



MACRO Voices

with hosts Erik Townsend and Patrick Ceresna

Matt Barrie: AI'll Be Back!

September 19th, 2024

Erik: Joining me now is [freelancer.com](https://www.freelancer.com) founder and AI expert, Matt Barrie. Matt, it's been eight months since we spoke about AI last, it seems like there's been a disturbance in the force, some kind of a blow off. I don't know if it was a top or just an intermediate step along the way, but it seems like the enthusiasm is starting to abate a little bit on AI. Is this the beginning of the end? The end of the beginning, or something different? Let's recap what's happened since we spoke last.

Matt: Well, it's certainly interesting times. I think perhaps the best way to understand where we are in the space is probably to just do a bit of a recap first on the fundamental breakthrough that's happened in AI and how that translates to the underlying economics. Because, while the breakthroughs have been astonishing and unpredictable, even for the inventors of many other systems themselves, once one understands the economics, I think your listeners will get a good feeling for what's actually going on in the space and how it may play out. Now, fundamentally, what's going on, and the big breakthrough that's happened in the last few years, is the ability for effectively, machine learning or artificial intelligence, to consume very, very large data sets, to train and do so in a way where the more data you feed it, the better the AI gets. If you look at one of the very common forms of AI that are out there, which are these large language models like ChatGPT, which I think it is probably the AI that most of your listeners will be familiar with, essentially what these LLMs are, at the very core, a next word predictor. So, you take a lot of training data, you train the model, you then give it some new input, which is what you type into the ChatGPT interface. And all it's really doing, fundamentally, is predicting the next likely bit from your question, which was effectively the answer. So it's a bit like a next word predictor. So you give it a sentence, and ChatGPT will predict the next likely word, next word after that, and so on. Now, the fundamental breakthrough that happened in this space was really what's called a Transformer, which was invented by Google, which allows neural networks or machine learning to consume large amounts of training data without getting lost, and to do so in a very highly paralyzable way. So that reduces the time to train it and partly the cost. So you take a large amount of text, say, a large amount of English text that you scrape off the internet or get from books or other places, and you train it, and in the first instance, the AI will complete the sentence in a way that the output looks okay, but you could tell it was obviously written by a computer. But then what the Transformers allowed these models to do is step up the amount of data you feed in by an order of magnitude. Step up by an order of magnitude to compute and the parameters in the model. And then all of a sudden, the AI starts to get really good at completing that sentence. You know, the English is perfect, and so forth and it can carry a conversation. It becomes a little bit mesmerizing. But you step up again and again and again,

and all of a sudden, that's where you get this sort of voodoo magic coming out the end of the model, where it can suddenly speak to you in Persian. It doesn't know math, it doesn't know math, it doesn't know math, suddenly, it can do University calculus. It can pass the bar exam. It can write the next Harry Potter book. It can do pattern recognition. It can bounce a pencil on a robot arm. It can control the HVAC system in the building. Because it's consumed so much data into this model that these abilities start emerging now. These abilities are emerging in a way where the inventors, they didn't predict them, they don't know how it's really happening. It's just kind of coming out with a certain amount of magnitude. And I think this is the important thing to remember, in terms of the fundamental economics, is that, the breakthroughs that we're seeing as we go from ChatGPT to GPT 4.0. And, we're waiting for GPT 5.0 to come out sometime, it's been at least, I think, since was it March 2023? We've been waiting for GPT 5.0 to come around. These models need to consume ever large amounts of data. They need to consume and do so with ever increasing orders of magnitude in terms of compute that has real constraints in terms of access to chips, access to data centers, access to energy, and fundamentally, you also need access to the actual raw data itself. I mean GPT 2.0, which was really only interesting to computer scientists, and came out many, many years ago that had about 1.4, 1.5 billion parameters in the model, and scraped out about 8 million web pages. And to give people probably a bit of an idea, there's probably about a billion websites on the internet, and about 200 million of them are basically active. GPT 3.0, when that came along, stepped it up by an order of magnitude. So, you went from 1.5 billion parameters, you can think of a parameters like a synapse in a brain, and it went from 1.5 billion to about 175 billion parameters. And the data it scraped down was about 400 billion tokens, which is about 3% of Wikipedia, about 8% of books, 22% of the web, and 60% of what we call a common crawl data set, which is what all these models are using the fundamentals to train on.

When GPT 4.0 came along, which is that big leap, where all of a sudden it could pass any exam that any human could take in the top decile, or sometimes in the top percent, it stepped up from 175 billion parameters, or synapses, to 1.8 trillion parameters, and it went from 400 billion tokens to 13 trillion tokens. And pretty much consumed very, very large percentages of the web. It added in Twitter, it's believed to add in Reddit, YouTube, etc., and so forth. But it consumed a huge, huge amount of data. Now we've been sitting around waiting for some time before GPT 5.0 comes along. And you think about, well, where does it go from here? Where does it get the data? Where does it get the compute? Where do the chips come from? And where does the money come from? I think the last estimate of the training model, the last training instance for GPT 4.0 was something like \$80 million, just to do a training run. And I think Gemini is estimated to be something around \$200 billion. So yeah, as we kind of step up going through the scale, you can see here, we're starting to reach some fundamental constraints in terms of economics and reality, in terms of where's the data coming from. If you scrape down the entire internet, I mean, there are some really good contemporary data sets, for example, in your phone, and there are many other data sets that we will consume. They're trying to get into video, which is obviously, there's a lot of data there we haven't really got into, in a very, very big way. But where's the money coming from? We've seen Nvidia's stock price go tilt, and I think they're punching about \$300 billion a quarter of revenue, but 46% of that revenue that's coming in is coming from four customers, because you probably guessed the likely suspects of who

those four are, because there's not many companies out there that can afford to buy the chips, set up the data centers and run these training runs. And so that's kind of why I think we're in this little bit of a lull right now, is because we're starting to see that we've kind of caught up to the available easy data, the available easy compute within the constraints of what companies have to spend on training runs. And now we're starting to jump up that next order of magnitude. And that's why, despite OpenAI promising all these things, I mean, what have they talked about since GPT 4.0 has come out? We've had talked about this advanced voice mode search GPT, which is their attempt to go after perplexity in providing a better version of Google. There's all this cryptic meme activity on Twitter, which I don't know if it's Sam Altman kind of just running a soft puppet account, or what have you, talking about strawberry, whatever that may end up being, Orion and obviously Sora, the video modality of the GPT model, where some pretty impressive videos were shown.

But then there's been nothing for months as OpenAI is just trying to figure out how to make it all work. And it's also very clear, I think, to anyone, the layman even out there, that there's a complete lack of sustainable competitive advantage, open source is catching up very, very quickly. Facebook is really leading the way there by open sourcing a lot of the Llama tooling that they're developing, in order to neuter the competition and make it really a war of attrition in terms of resources. And there's no business model right at the moment charging a few cents for an API call. The end reality is, there's no actual business model here with these foundational models. Now, there's going to be some incredible applications, which we'll talk about later with the AI and the incredibly lucrative opportunities. But in terms of these foundational models that are taking \$100 billion to do a training run, and those numbers going up by laws of magnitude, you've got to think about it, if your cost of goods is \$100 million to do a training run, you've got to generate probably at least \$500 million of revenue in order for that to be an economically viable activity as a business. So, we're kind of in this lull at the moment. There's certainly been some interesting things happening at Open Source. There's been some interesting bits and pieces coming out in the space around the edges, certainly in the image space, and the ability to generate high fidelity images of any particular type, and the reverse, in terms of being able to analyze your images and then kind of extract what's going on in the scene, some pretty amazing applications that are going to be quite possible. It's very clear that the text modality is pretty much solved, and now they're trying to chip away on video. So, some pretty amazing things have come out. But fundamentally, we're reaching these limits in terms of just the world. Compute, chips, access to data, and ultimately the underlying physics of all of that, which will get into energy and so forth, which I know we'll talk about later in the episode.

Erik: Matt, if I draw an analogy to the internet and the E commerce, the dotcom boom and bust that happened, the way I think about that is, you know, Wall Street actually had the right idea, which is, the internet's going to be a big deal. But what they did is they just threw money at anything that had .com in its name, without understanding anything about what they were buying. That led to this big boom cycle up to the dotcom bust. And I actually think of that not as the end, but really as the beginning. We were done at that point with the frenzy and the hysteria, and it was time to start thinking about which things we really ought to be making investments in that made sense. And I think a lot of the technology, in terms of Internet technology, that was

invested in after 2000 is actually what lasted. Is that the right analogy? Did we just have the first blow off of hysteria in the AI bubble, and now we're ready to get serious about figuring out how to apply this technology? Is that a good way to think about it, or where's the timing relative to that, or any other analogy you might want to use?

Matt: I think you're exactly right. I mean, if you think back to the late 90s in the Cisco bubble, I mean, Cisco's tagline was, *we network networks*. And when the internet was going nuts in the late 90s, what would be a better business model than being the company that supplied the routers that connected up every single computer on the internet, right? And you had this phenomenal bubble in the Cisco stock price, where you just thought, this is just never going to end. You know, everything is going to be networked. Cisco powers that networking. Every single node needs a Cisco router connected to every other node. This is going to go on forever. Cisco is going to be the most valuable company in the world, and you just have this massive, massive run. Of course, at some point, that burst, because...

Erik: Hang on, the fallacy of that, and I think it's very applicable here, is everybody was getting excited about Cisco when Cisco was doing nothing but making routers, and that is such a commodity that anybody can compete with, that hey really didn't have any unique advantage. And it seems to me, although I don't know as much about the Nvidia specialized technology, at the end of the day, they're making, let's call it, graphics cards on steroids. There's been a whole set of architectures around AI specific hardware, but at the end of the day, these are GPUs on steroids. These are processors. We're not talking about the invention of generative AI, or of general AI, or of any that. It's just the hardware that it runs on. But nobody can think of what to invest in other than the hardware that it runs on. Sounds like exactly the same thing as the Cisco bubble. And I would argue that with Nvidia, it's just the hardware it's running on, that's not really where the substance is, is it?

Matt: Well, I think it's an even bigger problem for OpenAI. Actually, if you think about OpenAI, what is their sustainable competitive advantage in terms of access to data, access to capital, access to compute? Fundamentally, Ilya Sutskever, who is one of the principles of OpenAI said that 95% of the space in AI can be understood, understood by reading 40 papers that are published out there in the world. And so it's very, very, very clear that, at least at the fundamental model layer, that there is no sustainable competitive advantage, really. Unless you've got, maybe some secret licensing arrangements with data sets that you've obtained somehow. And you're just seeing a complete decimation of the foundational, the business model of foundational AI models. As you know, a team in France comes along and releases their version of the model, Facebook open sources their version of the model. I mean, every week there's a team coming along that's releasing a new version and goes into the benchmarks and might beat based upon a math benchmark or what have you. I mean, Elon Musk just came out of nowhere with Grok, too, it's already beating in a bunch of the math metrics and domain expertise metrics and so forth. With Nvidia, with chip design and writing, sort of Moore's law. Fundamentally, there are some incredible challenges that you need to have in semiconductor lithography in order to get the feature sizes down and so forth. And this is kind of one of the last

fundamental areas that the US really has an enduring competitive advantage over, for example, China, in terms of chips and so forth.

So from my perspective, I see it more as Nvidia's customers. There's really four that generate half of the revenue, and those four are really driving these foundational models. Those foundational models really don't have a business model yet, and really don't have a competitive advantage in terms of intellectual property per se, outside of data set access. And so, the customers of Nvidia are really the problem in this particular case, in my mind. And that's now hitting a wall, and that's where Nvidia will have a problem. Because I think, the last time I saw the numbers were that these four companies are spending about, I think it's \$200 billion a year on AI in terms of CapEx, and about 15 billion, a quarter of that is going directly into Nvidia. And so, it's really just down to those four companies and whether they're going to keep spending, because startups don't really have the capacity to raise that sort of money.

Now, what that is leading to, actually some pretty perverse things. So it's leading to these ridiculous valuations in these AI financial foundational models, because they need to have these high valuations in order to raise a lot of money. Now, you don't know really what's going behind the scenes in terms of the financial engineering, like this whole unicorn phenomenon we've seen in the dotcom space. Are these companies really worth a billion dollars? No, they're not. And you can synthetically generate a billion dollar valuation by putting in all these terms behind the scenes, such as liquidation preference and ratchets and so forth. So ostensibly, you can have like this glory headline for fame saying you're a unicorn. But the reality is, when the company ends up selling, it's not going to end up in terms of the payoff, like you, like a normal company would. So they're coming up with these stupid valuations. Now, while they may be financial engineer, they're getting so large now they've got to promise the world in terms of a return. And Erik, you're no more better than anyone, you know, you've got blue sky, and then you've got reality, right? When the sky is blue, you can pitch blue as sky as you want. You can have the most amazing stories, etc., and so forth. And you can really pump a valuation to them to the moon. When you got reality, revenue, earnings or a loss, or what have you, that's when you have a problem. And a lot of these valuations come crashing down to earth. Now, OpenAI, to get to the next generation, they're now in the market, I think, with \$100 billion, touting \$100 billion, \$125 billion pre money, trying to raise funds. Because, ultimately, the amount of money they need to raise is going to be in the billions in order to kind of fund the next generation of model. Now, in order to justify a valuation of 100 billion, and by virtue of that, the competitors of the other foundation models, they need to promise the world. So that's why you're coming along with these ridiculous stories about artificial general intelligence and how we're only a few years away from the singularity, and whoever invests in this model is going to have literally infinite return. That's what they're touting. Infinite return, because once you get to artificial general intelligence, you have man and machine merge, we go into nirvana. And whoever invests in that is going to conquer the universe, right? I mean, these are absolutely ludicrous stories that are coming out, and that's what they're touting. And you know, as a result of that, they're also running around trying to scare every government of the world into thinking that they need to regulate the space.

So, Sam Altman did a tour around the world, I think was bigger than Elton John's Fairwell Yellow Brick Road tour, going to every government of the world saying you got to regulate this, etc., and so forth, trying to stop competitors coming in, because obviously, there's no sustainable competitive advantage. At the same time, trying to justify these ridiculous valuations, and that's going to be absolutely game changing. So, I think it's a pretty crazy situation to be in. I mean, I think only back in February, I think OpenAI was trying to raise \$7 trillion, I think they're trying to have a whole tool chain, including the chips. And that didn't really go anywhere. So, I think we're kind of reaching the fundamental limits of P.T. Barnum here in promoting what these technologies can achieve. Because, fundamentally, if you think about it, GPT 4.0, 1.8 trillion parameters, if you think Artificial General Intelligence, what they're kind of thinking here is that the human brain has somewhere between, say, 100 trillion and 1000 trillion synapses, right? So if you kind of take Moore's Law and compute, basically efficiency, doubling every 18 months or so, or say, two years to get from 1.8 trillion parameters of GPT 4.0 to maybe what would be roughly equivalent to the human brain, maybe there will be about nine doublings from here, and voila, suddenly we have artificial general intelligence. And the whole API story is kind of, it's kind of out there. But if you think about nine doublings of 100 million dollar training runs, and nine doublings of the data center capacity, and nine doublings of the energy requirements to run those data centers and the grid, and nine doublings in maybe the data, where's the data going to come from? You might start to think, okay, now I can understand why there might be a little bit of a slowdown in the space and a lack of delivery, in terms of fundamental breakthroughs since GPT 4.0. I mean, it's been some amazing things that are really big, the next big step everyone's been expecting with GPT 5.0, etc., because we're starting to run into math, we're starting to run to economics, and we're starting to run out of bullshit, in terms of being able to promote these stratospheric valuations.

Erik: Well, if I'm understanding and assimilating everything you're telling me correctly, Matt, this sounds like a recipe for the most massive monopoly ever, because it sounds like the name of the game, really, if I look past the hype around Nvidia stock, which is really just about the hardware. The real key to this, it sounds like is, whoever is willing to invest enough to build the biggest and train the biggest model, they won. And when you get to the point where everybody else looks at what somebody did and said, boy, we can't top that one. You know, we can't possibly raise enough money to do better than they just did, what they just did is going to be the biggest one. Then whoever's got the biggest ones, got the biggest one. Nobody else is buying any more Nvidia hardware in order to train other models to compete with it, because they're giving up on that. And you end up with somebody's got the big monopoly, and now you know the government's trying to break them up on antitrust complaints. Is that where this is headed? I'm just trying to see where we go from here.

Matt: I see it more as that. For example, Cisco comes out with a gigabit router. Everyone else can produce a gigabit router, and at the end of the day, the money is not really made producing gigabit routers. It's really the applications that run on top of those routers, which is the software layers and the applications that happen within particular industry domains. I think there's going to be incredibly lucrative investment opportunities that they're not going to be in the foundational models, and certainly not at these stupid valuations, where I don't understand how you could

possibly make a return on \$100 billion pre money valuation as an investor. And in fact, OpenAI is running around saying this interesting quote here. I wrote it down, if I can just find it. Investors are currently required to sign up to an operating agreement that states, "It would be wise to view any investment in OpenAI's for profit subsidiary in the spirit of a donation." And that OpenAI "may never make a profit," right? I think the real money is going to be made just as it was in the dotcom boom, in the transformation of every single industry back then into an internet business and a software business. It's going to happen over the next few years, with every single industry transforming, in terms of an AI powered applications in the space. And it'll be across everything. It'll be across every single industry domain.

So, for example, in real estate, how are houses sold online? How are they rented online? There'll be sort of AI powered applications such as, real estate agents hate running a rental role, speaking to tenants about broken taps and replacing the carpet and this, that and the other, there'll be AI agents on the phone dealing with that, or however they're communicating. Imagine all that and there'll be some companies that come out with some breakthrough software very soon to kind of power that. And there'd be incredible returns to be generated in that particular industry. But this is going to happen in every single space out there. And I think we are going to go through an incredible wave of transformation from small business to large business because of this. I mean, if you think back to '94, '95, I mean, 1994 was the year that kids had email addresses. 1995 was the year your grandmother had an email address. And every internet business, every business in the world, wanted to become an internet business, and to do so, they built websites. And you had this tremendous boom where this huge cottage industry came out of nowhere. You had one-man bands building websites. You had small agencies, large agencies, and Deloitte building websites and today, from Freelancer, for example, this is a very large part of what we do as our largest category, about 30% of the work we do is website design. The next big boom you had in transformation in this sort of space was with smartphones. You had Android and iOS came out, and you had a big app boom. It wasn't as big as website design, because not every business actually ended up needing an app. Today, it's about 16% of the jobs we do on Freelancer, but I think the big thing that is going to happen today, which even be bigger than web design, will be AI development. So you'll have, just like you've got web development, you've got app development, you'll have AI development, and every business, small and large, will have AI agents powering the customer support, taking orders on the phone, processing a credit card, putting an appointment in the calendar. For example, it could be a hairdresser, and instead of someone, a human picking up the phone and taking the booking, or restaurants or a hotel or what have you, it will be the AI doing that, or doing outbound sales, this, that, and the other, I think you'll see this incredible transformation that require huge amount of domain expertise about that particular business that's very local in nature. So it's not going to be so much, you know, Google's going to come up with a technology product, and you just install it and it just runs your business. It's going to be very, very, very customized, much like how a website gets developed. So, I think that this is going to be huge, and it's across every single industry, and I think there'd be tremendous investment opportunities using the applications of AI across all these different industry verticals. I think this is going to be very, very big, and it's going to be transformational, and it's going to be a real wow moment, I think, in the next couple of years.

Erik: But what you're saying, economically, is that money is going to be made by the systems integrators. It's going to be the people who hook up AI to your business, to my business, to everybody's business. Take your website that now has an ordering page and add an AI driven wizard or something to it that asks you questions and places the order for you, or something like that. That's where the money is going to be made. And you think that the models themselves become a commodity that is just assumed, or do the providers of the model still make a profit here?

Matt: Well, it depends how much of the value chain they capture. I mean, like everyone seems to be converging on the same business model, whether it's Google or OpenAI or what have you, they all seem to be converging on in order to deliver this infinite return. And on the \$100 billion valuations they're all converging on, well, we're going to develop AI that's going to make humans redundant in work, and therefore, we're going to capture trillions of dollars of GDP, because every job is going to be done by our AI, right? Now, as we've seen in the past, there's been many moments where probably the world thought that was going to happen. I could imagine, as the desktop computer kind of came out, a lot of people thought, well, this the end of the world. Computers kind of will take everyone's jobs. But, as we've seen in the past, more jobs get created by technology than gets destroyed. I mean, we're certainly seeing this with our applications of AI, like we have an AI agent framework on Freelancer that's starting to do tier one support and tier one sales. And the amazing thing that we've noticed is that because the AI is incredibly productive, and the AI will generally can't do everything, because there's things that need to escalate to humans to do, sometimes in terms of processing a payment or getting access to data that it doesn't have access to, and so forth, that it actually ends up generating more work for humans. Plus, you have the teams that are doing the prompt engineering, you have the teams that are doing the fundamental framework development and so forth. So at least in our experience so far, we have a lot more people employed as a result of deploying these agent frameworks and having them run, than jobs are being taken away.

Now, I don't doubt for a second that there is going to be dislocation and disruption in some job functions. For example, I think in particular, things like large call centers, where it got 1000s of people in a room answering a phone, or 1000s of people in the room processing an order or doing chat support or email support or ticket support, etc., and so forth. I think that's fundamentally going to change in a very, very big way. But what we have found, even in our support centers, is AI has allowed the deployment of agents doing forms of customer support that were not economically viable previously, and there's overhead that you have now with your team in terms of managing those AI's, which means there's more employment. So, for example, it would never be economically viable for Freelancer to have a induction specialist welcoming every single freelancer to the platform from emerging markets. You know, we get 25,000 sign ups a day. Hi there, Muhammad. You've come from Lahore, Pakistan. I can see you've signed up. You've got a great profile. You've uploaded your portfolio, you're a web developer. I noticed you say you've got Flutter experience, but I don't see in your portfolio any examples of that. You've got a few typos in your description, etc. Would you like me to kind of walk you around the website? I mean, that would never be economically viable no matter where you employ

humans anywhere in the world to do that. You can now do that with AI, and in fact, all the workload that comes out of the result of that means you will have escalations to humans to do other things. For example, this person wants to buy a membership, or this, that the other and so forth. So, a lot more work does actually get generated. So, look, I think you will, in large companies like your insurance companies, your banks and very much traditional customer service type organizations, you will have dislocation, and you will have some disruption. But I think everyone's job function is just going to move up the stack, and so people won't be doing kind of what they were doing before, just like before you came into the office and you had no computer, no desktop PC on your desk, you did one thing. You had typing pools in offices, etc., with secretarial pools and so forth. And the computer came along and that kind of disappeared, and people did different things. I think we are going to see just people being more productive, doing higher order types of work and the AI really powering them to do so. And fundamentally, I mean, how many people have, you know, ChatGPT has been out now for some time. How many people really lost their jobs from ChatGPT? How many designers that actually fundamentally lost their job because of Midjourney? I think very little. I mean, we did, we do surveys, for example, of the freelancers, and a letter survey we've sent out to 1000s of them, is that half of the freelancers report earning more money now because it's more productive, and about 27% saying they're only about the same. So, I think there's a lot of hype, and there's a lot of promise. The hype that's out there is really to pump these stress valuations. I don't think it's going to be a world ending moment anytime soon, simply because we've got economic and physical constraints.

Erik: Matt, where do you see this going, in terms of the models? You said earlier that we've reached a point already where each next bigger model is exponential increase in the amount of dollars and electricity and so forth it takes to train that model. Can we go from ChatGPT 4 to ChatGPT 5 to 6 to 7? I mean, at what point do you get to the point where these models are just too big and it's not possible to get to the next one? Do we stabilize on one at some point? How do you see this playing out?

Matt: Well, it's a rumor that GPT 5.0, well, I mean, I thought it was going to be out by now, but now I think you could go to the betting prediction sites, they're targeting March 2025. I think the really exciting thing that we're going to see in the short term is going to be what I call the Midjourney moment for software development. So just as GPT came along, and it offered an astonishing set of capabilities for people who write copy and all derivative works of text generation, and just as Midjourney came out, and now you can type in a sentence and get photographic quality, hyper realistic image of basically anything, and more importantly, be able to reverse those images into text and so forth. So you could have a video camera running, and you could ask, what going in the scene? I think there's some pretty even more powerful applications coming up from running these models in reverse. But I think in software development, we're just starting to see now the solving of a related problem, which is, how do you load these large code bases into the models, which is akin to the text input box you type into ChatGPT, which is maybe just a sentence. But when you load in a code base, it may be the same number of lines as the F-35, the jet fighter. So millions and millions and millions of lines of code into these models, so that when you put the input in, it can analyze the code base and

actually do something. And I think this is going to be an incredible moment, which may take this whole AI revolution on a pretty exciting path, where effectively, you can, for example, load in your code base. So, I could load the code base of Freelancer in, I could say, okay, go through all the backlog of all the bug fixes you wanted to do and all the feature requests that any product manager ever came up with. And as anyone who in software development knows, your backlog of things that you want to do and all the bugs you want to fix grows exponentially forever, because it captures not just all of that, but all the aspirations this, that and the other, but what if you could hit a button and all of a sudden go, okay, I've got my backlog, just get that out and do that. Now, go look at my competitors' websites and find every feature that they've got that we don't have, and go build that for me, and you hit a button. We're using the Angular framework, which is a sophisticated front-end framework for high performance websites. I want to move that to react, now this would be, normally a multi-year effort to do so.

Or maybe another way of explaining this is, I'm in AWS on Amazon, I think that actually was great in the past for scaling, getting going. I think it's really expensive now, I want to get off Amazon AWS, and I want to go into my own hosted metal inside a data center somewhere, and save two thirds of the cost. Now, that would be a multi-year effort to do any other way, but with AI powered software development, I can maybe hit a button and maybe an hour or so later, bang, the whole complex environment has now been retargeted to a different architecture or a different platform. That's going to be truly crazy. You're also seeing, at the same time with that the democratization of software development, in that anyone could talk to a GPT like interface and get software developed. We're seeing that with applications like Teal Draw, which is a whiteboard. With a whiteboard, you can just sketch out, as you go to a conference room, let's sketch out how maybe the this program or application might work, but you can hit a button called "make it real," and boom, it will write the software for you automatically. Or the average person to the street might be able to say, okay, I want to make my own business. It's going to be Uber for pets. I want it to be like Uber, but instead, I want to have the following features, etc., and so forth. Go make me an app and a website, and also, by the way, build me an AI custom support team and a marketing launch campaign and everything else like that, and you'll just be able to talk to the software and have it be generated on the fly. Now that's going to be pretty exciting. There's going to be some questions about whether you'll get a human in the loop to drive the tooling, because what we've discovered is that the more you know about the nomenclature of the particular space that you want the output of the AI to be generated in, the better the quality the result you get. So, for example, all the training guides that's going around about Midjourney are about, you know, what does 75 millimeter lens do? What does a 200 millimeter lens do? How does color gradient affect the scene? What's an extreme close up versus a mid close up, etc. So the main culture is really about being a director of the scene or a software architect and so forth. So it'd be interesting to see whether you still need a human to look, and we think for some time to come, you will need to do that, but it is going to lead to an explosion of software development and very rapid product iteration. And it will be an iteration of business models akin to before the dotcom boom and after the dotcom boom, right? And how rapidly you could innovate and deliver products and services to people now you could do so over the internet, versus when you had to do over the telephone or through a catalog or and so on. So I think that's going to be the really big, amazing thing that we're going to see soon.

Now, the question is going to be whether it's going to be cost effective to be able to run these AI co-development platforms or not. I think when people have been running the numbers on some of the early systems that have been promoted, like DevOn and so forth, that they are quite costly. And in fact, they can cost more than actual physical software developer in your office. And I think that's also part of the reason why we haven't really seen a lot of these video AI tools released, because they cost 1000s of dollars per minute, potentially to run, and multiple runs, because you might not get the right result you want the first time you run it, you have to kind of do multiple edits. And so really, we may be seeing whether the inference costs or the compute costs, or fundamentally, the data center and the power costs are within the budget of ordinary business to be able to run some of these things. But I do think you are going to see a pretty interesting transformation come out. And I think that's where it's going to come, it's really at that sort of application level for software development, and then the ultimate impact on industry segments as they transform from the effectively the dotcom model, to an AI powered future. And I think that's where the real money is going to be made, and the real incredible things we're going to see happening in the world in the next few years.

Erik: So would you anticipate new AI driven products being introduced that do these things, or is it more of a service, a systems integration service that you see as where the money is made in this?

Matt: I mean, really both. I mean, you're just going to have, yeah, every time we deploy AI through like a funnel of a service on our site, we see incredible uplifts and incredible affinity with customers. So, it's just going to be intense personalization, predictive capabilities, just an incredible ability to deploy wonderful products and services to people across every single industry. And it's going to make the dotcom version of the web look very, very, very basic and primitive and old school, and it's going to be very, very exciting. For me, that's where the money is.

Erik: If these models cost 100 million bucks to train, where's the logic, or the ongoing impetus to continue training them and continue making them better?

Matt: Well, that's exactly it, because a lot of these capabilities around things like software and so forth aren't coming from increases in the foundational models. They're coming from solving analogous problems in terms of, effectively, large context windows. So it may be that what we see is a stall in these foundational models, and I think we're starting to, that's kind of the feeling that I kind of have right now is that we're kind of crunching on that next order of magnitude. And that's why everyone thought GPT 5.0 was going to be out, and it's been, what, 18 months, and it's going to be another six months, and then maybe that does come out, maybe it will deliver a big leap in textual capability. I mean, we haven't seen the emergence, really, of creativity and invention out of these models, and that may be the big next leap that comes out. For example, feed in every single patent into GPT, every single bit of scientific research, every single academic paper, every forum that anyone's talked about this particular scientific endeavor, and then go look at everything we know about science and develop novel scientific breakthroughs.

We haven't seen that yet. In fact, we haven't even seen a bestselling author yet, which has been ChatGPT. Probably the clearest thing we'll see. I think that will prove whether AI has really leapt through that level of creativity, will be if songs start trending on Spotify, hit songs that everyone's listening to and they're completely AI written, I think then we'll probably start getting a feeling that, okay, maybe AI has kind of leapt through and can actually do that next level of creativity that traditionally we associate only with humans, and can really create new things that are that level of creativity that in terms of a breakthrough. For example, if I sit down with ChatGPT today and I say, help me write an advertising campaign that will win an award, that will go viral, that will be super funny and really resonate with people and modern, you can't get it to do that. It may be that my prompting skills are not good enough, but I haven't seen anyone do that yet, at least it's not obvious that anyone's doing that yet. So I think that's the big thing that maybe will come through the next foundational models, we may be able to see that level of invention and creativity come out, and as I've said before, if you start seeing some songs trend that are hits and they're complete AI, and then all the artists suddenly start having a real problem, then we might get there. But instead, it's going to be more the applications.

So, I think these foundational models might stall. It might happen very soon, maybe in the next year. And, I mean, there are some other things that will come out at the same time. I mean, the data set of coming out of handsets, I think, is very powerful. I'm surprised that Siri is so awful today in 2024, Apple has really dropped the ball in terms of AI. I have seen last week, the Anthropic and Amazon have linked up and the next version of Alexa will be powered by Claude. I mean, Claude is a classic example. Claude has overtaken, I believe, GPT, in terms of quality of output. I know it also goes to show just how low the switching costs are in this space. I was using GPT for pretty much every day in terms of getting my work done and help me write a legal document or help me write for an essay or what have you, and then the second Claude comes out., I switch over to Claude, and it's dramatically better than GPT. Though, I will say one thing is that, which, again, ties back to the economics of this particular space, and it's not the economics of the training. So up until now, we've been talking about the training costs \$80 million to train GPT, for \$200 million rumored to be for Gemini, but the inference or the actual execution, the writing costs of these models, which I hinted on with the video and the software development, may be too expensive to actually, at this point in time to run these things. It seems that both OpenAI and Anthropic have a real problem in providing a stable model, at least in the consumer version, which you or I access and pay the \$20 online. You use the models, they're good for a while, and then they just go to crap. And this is reported all over the internet. So, for example, GPT started using listicles. So you asked a question and it helps you write an essay. And it used to produce a nice page of text, and then all of a sudden, it just started summarizing the output and giving you dot points, and not really writing what you want, but instead telling you how to go about writing what you want and trying to get you to do the work. And so, you'd sit there going backwards and forth, going, no, can you please write it for me? Can you please write the legal document? Not give me bullet points on how I should write a legal document, or tell me that I should go talk to a lawyer or this, that and the other or otherwise, tell me that you don't want to do the job. GPT did it, and then now Claude is doing it, where it just seems that one of two things are happening behind the scenes with these companies. One could be that

what they call RLHF, for reinforcement learning through human feedback, which is effectively, there's 1000s of 1000s of humans that are actually involved in the training of these models.

In fact, my company, Freelancer, provides them for one of the large foundational models, in that the fine tuning, or the safety training of these models, humans will look at the output of the model. They'll give it a left version and a right version. Do you prefer the left version of the right version? So, for example, it'll give you an example output of someone asking a question, and do you prefer the answer A or answer B? And humans will repeatedly go and perform left or right, left or right, left or right answers. And that will provide quite a dramatic improvement in terms of quality of these outputs. It could be that either that RLHF training, particularly in biasing the model for political correctness, avoiding answers about drug creation or criminal activity, or even political answers, GPT will write you a song about Joe Biden or Kamala, but it won't write you a song about Trump, for example. It could be that these biases, which are definitely shown to make the models worse in many ways that are unknown in terms of the fundamentals, but certainly evident in the output of the models. It could be that safety training is making these models worse, but I think more likely what is actually happening is the actual running costs. These models are too expensive, and what both OpenAI and Anthropic and other models are doing in order to save money is that they're biasing the output of the models for terseness to lower the inference costs or the running costs. And so that's why GPT, all of a sudden, starts producing summary dot points, and Anthropic does the same thing, because the business model is not there. You know, charge \$20 a month to run these models and do so and make a profit, is just not, it's not working. So, there is a fundamental problem, even with the models running today, that can't get, really, get stability. So, it'd be interesting to see where all this goes, but I do feel, I get the feeling, I'm sure everyone else gets the feeling, that we're starting to reach some limits.

Erik: Let's touch on some of the things that we both predicted in our previous interviews. We thought that there would be potentially, an epidemic of online scams enabled by AI, I have noticed that the online scams have gotten to have better grammar and better spelling. So, it feels to me like they're definitely using it at least that far. But as far as the really big epidemic of scams and shams online, doesn't seem like that's really materialized. Any idea, why not?

Matt: Well, I mean, scamming on the internet is a huge industry. It's the number one industry out of Nigeria, for example, 419 scams. Certainly, we are seeing that, and I think there's many different aspects to that. One is the scammers are clearly using GPT to generate conversations that are more believable when they're ripping off their victims. It's happening at scale on dating sites, for example. Ironically, you and I talked about potentially one of the big dangers of the world being AI girlfriends and basically some evil mastermind out of their island lair making all of the world's computer geeks fall in love with AI girlfriends and slowly twisting them to do certain things and then ending up controlling the oil. It turns out AI boyfriends are actually more powerful in terms of these dating scams than AI girlfriends, simply because AI boyfriends are far more empathetic, and that appeals much more to the female mind than AI girlfriends, where I think the guys just want to see video high quality videos and images that go in various states of undress. There's an interesting one, actually, I saw this week where some poor senior lady was

ripped off by a scammer who was actually using AI video to conduct the scam. So, in addition to seeing some photos which have probably just taken off the internet, of some gentleman, they claimed that he claimed that he was on an oil rig. He had sporadic internet connection, you know, would send GPT written emails to the lady and chats, etc., and so forth. But he would also get on the video call with her. Occasionally, these video calls had to be pre-arranged. So obviously, the scammer was kind of setting up the system and getting it ready to operate, etc., but he would actually run chats with the lady and talk to her over video. Now, there's some pretty crazy AI video software out there that could take URI and turn it into a streaming video avatar of anyone. So, you could be Taylor Swift, or you could be Claudia Schiffer, or whoever, and this particular scammer was actually talking to the lady as the gentleman in the photos that were sent across. Now, she caught him out, because at one point the video glitched, and there was someone in Africa with a sheet on his head, and that's how she picked it up. So, wow, what's going on? Am I being scammed? I do think that this is going to get more prolific.

There are certainly stories out there about companies getting scammed, where millions of dollars were transferred in a wire transfer after a video call was done with someone in the finance team. There's a lot of payroll fraud that's happening out there, where, and this is a big one that your listeners should be very much aware of, where it's quite an interesting scam, where, simply, the scammer will research someone on LinkedIn that's in a high paid role at a big company, then find out who the payroll officer is, and just send a fake email to that person, saying, hey, I just want to change my bank account for my next pay run, what do I need to do? And then the payroll office will say, just tell us your bank details, and I'll send it across. And I've just changed the payroll. And of course, then the next payroll runs, and maybe \$10,000 gets transferred to the scammers' bank account. That scam is a multi-billion dollar scam. It's GPT powered now, and huge. And I think this is going to get very, very bad. And I think it's going to get very, very, very bad, in particular with AML, a lot of platforms such as mine have to collect ID from people in order to process financial transactions. So, you need to have a driver's license or a passport provided and uploaded. Those documents now, are actually able to be synthetically created on the fly, using, not AI, actually, but ray tracing. So there's a bunch of websites out there, particularly run out of Russia, that you can type in, you know, Erik Townsend, a date, etc., and so forth, and look at a little photo of yourself. And it would just generate a photo realistic passport, photorealistic driver's license, photo realistic what have you. The defense against that was upload a photo of you holding your ID at the same time as your ID, or get on a video call and, talk to me about something while showing your ID, that now looks like it's going to be defeated, thanks to AI and the ability to stream someone's, anyone's image, and basically do a face swap or a body swap, etc., on the fly. So, I do think that this is going to be a real problem, and fundamentally, I am not sure how it's going to be solved in terms of authenticating people on the internet properly.

I do think, and this kind of touches on a little bit another aspect we talked about, I think, in the last interview, where we talked about the dead internet theory, and that potentially, in the near future and maybe even today, you know, 95% of the content you see on the internet may be produced by AI, whether it's something that may be easily understandable, such as AI content on websites, you know, all those blog posts and all those marketing copy you see on websites,

a lot of that today will be AI generated. For a while, Google tried penalizing companies that did that, and then they just gave up and tried a bunch of tooling for people to do it, because they just knew that the floodgates were open. They couldn't stop it. But even in chat forums and on the Washington Post comment section or Facebook comment, if you scroll through Facebook, some weird stuff happens now. You scroll and you'll see it's trying to suggest to you now all these weird groups, like, here's a group on, I don't know, Roman history, or, here's a group on architecture, what have you. And then you see all these weird comments. And it just looks strange that someone, you know, an AI photo, obviously, AI photo, might be uploaded a birthday cake that looks off, and then there'll be hundreds of comments about that birthday cake. And you just go, this is so weird and a lot of that is possibly just being generated by fraudsters and scammers. Because, there's quite a lucrative business model in ad fraud on a lot of these platforms. So I think the scams are going to get quite bad, and I think they're going to get worse, and I think we don't really have the capability to defend against it. At the moment, there's obviously reports in the media about parents being scammed by a caller ringing them up, they've got an AI version of their daughter or son's voice saying, I've just been in a car accident. Can you please send somebody, send a lawyer, I'm in jail, you know, I need to get out of jail. Can you please send it immediately? I've been traumatized, etc., and there's been some gentlemen that actually, I think, went in front of the Congress or whatever, got up to testify on kind of what happened to him, when that happened to his son. I think one bit of advice is probably to share a code word with your family right now ahead of time and say, look, if you ever don't think it's me, and a phone call kind of comes in whatever, just ask what a code word is. And that's a way of potentially defeating against it. That's sort of a one-time pad. But I think fraud is going to get really, really big with AI, I think we're starting to see some things, but it's going to happen at scale. I certainly want to be on a dating site right now and talking to people, because I'd anticipate that probably a lot of people on those dating sites are actually not real humans.

Erik: So where do you see this AI story playing out over the next couple of years? It seems like we've just had the Cisco router bust where, okay, Nvidia hardware is not the end all be all to everything AI that's starting to blow off. Where does it go from here? Are we seeing the equivalent of the dotcom bust at this point? Is it just getting started? Is there another wave? What happens next?

Matt: I think, as the dotcom led to a whole SAS revolution and we had the bust of 2000, 2001 then, certainly through the next couple of decades, you had Facebook, Twitter, and you had all these Google kind of taking off again, etc., and so forth. I think we're going to see a huge amount of venture creation and transformation and so forth. But it's going to be in, as I said before, the applications across every single industry. I think software development is going to be crazy. The ability for anyone to generate product service is going to be democratized and very inexpensive and very, very accessible. I think there's going to be an explosion in entrepreneurship and business creation for a few reasons. One is, it's going to be so easy and cheap to do so. And the second is, there will be some people out there that will need to change in terms of the nature of their work as AI becomes more prolific. You know, those that used to do copywriting need to become more like editors. Those that did illustration need to become

more like creative directors, those that are like software programmers today, just like we don't program an assembly language anymore, we may not program in Python in the next year or two. It may be that we become more like product managers and we talk in a much higher level way in order to get software developed. I think that's all going to happen. So, I think a lot of the magic is going to come, but I don't think it's really by holding our breath and waiting for the OpenAI to come up with the next model. I think it's going to be really now, the application layer, or the system integration layer. And I think that, just for example, this whole agent-based framework that's coming, I think that's going to be very transformative. I mean, every business in the world is going to want someone to answer the phones, take an order, put a booking in the calendar, and I think that that's going to be the first place we're going to see huge transformation. It hasn't happened yet. I mean, there are some companies like Klarna that keep issuing press releases saying, oh, AI's customer supports allowed us to cut our job workforce from 5000 to 4000 or what have you. But when you check sort of glass door, what's really going on is the staff say, oh, it's just a bad economic environment. They're cutting benefits as well. And it's not really the AI, but I think some big transformation is coming. I think it would be bigger than the dotcom boom, but it's going to be really in the application level, not at the foundational model level.

Erik: Matt, let's talk next about the amount of energy and computing power that all of this is consuming. We're getting to a point now where a lot of people are recognizing the AI trade as more of an energy trade, and there's actually people that are looking at nuclear data centers as an option for powering AI. Where do you see all of this headed, in terms of AI's demands on electricity? How is it going to be reconciled with the industry's ability to provide that computing horsepower?

Matt: Well, I watched an interesting video, actually the other day by Eric Schmidt. He was talking at Stanford, and I think they've actually pulled the video down, because he said a few controversial things in it, particularly about Google missing the boat because they preferred work from home, over innovation. But he said the American grid right now doesn't really have enough capacity in order to deploy the data centers that are needed to power the next generation of AI foundational models. In fact, he said what America needs to do, in order to power the next generation, is actually become best friends with Canada, for access to hydroelectric power, because the US doesn't have enough power. And I've got a very good friend who runs a multi-billion dollar data center company in my region. And he says, anecdotally, he gets called, he's in Australia, he gets called by the market operator quite regularly, saying, hey, can you please turn on the gas to power your data centers? Because if you don't, you'll brown out the electricity grid in Sydney or Melbourne when he builds a new data center now, that consumes 300 to 400 megawatts of power, and there's simply just not enough capacity in the grid to be able to build these things. As everyone's well aware, building energy generation capability, that's a multi-year exercise. I think the fastest you can build a nuclear power plant, I think the South Koreans do it in seven years, you're probably a much better, much more on top than I am in other more bureaucratic locations, decades. So you do fundamentally need, if you're going to run these huge training runs and also have the inference running, and some of these models are more expensive to run than their previous generation by

quite a factor. I think GPT 4.0 is about triple the cost of GPT 3.0, in terms of the actual compute power, and as a result, the energy costs to run the GPT 3.0, you need to have access to large amounts of cheap energy. So you have to go nuclear, or, be lucky, and have hydro nearby. And so really, this build out in data centers, where the demand at the moment is somewhat exponential, is more of an energy problem than anything else.

Erik: Well, I definitely agree it's an energy problem. I just wonder about the solution to that problem, because unfortunately, the technology that we really need, in my opinion, to build nuclear data centers is some of the newer generation IV nuclear technology that's not really ready yet. And I have talked to your friend in Australia who runs data centers to try to do, the let's build a KEPCO conventional, old school nuclear plant, that takes even the Koreans going the fastest that they can, could do it in four or five years, as opposed to seven or eight for American builders. But still, it takes too long, it costs too much. It's not really the right technology anyway. The getting to the right technology, the industry's not ready for that yet, and it leaves me wondering. I think what's likely to happen here is the guys that are in the data center business are going to find themselves in the advanced nuclear technology business, because they're going to recognize the only way they're going to get the energy they need, is going to be to fast track some of the developments that are already underway in the advanced nuclear space, but that aren't quite ready for prime time yet. So, I'm actually looking to see the data center and IT crowd get involved in the advanced nuclear energy business. We'll see if that happens or not.

Matt: Well, that's exactly what you're seeing. You're seeing the CEOs of these data center companies going out there, and they're not talking about data centers, they're talking about nuclear energy. And I think that's going to be the future, is you're going to see those two industries merge.

Erik: Matt, a minute ago, you said something about ChatGPT 4.0 moving to ChatGPT 5.0, maybe triples the cost of operating that model. Well, hang on a second, if charging 20 bucks a month to access ChatGPT 3.0 or 4.0, I can't remember which version you got for your 20 bucks a month, if that wasn't really working and that wasn't producing a profitable business model, then I can't believe that we're going to get 60 bucks a month for a fancier model. How is this going to work? What's the business model that's going to keep this stuff going?

Matt: Exactly. I mean, the average person in the street isn't using ChatGPT at all. I mean, the average person in the street may have heard about it, may have seen it. Perhaps someone showed it to them once. Maybe they logged in and played with the free version for five minutes. But the average person in the street is not using these models at all. And then, of the people that are using these models, the minority are paying the \$20 a month. And certainly, the future is not 60 bucks, \$200, \$500 a month. It just doesn't work so fundamentally, and this is the problem with these foundational business models, is they need to capture more of the value chain and to capture more of the value chain. That's why they're all heading the direction of, well, let's produce AI agents. These AI agents are going to replace certain job functions. We'll start maybe with customer support, and instead of you hiring a human at your hairdressing

salon to answer the phones, you're going to pay us to run an AI agent to answer the phones and take a booking and process a credit card. Now, the problem is that, in order to do that really effectively, you need to have a lot of domain knowledge about that hairdressing salon. Who works in the salon? What haircuts do you make? What do you know about the local community? What do you know about what we do? So that is not going to be solved by a point, click, install, bang. You know, phones are getting answered because the customers will hate that, right? You want to say, hey, is Sarah working today? And I really want to have a chat about the local market, or maybe your last haircut you had, or what have you. So that is going to require a lot of data about the salon, et cetera, put into that software. And that's why it's going to be more akin to the web development industry, where you need to have a web developer in there and do a lot of work to load in your data into the web pages, as opposed to hit a button and bang, you get a website, which is done in a second, and you just do it yourself. So, they've got to come up with a way to capture more of the value chain. Because certainly, making sense for API call or a few bucks per month, there's no business model there, and that's why I don't think these foundational models are going to make any money, and it's going to be instead, the smart companies and startups and businesses that are incumbent, that transform themselves, that are going to make the real bucks from this AI revolution.

Erik: With respect to using what's there in ChatGPT 3 or 4, or whatever version you can get access to now for 20 bucks a month, should we be using it? Because it seems to me, the way I feel about this, I went and got ChatGPT for a month, I tried it a bunch of times. I thought it was very educational to see what it's capable of doing. I figured out how to load the plug-in that allows it to search the web. It became much more functional when I did that. But honestly, after I'd given myself the tour, I didn't feel any real desire to keep it. Now, I've talked to other people who did keep it primarily as a replacement for Google. Instead of doing a Google search to find something, they do a ChatGPT search. I guess they feel that they can give the prompt more specific instructions than Google allows in the way that they specify search terms in Google. But from what I've heard about these things, they're changing so frequently that unless you're really into it, unless you're an AI aficionado dealing with the constant changes in the interface and the thing that you finally learned how to use, changing and becoming something else, almost seems like it's not worth it to me. Am I missing something? Is there something that everybody ought to be using? You know, your life would be better if you were on AI?

Matt: Look, it's absolutely transformative for me. So, I use these LLMs, whether it's GPT or, more recently, Claude, to produce any sort of written document. And it comes down to how good you are at the prompting. There's a classic example I think someone showed, where they said, okay, let's get GPT to write a tagline for a fashion business. And the average person on the street would go, please write for me a tagline for a fashion business, and you'd get a pretty garbage tagline. And then someone said, imagine you are a marketing person. Now, write for me a tagline for a fashion business, and you've got a slightly better tagline for a fashion business. But then someone said, no, if you write the prompt like this, imagine you are Gianni Versace. You are the world's greatest fashion designer. Now please construct for me a tagline for a fashion business, and then out comes this beautiful tagline that has an emotional connection to the brand and everything you're kind of looking for. So a lot of it does come down

to how good you are at the prompting. I mean, I use it all the time for all sorts of stuff. I use it for drafting of legal documents, is incredibly good at doing that, because it's read all the legislation, it's read all the cases, it's read all the discussions in the forums, it's read all the Twitter commentary, etc., and so on. And so, you've got to know how to prompt it in a certain way, and it will write for you things like that. I write a bunch of essays. Sometimes those essays get very, very long. I write these stupidly long form essays, in fact, for background for your listeners, before I got onto your previous interviews done with me on AI, I'll typically write a long form essay on AI, and there's two of them on Medium, which are really the companion articles to our two previous interviews, if someone wants to go to the background. But these essays can be quite long. They can be like 80 pages long. And so sometimes when you try and restructure those, it can be quite cumbersome to try and get the better flow. And if you write the prompt in the right way, it can be incredibly productive for getting things like that done, for helping writing marketing copy, and it can do so in a very sophisticated way. I mean, I did a slide deck the other day for the innovation contest cyber freelancer, which is where we will do things like run a ten million dollar contest for gene editing, the central nervous system of humans, etc. And just trying to explain that in a way that resonate with the C-suite, and it did a phenomenal job if you craft the prompt in the right way. Now, my only frustration with providing some specific pointers into what your listeners should do right now, and two weeks ago, I would have said, go get Claude. Go to claude.ai, pay the \$20, because the \$20 model is like night and day compared to the free model. And get in there and really just try, and when you write the prompts, don't just say, please do this for me. You kind of just get it to role play a little bit, imagine that you are the following. Now you're an expert in California law in employment. Now, please write for me an employment agreement suitable for the employment of a, I don't know, software developer at a large technology company such as Microsoft, etc., for my business, which is called blank, right? And it will produce an incredible output for you, for you. Now, the only frustration I have with that right now is, when I use the Anthropic model as in the last two weeks, they've done something to nerf the output or make it less verbose, and now it becomes incredibly frustrating exercise to actually get it to produce what I want it to do. But, I do think these tools are pretty incredible. They're a great productivity tool. And I do think that you should give another go and get in there and really have a try and figure out with whatever workflow you have, to kind of make it part of the everyday order of course of business. But they do get frustrating because there is a lot of model drift. Whatever they're doing behind the scenes, they'll work really well for a few weeks, and then you'll have to, then you'll switch to the next company's model, because they're out there trying to promote that, and they're running at a loss making way. And so, you get a much better output than when they try and run it for profit and the model isn't so stable anymore.

Erik: Well, it sounds like, if I'm not trying to be on the leading edge of everything, I ought to pick the most stable, well established, large model and then learn how to write prompts for it efficiently. So, which one should I pick? And where do I learn how to write prompts for it efficiently?

Matt: Lately, I've been using Anthropic and Claude and paying the 20 dollars. And in terms of image generation, Midjourney. And now Flux is pretty crazy for generating images, and that's basically it, in terms of what I use on a day to day basis. And I'm just imminently waiting, and I

get in front of my entire company every Friday, in front of town hall, and I just say to the engineers, any minute now, something's going to drop for software development. I want you guys on the leading edge, make sure you're watching everything that's being produced, all the tooling that's coming out, whether it's cursor or what have you, because any minute now, there's going to be a new tool dropping in. Soft development is going to change everything.

Erik: Matt, I hate to bring this up, but one of the timely conversations these days is warfare, and of course, AI has been promoted as having a lot of applicability to warfare, automating the way some battles are fought, so that even life and death decisions may be made without human involvement at some point. Where is all of this headed? Obviously, the best of it, I'm sure, is super classified, top secret, but what do we know about it? Where is it headed, and what information is available?

Matt: Well, I think, ironically, probably the biggest killer app right now for AI is probably flying drones in war. I think anyone that's on social media and kind of not watching what's going on in Ukraine right now has seen that the battlefield has been transformed by these low cost, quadcopters strapped with some C4 and a wire out the front that, when hit, will trigger the explosive. And there's some pretty horrific, very, I don't know, eye opening, I guess, videos of tanks and so forth being effectively neutered by these low---cost Chinese drones flying in and really having air superiority over the battlefield. I think the next thing you're going to see imminently are these quadcopters being controlled by AI. And it's a pretty scary thing to think about. I think there's a black mirror episode on this of what they're called slaughter bots, where, you had a little quadcopter that was controlled by AI that would just hone in on someone and just like a suicide bomber, come in and blow them up. And I think this is going to completely transform warfare. Obviously, it's all about countermeasures and counter countermeasures and so forth. And maybe there's going to be some way in which you'll be able to shoot these things out of the sky, et cetera. But I do think, certainly, you are going to see a huge application of AI on the battlefield.

The other thing that's happening at the same time is, you're seeing quite some scary videos coming out of China about these robotic dog platforms, which are being produced en masse, in the 1000s, where which are being, starting to be deployed in war with a gun put on top this little robot dog. And instead of sending the troops over the southern side of the hill to capture a particular position, you'll send a bunch of robotic dogs with guns on them. I think certainly the future of warfare is manufacturing supremacy, and it's going to be AI powered, whether it's quadcopters, whether it's robotic dogs, whether it's jets that are being controlled, and the dog fighting is no longer done by humans, it's done by AI. And quite a number of these platforms look like they're going to head into the commercial and consumer space. I think there's another one called Neo, which was a humanoid robotic platform, was even announced this week, which is supposedly going to be in the home and doing your ironing and doing your laundry. Although it's going to be pretty unsettling for a few people to have that in your house at night time, and you might hear a bump in the night and go, is that the robot platform hacked and is going to come after me with a knife from the kitchen? I don't know, but I think that's going to be a big transformation that I think is happening right now on the battlefield.

Erik: Well, Matt, I can't thank you enough for another terrific interview. Before I let you go, though, I want to come back to your company, [freelancer.com](https://www.freelancer.com), which is ticker FLN, Fox Lima November, on the ASX, Australian Stock Exchange. You mentioned earlier that the market is going to be for systems integrators. I would think a lot of those systems integrators don't have to be great big, you know, big six firms, but they're going to be freelancers, like the folks that can be hired through your website. So if I want to have these things, if I want to have AI chat bots on my website for my small business, do I go to Freelancer in order to do that? And if so, do I just work directly with freelancers? Does your company do anything to provide a filter or a reference guide to tell me who I need to hire and what they need to do for me? How does this work?

Matt: Well, absolutely. I mean, just as we saw businesses transform with web development and then subsequently transformed with app development, the next big thing is AI development, and you're going to go to [Freelancer](https://www.freelancer.com) to get these apps, these AI apps, kind of built for you, and whether it's your AI powered customer support, AI powered sales and so forth. You can just go to the homepage of [Freelancer](https://www.freelancer.com), and there's a whole section on an AI services marketplace where you can get AI appointments that are developed for you by a freelancer, or an AI lead generation agent developed for you, or customer service, or whatever it may be. But I think one of the greatest transformations that's going to happen over the next year or two is going to be businesses, small and large, transforming parts of their operation with AI agents. And where do you get that done? You'll get that done at Freelancer, just the same place you get web development done and your app development done.

Erik: We're going to wrap it there for this week's special episode of [MacroVoices](https://www.macrovoices.com). We'll be back with our regular show format and Patrick Ceresna as co-host next week.