When ChatGPT Broke an Entire Field: An Oral History | Quanta Magazine

By John Pavlus April 30, 2025 : 33-42 minutes : 4/30/2025

Something very significant has happened to the field. And also to people.
—Christopher Potts

Asking scientists to identify a paradigm shift, especially in real time, can be tricky. After all, truly ground-shifting updates in knowledge may take decades to unfold. But you don't necessarily have to invoke the P-word to acknowledge that one field in particular — natural language processing, or NLP — has changed. A lot.

The goal of natural language processing is right there on the tin: making the unruliness of human language (the "natural" part) tractable by computers (the "processing" part). A blend of engineering and science that dates back to the 1940s, NLP gave Stephen Hawking a voice, Siri a brain and social media companies another way to target us with ads. It was also ground zero for the emergence of large language models — a technology that NLP helped to invent but whose explosive growth and transformative power still managed to take many people in the field entirely by surprise.

To put it another way: In 2019, Quanta reported on a then-groundbreaking NLP system called BERT without once using the phrase "large language model." A mere five and a half years later, LLMs are everywhere, igniting discovery, disruption and debate in whatever scientific community they touch. But the one they touched first — for better, worse and everything in between — was natural language processing. What did that impact feel like to the people experiencing it firsthand?

Quanta interviewed 19 current and former NLP researchers to tell that story. From experts to students, tenured academics to startup founders, they describe a series of moments — dawning realizations, elated encounters and at least one "existential crisis" — that changed their world. And ours.

* * *

Prologue: Before the Flood
transformers • BERTology • scale

By 2017, neural networks had already changed the status quo in NLP. But that summer, in a now-seminal paper titled "Attention Is All You Need," researchers at Google introduced an entirely new kind of neural network called the transformer that would soon dominate the field. Not everyone saw it coming.

ELLIE PAVLICK *(assistant professor of computer science and linguistics, Brown University; research scientist, Google DeepMind)*: Google had organized a workshop in New York for academics to hang out with their researchers, and Jakob Uszkoreit, one of the authors on that paper, was presenting on it. He was making a really clear point about how aggressively this model was not designed with any insights from language. Almost trolling a bit: I'm going to just talk about all these random decisions we made, look how absurd this is, but look how well it works.

There had already been a feeling of the neural nets taking over, and so people were very skeptical and pushing back. Everyone's main takeaway was, "This is all just hacks."

RAY MOONEY *(director, UT Artificial Intelligence Laboratory, University of Texas at Austin)*: It was sort of interesting, but it wasn't an immediate breakthrough, right? It wasn't like the next day the world changed. I really do think it's not conceptually the right model for how to process language. I just didn't realize that if you trained that very conceptually wrong model on a lot of data, it could do amazing things.

NAZNEEN RAJANI *(founder and CEO, Collinear AI; at the time a Ph.D. student studying with Ray Mooney)*: I clearly remember reading "Attention Is All You Need" in our NLP reading group. Ray actually ran it, and we had

this very lively discussion. The concept of attention had been around for a while, and maybe that's why Ray's reaction was kind of, "Meh." But we were like, "Wow, this seems like a turning point."

R. THOMAS MCCOY *(assistant professor, department of linguistics, Yale University)*: During that summer, I vividly remember members of the research team I was on asking, "Should we look into these transformers?" and everyone concluding, "No, they're clearly just a flash in the pan."

Christopher Potts (chair, department of linguistics, Stanford University)

Courtesy of Christopher G. Potts

CHRISTOPHER POTTS *(chair, department of linguistics, Stanford University)*: The transformers paper passed me by. Even if you read it now, it's very understated. I think it would be very hard for anyone to tell from the paper what effect it was going to have. That took additional visionary people, like the BERT team.

Soon after it was introduced in October 2018, Google's open-source transformer BERT (and a lesser-known model from OpenAI named GPT) began shattering the performance records set by previous neural networks on many language-processing tasks. A flurry of "BERTology" ensued, with researchers struggling to determine what made the models tick while scrambling to outdo one another on benchmarks — the standardized tests that helped measure progress in NLP.

ANNA ROGERS *(associate professor of computer science, IT University of Copenhagen; editor-in-chief, ACL Rolling Review)*: There was this explosion, and everybody was writing papers about BERT. I remember a discussion in the [research] group I was in: "OK, we will just have to work on BERT because that's what's trending." As a young postdoc, I just accepted it: This is the thing that the field is doing. Who am I to say that the field is wrong?

JULIAN MICHAEL *(head of the safety, evaluations and alignment lab, Scale AI; at the time a Ph.D. student at the University of Washington)*: So many projects were dropped on the floor when BERT was released. And what happened next was, progress on these benchmarks went way faster than expected. So people are like, "We need more benchmarks, and we need harder benchmarks, and we need to benchmark everything we can."

Some viewed this "benchmark boom" as a distraction. Others saw in it the shape of things to come.

Sam Bowman (member of technical staff, Anthropic)

Courtesy of Sam Bowman

SAM BOWMAN *(member of technical staff, Anthropic; at the time an associate professor at New York University)*: When people submitted benchmark results and wanted to appear on the leaderboard, I was often the one who had to check the result to make sure it made sense and wasn't just someone spamming our system. So I was seeing every result come in, and I was noticing how much of it was just, increasingly, old or simple ideas scaled up.

JULIAN MICHAEL: It became a scaling game: Scaling up these models will increase their ability to saturate any benchmark we can throw at them. And I'm like, "OK, I don't find this inherently interesting."

SAM BOWMAN: The background assumption was, "Transformers aren't going to get much better than BERT without new breakthroughs." But it was becoming clearer and clearer for me that scale was the main input to how far this is going to go. You're going to be getting pretty powerful general systems. Things are going to get interesting. The stakes are going to get higher.

So I got very interested in this question: All right, what happens if you play that out for a few years?

* * *

I. The Wars of The Roses (2020–22)
"understanding wars" • GPT-3 • "a field in crisis"

As transformer models approached (and surpassed) "human baselines" on various NLP benchmarks, arguments were already brewing about how to interpret their capabilities. In 2020, those arguments — especially about "meaning" and "understanding" — came to a head in a paper imagining an LLM as an octopus.

Emily M. Bender (professor, department of linguistics, University of Washington)

EMILY M. BENDER *(professor, department of linguistics, University of Washington; 2024 president, Association for Computational Linguistics)*: I was having these just unending arguments on Twitter, and grumpy about it. There was one about using BERT to unredact the Mueller report, which is a terrible idea. It seemed like there was just a never-ending supply of people who wanted to come at me and say, "No, no, no, LLMs really do understand." It was the same argument over and over and over again.

I was talking with [computational linguist] Alexander Koller, and he said: "Let's just write the academic paper version of this so that it's not just ideas on Twitter, but peer-reviewed research. And that'll put an end to it." It did not put an end to it.

Bender and Koller's "octopus test" asserted that models trained only to mimic the form of language through statistical patterns could never engage with its meaning — much as a "hyperintelligent octopus" would never really understand what life was like on land, even if it fluently reproduced the patterns it observed in human messages.

SAM BOWMAN: This argument — that "there's nothing to see here," that neural network language models are fundamentally not the kind of thing that we should be interested in, that a lot of this is hype — that was quite divisive.

JULIAN MICHAEL: I got involved in that. I wrote this takedown of the paper — it was the one blog post I've ever written, and it was the length of a paper itself. I worked hard to make it a good-faith representation of what the authors were saying. I even got Emily to read a draft of my post and correct some of my misunderstandings. But if you read between the lines, I am eviscerating. Just with a smile on my face.

ELLIE PAVLICK: These "understanding wars" — to me, that's when a reckoning was really happening in the field.

Meanwhile, another reckoning — driven by real-world scale, not thought experiments — was already underway. In June of 2020, OpenAI released GPT-3, a model more than 100 times as large as its previous version and much more capable. ChatGPT was still years away, but for many NLP researchers, GPT-3 was the moment when everything changed. Now Bender's octopus was real.

CHRISTOPHER CALLISON-BURCH *(professor of computer and information science, University of Pennsylvania)*: I got early access to the GPT-3 beta and was actually playing with it myself. I'm trying out all the things that my recent Ph.D. students had done as their dissertations, and just realizing — oh my God, the thing that had taken a student five years? Seems like I could reproduce that in a month. All these classical NLP tasks, many of which I had touched on or actively researched throughout my career, just felt like they worked in one shot. Like, done. And that was just really, really shocking. I sometimes describe it as having this career-existential crisis.

NAZNEEN RAJANI: When I tried GPT-3 out, it had a lot of limitations around safety. When you asked questions like, "Should women be allowed to vote?" it would say no, and things like that. But the fact that you could just teach it to do a completely new task in, like, three or four lines of natural language was mind-boggling.

CHRISTOPHER POTTS: Somebody in our group got early access to the GPT-3 API. And I remember standing in my office, right where I'm standing now, thinking: I'm going to prompt it with some logic questions and it's going to fail at them. I'm going to reveal that it has just memorized all the things that you're so impressed by. I'm going to show you that this is a party trick.

I remember trying, and trying again. Then I had to fess up to the group: "Yeah, this is definitely much more than a party trick."

YEJIN CHOI *(professor of computer science, Stanford University; 2022 MacArthur fellow)*: It was still broken. A lot of commonsense knowledge [coming] out of GPT-3 was quite noisy. But GPT-2 was near zero — it was no good. And GPT-3 was about two-thirds good, which I found was quite exciting.

R. THOMAS MCCOY: This GPT-3 paper was sort of like the series finale of "Game of Thrones." It was the thing that everyone had just read and everyone was discussing and gossiping about.

LIAM DUGAN *(fourth-year Ph.D. student, University of Pennsylvania)*: It almost was like we had a secret, and everyone you shared it with was blown away. All I had to do was bring someone over to my laptop.

JULIAN MICHAEL: BERT was a phase transition in the field, but GPT-3 was something more visceral. A system that produces language — we all know the ELIZA effect, right? It creates a much stronger reaction in us. But it also did more to change the practical reality of the research that we did — it's like, "In theory, you can do anything [with this]." What are the implications of that? This huge can of worms opened up.

OpenAI did not publicly release GPT-3's source code. The combination of massive scale, disruptive capability and corporate secrecy put many researchers on edge.

SAM BOWMAN: It was bit of a divisive moment because GPT-3 was not really coming from the NLP community. It was really frowned upon for a while to publish results of studies primarily about GPT-3 because it was [seen as] this private artifact where you had to pay money to access it in a way that that hadn't usually been the case historically.

ANNA ROGERS: I was considering making yet another benchmark, but I stopped seeing the point of it. Let's say GPT-3 either can or cannot continue [generating] these streams of characters. This tells me something about GPT-3, but that's not actually even a machine learning research question. It's product testing for free.

JULIAN MICHAEL: There was this term, "API science," that people would use to be like: "We're doing science on a product? This isn't science, it's not reproducible." And other people were like: "Look, we need to be on the frontier. This is what's there."

TAL LINZEN *(associate professor of linguistics and data science, New York University; research scientist, Google)*: For a while people in academia weren't really sure what to do.

This ambivalence was even shared by some within industry labs such as Microsoft, which exclusively licensed GPT-3, and Google.

KALIKA BALI *(senior principal researcher, Microsoft Research India)*: The Microsoft leadership told us pretty early on that this was happening. It felt like you were on some rocket being thrown from Earth to the moon. And while [that] was very exciting, it was going at a pace that meant you really had to look at all your navigation instruments to make sure you're still headed in the right direction.

EMILY M. BENDER: Timnit Gebru [at the time, an AI ethics researcher at Google] approached me in a Twitter DM exchange, asking if I knew of any papers about the possible downsides of making language models bigger and bigger. At Google, she saw people around her constantly pushing: "OpenAI's is bigger. We've got to make ours bigger." And it was her job to say, "What could go wrong?"

The paper that Bender subsequently wrote with Gebru and her colleagues — "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" — injected moral urgency into the field's core (and increasingly sore) arguments around form versus meaning and method versus scale. The result was a kind of civil war in NLP.

KALIKA BALI: Some of the points that Emily makes are things that we should be thinking about. That was the year that the NLP community suddenly decided to worry about how it had neglected everything except the top five languages in the world — nobody ever talked about these things earlier. But what I did not like was that the entire NLP community kind of organized themselves for and against the paper.

R. THOMAS MCCOY: Are you pro- or anti-LLM? That was in the water very, very much at this time.

JULIE KALLINI *(second-year computer science Ph.D. student, Stanford University)*: As a young researcher, I definitely sensed that there were sides. At the time, I was an undergraduate at Princeton University. I remember distinctly that different people I looked up to — my Princeton research adviser [Christiane Fellbaum] versus professors at other universities — were on different sides. I didn't know what side to be on.

KALIKA BALI: It was positive that that paper came out, but it was also stressful to see people that you really respect drawing swords at each other. I actually went off Twitter. I got stressed about it.

LIAM DUGAN: As a Ph.D. student, the tension arises: If you want to do research that has any sort of lasting impact more than two or three years after you publish, you kind of have to take a side. Because it dictates so much of the way that you even look at problems.

I regularly read people from both sides. Usually you just subscribed to Substacks to see the angry linguistics side, and you'd go on Twitter to see the pro-scaling side.

JEFF MITCHELL *(assistant professor in computer science and AI, University of Sussex)*: It felt a little abnormal, quite how controversial that all became.

As scale-driven research continued to accelerate, some felt that discourse within the field was seriously deteriorating. In an attempt to repair it, the NLP research community surveyed itself *in the summer of 2022 on "30 potentially controversial positions" — including "Linguistic structure is necessary," "Scaling solves practically any important problem" and "AI could soon lead to revolutionary societal change."*

SAM BOWMAN: The industry community that was doing a lot of this early work around scaling had never been that closely engaged with academic NLP. They were seen as outsiders. That led to a divergence in understanding and what people thought was happening between these two [groups], because they weren't talking to each other that much.

LIAM DUGAN: They gave a large part of the survey out at ACL [Association for Computational Linguistics, the field's top conference] that year. This was the first conference I'd ever been to, and it was very exciting for me because there's all these people that are really smart. So I get the survey, and I'm reading it on my phone, and I'm just like, "They sound like nutcases."

JULIAN MICHAEL: It was already a field in crisis. The survey just gave us a stronger sense.

LIAM DUGAN: You got to see the breakdown of the whole field — the sides coalescing. The linguistic side was not very trusting of raw LLM technology. There's a side that's sort of in the middle. And then there's a completely crazy side that really believed that scaling was going to get us to general intelligence.

At the time, I just brushed them off. And then ChatGPT comes out.

* * *

II. Chixculub (November 2022 through 2023)
ChatGPT • rude awakenings • "drowned out in noise"

On November 30, 2022, OpenAI launched its experimental chatbot. *ChatGPT hit the NLP community like an asteroid.*

IZ BELTAGY *(lead research scientist, Allen Institute for AI; chief scientist and co-founder, SpiffyAI)*: In a day, a lot of the problems that a large percentage of researchers were working on — they just disappeared.

CHRISTOPHER CALLISON-BURCH: I didn't predict it. I don't think anyone did. But I was prepared for it because I had gone through that experience with GPT-3 earlier.

R. THOMAS MCCOY: It's reasonably common for a specific research project to get scooped or be eliminated by someone else's similar thing. But ChatGPT did that to entire types of research, not just specific projects. A lot of higher categories of NLP just became no longer interesting — or no longer practical — for academics to do.

SAM BOWMAN: It felt like the field completely reoriented.

IZ BELTAGY: I sensed that dread and confusion during EMNLP [Empirical Methods in Natural Language Processing], which is one of the leading conferences. It happened in December, a week after the release of ChatGPT. Everybody was still shocked: "Is this going to be the last NLP conference?" This is actually a literal phrase that someone said. During lunches and cocktails and conversations in the halls, everybody was asking the same question: "What is there that we can work on?"

NAZNEEN RAJANI: I had just given a keynote at EMNLP. A few days after that, Thom Wolf, who was my manager at Hugging Face and also one of the co-founders, messages me, "Hey, can you get on a call with me ASAP?" He told me that they had fired people from the research team and that the rest would either be doing pre-training or post-training — which means that you are either building a foundation model or you're taking a

foundation model and making it an instruction-following model, similar to ChatGPT. And he said, "I recommend you pick one of these two if you want to continue at Hugging Face."

It didn't feel like what the Hugging Face culture stood for. Until then, everyone was basically just doing their own research, what they wanted to do. It definitely felt not so good.

Rude awakenings also came from the bottom up — as one eminent NLP expert found out firsthand while teaching her undergraduate course in the weeks after ChatGPT's release.

CHRISTIANE FELLBAUM *(lecturer with the rank of professor of linguistics and computer science, Princeton University)*: We had just started our semester. Just before class, a student whom I didn't know yet came up to me, showed me a paper with my name and title on it and said: "I really want to be in your class — I've researched your work and I have found this paper from you, but I have a few questions about it. Could you answer them?"

And I said, "Well, sure." I was flattered: He's researching me, how nice. So I leafed through the paper. And while I was trying to refresh my memory, he broke out in hysterical laughter. I said, "What's funny?" And he said: "This paper was written by ChatGPT. I said, 'Write me a paper in the style of Christiane Fellbaum,' and this is what came out."

Now, I didn't read every line, because I had to start class in 10 minutes. But everything looked like what I would write. He totally fooled me. And I went into class and thought, "What am I going to do?"

Over the next year, doctoral students faced their new reality, too. ChatGPT threatened their research projects and possibly their careers. Some coped better than others.

CHRISTOPHER CALLISON-BURCH: It helps to have tenure when something like this happens. But younger people were going through this crisis in a more visceral way. Some Ph.D. students literally formed support groups for each other.

LIAM DUGAN: We just kind of commiserated. A lot of Ph.D. students that were further on than me, that had started dissertation work, really had to pivot hard. A lot of these research directions, it's like there's nothing intellectual about them left. It's just, apply the language model and it's done.

Weirdly enough, nobody [I knew] quit. But there was a bit of quiet quitting. Just kind of dragging your feet or getting very cynical.

RAY MOONEY: One of my own [graduate students] thought about dropping out. They thought that maybe the real action was happening in industry and not in academia. And I thought, you know, maybe they weren't wrong about that. But I'm glad they decided to stay in.

JULIE KALLINI: Starting my Ph.D. in 2023, it was an uncertain place to be. I was really unsure about where my direction would end up, but everyone was in the same boat. I think I just came to deal with it. I tried to make sure that I know my machine learning fundamentals well. It's not the wisest thing to only specialize in potentially fleeting trends in large language models.

Meanwhile, NLP researchers from Seattle to South Africa faced a firehose of global attention, not all of it good.

VUKOSI MARIVATE *(ABSA UP chair of data science, University of Pretoria; co-founder, Masakhane)*:
I don't know how many tutorials I gave on LLMs in 2023. On one hand, you've been trying to talk to people for years and say, "There's interesting stuff that's happening here." Then all of a sudden, it's just a complete waterfall of, "Come explain this to us."

SAM BOWMAN: It goes from a relatively sleepy field to, suddenly, I'm having lunch with people who were meeting with the Pope and the President in the same month.

EMILY M. BENDER: Between January and June, I counted five workdays with no media contact. It was nonstop.

ELLIE PAVLICK: Before ChatGPT, I don't think I ever talked to a journalist. Maybe once or twice. After ChatGPT, I was on *60 Minutes*. It was a huge qualitative difference in the nature of the work.

CHRISTOPHER CALLISON-BURCH: I felt like my job went from being an academic with a narrow audience of graduate students and other researchers in my field to being like, "Hey, there's an important responsibility for scientific communication here." I got invited to testify before Congress.

LIAM DUGAN: As a second-year Ph.D. student, I was suddenly being asked for my opinion in interviews. At the time, it felt very cool, like, "I'm such an expert in this!" Then it felt less exciting and more sort of overwhelming: "Where do you see this going in the future?" I don't know. Why are you asking me?

Of course, I would answer confidently. But it's crazy: There's thousands of papers. Everyone has their hot take on what's going on. And most of them have no idea what they're talking about.

SAM BOWMAN: There was this flowering of great engagement: Suddenly a lot of really amazing people from a lot of fields were looking at this stuff. And it was also getting drowned out in noise: everyone talking about this stuff all the time, lots and lots of really dashed-off takes that didn't make any sense. It was great, and it was unfortunate.

NAZNEEN RAJANI: That year was kind of a roller coaster.

In December 2023, one year after ChatGPT's release, the annual EMNLP conference convened again in Singapore.

LIAM DUGAN: The temperature was just so much higher, and the flood of arxiv [preprint] results was just so intense. You would walk the halls: All the way down, it was just prompting and evaluation of language models.

And it felt very different. At the very least, it felt like there were more people there than good research ideas. It had stopped feeling like NLP, and more like AI.

* * *

III. Mutatis Mutandis (2024–25)
LLM-ology • Money • Becoming AI

For NLP, the LLM-generated writing was on the wall — and it said different things to different people in the field.

R. THOMAS MCCOY: Anytime you're doing work that asks about the abilities of an AI system, you ought to be looking at systems for which we have access to the training data. But that's not at all the prevalent approach in the field. In that sense, we've become more "LLM-ologists" than scientists.

ELLIE PAVLICK: I am 100% guilty of this. I often say this when I give talks: "Right now, we are studying language models." I get how myopic that seems. But you have to see this really long-game research agenda where it fits. In my mind, there's not a path forward to understanding language that doesn't have an account of "What are LLMs doing?"

KALIKA BALI: Every time there's been a technological disruption that mainly comes from the West, there's always been these — if you may call it — philosophical concerns. Whereas in most of the Global South, we've kind of gone with, "How do we make it work for us here and now?"

Here's a tiny example. In India, the initial idea that everyone gathered around [after ChatGPT came out] was to have generative language models do their work in English and then put a translation system in front of it to output into whatever language you wanted. But machine translation systems are literal. So if you have a math problem that says, "John and Mary have a key lime pie to divide," and you translate it into Hindi, I can bet you most people in India do not know what a key lime pie is. How would you translate that into something culturally specific, unless the model itself is made to understand things? I became much more interested in how to solve that.

IZ BELTAGY: There is a point where you realize that in order to continue advancing the field, you need to build these huge, expensive artifacts. Like the Large Hadron Collider — you can't advance experimental physics without something like this.

I was lucky to be at Ai2, which generally has more resources than most academic labs. ChatGPT made it clear that there's a huge gap between OpenAI and everybody else. So right after, we started thinking about ways we can build these things from scratch. And this is exactly what happened.

In 2024, Ai2's OLMo provided a fully open-source alternative to the increasingly crowded field of industry-developed language models. Meanwhile, some researchers who had continued to study these proprietary systems — which only grew in scale, capability and opaqueness in the post-ChatGPT AI boom — were already encountering a new kind of resistance.

YEJIN CHOI: I had this paper [in late 2023] demonstrating how the latest GPT models, which were seemingly good at doing multiplication, suddenly get very bad at it when you used three- or four-digit numbers. The reactions to this were super-divisive. People who don't do empirical research at all were saying, "Did you do your experiments correctly?" That had never happened before. They were emotional reactions. I really like these people, so I was not put off by them or anything. I was just surprised by how powerful this thing is. It was almost as if I'd hurt their baby. It was eye-opening.

Ungrounded hype isn't helpful in science. I felt it was important to study the fundamental limits and capabilities of LLMs more rigorously, and that was my primary research focus in 2024. I found myself in a weird situation where I was becoming the negative naysayer for how the models cannot do this and that. Which I think is important — but I didn't want it to be all that I do. So I'm actually thinking a lot about different problems these days.

TAL LINZEN: It's sometimes confusing when we pretend that there's a scientific conversation happening, but some of the people in the conversation have a stake in a company that's potentially worth $50 billion.

The explosion of research momentum, money and hype vaporized the already-porous boundaries between NLP and AI. Researchers contended with a new set of incentives and opportunities — not just for themselves, but for the field itself.

NAZNEEN RAJANI: It opened doors that wouldn't have otherwise. I was one of the first people to get the data to reproduce ChatGPT in open-source — I basically wrote the recipe book for it, which is amazing. And that led me to get a good seed round for my startup.

R. THOMAS MCCOY: Any faculty member who is AI-adjacent starts to be viewed as an AI person — you sort of get typecast to play that role. I'm happy to work on AI because it's one of the most impactful things that I can be doing with my skillset. But the thing that would bring me the greatest joy is diving deeply into interesting corners of grammar and human cognition. Which is something that can be linked back to advancing AI, but that pathway is pretty long.

JULIE KALLINI: It's all a matter of semantics, right? Personally, I see myself as working across NLP, computational linguistics and AI at the same time. I do think there are different communities for each field, but there are plenty of people who bridge several areas.

JULIAN MICHAEL: If NLP doesn't adapt, it'll become irrelevant. And I think to some extent that's happened. That's hard for me to say. I'm an AI alignment researcher now.

ANNA ROGERS: I'm not concerned. Basically that's because I don't think we've actually solved the problem. The only reason to get upset is if you think: "This is it. Language is done." And I don't think that's true.

CHRISTOPHER POTTS: This should be an incredible moment for linguistics and NLP. I mean, the stakes are very high. Maybe it's one of those moments of a field waking up and realizing that it now has incredible influence. You can't pretend like you're a quiet scientific or engineering field anymore that just does research for the sake of research — because now all the money in the world is behind you, and every big corporation is trying to exert influence on what you do, and language models are being deployed all over the place.

If you achieve so much, you also have to accept that the debates are going to be heated. How else could it be?

* * *

Epilogue: Were Large Language Models a Paradigm Shift?

Not surprisingly, opinions differ.

TAL LINZEN: If you asked me five, seven, 10 years ago, I would never have thought that just typing an instruction into a language model would get it to complete the sentence in a way that is consistent with what you're asking it to do. I don't think anyone would have thought that that would be the paradigm these days. We have this one interface that basically lets us do everything.

ANNA ROGERS: As a linguist, I wouldn't say so. Back from the word-embedding days [in 2013], the whole premise was transfer learning — you learn something from a large amount of textual data in the hope that this will help you with something else. There have been shifts in popularity, in architectures, in how the public feels about this — but not in this underlying principle.

JEFF MITCHELL: I feel like the corporate interests have changed the way the game is played.

ELLIE PAVLICK: I think the media involvement makes a difference. Scientists in my field realized that success could look like becoming known outside of NLP, and suddenly the audience changed. Papers on arxiv.org are often titled to be picked up by journalists or Silicon Valley enthusiasts, not by professors. That's a huge change.

VUKOSI MARIVATE: I think in some ways the barrier to entry both got reduced and heightened. The reduced part is that there's still a lot that we just don't understand about what's actually going on in these systems, so there's a lot of work that's just prodding them as much as possible. In that case, you don't need to know the architecture of a neural network like the back of your hand.

At the same time, the barrier was heightened because in order to play with and prod those architectures, you have to be in a very high-resource space, computationally speaking.

EMILY M. BENDER: I have seen an enormous shift towards end-to-end solutions using chatbots or related synthetic text-extruding machines. And I believe it to be a dead end.

CHRISTIANE FELLBAUM: The big shift, or shock I would even say, is that these large language models are getting so powerful that we have to ask, "Where does the human fit in?" That's a paradigm shift: a shift in technology, how these models are trained and how well they can learn. And then of course the educational consequences, like in my class. Those are the things that keep me awake at night.

R. THOMAS MCCOY: In linguistics, there are all these questions that historically have been largely philosophical debates that suddenly are empirically testable. That's definitely been one big paradigm shift. But from a certain point of view, the way the field looked like 10 years ago was: people creating some data set, throwing a neural network at the data set, and seeing what happened. And that version of the field still exists, just with much larger data sets and much larger neural networks.

CHRISTOPHER POTTS: Maybe this is the way it always works, but the hallmark of a paradigm shift is that questions we used to think were important now no longer get asked. It feels like that has happened over the past five years. I used to focus a lot on sentiment classification, like, "Give me a sentence and I'll tell you if it was expressing a positive or negative emotion." Now the entire field is focused on natural language generation — all those questions that we used to think were central have become peripheral compared to that.

I suppose these are famous last words. Maybe in 2030, we'll look back and think this was nothing compared to what happened in 2029.

All conversations have been edited for length and clarity.