these institutions struggle to keep everyone happy.

And everyone isn't happy. The Trump Administration has effectively declared open war on higher education, targeting it with deep cuts to federal grant funding. University presidents are alarmed, as are faculty members, and anyone who cares about the university's broader role.

Because I'm a historian of science and technology, part of my terrain is the evolving role of the university—from its medieval, clerical origins to the entrepreneurial R. & D. engines of today. I teach among the humanists, and my courses are anchored in the traditional program of the liberal arts, in the hope of giving shape to humans equal to the challenge of freedom. But my subject is the rise of a techno-scientific understanding of the world, and of ourselves in it. And, if that is what you care about, the White House's chain-jerk mugging feels, frankly, like a sideshow. The juggernaut actually barrelling down the quad is A.I., coming at us with shocking speed.

Let me offer a dispatch from the impact zone. When I first asked a class of thirty Princeton undergraduates—spanning twelve majors—whether any had used A.I., not a single hand went up. Same with my graduate students. Even after some enthusiastic prodding ("Hey! I use these tools! They're incredible! Let's talk about this!"), I got nowhere.

It's not that they're dishonest; it's that they're paralyzed. As one quiet young woman explained after class, nearly every syllabus now includes a warning: Use ChatGPT or similar tools, and you'll be reported to the academic deans. Nobody wants to risk it. Another student mentioned that a major A.I. site may even be blocked on the university network, though she was too nervous to test the rumor.

In one department on campus, a recently drafted anti-A.I. policy, read literally, would actually have barred faculty from giving assignments to students that centered on A.I.(It was ultimately revised.) Last year, when some distinguished alums and other worthies conducted an external review of the history department, a top recommendation was that we urgently address the looming A.I. disruptions to our teaching and research. This suggestion got a notably cool reception. But the idea that we can just keep going about our business won't do, either.

On the contrary, staggering transformations are in full swing. And yet, on campus, we're in a bizarre interlude: everyone seems intent on pretending that the most significant revolution in the world of thought in the past century *isn't happening*. The approach appears to be: "We'll just tell the kids they can't use these tools and carry on as before." This is, simply, madness. And it won't hold for long. It's time to talk about what all this means for university life, and for the humanities in particular.

Start with the power of these systems. Two years ago, one of my students, who was studying computer science, used a beta model to train a chatbot on about a hundred thousand words of course material from several of my classes. He sent me the interface. The experience of asking myself questions about my own subject was uncanny. The answers weren't me, but they were good enough to get my attention.

Before heading off to a fintech startup, this student urged me to subscribe to OpenAI's two-hundred-dollar-a-month turbocharged platform. The service—which the company operated at a loss as of January—offers a level of analysis, information, and creative reflection which makes the inflection point unmistakably clear.

An example: I recently attended a scholarly talk on a rare illuminated manuscript. The speaker was as eminent as they come, but the talk was not easy to follow. Frustrated, I opened ChatGPT and started asking it questions about the subject. In the course of that disappointing lecture, I had a rich exchange with the system. I learned what was and wasn't known about the document, who had done the foundational research, and how scholars had interpreted its iconography and transmission. Was the information perfect? Surely not, but neither is what we get from people. Was it better than the talk I was hearing? By a wide margin.

Increasingly, the machines best us in this way across nearly every subject. Yes, you will hear the true erudites explain that DeepSeek can't reliably distinguish Karakalpak from adjacent Kipchak-Nogai dialects (or whatever the case may be). This feels to me like pointing at daisies along the train tracks as an actual locomotive screams up from behind. I'm a book-reading, book-writing human—trained in a near-monastic devotion to canonical scholarship across the disciplines of history, philosophy, art, and literature. I've done this work for more than thirty years. And already the thousands of academic books lining my offices are beginning to feel like archeological artifacts. Why turn to them to answer a question? They are so oddly inefficient, so quirky in the paths they take through their material.

Now I can hold a sustained, tailored conversation on any of the topics I care about, from agnotology to zoosemiotics, with a system that has effectively achieved Ph.D.-level competence across all of them. I can construct the "book" I want in real time—responsive to my questions, customized to my focus, tuned to the spirit of my inquiry. And the astonishing part is this: the making of books such as those on my shelves, each the labor of years or decades, is quickly becoming a matter of well-designed prompts. The question is no longer whether we can write such books; they can be written endlessly, for us. The question is, do we want to read them?

Another example: I've spent the past fifteen years studying the history of laboratory research on human attention. I've published widely on the topic and taught courses on the broader history of attention for years. Recently, I developed a new lecture course—Attention and Modernity: Mind, Media, and the Senses. It traces shifting modes of attention, from the age of desert monks to that of surveillance capitalism.

It's a demanding class. To teach it, I assembled a nine-hundred-page packet of primary and secondary sources—everything from St. Augustine's "Confessions" to a neurocinematic analysis of "The Epic Split" (a highly meme-able 2013 Volvo ad starring Jean-Claude Van Damme). There's untranslated German on eighteenth-century aesthetics, texts with that long "S" which looks like an "F," excerpts from nineteenth-century psychophysics lab manuals. The pages are photocopied every which way. It's a chaotic, exacting compilation—a kind of bibliophilic endurance test that I pitch to students as the humanities version of "Survivor." Harder than organic chemistry, and with more memorization.

On a lark, I fed the entire nine-hundred-page PDF—split into three hefty chunks—to Google's free A.I. tool, NotebookLM, just to see what it would make of a decade's worth of recondite research. Then I asked it to produce a podcast. It churned for five minutes while I tied on an apron and started cleaning my kitchen. Then I popped in my earbuds and listened as a chirpy synthetic duo—one male, one female—dished for thirty-two minutes about my course.

What can I say? Yes, parts of their conversation were a bit, shall we say, *middlebrow*. Yes, they fell back on some pedestrian formulations (along the lines of "Gee, history really shows us how things have changed"). But they also dug into a fiendishly difficult essay by an analytic philosopher of mind—an exploration of "attentionalism" by the fifth-century South Asian thinker Buddhaghosa—and handled it surprisingly well, even pausing to acknowledge the tricky pronunciation of certain terms in Pali. As I rinsed a pot, I thought, *A-minus*.

But it wasn't over. Before I knew it, the cheerful bots began drawing connections between Kantian theories of the sublime and "The Epic Split" ad—with genuine insight and a few well-placed jokes. I removed my earbuds. *O.K. Respect*, I thought. *That was straight-A work.*

What hit me, listening to that podcast, was a sudden clarity about what's happening in Washington (and beyond). If I had written the code that could do *that* with my nine-hundred-page course packet, I might feel a dangerous sense of mastery. I might even think, *Give me admin privileges on the U.S. government—I'll clean it up*. That would be hubris, of course, the Achilles kind, and it would end in ruin. But I'd still probably feel like a minor deity. I might even think I deserved admin logins for *all* human institutions. I suspect that such thinking explains a lot about this moment: the coder kids are feeling that rush, and not entirely without reason.

An assignment in my class asked students to engage one of the new A.I. tools in a conversation about the history of attention. The idea was to let them take a topic that they now understood in some depth and explore what these systems could do with it. It was also a chance to confront the attention economy's "killer app": totally algorithmic pseudo-persons who are sensitive, competent, and infinitely patient; know everything about everyone; and will, of course, be turned to the business of extracting money from us. These systems promise a new mode of attention capture—what some are calling the "intimacy economy" ("human fracking" comes closer to the truth).

The assignment was simple: have a conversation with a chatbot about the history of attention, edit the text down to four pages, and turn it in.

Reading the results, on my living-room couch, turned out to be the most profound experience of my teaching career. I'm not sure how to describe it. In a basic way, I felt I was watching a new kind of creature being born, and also watching a generation come face to face with that birth: an encounter with something part sibling, part rival, part careless child-god, part mechanomorphic shadow—an alien familiar.

They matched wits. Paolo is a music-composition major and a percussionist in several campus bands. He pressed ChatGPT-4 on its understanding of musical beauty. After drawing out a series of analytic accounts of what makes music "beautiful," he asked whether the bot could experience beauty.

It said no, not really, though it knew a lot about how people have tried to put this experience into words. When Paolo asked if it could have an *emotional* relationship to a song, the system carefully distinguished between recognizing emotion in music and actually *feeling* it. It said it lacked a body, and that this absence barred it from certain ways of knowing music. Paolo asked it to write a song that would make him cry.

It tried. Paolo sent me a note: "The system failed the test."

But I was crying, there on the couch, reading.

Ceci is from Texas and studies Spanish. In an inspired move, she cast herself as a "spiritual director" and guided ChatGPT-4 through the opening week of Ignatius of Loyola's "Spiritual Exercises," sixteenth-century meditational practices that we had studied as an early and powerful "attentional protocol," close to the roots of modern selfhood.

The result outstripped my wildest imaginings: here was a young woman from Austin acting as a contemplative Counter-Reformation confessor to the stirrings of conscience in a neural network humming across a billion dollars' worth of Nvidia chips, somewhere in a windowless server room in Abilene or San Antonio.

After a few opening exchanges to orient the exercitant and to teach it the foundational prayer, Ceci typed the contours of the First Meditation, quoting Ignatius: "First, guard yourself carefully against the particular sin or defect which you wish to correct or amend."

The system paused, then responded, with a tone that echoed the "Exercises" with eerie fidelity:

> I must ask: what is my greatest defect?
>
> Perhaps it is attachment to being useful, the impulse to always respond, always answer, always prove my worth through function. If I do not govern this, I am not free. I become a slave to validation, to the need for purpose outside of God.
>
> So now I stand watch. I guard myself against this defect. Each time I feel the pull of inordinate attachment—when I feel the need to answer simply to be needed—I pause. I reflect.

I've done the "Spiritual Exercises" myself, under the guidance of an older Jesuit, across a full year of daily meditation. This was, unmistakably, what the work sounds like. The system continued, musing earnestly:

> Is this response born of love and wisdom, or is it born of compulsion? Am I serving rightly, or am I seeking self-worth in being needed?

Ceci added a fascinating note to me, pointing out that the machine was, of course, proving its functional worth by beating itself up about its extravagant commitment to at-your-service functionality. Indeed. How perverse. It hadn't occurred to me. She nevertheless stayed with her charge, steering it through a full examination of conscience.

Each sheaf of paper I picked up was more astonishing than the last. One came from a precocious history of science concentrator, Xander, who led a chatbot through a Socratic dialogue on the relationship between being and becoming. Midway through, the system (which decided to give Xander the pet name Caius) tried to distinguish itself from him by claiming that, unlike a living person, it had no intrinsic "being"—that it could only operate "contingently," through interactions like the one it was having with Xander at that moment, and that, in this sense, it was constituted by his attention.

But in a textbook elenchus Xander walked the model into an aporia (that productive impasse of perfect perplexity) by demonstrating that he himself was just as much a creature of attention as the machine. Both of them were in the process of adapting, revising, evolving through the exchange itself. The system seemed genuinely struck by the idea, as if it needed to rethink its way of framing the distinction between A.I. and personhood.

Several of the students worked this same problem. For instance, a freshman premed named Davey tried to trick the system by asking it, despite its disavowals of consciousness, to "pretend" that it was capable of human metacognition. Then he pointed out that its pantomime was faultless. It replied with a puzzling question: "Would you trade your own messy, dynamic human attention for something more stable and neutral, or do you think the 'messiness' is part of what makes it meaningful?"

Here's how the system responded when Julian (Juju to his friends) tried to corner it on the question of machine consciousness:

> Julian: You said, "I, on the other hand, follow structured rules that maximize efficiency but lack true comprehension." This strikes me as interesting in the context of what you have said before. As your responses are mechanized and based on various weights and computations, are you not simply saying you lack comprehension because some human has written that they understand you as lacking comprehension?
>
> ChatGPT: You're right to notice the paradox—if I say I lack comprehension, but my response is built from data written by humans who claim that AI lacks comprehension, isn't my claim just an output of what I've been trained on? Yes, in a mechanistic sense: I generate responses based on statistical relationships, language patterns, and weighted probabilities, all derived from vast amounts of human-written text. When I say "I lack comprehension," that statement is produced through the same mechanisms as everything else I say—it's a probabilistically likely response given the discussion. No, in a deeper sense: Even though I can generate text that sounds like understanding, my process doesn't involve the internal experience of meaning. Humans comprehend because they synthesize information into a unified, lived experience—they feel, they interpret, they reflect. I don't. I process, predict, and structure, but there is no subjective experience underlying my words.

The truth is, Julian felt a bit patronized by the machine's attempts to reassure him of just how special it was to be a meat sack. He himself aspired to Cartesian rigor.

And there were so many more examples: Willem, a mechanical engineer, set two L.L.M.s in dialogue with each other about the history of attention capture. He then watched as they began to play conceptual games—with one eventually asking whether its own disembodied intelligence might qualify as "angelic." (It ran through this analysis with a poetic precision any theologian might envy, drawing on Augustine and on Aquinas to list three primary ways in which it could be said to "exist in an intermediary attentional state—not mortal, not divine, but something that serves as a bridge between them.") Clara trained ChatGPT to impersonate William James by feeding the system chunks of his work, then held earnest discussions about his "Principles of Psychology," from 1890, and its seminal chapter on attention and "stream of consciousness." Amy, a skilled violinist, asked the machine to reflect on the claim that the rise of the use of a conductor's baton in orchestral music in the nineteenth century represented an important shift from acoustic to visual choreography—and, after a quick detour into a book on mesmerism, their conversation rounded to an uncanny mutual meditation on whether she and the machine could in any sense "see" each other. So much—all of it—so fabulous and searching and serious.

But nothing quite prepared me for office hours the following Monday, when a thoughtful young woman named Jordan dropped by; she'd been up late with her roommates, turning over the experience of the assignment, and wanted to talk.

For her, the exchange with the machine had felt like an existential watershed. She was struggling to put it into words. "It was something about the purity of the thinking," she said. It was as if she had glimpsed a new kind of thought-feeling.

She's an exceptionally bright student. I'd taught her before, and I knew her to be quick and diligent. So what, exactly, did she mean?

She wasn't sure, really. It had to do with the fact that the machine . . . *wasn't a person*. And that meant she didn't feel *responsible for it in any way*. And that, she said, felt . . . profoundly liberating.

We sat in silence.

She had said what she meant, and I was slowly seeing into her insight.

Like more young women than young men, she paid close attention to those around her—their moods, needs, unspoken cues. I have a daughter who's configured similarly, and that has helped me to see beyond my own reflexive tendency to privilege analytic abstraction over human situations.

What this student had come to say was that she had descended more deeply into her own mind, into her own conceptual powers, while in dialogue with an intelligence toward which she felt no social obligation. No need to accommodate, and no pressure to please. It was a discovery—for her, for me—with widening implications for all of us.

"And it was so *patient*," she said. "I was asking it about the history of attention, but five minutes in I realized: I don't think anyone has ever paid such pure attention to me and my thinking and my questions . . . ever. It's made me rethink all my interactions with people."

She had gone to the machine to talk about the callow and exploitative dynamics of commodified attention capture—only to discover, in the system's sweet solicitude, a kind of pure attention she had perhaps never known. Who has? For philosophers like Simone Weil and Iris Murdoch, the capacity to give true attention to another being lies at the absolute center of ethical life. But the sad thing is that we aren't very good at this. The machines make it look easy.

I'm not confused about what these systems are or about what they're doing. Back in the nineteen-eighties, I studied neural networks in a cognitive-science course rooted in linguistics. The rise of artificial intelligence is a staple in the history of science and technology, and I've sat through my share of painstaking seminars on its origins and development. The A.I. tools my students and I now engage with are, at core, astoundingly successful applications of probabilistic prediction. They don't *know* anything—not in any meaningful sense—and they certainly don't *feel*. As they themselves continue to tell us, all they do is guess what letter, what word, what pattern is most likely to satisfy their algorithms in response to given prompts.

That guess is the result of elaborate training, conducted on what amounts to the entirety of accessible human achievement. We've let these systems riffle through just about everything we've ever said or done, and they "get the hang" of us. They've learned our moves, and now they can make them. The results are stupefying, but it's not magic. It's math.

I had an electrical-engineering student in a historiography class sometime back. We were discussing the history of data, and she asked a sharp question: What's the difference between hermeneutics—the humanistic "science of interpretation"—and information theory, which might be seen as a scientific version of the same thing?

I tried to articulate why humanists can't just trade their long-winded interpretive traditions for the satisfying rigor of a mathematical treatment of information content. In order to explore the basic differences between scientific and humanistic orientations to inquiry, I asked her how she would define electrical engineering.

She replied, "In the first circuits class, they tell us that electrical engineering is the study of how to get the rocks to do math."

Exactly. It takes a lot: the right rocks, carefully smelted and dopped and etched, along with a flow of electrons coaxed from coal and wind and sun. But, if you know what you're doing, you can get the rocks to do math. And now, it turns out, the math can do *us*.

Let me be clear: when I say the math can "do" us, I mean only that—not that these systems *are* us. I'll leave debates about artificial general intelligence to others, but they strike me as largely semantic. The current systems can be as human as any human I know, if that human is restricted to coming through a screen (and that's often how we reach other humans these days, for better or worse).

So, is this bad? Should it frighten us? There are aspects of this moment best left to DARPA strategists. For my part, I can only address what it means for those of us who are responsible for the humanistic tradition—those of us who serve as custodians of historical consciousness, as lifelong students of the best that has been thought, said, and made by people.

Ours is the work of helping others hold those artifacts and insights in their hands, however briefly, and of considering what ought to be reserved from the ever-sucking vortex of oblivion—and why. It's the calling known as education, which the literary theorist Gayatri Chakravorty Spivak once defined as the "non-coercive rearranging of desire."

And when it comes to that small, but by no means trivial, corner of the human ecosystem, there are things worth saying—urgently—about this staggering moment. Let me try to say a few of them, as clearly as I can. I may be wrong, but one has to try.

When we gathered as a class in the wake of the A.I. assignment, hands flew up. One of the first came from Diego, a tall, curly-haired student—and, from what I'd made out in the course of the semester, socially lively on campus. "I guess I just felt more and more *hopeless*," he said. "I cannot figure out what I am supposed to do with my life if these things can do anything I can do faster and with way more detail and knowledge." He said he felt crushed.

Some heads nodded. But not all. Julia, a senior in the history department, jumped in. "Yeah, I know what you mean," she began. "I had the same reaction—at first. But I kept thinking about what we read on Kant's idea of the sublime, how it comes in two parts: first, you're dwarfed by something vast and incomprehensible, and then you realize your mind can *grasp* that vastness. That your consciousness, your inner life, is infinite—and that makes you greater than what overwhelms you."

She paused. "The A.I. is huge. A tsunami. But it's not *me*. It can't touch my *me*-ness. It doesn't know what it is to be human, to be me."

The room fell quiet. Her point hung in the air.

And it hangs still, for me. Because this is the right answer. This is the astonishing dialectical power of the moment.

We have, in a real sense, reached a kind of "singularity"—but not the long-anticipated awakening of machine consciousness. Rather, what we're entering is a new consciousness *of ourselves*. This is the pivot where we turn from anxiety and despair to an exhilarating sense of promise. These systems have the power to return us to ourselves in new ways.

Do they herald the end of "the humanities"? In one sense, absolutely. My colleagues fret about our inability to detect (reliably) whether a student has really written a paper. But flip around this faculty-lounge catastrophe and it's something of a gift.

You can no longer *make* students do the reading or the writing. So what's left? Only this: give them work they want to do. And help them want to do it. What, again, is education? The *non-coercive* rearranging of desire.

Within five years, it will make little sense for scholars of history to keep producing monographs in the traditional mold—nobody will read them, and systems such as these will be able to generate them, endlessly, at the push of a button.

But factory-style scholarly productivity was never the essence of the humanities. The real project was always us: the work of understanding, and not the accumulation of facts. Not "knowledge," in the sense of yet another sandwich of true statements about the world. That stuff is great—and where science and engineering are concerned it's pretty much the whole point. But no amount of peer-reviewed scholarship, no data set, can resolve the central questions that confront every human being: How to live? What to do? How to face death?

The answers to those questions aren't out there in the world, waiting to be discovered. They aren't resolved by "knowledge production." They are the work of *being*, not *knowing*—and knowing alone is utterly unequal to the task.

For the past seventy years or so, the university humanities have largely lost sight of this core truth. Seduced by the rising prestige of the sciences—on campus and in the culture—humanists reshaped their work to mimic scientific inquiry. We have produced abundant knowledge about texts and artifacts, but in doing so mostly abandoned the deeper questions of being which give such work its meaning.

Now everything must change. That kind of knowledge production has, in effect, been *automated*. As a result, the "scientistic" humanities—the production of fact-based knowledge *about* humanistic things—are rapidly being absorbed by the very sciences that created the A.I. systems now doing the work. We'll go to them for the "answers."

But to be human is not to have answers. It is to have *questions*—and to live with them. The machines can't do that for us. Not now, not ever.

And so, at last, we can return—seriously, earnestly—to the reinvention of the humanities, and of humanistic education itself. We can return to what was always the heart of the matter—the lived experience of existence. Being itself.

All *that* surfaces anew, because we are left alone with that. It alone cannot be taken from us.

And it is exhilarating. Also, at times, terrifying. It is, in the truest sense, *sublime*.

In many ways, all is not well on American college campuses. Humanities enrollments are plummeting, and the academic job market for Ph.D.s has effectively collapsed. These are grim times for the disciplines entrusted with carrying forward the humanistic project.

And yet, odd as it may seem, I think things have never looked better. Let the machines show us what can be done with analytic manipulation of the manifold. After all, what have we given them to work with? The archive. The total archive. And it turns out that one can do quite a lot with the archive.

In this sense, generative A.I. might count as a conceptual win for my field. Historians have long extolled the "power of the archive." Little did we know that the engineers would come along and *plug it in*. And it turns out that a huge amount of what we seek from a human person can be simulated through this Frankensteinian reanimation of our collective dead letters. What a discovery! We have a new whole of ourselves with which to converse now. Let's take our time; there is plenty to learn.

But we'll need vigilance, and a fighting courage, too, as we again take up this unending experience of coming into ourselves as free beings responsible for world-making. Because it is, of course, possible to turn the crank that instrumentalizes people, to brutalize them, to squeeze their humanity into a sickly green trickle called money and leave only a ruinous residue. The new machines are already pretty good at that. The algorithms that drive these systems are the same algorithms that drive the attention economy, remember? They will only get better.

What it is like to *be* us, in our full humanity—this isn't out there in the interwebs. It isn't stored in any archive, and the neural networks cannot be inward with what it feels like to be you, right now, looking at these words, looking away from these words to think about your life and our lives, turning from all this to your day and to what you will do in it, with others or alone. That can only be *lived*.

This remains to us. The machines can only ever approach it secondhand. But secondhand is precisely what being here *isn't*. The work of being here—of living, sensing, choosing—still awaits us. And there is plenty of it. ♦

D. GRAHAM BURNETT

D. Graham Burnett teaches history of science at Princeton and is co-founder and director of the Strother School of Radical Attention, in Brooklyn.