Alastair Unwin (Polar Capital) – Investing in an AI-first World

Tom Yeowart: Ali, welcome to the podcast. Thank you very much for coming on.

I think you did history at Cambridge. George and I are both history grads. Good choice. And then you ended up, I think, starting your career with Herald as a technology analyst before moving on to Neptune and finally ending up on the Polar Cap technology team in 2019. But just going back to the beginning, how did you end up in investing? Why technology? And also, just how you've evolved as a technology investor over that time.

Ali Unwin: It's been a really fun journey, but it didn't start terribly well because I graduated in June 2009, which you may remember was not one of the great periods for the graduate job market, particularly in financial services. So, I'd been lucky enough to intern at YouGov, the political pollster. And they thankfully said that I could come back as a sort of grad and I worked for Stephan Shakespeare, who's the CEO and founder, and actually still the CEO now, which was a wonderful first job. We did polling for number 10 and we did some other interesting stuff, but we did more and more with Bloomberg and with financial markets, and I became increasingly interested in how I could move into that.

And Herald was one of YouGov's investors at that time, which is an unusual route into the industry. I'd known Katie a little bit as an investee company. And she was obviously an extremely well-regarded investor. So, I was very lucky to convince her that I could go over and start my career there. So that was a wonderful place to start.

In terms of why tech? I'm not a technologist. I'm another historian, and Ben's an historian as well, the lead manager of PCT. A plague of historians. What's happened. Exactly. But my interest or fascination has been more on the economic history side and the idea of technology as one of the key drivers of history, the sort of lever of riches as Joel Mokyr has it. And it's just been a fascinating area to work in to see how technology is to an increasing degree, even today, driving the global economy and driving productivity growth. But it's just, to me, the most interesting, most exciting, most important sector from that perspective.

Tom Yeowart: And I imagine Katie Potts is a very different investor to the likes of Ben Rogoff, who you mentioned. What have you learned along the way

from working with the likes of Katie and Ben and the rest of the team at Polar Cap?

Ali Unwin: I've been extremely lucky and to have come attacking tech investing from several different perspectives. Firstly, at Herald, which was very, very bottom up. Lots of company meetings partly because of the sort of holding period that Herald has when liquidity is somewhat more difficult at the kind of small and micro-cap end of the market. So, we would be seeing eight or 10 companies a day at some points and just an extraordinary range of businesses. And Katie can read a balance sheet in about 10 seconds in a way that I still struggle to. But the idea that you could see companies over the years and learn what worked, learn who was reliable, who was less reliable, learn what kind of things you wanted to look for, what the red flags were.

So that was an amazing bottom up grounding I suppose. And then Neptune was more of a top-down shop, which was very macro led. It was young, bright people who were given money to run relatively early in their careers with somewhat sort of sink or swim approach but was extremely dynamic.

And then I was lucky enough to move eventually to Polar, which is actually a bit of both. Where we're a team of 11 people, we have a very robust process. We're doing a thousand company meetings a year, plus. There's a huge amount of bottom-up work. But there is also a thematic framework and some top-down work. So for the trust, we do a strategy piece every year for our board where we go through all the themes we think are important, all of the drivers of the markets we're in, where we think the markets are going, what's worked, what hasn't worked. The idea that tech is now pervading across all sorts of different sectors and that's been really interesting. So, I've learned a huge amount from all sides really. And but still feel like I'm relatively early in that process in many ways.

Tom Yeowart: You touched earlier on the dynamism of technology investing and it's interesting to your point that technology is broadening out. I'm just interested in what mindset you need as a successful technology investor given that dynamism but also given technology is relevant across pretty much everything you can think about.

Ali Unwin: I feel a bit, premature in my career to answer what makes a very successful technology investor. But I think some of the things we've observed and ideas we've coalesced around at Polar in particular are firstly that technology change is non-linear. And that makes it quite different from other sectors.

That means mean reversion is not the base case. Winners tend to win by more. Winners take, if not all than most. You've seen that in smartphones. You've seen that in internet categories. The debate at the moment is whether we're seeing that in AI and there are certainly some signs that, that looks like being the case.

And so, you need a process that can take advantage of that and deal with that reality. And that means you have to have a process that doesn't miss the big winners. You have to have a process that can get you out of companies that are not going to be winners, even if they may have other merits for other investors or from other points of view.

So, I think those things have been very important. The other thing which is more of a sort of Polar mindset in particular is that hubris is your enemy in tech. The sector is full of very smart people doing really interesting things, that always have the answers. And no one will ever tell you they're the wrong side of any tech trend, of any trade. And these are some of the most convincing people in the world. We have to not be hubristic in our own interpretation and look for where we can be wrong and look for where the challenges to our thesis are. Which isn't the same thing as not having conviction, but it can mean that the best stocks to get exposure to a theme, the best conduit for a new technology cycle may change.

They may not be the first generation of them, they may not be the second. We saw Google was not the first search engine. It wasn't even clear that search was going to be the way we would access the internet, for example. At the beginning of new technology cycles, when things are still in flux, it feels very important to retain an open mind. The ability to change your mind. Liquidity as something that you need, not just to move the portfolio around, but actually as part of your risk management, the ability not to get boxed in. And I think that's been something that's come out of working at Polar, which is an unusual focus on that sell discipline as much as the buy. Everyone loves talking about the winners and the buys but actually getting out of the stuff that isn't working, recycling the capital has been a really important part of our process.

George Viney: We'd love to explore your views of AI. It feels like the dominant theme of the day, maybe of the computing era. We're coming up for three years post the launch of ChatGPT. Where do you think we are in the development of this technology? How do you see its significance? And how do you separate the hype from the reality?

Ali Unwin: Well, being an historian, I'd go back a bit further to put it in its proper historical context, which for us would be that AI has been a domain of

intellectual discovery and work since roughly the fifties. There was a conference at Dartmouth, a famous conference in 1956, where they thought they'd solve AI in a summer, which unfortunately didn't happen. And then since then, we've had various bits of AI progress, various AI winters where we haven't made much progress. There have been different approaches that have been tried and ultimately failed until something happened in 2017, which was the emergence of a new way of looking at AI with the emergence of something called transformer models, which came out of Google. And this provided a new way to build AI models, which meant that they could get bigger much more quickly. And since that point up to and past the launch of ChatGPT in November '22, we've seen the continuation of something called scaling laws.

And these are not laws of physics. These are observed statistical relationships between the amount of data you use to train a model, the amount of compute you use, and the number of parameters in the model or the model's sophistication if you like. And as those things have grown, the output of the model has improved. The test loss ratio has fallen is how it's technically termed. Those have continued to hold as you increase the amount of compute, increase the amount of data, increase the parameters of the model, all the way through to date actually.

Then last year in 2024, we had the arrival of a second scaling law, test time scaling, which is not just the fact that bigger models perform better. It is that the longer that the model quote unquote thinks about a problem, the more cycles it goes through at test time, at inference time, the more tokens it uses, the better the output as well. So bigger models are better, models that think for longer are better, and that has some pretty meaningful implications downstream for things like agentic AI and even AGI one day.

So, there's been a huge amount of change. Progress at the fundamental level continues, and that for us is something we spend a lot of time focusing on because if scaling laws were to slow down significantly, then we definitely have the wrong portfolio, and we would have to change it. For now, we think they are continuing. We think the industry participants suggest they are continuing from their words, but more importantly their actions which we can get into. But that has been the prime driver of what has changed in AI to get us where we are today. That's almost the why as it happened now.

Tom Yeowart: What do you consider in terms of the continuation of that exponential curve and, perhaps some of the potential bottlenecks that will slow that rate of improvement down? Obviously, there's been a lot of discussion on the quantum of data, the shift to synthetic data, I guess some of the sort of more

physical energy related bottlenecks. But as you and your team are thinking how sustainable this exponential path is what are the key things you are looking at and thinking about?

Ali Unwin: I think in any new technology cycle almost the more important or the more profound, if that's not too grandiose a term, it is, there will be constant challenges to the trajectory of the technology's improvement and the improvement matters. If you believe that AI is best understood as a general-purpose technology, which we do at the moment, then it has to be pervasive, it has to continue to improve and it has to have spillover innovations. And so far, those things look like they are all in place for this to be the next general-purpose technology in the style of the internet or steam or steel.

So, while there have been bottlenecks around the amount of data you might use, where you get the power, chip supply issues, in scaling Nvidia racks, all these things are quote unquote unique but normal for a new technology cycle. And one of the challenges when something like DeepSeek happens is that on the face of it, it sounds very bad that a model was trained for $6 million rather than a billion dollars, and it wasn't using the latest Nvidia GPUs. And how can this be good and isn't Capex going to zero?

Very quickly, but not instantly because Nvidia's share price move told you it certainly wasn't instant or instantaneous that people did understand this, you learned that this was a text only model. And actually, this was trained on quite a lot more GPUs than we thought, and it was only the final training one, et cetera, et cetera. And the market quite quickly comes to the view that this is another plot on the curve. And it is a challenge in technology, particularly this deflationary nature, because when GPT-4 came out, it cost about $50 for a million tokens, and a token is roughly a word, and it now costs about 15 cents. And that's using GPT-5, which is a far superior model.

That level and that speed of deflation is disconcerting because if you used to charge X for it, if you're getting a thousand times cheaper, it feels like, how can you maintain pricing? And of course, the answer is Jevons paradox, that as you make the use of the resource, you will see an explosion in demand. And we've seen that with other technology cycles, but you see the cost come down before you see the demand response. And that is a feature of technology cycles. We've seen that with the cloud. We've seen that with the internet and bandwidth. We've seen that with electricity; we've seen that with lighting even. This is the shape of technology change, but it does not make it necessarily easier through the sort of pointy bit where you are seeing the cost down, but not yet seeing the revenue come.

And we can talk about ROI on AI and how people think about that, but I think for us the focus is on what are the participants who have the best visibility, the best vantage point if you like, which for us is the hyperscalers. What are they doing with their own capital? What strategic moves are they making? So, the real proof point for DeepSeek not ending the AI trade was not that we felt better about it, or we did a load of work and tried to come to a view of what it meant. It was that all the hyperscalers raised Capex after Q1, all the hyperscalers raised Capex again, after Q2, you have to take the sort of inference to excuse the pun from that, that nothing has changed. Or rather their conviction in it or their need to invest in it, or their desire to invest in it has actually increased.

George Viney: How do you judge this technology and the capital that forms around it versus prior precedents of technological change when the capital is abundant and the hyperscalers you referenced, have enormous resources to throw at this and that ChatGPT and large language models in their current form at least, are being distributed through a very mature distribution system, through the internet that's already broadly adopted. How do you think about that in terms of the speed of the Capex cycle and the speed at which the technology is distributed and adopted?

Ali Unwin: The diffusion rate of general-purpose technologies has been coming down over the past 150 years. If you think it takes steam 80 years to get to let's say 50% penetration, it takes electricity, 40 odd. Like the ICT revolution as it used to be called, was 20. Cloud and smartphone was about 10. Is five or six a good starting point for AI? We think that's probably a reasonable base case. I think the data we've seen so far ChatGPT at 700 million weekly active users, and bear in mind that is 700 million weekly active users into an internet user market that is already saturated with Meta and TikTok and video gaming and a million other things to do.

So, they've managed to get to extraordinary scale and usage, and they put out some data last week about how people are using it. But the fact is that we've seen extraordinary adoption, and that shouldn't be a total surprise because the base rate is going down for how quickly things are adopted anyway. But also, a lot of the building blocks, there are 7 billion smartphones with broadband internet connections in the world. All of the data that we've had built up over the internet to train the models over the last few years, all the investments that were made during COVID to digitize the, at least the front end of a lot of businesses and functions, have been made.

So, you are standing on the shoulders of all those prior investments to your point, it is a mature distribution network. Could that compress the capital cycle

for its distribution? Potentially. But if we look at where we are in terms of the proportion of GDP spent against other general purpose technologies or massive reallocations of capital to a new infrastructure at the macro level, we still think we are quite early and it hasn't been going on that long.

People look to railroads where you get to in aggregate, I think in the 1840s, 20 or 25% of GDP is spent on building the railways. Now that ends in a terrible bubble. As indeed many historical technologies arrivals have been commensurate with, and this one may end in a bubble as well at some point. We think we're a long way from that today, and we can go into why we think that. But for us, the existing infrastructure is the mechanism through which this technology will be spread. But actually, what's happening now is we're building out a new infrastructure. AI is not going to be run on the same infrastructure as the cloud, and that is why you see the CoreWeave's and the Nebius' and Oracle's extraordinary bookings number. That is a function of the fact we are building a new parallel compute architecture, which is not like what came before. And that is why we're getting to these very, very big numbers very quickly.

The capital formation that has to happen around that. It's common with new technologies, innovative, sometimes a little bit unnerving saying on the back of a hundred-billion-dollar investment from Nvidia in OpenAI yesterday. These things are new. When Google came to market, they came through a Dutch auction IPO process that no one knew. They had A and B share classes. Everyone was like, what on earth's going on here? There was share-based comp, which I remember having endless conversations about, and in the end, for the winners, it didn't really matter. And for the losers, they lost anyway. So, there are always these new features that go along with it. They feel uncomfortable. Financial innovation is a sort of two-sided coin, can enable things that weren't enabled before, but bring risks that obviously we don't really know how to price necessarily.

George Viney: One feature that's striking this time around is that so much of the action is in private markets and that as public market investors we can participate through various ways. And clearly Nvidia is the standout success story at this point of this technology cycle. But one of the reasons I think why Oracle's shares did what it did on that day is because it was a way of finally investing in OpenAI.

You've got this beast of OpenAI that sucks up attention and capital from an amazing array of different investors, venture capitalists obviously, but sovereign wealth funds and so on. And you mention the signals that come from the hyperscalers. The signals are diffused somewhat from private markets into

public markets. How do you assess all these developments from a public market investor perspective when so much of the action is happening in private markets?

Ali Unwin: It's a, it is a very good question. I think while we are in the infrastructure build phase of this technology, actually as you say, a lot of the action is happening in public markets and Nvidia is the most obvious and largest by market cap example, but actually we are redeploying capital into a hardware cycle.

And that has been a challenge for many investors who, including myself, who have only invested through a software cycle where high ROIC and asset light businesses were good and everything else was bad. We are now in a hardware cycle where actually for a period as there was in the mid-nineties, you can get paid for having the best-in-class hardware. And you can build custom hardware and there can be a return on that. And there are shortages everywhere and people need power, and they need chips. And networking stocks making money for the first time in 25 years. By and large these are public market things.

So, the infrastructure trade has been actually pretty largely felt in the public markets. And we've seen, the markets at all-time highs, the trust is at all-time highs because of that. In terms of the lack of visibility or signal from what's going on, because OpenAI and Anthropic and some of these labs are private, xAI as well. There will be some signal that we don't get that maybe investors who are around those tables do get. But equally, you can still stay pretty close to this stuff. A lot of this is playing out on X and on blogs and podcasts around the world all the time.

Eventually as well, we think these companies will come to market. Now the timeline on that and what that'll look like and how much will be left, remains TBD. But we are confident that if the capital requirements for the reallocation of capital, if you like, from labour to compute, to put it crudely, will require public markets at some point.

Tom Yeowart: There's obviously a lot of people making the contrast with the 1990s and the technology bubble more broadly. And just want to dive a bit deeper into what gives you that confidence that there will be the demand to match the investment and ultimately earn a good return on all that Capex?

Ali Unwin: I think the analogy we look to is the mid-nineties, actually before the dot.com boom went crazy. And that very much was a bubble by the end. But in the mid-nineties, between '95 and '98, the NASDAQ returned 350%. So, it's a

phenomenal period for returns. But during those three years, there are seven drawdowns of more than 15%. That is just part of the course. The industry is moving at such speed, the capital deployments so fast. Things will go wrong. All of these things are normal, but the volatility they bring is in our view and we would say this better held through a diversified product with specialists being this all they do. If you are rifle shooting these companies, that can be quite difficult because you may not end up A), in the right ones, or B), knowing when to get out when it goes wrong.

The ROI question, which we're asked in many ways very frequently because you see Capex estimates go up and up and we're now at whatever, it's $450 billion for next year for the hyperscalers. Like these are very, very large numbers. I think that's triple what they were spending just before ChatGPT came out. So, we've seen a step change and even a narrative change from these are capital light businesses with huge user bases and high incremental ROICs to now something that looks much more capital heavy and much more build it and you will monetize it at some point.

I think you have to step back and think about the motivations for that Capex for the companies who are making it, and the sovereigns who are making it, and indeed everyone else who is making it. For the hyperscalers in particular, these are companies who, in economists' terms, are earning quasi monopoly rents or monopoly rents, and AI looks disruptive. And if it is disruptive to those quasi monopoly or actually monopoly rents, then it is a big problem. And the price that you will pay and the investment you will make to protect your search business or to protect your social networking business is very high. That is an existential we spend to defend. That has been arguably the knee jerk driver of the rate of capital formation and capital deployment.

You also have a race on. This is why we always end up coming back to the scaling laws. While the scaling laws hold. And as long as someone is willing and able to make those investments. And the scaling laws say that if you make the investments, the models do improve. If you don't make them, you are letting someone else win. So, so long as OpenAI is investing, Google has to invest. You've already seen in the data that the Google share of online query activity, if not search itself is dropping pretty fast. That is a huge problem. That's a terminal value problem for Google. It may not matter for numbers the next 12 months, but that is a terminal value problem. Now obviously Google, you have opportunities on the other side of that as well. So, all is not lost by any means.

I think the second category is the offensive, which is what people normally mean when they ask what's the ROI? Where's the revenue come from? Well,

we've seen $13 billion of AI revenue last December quarter in Microsoft, which has grown significantly since then. We've seen Meta talk about $20 billion of uplift to their existing business by better targeting, better optimization. Like these are big numbers. They're not numbers that justify 500 billion in Capex yet, but they are big and they're growing quickly, and they're companies that have huge user bases.

And if you can get AI to give you another turn of the crank or two or three turns of the crank, you get to big numbers quite quickly. So you have the defensive drivers, the offensive drivers but I think most importantly, and this is the bit that's hardest to price, but they do talk about the leadership of these companies, Mark Zuckerberg will talk about this very openly on earnings calls, is the optionality spending. The return to getting to AGI, getting to super intelligence, being a leading player in agentic AI. The race, if you like. That is very hard to price. And I think Zuckerberg was interviewed quite recently where he said it would be a shame to misspend a hundred billion or 200 billion, but he said not as big a shame as not spending it and missing out.

And I think that's something that we take seriously that some investors don't. And I understand why they don't because it seems a bit sci-fi and AGI being this position where essentially a computer can perform at human or better level or expert or better level in anything. And then beyond human capabilities in anything. Dario Amodei's data center full of geniuses kind of idea but in your pocket is a bit mad. And then you have cognition or intelligence or knowledge work no longer being a constraining factor on the production function of the economy, which to me is a very, very interesting dynamic. And there'll be other bottlenecks for why we I'm not sure we do get to 10% GDP growth or even faster as some of the accelerationists would have it, but I think there is a strong possibility that AI is able to deliver a material uplift to productivity growth.

So, I think there's a lot of case for optimism and when you put all those motivations to spend together, and everyone else is, how can you not spend. The ROI is seen in that context. The problem is that's what people say on the way up until at some point the money runs out.

George Viney: So where do you put hubris in that? The company I'm thinking of is xAI, because A, they're private, so they're out of the line of sight of many public investors, but they don't seem to have an obvious path towards monetization. But there's an empire of investments and technology that may get used in the future. And they're led by somebody who has enormous ambition. It feels like this year and next, it's the major test of scaling laws because they're building out these huge training clusters to really beat one another at who can

build the biggest training cluster to figure out where scaling laws begin to fade. So do you think that 2026 is the year to decide whether these models need to converge and therefore consolidate because we don't need 12 providers of foundation models, we probably need three or four. Or do you see a different path where this whole thing can continue to go because we're going to find out the results of this experiment and that could lead to outcomes that we're not really understanding at this point.

Ali Unwin: It's a very astute observation that if you think about Grok3 trained on a hundred thousand coherent GPU cluster, Grok4 trained on 200,000 GPU coherent cluster. The next generation will be trained on Blackwell and then Nvidia is putting out a new chip architecture every year. The chips are moving on at an incredible pace. The networking is lagging that but catching up quite quickly.

So long as you can physically build those larger clusters, you can supply them with power. You have a powered shell, which is a variable constraint. Again, there has been a supply response to that because when you are willing to pay twice the going rate for power, we'll turn back on Three Mile Island. You do see a supply response, so I'm a bit more sanguine about the bottleneck bit of doing it.

The amount of capital you start needing, and this is why we're seeing these mega deals gets very big, very quickly because scaling laws are non-linear as are many things in tech. At what point do you run out of capital? I think won't be a problem for '26 or probably '27, but you are getting to some very big numbers. And then you have to start thinking about is this more of an economy level reallocation of resources just to compute ultimately and compute and power and storage and network and everything around it really to compute because we are still seeing these models get smarter and more and more capable.

And I think that will depend on how good GPT-6 is or how good Grok5 is, or the next generation models. If these models are an order of magnitude better again. And you'll know that from benchmark performance, you'll know that from scientific breakthroughs starting to happen. OpenAI have talked about, you could basically hire their latest model at $10,000 a month as the most intelligent developer you've ever had. If you actually see people doing that, AI moves technology from being a 5 trillion a year spending category that's growing and really important to something that is eating into the global wage bill, which is in the sort of 30-40 trillion quantum.

At that point, your 500 billion of spend looks quite small. And if the models are getting better, then the aperture is opening at the same time for the number of tasks or occupations that are in scope. If the models slow down and everyone levels off and you have time to catch up, then that's not the case. So, it really does come back to the idea that models will continue to improve. At what point the actual ceiling will be reached in the amount of capital that can be deployed. I think we don't honestly know yet. You've started to see these big deals, but there are huge pools of capital that have still not really been tapped in this.

I think Morgan Stanley had an interesting note out talking about the need to find or get $3 trillion to pay for AI investment through 2028, I think it was, about half of it comes in the hyperscalers themselves, because these are the most cash generative AA or AAA credit rated businesses. So, in that sense, it's very unlike 2000, just to touch on the bubble question, when you had much lower quality credit and it was all coming on the balance sheet of new telcos and that kind of stuff. That is half the source, the other half will be third party investors, debt markets, whether that's asset-backed securities, sovereign wealth funds, we don't yet know.

But the fact is, if these companies believe there is a return on that investment in GPUs, they will continue to make it so long as they have the capital. That's our base case.

Tom Yeowart: Do you expect there to be more deals like the one Nvidia's just made with OpenAI, where the funding comes from interested parties within the ecosystem. There's an interesting circularity to the investment where obviously Nvidia is benefiting hugely from this Capex explosion and can see a very direct return on investment. The promise obviously is that as the models get better, you expand into a broader audience and you tap new revenue streams, but it's a promise which you have to take on faith. And so, as the sums get bigger, do you expect more and more internal financing within the AI ecosystem?

Ali Unwin: There may well be some of that. I would expect there will be more deals where tech corporates with very strong balance sheets stretch them further or push them a bit further because the race is on and the prize for the winner, if you believe scaling laws hold and we reach AGI or super intelligence is vast and it's in the trillions or tens of trillions, even as we've talked about.

The other place the capital could come from. And we're seeing this more and more is the disruption AI is bringing to other parts of the market is starting to show up and it's starting to show up in areas and ways that I think people didn't really expect. So, in our team, we've run an AI fund for the last eight years,

which is a global equity fund, which was set up to try and capture the beneficiaries of AI outside of the tech sector primarily. And in the early years, that was quite difficult because people were using machine learning and image recognition stuff and there were some robotics bits in and automation, but it was a challenge. That is becoming much easier. The pipeline for that fund has exploded because you're seeing AI appear everywhere.

But it's getting to a bit more of a pernicious phase of AI and some of the capital that has been allocated to, you pick on the stocks that have gone down, but a Gartner or a Wolters Kluwer. Fantastic, storied businesses with fantastic ROIC and great histories, are getting torn to pieces because people have realized that AI is competitive. The idea of terminal value risk appearing on people's doorsteps that no one thought it would appear. Proprietary data, which was supposedly the great moat that would keep you safe and you could generate more value by becoming a decision-making tool rather than just a source of data. AI is now chipping away at that moat.

And I think investors will increasingly feel the effects of AI negatively in their portfolios. Perhaps before they feel some of the positive sides. And that may also unlock capital that would've been deployed into other, as we would see it, legacy industries that can be deployed in AI.

Tom Yeowart: Do you think with the sort of foundational models it is winner takes all, or do we end up with, whatever it is, 10, 11 or fewer models all similar to a certain extent. I think it was Dario Amodei saying they're like humans. We've all got different personalities. And so even if they're similar, they're also different. I guess the question is around you spend all this Capex, do you end up with something that is quite commoditized? And unless you get to AGI where maybe its winner takes all, actually all that investment, again, struggles to make a return because there's just too much competition.

Ali Unwin: I think the commoditization of models question is probably, along with the depreciable or the economic life of the GPU, is one of the key questions. My own view is that these companies absolutely believe that these models are not going to be commoditized. Or rather, as you say, the leading foundation models will not be. If you want to use these models for document summarization, that's a commodity already. If you want to use them for image recognition, that's a commodity. If you want to use them to cure cancer, that is not a commodity. And the best model may be able to do that in a way that the second best can't.

If you believe we move to more of an agentic economy where we all have agents acting on our behalf, if my agent has a n IQ of 200 or sort of efficacy value of 200 and yours is 193, mine's going to win all the time. And so, it depends whether you think the race is at the margin or whether you think the race is in the sort of stock. And our view is the race is going to be fought at the margin so long as scaling laws hold, of course, because if scaling laws don't hold and it's going to cost $10 trillion to train the next best AI model or the next Grok5, it just isn't any better than Grok4, then everyone gets to pretty much near the frontier and there's no differentiation. And the sort of limits of what AI can do are found. We just don't think we're there today. And the fact that you've seen Jensen putting up a hundred billion dollars of his own capital to go after this and to make sure OpenAI's got the compute says that they don't think we're anywhere near the end of this either.

They can be wrong, but that is the signal we take from that.

George Viney: Can we dwell on Nvidia a little bit? They play a fascinating role because they are by far the leader currently in that subset of hardware, which is so critical to all of this. They are also highly motivated to diversify their customer base, and you see that with the relationship with Oracle to the disadvantage arguably of the other hyperscalers.

Custom chips come into this conversation as well. And then there's, as you mentioned, this risk of obsolescence and Jensen Huang would talk about himself being a chief revenue destroyer as the new chips come in, destroying the revenue of past generations of chips, which everybody's been buying. How do you see all of that playing into what you've been talking about as it relates to scaling laws and the ROI question?

Ali Unwin: I think the depreciable life question is very, very important. As you say, Nvidia is incentivised to create a much more diversified group of buyers. They don't just want three. They're also incentivized to not help cloud providers who are producing their own chips. So, Nvidia have somewhat unfairly to AMD in my view, always talked about the TPU at Google being their biggest competitor. And I think a lot of the bull case on Google and AI now is that they have their own chips and they're not paying the Jensen tax of the gross margins that Nvidia have. And so, compute is absolutely central to the scaling laws.

The other thing that Jensen wants to do is to build this new parallel computing architecture. And when, two years ago, I think Nvidia said this is going to be a trillion-dollar opportunity over the next few years to rebuild what we have. This CPU based architecture, which does serial compute, which has been great for

the cloud. What's coming now is entirely different and we've likened this to a Toyota Prius versus a McLaren F1 car with the sort of Toyota Prius being cloud computing that is built for everyone. Lowest power per watt kind of good utility scale. Pretty good, but not amazing at anything. Parallel computing is a new dense form of compute, which is really quite different. And here you are not solving for power usage. You are solving for pure performance. You're not even solving for cost, which is unusual.

That means that every single bit of the stack, and in Nvidia's case, it's not even a server. It's a whole rack of servers. Then you have all the cabling that goes together, the networking that goes around it. They're selling you an entire package. They have also committed to an unbelievably aggressive roadmap for what they're going to deliver every 12 months down the line. So, you will want to buy the new generation because it will deliver you so much more value on a total cost of ownership basis on a per flop basis, if you like. The economics of doing so will be compelling even more than building your own chip, even more than using an AMD chip that has a 50 gross margin, not a 75 gross margin.

And the reason for that is that they can sell you the whole thing and it all works perfectly together, and they can make the networking work brilliantly. The opportunity for us is that this new architecture has a in many cases, completely different set of players than the cloud architecture. Cloud is all about cost down. Get 10 PCB suppliers in Taiwan. Get them all to cut each other's throats and destroy the margins. Hyper cyclical, horrible business. Beat up the networking players. That's the whole game. It was all about cost down, power down.

AI is very different. I want best in class substrates. I want best in class PCBs. I want best in class server rails. A lot of these companies are in Asia. High bandwidth memory, for example, like memory's been a difficult, choppy commodity market. And it feels like we're going now into a, we've certainly had a period where you've had more pricing power in high bandwidth memory, and SK Hynix having been the dog of that industry, it has had a period in the sun. So, everything is getting repriced in this new architecture. And that is why we see this both in Nvidia terms, but actually it's not just Nvidia. It's the whole parallel computing architecture that we are moving to being very different. And this being primarily a hardware centric cycle for now. So, when we ask about what are the ROI, what are the applications, what are the use cases, you can't really know.

Like the big markets that come after general purpose technologies come and continue to improve. You don't know. When steel comes along, you can see that the Chicago Home Life Insurance building in 1885 is built and it's a bit taller

than the one next door and a bit cheaper to build. You can't see that buildings are going to get 25 times larger over the next 50 years. You can't see the Manhattan skyline being transformed, the price of land. You can't see the can goods industry arriving. You can't see the auto industry, none of that. It's all downstream of steel. The first market for steel was replacing iron bridges with steel bridges. Iron rails with steel rails.

That's fine. And that's the sort of visible market where we look today about chatbots versus search or something like that. I think that we will look back on that as being like the sort of surfing the web in the mid-nineties or the information superhighway. I think it will feel that the first ChatGPT interface will be looked back as quite quaint, because in an agentic world you're not going to be chatting away with a bot as if just a better version of a chat bot that you had before. I think things will be quite different.

Tom Yeowart: What does this mean for the hyperscalers themselves? Obviously, Oracle came out with some pretty extraordinary projections the other week and if they deliver or get close to delivering on those projections and they build this significant parallel compute infrastructure to service OpenAI, which they're trying to do. What implications does that have on the other hyperscalers and their economics and their ability to win in this new paradigm?

Ali Unwin: I think the fact that you include Oracle in the hyperscalers tells us how far we've come, how quickly, because I think 12 months ago, that would not have been something that you would've said because Oracle was emphatically not a hyperscaler. They were very much the people who had failed really in cloud.

But I think the implication for the other players is potentially quite bad. When AI arrived, there was an assumption that Amazon will be a good conduit for AI because computing happens in the cloud. All the data's there, they've got loads of GPUs as well. Models are going to be a commodity, so they're actually going to be well positioned. Not as clear that's the case now. If you believe this is a wholly new architecture and if you visit a CoreWeave or a Nebius, whatever, you think of the stocks and the business model and the debt load, they will tell you that you are building and designing these architectures with a completely different set of priorities because this is a new architecture and all of the old stuff it was designed for the old stuff.

And it is a Prius and F1 world. That leaves the hyperscaler in a position where they still having to invest in the back book, if you like, of cloud computing. You still need your storage and serial compute and all that stuff. But really all of the

incremental action is happening in parallel compute. And there your position is nothing like as strong. One of the analogies we've used is that when the cloud came along, Intel, which was the dominant player of on-premise computers, both in server and PC. Intel was still, I think, the largest provider of chips to the cloud, but it was no longer the monopoly provider. And you had AMD, you had Nvidia, you later had ARM come along.

And it didn't take long for the market to sniff out that Intel was not going to be a good conduit for the cloud. And I think in 2015 when cloud penetration of compute is about 9%, economically not that significant, and Intel's business is still pretty good, Intel stops outperforming the SOXX. It doesn't actually go down in absolute terms, but it just stops outperforming. So having been a leader, it becomes a bit of a laggard, and then finally it becomes a loser. You go from being the dominant player in the old world, to fighting for everything at the margin, to actually being slightly left behind.

And there is a risk to an AWS and even to a Microsoft, however brilliant they have been at securing the OpenAI relationship, and that now has its own challenges, that they end up the same way. That they were brilliant conduits for cloud computing, and they are just not going to be as good for this new world because you are solving for different things.

George Viney: Where does Google sit in all that? You mentioned they've got the chips. They're in rare company when it comes to that. They have a very strong lead in, for instance, autonomous vehicles. They have the know-how; they have the infrastructure. There's been questions about whether they have the managerial capacity to deliver. And there are signs though that they are beginning to ship product at a faster pace. They seem to be straddling the old world and the new. Can they make it?

Ali Unwin: Microsoft made it having missed mobile and really missed the consumer internet fairly spectacularly. And then Steve Ballmer went, Satya came in, and they did get there in cloud, so it can be done. Oracle having missed cloud and Larry Ellison saying it was a joke has made it to AI at least seemingly from the early indication, so it can be done. I think the challenge for Alphabet is the Google business. There was an interesting note saying that ChatGPT is the best thing that's ever happened to Google because they have finally been forced to ship some product and actually make some changes and take a few risks as opposed to just feathering the nest in the kind of dominant economics of search, best business model ever created, et cetera.

Will they be able to cannibalise, and I think that is the right word at this point, the traditional search business, and particularly the economics of that business being so good, in favour of something that will be more contested. Google's going to be a big player in AI. They've got the distribution, they've got the people, they've got the chips, they've got the data. They've got a huge number of advantages, but they'll be competing with OpenAI and OpenAI so far looks like they're leading, and we know that leaders tend to continue to win in tech or certainly historically have continued to win. Winners have taken most.

And that leaves the base case on Google, the null hypothesis, that Google won't win in the consumer application of AI. If they can get an AI agent in everyone's pocket. And actually, all the data that sits in your Gmail and your Google Docs and other applications. And your trusted relationship with Google's brand and their ability to deliver. If that can get people into an agentic way of transacting on the web say or searching. If the agentic case takes over what you did before, then they can win. That is a difficult thing to navigate. I think the economics of search have held up better than people, including us expected.

Until you realize that OpenAI doesn't have an advertising product at the moment, there's nowhere else for the money to go. And this was a little bit like TV advertising actually kept growing at a sort of GDP rate for really quite a long time. Even when the internet came along, it was all a bit strange. And of course there was no way to do reach. So, TV ads stay at 65 billion and then the ad market goes to 300-500 billion quite quickly and it's just TV ads stays at 65 billion. Does search become that market at which point Google stock is just less interesting because you have a legacy business that isn't growing.

The Waymo stuff is phenomenally exciting. I don't know if you've been in a Waymo car, but it's quite extraordinary. And someone pointed out that we lament the lack of AI impacting the physical world, but you can actually take a driverless car from one side of San Francisco to another. Whereas today you can't actually get an agent to book you a hotel with any great success. Now we are nearly there, but nearly isn't there. And so there is an interesting case if Google can solve some of those problems, then they do have the potential to lead. But I think our default would be that the search business will look very different in three- or four-year's time and they'll be fighting jump balls for everything on the new side of things.

Will they win? Not sure. My honest bet would be probably not. Where that could be very wrong is if this comes down to a battle of cost per token or cost of compute and Google there does seemingly have an advantage there. And if that is what matters in the end, which again, being lowest cost producer hasn't

particularly mattered in tech. Apple is not lowest cost producer. Microsoft is not lowest cost producer. If that was how this played out, then Google could win, and Google's agent would be better because their cost of compute was lower.

George Viney: Is it simply too early to talk about applications beyond coding, some applications in customer service and so on. The companies that are developing product that's spreading through their customer base, whatever industry they may be, and monetizing effectively because the product is so good. We've seen that in coding. We haven't seen it in many other places. What's your perspective on the timeline of adoption? Do we just have to wait, or can you see things developing now, which gives you confidence that the applications will be there?

Ali Unwin: Developers are an unusual group because they like adopting new things. The work of coding happens to be extraordinarily well suited to AI models. And they continue to get better and better. And that has flowed through directly into products like Claude Code, like Cursor, and where you've seen these companies grow to very fast ARR to a degree that we just haven't seen before. And with headcounts, that would've been unfeasible. I think customer service is, as you say, the second area where this has been most obvious. Deflection rates are easy to measure. This is a cost center. You can look at the amount of people you can reduce. This is also a job no one likes. Turnover rates are 50%. You can see why those have been the earlier use cases.

When this turns up in other areas, will be a function of improved models because one of the things we look at is how long a task can you perform. Because an error rate obviously can be very low. But if you are doing it for an hour and there's one every 10 seconds, it compounds up to becoming useless. And the bar, as with driverless cars, is not human performance. It's not even twice as good as human performance. It's probably 10 or even a hundred times what humans can do before the economics become so compelling that either you do it or a new competitor who is running an asset management firm like Polar can do it with a 10th of the headcount that you can and suddenly the clearing price has changed.

I think the place this will show up more rather than revenue is in the cost base. And you've seen that in the hyperscalers already. These are companies that have grown revenue 40, 50% over the past two or three years. And have grown head count five. You are seeing revenue per employee become an interesting metric for companies who get this.

Adoption rates tend to go through s-curves. So, you tend to not see much progress for ages, it's all a bit surprising and then suddenly, oh my goodness it clicks, and the rate of adoption actually accelerates.

George Viney: Does that apply to the consumer side as well when new products come out? Really, it's the new delivery, the new models and what they can do will really drive adoption.

Ali Unwin: OpenAI put up some data on what people are using ChatGPT for and the interesting thing is that the non-work use cases are growing quite a lot quicker than the work use cases. Consumers are finding value in this and if you look at the cohort analysis, the longer people are ChatGPT users, the more they use it.

And as a growth investor, that is just the thing you always want to see because then you can start to extrapolate down the line, what could that look like? And the mix of tasks is quite interesting. So, three quarters of usage roughly is non-work. And it is things like I think they talked about practical guidance, which is what you might have used a YouTube video for, or ask someone or looked up in a book even many millennia ago or gone to an expert for.

So, there's practical guidance, there's writing and editing was another big category. And then there is this sort of information seeking as in what probably would've been a search or more likely a series of searches on Google, is now a tell me where to go on holiday. This is my family. I like these things. And the answers you're getting are already phenomenal.

One of the things that came out around ChatGPT-5 is this looks more like an operating system than just a model with a chat interface sitting in front of it. That is a slightly different thing. And working out how you best might want to use that and the memory that it might have of your preferences. Those things may look more like traditional moats for a consumer facing business. If I am the trusted agentic partner of the consumer who is doing their buying on their behalf, who's advising them, who's telling them how to raise their kids. We're going to see a huge amount of innovation in the use of these models while they are continuing to improve.

I think this is a Dario line. The models you use today are the worst they will ever be. And that is a difficult thing to get your head around because you're trying to imagine a future which is already difficult and different, but with an improvement curve that is extraordinary. So, Moore's Law, which changed the world and just driven the tech market and the tech story since the fifties

improved at about an order of magnitude every decade or 1.5 orders of magnitude every decade. AI models at the moment are doing somewhere between half and an order of magnitude every single year.

It is a different slope and that makes it very disconcerting. It means that there are more questions than answers on the downstream effect and candidly is why we focus our investments on the infrastructure bit. Because we don't honestly know what the applications look like down the line. We couldn't have told you when the mobile phone's invented that Uber's going to come along or Airbnb. You just can't know. But what you can do is observe I think earlier who isn't going to win or who is unlikely to win. And try and get out of those stocks. And the aperture of that is expanding very, very quickly. And then keep a very open mind on the repricing of assets that might fare better in a new world.

We use the phrase an AI lens, which really means thinking first at the very start of when you're looking at the company, where does this company sit in an AI first world, in an agentic world, in an AGI world. What does it look like? Is that bad or good? I think for us, that's the right question to be asking.

Tom Yeowart: Ali, turning to our closing question, if you could go back to the start of your career, what piece of advice would you give yourself?

Ali Unwin: I think the thing I wish I'd done is bought PCT because I actually checked today and it would've been up 2000% from the start of my career, so that would've been the smartest thing I could have done.

I think the piece of advice I would have given myself, would've been to not confuse not knowing how things are going to pan out. Not to confuse that as being the same thing as not knowing what's going on. So, the idea of a healthy scepticism not being the same as a lack of conviction. And that's a difficult thing to understand in a n industry that values conviction and views and love this, hate that, love it at this price. Must love it even more here. As opposed to things are in motion. The best bets today are this, but the stocks don't know you own them. You can change your mind, and you can keep high conviction that tech is going to transform this industry or that this is the way things are go but you don't have to be dogmatically convicted in it in order to understand it.

And I think that's been a tricky thing to realize because it's nice to understand things and it feels better when you have a strong view, and all the data points confirm it and all the other ones are wrong and that kind of thing. I think sitting with ambiguity is a very difficult thing to do as an investor and I'm still learning, I think, how to do that.

Tom Yeowart: Yeah, great answer and thank you for coming on.

Ali Unwin: Thank you very much for having me.