

5.6 Sol + Superapp Release, AI 2040, Tibothy Joins, New Robot Hand Alert, Meta Releases Muse 1.1

147 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=S8DGGHYV-YY](https://www.youtube.com/watch?v=S8DGGHYV-YY)
<https://www.youtube.com/watch?v=s8dgghyV-yY>

SUMMARY

The discussion covers a wide range of topics in the tech and AI landscape, focusing on recent developments in AI models, their applications in business, and the evolving dynamics of consumer brands. Key highlights include the launch of OpenAI's GPT 5.6, advancements in AI-driven advertising, and the ongoing shifts in consumer behavior and brand strategies.

- OpenAI's GPT 5.6 introduces significant improvements in coding, cyber capabilities, and multi-agent collaboration, enhancing user experience and efficiency.*
- The AI landscape is evolving towards more accessible tools, allowing users to interact with models in a natural way without needing extensive technical knowledge.*
- AI-generated content is becoming a staple in advertising, with brands leveraging it for static ads and video content to increase engagement and conversion rates.*
- The consumer goods market is witnessing a generational shift, with brands like Coach and Ralph Lauren outperforming traditional luxury brands like LVMH.*
- Live shopping has not gained significant traction in the U.S. compared to China, where it has become a native consumer behavior.*
- The integration of AI in demand forecasting and inventory management is revolutionizing how brands operate, leading to more efficient supply chains.*
- There is a growing interest in hyper-personalized advertising, with brands exploring the use of AI to create tailored experiences for consumers.*
- The potential for infomercials and long-form advertising is being revisited as brands seek innovative ways to engage audiences.*

I see a large IPO on the horizon. You're surrounded by journalists. Hold your position. Overnight success. 3, >> right?

There's misinformation clearing order inbound. >> Let's just roll. >> We are surrounded by journal. Hold your position. >> Come on. Get up. Trust the experts here.

We are experts. Founder Mal. I see multiple journalists on the horizon. Stand by. UAV online. >> Blaze.

Double blaze. Triple blaze. Double kill. Fight is w team deathmatch.

We are experts. Triple blade. Let's just roll. Right. Marky clearing order inbound. >> >> Get up.

>> You're surrounded by journalists. Hold your position. >> Strike one. >> Strike two. Activate golden retriever mode.

>> Mark clearing order inbound. >> 5. I see multiple journalists on the horizon. Stand by.

YOU'RE WATCHING TVPN. >> Today is Thursday, July 9th, 2026. We are live from TVP Ultradome, the temple of technology, the fortress of finance, the capital capital. Let me tell you about ramp.com. Time is money. Save both. Easy to use corporate cards, bill pay accounting, and a whole lot more all in one place. I need some soundboard. Here we go. Yes. Today on TVPN, we're talking about model mayhem. Everyone's launching new models. Slow summer, but not for the AI race. You got XAI unveiling Grock 4.5, the first model spec built specifically for coding and AI agents developed in collaboration with cursor. Talked about it a little bit yesterday, but we have some more benchmarks. some more discussion on the timeline about where this model fits in on the paragon frontier. Also why it might be outperforming so well on cursor bench.

lots of debates there. Meta announced Muse Spark, a new agentic coding model with Mark Zuckerberg returning to X for the first time in basically a decade. Three years ago, he posted one joke post about launching threads, but he has not been an active user, but the AI vortex sucked him in and he's got a post. >> Oh, I think he's an active user, John. >> You think so? >> He's just not an active poster. >> He's just not an active contributor.

>> You're calling him a lurker. >> I'm calling him a lurker. >> You're calling him a lurker. >> I'm calling him a lurker. I think he's absolutely glued. >> You think so? >> I think so. >> You really think so? >> I think so. >> I feel like I don't know. So busy. So much other stuff going on. I feel like he I feel like most people >> the busiest people I know. >> Okay. >> Are not active on X.

>> Yeah. >> But they >> they are on X a lot >> sometimes. But there are there's a different class of person. You can just quiz you can just quiz them. screenshots come to them via Slack or or via text message because they have a team that's monitoring the timeline and then is delivered. This is the important this is the important stuff. >> They're calling him Mark Lurkerberg. >> but the other big news, Open AI just released GPT 5.6. Let's go. Let's a new general purpose model with expanded coding and agent capabilities alongside GPT Live which we talked about yesterday. a new real-time interactive voice experience. reactions are great to 5.6. Bunch of interesting details here. you had people have been identifying that while there is a frontier and there are a just a few companies that are actually on the frontier. the the the frontier is spiky and they have different flavors to them and different different reasons to pull different tools off the shelf.

people are drawing analogies between Fable 5 being some, you know, recluse genius and 5.6 being a, you know, a collaborative co-orcker that you love chatting with or something like that. >> I said I don't know how else to describe it, but Fable 5 is like Kendrick on Good Kid Mad City and 5.6 Soul is like Chief Keef on Finally. >> Now it makes sense to me. Thank you. So, I just wanted to put it into, you know, 2010 hip hop hiphop like

terminology.

>> Really, really clear there. Thanks for clearing that up. >> I mean, the funny thing is that will be >> very explicit for like a hundred people in the whole world. >> Yeah, probably. >> this one's for you. >> the fun the the most interesting benchmark to me has always been Arc AGI V3. We've interviewed the team over there many times and had a lot of fun understanding what goes into that that benchmark. And 5.6 soul scored a massive 7.78% which is tiny considering that the whole point of arc AGI is that a human should be able to get 100% on it and basically any human. So it is a true test of AGI in this sense of you know can you give this test to just actually anyone not you know the the the crazy math projects the crazy hard programming projects the hacking all of that stuff is very economically valuable of course but there's a more interesting question where you know when there's less of a spiky frontier and there's just this question of what is something that anybody can do that AI can't because we've been searching for those and the ARGI team has done a fantastic job building out these puzzles that AI has historically struggled with. ARGI one, the model sort of climbed. Two, became a little bit more complicated and now three, we're starting to see glimpses of progress, although 7.76% isn't 99%. We're nowhere near saturation, but it's still a huge jump.

Opus 4.8 had 1.5%. So, GPT 5.6 six soul is showing more generalization, more spatial reasoning, more puzzle solving abilities. So fun fun stuff. I am trying to refresh my timeline, but the blog post is also very very fun because it includes games. I'm a big fan of the the GPT 5.6 launch games. I got immediately sucked into the to the the sailing mininame which is like very high fidelity but also delightful to actually play.

>> Should we play it? >> Yes, we should definitely play it. Yeah, Saltwind. You you you guys play it. I want production team to see what they can get. I think my time was 25 seconds. >> And is this hosted on a on a site? I think this is I mean this is hosted on the open blog, but I think that the idea is that you could vibe code this in the latest GPT 5.6 in the app in chat GPT and then deploy it and have someone Are you trimming the sails appropriately because it looks like you're losing speed. You're losing wind. It's not working.

>> I'm going to smoke you. I got 25 seconds. Wow. Amateur hour over here. Look at this. >> Boost. >> Yeah. Yeah. well well the whole the whole game which you probably missed is that there is a little bar there where you have to trim the sails to be in the sweet spot of the wind while you're turning. So as you turn see the bar there's a recommendation for where you put the sails. You got to keep that line in. See it's moving over. You got to you got to press the down S. Yeah, exactly. Keep trimming those sails while you steer the ship.

>> This stuff is very very fun. I am trying to open >> one interesting data point from the live stream which was just an hour ago. They said already soul has been transforming our research program as one example GPD 5.6 soul autonomously post-trained 5.6 Luna. >> Yeah that's fair. >> A lot of people are having fun with that. Dylan Field says a lot of people want to compare Fable versus 5.6 Soul. This is a mistake. their apples and oranges. Despite all the research achievements, we are still very very early in exploring the tech tree for model training.

>> Cool. Sorry, I'm just getting set up again. what else is in here? oh, yes, I I I do think that didn't didn't, Dylan Eper Scott write something about this? What was the the the the the essay he wrote about interactive memes

and this idea of like generative AI enabling these vibecoded mini games? Like we've been seeing a bunch of them with like the copy bear simulator, the coconut simulator where it's something that's just a joke that's funny for like a few people, but and normally you would instantiate that in a in a tweet or maybe if you were getting really crazy, you'd do a Photoshop edit of a meme, but now you can go and create a full mininame, something that runs in the browser and soon something that runs in Unreal Engine and can actually be distributed on the Steam store. We're already seeing that with like the data center simulators and all these funny simulator games that are going on Steam.

all this all the all the advances in the coding models certainly speeds up the ability to actually deliver polish software. I'm I'm particularly excited for like >> Yeah. Dylan's title was the future of entertainment is interactive. >> Yes. Yes. But but yeah, that that's part of what I honestly love about AI is there's a lot of things you can make now that never would have made sense to make because they would have taken you >> four days and it was good for like a small laugh and now you can do it in four minutes and it's just fun. Yeah, I I I think there's going to be there's if you have some sort of like small custom some sort of custom functionality in your business. it feels like there's >> is this the David Senra simulator?

>> Why is this David Senra >> late nights in a Miami abandoned apartment complex >> in 2015 just recording podcasts and reading? >> This is very creepy like horror back rooms liinal space game. Stanley Tang, co-founder and CPO over at Door Dash, says, "I have an insane magic trick that so far none of the models can figure out, including Mythos. It's a bulletproof trick that I've shown to a 100 plus people, including magicians that couldn't figure it out. It's not anywhere on the internet. Only way to know it is through first principles reasoning."

>> Told everyone I'll believe in AGI when it can crack this trick. Well, GPT 5.6 just did. >> How? >> I want him to I want him to actually open like Okay. like give us now that now that a model cracked it >> because I feel like a lot of magic tricks are like slight of hand. So is he uploading a video or something like >> Yeah. So John Palmer says I have a hilarious joke that so far none of the models think is funny. It's a bulletproof joke that I've told to 100 plus people including comedians and no one laughed. It's not anywhere on the internet. Only way to know it's funny is a first principal sense of humor.

>> Told everyone I'll believe in AGI when it tells me a joke. The joke is funny. Well 5.6 six just did. >> Huge, huge news. Huge news. >> Huge. >> yeah, people are going back and forth. GPT 5.6 is a Porsche. Fable's like warp drive. I had a different experience. Fable is an F1 car. 5.6 Soul at Ultra is a Tesla Model X Plaid. Does it find things that Fable misses during plannings and coding? Yes, most of the time. But for the hardest problems, does Fable routinely find things that Fable that 5.6 doesn't? Also, yes, some of the time. is 5.6 way faster and affordable? Yes, with an unlimited token budget. What am I currently using 95 95 plus percent of the time? GPT 5.6 from CQI Chen. So, interesting take that the Parade of Frontier is alive and well and everyone's duking it out for their slice of the the AI opportunity. Very interesting seeing how the how the market share is shifting while during a time of acceleration. you have multiple companies that are growing revenues, even accelerating revenues, while market share is declining because the overall market's growing so fast that that if you're only growing at 300% and someone else is growing at 400%, you're losing market share, but you have like one of the greatest businesses by modern metrics. Very very interesting dynamics in AI. It's also funny

because yesterday with Ben Thompson, you were like a slow summer and then in the span of 24 hours we get four five muse 1.1.

>> Yeah. I mean this I don't know. This isn't this isn't as dramatic as the AI talent wars. It's not as dramatic as Yeah. Yeah. this is this is new technology. and and there's only so much to there's only so much of a take to be given around these things. Although AI 2040 launched today, the sequel to AI 2027, that's something that's more of a thoughtprovoking piece that you can debate and interrogate and and talk through. I'm sure we'll go through some of it because they pose a couple interesting couple interesting ideas of where where AI might go and where they want it to go and how they want the industry to develop. sort of advocating for a slowdown generally, but it's it's an interesting way they puzzle piece all the different geopolitical chips on the table around. what else? of course people are joking about the lead is widening because the the anthropic and open AAI version numbers over time.

GPD6 is predicted and it is it that the the model numbering we were talking about this this morning that the numbers they sort of don't mean anything anymore. Do the model numbers mean anything in particular? It used to be the the model number was the pre-train and then the and then the version number was the post-train, but then that sort of got flipped around and now it's just like are you do you feel like you're competing at a four class or a five class. So I wouldn't be surprised if we saw like Muse Spark not not release Muse Spark 2, but Muse Spark 6 or five and jump straight. I mean, Samsung wound up doing this where they jumped to the year like sort of like the car manufacturers where you know, there's a 5 series BMW, but then there's also just the 2027 because that's the actual model year that's relevant.

>> 2027 5 series. >> Yeah. Which is sort of odd. and we're sort of like duking it out between those. Do you have >> Yeah, I mean I think post-reasoning models you just have like a different way to scale the models besides just pre-training. So it's hard to bake that all into one number that like is you know evocative of of both those like two ways. >> Yeah. So the number is is becoming closer to the year in the in the second decade of the 21st century.

Basically it's just like is this on the frontier in 2026? You'll probably see a six by the end of the year in front of the models that are leading in in the year 2026. Something like that. I'm very interested with Google strategy because the the rumor is that 3.5 Pro will be coming out this ne next week I believe. But I it was it's very odd going into the Gemini app right now and seeing that there's 3.5 flash but then you have to go back to 3.1 Pro.

I think 3.1 Pro is the most advanced model, but they default you to 3.1 flashlight. And I would expect them to jump just forward to four, but I think that they're going to do 3.5 Pro, but it's been a little bit of a slower cycle there. as silly, I mean, the OB obviously all these numbers don't really mean anything. They're marketing terms. but they they I I still think they do actually stick in people's mind and so there should be some strategy around them. but anyway, before we move on to our next story, let me tell you about the New York Stock Exchange.

Want to change the world? Raise capital at the New York Stock Exchange. so, Mark Zuckerberg is on a press tour. He's talking to the legacy media for the first time in a long time. Andrew Bosworth, the CTO of Meta, also

did an interview, with the, head of the Atlantic. dug into some of the launches around the glasses and then also had a whole discussion in that podcast around the goals of the keystroke logging thing and it was it was interesting. I mean it was framed as like you know like a tough interview around surveillance in the workplace and it certainly the headlines were very scary. I don't know where I sit on it because I've I I kind of always assume that everything you do at work is logged in the sense that like if you're on a work computer and every web page you visit is is going through the network and monitored for traffic and security purposes and all the code you write and all the emails you write and all the documents are stored in a shared document. It's like it's it doesn't seem that crazy to go to keystrokes because everything is already so monitored. but he was framing it as more of an experiment, something that they weren't sure was going to pan out.

something that they allowed everyone in everyone at Meta. So there were there were certain sections of the workforce that were by default opted out. So anyone who was working on like confidential or sensitive information was opted out of that program by default. he said he himself, Andrew Bosworth, was opted out of that program because he has a bunch of legal holds because they're getting sued all the time. So, so they can't be recording everything I guess that he's doing because then that would be admissible in court. And so all of a sudden the the lawyer who's suing him would be would say, "Okay, great. In the email, you said you know, we we don't want to do this, but before you >> Well, let's see your writing process.

>> Exactly. Yeah. let's see what you what sentence you typed and then deleted like what word did you use before minimal impact did you say medium impact or whatever you know so so he was opted out and apparently I I think all of the Meta employees who were part of that program were able to were were able to just turn it off indefinitely like you could toggle it on and off and and the idea was that they wanted to collect information on how work plays out over like a 12 to 18 months per 18month period and they couldn't get that from any sort of data labeler because they needed to have very high-skilled workers actually chopping wood on projects for a long long time to see how projects go from start to finish. so so basically like how do you compact the longest possible rollout not just a like a single chain of code but an actual series of of meetings and decisions and tradeoffs and everything that goes into making a decision in a white collar workplace like how do you actually reason through all of that.

it's it's hard to to distill that from just oh well the code got written this way so that's the right way to write the code that that might the code might have gotten written that way because a lawyer said hey oh we have to do this and then the marketer said oh well well you know we have an activation with this person so we need to integrate it this way and then the business people came in and said oh well like the margins will be better if we write it this way and so it's not entirely first principles software engineering all the time when you're actually building real products. So, interesting to see him sort of, step into the, you know, a tough interview and and and, and and sort of lay out his his side of the story. But Mark Zuckerberg is in Bloomberg today, pledging aggressive pricing with Meta's first payto-use AI, which is a funny framing for just an API for a model.

but that's the way Bloomberg put it. in a crowded market for AI tools, Mark Zuckerberg wants to win on price. Meta Platforms unveiled a version of its most advanced artificial intelligence model, Muse Spark 1.1, that includes a new paid tier for developers, marking the first time Meta has charged businesses for access to its

models and providing a new revenue stream. It'll be among the most affordable options on the market, Zuckerberg said in an interview ahead of the release. quote since this is not an open source model this is I think the first time that we're doing a real serious API referring to the API used to access Meta's AI and the pricing is going to be very aggressive and attractive makes sense I mean they own the data centers they're very efficient at building data centers they should be able to serve a model efficiently the new model standout improvement is is in its agentic capabilities the meta chief executive officer said agents are a big theme of AI this year with the label will apply to systems that can complete multi-step tasks on behalf of the user. Zuckerberg described Muse Spark 1.1 as having quote state-of-the-art or very close to it agentic reasoning and tool use. The model is also greatly improved when it comes to coding and Meta employees are using it internally to build products and features for various apps.

>> Yeah, my big question is how quickly do they move all of their internal workloads onto their own models? they're buying they're buying they're getting access to models through Google Anthropic and OpenAI. >> I think that a lot of companies will look to Meta's own actions as a way to >> basically validate whether or not they should be using this model themselves, right? Because it was just, >> you know, within the last month that Google had said like, "Hey, we don't have Yeah.

>> capacity. We don't have enough capacity for all of Meta's demand for our models." Yeah. >> And so, yeah, they can't they can't get enough AI elsewhere, at least from some providers. And so, how much of their workloads will they be able to run themselves is the big question. >> There are a bunch of bull cases. First, I'm going to tell you about Figma. Agents meet the canvas. Your AI agents can now create and modify your Figma files with design system context. So, yeah, Meta was one of the first companies to sort of, reportedly be token maxing and have a leaderboard and all of that. if you have your own model and your own data centers, the incentive to token max is much much higher because you're just paying the electricity on the cards that you're already depreciating. So, you should sort of lean a little bit back into that. Not that you want to be fully token maxing, but you do want your employees using the tools that you've built. as efficiently and as effectively as possible. It's just way cheaper to explore when you're not paying margin on another closed source model and you're you're not paying anything else and you're actually improving the model. so makes a lot of sense for them to roll this out broadly. the the interesting take that Ben Thompson had which we didn't get to yesterday because we w up spending the whole interview talking about Xbox. But the the interesting dynamic is that when you are willing to sell API access, you're willing to sell compute directly and then you're also using your own tool internally. It creates this economic incentive internally that you you you have an incentive to always go with the most profitable the most the the most economically efficient outcome. That can be very good for business very good for the investments that they made. The the trick is that you can wind up in a little bit of a situation where your business team or your your enterprise sales team goes and sells all your compute capacity or all your chips and then internally your team is frustrated that they're not making enough progress. So there's a little bit of a dance there, but in general it it's a forcing function on the internal use of their tools to say, "Hey, wait, why why is someone willing to pay five times as much than what we're willing with the value that we're creating here.

Like we spent a billion dollars on energy consuming our own LLM and someone showed up and said wait we'd pay you 5 billion for that same compute power to run a different model and do a different task. It's like why is

their model not economically valuable internally? That would be the question. the flip side is that they do have low cost. So they should be able to say, "Oh yeah, we actually did we yeah we we we inferenced Muspark 1.1 internally and we improved the ad model and boom, we made a bunch of money."

>> And and these are the same trade-offs and decisions that every lab is having to make is how much how much compute do we allocate towards research towards internal use towards to the API to subscriptions to free plans etc. Yeah, there was that funny semi analysis deep dive into anthropics forecast and in there I mean some staggering numbers really really optimistic but the the flip side was who was Ed Zitron was was taking shots at the fact that they had >> EBTI >> EBTI earnings before >> training >> training inf no training in training inference and >> and everything no earnings for training interest and taxes. and what was odd about it was that Ed Zitron was was was saying it's like the new comm community adjusted EBITDA and it is always odd when a new non-GAAP metric pops up. in this case, I think it makes a lot of sense because training runs do fit a depreciation profile. It's a little bit different. I don't know why you wouldn't just put it in in depreciation though, like just figure out how to account for training runs through a depreciation schedule and then maybe it's like a non-GAAP depreciation metric, but it's still in there instead of trying to get everyone up to speed on a different a different like sounding phrase entirely. also, you could just do IBITRA instead of Ebida, like earnings before interest, taxes, and training, TRA. Yeah. and that might be a little roll off the tongue a little bit easier, >> but I think semi analysis likes to be a little cheeky.

>> That's true. That's true. but the but the other funny thing is that Ed Zitron is taking shots at at this like oh like the the EB EBTIT is like so ridiculous, but like in that forecast they had like net income of a billion dollars in a quarter. So it's like it's like okay well yeah you have this like funny metric that you could value the business on to get crazy and valuation but if the business is making money and and actually like generating net income you're not in a disastrous financial position. So you know you you could debate about the valuation but you shouldn't the the whole idea of like this oh it's like such like it's it's a very very different discussion than than we work which was not cash flow positive which was not net income profitable and and was using those terms to sort of skirt around the losses that were acrewing from the business. so, >> yeah, I was looking back at, Ben Thompson's, earnings transcript or a script that he wrote for for Mark Zuckerberg. He has a good, segment on why AI matters.

>> Ben writes, "For, forgive the long preamble, but this is necessary context for me to properly explain why AI is so important to Meta and why I'm making the right choice to invest so heavily in both talent, infrastructure." and he goes on and on and on, but he says, "What I've come to realize as I've embraced our status as an entertainment provider and ad purveyor is that our nature as a digital business, non-withstanding, we are remarkably well placed to thrive in an AI era. Remember what we learned about humans? They are obsessed with other humans and they want to connect with them. That obsession and desire only going to increase as we interact more and more with AI. AI is going to make our properties more essential, not less. Moreover, AI is a productivity tool, but productivity is not the end- all beall of the human experience. I've talked over the last year about building super intelligence that helps you get things done. But that's a business story. What we can do uniquely is gives give people the experiences they want from connection to entertainment to shopping when they are off the clock. The fact that we are investing in AI but not selling solutions to businesses is actually one of our business biggest advantages. So, of course, this is just a a sort of fanfiction for an earnings transcript. Meta

is in fact selling to businesses now. but who knows over time how big will the API business be relative to how much value they can unlock across their broader Yeah.

>> business with all of their infrastructure. >> Well, let me tell you about console.com. Console builds AI agents that automate 70% of IT, HR, and finance support, giving employees instant resolution for access requests and password resets. we have the perfect guest to talk about all of this with because Eric Sufer from Mobile Dev Memo is with us today in the TDP Ultradome. Let's bring in Eric. How are you doing? Good to see you again. And congratulations.

Is it Dr. Seufort now? >> No. Masters, right? >> Can you hear us? >> Can you hear us? Okay, let's bring him back in in a second because I want to hear his take about Meta's vertical integration specifically with regard to their image model because their image model has the potential to feedback in the ads product where they are getting a signal from what ads are performing, generating new images and then using that as fine-tuning data and reinforcement learning data for the image model. And it's a very different cycle than what you see in shap images where you know they're trying to be useful, educational, sometimes funny. Same thing with nano banana. But meta has a different job to be done, a different process and a different different goal because that model although it might wind up being something that people use just to post on Instagram and people might just use it like any other image model. The real killer application of that is in the ads manager. And we also have Sean Frank coming on later in person to tell the more business side of that story. he's built Ridge into nine figure DTOC brand.

Obviously has been deeply ingrained in the meta ecosystem for probably over a decade now and and can comment on everything that's happening in AI generated advertising. But I believe we have Eric Sufer lined up now with audio. We'll bring him in to the TV. Eric, how are you doing? >> Hey guys, thanks for having me. It's good to be back. Sorry about the the technical >> mishap. It happens. It happens. >> it's great to have you here.

Question, did you did you rebrand mobile dev memo? >> What? No, >> to Heracles. >> To Heracles or is this is this an error on our side? >> This is an error on our side. It's mobile dev memo. Herac Hercules is the fund. Is that right? >> Yeah, the font. Yeah. >> Yes. Got it. anyway, we were just talking about Meta. They had just launched, Muse Spark 1.1, their LLM. They're selling that over API. They're also selling some compute. But I want to I want your take on the image model specifically and why that is an important technology for them. Why vertical integration makes sense there.

Why specifically is is their image model like it it feels like it has a different business case around it than a nano banana or a chachi images. >> Yeah. So I think this is brilliant, right? like this actually gives them the narrative firepower that they need to to sort of undermine the skepticism that investors feel about the AI investment. Right? So like I was at this dinner I do these dinners a lot like these ideas dinners >> like a a research company will bring me in and talk to like a bunch of their hedge fund clients and like I was talking about you know why these investments that Meta is making right now are bearing fruit right now. like 33% advertising revenue growth last quarter on \$55 billion in revenue.

That's incredible. Like you look at Google search was 19%. Amazon which is much smaller was 24%. They're outgrowing everyone except for AppLin and Reddit. And people don't believe it. And I asked somebody okay what would it take to convince you that these investments are actually productive in this moment in time and they said 40%. >> 40% growth on 55 to 60 billion in revenue. Where did that number come from? This is pulled out of thin air. This is what they need to do. They need to be able to point to something and say, "You see that ad that was created by our AI investments." When you talk about GEM, Gem is a foundation model.

MetaTrained a foundation model for ranking. That's important. But you can't see it. It's hard to convince. First of all, the research like, you know, a lot of the research analysts, >> they're really smart people, but they operate in this paradigm of like spreadsheet says this. I get it. I can understand why ranking investments would actually be really beneficial for the company. But I put a number in the spreadsheet and it spits out something. that's the tool I have to work with. And like these are really smart people. I think they conceptually get why these are good investments, but they have nothing to sort of tie it to that's quantitative. I think when you can point you can point to the ads was created and you say look this ad was created with data that only we have. It's the image model that we built the foundation image model that we built that is trained trained not fine-tuned but trained on our own data can be verified against that. We've got our own custom evals.

Everything about the training was built with data that only we have. no other Frontier Lab can can can fine-tune their own models for this use case. Only we can do that. And then I think if they can point to the output and they say that ad that you saw in your Facebook Instagram feed was created only as a result of our ability to train on this data that only we have then I think you can kind of make the case. So I think bundling those two integrating those two things together is like the really smart move. And I I I I kind of, you know, I understand like they, you know, they restarted the AI efforts and this is the whole, you know, this is the MSL rebrand. But my sense is like this might be more convincing than anything that they've been able to say with the ranking infrastructure and the transfer learning infrastructure. I'm talking about Lattice. I'm talking about GEM. So, we'll see. But my sense is like if you can actually point to some output and say look, this only exists as a result of our ability to train this model on this proprietary data that no one else has. My sense is you can make the argument more robustly.

>> And if you ask advertisers what is the bottleneck to spending more on meta, they will almost always say it's great creative. And so there's yeah, >> there's a there's a path to sort of just like unlocking removing that bottleneck so that the constraint just becomes how much revenue do you have and that will just become a proxy for how much you can spend with us. >> Yeah. It's like what's your bid? Well, your bid should be the value that you get back. I mean that's like the whole point of a second price auction is that you should build your bid your true value because you you're going to make money if you win.

>> But like the thing is like they built like a lot of their initiatives right so gem foundation model for ranking Andromeda is is is a whole system for doing retrieval and then lattice which is transfer learning but like the whole point of Andromeda was they were reacting to a lot more creative being deployed. Now the creative being deployed though was created with third party tools right this creative being deployed is being built without the benefit of the actual performance data.

If you actually want this to work, you need to train it on the rorowaz. You care about the end result. You care about what the person does when they go to your website or your app. And that no one has that except for meta at that volume because they get it passed back to the cappy and the pixel, right? And so their ability to build this foundation model I think really unlocks a lot of value and you know I think you'll probably see that show up in you know in the in the revenue growth. Now do they hit 40% to satisfy these these needs of these hedge fund people? I don't know. But like my sense is if you can point to the output and you can say look this addresses the core bottleneck which is we need a lot of creative but it needs to be built with the knowledge of what actually drives the outcomes and not just a bunch of variations that clog up the system.

>> I want to push back on that idea that no other lab can do anything like this. What about Deep Mind? What about YouTube Nano Banana V3 video generation? it does feel a little bit farther off. And also, it feels like the hedge fund analysts that you're talking to aren't asking the same questions of Google's investments in AI because they have such an incredible business with Google Cloud Platform and they're able to strike these massive compute deals. So, they have some offtake there. But how is how is the situation different in Google?

>> Because they don't have a history of tilting at windmills. They don't have the they don't have an albatross around their neck, which is metaverse. That was a misadventure. It cost a lot of money. It never resulted in anything meaningful. Right now, I would actually make the point that that whole rebrand was just a distraction. It was a smoke screen. They had to get away from the whole Facebook Files thing. And that took everyone's eyes away from, you know, that that scandal.

>> And was it worth it? I don't know. If you achieved that, which they kind of did. They kind of did that with the Meta rebrand. maybe it was worth it, right? you know >> also also the name even if you ignore like the actual investments they made in in metaverse and like horizons and things like that like Meta Platforms is is the metaverse like it is the place that people exist online. So like the name makes sense even if you ignore that and it makes sense for the strategic reason like you said to get away from the Facebook files. So like they could have just rebranded and never done the metaverse and they'd be in a much better position, right? to have that >> sort of sort of >> but you can't really do the rebrand unless you tell the story or else everyone accuses you of what you just described, >> right? Yeah.

>> I mean, you got to put your money where your mouth is. >> Yeah. but on on video generation in particular, do do you feel like we are further out just further away from that? I was I was demoing the latest VO model and it's good but it's still clockable as AI generated. There's still I had some cars spinning around and you know it's three cars and then it's four cars and then it's two cars and they're sort of melding into one another. It's an incredible video model definitely state-of-the-art but I don't know that that that's ready to be deployed in YouTube across a ton of video ad impressions. So do you have a timeline for that or do you have some thoughts on when that will actually be important to the business? Because you have to imagine that Instagram video ads perform better than just image ads and so they'll try and do this. But what how are you seeing the AI video advertising model evolve >> here? You got to look at Bite Dance.

What Bite Dance is doing is incredible on this front, right? So they put out this paper a month ago. I did a

summary on Twitter and LinkedIn, but what they're doing with Tik Tok shop is these real like you know basically photorealistic 3D avatars that are selling stuff, right? Like so infomercials and they've they've built custom models to build those ads and those are all ads. A lot like I mean not all of it but like if you go on Tik Tok and you're looking at the Tik Tok shop stuff a lot of it is AI generated >> and so what they they in this paper that I summarized like they invested in you know in in essentially like fine-tuning this model to make sure that there was no collision with the hands. What they were finding is that like so when you get in that uncanny valley situation where people can tell it's AI then they turn off. Like there's a lot of research that's been done into this. If people know that it's AI, they penalize the ad.

But when they don't know it's AI, the AI ads outperform the human creative ads. It's really fascinating. But so what they found was like when you saw the collision between someone holding something and their hand and the object, then people the the the click the rate drop, the conversion rate drop. But so what they did was they fixed that. So they built this whole like visual interpretability model that just focused on that with it with with an expert like in the model. And so it just addressed the hands. And so like if but that's use case specific. you needed a general purpose model that's going to build photorealistic video. We're probably pretty far off of that for like all ads of all types. But I think with something like, you know, okay, well, we need human photorealistic kind of like infomercial style. I think you could get to that point now. And maybe we were there now. And maybe Bite Dance is there now. But I think it takes a lot of investment, right? I mean, they had something if I remember correctly from the paper, it's been a while since I saw like 12,000 hours of live human product interaction. So, I mean, it takes a lot of data to do that, right? And so, you know, it's just it's whatever you want to invest in. I think if YouTube wants to do that for a general purpose photorealistic video ad tool, we're probably pretty far off.

>> Mhm. >> how do you think that the market will react if Muse 1.1 is like a very much like a base hit on the API where they they get some com, you know, big companies to move over some workloads, but it's not, you know, this runaway hit. And and I say that because so many models that have been good, not great, have a little demand just because there's a lot of demand for AI, but but they don't sort of like have these sort of breakout revenue charts.

>> Well, you know, they clearly are going to be very aggressive on the pricing. I mean, they talked about that today, right? And so my sense is like what you're going to start seeing is that people don't need to operate at the frontier. and you have a lot of use cases that work just fine with and and and basically are just good enough, right, with some legacy model. And so it just comes down to then, okay, well that's commodity and so is it priced like a commodity? Like like if you think about like and also like I think we're going to see a lot more like people relenting from needing to be on the on the frontier when they've built stuff using a model that then gets upgraded. And the whole idea there is like well am I going to upgrade this tool because I have to adapt it to the new model, right? like if I'm if I if I you know because essentially like the model name like if you if you use like vertex AI right you just got a model name as a variable like you're sending this you know system prompt to to you know Google but like it's just a variable you could swap that out in 30 seconds but the the fact of the matter is you're sampling from a new distribution if you do that and it's going to change the output it's going to qualitatively change the output and it might change it in quantitative ways that like with retention and engagement and so the thing is like if okay well now we're talking about this big cycle it's a new product

development cycle because I have to adapt this product that I built to this new model and the output that it provides, right? And maybe it was working perfectly. It was working exactly as I expected it to before. Now I've got to invest a bunch of hours, a bunch of engineering time in adapting it to this new model. So even if it's just a swap of a variable name, it's still a whole lot of testing, QA, determining like how that impacts long-term potential. Going to do AB test. So my sense is like you're going to see a lot more people just saying, "No, this works fine. This is perfect."

And and getting more getting like more robust output. Let's say the token price is exactly the same. Getting more robust output wouldn't benefit me. Why am I going to invest the resources into adapting to the new model? >> Yeah, but isn't that so so if they're going after workloads that are running on old models that are working fine >> and it's a lot of work to switch over. >> It's not a lot of work to switch over.

>> No, it's it's some risk. >> It's fast to switch, but then again, you have to go through this process of like QAing and running it through your own benchmarks and all this stuff. Like >> the question is like >> how I I I just don't know. Like I I'm I'm thinking about a scenario where like a year from now we're sitting here and like Meta has a \$3 billion AI API business and the analysts that you're talking to are like >> that like to me they're like >> that doesn't get you to 40% >> year-over-year revenue growth on the on the business overall. So it like doesn't solve at least what those analysts in particular were talking about. and like zero to three billion on like a new business line would be crazy and is like, >> you know, only been done a handful of times throughout history over the last few years, right? Yeah.

>> So, >> I'm just saying like >> we're we're there's such big numbers now that there's a possibility where you have like this incredible breakout revenue growth, but it doesn't actually move the needle enough that that the market still says like, "Hey, we're we're not super confident about about about like the next capex cycle, right? This 2027 numbers that are coming out." >> Well, that's and that's the problem with this whole business line in the first place. Like, I think it's a mistake. I think it's a capitulation. And I think you're going to get much more value out of that compute if you apply it to your own core business which is advertising.

My sense is you get better growth. But like the problem is the investors don't buy that right now. Like they've got a narrative issue. It's not a it's not a productivity or a competency issue. It's a narrative issue. And like the problem is like you know and you know you you you cited Ben's brilliant essay from yesterday and the day before about like what suck should say. He should say that like he Ben is totally right. He should say that. He should come out and say look we like senator we run ads.

Senator, we run ads. When he said that, I was at F8. It was like the next month. Every Facebook employee is wearing a shirt that said, "Senator, we run ads." They know what business they are in. Zuck seems to be confused about it. Like, I don't understand. >> Here's a question. Here's here's a big question that'll be interesting. This year, they're going to spend I don't know how much they're going to spend on external models. Like I would say I would I would expect them to spend like maybe like \$10 billion, right? Like something in the range of 10 billion dollars from from Google, Anthropic, and OpenAI. If next year they can say we're not spending money on any external models >> then then that could help them with the narrative issue of saying

like hey like we're we're invest we're basically getting instead of having to give this money to other businesses we're just using our own infrastructure it's a lot more efficient we can like token max we can use way more tokens we can do way more workloads and so that's potentially it's potentially setting but but the question is can they actually move all the workloads that are on these other models to their own models.

>> Well, so that's where you actually do need the frontier, right? Like coding tasks like you actually benefit from having the frontier, but if you're talking like customer support stuff, right? Like that doesn't need a frontier model. That could use like a three or four, you know, sort of like release back model and it'll be it you'd be producing reliable results that you know work like you've measured, you tested and you don't need to upgrade like the customer support or like the chatbot, you know, for customer support integration to the to the the bleeding edge model every single time.

But like coding, yeah, if you're actually using these models to build models, you probably want the best of the best, right? And like what Meta is doing is really actually at the frontier with like integrating agents into the coding workflow. If you look at like their system they built called Confucious. It's like a self-learning agent like that actually helps them deploy better. They've got a whole pipeline for data science and machine learning tasks that helps them decide like okay which which which where should we even apply this where should we even do testing right because that's actually takes a lot of time like you just figuring out what kind of experiments what kind of tests you want to run they built a whole pipeline around that that's all driven by agents right so my sense is like there's where you want the the the top of the line and maybe their own models don't perform best there but like also I don't know how how excited investors are going to get when you say that we cut expenses I think they really need to see the revenue growth at the top line and so my sense is like you that you get more of that by just making the ads platform better. And they've done that. All they need to do is say, "Look, we can," if they could forecast out that growth and say, "Look, we're really dedicated to this. This is what we're pointing everything at." My sense is you could get investors excited over time. You keep printing 33% or whatever every quarter, like you're going to get investors excited after some time. If you start saying, "We're going to compete with Coli. We're going to build an AI pendant hardware." Like, they're not going to get excited.

They're going to think you don't know what to do. And they're going to think you've got all this comput capacity, you don't know how to use it. help us understand the the prediction markets play is that I think the the answer that we landed on was it is just in Meta's nature to copy the new hot thing. So regardless of what it is, we're just going to do it, right? And you can see the history of like these shots on goal with like every new hot thing in consumer, they just build a version of it or they try to buy it. So I think that's the most simple explanation.

John was trying to explain it as like maybe there's some way to do it with like there there there is no dollars and it's just for like social status you know who who can be you know an oracle. I think that that was we got more news that maybe showed that that wasn't the case. but then when you look at again you look at the market like you look at the total market for like gambling and again even if they got a meaningful amount of that market they would not really move the needle in the way they need to on the core business and they would invite all these new regulators that are now saying like you're not only trying to harvest you know my teenagers attention

and making them you know sad about their life you're also getting them like it just feels like it opens up this huge can of worms for no reason.

I It's just a totally misdirected move. Is that really the space you want to go into right now? That is so politically fraught. That is such a political hot potato. Why do you even want to touch that? I would steer clear of that. I would say, "Look, Facebook apps time well spent. You connect with friends." All this gambling. You get addicted to that. Don't spend time there. Spend time in Instagram. It's more wholesome. Why are you going to touch that at all? Are you just going to open the door to more scrutiny? That's insane. I really have no idea why they're even talking about that. It doesn't make any sense. But they've got like, you know, look, they said we're going to we're going to publish a lot more apps. They've got the new pocket app. I mean, it seems kind of interesting. Maybe some of this stuff sticks. I think they could just try a bunch of stuff. But why would you touch the most politically toxic area right now, like when you could just be touching anything else?

>> Yeah. take us through the prosperous society. I wanted that's originally why I wanted to bring you on. I want the thesis. I want to dig into it. because I I I found the it's a three-hour, four-part podcast series. because you've written a lot about it, but introduce it for those who haven't been following along. >> Yeah, Prosperous Society. I started out I wanted to write an economic bullcase for AI. We've heard all of the, you know, bare cases. We've heard all of the, you know, doom narratives around the around the economy. It's going to just basically displace all white collar work. You know, you're going to get Door Dash created with a vibe coding session and so all these companies going to go out of business. And I wouldn't make the case that's probably not going to that's probably not going to happen. And actually, there's a lot of reasons to be optimistic, right? And I think, you know, in writing that it ended up becoming, you know, we live in a in a sort of like very pivotal moment, I think. You know, if you look at the elections, you know, in the sort of the House primaries in New York, you look at what's happening with the the, you know, the New York City mayoral election, you look at what's happening in the LA mayoral election, like there's there's this this this sort of moment where these sort of like impulses against capitalism have become a lot more popular, right? And like there's reasons for that. And you know, you know, I'm not an expert on those reasons, so I won't delve into them. But I think, you know, ultimately it's a mistake to go down that path. And the thing is, like my sense is a lot of the AI or the anti-AI narratives are actually have nothing to do with AI, right? That's just seen as an avatar or like a bogeyman for capitalism. And so what I wanted to do was sort of like anchor this economic defense of AI, this economic bullcase of AI in, you know, the sort of like liberal tradition of of of the Western world. and and and in in in doing so like you could say like you know you could anchor it to these these sort of like these these great thinkers you know the sort of like enlightenment thinkers and and sort of the economic giants that have built you know built the sort of intellectual framework that our that our western civilization is is based upon and people can say look well look you've you've misinterpreted them and so okay well maybe but but if that's not the case and they you're going to make them drop the mask you're going to make them drop the mask and say no that's not my problem with AI I just don't want to live in a liberal economic society based on these western thinkers. And I think if you actually kind of force that to be articulated out loud, you make a lot of progress against the anti-AI narratives.

But the whole point of the prosperous society is my sense is, you know, a lot of these AI investments, they're going to push the economic constraints away from production and towards just distribution, right? They're

going to make distribution the binding constraint. And so because you just have this this this flourishing of of content creation and and and and so when that becomes the actual problem with distribution and these AI investments go into things like ads platforms, digital advertising, you know, REXIS, recommend recommendation systems, then actually commerce the economy becomes much more efficient and it's a really good thing and you generate a lot of value by pushing that binding constraint to the distribution layer and you get then as a result you get a lot more heterogeneous product development because you can actually reach those people economically, right? So like what is the constraint now? It's like well can I reach a big audience, >> right? Can I reach a big enough audience to support a business because well I've just got this kind of like blunt tool.

But if these AI investments go to making recommendation systems better, digital ad systems better, reaching these these these these pockets of people that have these very specific interests that were totally unserved before, then you enable a lot more commerce, right? Yeah. >> And and then and then you know you support this like flourishing of people making this wide diverse variety of goods. You get rid of this idea of like the paro principle. We have to serve you know the the 20% that supply 80% of the commerce. Well, no. Now you can serve everybody and they can pay what they're willing to pay for these things. You reduce consumer surplus sorry you reduce consumer surplus. You just you create this flourishing of of of everyone getting exactly what they want, right?

And that's the prosper society. And so the way I frame it is kind of a reaction or like call call it a conversation with John Kenneth Greath. He wrote the affluence society, right? But this was written in the post-war economy. >> It serves kind of as the degrowth or handbook. And my sense is like you know John Kenneth Gre is a brilliant man. I'm not saying he's wrong or he was wrong when he wrote the book, but I'm saying it just doesn't apply anymore, right? He was talking he had this idea of the dependence effect. Advertising actually is a way to whip up demand so we can maximize production because that was what he called the conventional wisdom at the time. you should be maximizing production. But the reality is in the time he wrote this book 1958, you had people moving to the suburbs, getting big houses. They, you know, the GI Bill helped them buy these homes. You had the suburb, the idea of the suburbs was was being deployed. And so, you know, people needed washing machines, they needed cars, they needed refrigerators for the first time. And so, you had these big companies that made these mass market goods. They advertised in mass market media. And they and and John Kenneth Grey's idea was like, well, that's just creating demand. The the there is no actual inherent demand for these things.

it's creating it through advertising. And my point is the opposite. You know, we don't have this homogenized society anymore. And we don't have people that need these homogeneized goods anymore. And we have a lot more particular specific media now. And we can reach people and advertise to them the things that they have demand for for products that that weren't economically viable prior to these systems, these distribution systems. And that is the prosperous society is being able to reach people to meet the demands that they have with the products that they couldn't access before with ads that they wouldn't otherwise see absent these systems. And so I think it's it's you know I think it's a very it's my sense is like you can make a very credible bull case that that's what AI delivers to us and it's not about wiping out white collar labor because the reality is like that's going to create more jobs and we're seeing that now like there's no there's no justification for that skepticism. it just doesn't exist. We're seeing an increase in hiring, maybe not at the entry level, and you could >> discuss if there

should be some intervention there, but my sense is like AI actually, if you look at the data, and there's a Financial Times article about this the other day. If you look at the data, it doesn't support that bare case. And so that bare case should be absolutely eliminated as something that even enters the conversation.

>> Yeah. No, I completely agree. That was an amazing speech. I don't know if I have anybody else. Yeah. No, I I love this idea. We've been talking about it a bunch. just more customization, the long tail of commerce getting even longer and it feels unfathomable like you know specific shirts that are just for you designed and targeted to you on Facebook. We've seen that but like it can in fact get more personalized. Well, yeah. I mean, we've we've seen this with like content in these platforms, right? They're very good at they're very good at like find they they they are very good at serving you.

>> They can serve you a video that has 50 views from a new channel on YouTube and you'll be like, "That's interesting. I will watch this." Right? And it and it's a niche that is like so small it never could have existed in the era of, you know, radio and television. And, you know, seeing that seeing that trend accelerate, it reminds me I was I I was wanting, you know, you know, these like kids RC like ride-on cars. I got like incredibly frustrated with these cuz I've tried a bunch of the different brands, I've tried spending like \$800 on them and, you know, \$400 and all of them just suck. Like the kids, even when you have the you're driving the kid on the controller, the kid can still hit the gas and just like run into stuff and, you know, it's just absolute chaos. I'm like, "What is the version of this that is like, you know, you know, the the top-of-the-line version of this cuz I want to get it. I I use these things a lot." And I searched around, couldn't find it anywhere, and then John just like >> within 24 hours, I got served exactly what he was looking for. I don't even know how how it got served to me, but it was fantastic. thank you so much for taking the time to come chat with us. You back. Always always a great time.

>> My favorite my favorite conversations. And when you're on it makes me feel like we're on we're on we're actually you know on Sports Center >> because most people that have talked about this stuff are not high energy and you're just like full on sports center. It's amazing. I love it. >> Well congratulations all the progress and and the prosperous society. Go listen to it. It's three-hour four-part series. and sign up for mobile dev memo if you haven't already. Of course you should. But thank you so much for coming on the show Eric. We'll talk to you soon. Cheers. Have a good one. Let me tell you about public.com investing for those that take it seriously. They got stocks, options, bonds, crypto, treasuries, and more with great customer service. And our next guest is Burnt from 1X. He's the founder and CEO. How you doing? What is that behind you?

>> Welcome back to the show. Thank you so much. That is incredible. >> My hand and then it's like Neo's hand. >> Yes. Introduce the launch today. What happened? Tell us about it. >> I mean, we've been cooking on this for quite a while. So, super excited to finally show this to the world. And also, it's just so exciting because we're so close to shipping now. >> So, everyone's going to pick this apart anyway. So, we don't need to be careful anymore and we can just like open it up and to me it's this beautiful machine that >> becomes almost more art than engineering right at this point. And >> yeah, we're we're excited to show people what we've been cooking and excited to put it in people's hands.

>> And I think also it has a special place in my heart because I've been working on this problem for more than a decade. Mhm. >> And one thing that Neo and 1X has really been pushing is how to use highly miniaturized high power motors and tendons to create these machines that mimic humans. And the hands is the kind of culmination of all that work, right? It's where all of the complexity comes together to meet the world. And hopefully we've created something here that can really remove that final barrier for how intelligent our models can become, right? Like so much of human intelligence comes from our ability to probe the world for truth and to really figure out how the world works through our hands.

>> Okay. Take us through the full journey of developing hands for Neo. A lot of people in tech love to you know follow an Elon style playbook. Just make the most simple version of something. Simplify simplify simplify. This looks incredibly beautiful but also incredibly complex and so I want to understand like how you got here basically like version by version. >> So it is a very complicated hand but in my opinion it is the simplest version of it that exists that is good enough to do what needs to happen. So first of all 1X is fully vertically integrated. So we do absolutely everything inhouse and in our factory here in California where I'm sitting right now. We do everything from designing the production processes to producing our own motors, our own tendons, the entire system, right?

Sensors, electronics, everything inhouse and that allows us to iterate very very fast because I do really believe in like you have to have this first principles approach right so start with like what what am I trying to solve? We're quite lucky in that we have humans to look at with respect to how do you solve this problem? Nature did a pretty good job. >> So when you say you're looking at humans, what type of training data are you using? What's useful? What are you discarding from just there's a lot of hand videos on the internet. I'm sure people doing all sorts of things. You can do tea operation. You can have people wear gloves and do motion capture. You can do simulation. X-ray.

>> Yeah. X-ray. I mean there's so many different ways. Like are you using everything? Is there one path that you found specifically valuable? So I think actually it starts from like starting with first principles right so 1X has always been about we want to design robots that are safe so they can live and learn among people >> but also because safety is what allows us to learn so we probe the world for truth and of course if our fingers break while doing that or we break whatever we're trying to touch it doesn't work.

So you need to design these beautiful kind of like compliant soft systems that force can flow both ways because you're both seeing with your hands and acting with your hands. >> And that's really what this is all about. So like how do you create that? And you kind of have to look at how nature works like our muscle. Nothing really moves fast. There's no gears, no nothing like this. So we've designed this from these first principles. But we go way deeper than that. I think what we haven't talked enough about yet and we'll share more about this later is like how incredibly seriously we take closing the gap towards a human and it's not to look like a human. It's because we want the work system to work like a human. So even like if you look at Neo's hand behind me here, we worked so deeply on how do you make these fingers nonlinearly just like >> be compliant exactly like a human finger. Because if you get all these details right, you can take all of the video that's out there on the internet.

You can train huge world models based on this and it just works on a robot. And that's what the next world model lab is about. So to enable that general intelligence for robotics, you need to design the robot so that it interacts with the world exactly like a human. And then you want to do that with the least amount of complexity possible. And that's essentially what we have here. But of course, the complexity of that is pretty high because you're mimicking a human head.

>> Let's talk about grip strength. We got grip strength testers here in the studio. is this an important benchmark clicking these together? How strong is the hand currently? Where do you want to go? Because it feels like there's a trade-off there where if the hand like the stronger you make the hand, the heavier, the more the more dangerous it could potentially be. At the same time, there's certain tasks that you expect a certain level of grip strength >> also. Yeah. Yeah. Yeah. It'd be interesting to understand how often tasks come up in your daily life where you need like insane grip strength, right?

>> It's pretty rare, I think, but certainly in, you know, industrial capacity or even around the home, picking things up, moving a chair, you need to be able to grab it without dropping it. It's a safety issue at the end of the day. >> Yeah. I think it actually appears quite often, but you don't think that much about it because you don't do it for a long period. like, >> yeah, you're not grasping that hard, but then something starts slipping and you tighten your grip or like you're actually using quite a bit of force. So, the hand is roughly the same. Yeah, that's a good one. The hand is roughly the same strength as an average human.

So, and that's so you I mean it needs to be able to do the full capabilities of a robot, right? So, the robot can deadlift 150 lbs. So, the hands need to hold the bar of 150 lbs. Not because deadlifting is useful in everyday life, but because it's a good metric for like how capable we are. >> So we really worked hard to make that kind of power to weight ratio also about the same as human.

>> So if you look at the general hands in the market right now, >> this thing is roughly three times as high force as the other hands. Y >> and that is really also something that's going to enable a lot of new applications, right? Because in the end your AI will be as smart as the diversity of the experiences that you have lived and experienced. Like diversity of data is directly correlated with the intelligence of your model.

And if you are as a third as strong as a human in your hands, >> there's a lot of tasks you just can't do. >> Yeah. I have one more. it feels like you have jumped to the frontier of hands specifically. is I saw people joking, can I just buy the hand? Obviously, they're making probably rude jokes, but is there a world where you partner with other robotics companies to sell a piece of your hardware, maybe just the hand to someone else that already has a wheeled robot, but it needs a hand. Is there a world where you're selling parts of your technology, or do you want to be vertically integrated from end to end the full experience?

>> I think there there is a world like this. Mhm. >> I do think it's very important though that like we want to we we're we're about to also launch Neo as a platform where we we're going to invite everyone in to build on this and having like a homogeneous platform that everyone is building on is so incredibly powerful because that doesn't exist today and that really allows you to do benchmarks across systems like you have in the rest of the AI

community. but that being said, it's not a hill we're going to die on. like if the if the collaborations are the right types of collaborations, we just want to make sure we can scale our manufacturing and get as many out there as possible and build the ecosystem and really give robotics all the love it deserves, right?

>> Yeah, totally. >> We just want to accelerate the path. >> Yeah. >> Amazing. timeline around shipping, what what's what's the what's the update there? I know a bunch of people that are that are in line that have that have ordered. So, everyone's ready. >> So, yeah, I'll be kind to my team and not give you a specific date, but we have promised that we are going to ship this year and we will ship this year.

So, we're going to keep that promise and it's going to be incredibly exciting. And like I said, like the reason we can be so open, right? So, you can just read into that. Like I said, the reason we can be so open is that this is about to ship. So, people will pick it apart anyway. >> And I do think this is going to be so big, right? Like as AI now becomes physical, it's really hard to understand what kind of impact that will have. Mhm.

>> we're getting so much interest from let's say wet labs that want to have their AI for science actually design manufactur and run experiments. there's like hospitality elderly care. You have the home that we're already working towards. Like there's this enormous surface area and it's going to happen a lot sooner than people think. And I think right now it's just about really growing the pie and making sure that everyone has platforms that they can work on to to solve these hard problems.

>> Yeah. Amazing. How will I mean the last question I have is like like this this feels like a a technology that even after you solve development design and the the the AI that powers all this like it is much more gated by the real world and thus we would see like a slower takeoff like what we've seen with Whimo. It's it's you know everywhere in San Francisco but as you go around the world you don't realize that cars can drive themselves. whereas you know chatbt.com was available in every country and it was just like the touring test is passed for everyone at the exact same time and that's that feels impossible in robotics in the physical world. But do you have a different view of it or am I roughly correct with that prediction >> the ramp is going to be slower but the total uptake is going to be way way way higher. Sure.

>> Right. So like if you think about and I'm I'm very bullish on this like I think it's just a couple two to three years away but even if it's a decade away like robots will build robots. >> Yeah. >> And we're already working on this in a factory but they won't just build the robots they'll build the data centers the chip fab the energy infrastructure get into mining and refining and this full automation of the physical substrate that enables everything including intelligence >> that can only happen with robotics.

>> Yeah. >> And that's going to look like this right. >> Yeah. So you you need to kind of like enter that curve and I think the uptake ramp is going to be slower in the beginning but way way way higher as you kind of like hit vertical on the curve. >> Yeah. >> And I think this is also where it gets extremely interesting. Right. I'm back to like h how we're going to solve some of the remaining problems in science.

How we're going to create an actual true abundance of labor across society. >> Yeah. >> this is only possible if

you automate the physical substrate. So, it's going to take slightly longer, but it's also worth it because the impact is so tremendous. >> Yeah. And it still should be an exponential curve because once you get to the point where you know five robots can make one more robot in a month, then you wind up compounding and the exponential just grows and grows and grows. Fascinating. Very exciting times.

Congratulations and thank you so much for coming on the show. I'm super excited you guys shared this and you guys continue to have the the best aesthetics in robotics by by 100x. >> Yeah, it makes me feel more much more C3PO than Terminator, which I think is the right the right direction to go. >> The hands a little% >> the hands a little Terminator, but yeah. One one should not underestimate how important it's going to be to do this together with people in the sense of like adoption needs to come through making everyone used to this technology.

Right. Totally. we we want to make sure everyone understands how helpful this can be and really make sure that we don't >> don't hit any barriers where like this becomes something that people don't want because it's such a great opportunity and we want to make sure we can accelerate the path. >> Yeah. VR you know VR was useful in in certain pockets but it w it was awkward and it was never adopted and it was always seen as like this this like yeah very like niche technology and I I think the the aesthetics are underrated. So congratulations on nailing them. Thank you so much for coming on the show. We'll talk to you soon.

>> Great stuff. >> Have a great day. >> Awesome. Thank you guys. Cheers. >> Bye. >> Let me tell you about Codeex. Codeex is a powerful workspace for getting work done with AI agents. Whether you're writing code, analyzing data, creating content, or automating business workflows, Codeex helps you move projects forward from start to finish. And we have Tibo joining in just 40 minutes or 30 minutes to give us the update on 5.6 and what's going on in Codeex, of course. But first, we have Josh Lindri from CIA. He's the head of podcast development here to give us an update on all things podcasting. Suited up, looking good.

Josh, how are you doing? Welcome to the show. >> I'm doing well. You guys are always looking good as well. >> Yes. great to have you on the show. Great to hang in France just a couple weeks ago. >> That's great. >> give, an introduction on yourself, how and when you got into podcast, and then we'll talk about where we are now. >> Yeah. >> Yeah. It's been a wild ride. 12 years for me. I started representing podcast 12 years ago. I was a music agent at a boutique music agency in Seattle, booking tours for indie rock bands. And I used to listen to podcasts all day while I was routing tours and had the idea that maybe podcast could have agents. Started cold emailing podcasters and was surprised to discover that there was a business there for me.

>> and very surprised in that time to discover that there wasn't really much of a business infrastructure yet. >> so there was a lot of opportunity. spent the next several years signing podcasters within that agency. and then in 2018 I met with 11 different agencies and ended up joining CA started the podcast department here. and we have a great team people focused on podcasts. the podcast department is within our greater creators department that works with all kinds of different creators like yourselves.

>> and it's been a really wild ride. I mean, when I got into podcasting, the estimates I've seen is that the global

podcast advertising industry was worth about \$45 million. And Owl and Catus put out a report that last year it was worth 9.2 billion. So, you know, I I expected there was going to be a lot of growth in this space. It seemed like a great growth area. Thank you. >> but I never expected this level of growth. I mean, it's been a really tremendous wild ride. I had started working with some podcasts around the same time a little bit later. I think it you were 2014 is that?

>> Yes. >> Yeah. So, I I probably started working with with podcast in like 2016, but even then I would meet a show that today was probably like a \$10 million a year business and they would have zero revenue, but they would have this like rabid fan base and maybe they'd have one sponsor, which was just someone in the audience that reached out and was like, "Hey, can I send you some some free stuff if if you talk about it?" And they'd be like, "Okay." And then and then, you know, fast forward to today.

those kind of properties are are super valuable. what what are you seeing what are you seeing now like what's coming down what's coming down the pipeline net net news shows you know we've covered a lot of the evolution of you know formats how podcasting obviously interacts with with live streaming in our case but what do you see coming down the pipeline >> yeah I mean you know it's been clear for a few years now that video is going to be a bigger part of podcasting and we're really seeing that come to fruition in the past year. It's like a major inflection point right now. now to be clear, like it's not that video's replacing audio. Both can continue to exist together. There was a recent study from Edison that said that the majority of podcast consumers do both.

>> Sometimes they do audio, sometimes they do video, which that's my experience. I live in LA, so I have a long commute and so I I like listening to podcasts in the car, but I also like watching podcasts at home and in the office, you know. but it's created a really interesting moment in podcasting where there are some things that are fundamentally different about digital video versus digital audio. For one thing, advertising looks really different, right? The ways you can integrate with brands looks very different versus the audio space tends to be much more dominated by your 30-sec pre-rolls and your 60-sec mid rolls and so on. The video space tends to have a lot more custom integration with advertisers.

and it also discovery looks really different in video. You know in the audio space in terms of breaking stuff through there's a lot more spend that is required in sort of the same way that you might market a TV show or a movie right whereas in video there's a lot more clipping. There's really seamless integration into social media. and so if you're trying to break through with a new podcast in 2026, you should have a really strong reason why you're not a video podcast or else you should probably have a video podcast.

>> Yeah. somewhat related to the video podcasting thing, a big trend out of can that Colin Samir and others were talking about was, that many podcasts are sort of reformulating as shows. we we we've done this where we think of this more as a show than a podcast because it's a live show. There's a lot else going on. Of course, like it's available as a podcast. And then you also think of like Subway Takes. It's Emmy nominated now and it's a it's very much a show, but it's also an interview that's sort of like a podcast. And I'm wondering how you're perceiving the definitions changing, evolving, just this idea of what does it actually take to to deliver a show as opposed to just a podcast in the modern era.

>> Yeah, I'm with you. I think that's the right thinking to not try and put it too much in a box, right? Because the lines are just getting so blurry between what's a podcast, what's a TV show, what's a series of reels, right? What's a YouTube channel, in your case, a live stream. >> I do think that the word podcast has a bit of utility for me just because >> it's sort of like I know it when I see it, right? People talk about podcast, people like podcasts, but I mean truly the line is really blurry. I mean, you know, I think we talked a bit in can about Oprah's podcast, which just moved over to Amazon, right? And I mean, Oprah is the queen of television, right? And and this is where she's putting her energy.

>> and you know, who's to say like if you're watching it on Prime Video versus if you're watching it on your phone, is it a podcast if you're watching it one place and it's a TV show if you're watching it the other? I mean, I don't know if it necessarily matters, but I think that it's an amazing time to be a media consumer because media is meeting us where we are. >> Yeah. Talk about the landscape of these podcast distribution deals that are happening. Pat McAfee with ESPN, Oprah, you mentioned, is doing one.

There's Netflix is entering the space. Spotify went with Joe Rogan very early on. some of these platforms just sort of get the video podcast for free. Like YouTube is you know, the default for most creators. But what are the larger companies looking for when they want to go deeper with a creator who might just not be ready to graduate from the self-s serve options or maybe they don't have a self-s serve option if it's linear TV or a platform like Netflix.

>> Yeah, it's funny. I mean it's a really unique marketplace because as you kind of overviewed the different buyers in my space are just drastically different in terms of their business models you know so it's really hard to compare apples and oranges when you're looking at one of our major buyers say is Sirius XM which has a really significant satellite radio business comparing them to Amazon right one of the biggest companies in the world and primarily you know doing e-commerce right >> they both have drastically think different things they need out of it let alone now you have Netflix and Hulu entering the space, right, that that have different KPIs in terms of what they're looking for. So, I think part of the the challenge and the joy of representation in this space >> is understanding who the different buyers are and what their needs are and understanding your client, the talent, >> right, and what they want, you know, and it's it just there's no one deal that makes sense for everyone. And there's no one-size-fits-all in podcasting, which I think is great ultimately because it means we can paint with different paint brushes for different types of shows.

>> Yeah. >> Can you do you feel like you can identify if somebody's going to be a star based on their first ever episode? >> Well, I think that I I think that taste is really important, you know, and for me, like I I will take bets on stuff that I think is fundamentally good. I mean, >> when I was a music agent, I was picking bands based on the bands that I liked, and I just had to hope that other people would like those bands as well at some point. And I I still believe that, you know, in podcasting, I think that there there is room for taste. There's certainly stuff that I don't represent that has that does good business, but it just wasn't right for me. And that's fine. You know, I think that as soon as you divorce as soon as you divorce your love of the medium from the business, then what's the point, right?

Like if we're just here to make money, maybe we should be working in finance or in tech or something, right?

>> Yeah. Maybe we should all be wearing suits. >> There you go. >> Look, I look at me. I'm just here to have fun.

>> Yeah. >> How are you thinking about live streaming and maybe more pure play live streaming? I'm thinking about the video game creators, the ninjas, the the shrouds, the folks who spend 12 hours, eight hours a day live streaming political commentary, business analysis, anything. It's such a different business model, such a different community, sometimes much tighter audiences, but incredibly engaged. is that a muscle you want to build? Is that something you're thinking about growing? what trends are you seeing there generally?

>> Yeah, I think the live streaming space is fascinating, right? And obviously this has been happening for a while, but I think the moment that we're talking about of the merging of all the different things is really playing out for live streaming too >> for yourselves and for other shows. You mentioned political shows where people need real time news. I don't think your fans want to wait an entire week necessarily to get your take on something when it develops.

>> it's a lot of work. I think you guys know the amount of time and energy you have to put into doing this, right? It's really impressive and I applaud you for it. >> but it really creates a lot of different opportunities. I think something else that you guys are doing really well is that you do the live stream, but you also cut this up into audio episodes and video episodes on YouTube and so on. So, there's so many different ways to reach your audience.

And I think as a creator today, the more that you can be flexible to meet your audience where they are, the better chances you're going to have of succeeding and breaking through. So I love to see what you're doing and I think that you're on the tip of the spear right now in terms of the new type of experimentation that's happening in the streaming space. >> Yeah. talk to me about how you're thinking about working with a celebrity or group of celebrities that wants to get into podcasting. It feels like there was this crossover moment where podcasting was a backwater. Then it became cool. Maybe it was around COVID, but we got a whole bunch of celebrities crossing over. Some of them did extremely well, won awards.

Smart list is huge. and but it feels similar to celebrities launching brands where there's still going to be a power law. It's not just it's obviously a huge advantage to have an audience already, but not every celebrity is going to have a hit podcast and vice versa. Not every podcaster is going to work on TV or wind up starring in movies. how are you assessing the the the reverse transition from you know Hollywood or TV or films coming over into the podcasting world successfully?

>> Yeah, I think there needs to be a reason to be for any given podcast. There was a level of experimentation like you mentioned around COVID where a lot of folks launched podcasts that maybe didn't pan out. >> Yeah. And I think maybe some of those had to do with the the the reason behind the podcast and the idea of the podcast, you know, wasn't as sought out, right? I think a good example of it working would be Julie Drifus's Wiser Than Me where that was driven by her desire to hear from older women that are largely ignored in our culture, >> right? So, she had a reason that she wanted to make it. It wasn't like she would just come into the space because, you know, some enterprising agent myself told her that she should make a podcast and she could

make money, right? She was there for a different reason. The things I see succeeding have that reason behind them. I mean, you mentioned smart list, right? Another co, you know, project that that began of co and it came together because the three of them wanted to hang out, >> right? and they wanted to create something for people who were locked at home, you know, and it and that that like genuine friendship between them is the basic building block of what it is, you know. So, a lot of what I do when I talk to celebrities about podcasting is try and get to the core of what it is that they want to create and what their purpose is for coming and then we can craft all the business and everything else around that like central seed of an idea.

>> What's the secret to a successful podcast tour? I mean, back to Smart List, I feel like they've been extremely successful at engaging the community off of the internet, which is interesting because they started off the internet, they went to the internet, then they go back into the live tour. but what what what makes for a successful podcast tour? >> Yeah, I think that your relationship with your audience is so strong in podcasting and it depends a little bit by format. you know, I think the level of engagement for more format driven shows maybe a little bit less than more personality driven shows, but something that I saw really early on getting into the space and coming from a touring background, I was booking tours for podcasts and you know, one of the first live podcasts that I went to was early client of mine, stuff you should know, which I assigned by emailing info at stuff you should know.

>> it was a really different time. but then I I went and saw them do a show in Vancouver and they had a Q&A at the end of the show. keep working up to the microphone and this this woman gets up to the microphone. She's like 22, looks normal and nice. As soon as she gets on the mic, she is balling cuz she is talking to Josh and Chuck from Stuff You Should Know. And it was a real like light bulb moment for me, right? Cuz that's a educational podcast, right? But for her, she was explaining on the microphone that they spend so much time in her ears that she has this relationship with them and it was like she was finally meeting these friends that she's had for so long. And I think that's that's when you get people who are not just willing to buy a ticket, but to travel four states over to make it to a live show. And you know, we we look at data for download performance in a market before and after a live show. And we see, you know, in some cases like a niche show can have really solid touring business because they're converting like 50% of their listeners in a market are turning out for these live shows. And we've done analysis too where we look at zip codes of of ticket buyers and the the number of people who are traveling long distances to come to these is is pretty incredible. So, you know, I think that's the key thing is if you have that relationship with your audience, if you let your personality be a part of the podcast, people are going to want to come out for that.

>> how are you thinking about international just the like the puzzle? Because if your business is if your podcast is aligned to specific products that are maybe only sold in the United States, it can be hard to monetize. You might need to go on a tour. An international tour can be more expensive. At the same time, I'm sure from musicians that you've worked with, you've solved the puzzle of what an international world tour looks like that is successful. So, how do you think like the the that that that the the future era's tour of the podcasting world will play out?

>> Yeah, I I think that the international touring is I mean you have >> it see it feels like it's early. Am I right in

that or am I just not aware of there was already the era tour of podcasting and like Smart List did it and I just wasn't paying attention to their trip to Japan or something? I would say there's only one era tour, but you but you are seeing a bit of international touring happening already.

I mean, you know, many years ago, we sent stuff you should know to Australia and it was such a wildly successful tour. You see a lot of American podcasts doing the UK, a little bit in Europe. I mean, one of the challenges is when you're crossing over into markets that are primarily not English language speaking markets where you might have listeners, but you might have a different type of relationship with them. >> that can be a real challenge. But especially going from one English language market to the next, it can really work.

>> but you do see audiences becoming really segmented between different markets. >> if you look on any given day at the charts in the UK and the charts in the US, they're probably going to look pretty different. There's going to be crossover for sure, >> but a lot of different stuff changes out even though we speak the same language and understand each other very well. There's just different sensibilities from one market to the next, you know.

>> but I mean, you mentioned advertisers, right? the US ad market is a big leader in podcasting, right? Because this is a consumer that a lot of brands want to reach. >> And so, you know, as we've built our business internationally and signed podcasters from all over the world, you know, there's still at this point a real value into having a foothold in the United States in terms of reaching audience. >> I think one of the most exciting future change areas in podcasting is going to be some seeing more of these markets come alive. M >> and if I was, you know, an investor looking for a place to start up, I would be looking at India right now for instance, you know, right whereas there's just a little bit less saturation and more opportunity for extreme growth. Australia, I think, is a really interesting market. I made it out to Sydney last year for South by Southwest Sydney, and it reminded me a lot of the podcast market in the US 10 years ago, 12 years ago when I got started. And so I think there's a lot more that's going to come online for those markets as the brands locally and the advertisers in those areas start to realize that this is a great way to spend money and to reach audience. But it was an education process in the US to get brands to spend here and trust this market and it'll be an education process and a marketbymarket basis.

>> Last question. how are you talking to or I don't even know if you can talk about this, but folks who work for large media companies, they have been, around for the pivot to video. We're going to put you on camera. We're going to set you up with a podcast. They build an audience and they're ready to venture out on their own. We've had Ashley Vance on this show, Joanna Stern, Eric Newcomer. there's been a whole host of these folks who sort of grew audiences and learned the skills of content creation, whether it's Advice or Vox or any of these platforms, and then they go independent. If you're having a conversation with them, what how are you talking to them about why they might want to do that or why they might not want to do that?

>> Yeah. So, every podcaster, every creator is an entrepreneur, right? which can be really scary if you're used to getting a paycheck every single, you know, bi-weekly from a big media company. >> But it's high risk, high reward, right? Because once you launch your own show, you own that audience. and no longer are you at the

whims of, you know, the executives that you work for. >> and actually capturing a smaller audience can be more lucrative for you because you're capturing more of the revenue that that audience drives.

Right? This is a conversation I have with folks all the time who are at legacy media companies trying to decide what their next steps look like. especially in this really fast changing landscape where some folks are forced out necessarily when they necessarily want to make that choice right away. >> so it can be a really scary transition. I'm very empathetic for people who are going through it. I don't begrudge anyone who decides that they want to keep working in legacy media. I don't think that it's doom and gloom for legacy media, right? I think that there's still room for great journalists on television and on radio, print journalists and so on. but for folks who are entrepreneurial and want to build their own thing and own their audience, there's incredible upside and opportunity.

>> Yeah. Yeah. It's exciting times. Well, thank you so much for taking the time to come check. >> Yeah. Great to have you on. >> Have a great rest. Thank you so much for having me. >> The godfather of podcasting. >> It's true. It's true. >> we will talk to you later. Have a good one. >> Great to see you. Thank you so much. >> Let me tell you about Crowd Strike. Your business is AI. Their business is securing it. Crowdstrike secures AI and stops breaches. Up next, we have Jeff Morgan from OAM. He's the co-founder and CEO. It's his first appearance of the massive series B. Jeff, how are you doing? Welcome to the show.

>> Thanks for having me. Doing great. So excited to be here. Big fan of the show as well. >> Thank you. Please introduce yourself. Introduce the company. I want to get to the bottom of how the hell you got 85% of the Fortune 500 using your product, but first introduce yourself and the company. >> Yeah, I'm Jeff. I'm the CEO and co-founder of Alama. Lama is the largest network for developers to access open models. You can download Alama, get connected right away to open models like GLM52. And you can download them locally and run them right on your laptop for more kind of edge low latency use cases.

>> Okay. tell us about the round. Jord's warming up the gong. How much did you raise? who from >> 65 million the lead was Tamas Tongas and we had existing investors participate too like >> awesome fantastic so tons of GitHub stars why how does this fit into like the the value ad to businesses versus just downloading the model themselves from an open platform getting the the actual the weights themselves, deploying things or just going with an API like how are you talking to Fortune 500 customers about how you will actually improve their experience with something like GLM 5.2? You know, the power of open source is how fast it can build trust with developers around the world. It ends up a large number of developers work at work at companies that work at, you know, a lot of those in the Fortune 500 or global 10,000. And open source has this special, you know, capability where you can deploy it in your own environment and not have to think about a ton of security and compliance. And what that means is look like you can take an open model which already has open weights and that's kind of why open models are perfect for these use cases and run them without any approval really as a developer and and get real successful results all without you know having to expose your data or you know even incur big costs. So you know that's really been the driving force of being able to to get into these large businesses. and to your point look like you know just the weights being open isn't enough. You need a way to deploy them, to run them, to make sure it works on your hardware. For the cloud models, you know, you

really need to make sure that you're running them in a secure environment where your company can access them. You know, a lot of a lot of businesses we talk to, especially Fortune 500, they need these open models hosted in the US and in Europe. And that's just a requirement and it's it's a need they have. And so, you know, put put that all together is it's so surprising and incredible how fast it can get adopted. So what does I mean you say you're you're you're talking to these Fortune 500 companies but I imagine that the fact that you have so many GitHub stars means that there's a lot of self-s serve activity. when does a customer cross over to an enterprise relationship with you?

>> Yeah, generally it starts with the individual dev, right? They bring it they bring it to work. A lot of them use it, you know, for personal productivity and they bring it to their team. And once you're using a team, it's not a one one person story anymore, right? There there's a team there, right? and everything from security to technical architects, IT teams, you know, and and these folks, they they need not just a product that's really, you know, easy to use and self-s served, but they they need a solution that really kind of end to end covers things like safety and monitoring and logging and data storage and protection. Like these are all components of a successful, you know, agent deployment. And so, you know, that's when it becomes a multi-party environment. And, you know, as we know, like that's when you need a solution. and and on our side, you know, we need a team to be there to help help those customers.

>> What what set of models are you most excited about for the back half of this year in the open weights world?
>> I mean, I think with GLM52, we just had another massive moment in open models and you know, by far at least from what we know publicly is the highest token volume of accessing GLM52 and so I'm excited for that because I think there's going to be a series of new models that are long horizon. and they're focused on these really hard agentic use cases and there's going to unlock so many use cases in enterprise that you know the prior generation of open models couldn't you know and the gap between open models and the frontier models is shrinking and so you know I think at that point we're able to get to these incredible use cases that just weren't there you know three or four months ago >> take me through some of the game theory in the opensource community around those rumors that we heard that there might be export controls on open weights models coming out of China soon if if if we stop getting frontier or or or near frontier opensource models from China for free is is the next step that you would see an American company step up Nvidia or maybe Meta changes their strategy. How are you thinking the the open source ecosystem would evolve if China changes their strategy? Yeah, you know, we like to work backwards from from our customers. What are they trying to do? And they >> for by and large, you know, they may have preference on specific, you know, geographies where the models are from, but by and large, they're adopting both, right? They're and some mix of open models and frontier models as well. And to your point, like I think the US models are absolutely stepping up.

They're incredible. The Neutron 3 Ultra model is just amazing and is able to accomplish some of these longunning Asian tasks. and then also, you know, one of the most downloaded models on Alama is a US model. It's the Gemma models and you know this is like a super amazing team at at Deep Mind that's putting them out. The new ones are you know agent ready like they can run coding agent loops they can accomplish much harder tasks and so look I think it's really up to the customer if they want a US entirely US built model designed from scratch that's there.

if they want a Chinese model which is often the case they it's less about where the model's from. It's like where does it run and is it running next to your data which you know a lot you can deploy it locally. and then are you able to deploy with safeguards and ends up a lot of customers, they're not looking for like where the model's from. They just want to make sure that they're running it properly and safely. so they, you know, they can have a an understanding of what's going, you know, what could go wrong, but what could go right and and there's a lot of safety tooling that can be deployed to help with that. they have tons of appetite for that. What do you see your role as in terms of benchmarking, reality checking, vibe checking, different models, helping enterprises that work with you to make the right decision, pick the right tool for the job?

>> You know, our job fundamentally is to connect the 9 million developers on OAM to the right model for the right task. And that's that's step one. And so just by having that sheer volume and that this critical mass of devs, we're able to already understand just from you know our community which models are performing right for the right tasks. That that's a starting point. I think from there it's really collaborating with the model labs and we're launch partners with every major model lab and just making sure that you know the best parts of the model are shining through through a lama. including what are they capable for, what are their benchmarks, how can customer customers benchmark it for their own use cases.

it all comes down to a lot of software tooling and and and and and you know, a community and a and a network and that's, you know, what we've built and and is what makes Alana special for developers. >> Got it. \$65 million raised. Is this like what are you using the money for? Because you don't have the crazy training costs because you're more of a gateway. is this headcount off >> 2,000 BDRs? >> Is that banging down?

>> No, it feels like you have also like bottoms up adoption with developers. you maybe need a lighter sort of like sales force to to actually capitalize on that. >> Yeah. What what Yeah. What are the next 12 to 18 months look like for you? >> Yeah, you hit the nail on the head. Look, we we put out a you know on our site, hey, we're launching a teams plan. We were inundated with thousands of teams that want to use a Lama and you know that's going to that that's the core mission. It's like look we've got this critical match of devs. How do we go solve problems for businesses? Back to what we were just talking about.

>> And that takes a team. So obviously we're expanding. We got here with 14 people to to company of this magnitude. but there's a much bigger team to build around market. >> Very cool. >> And then of course you know one thing Oama does very special for the larger open models is we host it on US and European servers. Okay. a lot of the consumption of open models is going to China or is going to servers where it's there's there's no data retention guarantees and that's so important for companies and so that's a compute investment we're making and really enabling you know every business in the world to access the most powerful models on compute that's secure and safe in the US or Europe. will we ever settle the debate on whether the gap between open and frontier models is closing or widening because depending on what sort of group somebody is a part of they tend to have one one view or or the other.

but it but I think in in reality it's probably always kind of going like going like this to some degree. But what's your view? >> Yeah, I think you're right. It's oscillating. I mean, I'm a daily JLM 52 user through Alama right

now and it's replaced 80% of my coding work >> and I think a lot that's going to be true of a lot of customers. as for the gap, like to your point, I think it it may widen, it may shrink. I think overall it's shrinking. but but ultimately customers are going to use a mix and you know, for the bulk of their use cases, they're going to reach for these open models because they can tune them to be much faster. They're obviously much cheaper. and there's always going to be use cases where you need the frontier. I don't know if we'll ever settle the debate. I think ultimately the gap will will continue to shift. I think that's what makes it exciting, right? It's like every every 3 months we're able to do something new.

We're able to run better agents and quickly open models will catch up and really enable a whole wave of customers that want to run open models to do that you know in their own environment or to customize it to the point where like they can even make it more more powerful. The last thing I'll say too is, you know, customers are readily taking these open models and customizing them and they're actually getting better results often than, you know, just a stock Frontier model and, you know, I think we're just at the beginning of that transformation.

>> Very cool. Well, congratulations on the progress. Awesome to meet you. Congrats to the team for coming on the show and have a great rest of your week. We'll talk soon. >> Goodbye. >> Awesome. Thanks. Cheers. >> Let me tell you about Shopify. Shopify is the commerce platform that grows with your business. lets you sell in seconds online, in store, on mobile, on social, on marketplaces, and now with AI agents. Jordy, there is a story. Oh, wait.

We actually have our next guest. >> Guest of honor. >> Now, TBO from OpenAI is the head of core products and platform. Tibo, how are you doing? Congratulations on the launch. >> Hey, thanks. Doing well. >> give us the highlights. >> How did you sleep last night? >> Do you sleep? Do you sleep at all? >> Yeah, I do sleep. Currently we have like, you know, maybe five to 10 war rooms going. So, you know, it's like a little bit it's intense. It's a sport.

>> Yeah. >> you know, we heard >> Well, and you guys like to make it hard on yourselves by launching every every time there's a launch day. It's like, you know, 15 new things. So, it makes sense that there's, you know, close to equal amount of war rooms. >> Yeah. Let's start with 5.6 Soul, though. I want to I want I want you to identify for me like what is sticking out? What are the most cutting edge capabilities that really stuck out to you? The latest unlocks from the frontier of the actual model. Then we can go into codecs and voice model and everything else and how things come together and how these are used. But first, just from the raw model capability, what was most impressive to you? What was most exciting?

>> Yeah, it's it's it's actually really hard to answer that question because when we were looking at the benchmarks trying it, you know, we were just like blown away by it all. is just better at coding, better at cyber, better at everything like long context, producing documents, better at, you know, having taste, website generation. and then for the first time, we also really cracked, I think, multi- agent setups, which we shipped as the ultra mode. And when you just see that going and you know, you've got like eight agents collaborating together, communicating, and you know, getting the same work done like faster, it's just you just feel like, wow, you know, this is like another way to scale test time compute. Yeah.

>> but overall just amazing workhorse. Feels way way better than 55 and anything else we've produced so far.
>> So in January there was a project that was getting some attention called Gas Town talking about all these different sub agents. You had PC cats and all sorts of different abstractions. It felt highly technical and it seems like soul ultra is a way to abstract that away. Is that is that deliberate? Is is there I guess the question more broadly is like is the level of like prompt engineering are we are we leaving that era or or will there always be some cycle of you know if you get really good at using soul ultra you'll have a better experience because you'll be able to give more fine-tuned fine grained prompt and and direction to the model.

So one thing that you see as well with solo is its inability at understanding human intent and you know you need shorter prompts you don't need to explain yourself in that much detail and so like you know just gets it and then goes and like does like a very complex thing >> you know you saw the prompt for like post training the Luna model which is like super crisp and then it does that like but it actually worked for many days and >> this is also like with us launching chat work it's it's about making it accessible for everyone and you you don't need to have like a PhD to use this model. It's just like >> it should just behave like just like another super super smart human and you know just kind of get you in the moment and that's what we're striving for. Of course, if you know you really push it to the limit, you're always going to find new setups and this is like also a very exciting space. You know, we continue to develop like also Codex in the open source. and then we're seeing like you know all sorts of all sorts of novel ways to set up these agents and models so that you know you can get results in like cyber security and like you know all these other more nuanced and complex things. but for everyone you know you should just feel like you know a ton of power out of the box.

>> Talk about what was important at the product layer. Fundamentally, what I think people want out of products is to just be able to talk to their computer like a really smart co-worker and be able to get things done. But then you're dealing with, you know, so many users over here, millions of users over here trying to combine it and condense it into something that's simple and obviously simple things end up being, you know, exceptionally complex to actually create.

>> Yeah. So, if you look at it, it's deceptively simple. You can open it on your phone. It's catchup work. just toggle it and then there you go. You connect it to the things that you already have, your email, calendar, you know, your docs. and then suddenly you're like, okay, wait, I can ask it to process all of this information that I had over there that I had to manually like do all these things myself and it can just do all of that and it's just like on the go and it's like on your phone in the chat app you already have installed. I mean, that's the that's the beauty of keeping it very simple. At the end of the day, we want it to be just a normal conversation between you and the agent. This is also why we decided to, you know, ship it like just in >> town.

talk about progress in computer use. what is actually driving progress there? Is this just something that sort of comes for free with scale and and model advances or is there deliberate data collection that's happening and some sort of flywheel that's unlocking new capabilities in computer use? >> Yeah, we've done a lot of effort bespoke effort on Windows, Mac and like mobile computer use also like phone use as well. And so there's an entire team working on this. It doesn't just come for free. But what does come for free is like every time we push the efficiency frontier and the model gets like you know more efficient like at thinking and acting and it just you

know costs less tokens and it gets compressed in time.

It also gets better at computer use because it reduces the latency and reduces the cost and so the two compound like you know we have a lot of gains that we're getting from you know also like visual understanding and every time you know it improves like the model just gets more precise so it doesn't have to correct itself and it was like maybe it misclicks the button and just like oh just like wait I have to redo that. So every time it's like more accurate and you know more token efficient computers definitely benefits from it and when you compare it to 55 it's like you know just really like three times faster. So you know we're not at all hitting a wall here in like how fast we can do computers.

>> Yeah. Can you talk about how the role of member of technical staff is evolving because you're you're you're talking about soul ultra going off and working for days at a time and at a certain point your job sort of evolves to if you have a launch tomorrow, don't kick off a task that takes four days. Even if the model's capable of it and will deliver something great in four days, you need it tomorrow. And so you have to size your workloads appropriately. How is the how are you thinking about sizing work and and actually delegating the right the right chunk of work at this stage?

>> Yeah, I find your question very interesting because it actually highlights like a shift in our thinking over you know since we had five sixes. >> You don't really instruct it necessarily you know for a task that's going to take four days is that you know you tell it all the information that you have. So, you know, you're like, "Hey, I have a launch tomorrow." Yeah. Right. And then like keep track of the time and like understand that, you know, the PR needs to land like, you know, by midnight or like by 2 a.m. It will just like reason over it. You're not the one that needs to manage like, you know, all of that extraneous like complexity.

>> Yeah. >> And so that's it. What you're seeing as well is like, you know, your relationship with the agent like changes over time as it gets more intelligent and you're just like, "Oh, yeah. I can just talk to you like, you know, another like, you know, super super smart human." >> Yeah. Yeah. talk about the efficiency of the model, what work went into that, why why it matters, you know, what kind of conversations you're having with, you know, big customers, all that stuff.

>> Yeah. So what matters a lot right now is sitting at the frontier and you know getting the max capability when you want it but also for your normal average day-to-day task is you know being super efficient and not just for latency just because also we're seeing so like we had this era of token maxing and then you know we've been talking a lot with you know all old all the companies and enterprises that we're working super closely with and then they were like oh it's just a little bit you know maybe out of control it's like know what we want is like you know we want a highly efficient model that is you know steerable controllable we want the right you know spend control dashboards and so we also have all of that. you can you can look at your spend, understand the ROI, but also you can rest knowing that you know this is actually like a super super efficient model and so it gets the job job done with you know way way way fewer tokens which you know to you you know means that you know you have to pay less u for the same results which is super important and this is really the theme I think of the year is like you know that being on the efficiency of like you know performance and and cost. What about if

you want to spend more for faster performance? What does the future of either ASIC enabled, Cerebrus enabled, spark and and fast mode? What do you want to see develop there either immediately or over the next couple weeks?

>> Yeah, I think what we are really working towards like a buffet of options, right? So for for your normal like you know interactive task is that you know you're going to use five six soul like on medium or on high and you're going to have like an amazing time. If you have a really hard problem and you're trying to for example find a cyber vulnerability in something and so you're going to run ultra you're going to run it like for two days and it's just going to like leave no stone unturned then you know invent novel techniques and you know you're going to be like absolutely blown away by what it comes up with. and but a lot of times you also just need speed, you know, like for some of the SES that we're dealing with is like, you know, we love working off of like the Cerebra's version of this, which is like at, you know, about like, you know, 750 tokens per second, which is an order of magnitude faster than the the default version that we have on the on the API and in the product today. And this is just really situational or if you just want the very best and like absolutely no compromise. It does come at a cost.

>> Of course, >> a lot of people were feeling left out this week that weren't in the early access program. What makes a good early access partner? I'm sure your DMs are just fil, but what makes a good partner to to the to the product and the research team? >> Yeah, we really try to go as broad as possible. it is quite a bit of effort to manage and then also like we're getting all that feedback and incorporating it and we work very closely with the folks in early access. For us, it's just really about realizing whether it is as good as we think it is, right? You know, you're like so close to the model. You train it, you know, you've incorporated like all that feedback, all your dreams, visions into this model and then you've played with it for a little bit. And then when we give access to, you know, folks outside of OpenI, it's like, you know, the first time where we have like a an unbiased look, you know, where people use all sorts of models and different different harnesses every day.

And so it's just kind of like this awesome you know is it actually as good as we think it is? Is you know what are the things that we missed? And so it's that you know high high bandwidth engagement, good feedback and then you know that sort of people who have shown to be unbiased in the past and you know talked honestly about you know all sorts of models and all sorts of harnesses. h how do you how are you thinking about the trade-off between mobile, cloud, desktop, the Mac Mini that went mega viral last year? do you think that we'll stay in a hybrid pattern for the foreseeable future? Is it a person byperson basis? Do you have a uni grand unifying theory of how agentic work happens in the future?

>> Yeah, the way we think about it is no compromise. So, you want to be able to use the same your same AI partner, you know, like on your phone on you like on the go. It's just like I go walk in the park, I want the exact same thing. >> I want it on my laptop. I want it, you know, at home maybe like running in a Mac Mini. And >> it is more that it needs to be able to have access to all the things that are important in my life and, you know, not be constrained by the physical, you know, boundaries of like, you know, just like, hey, I started this prompt or I started this conversation on my phone and it's just like now it's stuck on my phone.

>> Yeah. you know, we want it to be uncompromising and so your AI your ideal AI partner I think just you know has access to everything all the time and just you know processes the information as needed and then you know can act in a safe way and controlled way you know so that you know you always understand like what it's trying to do and you know if there's like something risky you can you know approve it or you know ask it to change tac and there I also think you know like the mobile has like a big role to play right so if you're if you know five six is you know busy like you know working on something and then you know you just go out to dinner like you know it should ask you for permission to do something >> when you're there you know you don't need to be stuck on your laptop.

>> Yeah. >> Fast forwarding six or 12 months how is how how important is voice to someone's day-to-day experience with chatbt works >> I don't think we need to fast track like we shipped voice yesterday. no, I know, I know, but but you know, you assume like, you know, often times like something ships, you know, and it takes a little while, >> everyone has to go on holiday break, >> a three-day weekend, and then everyone can test out the latest and and integrate it into the workflows.

>> Yeah, just, you know, open a chat with the app like using latest voice. We also demoed it in live stream this morning and >> it's very sort of super enjoyable, like magical experience when you first experience it, but also like, you know, on the fifth time as well. It's going to be part of like, you know, day-to-day experience, you know, of like how you work with these systems. >> We don't we don't have it yet in the desktop app. but this is something that we're working towards. And >> when when you experience it is it it's a like a modernday like Jarvis, right?

It's like, you know, you just talk to it, you just walk in your room and, you know, suddenly it's just like doing things on your computer, you know, with the same level of like precision and power, you know, that you currently have over text. >> Yeah, it's amazing. >> fantastic. Congratulations. Thank you so much. Hopefully you can get some sleep. I'm sure crazy. >> Many big days to come. >> Back to the one of five war rooms, whichever one you'll go to next. have a great rest of your day.

Congratulations and we'll talk to you soon. Ta. >> Cheers. >> Have a great one. Goodbye. >> Let me tell you about Railway. Railway is the all-in-one intelligent cloud provider. Use your favorite agent to deploy web app servers, databases, and more while Railway automatically takes care of scaling, monitoring, and security. And while we bring in Sean Frank, I'm also going to tell you about Cisco, critical infrastructure for the AI era. Unlock seamless real-time experiences and new value with Cisco.

Sean, how you doing? Welcome to the show. Great to see you again. First time first time in the actual studio, right? >> Second. I did the Shopify episode. >> Oh, that's right. That's right. >> He took the time out of his Black Friday. >> That was great. Yeah, but we Yeah, we bounced around a lot. It was a It was a big It was a big day. That was a good one, though. >> Hopefully, we'll do it again.

>> Yeah. Another >> And And Black Friday was a record for you, right? Right. >> Of course. Well, >> yeah. Yeah. Have you Have you had a Black Friday that wasn't a record? >> No. Every year it's gone up >> every year.

>> And it's not going to stop this year. >> Only over here at the Ridge Wallet Ridge, right? The portfolio is growing still. >> Yeah. The wallet business is it's a great business. It's like a 100 million plus a year. But the growth is now like all the other stuff we do. So, we have like a travel line. We have >> luggage.

>> We do like 10% of all men's wedding rings in America. So, we have like a huge >> double digits of the TAM. >> Yeah. It's like a very boring monopoly. We sell a ton of like men's wedding rings >> and they're coming for it all. >> Oh, yeah. Soon. >> you won't be able to get married. They're actually going for regulatory capture. You won't be able to get married if you're not planning to use a ridge ring.

>> Yeah. >> I'm I'm like a big pronatalist now cuz I'm like, get married, have kids, buy a ring. >> Okay. So, you're behind all of that. Every time I see some podcaster about the the fertility crisis, it's you. you're behind in funding it all. >> And the big thing now is like the tech business for us. So like we do like phone cases, that type of stuff. >> It's like 12 months old and it's like already like \$100 million a year. So >> just adding a bunch of new random little widgets to sell.

>> That's insane. okay, so take me through the demand generation side of the business. Like what is actually changing? obviously there's AI generated advertising, AI enhanced targeting. There's all sorts of different stuff. But like from your daytoday, from the platforms you're advertising on, like what has been the most material change over the last 12 months? >> Well, it's a big day to be here, right? Because big Meta announcements. Yeah. People were like dancing on Meta's grave like three days ago and and now everyone's stoked on it.

>> And now you have a 1% god candle. >> Totally. >> wait, wait, is that specifically about the LLM or or the image generation model? Because the image generation model seems much more impactful to the advertising business than the agent coding capabilities. >> Well, yeah. >> Yeah. >> Well, I would say the Manis actually like really helped unlock like a lot of like stuff inside the ad account >> using the ad manager more. >> Yeah. Yeah. I think like it just democratized a lot of tools that like the best ad managers were already using.

Yep. but really we just want Meta to continue to get better. Yeah. Right. Over the past three months, they've rolled out a lot of changes of like the actual ad algorithm. Yeah. And it was really bad for a lot of people, but we just had like the best Q. Oh, it was awesome. Yeah. >> Interesting. >> I think they're getting >> So, so do you think that it was actually bad for some people or they were just going through a slump in their business overall? They had like business problems that they wanted to blame on the ad changes.

>> Well, dude, yeah. I mean, what's new, right? It's like every so if your business is going bad, it's everyone else's fault. but really like you know, they do like the Meta Performance Summit every May and they've just rolled out so many great changes to the ad algorithm that like I think you're getting better impressions with better people and >> I the more compute and AI they throw out, I think it's just going to get better and better and like we're getting to like a future where like the perfect impression at the perfect time with the perfect ad that's like customized to the person with AI that's coming. Yeah. and clickthrough rates will go up, conversion rates will go up, and CPMs will go up with it. But I think there'll be like a moment of arbitrage there.

>> Do do you think do you think Zuck is doing enough to actually message that to customers like you? Because Ben Thompson put out this piece on Monday sort of an earnings call transcript that he wrote in the voice of Mark Zuckerberg and his pitch was to investors telling the investors hey look we are investing a lot in AI but it's all in service of the ads business which is great which we do take seriously. We have taken some side quests done some metaverse some VR stuff but this investment that we're making right now you shouldn't beat us up in the public markets because it's going to come back huge. We're growing really fast on the ad side. 33% is massive at that scale and it's going to continue to double down. But I'm wondering if if advertisers are are receiving that signal from Meta that it is going to get better. It is some place that they should be spending more time and more dollars.

>> Yeah, we're a captive audience. So I don't think he has to message to us. >> He doesn't he can just ignore you. >> Yeah. Like the best thing he could do is get people to spend more time on their app, right? And then better understand who those people are and what they're in market for. And if he delivers those things, the ad dollars will come. because besides that, like where else are we gonna spend money? Like Tik Tok shops actually doing great, but it's still a really small business, right?

Like the GMV this year in America might be 20 billion. Really? And Amazon does that like every four days or something? So it's like it's still a very very small business. >> Why is Tik Tok shop so small? Is it is it a separate panel or something? I feel like Tik Tok still has so many so many impressions, so many videos that are being served. Do you do you have an idea of why that business is? I think it's it's not necessarily I think it's primarily that content platforms have been a place you discover products but not where you transact like Meta has done a lot of had a lot of efforts in shopping in app they never you would have thought that they would have clicked harder right it's like you have this captive audience that uses your product for an hour a day to discover things to buy >> and they're not buying that many products in the app.

>> Yeah. >> Yeah. And Tik Tok shop is delivering way more value than 20 billion in GMV. like a a brand is like comfort hoodies. I'm sure you guys have seen them. They're doing a billion a year right now. They're four years old. >> Last time last time last time you told me about them, I think it was like 400 or something like that. >> They they've more than doubled. Yeah. It's crazy. And all it is is like Tik Tok shop affiliates. So people posting thousands of videos a day. They take those the best videos of it. They put it in the Tik Tok shop GMV Max. They run that. and their Tik Tok shop business might do might do 100 million a year, but the spillover is crazy. Everyone goes to Amazon. Everyone goes to your website. Got it. So, like they've just done a horrible job actually capturing the value they're generating.

>> Interesting. But they're getting a lot of impressions that are driving a lot of purchases. What's the sweet spot for price point on Tik Tok shop? That hoodie company. Is that a \$70 hoodie? >> It's cheaper, right? So, like it to a successful product Tik Tok shop is female focused impulse buy. So, like, you know, \$20, \$30, \$40, something like that. And then you have people to make outrageous claims, right? >> And so, comfort hoodies is like, it's a hoodie that cures anxiety. So, like, that's pretty good. Like, it's a pretty good value.

>> People claim the FDA might, wake up to that one itself. >> All the affiliates are making it, so who cares? >>

Yeah, somebody should care, but we'll see. yeah, a little bit of a gray area. Interesting. are there any I I I want to know about when you're using AI personally inhouse or someone on your team is using AI like the models or the products like chatbt, claude, etc. versus you feel like you're getting AI for free because you're using Mailchimp and Mailchimp integrated AI or you're on Shopify and Shopify gave you an AI feature for free.

>> Yeah. You know, I I don't want to talk too bad about any sponsors. I don't know if Notion sponsors you guys, but Okay, great. >> you would actually hope that more of these companies would be rolling out AI faster and better and more useful because like you know, we're using it internally a ton directly with the models and like inventory planning and buying is a solved problem. Okay. Is like the biggest problem that has plagued Yeah.

>> like demand forecasting. >> Yeah. Like all of that is like >> and then sending up is sending a proposal to a supplier or or in your case like the actual factory to to understand how many you're going to sell by month, when they need to be delivered, what with the shipping timelines, all of that's just a bunch of Excel work normally. >> Yeah. And it was huge teams and if you got it wrong, it bankrupted your company, right? Like you have bad inventory or you get like shorts in December, right? Like it ruined a bunch of businesses. AI has totally solved that. Working directly with the codeexes of the world, right? And just putting in all of your your business information.

such a huge unlock for the global economy. That's like crazy to think. >> We got to tell we got to tell consumers that don't like AI, your favorite brands are doing perfect forecasting. And that means that you are going to get the perfect article of clothing right before your trip to Hawaii. Even though you waited until 3 days before to order it, >> you'll always have the right sizes in stock. That is coming because of Codex.

>> That's crazy. >> this is this is this is like part of what the souvenirator was talking about, right? Like the the one of the Eric Su mobile bubble dev memo. he he was just talking about like actually if you have like better advertising because of AI, it will create a like like it will drive economic growth purely because you have more and more of these like niche businesses that might have an audience of 50,000 people in the whole world. Historically, you couldn't build that business because it would have been impossible to find that 50,000 people out of billions. Now, you can.

And then this is again an accelerant of like so much revenue is lost every day. So much purchasing activity doesn't happen because people just can't buy the stuff that they want because the brand >> didn't properly forecast and and is still kind of doing like fly by wire. >> Totally. The wrong sizes, the wrong things at the wrong time. like with how expensive it is to get stuff on shipping containers and how long it takes. and there was a lot of companies spent a lot of money trying to solve this. There was demand software that did billions a year in revenue.

as long as you have like a system of records, so like a clean data warehouse and you have your all of your sales from Shopify, Amazon, whatever else. You port all of that into a codeex and it will it totally has it figured out. And if you have to make those small tweaks like oh actually last year we were on a promo, this year we're not going to. Oh, so the multiplatform thing is big here because I would have I was just about to ask you like should

Shopify roll out a demand planning tool for Shopify merchants, but they probably Amazon has sharp elbows and won't give them all the data just via a simple integration.

>> Yeah. It all has to roll up into the harness. It all has to go into codec >> and so and so you have to get into a data warehouse and then you have to point an LLM at that. >> Yeah. So like you know we're not like you we're not like a large buyer of software like you know we we have Shopify data warehouse whatever else but as long as you have those like basic things and you put it into a harness it's like >> but your IT spends like yeah probably like less than 1% of revenue. Totally.

>> Yeah. Exactly. Which is which is the good benchmark. You don't want to be like building every system custom from scratch. >> Yeah. And people are so excited because they can vi vibe code everything, right? But like you know judge me reviews is like \$5 a month. So, I'm not going to vibe code my own reviews thing, right? I'm just going to do this. >> Yeah, just pay that. Yeah, >> but but yes, >> there's some review plugins that are very expensive. So, >> yeah, Bizarre Voice Horrible like Yeah, Yapo has a bad reputation.

>> I've gotten I've gotten fleeced a couple times. >> what's new in manufacturing land? Oh, >> you know this >> you you guys were trying to develop US manufacturing like way before like American dynamism was like a category like this is something that you guys have been like exploring and dabbling with for a long time but what's what's the latest? >> Yeah, so like we actually worked with the government in like 2022 to get something called like a general exclusion order. So like we have a we can we can very easily now get people to get banned from importing like if we if they value our RP. But to do that you have to prove that you're like a important part of the American economy.

So to do that we actually bought like the largest independent watch maker in America. It's called like a it's called FTS five time Solutions. They're in Arizona. So we own them. So if you ever buy Americanmade watch I probably made it. but it's a really hard business. That's crazy. Watches suck. so we do wallet production there. Okay. you know, we probably spent two or three four million dollars like getting that whole thing set up to actually produce wallets there. But then like the Trump tariffs made it really hard to get steel because there's a huge global tariff on all non-American steel.

But that means everybody wants American steel. So now it's really hard to get it. And we have to wait for those things to work themselves out. But >> look, most manufacturing is already very automated. Like you guys have spent time in China. I'm going back there next month. It's like they don't have that many people in factories. It is robots and assembly and that can be done basically anywhere >> and then it's just getting the raw goods to wherever you actually want to manufacture stuff. And with steel and batteries, we can't do that in America yet.

>> So we we have to build a supply chain. It'll be like a 10-year thing, but >> the future is going to be hyper local manufacturing for sure. >> I like the idea of you getting into steel manufacturing. That would be electric. Just make your own steel fully vertically integrated. Is now the best time in history to start a consumer brand? >> Well, I would say probably like January 2012 when Facebook ads just rolled out. That was probably the best time.

>> But now is the second best >> for sure. I mean, one, it's moed from AI. Like I would hate to be trying to sell software right now, right? And a lot of services. I'd hate to be in that business. People are going to buy stuff forever, right? You have birthdays, you have Christmas, the American consumer is still incredibly strong. We really just had the best Q2 of all time. And that's with a war in Iran like in >> and inflation and all sorts of stuff.

>> Yeah. And like it all actually like looking at >> consumer confidence is so low even though consumer spending is holding. you you you there's always nervousness about will there be a pullback? >> Yeah, there's been nervousness for like six years. But I'm telling you, the Ukraine war in 2022, there was a noticeable decrease in e-commerce activity. And right now, it's not happening. Things are actually ripping. So, I think it's a great time. I think >> total vibe session.

>> Yeah. >> Total disconnect between what people say in the Pew Research and then what people actually do. >> If I had my Shopify notifications on, it would just be chiming all day right now. I think people are >> Yeah, it's it's definitely a vibe session. >> so it's a great time to be selling stuff. >> That's great. What are you thinking on the creative side? Are you Higsfield maxing with like all these workflows to generate endless AI videos? Are you seeing progress in AI images? what what's working? We just talked to the suenator about the this study that showed that when AI is clockable, it underperforms, but when it's not identifiable as AI, if it's just a product image and it just looks indistinguishable from CGI or a photo, it overperforms sometimes.

>> Oh, dude. pull up my Facebook ads library and it is tons and tons of AI generated static ads. Okay. Right. >> Static. >> Yeah. Because the the statics are actually like you can build like ad factories like totally automated. So you know you take a hick field, you use an MCP, you bring it into like your harness of choice like a codeex and you can generate 10,000 static ads if you wanted to, right? And we just have that running 247. They get pumped into a Facebook ads library to test it and then the winners go to a different ads library to like a different ad account to actually scale those up. so the static stuff is totally solved and >> I would hate to be trying to do ads without it right now, right? I actually built a spreadsheet yesterday or like a presentation slide by hand and I felt like I was a caveman. It's like that's like what working with like the ads of like the past were like video we still do a lot of it. It's mostly just as like cut scenes in like a hyper cut. So like we want to add motion or somebody talking or whatever. We'll do a lot of that. but it went viral yesterday. The like the Seinfeld fully AI episode >> that is really really good. It's getting very close to actually being indistinguishable.

>> I thought it looked really good. I thought the editing pacing was wildly off. Yeah, I agree with that. I thought there was like the gaps in the humor. Like the pacing is so key to that. It was like but from a fidelity perspective, it looked indistinguishable from the show. >> Yeah. So, it's like it's very inhuman in like the the way they talk or whatever. But like that's going to be fixed. like that's coming in two more updates, I'm sure. And then and then it's like, yeah, all video. What about using AI either to write deterministic scripts that can assemble hyper cuts in different because if you have a picture of the wallet, a picture of a person putting in their pocket, a person stepping out of a car, a person on a beach, you might want to sequence those 1 2 3 4 2 1 3 4 3 2 1 4 and get every possible variation on the th those different video clips. Are you using AI? Do you already have a system for that? Is that already automated? Is there what's the future of that?

>> Yeah, that's still very human in the loop, right? So, we have like, you know, two amazing editors who would do all the hyper cutting themselves and now they are using Hicksfield and just getting, you know, hundreds of more variations. Yeah. Going through those in the upper mode to the app. >> Got it. Got it. And they can probably even do some like style transfer and filtering on top of the raw footage that they have to like >> Yeah. And sometimes like, you know, it's a robot. It'll it'll make stupid decisions. It's like it'll it'll go like, you know, wallet to something, you know, falling in the ocean. It's like, what the hell are we doing?

>> Yeah. It's just complete hallucination. >> 5.6 six. People are reporting that it's it's working on video editing workflows now in a way that a lot of other models haven't. So, >> I wonder how that will actually play out because there's one world where you're just literally opening Premiere Pro and saying like move the mouse cursor and make the cut in the footage. And then there's another one where you're like editing the underlying file and then you're watching it in Premiere Pro because most of these video apps, they sort of represent the file as like a structure of folders or you know a bunch of JSON or something. So you you you can manipulate things multiple ways, but we'll be interesting to see where that where that goes even if it's just for like those that the the re configuring sequences and whatnot. Are there any vibecoded e-commerce plugins or add-ons or tools or software that have stuck out to you as wow like this thing it was in the \$1,000 a month category now it's in the \$5 a month category or or this is a new capability that's unlocked that's still something you wouldn't roll your own but you would buy outside >> you know not really but like a lot of the playbooks that people share. It's like, you know, how to build landing pages in one prompt or whatever. And like that is going that's like going to one shot companies like Shogun or like there's all these like landing page builders and it's like they're just, you know, drag and drop tools, but you're getting way faster, way more responsive, way better stuff just out of codecs and it's like it's completely on brand with your assets and it's like look that is vibe coded, right? It's and we we're going to launch 50 landing pages next week and it's all vibe stuff like that.

>> What's the value of landing pages these days? Is that's is that critical to the funnel? Is that critical to the >> Oh, for sure. >> Yeah, >> dude. You guys had Hermosian here like two days ago. I listened to Finn. It is >> he's got 1 billion now. He's going to get a landing page for every human on Earth. That's actually where we're going. >> Yeah, basically, right? >> census records, >> but it's like you need the offer to get the click, right? And then you need the landing page to inform them and then the like as fast as you can get them to making the purchase as possible, right?

>> in sales, they tell you to use the person's name a lot. I'm really happy that you're here, Sean, cuz I want to talk to you about this. do you think we get to the point where like digital platforms end up, you know, and ads and landing pages are like using the individual's name >> to an extreme degree? Like because you get down to targeting one person. Why wouldn't you say >> that long time ago? >> Yeah. Yeah, but would would eventually there's like a >> there's a there's a flipping point where maybe it just is so effective >> that it makes sense for Facebook to enable.

>> Well, all AI cold email right now already puts your name in there and there's been beta tests rolled out where they're using people's faces in the ads. It's like like >> Oh, yeah. We saw that on on remember? >> Yeah. Like they I think they were trying to sell a pair of the meta glasses where it's like showing like you calling your you

know and they can tell like who your significant other is just based on your activity >> on Instagram.

>> Yeah. So Facebook has all that information. It has your face. It has who you're talking to. It has who your wife is. So it's like they they look hyperpersonalization is coming for the entire web. And >> you know if you're going to be shopping it's going to show what you're going to look like in the clothes or what your dad's going to like when he gets the wallet. I definitely think that's happening. Yeah. I mean, if the price of a cold call goes to a penny, do you think you'll be cold calling people?

Something just came across my desk. I got a wedding ring here for you. >> We've talked about it. >> Talked about it. >> Yeah. There's a thing called like a ringless voicemails. Well, like, you know, they'll just mass drop those off. >> And they put people's names in there. It's like, >> "Hey, John, we got this thing for you." Right. And it's like, you know, it's no it's one way, but yeah, the two way is definitely coming.

>> Interesting. >> What's happening in luxury? LVMH and Caring are down like 25 and 20ish percent. >> Well, we talked about the last time I was here. I was like I'm like, "Oh, yeah. Gucci's getting crushed and and it's like, yeah, it's going to continue to happen." >> Why is that? >> You know, it's I think it's just a generational change is is really what it comes down to. Like each one of those brands have good assets, but like Richmond is is still tearing, right?

Hermes is doing great. Coach is on like a generational run. like the best performing stock over the past like two years, right? I think last year it beat Nvidia in performance. No way. Ouch. >> Yeah. So, it's publicly traded under Tapestry. You pull them up. they own Kate Spade, too, but Kate Spade's like a nothing business. so, it's just it's just a generational rotation, right? Ralph Lauren also on a tear. So, you talk about like American Dism, there's American luxury. Like America wasn't old enough to have luxury brands, but like LVH bought Tiffany's and now it's like look, Ralph Lauren's crushing, Coach is totally crushing. And I think the old stodgy LVMH's of the world, the Gucci's of the world, one, they got overexposed.

Like they really rely on the middle class. And if you ever look at like their sales demographics, it is like 80% of the revenue come from people making under \$12,000 a year. And it's like that's just a big disconnect when you're trying to sell whatever, right? so there's going to be a shift towards like, you know, true luxury. There's also some concern that like El Cton is just personally investing in all the great assets and not bringing them into the portfolio. like they own crumbarts and they they haven't brought that into LVMH.

>> Oh, interesting. >> Yeah. >> Okay. Well, but when you say personally investing, it's El Caterton is investing. Got it. Okay. >> And El Catton is the investment vehicle >> of LVMH. Yeah, that makes sense. And I know that they do a lot of deals in the private equity world, but Oh, yeah. Why Why doesn't LVMH >> like you imagine the Arnos are like, "Yeah, we want to own the hottest." Like, you would think they would do everything they possibly could to own Chrome Hearts. There's a divestment in brands right now. They're actually trying to push stuff out of their portfolio. I think they just liquidated Off-White, right? And Off-White ended up being like a Target like Costco brand. Like there's Yeah, there's like photos of like Costco Off-White like big pallets getting delivered. So, they're actually doing a divestment right now from brands. and you know, Tapestry just

devested from Steuart Weissman. So, like they're actually trying to go like bigger down on winners. But yeah, then there's just like a whole like you know, Mad Happy is a great brand. It's invested via Elcadon, not inside of the portfolio.

Interesting. And the whole idea was there was supposed to be a scout fund to then bring it into the main portfolio. But if you're the North family and you own 90% of LCON and only 45% of LVMH, it's like what's what's the incentive? >> Interesting. Yeah. Yeah. >> But look, I mean, you could like Richmond owns Cardier. Cardier is having a great time. you know, they own Van Clee. Van Clee is still having a great time. But those are both of those are way more true luxury brands than an LVMH. Yeah. So, >> have we hit peak timu and sheen?

>> Oh, yeah, dude. I mean, Trump got rid of, section 321, dimminimus. That totally crushed those brands. It's like I I mean, >> you know, team, >> where do you go if you want to shop like a billionaire? >> There's nowhere else to go. >> Yeah, You're right, man. I don't know. Bridge.com. no. look, I mean, they they're still huge. They they run a lot of ads and like Team Moose publicly trade under like PDD or whatever. So like they have a huge business in Latin America, they have a huge business in Southeast Asia, but like you their whole arbitrage was just flooding into America with with low cost goods >> direct from the factory, right?

>> Yeah. And it just kind of that kind of blew up. >> It's gone. Interesting. have you been surprised that live shopping has been slow in America? >> No. >> No one was predicting this from China. Oh, if it's big there, it's going to be big here in a year and it feels like it's been very slow. But what what have you seen on the live shopping? >> Yeah, look, people are very excited about whatnot. Yeah. and >> they are putting up impressive GMV growth, but it is very much dependent on like the trading card bubble. It's like that's that's what it is, right? So, it's live shopping in America.

>> Yeah, this is what I was saying yesterday. It's like it's effectively >> I'm not going to call it gambling, but there there's some speculation happening on >> there's a there's a massive I would say spec you would want to figure out what percentage of GMV is driven by speculation and I would expect that it's significant. >> Yeah. Like people aren't going on there to buy their everyday essentials, right? They're going on there for the excitement. Like it is an auction.

>> Whereas in China there's live shopping where you can buy a tomato >> and that's just not happening. and women are buying dresses and it's the whole thing. >> Yeah, of course. The whole person put picking up one piece of clothing, putting it down, picking up the next one. we've seen that video. and we have yet to see that in America. >> Yeah. >> Why is that? >> I think I mean, dude, >> Americans can't watch anything at the same time, right? Like we are also like consumption based on like like on demand everything, right?

>> everything is catered to us all the time. I just don't think there's any value to it actually being live. We're we're doing Tik Tok shop lives and we drive like \$300 to \$400 in in revenue per hour that we're live. So it's like some people are buying stuff when it's live, but the real value is like getting all that content and just running it whenever you have four hours at night to to go log onto something, right? And I really just think it comes down to there's a lot more people in China and they are spending a lot like the the screen time per person is still way

higher over there and just you know it developed over there. It's more of like a native sport.

>> Yeah. Have you looked into advertising on Netflix? >> we have looked at it. the CPMs are not very good. >> Okay. >> And here's the thing is we buy a lot of TV ads and we buy them like like directly through the networks. So like you know Fox will have something like we just got an email for the World Cup and it's like >> you could run a 30 secondond spot during the World Cup the USA game and it was like a \$25 CPM, right? and Netflix >> they you'll get a lower tier of consumer because it's price gated, right? It's the lowest tier consumer gets the ads and they want a \$45 CPM, right? It just it doesn't make a lot of sense.

>> Yeah, that makes a lot of sense. what about >> infomercials? Have you ever thought of running one? It's like a full hour, middle of the night type of thing. I've >> That's the new >> I've wanted to get. I've wanted to It feels like a bucket list item on an entrepreneur. >> when you have the studio, bro, let's shoot one. >> We should shoot one. I I'm I'm down to be >> one hour pitching you a three-hour live stream. I interest you in hitting the subscribe button, potentially leaving us five stars.

>> I wanted hyperpersonalized infomercial, though, where somebody's just like falling asleep on their couch. It's 1:00 a.m. and you're like, "Hello, Sean." >> I I feel like I've seen some of those YouTube experiments ads where they they didn't gate it and so if you didn't click skip, you'd wind up watching like 12 minutes of an advertisement. >> Oh, dude. know I IKEA did that but it was like 12 hours IKEA's like we're going to read every product we have but please click skip right >> fantastic >> >> but you've never done an infomercial >> no but I met a guy one time who made like you know his whole business \$80 million a year was selling hoses on infomercial >> and he's like yeah I have a great hose and I just sell it on infomercials and then he gets into Home Depot or whatever so look you can make a lot of money in a lot of different ways so >> yeah but you got to you got to focus a little bit you got to get things going >> yeah right now I'm trying to sell like a bunch of random accessories to men and my goal is to get to a billion a year in annual revenue, I'm like three or four years away. And if I do that, my life's good.

>> Fantastic. Well, thank you so much for coming on the show. >> You want to close it out with us? >> Yes. >> All right. >> We got one last advertisement. It's for MongoDB. What's the only thing faster than the AI market, your business on MongoDB. Don't just build AI, own the data platform that powers it. Thank you so much for coming on the show. Leave us five stars in Apple Podcast and Spotify. Sign up for a newsletter, tbp.com. And we will see you tomorrow at 11:00 a.m.

Sharp. Goodbye.