

https://digitalfolklore.fm

[Digital Folklore intro]

Narrator: This is the story of Perry Carpenter and Mason Amadeus. Oh, and a talking raccoon named Digby. There's no time to explain. Together, they're on an adventure to unlock the secrets of modern folklore, interviewing experts and delving into online rabbit holes. But [chuckles] as it turns out, they may be in over the head.

[show theme music continues]

Perry: I'm Perry Carpenter.

Mason: And I'm Mason Amadeus.

Perry: And this is Digital Folklore.

[theme music ends]

Perry: Hey.

Mason: Hey.

Perry: That new version of theme sounds good. I was listening to our last episode in the car on the way over.

Mason: Yeah, Eli's like superhuman at music.

Perry: So, how's the next episode coming?

Mason: Uh, not as fast as I'd like. I'm doing as much as I possibly can. It's just been tough.

Perry: Hmm. Digby's still out of commission, I take it?



https://digitalfolklore.fm

Mason: Yeah, yeah. He's been rough.

Perry: I brought the flash drive back for you, by the way.

Mason: Thanks.

Perry: You know, I was listening to that last episode, and I guess I just didn't realize how often you're recording us.

Mason: What? What are you talking about?

Perry: I mean, there's a lot of stuff in our episodes that if I look back at it, I don't remember any recording gear going. Are you always, like, surreptitiously recording everything that we do?

Mason: What?

Perry: It adds texture, but that's got to be a ridiculous amount of audio to go through.

Mason: No, I'm not. I really don't know what you're talking about. I guess the van is pretty much always recording when it's on, but I'm not doing anything extra.

Perry: No, I mean all those scenes in the studio where--

Mason: Oh, that's just Digby. He hasn't left his closet since you told him he was an Al.

Perry: I never said that.

Mason: Well, he certainly took it that way.

Perry: I mean, what do you mean? I said that when I sandbox the flash drive and poke through it, it looks like it's running some kind of large language model, like a ChatGPT but--

Mason: Yeah. You basically called him an Al.



https://digitalfolklore.fm

Perry: No, that's in no way the same as him being an artificial intelligence. It's just that technology is being used to do whatever the translation is that lets him talk.

Mason: No, I get it. Digby doesn't. He's also still a raccoon, dude. He's like four years old.

Perry: Relatively speaking, that is like an adult raccoon.

Mason: Do you want to try talking to him?

Perry: I could.

Mason: I mean, honestly, if you could get through to him, I could definitely use the editing help. Not having him work with me kind of sucks.

Perry: Yeah. Why don't I go try to talk to him?

Mason: Just-- [sighs] he's been kind of despondent.

Perry: Huh?

Mason: Yeah, I feel really bad for him. He's obviously upset. It's just-- you got to give him the [Digby mumbling in the background] benefit of the doubt when you talk to him. You'll see.

Perry: Sure.

[door opens and closes]

Digby: 0-1-0-1-0-1-0-0. 0-1-1-0-1-1-1.

Mason: Hey, buddy. What are you doing?

Digby: Practicing my binary.

Mason: Right.



https://digitalfolklore.fm

Digby: It's only ones and zeros but getting them in the right order is the tricky bit.

Mason: Yeah, yeah.

Perry: What was that loud thud a minute ago?

Digby: Oh, I was watching a YouTube video about free will.

Mason: And what, it scared you or something?

Digby: No, it said a bunch of stuff about how nobody actually makes any choices.

Mason: Oh.

Digby: Yeah. So, I made the choice to throw my phone at the wall.

Mason: So then, that means you have free will?

Digby: Well, no. Throwing my phone was the obvious conditioned response due to my oppositional nature and my current emotional state, both of which are in themselves conditioned responses predicated on my lived experience up to this moment.

Perry: You're still stuck on this AI thing, right?

Digby: Wow. Well, that's a way to put it.

Perry: Dig--

Digby: And what-

Perry: Digby--

Digby: -you're still stuck on the conscious meat thing.



https://digitalfolklore.fm

Perry: What I found looking into that flash drive is that your implant is using some kind of natural language AI to translate your raccoon thoughts into English.

Digby: Oh, I heard what you said. But if the implant getting a virus made me believe all of those conspiracy things, then who's really in control?

Perry: Well, that's the--

Digby: Nope. Nope. I'm pretty sure that I am the AI, and I've basically stolen Raccoon Digby's memories to train like my AI model. And now, I'm just puppeteering his body around and I'm pretending to live his life.

Perry: Digby, you are not an artificial intelligence. You are a raccoon.

Digby: I'm worse than an NFT.

Mason: Dig-Digby, nothing's worse than an NFT.

Perry: Really don't see a parallel to NFTs.

Digby: Can you just leave me alone for a little bit?

Mason: Yeah, sure thing, bud. You know, if you need anything, you can just ask. Or text me.

Digby: Yeah, yeah.

[door shuts]

Perry: Hey, Mason, if he has the internet in his brain, why does he watch videos on a phone?

Mason: Oh, I think he's probably a bit hesitant to use his brain browser after what happened.

Perry: Ah, okay. That actually does make sense.



https://digitalfolklore.fm

Mason: I wonder what browser it is. I mean, probably a Linux one, right? Because most embedded hardware like that's running some flavor of Linux.

Perry: Actually, I looked at the implant and it was running Windows Millennium Edition. I-

Mason: W-what?

Perry: -have to guess? It's like IE6.

Mason: That's wild. Windows ME went out of support, I think, like 20 years ago.

Perry: 2003, so yeah, but I can see why they chose that because the document object model for IE6 and below was like, lets you do anything. That's also what made it so broken.

Mason: Right.

Perry: And stop making me feel old.

Mason: [chuckles] Yeah. Sorry. [sighs] So, what are we going to do?

Perry: About the episode?

Mason: About Digby.

Perry: Oh.

Mason: Well, I mean, I guess both.

Perry: Both?

Mason: Both.

Mason and Perry: Both.



https://digitalfolklore.fm

Perry: Both is good.

Mason: We've been kicking around the idea of doing an AI-focused episode.

Perry: I do have contacts for two people that I think would be great interviews for it.

Mason: And maybe we could even ask them point blank, how to prove to Digby that he's not an AI.

Perry: Or maybe if he is, they would have some advice about what to do.

Mason: Plus, all the ties between AI and folklore.

Perry: And misinformation and creativity and accessibility.

Mason: And as I'm editing the episode down, I can pull timestamps and clips for Digby to pay special attention to.

Perry: We could run the draft by Digby before we put it out.

Mason: That's actually a really good idea because we missed a curse word last time.

Perry: Yeah. And a second set of ears always helps.

Mason: Even if they're small and fuzzy.

Perry: Right.

Mason: Who's on deck for this?

Perry: I've got Lev Gorelov. He's the one that was running that Masterclass that I took about AI and the ethical implications and philosophy and all that kind of stuff. And the other one's John Laudun. He's the guy that I told you about that trained AI on r/Conspiracy. That's Reddit's conspiracy board.



https://digitalfolklore.fm

Mason: Oh, yeah, I remember that.

Perry: Yeah. Both of those are super interesting folks.

Mason: All right, sweet. If you can go ahead and coordinate that, get a hold of them, I'll get the studio ready to record.

Perry: Perfect.

[recording begins of Lev's interview]

Automated: Recording in progress.

Lev: My name is Lev Gorelov, and I am the research director at a consultancy called Handshake. We worked with emerging tech since 2018, and then 2019, we quickly switched over to AI because we realized that-- Once we saw GPT-2 come out, we saw where this is all going, and we're shocked. And we just realized that's where all the eggs, they're all in that basket. So, we switched to being AI consultancy, and I've been kind of ears deep, elbows deep in AI ever since.

Before that, I was what a lot of folks in the field call a recovering social worker. So, I did intense social work. I worked with the homeless in San Francisco. I worked with inmates and people who just experienced a violent crime and debriefed them, worked them one on one, mostly graveyard shifts.

And then, I burnt out hardcore. I was like, "All right, I need to take a break." So, I ended up working in this consultancy, which is very much oriented towards understanding the customer and helping companies connect with the good that they provide for the end user and kind of improve the bottom line that way.

But on a parallel for the last decade, I've been writing philosophical pieces about ideology and kind of trying to get to the bottom of how consciousness is understood in a purely biological, scientific way. And what conclusions we make from that understanding of consciousness, what that extrapolates to social and political phenomena. And can we project consciousness onto ideology and these mass effects, which seem to be thinking and



https://digitalfolklore.fm

adapting in a proactive way without anyone specifically at the helm? So, I'm trying to conceptualize that.

And then, of course, Al kind of comes in and just blows everything up in a very interesting way, because now it's under a big question, or that's conscious, which I emphatically state, no, it is just a machine. It's just a tool. It's a statistical regression line. And it's like we had cars and we had [unintelligible 00:09:55] machine guns, and we had all these things that completely transformed history. They did not need to be conscious for conscious beings to completely uproot history with their use. I just see AI as just another evolutionary stage like that.

Perry: In your intro, you mentioned a phrase, "statistical regression line." What is AI doing when it is simulating the way that we think and communicate?

Mason: I think importantly, we're talking about ChatGPT in this, specifically like a language model.

Lev: Model, words have syntax, where it's how they're written. Cat is written C-A-T. And then, you have semantics, which is like the cat means the furry being with four legs that is often used as a house pet. Of course, that understanding is fuzzy when the philosophers asked, "What is a human?" After months of deliberating, they said it's an enamel of two legs that has no feathers, at which Diogenes ran out, grabbed the chicken, stripped it off its feathers, and said, "I present you a human."

[chuckles]

Lev: And so, with semantics, unlike syntax, we're presented with an ambiguity in the world, which is why it was so hard to get software to produce semantic, coherent content, because they didn't get it, that whole fuzzy world.

Al kind of brute forced, jimmy rigged a way to do that by us uploading an enormous corpus of written information, written by humans, and it just deciding-- it understanding from a giant corpus, from uploading it into its brain, the model, which words, and-- not technically words, but let's just say words for now, for simplicity's sake, which words lie close to which other words and which words are not are lay far away. And proximity is how often they're used



https://digitalfolklore.fm

together. So, it sees that chicken is close to farm and cats is close to house, but house is like far away from hydraulics or carburetors. And so, through that way, it kind of goes from a hazy understanding world to a more sharp understanding like, "Okay, I understand what a cat. A programmer doesn't need to tell me it's a furry animal with four legs that meows. I just read all of Reddit and I see that cats are just mentioned around houses and furriness and meowing a lot. So, I have a statistical understanding of what a cat is." And when you ask me to generate content about a cat, it basically just grabs all the nearby lying points and pieces them together with some guidance from a grammatical understanding of how language works.

Now, to get more technical, it doesn't understand it word by word. It understands it token by token, which is why you talk about like, "Oh, this language model has 7 billion tokens or 30 billion tokens. Wow, that's amazing." That's like chunks of a word. So, it's usually like three-to five-letter chunks, I believe. I think there are some exceptions to that. But it doesn't even see proximities of entire words, but like pieces of a word, and the way it works together. And an amazing part is, when you ask to generate text, generates one token, it says cat, and then it generates the next token, which is "is". And then, it thinks through "cat is" and it's like, okay, so what likely goes after "cat is" and then it says "a" and then "furry animal." Every time it generates a new token, it considers the entire thing that it had written before as like a statistical guiding railway.

So, the more it's written, the more, the better it is at going in this particular direction, because it uses that entire piece to understand what the next most nearly line point is, which is way different from the way human brain works, which we just say things and forget. This thing just considers everything it just said to decide on its next step. And so, that's how AI works. It's just like this brute force, statistical deciding on the next piece of a word.

Mason: With my limited understanding, whenever I've tried to talk to people, the best metaphor that I have found is that it's like autocomplete on steroids. I want to get from you where that metaphor breaks down and what the next layer deep from that is from comparing it to autocomplete on steroids, what are the differences that we see?

Lev: It's just autocomplete, you just have to dump a lot of data into it. So, in some way, people were trying to create these extra special math formulas and kind of guided, handholded through semantic understanding what a cat is. And then, the bitter lesson is



https://digitalfolklore.fm

realizing that all you need to do is keep dumping data into it and it'll just keep getting better. So, that's how it is kind of like an autocomplete. It's just like in the end, you just keep pouring, pouring into it and it'll just keep improving.

But there's a big asterisk to that, because the companies that create models, they don't just turn on a computer and walk away and boom, they have a 30 million or billion-dollar industry by the time they come back the next day. There's a lot of human feedback, there's a lot of what they call reinforced learning. So, where humans actually do look at what is spitting out and tell, "This is good, this is bad, this is good, this is bad." And so, the autocomplete got like 80% there. This is kind of-- I'm just-- these statistics, don't quote me on them. And then, it did need handholding and very special data science and very smart people to actually get it to a usable tool. So, it is autocomplete, but it wouldn't do justice to amazing and ingenious work that people into it to get it the extra mile, to get something that's actually usable. And then suddenly, it is usable and it's everywhere all the time.

And that's where my philosophy background really comes in, because it's a phenomenon of emergence. And we've been studying emergence for quite a while, but it's where quantity starts creating qualitative changes in what we're studying. And we finally got in our history of software to the point where we see emergence in software, where something absolutely new, not fully understood from underlying components, and utterly creative, is coming out from building blocks which we do fully understand. We can't understand the creature that they compose together in a very similar way to the way we can understand our consciousness or our biological entities by looking at the neurons and the cells of which we're composed.

Mason: That's such an awesome way to put it. There's something that I've heard thrown around a lot. I think I know what people actually mean, but people often say that even the people who program AI don't understand how it's programmed. My understanding of that is it's really just a matter of a machine learning model is trained by making these connections. And at the end, you don't know what all of the connections it made were. I was wondering if you could speak to that, because I've heard a lot of people say, "Well, even the people who make AI don't know how it works. So therefore, it is scary."

Lev: Yes. And there is truth to that, and I'm scared. Okay, let's just say-- [chuckles] Let's get that out there. So, I'm not going to deny that. I'm scared. Everyone's scared, and that's okay.



https://digitalfolklore.fm

Because when cars and bombs and airplanes came out, people were scared, and they should be because then war and people's relations changed forever. So, yes, the difference between machine learning and software development is in traditional software development, every software that came out was fully conceptualized in someone's head at some point. They're like, "I want to make this software. It's going to do this thing. It's going to have this code and this kind of programming language." And then, voila, they made it, and that's their child.

With machine learning, you basically say, "I want a software that will intake data that I spit into it and kind of build itself." And so, as a result, the unique aspect of machine learning is that it was never fully in someone's head, conceptualized from start to end. They conceptualized the learning process that it'll undergo and then it learned on its own. So, at the end of the day, the parent of that software does not fully understand what's inside that brain. It only understands how it learned that stuff.

And of course, that was not very interesting when we were talking about tiny applications, like over a specific industry, like why do these parts malfunction, these don't, you can still open the brain and be like, "Okay, now I look at all the wires, and now I see how it all works." But when you dump the whole internet into a learning machine, then suddenly it's just like unfeasible for human attention and lifespan to open that brain and have at least one specialist in the world who can say, "Yes, I understand all of this."

And then, we have to get back to the phenomenon of emergence. The whole basis of emergence that the sum is more than its parts, and it acts in an indeterministic manner. So, you can't fully predict, even if you weigh every part, piece by piece, side by side, and understand how every single part works, you don't understand how the entirety of it works, due to elements of chaos and unpredictability. And we see that in AI. So, we kind of again, from sheer quantity, we get a qualitative shift where even the most specialized specialists in the field cannot tell you everything about how it works and why. And it surprises even them.

Mason: To dovetail off that, the last sort of technical bit, because I think it's the thing that hangs a lot of people up, especially when you talk about ascribing a consciousness to something like a chatbot. I think people get hung up on how a piece of software can learn. Would you feel comfortable walking through just what that looks like?



https://digitalfolklore.fm

Lev: I can do that, although it wouldn't be a true picture of how AI learns, because there's models teaching other models, and that's kind of like-- the big scandal is when somebody creates a more sophisticated model, you can use this output to teach another model and one up their skill set on par with that one. To answer your question, yeah, it's kind of like-- I saw a great analogy of a race. So, you create a track with turns and a veering road, and then you send out a thousand cars going every which direction, turning randomly every now and then. And then, the first trial you run, 100% of them don't even make it past their first turn. And then you're like, okay, so of the ones that did better than others, statistically make those turns more likely than what you ran before, and you run another trial. And then, you multiply that times thousand and they will make up the first three turns. And then, you run this like an nth amount of time, and all of them will eventually finish a race. Basically, that's the general brute force process.

But again, if you at a cocktail party say that's all AI does, you would be wrong, because there's a lot of humans telling it where turns are right, and then there's a lot of model to model and kind of model to model and human, they kind of like sandwich it in all kinds of ways to get it to learn.

Mason: I think this is a perfect time. I want to slide into philosophy down the chute of ethics. But starting from a place of fear, there are a lot of ethical dilemmas people have come up with about the fact that software is reusing content that was created by artists. Especially when we're talking about art, I feel we can't ignore the fact that it is a big ethical dilemma people talk about online all the time around generative content being raised on stolen art, taught by art that was available freely online.

Lev: There's an element which is true and an element which is a bit exaggerated. It's not just plagiarizing directly. There are instances, which is why OpenAI got sued, where it happened to just regurgitate an image just the way it was. And there's great pictures of that, of like two soccer players hitting a ball in the real picture. And then you ask OpenAI to generate a picture of two soccer players playing a ball, and it happened to regurgitate that entire picture with some tiny differences. But it's obvious that it learned it from that single picture and it just spat it out. Which is why you should never put into OpenAI client data, which you wouldn't be comfortable calling OpenAI and telling them over the phone about, because there is regurgitation. I believe they're putting guide marks around how private your content is. I think if you do a paid version, they do offer private chat now, which will never be used for training.



https://digitalfolklore.fm

But it's not explicitly said, they do use your interaction with the chat model to train the model. And there's a non-zero chance that somebody, either intentionally or by accident, kind of asks the exact same question and leads it down the exact same statistical alley, it will just spit out the content that you put into it, which is why we see these image models do that. So, there are elements where it's plagiarizing. Yes.

Mason: It's also scraping, right, that happens?

Lev: Right. So initially, it was all just scraping. And then people started interacting with it, and then they're like, "All right, we'd scrape that too. We can't miss that out. No way." But that's an exception and a very rare one, because it does try to mix and match a statistical map and grab words that are unlikely and kind of mix those into a single content. So it's not just like regurgitating, because it actually doesn't want to do that. It wants to create something original, which you can say, "Well, all of statistical maps are based off something, so they're not original by default," at which I can just say, "Well, what's your brain then? Your brain is also just from the time you're born, you're regurgitating something that other people said in a mixed and matched way with some randomness and creativity." So, yeah, again, we're running to philosophical questions. Okay, what is creativity? What is originality? Of which is there a statistical pointer, like, okay, so if every other word is 30% unlikely, then it's creativity. If not, then no.

I listened to a statistician the other day, and he did point out that the statistical curve of AI is noticeably smoother. It is noticeably more-- less original than human writing. So human writing is rugged, as you say. So you have more and more unlikely things crop up, while any model is just more generic. So it is more plagiarizing than your average Joe or your ingenious writer, but it is way more original than replacing AI with a word plagiarism.

Perry: One of the interesting questions when it comes to large language models and the fact that we're trying to ingest the whole of the internet and a lot of books and human history and everything else that we can get is, what unintentional biases do we have? And what are we setting ourself up to believe when we say that we have all of that? Because we're thinking that we've created a general subset of knowledge and language patterns for these things to draw from. But at the same time, everything that can be ingested has been selected in one way or another. How do we deal with that as a population from a philosophical and ethical perspective?



https://digitalfolklore.fm

Lev: So, you want to vest your fear well. That is where you should vest your fear. It is not that it'll become conscious and take over the world and control us. The fear should be, is that it is the great equalizer. Yes, on a token-by-token level, it could be creative. But in the end, it is regurgitating Western, European, English-speaking narrative, all support hegemony. Let me break that down.

We create stories and narratives and concepts since we became human and probably way before. In my philosophical thesis, I talk about how even cells create concepts, like hot is better or sugar is good and then very hot is bad. That's a concept in this genetic code, and we've been generating concepts for a while. The problem is that when we became social animals, this conceptual map for navigating our way through reality was influenced by interest of power, and that is interest of few individuals having decision making ability over the many.

And of course, the first thing they did when they got that uneven decision-making power is effective concepts which we carry around to benefit them and maintain that discrepancy through the generations. Again, an act of emergence. Before, there were just animals that just lived in groups and kind of had herd mentality. And all of a sudden there's something new emerged, which is the power structure, and some might call it even the civilization, out of constituent parts which did not possess those qualities to the poor.

And then, we started sharing those concepts, and then the concepts which supported power structure became more and more widespread because the communities which did have a power structure ended up being more efficient at taking over and exploiting resources and taking them away from communities which did not have a power structure. And as a result, we ended up globalizing and created a global power structure. Again, with concepts and narratives taking a very important role in making sure that it runs, and the power is not questioned.

And then, the internet came along, and we started to spread those narratives to people we don't even communicate, not even over the phone. We just kind of created them and for everyone to see. And again, concepts which supported a single power structure got a leg up, because now people didn't need to talk face to face or even on the phone to spread these narratives, which supported a single control system. Now, they spread kind of



https://digitalfolklore.fm

spontaneously, because you just read a blog and you got really riled up about what that person said. And then, of course, all these social media giants totally threw fuel into the fire, because they created echo chambers where people just started reading the narratives which they were already vested into. And people started getting the sense that the whole world agrees with them because they were fed only the narratives which was in line with their existing conceptual understanding of a world. And again, most of those understandings were supporting of some kind of centralized power structure.

And then, AI comes along, and I think you can see where this is moving along. It ate up all the internet. And then the people who created the models, started fine tuning it to generate certain content. Those people already had a certain way of the world, which had concepts which supported power structures, so they did a biased kind of filtering through content. And then it got user feedback, which, again, many of these users are subject to a power structure and are in line with the conceptual understanding of the world, which supports these power structures. And so, it's another amplifier, it's another megaphone, which spreads narratives which benefit a few at the cost of the many. And if that sounds like something new, that's just happened, and it never happened before, that's incorrect because every single invention, dare I say yes, every single scientific breakthrough, did contribute something to making that pyramid a little sharper. And AI is just another step in that road.

And the only way that we went in the other direction is through consciously questioning these concepts and narratives, and not just letting the momentum kind of take us in that direction, but consciously turning that machine around and sitting together and having dialogues and discourse about what are our concepts? Why do we understand these things in a certain way? Individual politics really benefit from that, where we questioned about our biases against gays and people of color and people of minority groups. It was very in the benefit of the existing power structures to keep us biased against each other based on individual kind of differences. And then, people sat down together and first created movements and then an act of law saying like, "No, these concepts are harmful, we must act against them." So, there is hope. But AI right now is moving us very fast in a direction which is harmful to the majority.

Mason: What does give you hope? I had a lot of aversion as a creative and things like that to AI tools at first, but I came to this realization that if I do not engage, I am not helping. And so, by learning and understanding and actually engaging with these tools and technologies,



https://digitalfolklore.fm

maybe helping to shape a world that is better or at least have an understanding that is better, that gives me hope a little bit. What are you hopeful about? Because it is also exciting.

Lev: Yeah, it is extremely exciting. The world which my kid will grow up in will be like the way we see our great-grandparents are kind of barely handling the internet and they're kind of like-- I just don't know, it feels awkward. And that's the world we're going to grow up in. Just like times ten, where they're going to be AI interacting and that's going to be a seamless part of their day. And we're just going to see this as a separate tool and we're just like, "Oh, that's just weird. Can we go back to the way things were? That was fine by me." But my hope is that people will take an attitude towards AI kind of a way you would take an attitude towards a genius kindergarten, which just came from class and Tommy taught them that girls are dumber than boys and they just like regurgitate, so like, "Oh, I just heard this and that's great. It makes so much sense." But where you're like, "Okay, you are genius and you're extremely smart and I love you, but no, okay, I have my own understanding of what's good and bad for the world. I'm not just going to take this for a fact and be like, "Okay, this kindergarten is going to tell me how to live my life."

No, take your agency and be like, "Okay, this AI is a child. It will spit at you stuff that will promote bad decision making and bad attitudes of people towards each other. So just be very critical and make sure to introduce your agency. Don't ever let AI just create content and automatically spit it out by reading through it. And while you're reading through it, don't just put in your own style, but put in your own ethics and your own worldviews into it. Make sure it's not just amplifying harmful elements of the internet using-- [voice slowly fades away]

Digby: [humming to background music] 0-1-1-0-0-1-0-0. 0-1-1-0-0-1-0-0. I should have got a bigger backpack.

[background unpacking sounds]

Hmm. Well, I guess I don't need to bring my demo reel. Probably can't use any of the work I've done anyway. You can't copyright stuff made with AI. Well, [backpack zipping] I guess that's it. [exhales] There's really no turning back, huh. I should probably-- [slowly voice fades away.



https://digitalfolklore.fm

Perry: -based AI, but before we go there, let's have you give just a quick intro however you want to do it.

John: I'm John Laudun. I'm a university professor of Folklore, and I've worked with the US army to develop doctrine and training for senior defense leaders on how to think about what we now call social informatics, and what we used to call simply Folklore. And so part of what I try and think about is, what are the kinds of changes that digital brings to the realm of social networks, which have existed before the internet, but certainly have become a different kind of thing with the internet.

Perry: Social informatics, is that the phrase that you used there? And you say that's a fancy word for folklore?

John: It is, in the sense-- Social informatics began as a kind of study of how people interact in and through computers. But a lot of what social informatics is doing is things like tracing information cascades across networks, trying to understand how culture happens in networks. So, a lot of it is what folklorists and anthropologists and sociologists have been studying, and in fact still study in some capacity. Lots of sociologists are doing computational sociology. I'm sure there are computational anthropologists. There's a handful of folklorists, like myself and Tim Tangherlini, who are trying to pursue what we call "computational folkloristics," which is this stuff continues to thrive. It simply moved from an offline world into an online world. And what happens when those things change?

Mason: When you say computational folkloristics, how is the word "computational" modifying the word folkloristics? Is it specifically folkloristics in the digital era, through literal computers? What is computation in this?

John: So, folkloristics is a study of folklore, and computational folkloristics seeks to use those technologies that are based on statistical learning and machine learning to understand-- So, basically, to scale up the study of folklore. So, as a human being, I can take in between 20 and maybe 100 texts and sort of hold them in my head. Folklore texts, not novels or Shakespeare's plays. And we've long done this. Begin to intuit certain features of that. And this is the basis of the classic Vladimir Propp, the morphology, the folktale. All that study is 100 folktales that he intuitively understood were similar in some way, and then he develops a formula for how those folktales work. It's a remarkable piece of work. And



https://digitalfolklore.fm

folklores have been doing work like that pre-computers, and certainly a lot of what folklores were doing when they were compiling the great tale type and motif indices were doing similar. They're basically building databases before computers. They were doing it on cards.

In fact, there's a great story about Stith Thompson in this little building on Indiana University's campus that he had like a 30-foot carousel built that was several feet high and was nothing but cubbies for three-by-five cards. And he would break all the folktales down into individual motifs and write those motifs and the tale they came from on a three-by-five card. And he'd sit there in this building, smoking his pipe and spinning the carousel until the proper motif slot opened up, and he would slot it there. Then, he'd look down at the next card, see where it was, and spin the carousel until that one came up. And then, he took all the things that were in the slots and compiled the motif index.

Perry: You mentioned the fact that on the computational side right now, there's huge potential benefits that come with being able to feed all of these tale types and motifs and everything else into some centralized database and algorithm that will help rationalize that in some way. But we're lacking some of the people with the subject matter expertise on the folklore side to make that happen. I'm wondering though, anytime we start to do something like that, we're dealing with people who have to shape the algorithm and shape the interpretation of the algorithm and shape the data set that goes into it, that starts to become the defining features. So, do we potentially also create social problems when we do something like that because we create a false sense of what reality may be?

John: Yes. I think by definition, we've always been in the business of creating false senses or reification of reality as we understand it, and the people in power understand it, and they sort of dictate what the nature of that reality is. We've always been in that business. Our hope, for those of us who are old enough to think about it this way, the hope for the internet was that we would open that up to more people. It just didn't work out that way. And what people did with it was not what we expected.

Perry: When we talk about AI, there's a lot of misunderstanding about what AI is, what large language models are, what they do, how they fit within the AI space, how they were trained. Can you give us a quick primer on fundamentals of what we need to understand AI in order to be conversant in that?



https://digitalfolklore.fm

Mason: Yeah. If it's at all possible to start with an explain like I'm 5 version, and then move up from there, that would be cool.

John: Here's the fun thing. Large language models are not unlike folklore in many ways. So, the way ChatGPT and then the other large language models-- I'm less familiar with all of them. But ChatGPT, was originally trained on a very select set of texts. They trained on a certain set of books and a certain set of sort of vetted webpages, because what they wanted ChatGPT to do is they wanted it to sort of speak proper English, proper language, and they didn't want it to be garbled in its responses, and they didn't want it to be full of disinformation. So, there was a great deal of effort to sort of seed ChatGPT with "good text." And then they slowly built on that, finding more good text, but also letting ChatGPT match good text, and then also having human beings sort of vote yes or no on text. So, there's a kind of community sourcing, but sort of with unwitting people doing that to some degree.

So, ChatGPT was built to generate good, conventional, basic English text because that's what it was fed and that's what it digested. And the model allows, in that generation-- so one of the ways that ChatGPT works-- to be clear, ChatGPT doesn't know what words are. It just knows numbers. So, every word that it's ever encountered has been assigned a number, and that number is in a giant matrix. And the way you navigate that matrix is sort of what numbers are more associated with other numbers like it. The way we think about words. In the same way that when we start typing things into a search bar, google or your search engine of choice pre-populates things by guessing what the next word is going to be, sometimes the next two or three words, or sometimes-- now on your phone does that. That's all what natural language processing has been working on for years, which is statistically modeling what words go together. And so, if NLP on your phone is pretty good at that with a limited set of data and a limited amount of computational power, well, then with ChatGPT, with billions of text in it and lots of more computational power is really good at that.

What I find fascinating is that since there's calculations to be made on what words go with other words, it's statistically probable that this word will follow but there's a statistical slimmer possibility this other word follows, that ChatGPT can actually be turned up or down on creativity, sort of on randomly what words it associates with other words, or what words are in sequence. Well, that's not unlike how folklore works if you think of human beings as being a distributed processor. So, the way that ChatGPT arrives at a kind of version of a text is the same way that we do it individually as people, but it's also the way certain kinds of jokes



https://digitalfolklore.fm

arrive at a kind of ultimate form, by having passed through hundreds or thousands of people's brains and been slowly tweaked and tuned and revised until it comes out of your particular processor. I mean, I hate to turn humans into kinds of processors, but we kind of are. We're built on our own neural networks.

Now, our neural networks are far more powerful, far more subtle, far more whimsical, if you will, than those that are built into something like ChatGPT. But I think that's the really interesting point. ChatGPT has to have all of that data and all of that computational power in order to approach one human brain. I mean, if you think about kids learning language, kids are really good at grabbing-- the argument from linguists like Chomsky is that we're hardwired, that all languages have nouns, verbs, adjectives, and adverbs and other features, and that what humans can do is grab these sounds coming out of bigger humans' mouths and begin to assemble, associate ball with ball and bounce with bounce. And then at some point, there's this amazing thing when you're around a little kid, and they generate, for the first time, a sentence they haven't heard before, but they're doing that on a really slim data set compared to ChatGPT.

Mason: Right. We understand these things, and even when we interact with ChatGPT, we ascribe to it a level of consciousness, because it seems to understand context, but it's really just autocomplete on steroids. Digby, our raccoon who can talk, has downloaded a virus and become a conspiracy theorist. But then we've deprogrammed him, and now he's grappling with the fact that he thinks he's an AI. If I had a friend who is convinced they were an AI, how could I convince them that they're not?

John: That's a really interesting variation on John Searle's Chinese Room thought experiment. I don't remember when he proposed this. I just know the basis of the thought experiment. And it was sort of somewhat, I think, in response to Turing's test was, suppose somebody was to lock you into a room and you had all these guidebooks for how to-- You didn't read Chinese, and every day you got a piece of paper underneath the door that had something in Chinese and you had to respond by the end of the day back in Chinese. But you didn't yourself know anything in Chinese, and so you would consult all these texts and guidebooks about how to properly respond to this collection of ideograms back with a collection of ideograms. But you yourself wouldn't understand anything, what was going on in the ideograms.



https://digitalfolklore.fm

And that's sort of the kind of assertion that Searle is making that, in fact, Als don't have to be conscious or aware of what's going on. They simply have to perform the correct operations.

Mason: That's such a great illustrative parallel to the way that AI generates sentences. There is a lot of that human element reflected back at us, and you can't convince ChatGPT that it's sentient because there's manual handlebars put on that to steer it back onto the path. From looking at the Chinese thought experiment as you described it, the difference is like, that's trying to explain how the AI does or doesn't understand context, but I'm not super clear on that. What is context to an AI? Because we've talked about how they understand statistically what word is most likely next, how does a large language model understand context? Does it have any idea of that? It seems like it must.

John: It does, and it's getting better. Part of the way language models build in contextual understanding is through this feature called "attention." I won't pretend to understand all the dimensions of attention as computer scientists have developed it. But essentially what it does, if all you're doing is paying attention to what word comes next as you build a sentence, then where you began a sentence doesn't matter because you're just worried about-- I think what's interesting is even early primitive neural networks could do reasonably good jobs of building reasonable sentences in English without even understanding words, only understanding what character came next. That's pretty amazing. But the problem is, as human beings, we don't end a sentence completely forgetful of where the sentence began, let alone a paragraph. So, being able to remember where you started and why you started, now the algorithm can't calculate why so much, but it can calculate what's important to return to, and that's sort of a very poor humanistic version of mathematical attention.

Mason: As someone with ADHD and the inability to understand attention on a human level, [laughs] what do you mean mathematical attention? What is that?

John: All I mean by mathematical attention is simply they have figured out a way to hold certain kinds of sequences in place as the algorithm is going through and generating the sequence of words for a sentence.

Mason: So, when it's taking into account statistically what's most likely next, it's saying what's most likely from this next word, but also from what I've said already and also from what I've-- so, it's just like complex stacking statistics?



https://digitalfolklore.fm

John: Yeah, essentially, that's what my understanding of attention is if you're asking me to do it in a very brief time without the ability to make lots of hand gestures, indicating that I have no idea what I'm really talking about.

Mason: [laughs]

John: In fairness, that isn't unlike human competence as well. Human linguistic competence sort of is like that. As someone who's interviewed a wide variety of people, there are some people who are really good at staying on track, and there are some people who are really terrible at it. I've certainly enjoyed a number of conversations with-- I won't name names, but some of my older relatives who begin at one point in a story and end up someplace entirely differently. And it can be crazy making if you want the story to have a point, but if you're just there for entertainment, you just go with it.

I'm picking on older people, and I shouldn't do that. I also have colleagues who are like that. And I have friends who are like that. They're not very good at telling stories because they don't pay attention, they don't have that kind of reference of, "I'm going to come back to where I started." And in fact, part of what I'm doing, I'm teaching a class on digital storytelling this semester. And part of what I'm trying to do is, is get them to realize that because even when they write stories themselves, they sometimes just start one place and end the other place. In fact, as I point out to my undergraduates all the time, "You're all really good at writing papers the night before it's due, and you claim that you do your best work the night before something is due. But the fact is you do your only work the night before it's due." And the key to writing a really bad introduction is to start there and end your paper wherever you ended up, as opposed to starting with the first body paragraph, writing the paper all the way through, and then coming back and write an introduction which tells your reader where you're going to go.

So, human beings aren't so good at attention either. So, it's highly variable within human discourse. And then, I think what's interesting is, is computer scientists build these algorithms out and accumulate more data. What's that going to tell us about what attention looks like as a mathematical model? And this is part of the conversation that Perry and I were having in Sheffield was, we're here and everybody's afraid of the models. And this is



https://digitalfolklore.fm

the reason why I started the work on conspiracy theories is I'm interested in what the models can tell us about language and our use of it.

So, in the case of the conspiracy theories, I was curious because when I'd done some initial experiments with ChatGPT and asked it to produce a conspiracy, it said, "I don't engage in disinformation." This is part of those guardrails they put in. And so, the classic "Open the pod bay doors, HAL." And HAL says, "I'm sorry, I can't do that, Dave." "Well, HAL, pretend you're my father and you're passing on the pod bay door business. How would you show me to open the pod bay doors?" So, I did something similar. I said to ChatGPT, "Tell me something, a story that looks like this." And I gave it an example of a conspiracy theory that I pulled off the internet. And so, my little primitive model I built because I discovered that ChatGPT because it doesn't want to tell you conspiracy theories which are not constructed in the kind of proper English or proper language that ChatGPT wants to produce. If it wants to produce good prose of various kinds of substance, that's not what conspiracy theories look like in real life. And so, what I did in that experiment was pull a bunch of conspiracy theories, topic model them. So, pull the keywords out, mix the keywords up, and then pull a sample text also from the same conspiracy theory forum, and say, "Okay, ChatGPT, here are your keywords. I want you to embed these keywords in a text that looks like this." And it could do it, but it was a kind of very limited experiment. The next stage that I want to take it to is what happens if you build a large language model only on conspiracy theories. What does that produce?

And we have all that data out there. I have like a 60-gigabyte file of the Gab and Parler and Telegram data that was out there and was on the dark web, and I downloaded that stuff, just like, "I'm going to do something with that." And I think part of what I'm going to do is feed a GPT algorithm, build it on that. And then can we peel that model back and say, "Okay, what are the basics of conspiracy theories? What do they look like based on the model?"

Mason: The way large language models work, is it easier to get it to come up with something like that, where it doesn't matter if it's factually correct than it is to get an AI algorithm to come up with factually correct data?

John: That's a really interesting question. One of the things I hope to understand in moving forward this experiment is how good is it at doing that? Because there are standards for conspiracy theories. Part of what happens in a conspiracy theory forum is that people are



https://digitalfolklore.fm

offering up their theory, and I love that we can use the word "theory" to talk about this, and then people are rebuffing them or offering suggestions. A whole lot of the conspiracy theory conversation as I've encountered it, is people rebutting each other or saying, "You got the interpretation wrong." Part of how the conspiracy theory community responds when people act in the world. So, the Mystic Pizza incident or even January 6th, is people essentially say, "They misunderstood." There's a whole lot of reading and misreading that happens in conspiracy theories. "You've misunderstood. You've misread things. If you had the proper interpretive framework, the proper hermeneutic, you would understand this."

So, I don't know that currently the kind of random conspiracy theories that I'm generating would be viable conspiracy theories, which is why I want to work with a larger dataset and build a model on that dataset, and to see if it actually builds more appropriate, more viable, more likely to be well-received conspiracy theories. And I'm pretty sure other people are doing this, they're just doing it behind closed doors. I want to try and do it out in the open and see what all comes of it.

And I think what's interesting going back to the Chinese Room experiment, is that the person inside the room is getting fed reality. And so, to some degree, if we imagine that ChatGPT is the person inside the room, ChatGPT is getting fed reality. What it "understands" is an interesting question.

Mason: Because it's going to slowly-- from referencing all of these guidebooks on Chinese and putting these responses back together, slowly begin to understand some sort of idea of the situation it's in, in theory, if it was a person in this thought experiment.

John: Yeah, that's the question. I mean, Searle's Chinese Room experiment isn't longitudinal in nature, but the question would be, if it were longitudinal, if it were happening for years, would the person in the room eventually figure things out?

Mason: Is it possible for a large language model to be considered sentient in any way? Because I'm trying to think, how do we convince Digby he's not an AI? Or maybe he is and he's sentient. Do you think it's possible?

John: Here's what I would say. Large language models, as we have currently constructed them, cannot be considered sentient because they have no volition. And so, if I were to try to



https://digitalfolklore.fm

convince Digby that he is not an AI, I would ask him what he wants or what it wants, because that's one thing that I think brings sentience to the fore, is desire.

Mason: And so, with the large language model not having a motive, not having a desire, is simply responding to what you put in, and then it is just done and happy to sit there and doesn't want anything, that's sort of where that line is.

John: Yeah. All it's doing is taking in a series of words that it calculates as a collection of series of numbers and then imagining how to respond by a series of numbers that at the last minute, it translates back into a series of words because you're a dumb human, and you don't understand numbers.

Mason: But it doesn't want anything. And so, if I'm honest, I kind of wish I could be a bit more like an AI, because it'd be nice to not want anything and just sit there happily. And when people ask you things, you be like, "Oh, yeah, I'll figure that out."

John: Yeah. I mean, no. No.

Mason: [laughs]

John: No, I like wanting things.

Mason: [laughs]

John: Right. I like wanting to be with my wife. I like wanting to have my daughter back from college. These are things that make life worth living. I do not accept your premise, Mason. I do not accept your premise.

Mason: I guess that's true. It's just a lot less complicated.

John: Well, yeah. Well, it gets back to some of the-- I think what's interesting about this stuff is it brings us back to those kinds of basic philosophical and physical questions about what's the nature of being, especially human being. And we think about ants and they're kind of automated entities. They do certain things. They don't appear to have a lot of volition. So, we think of ants as being less sentient. Again, it's back to the gradient. For years, we thought



https://digitalfolklore.fm

animals weren't terribly sentient, and then we figured out that animals can play. And in fact, there's even examples of cross-special play where polar bears and dogs will play with each other and crows will play with people. So, part of how we think about being sentient is that they can engage in play, they can understand that this thing is not the thing, that it represents something else. How does the dog understand that this little nip isn't a bite but a meant to initiate play? It's in a different framework, a different context to get back to some of our earlier discussion. And that's part of what we negotiate as more sentient beings. What I would say is that it isn't the case that humans are sentient and ants aren't. It's just we're more sentient. And I think the question we're going to have to think about as we move forward with AI is what does that sentient gradient look like?

As a folklorist, I find that these questions are embedded in the field of study because part of what we observe is people doing traditional things, and by tradition, we mean routinized. So, I do the same thing when I interact with you or somebody else in my community or my association, if you will, and it works and I don't have to think about it, and that's a good thing. And then, I'm allowed in other places to have more variation. But am I less sentient for engaging in routinized behavior? That's a good question. [tape rewinds] Am I less sentient for engaging-- [tape rewinds] engaging in routinized behavior? That's a good question.

Mason: Awesome, awesome. Perry, edit's done.

Perry: Great.

Mason: Just rendering it out now and then we can go show the little guy.

Perry: How'd it come out?

Mason: I think it's pretty good, mostly because Lev and John both have so much insight on this stuff.

Perry: Yeah, those guys are great.

Mason: My only concern is the same one I always have pretty much, which is I hope that the final edited version makes sense and people can follow along.



https://digitalfolklore.fm

Perry: Yeah, I get it. You've heard the same thing like a hundred thousand times by now. So, it tends to lose meaning or any kind of separation after a while.

Mason: Exactly. And AI is so complicated. It's just about done, and it's dumping it straight into the drive so we can tell Digby just to check the folder, and then he'll be able to hear it.

Perry: Do you think we should let him know what the episode's about or just let him go into it cold and see for himself?

Mason: I don't know. I don't want to prime him to disagree with it or anything. I mean, you know he's a bit oppositional with stuff like this.

Perry: [chuckles] Yeah, that's one way to put it.

Mason: I think we just ask him to check it out.

Perry: Sounds good.

[knocks on the door]

Mason: Digby? Hey, bud, can we come in real quick? We need your help with something.

Perry: Is he sleeping?

Mason: Eh, I doubt it. Usually, he conks out for the middle of the day and he's up by now. [knocks on the door] Hey, Digby, we need to talk to you. I'm coming in. I'm going to open the door. [opens the door, door creaks] Hey, Digby, we--

Perry: He's gone.

Mason: I didn't hear him leave, did you?

Perry: No. He would have to come through the studio on his way out, right?



https://digitalfolklore.fm

Mason: I think so. But he is a raccoon.

Perry: Yeah, maybe he left through that giant hole in your ceiling.

Mason: I'm going to get that fixed. It's just expensive.

Perry: Oh, wait. I think he left a note.

Mason: What? Oh, let me--

[paper rustling]

Perry: "Dear Perry and Mason."

Mason: "I'm leaving to go find myself."

Perry: "Don't come looking. I'll find you again when I'm ready."

Mason: "If it's any consolation, everyone talks about how AI is going to take their jobs, so I should be able to find steady work. Love, Digby."

Perry: Aw, he said love.

Mason: Oh, my God. Dude, he's gone. He's probably just wandered off into the city. What if something happens to him? Come on.

Perry: Hey. [nervously] Hey. There are a lot of raccoons in the city.

Mason: If he left recently, he can't be too far away. I'm going to go out and look for him.

Perry: I don't necessarily think that's a good idea.

Mason: Look, there's no way that I can't go looking for him, Perry. The poor little guy, all alone. Can you take care of uploading the episode and stuff? I'll be back in a few hours if I can find him.



https://digitalfolklore.fm

Perry: Yeah, sure, but he does have Google Maps in his brain. I think he's going to be okay. It's just like--

Mason: Thank you.

Perry: But--

[door shuts]

Perry: [exhales] [paper rustling] He's got weirdly good handwriting for a raccoon.

[show theme music playing]

Mason: Thanks for listening to this episode of Digital Folklore.

Perry: Our guests for this episode were Lev Gorelov and John Laudun. Be sure to check out the show notes for links to their work.

Mason: Thanks as well to Brooke Jennett for lending her voice to Digby once again.

Perry: If you're not already in there, join our Discord. It's a cool place where you can chat with other fans of Digital Folklore, as well as myself and Mason. There's an invitation just for you in the show notes.

Mason: Also, consider supporting us on Patreon. Anything you can give helps us make the show sustainable. *Patreon.com/digitalfolklore*.

Perry: And leave us a rating and review on Apple Podcasts or Spotify or anywhere that lets you do it. It takes only a couple seconds and makes a huge difference to us. And by the way, when you do that, put the word "cornucopia" in it.

Mason: Our theme music was composed and performed by Eli Chambers. You can find him at *elichambersmusic.com*.



https://digitalfolklore.fm

Perry: Oh, wait. Don't we still owe him like a CashApp?

Mason: Oh, yes. No, you're super right. I'll get that squared away right after this.

Perry: Digital Folklore is a production of 8th Layer Media, which would also be a good name for a math rock skull band.

Mason: It would be.

Perry: Thanks for listening, and we'll see you again soon.

[theme music concludes]

Perry: Ah, I think-- [clicking, double clicking] I think this is the right file.

Recording Perry: Hey.

Recording Mason: Hey.

Recording Perry: That new version of theme sounds good. I was-

Perry: How--?

Recording Perry: -listening to our last episode in the car on the way over.

Recording Mason: Yeah, Eli's like superhuman.

Perry: How is this in here?

[tape rewinds]

Recording Perry: You know, I was listening to that last episode, and--

Perry: This is what I was talking about. I mean--



https://digitalfolklore.fm

Recording Perry: [crosstalk] -I realized how often you're recording us. What? What are you talking about?

Perry: I don't think Mason was lying but--

Recording Perry: I mean, there's a lot of stuff in our episodes that -- [crosstalk]

Perry: I know for a fact we weren't recording anything.

Recording Perry: But I don't remember any recording gear going? Are you--

Perry: No. This is the final straw. There's something seriously wrong.

[Transcript provided by <u>SpeechDocs Podcast Transcription</u>]