

Humanities and computational analysis

Interview with Mark Algee-Hewitt

By Bruno Cousin

While we have entered the age of generative AI, it is worth having a look at how computers can help literature scholars and intellectual historians to explain the life of concepts, aesthetics and genres.

Mark Algee-Hewitt is an Associate Professor of Digital Humanities in the English Department at Stanford University. The director of both Stanford's Center for Spatial and Textual Analysis (CESTA) and the Stanford Literary Lab, his work applies computational methodologies to the study of cultural and aesthetic objects, focusing particularly on patterns of language in large corpora of texts. Trained primarily in the analysis of eighteenth and nineteenth-century literature in English and German, as well as in computer science, his research has grown to investigate the covert patterns of language use in the contemporary media ecology that governs our world.

Algee-Hewitt has published widely in humanities and literary studies journals, both as a solo author and as a collaborator on large scale projects. At the Literary Lab, he has led research into such diverse topics as the language that governs the experience of suspense for novel readers, the relationship between authors and writers in online fan fiction communities, the ability of climate fiction to teach real-world environmental facts to readers through fictional narratives, and the use of racialized language in American literature. His book *The Afterlife of Aesthetics*, forthcoming, explores the rise of aesthetic concepts, particularly the sublime, in the 18th century, identifying persistent patterns that govern its usage that remain even after the concepts have fallen out of fashion.

As a fellow at CASBS, Mark Algee-Hewitt is working on a new project that uses semantic distance to trace the development of complex compound concepts, such as “human rights” or “political economy” across the eighteenth and nineteenth-centuries. While at the center, he has developed a novel visual method that allows researchers to identify both configurations of words that represent successful ideas and, more importantly, those that distinguish failed concepts. By studying these points of conceptual failure, Algee-Hewitt argues we can better understand why some ideas succeed and why some are relegated to the dustbin of history.

Books & Ideas: Why study philosophical and literary texts through computational analysis?

Mark Algee-Hewitt: There are a number of reasons why I find it personally helpful to use computational analysis to understand text in general, particularly literary texts, but also philosophic ones. But broadly, it's because it lets me ask different kinds of questions that I could ask through my training as a literary critic or as an intellectual historian. And this happens in a number of different ways.

For example, one of the most common ways that we talk about this has to do with scale. The easy way to understand that is that computational analysis lets us ingest and analyze texts at much greater scales than we ever could as readers, not just tens of texts, but hundreds, thousands, tens of thousands, even millions of texts.

We can look at a single phenomenon over the course of that many texts, and really come to a much broader understanding of what's going on in, for example, the literary field in a particular century, in an intellectual, historical domain. And that is a revolutionary way of approaching literature. It's a revolutionary way of approaching text in general.

It gives us a whole new kind of evidence that we can use. Now, it's often overlooked that the scale thing happens at the opposite end too, instead of just many, many texts, we can use computational analysis to pay attention to things that fall well below the threshold that we as readers are trained to notice or even understand.

So, for example, I can pick an amount of my favorite novels and I can tell you what it's about, and maybe I've read it three, four, ten, twenty times. But if you ask me how many times the word « the » appeared in the novel, I would have absolutely no idea. And you may say that, why would you? Why is that remotely important?

It actually turns out that it kind of *is*. One of the ways in which computational analysis is used is in what's called stylometry, and authorship attribution. And it turns out you can tell who wrote a text by the number of times the words *the, a, I, he, she* – all of those really common function words – appear in it.

And again, these are things that we are not trained to recognize as readers. And so computational analysis lets us pay attention to these really, really minute scales in ways that we couldn't in any other way. And so between the really, really big and the really, really small, this gives us a whole new domain through which to approach the study of text, and again, for me, the study of literature.

And so I can understand a phenomenon like the sublime. I can understand the evolution of a concept across a couple of centuries in a way that I could not, simply by using a kind of critical reading-based approach. Now, that's not to say that critical reading is no longer important. It's absolutely essential.

It's how I come up with the questions that I want to ask. It's how I interpret the results from the computational analysis that I pull off. So that's still really central. But what computational analysis brings to the party, and I want to differentiate, it's not just computational analysis, it's also quantitative analysis. It's turning text into numbers that then becomes computationally tractable and then traceable at these different scales, at the macroscopic and microscopic scales. By doing all of that, I can ask questions like : what happened to the sublime?, in a way that I couldn't through any other kind of method or approach that we as literary critics have tried out.

In many ways, this is kind of a fever dream of Russian formalism. Early 20th century literary criticism dreamed of us being able to come up with these different kinds of questions through a mathematical or a quantitative approach to literature; digital humanities, and computational literary studies in particular, is finally letting us get at those kinds of questions in a way that we just never have been able to before.

Books and Ideas: Can you tell us about your current research on failed concepts?

Mark Algee-Hewitt: I'm really interested at the moment in studying what I'm calling failed concepts. So I talked about political revolution and measuring how those two terms evolved over the course of the 18th century and eventually merged together into a concept that unites them into a single thing. That's what I would call a successful

concept. You start off with these two different tracks of intellectual investment, and over time they merge together.

They assume more and more of each other's identity until they become a single thing: a complex-compound concept that is, as John Locke describes them. What I'm interested in is when that doesn't work, because I think by studying failed concepts, we can actually think about why the ones that were successful were successful. What did they have that the others didn't?

So, for example, moral inadequacy is one of my favorite examples of a failed concept that I found. And again, this is only because I can measure semantic distance over time by using these computational methods in the surface of intellectual history. I can trace these two apparently – from our perspective – unrelated words. And see, actually, around the middle of the 18th century, they started to come together.

They were a thing: moral inadequacy. Could we judge the ethics of an action based on a kind of simple accounting of how much good it did? And it looked for a little bit like we could. But then, if you trace the semantic distance around 1750, that falls apart again. It was coming together and then all of a sudden they diverge and there's a number of different explanations of what happened there.

Kantian, critical or categorical imperative comes into play where, all of a sudden, we can understand the morality of an action: not by looking up the good it does, but by a kind of ontological « Is this good if applied to every case? », for example. And so these kinds of underpinning philosophic, social, even technological changes that happened towards the middle to end of the 18th century drive a lot of these conceptual evolutions, such that the ones that fail really give us a better understanding of why they fail, and therefore an even better understanding of why some of them succeed.

Books and Ideas: What does the evolution of the category of the sublime tell us about the transformations of the literary field?

Mark Algee-Hewitt: In a very real sense – and this is the argument that I make throughout the course of this project – the sublime gives us the literary field in the 19th century in our modern understanding of what literature is. That's the big controversial claim. And what I mean is that the sublime becomes a critical category on which the differentiation of different kinds of writing hinges: something that separates out what

we might consider to be literature, the kind of culturally emulative writing that everyone wants us to be thinking about, that we teach in schools, that we encourage people to read for their betterment... and the mere entertainment, the drivel that gets published, in reams and reams, that that is somehow culturally destructive, the things that we don't want our children coming in contact with. And of course, in the 19th century, without television, without video games, without film, this differentiation happened in the realm of literature. And the sublime has always interested me because the sublime is this ultimate figure of unrepresentability.

According to Kant, for example, the sublime is what happens when the imagination fails and the understanding fails and reason has to come into the rescue. In other words, it's the way that we describe something that we can't describe, but we can describe our inability to describe it and therefore get at something essential about it.

At the beginning of the 18th century, there's a really key technological change, and that has to do with what's called, in literary studies, in history, in media studies particularly, the explosion of print.

Now, print is a technology that's been around for quite a while at this point, a couple of centuries, but it's still relatively rare, that is, more and more is being printed, but it's really in the 18th century, if you graph the number of works that are published, you get that kind of hockey-stick effect that just really takes off, exponentially. And one of the ways that we can think about this kind of mechanism of differentiation is that it used to rest on that differentiation of technology: things that were printed, because printing is expensive, because bookbinding is expensive, because books are beautiful and leather-bound, and only the very, very rich can afford them... Or lending libraries, for example, which come into being, a little bit earlier than the period I'm looking at, that does a lot of work in differentiating the good kind of things, because those get printed, and the bad kinds of things, because those either don't get printed, they are circulated in manuscript form, or they're relatively cheaply printed and circulated as broadsides and newspapers.

But as printing becomes more and more accessible to more and more people, because as the technology evolves and more and more is printed, it can be done at greater and greater volumes, all of a sudden, that doesn't work anymore. The streets are flooded in the way that Dryden described it, with cheap novels, with cheap literature, with cheap tracks and books.

Then people don't have the old ways of understanding, you know, this is what should be read and this is what shouldn't be read. And so a lot of the 18th century is kind of the search for how do we make this differentiation: how do we tell the good stuff from the bad stuff. And this is what eventually turns into the discipline of criticism, the discipline of literary study.

This is its genesis. This is originally what it's designed to do. How do you tell the good stuff from the bad stuff? And in tracing the sublime, which is the key aesthetic concept of the 18th century, it takes off in the same way that print does, like nothing else. People talk about the beautiful, quite a bit in the 17th century, and they continue to talk about the beautiful quite a bit by the end of the 18th century.

But the sublime goes from rarely used in the 17th century to one third of every book published in Britain in 1798 using the word at least once: it's everywhere. And so this project is really trying to grapple with: where does it come from? Why does it get so popular? And then, where does it go? What does it do, as this kind of aesthetic concept that exists into the early 19th century and then kind of dissolves? And the argument that I make is that it dissolves into criticism, it dissolves into this attempt to anchor ameliorative good writing literature from the bad stuff. And that's what the technology of the sublime does. The good literature represents this kind of unrepresentability, using the techniques of the aesthetic sublime, using the techniques that give this kind of Aristotelian catharsis to the reader, that teaches them about the unrepresentable and makes them feel this incredible passion or transformation - *transcendence* really is the word that is usually used to describe it. Whereas, you know, bad literature might give you some cheap thrills, but it's not truly sublime.

And so by tracing the evolution of this aesthetic concept, how it grows, how it changes, how it embeds itself in different kinds of writing, in different kinds of contexts, and looking at what it keeps as it moves into the 19th century, this attachment to poetics, this attachment to representability, but also this attachment to affect, to transcendence... and then looking at what it sheds, a lot of its religious connotations and a lot of its rhetorical connotations, you can really see how that becomes the anchoring concept of literary criticism and by extension, the anchoring concept of literature in the modern sense itself.

Books and Ideas: Switching now to another of your research projects, how do novels about climate change weave together scientific accuracy and fiction without disinforming the reader?

Mark Algee-Hewitt: The key there is understanding what fiction is, and this is one of those concepts that really is embedded in a literary critical understanding in a different way than is embedded in the public consciousness, because we tend to use *fiction* casually in terms of talking about things that aren't true. And this is where this project runs into a really interesting point of differentiation, because I'm really interested in fictional representations of real-world facts, as opposed to climate disinformation, which is everywhere, which I would call real representations of fictional facts.

And in fact, not all of the novels are doing that real fact, fictional representation thing. For example, Michael Crichton's *State of Fear* is one of the rare examples of an anti-climate change novel. In other words, Crichton's goal is to try and convince us that climate change isn't happening, or at least man-made anthropogenic climate change isn't happening.

And so he puts all the resources of fiction to work in this opposite way. Although most of the novels that we look at actually do represent climate change as anthropogenic and try and teach readers real-world facts using fictional worlds, and again, the key there is the fictional world, because what the fictional world allows the narrator of the book, and by extension, the author of the book to do, is to slot those facts in the text in such a way that it becomes easier for readers to absorb them.

And maybe not just absorb, but maybe also believe them, to internalize them in ways that don't necessarily put people's guards up, particularly with such a fraught subject as climate change, when so many people have already very stated, very strong opinions about the subject. And in fact, we can look at how different books use these facts and embed them in their narratives differently, in terms of how successful they are at getting around those kinds of guardrails that people have, based on their preconceived notion of that concept.

So, for example, one of the ways in which this happens is through the mechanism of world building. And you can see this in popular media all the time. You've got a young, naive protagonist, who comes to an older, wiser mentor at some point. And that older, wiser mentor teaches the young protagonist, and – by extension – the reader, the watcher, the audience, all the rules of the world that they inhabit.

So, for example, in the Star Wars movies, in the very first movie, Luke Skywalker goes to Obi-Wan Kenobi. He's a young, naive farm kid. Obi-Wan Kenobi teaches him what the Force is, what the Empire is, what the Rebellion is, and, by extension, we as the audience learn all of this along with the protagonist. And that's a really powerful narrative mechanism.

And you can see that being used in many of these climate-change novels, which start out with a naive protagonist that comes to an older mentor figure who teaches that protagonist how the world is changing, how climate change has affected this and that, and how it's all related to these series of man-made activities. That's really a kind of obvious example of the way this kind of mechanism works.

But by embedding it within these narrative conventions that we recognize, that we as readers, as watchers, as consumers of media understand how to read, how to interpret, and how to internalize... By embedding it in that technique, it can teach us these facts without necessarily being preachy about it, without creating, for example, an appendix, of just straightforward, dry facts that we're expected to read as a prerequisite for understanding what's happening in the book. And that's what I mean by this understanding of what fiction is; because it's not simply the creation of something that's untrue.

You can think of fiction as a kind of thought experiment. It lets authors, and so many of these books that we're studying are speculative fiction, science-fiction, etc., so it lets the authors of these speculative fictions create this playground where they can spin out the consequences of the facts that we have received today. So, in fact, I would suggest, based on the work that we've done and the modelling that we've done of these novels, that the most successful novels, the ones that are most able to teach readers about climate change, the ones that are most able to successfully convey these facts, are the ones that don't set their worlds in the present.

They don't try and teach us what's going on right now. They work towards the future. They work towards some kind of counterfactual world, some kind of alterity, some fantastic world, where they can play out the inevitable consequences, as they see, of the real-world climate situation that we find ourselves in today. So in that project, when we compare something like Michael Crichton's *State of Fear*, which is set very much in the present and is very concerned with dry recitations of facts... Crichton doesn't use these kinds of narrative techniques ; instead, he has characters get up and

sort of just like monologue about how climate change is all made up and environmental terrorism is ruining the world. It's not really the most engaging narrative. And you compare that to something like N.K. Jemisin's *Broken Earth* trilogy, which is really more of a science fantasy than science-fiction. It's set in a completely different world where the laws of physics operate somewhat differently.

But what she is really expert at is taking a lot of the climatological situation that we're in now, these kinds of facts about the climate, and spinning it out in this world that she's created, such that we can see the consequences, such that we can see what is happening. But because it's so displaced, because it's not set within our world, we don't immediately register that as being about our situation.

That sort of comes in later, and that's a much more effective way to teach audiences, to teach readers these facts, than the kind of recitation that happens in the less successful novels.

Books and Ideas: Can digital humanities researchers help generative AI developers?

Mark Algee-Hewitt: The question of AI, or large language models as the technology of current generation AI is based on, is one that's actually kind of fraught within the digital humanities world. This is technology that we've been using in a very rudimentary form for a number of years: word embeddings, language models, the underlying technology to current generative AI engines, is exactly, for example, in my project on the evolving concepts, which lets me measure semantic distance based on contextual similarity.

Now, contemporary AI models do that in a much more sophisticated way, but the underlying model is still relatively compatible. And yet what's happened in the AI world, particularly since the advent of ChatGPT and the way it took the public by storm three or four years ago, is that so much energy has focused on expanding the models, giving them more and more training data to use, because that has proven to really be a great return on investment. These models get much more sophisticated the more data that you give them. But the problem that we're running into, is that not all of that data is good. More for a while was just better. But it turns out that that hits a lot of diminishing returns.

And that's where a lot of the cutting-edge research within digital humanities is happening right now. Instead of simply taking one of these AI models and training it by feeding everything into it, what if we took a slightly more sophisticated approach to the training data? What if we fed it particular domains of data in order to create a specialized model that was really, really good at one thing, or that new one thing?

So, for example, in my world, there's a lot of interest around what are called historical models. One of the challenges of using AI models, for example ChatGPT, to do historical research is that they can't forget. So you could ask a model about things that happened in the 19th century. And overall it will be right. But if you keep probing it, for example, on the moon and the way it was understood in the 19th century, at some point it will sneak in something that came from the Apollo lunar landings, because that's just in its training data and it doesn't have a good way of differentiating : oh, wait, in 1899, people didn't know what the moon was like because we hadn't been there.

And so there's really interest in training historical models, which is just training the models based on historical data so that we could simulate a reader or a thinker from 1899. Now, the challenge there is that it takes so much data to train these models that we don't have the volume of historical texts that we do of contemporary texts.

And so a lot of the cutting-edge research around how to do this training work has to do with whether or not we can more intelligently train a model with much more cleverly chosen but highly selective texts, such as what might be available for historical models. Or do we take a contemporary model and try and fine-tune it using historical data?

This is still an unresolved question. But the fact that we are asking that question, the fact that we're experimenting with both of these things, has a lot to do with the ways in which AI researchers and AI developers are actually collaborating with digital humanities researchers, who are the ones asking these kinds of questions and who want to use generative AI as research tools, because they represent a kind of untapped pool of naive readers.

So much of the operationalization that goes on in the digital humanities has to do with annotating texts. So I have a project on identifying domestic space in novel, seeing if we can train some kind of model, to tell whether or not a scene is set within a domestic space or not. And to do that, we had to tag thousands and thousands of

example passages as domestic or not, feed it into a machine learning model, which then actually did a very good job of being able to differentiate it.

That took us probably the better part of a year. Now, imagine if we could just give all of those passages to ChatGPT, have it annotate it, and then use that to build the machine learning model that we wanted to make in the first place. That would be amazing, it would revolutionize the kind of work that we can do. But it's not quite there yet. And that gap between what it's capable of now, and what it might be capable of in two, three, five, or even ten years in the future: a lot of that research is happening not just by shovelling more and more into the model, but instead by thinking cleverly and thinking critically about how we train the models, how to teach the models to be better readers, and then, by extension, how to be better writers.

And that is coming a lot from the kind of work that my field is doing. A number of collaborations have sprung up around the country and around the world, between digital humanities researchers and developers of AI, in the hopes that we can use the kinds of techniques that we've refined and honed from doing computational studies of literature, of history, of geography, etc., of texts in general, and these new AI models that we want to be slightly more clever as they move towards more agential or research-oriented purposes.

Published in *booksandideas*, 3 October 2025.