John Thornhill (00:00:08):

Good afternoon. I'm delighted to be participating in this remarkable conference. And in this particular session, I think we only have an hour, as opposed to all the others, which are one and a half hours. So we're gonna have to talk at 1.5 times if you'd like -- you listen to podcasts like that. My name is John Thornhill. I'm innovation editor and tech columnist at the Financial Times. And like everyone else these days at the FT I'm writing a lot about artificial intelligence. This panel is called Unlocking Potential, Harnessing the Technologies for a Prosperous Future. And we have four fantastic speakers on this theme who are all bursting with human intelligence. All their titles, names are above us on the screen, and you can read more about them in the programme. I'd like to frame this discussion a bit by posing three kind of overarching questions for consideration.

(00:00:59):

One: everyone is getting wildly excited about AI, in particular generative AI. But how transformational is it really likely to be? A few years ago, the great economic historian Robert Gordon, dropped by the Financial Times. I'm sure many of you have read his wonderful book, the Rise and Fall of American Growth, which chimes a lot with what Tyler Cowen's original thesis about the great stagnation. And he looked at my business card and he said, ah, innovation editor, let me know if you find any. So I'd like to know, can I now go back to Robert Gordon and say, AI is an innovation comparable to electricity or flight or penicillin? How transformational is it gonna be? Or is AI just the world's worst misnomer? And we would get a lot less excited about it all if we just called it computational statistics. Second, the social and economic impact of technology.

(00:01:58):

Every new technology has created something of a moral panic. For those who haven't read it, I'd recommend the Pessimists Archive, which has a terrific collection of moral panics about the terrible things that bicycles and cars and electronic calculators are gonna bring to us. But I'm very struck by the speed and the scale of the moral panic about AI. Or is it just genuine concern? I mean, this morning when I woke up, I was listening to the radio and was listening to the people talking about the Hollywood writers and actors strike. And Fran Drescher, who is the president of the Actors Union, was saying, it's a terrible thing when big business tries to replace everyone with digital and robots and AI. And she said, if we don't stop this maniacal need to make money over allowing people to make a living, it's going to be a dystopia.

(00:02:48):

On the other side, there is a lot of utopian talk, some talk about AI leading to an economic singularity, a world of super abundance or fully automated luxury communism. Is this a revolution? If this is a revolution, then we might ask the old Leninist question: who, whom, who is going to win? And the expense of whom? The third issue is picking up on some of the discussion that we heard last night. How do policy makers react to all of this? What do they have to do to maximise the positives and minimise the negatives? As Sam Bowman said, the natural instinct of Europeans is to regulate the hell out of all this. Is that really necessary? We don't want to stifle innovation, but we do want to stop bad innovation. Is a risk-based approach a good idea? Or should we just update existing regulation? Or do we have to come up with wholly new regulatory agencies because of the novelty of this technology? So, to answer all of these questions and many more we're gonna start with Matt. Tell us about how you see AI, how transformational it is. Let's knock those three off.

Matt Clifford (00:04:00):

Great. And we've got got three minutes for that. You said how transformational is AI, you know, is it the same order as penicillin and electricity? I think the answer is clearly no. It's much more important than that. You know, perhaps the most important thing our species will ever do is create intelligences that are equal to or greater than our own. And I think what's really interesting about the current moment is that people have been saying that for a long time. You know, there's obviously a long history of artificial intelligence, but I think we are sort of talking within the first year in, you know, human history where there is actually a plausible path to that transformational moment that doesn't require you to sort of hope that we come up with some sort of big bridging idea.

(00:05:04):

And, just very briefly on that, and then I'll let Mark pick up the hard questions. I think the bit that is probably still underrated, which I realise there's now a lot of talk about AI out there, but I still think it is probably massively underrated, is that the current paradigm in AI, around these very, very large models, has some very weird properties that no one really expected quite a short time ago. And that is the models that we are currently training, the largest models that are being used, things like GPT4 and, Google's equivalents and Anthropic where Logan is equivalent. The kind of really bizarre and amazing thing about these models is that so far the performance of these models, the capability of these models, it predictably improves as you apply more computational power to them, and you give them more data.

(00:06:07):

And so far we've not seen any sign that that flattens off. And I think the reason I say it's underrated is I think people look at the state of the art that they're able to play with today and say, wow, that's really amazing. And they don't realise just how little resource was used to train those models compared to what's coming very, very quickly. So the best rumoured estimates of how much computational power is used to train GPT 4 which, by the way, unless you're paying for chat GPT, you're still probably thinking about GPT 3.5, which is much less capable. But GPT 4 was trading with about $60 million. It's pretty plausible that Google will spend a hundred times that next year on compute. And in a world where, if you believe me that there's no flattening off of these curves of capabilities improve as data and, and compute increases.

(00:06:58):

I think we're probably just, most of us, even those of us who spend most of our time sort of thinking about this underwriting the extent to which capabilities might improve over the next five years. Half of the money that's ever been invested in private companies building AI has been invested since chat GPT was released. And so I think we're just on this extraordinary exponential curve. I'm not blind at all. I sort of deliberately, my first sentence was provocative, but not blind at all to some of the challenges we might hit, some of the walls and pitfalls we might come across and some of the social and regulatory challenges. But I think as a starting point, thinking of this where we are today, not as a moment, you know, that's caused by chat GPT, but as an exponential curve. And this is like February 2020 in Covid, I think is the right way to think about this.

John Thornhill (00:07:49):

Great. Logan, could I come to you? As Matt was saying, you are working at Anthropic, which is developing large language models. How does the world look to you?

Logan Graham (00:07:58):

Let's see, you mentioned at the beginning: are these things equivalent to sort of electricity, at least in the same class of how we should think about them? I would propose something yes, and a little bit more. So we usually think about electricity, etc. as general purpose technologies. It seems very clear that AI is a general purpose technology. But it's distinguishing feature is AI has as part of it all the components that we use to develop general purpose technologies. And so, you know, the implication there is intelligence then allows you to improve things. And if intelligence is the thing that you're trying to improve, what happens if it keeps going upwards and gets applied to itself. So in some sense, general purpose technologies already do this. Like our ability to produce electricity has gone up significantly, but only going through the sort of unsuspecting host of a human to then think about how to make electricity better to then go back and make better machines to make better electricity.

(00:08:59):

If you remove the intelligence layer in between and you have a technology that can sort of recursively self-improve, then you're in just entirely different, entirely different world. The utopian case for this is the most sort of significant degree of sort of up-winger progress that we've ever seen. The really utopian cases accelerate and automate all science forever for all humanity, become space faring, post-humans living lives of unbelievable abundance. And the dystopian case says, well, there's quite a lot that we apply intelligence to that gets used to destroy everything. And so where I work and why I do what I do is I really like the good stuff to happen, but if we had, you know, a hundred dollars to allocate, we're probably underspending our chips on making sure that it doesn't all stop and the party is over and we all go home.

(00:09:51):

So let me be very maybe concrete to just set the stage for risks that are useful to talk about. There are large areas of science that we use, that we've proveably used to develop tools to give us an arsenal to destroy the world many times over - things like nuclear weapons. There are other types of weapons, for example, chemical, biological weapons, that are extremely sophisticated that we have barely started to master. If you can imagine a world where you have autonomous state level hacking programmes, and there's not a single human behind them, then can you trust anything at all that's ever connected to the internet? What happens if this stuff becomes integrated into everything we ever use? There are some paths out of this. The general sort of framing people use is the sort of path to alignment or safety.

(00:10:44):

And there are two paths there. There's the technical path: can we make these systems provably steerable? So they do the things we always want, and not at the expense of killing all of us. The second path is the coordination path of even if you figured out how to make machines aligned with humans, you need every human who's ever building any of those machines to make sure that the machine they're building is aligned. And that's largely I think a social and political one. So what do you do about that? Maybe we can dive into it, but I would propose maybe the single best starting point here is we are at this place where we're all kind of wowed by these things, and quite literally, you or I or nobody else, despite the amount of talking head hours that are spent on GPT and AI, we have no idea what these models are capable of. Genuinely. There's no real evaluations that are seriously done to figure out the extreme good and the extreme bad that these models are capable of. We should probably be running these models through as many tests as possible to see what they can do, and in particular for the areas that we're most concerned about.

John Thornhill (00:11:57):

Okay. Lots to come back to on that. Mark you're also working at an AI company faculty, which is also at the kind of leading edge of a lot of this technology. And you've been working a lot with the kind of public services in particular over the past few years. Can you tell us what is your take on all this?

Marc Warner (00:12:14):

Well, I agree with a lot of what Logan and Matt said, but I've been asked to talk about sort of some of the more, slightly more prosaic stuff. So, how do we actually use this technology to make things better? So maybe a particular example that Faculty's most famous for is the work we did during covid to help the NHS predict how many patients, Covid patients they were gonna see across the country. This enabled the NHS to make a bunch of better decisions about where to send medicines, ventilators, patients. I think there are enormous opportunities to find safe, connected AI systems across public services and to make much more intelligent decisions around how we use these things in government. A couple of or three, potentially like slightly more visionary options in education. I think it's very clear that we're gonna be able to create a personal tutor for every student everywhere in the world.

(00:13:10):

In health, I think we're gonna be able to have an AI advisor for every doctor helping to suggest new ideas and checking that they haven't missed anything. And finally, in culture, I think we'll see an unleashing of creative forces as we get movie level special effects in the hands of YouTube creators. And I think that promises, at least for the first time in a decade, something other than Spider-Man or Fast and the Furious in the cinemas. So I think locally in the short term, we have very good opportunities to make use of the kind of current technology to make these safe, connected, human first systems. Over the next few years, there's gonna be some very, very powerful new applications that are just going to fundamentally transform how we do some fairly basic things in society and the quality that we can do those things with. And then, you know, open question where we go into, into the kind of world that Logan's talking about.

John Thornhill (00:14:06):

Okay, great. Saffron, you work for something called the Collective Intelligence Project, which is a very intriguing name. Can you tell us a bit about that and how do you think we can harness AI for good?

Saffron Huang (00:14:18):

Yeah, so I started the Collective Intelligence project about a year ago to work on how we can improve our collective decision making capacities over the frontier of technology. So thinking a lot about AI. I used to work at DeepMind doing research and have kind of gone between AI governance and AI research in the past few years, and I think there's no guarantee that AI is going to be deployed in the collective benefit. There's so much I can say, but you know, AI is lots of different things and our large language models or generative AI is one very specific manifestation of them. And not everyone believes that that is the specific paradigm that is going to get us anywhere near the sort of theoretical, like high, high levels of intelligence that my co-panelists are talking about.

(00:15:14):

But I think in terms of, you know, how do we actually invest in the applications of AI that improve science or improve manufacturing? How do we, instead of investing in things like the most visible manifestation of AI that we see everywhere today, which is, you know, the Facebook or the Twitter newsfeed. I mean, that's the real AI that we see, and that's the one that makes the money. And so I think what is interesting is, how can we, because labour is taxed at a higher rate than capital and many other distortions of innovation, there's a lot of people that I'm seeing who are really excited about automating people with AI, about doing things more efficiently, about kind of using LLMs to generate the kind of things that humans would've done anyway, which is great and fine.

(00:16:06):

And, you know, efficiency is good, but I'm really interested in how can we direct AI towards the applications that do qualitatively new things? And I don't think we can take that for granted. I mean, Inflexion just raised a billion dollars to build a personalised AI assistant, which is cool, but it's not like mining uranium or something. So I think just thinking about theoretically this really sexy idea of what intelligence could get us and how it could tell us to do everything, but actually, you know, the language models that that are being trained on the internet may even bottom out because they start training on their own generated data and then they can't get any better. I don't know. I think as someone who's worked in research, a lot of researchers tend to be more pessimistic about AI because they're seeing the struggles up close.

(00:16:59):

So I could be for sure biased on this, but I do think the kinds of intelligence we're developing right now are a very specific kind. So I think I derailed a little bit, but essentially at CIP we're thinking about how do we ensure that technology is developed in the collective benefit? There's one part of this, which is how do we even know what's in the collective benefit? What do people want? How do we have more clever ways of -- I I think, you know, the democracy that we have today is not very good at actually aggregating people's preferences in high fidelity and useful and up to date ways. Are there ways we can improve that? And specifically for things like AI alignment, you know, if we want to align AI to humanity's values, what are those values?

(00:17:44):

How do we get them, you know, who is being asked them? Like we are trying to set up real infrastructure for doing that, both for alignment and for sort of like other questions around technology. And then the second pillar, so there's sort of elicitation of information and values and preferences. And then the second pillar is, you know, what are the incentives structures and the containers within which we develop technology? And I think, you know, even if we have the the safety scientists and the researchers who have created these amazing methods for aligning AI, how do we know that they're going to be implemented? And I think that's the question of incentives. And the questions also of how do the maximum number of people benefit from this are really unsolved. So OpenAI, for example, has clauses in their charter where they say, okay, if we generate a shitload of money from this, if we have infinite profits after a certain amount, we're going to redistribute the money, it's called the windfall clause.

(00:18:49):

And that's never been tested. We're not sure if it would be really interesting if that happened, but I think if we got to that world where they had that much money, the world would look very different basically. And it's not clear that that would actually be implemented. So I think, you know, thinking about are there other mechanisms that we can have for sort of governing transformative technologies that can be implemented in, you know, bylaws and charters and structures that can sustainably guide technology in the direction of people's values. And I think that I'm probably not as optimistic as you can tell, I'm probably not as optimistic about AI as other folks, but I think that's because AI is a reasonably general purpose technology, but it's also purely digital. And there's a really great piece that went out recently on, you know, why transformative AI might be really, really hard to achieve to your first point.

(00:20:00):

And one of the, one of the points that was made in this piece is by Arjun Ramani and Zhengdong Wang. And one of the points is that if the economy is incredibly unbalanced, so if you have amazing progress in AI, but very little progress in other sectors, the other sectors are going to be the bottleneck. So I think thinking about, again, if we're just deploying AI in recommender systems and, you know, news feeds and things like that, rather than in things that look more like AlphaFold, that's not going to achieve the sort of transformative economic growth that we're looking for. And I have more things, but I'm gonna stop there.

John Thornhill (00:20:46):

Ok, let's pick up on that point you're making. I mean, the theme of this conference is the great stagnation and what we can do about it. Last year I reviewed a book by Nouriel Rubini, the incredibly gloomy economist who wrote a book called Megathreats, the 10 Trends that Imperil Our Future and How to Survive them. And when I read the book, there are 313 pages on the threats and only seven pages on how to survive them. And in those seven pages, basically only two answers. One was kind of fusion energy, which he thought was gonna come along. And the second was AI, which is gonna Hail Mary pass for humanity. And his point was that if AI can help us get to 5% global growth, pretty much all of the other problems - sustainable growth, he adds - pretty much all of those other problems then seep away. So what do we think about AI as a productivity tool? How much is it gonna stimulate the economy and solve a lot of the problems that we've been talking about today?

Marc Warner (00:21:46):

We have sort of tested it with some of our software developers, so I know it's a very narrow circumstance, but a decent, probably two thirds of our software developers think it gives them somewhere between a 20 and 50% increase in their productivity in writing code. And about the same number, I think it was about 70% think it just makes their code writing experience more fun. So for at least some part of the economy, we are already, you know, very important part of the economy, we're already seeing productivity benefits. And in case you probably haven't met them, but faculty software engineers are pretty cynical. These are not people who would say it's useful if it wasn't. So I do think that it's gonna be a fairly big deal.

John Thornhill (00:22:34):

Matt, what do you think on that?

Matt Clifford (00:22:36):

Yeah, I mean, I think Saffron's right to raise some of the physical bottlenecks that will hit, you know. I don't think you're gonna. For all that I believe that the next generation models are gonna be very powerful, I don't think you're gonna get to like 50% GDP growth in a year for exactly the reasons that that Saffron raises. But I do think that, I think just at a very basic level, even more basic than helping you write code, it's amazing how many of the jobs that probably people in this room do are effectively summarising information, compressing or decompressing information. And I think, like even in the most pessimistic case, even if we never got any improvements, scientific improvements in AI from today, simply the productization of what we have today as like copilots (is I think the metaphor that most people like to use) suggests to me that you ought to be able to get a lot more particularly out of people whose time the market values most highly now.

(00:23:35):

And so I think I find it hard to believe that that's gonna be the ceiling, but I think that it's a really good thing that we are productizing what's out there. I think that's a good thing, both from a safety perspective because I think it helps us learn a lot about - to Logan's point - it helps us learn. We don't know enough about the capabilities of current models. I think we should see this as effectively an empirical problem where we should learn. But you know, I think it's almost certainly gonna be a huge productivity boost simply to have, you know, large language models in your email. How much time do you spend on email? It's an extraordinary thing. Even if that's the limit of it and it feels very unlikely that's the limit.

John Thornhill (00:24:20):

Logan, I want to pick up on something you say, which is that we have no idea what these models are capable of, which is a slightly unnerving position to be in, isn't it? And I mean, clearly there's a whole bunch of AI specialists who wrote that letter calling for a six month pause. I do feel incredibly sorry for the policy makers who are trying to make sense of all this at the moment. On the one hand, they're told that this is a fantastic productivity tool. On the other hand that we don't actually know what this technology is capable of. And there's some extraordinarily scary scenarios out there about what it could lead to. So help us, what should policy makers be thinking about this, how should they prioritise?

Logan Graham (00:25:00):

I'll provide a one sentence answer to your previous question, and then I'll say that, which I think is helpful if you take electricity, the steam engine, and I believe the calculation involved. So the general trend of robotics and automation in the post 1950s, and you took the sort of median estimate of their per capita, or their addition to per capita growth, and you said that AI's at least gonna be the median estimate, then that's somewhere around 1% extra GDP growth per year. And in the UK that would be by 2030, something like 200 billion pounds extra for free every year, which seems pretty nice. What is that, £3000? We'll take it. Yeah. 3000 pounds per person. But okay, so what do we do about these? I think an illustrative example is the EU AI Act, which for many years, you know, is this, this isn't being live broadcasted, is it? Is it going to be? Okay, great. I'll be a little spicy, I think. And this is a Logan perspective and not an Anthropic perspective. It's being recorded. That's fine. Okay.

Matt Clifford (00:26:10):

Do we dial it down slightly?

Logan Graham (00:26:14):

I'm not very convinced things like the EU AI Act came from a genuine motivation of saying, wow, I am extremely scared, and there's lots that we need to do. I think it more likely came from the motivation of AI is interesting, I'm an MEP or something, I'm going to get my name on this and do something. A lot of heads in the room nodding.

Matt Clifford (00:26:34):

That's the spicy take?

Logan Graham (00:26:36):

Either to confirm or to say I understand government. That's definitely what happened. I'm convinced that this is what happened and what did they have to do two and a half years on to that process? They had to basically scramble and rewrite it. Everything that they had done up until then suddenly became, you know, totally non-applicable because they realised the real thing was maybe this thing that's happening right now. So I'd say position number one is maybe we should have a bit of epistemic humility and make sure that we are reacting to the right thing at the right time. From my perspective, this does seem like the right thing at the right time. However, I don't think we know enough to figure out what the right reaction is. And you know, this will make other government people's heads in the audience nod, which is to say, when you're in government every day, you feel like, okay, I have these tools in front of me and it's broadly having meetings, writing regulations, and doing some other things.

(00:27:31):

My daily activity is to do something with one of those. Almost never is it, I'm going to ask myself the question of what do I need to know that I currently do not know that would make a certain regulatory or whatever intervention dependent on it. And I think that's the position we should be in. What government probably should do immediately, which I think you can do within three to six months, is develop a series of evaluations, whether they are, I think national security risks is probably like a really good first place to start. There's lots of other areas as well. And just let's get on the sort of playing field of what these models are capable of. Here's the second thing that I would do. I don't know if government is going to do this, but I would urge industry to do this, is try to build sort of, you know, lenses into the future before it happens.

(00:28:21):

There are some technical development paths of models today that I think will change everything. Like you were probably used to going on to a chat interface and having a conversation. I strongly suspect that that is going to be dwarfed by the next kind of interface of interaction. I think one of the clearest ones is the ability to have your own agent running around in the background, not waiting for your response to probe it to do something. You will say, give me a review of the current state of the literature in this area, and that agent will read 20,000 scientific papers, maybe browse the internet for a little bit, download some scientific simulation software, run hypotheses of its own and come back to you. And that will take potentially three or four orders of magnitude more computation than just sit spitting out a 500 word response. We need to figure out what the world looks like under that paradigm. And I think, you know, governments and industries should probably be compelled to go figure out as fast as possible, as safe as possible, how to do that.

John Thornhill (00:29:18):

There's some great brains in the room on this subject, so I want to bring you in very quickly, but I'm just gonna ask Saffron for a final question from me, which is we've already heard a great kinda spectrum of concerns about AI. And I mean, in a way it goes from the very here and now concerns about kind of algorithmic bias in hiring or judicial sentencing kind of issues, right through to the whole. We're gonna have a rogue super intelligence that's gonna turn us into paperclips. Where, what should policy makers be focusing on? Everyone is talking about regulating AI. We know what we don't like, as Logan has made clear, but what should we like?

Saffron Huang (00:29:59):

So I think I would largely agree with Logan, there's a missing piece of what's actually going on that we don't actually know in terms of how generative AI is being used and trying to forecast the future and forecast the effects. And so I think the thing that policy makers should be doing is really putting money into developing evaluations and metrics and monitoring and auditing and benchmarks and all of this sort of critical safety infrastructure for just being able to understand what's going on so that you can then do something about it. I think in terms of the kind of extreme risk, or the sort of longer term risks to more like near term risks, I mean, the future is unfalsifiable. We don't really know what's going to happen. I definitely think that extreme risks are possible, but it's not clear. Nobody knows the exact sort of percentage or amount that you should be putting into it. So I think it's like, put money into it, but also put money into all the other things. I think all these things are connected anyway. You know, if we are able to monitor and understand narrower term things and harms that are actually happening, then we probably have a good headway on the infrastructure that is needed to understand and monitor and do something about longer term, more extreme harms.

John Thornhill (00:31:30):

Okay. Great.

Marc Warner (00:31:31):

Oh, John, can I just come in on that? So just a distinction. I agree with everything Logan said, but it's important to note that there are a totally different category of algorithms, the sort of narrow algorithms that are being used most of the time in the circumstances that I'm talking about that have none of the properties that Logan's talking about, we really do understand them very well. Their behaviour is completely bounded. I mean, there are obviously risks of misuse in the sense that somebody can put them to harmful use, but they should be thought of as tools like pens rather than these kind of more general systems that are extremely hard to understand and have all the characteristics Logan was talking about. So I think it's very important to make this distinction very clearly between these narrow bounded systems and these more general, much less bounded systems -- large language models being a good example.

(00:32:23):

And basically it's crucial that we go fast on the narrow stuff that we really do control and we really do understand well. And I would advocate for being careful on the general stuff. Now that exact bounding line is slightly hard, probably you do it crudely to begin with by some kind of compute metric. Obviously pretty quickly you have to make that much subtler. But in the context of the kind of great stagnation that we are talking about, to today, I don't think we can afford to sort of let improving breast cancer screening in the NHS fall by the wayside because the very real concerns that Logan's talking about. And so the way I try to say this to people is, you know, it's okay to care about safety on two different timescales and think about them slightly differently. So it's okay to care about car seat belts at the same time as caring about climate change, people often try and frame those as kind of mutually exclusive. You can only care about one or you can only care about the other. It just doesn't make sense. It's not how we do anything in the rest of our lives. It's fine to care about both, but we can make much, much stronger statements about the narrow systems. And so we should go faster there

John Thornhill (00:33:39):

In that sense, isn't the EU trying to do a good thing in categorising risk? And it's saying for certain categories of applications and narrow AI -- get on with it, it's kind of permissionless innovation in that sense. But in these specific areas which we think are high risk, which are gonna affect people's lives, we apply a far higher standard.

Marc Warner (00:34:02):

I'm not an expert in the law, so I don't wanna speak too strongly about it, but my understanding is that even in the narrow domains, they are regulating AI as a relatively cross-cutting technology. I think that's a mistake. It's like trying to regulate steel, you know, it matters greatly whether somebody's gonna make a gun or like, you know, a girder. If you try and regulate at the level of steel, you either pick kind of very lax regulation that kind of suits girder manufacturers and you get loads of girders, you also get loads of guns, or you pick very harsh like regulation that restricts your guns, but then you restrict your girders. And so just regulating at the wrong level of abstraction is a mistake. It's only in the context of the application that you can even assess the trade-offs of the harm and the benefit that regulation is designed to mitigate.

(00:34:57):

So I don't want somebody who's making an algorithm that picks a red jumper or a blue jumper at the top of a website to have anywhere near the same checks as somebody who has to like make a breast cancer screening algorithm in the NHS. That would be ludicrous. But if you regulate AI as a kind of cross-cutting technology in the narrow domain, you end up in those kind of slightly weird situations. In the general domain, everything's different because in the general domain, it's actually quite unclear if somebody just makes something and then puts it out in the world - like if you just open source a dangerous model, you know, one that can let you write bioweapons recipes or whatever - where the liability lies. So I think in this kind of like general domain of algorithms, I think it's much less clear how you regulate, but those two have to just be kept separate.

Logan Graham (00:35:52):

One quick point: it would be a bad world if we waited for governments to do the right thing always at the right time. There is— sometimes the amount of thrashing about is actually the best indicator that we're doing the right thing and just having a lot of conversations and, you know, rewriting regulation, anything.

John Thornhill (00:36:10):

Thrashing about is a good idea. Right, Timandra.

Audience 1 (00:36:17):

Hi, Timandra Harkness, kind of writer and stuff. I'm really glad to hear Mark distinguishing between types of AI, 'cause I think especially at the moment with all the large language model excitement, there is a bit of a tendency to talk about it as if it's all the same and it's either the singularity or it's just a glorified text predictor. Both of which are slightly wrong probably. But I'd like, in the context of stagnation, maybe to say we're talking about two slightly different things in terms of economic effects, because obviously in a very innovative generative AI especially, but other kinds of AI that's also very cutting edge, we're talking about things that have great potential for the future and possibly for generating products and, you know, maybe digital assistance and all sorts of things we don't know about.

(00:37:06):

But if we're talking about improving productivity, then I personally don't think that having LLMs in my emails is gonna save me any time before emails. What did we spend all those hours in the day on? It wasn't exchanging random messages with everybody we've ever met, and half the people we haven't met. It wasn't, we did other things. Because we've got email, we spend lot of time doing email. If we get LLMs in our email, we'll spend the same amount of time doing email. It'll just be looking over drafts, I dunno. But I think that is all great, it's all long-term blue sky. I'm really excited to see where it'll lead. But I think there is a lot of stuff that AI is doing and could be doing for us now that really could and does improve productivity. And we're not applying it enough.

(00:37:51):

I mean, we were discussing in one of the breakouts that we're kind of the least automated mechanised economy in Europe probably, and we could be doing more of that already. And we were also this morning talking about the different areas of the economy, which add the most value per head. And they were industries, they were things like mining, manufacturing, and those things have been using AI for a long time. Certainly we could be using that more to make things safe and more efficient and I think it would be great to see a bit more attention on applying the stuff that is near-at-hand and already here to those very productive parts of the economy. And by all means, get on and do creative mad things with the other AI and let us waste our increasing amounts of free time and money on those, and I welcome that.

John Thornhill (00:38:44):

Okay. Anyone want to respond to that?

Marc Warner (00:38:47):

Faculty is very willing to sell you any of that that you choose to purchase. I mean, that is what we do for a living in some of the domains you're talking about. So we like to think that we hover this kind of cutting edge to real world problems. So we do our own, uh, research at the scientific frontier, but we are deploying that into real world circumstances in both the public and private sector. So, you know, very, very happy to talk to anyone who needs that.

Matt Clifford (00:39:13):

I mean, if it turns out that the great stagnation is caused by people sending random emails, my first advice would be to stop doing it. And so I agree that, you know, that's probably the the first thing I would deal with before I even try and automate that.

John Thornhill (00:39:29):

And for planning.

Marc Warner (00:39:30):

Yeah.

John Thornhill (00:39:31):

Matt, can I just follow up on that? I mean, you are wearing several hats these days, but one of them is you are the chair of ARIA, which is kind of Britain's moonshot agency, Britain's equivalent to DARPA. So, your agency has the ability to write very big checks. Tell us how are you gonna spend them?

Matt Clifford (00:39:49):

Well, I mean, fortunately I am not, and that's a big part of the model of ARIA. You heard a lot about it from James earlier. The ARIA model is based on the idea that what we really need to is unleash talent. It's designed in a very bottom up way; the ARIA model is to hire programme directors who are themselves, uh, world class scientists who have vision, a vision of some capability that can be kind of dragged forward into the present by in turn unleashing other scientists' talent. And, to hark a bit back to the last panel, I suppose one of the core beliefs that underpins ARIA is that right now we don't let our best scientists work on their best ideas.

(00:40:39):

Like, the incentives are simply not there for that to happen, certainly early in people's careers, and actually all the incentives pull towards incremental innovation, incremental science that's most likely to generate another publication, another citation, publications and citations are the real currencies of science today, which I think is a devastating drain on scientific productivity. And so what will work at Aria really will come out of the imagination and the skill and the vision of those programme directors. And their job is to say, what is the coalition of researchers across academia industry, non-profits, that we can pull together to make these capabilities happen? So we've just hired them - I can't remember if I'm allowed to say this - we have some number of programme directors coming in September across a really diverse range of scientific fields, including AI, and we expect to start distributing the money kind of late this year, early next year.

John Thornhill (00:41:44):

Lady at the back there please.

Audience 2 (00:41:51):

I've got a question about LLMs specifically. So I work for a small EdTech company and our coders certainly use copilot and find it very, very useful, so I absolutely see the benefits of that and reducing the time to code. But what we do is we specialise in assessing children and adults' writing assessments. And we felt this was a really sensible place to use large language models because writing involves language, it's something that humans are actually quite bad at, and it's very time intensive. So we thought this would be a brilliant kind of use case. We trialled GPT-3 with it and it wasn't very good. We trialled GPT-3.5 and just this week GPT-4, and they were actually worse. So the performance changed, but it got worse.

(00:42:37):

And I think this is one of the issues with AGI which is that, you know, you can just say: 'writing assessments is a pretty trivial thing, it's doing all these other brilliant things.' But if you have a general intelligence model, there's no guarantee it'll be good at the narrow specific thing you want to do. And we've had a lot of success using narrow AI models and we feel like we can trust them more, we can design them to do exactly the thing we want to do and they do it. Whereas, you know, the AI models just feel very, very volatile and there's no guarantee they'll do what you want them to do. We've worked out four criteria that are really important to us, and I don't mean just us, but other people we know working in this field.

(00:43:13):

And it's: we want something that can work at reasonable scale; that's reliable; that needs limited human oversight; and can work in a relatively high stakes situation. And at the minute for us, we haven't found an LLM that does that. So I guess my question is: on those four criteria, what are the current sectors that need those kind of four criteria - I still think there's others - where you think LLMs are working? What are the use cases for LLMs with those four criteria where they're working at the moment and, and not, like, 'oh, but it's gonna happen in a year or two' or 'wait for GPT-5', because we feel like we've gone through three models and it's got worse, so you can't guarantee the next one will be better. So that's my question.

Matt Clifford (00:43:57):

Mark will sell you one at a reasonable price.

Marc Warner (00:43:59):

Very, very reasonable price.

John Thornhill (00:44:02):

Logan, do you wanna have a go at answering that?

Logan Graham (00:44:04):

Yeah, sure. The first thing to say is: the way you have to interact with current models is different than the way you used to interact with say, GPT-3. It's like a kind of different paradigm. Literally the way you prompt these models has to change. There've been a handful of times where I've worked with people one-on-one and in 10 minutes they've gone from, this is a total regression, I'm just, I'm not gonna use it anymore to, wow, this solved my problem entirely. So it could actually come down to implementation details. We're at, like, the early stage of using this thing that we have sort of barely adapted to. Like, can you imagine what was it actually like the first time that search engines came about? The first paradigm that we used for search engine, it was actually just big lists of pages with a bunch of links. And then eventually we realised, well, you should type in some text and that text should give you back some results. And then people would type in text badly. And now we've just totally changed the way we interact and query Google and it's like a new interface or paradigm of how we use this intelligence tool that—

Audience 2 (00:45:05):

Do you have a specific example of those four criteria?

Logan Graham (00:45:08):

Remind me of the four.

Audience 2 (00:45:10):

Reliable; relatively high stakes; reasonable scale - so I'm talking like [inaudible]; and limited human oversight.

Logan Graham (00:45:21):

Yeah. There was a person I was helping who was a trained scientist who had built and exited a multi-hundred million dollar company. What he wanted to do was build a system that would read several thousand academic papers, extract all the implementation details of a machine learning model or a computational model and algorithm, and then automatically write the code for every single one and then surface all the review insights. So if you're building a company off of that, you're gonna use this code. It's pretty high stakes and I think it fits all those criteria. And in 10 minutes, we went from 'this model is producing garbage' to 'this model is regularly and reliably re-implementing a thousand papers worth of code in something like, you know, a dozen minutes worth of API calls.

John Thornhill (00:46:08):

Okay. Keen to get some more questions in. So gentlemen down here then a lady.

Audience 3 (00:46:13):

Conrad Griffin, Morgan Stanley. So I'm sorry that this is not specific necessarily to AI, but I noticed that, Mark, when your company becomes a unicorn, you'll be going to the NASDAQ to list. I'm sure, Matt, many of the entrepreneurs that you interact with will ultimately move to the United States. DeepMind was a British company now owned by an American company. So bringing back the conversation to how Britain will benefit as to opposed to the world more broadly, from AI applications: your thoughts on reinvigorating UK equity markets for startups in particular, also for this new wave of AI companies coming around?

John Thornhill (00:46:52):

Saffron, do you want to start on that then, Matt?

Saffron Huang (00:46:56):

I think that's a very good point. I think that, um, you know, I've been impressed at how quickly the UK government has jumped on kind of putting money into AI via things like ARIA, things like the Foundation Model Task Force, I think that's a great start. I am not an expert on innovation policy in the UK so I'll pass it on to other folks who have thought much more about startups and how to encourage them.

John Thornhill (00:47:32):

Okay. Matt, I mean, you run Entrepreneur First. You're helping to develop a lot of these companies.

Matt Clifford (00:47:37):

Yeah, I'll be honest, I've never really been able to get anyone to explain to me why we should care where British companies list. It seems like that matters a lot if you work in an investment bank, but I can't see why it matters at all if you're an entrepreneur or a shareholder or an employee. And, um, I think one of the best things the UK could do is shrink the size of its financial services industry and have more of those people become founders. So, I kind of hope we do nothing and it shrinks significantly as a result. Become a founder! It would be great to get more people out of these zero sum games into like doing something really productive

John Thornhill (00:48:15):

The lady there.

Audience 4 (00:48:16):

John, you mentioned the screen actors guild strike, wanting contractual reassurance that their jobs aren't gonna be stolen and they're going out because the writers - the TV writers and and movie writers - have done the same. Directors are probably gonna go out and the last time that group went out was in the late nineties. It actually prompted the birth of reality TV. It is possible that this set of strikes prompt a speeding up of AI content. That's not my question though, although interested in what people think. Excellent example of union counterproductive action. But my question is this: Netflix have invested 6 billion in the UK over the last four years, they've invested in studio space, they've invested in creative talent in the UK. They've made a really big punt, biggest investment outside the us. I'm genuinely interested in the panel's sort of knowledge or assessment of: do you think that is a company that's just misunderstood - so as Logan was implying this earlier - just how quickly and how big the change is going to be? It's possible - they're pretty big now Netflix - maybe they're suffering from the same complacency and sclerosis as large public sector organisations. Or do you think it's a Tyler Cowen optimism about UK soft power and our storytellers and our creative talent and it's worth it? Or do you think that, you know - and my hope is this - my hope is that they've made a really strategic punt - in the knowledge of everything that's been said on the panel - that AI and human creative talent will continue symbiotically and the UK is a place to do that. Because that's a pretty good outcome.

John Thornhill (00:49:50):

Logan.

Logan Graham (00:49:53):

Sorry, so the precise question is - you're using Netflix as an example - is that an indicator that they have seen that there's something significant and specifically the UK is a good place to do it? Is that right?

John Thornhill (00:50:02):

Or are they wrong and they're not understanding, and they've put 6 billion into studios. And human creative talent in the UK—

Logan Graham (00:50:15):

Got it. And the counter example is: 'well, why would you do that if you could just automate it and that kind of thing'? The first thing I'll say is, that under the flag bearing leadership of Munira from 2019 to 2022, the UK became and transformed into an extremely desirable place for things to happen generally. So you're gonna see like a good base rate of activity and investment. And for what it's worth, the really high value, extremely creative, deep intellectual product industries, the UK is like firmly at or near the top of the list for everybody around the world for very specific reasons, like a long history of creating those people and wanting to invest there. I think it would be hard to claim that for Netflix to have gone from a quirky company that your friends used to order DVDs from to now the most influential content producer via strictly a digital medium, would be an indicator that digital-plus mediums are not an indicator of the future.

(00:51:23):

I think that would be a hard thing to make. But I I strongly suspect that, give it a couple years, and maybe the 5 billion of that will go into GPUs and a billion will go into the humans that they're simulating.

Matt Clifford (00:51:34):

Can I add just one thing on that? I may be stealing this argument from Tyler - I dunno if you're still here - I thought worth saying that in case he is. I think it could well be a pretty smart move in the sense that, you know, I think we're probably gonna see a very strong preference for humans to watch humans do things even when they are automatable. But particularly in arts. And the specific argument I think I'm seeing from Tyler is the point that no one watches chess computers play against each other, even though they're much better than humans. And the complete dominance of machines in chess hasn't ended the interest in watching humans do that. And I suspect the same is probably true of most creative arts. I don't think that—

Marc Warner (00:52:23):

People do watch quite a lot of cartoons.

Matt Clifford (00:52:25):

Yeah but humans make those cartoons, ight? As in like, I think I can imagine there being a movement, quite a strong movement, maybe not the majority, but I can imagine there being a movement in the same way that you pay more for handmade and you know things are given extra credit by consumers for being not mass produced, kind of coming out of a handcrafted thing. I suspect the same will be true for art, even in a world of full automation.

Marc Warner (00:52:52):

But I don't think—

Matt Clifford (00:52:54):

I'm glad we could disagree on something.

Marc Warner (00:52:55):

—Pamela's question is exactly about full automation. It's whether 6 billion is a smart investment. Like to Logan's point, I would foresee for the next few years that it's gonna be humans operating with machines, like the storyiess GPT-4 writes are still pretty mundane. They have the form of a good story, but they are definitely not. And so I imagine the future being more like being able to write your text story into Midjourney, get back a whole movie with beautiful special effects, and that just being in the hands of every YouTube creator and influencer. That's more where I'd be slightly concerned if I'd put that 6 billion into studio facilities, because it seems to me like— I buy that there'll be some auteur-type films, but I suspect it'll be more.

John Thornhill (00:53:51):

Alright, we're running out of time. Squeeze in one last question. This gentleman here. Please keep it short.

Audience 5 (00:54:01):

What does the panel think about Sam Altman's call for AI licences? And how would the panel recommend policymakers, law makers against regulatory capture?

John Thornhill (00:54:14):

Who would like to take that one? Saffron?

Saffron Huang (00:54:18):

I'm actually—

Logan Graham (00:54:22):

I should probably recuse myself from that answer, mostly 'cause I personally don't think I have much smart to say about it. I think—

John Thornhill (00:54:46):

I reframe the question a bit? I think very often the problem is not, in this field, regulatory capture so much as industrial capture: that so much of the expertise in this field is in the private sector and amongst the west coast companies in particular that the public sector and the universities are devoid of really understanding what's going on at the cutting edge of this technology. Isn't that a problem, Logan?

Logan Graham (00:55:17):

I mean, there are incredibly smart people in academia. All of them were in academia until a lot of them started being in companies. But it's, it's not even necessarily that or the pay, it's that it literally takes potentially billions, as Matt was saying - would imply $6 billion worth of compute from Google - that when I was doing my PhD, there were 32 GPUs that were four generations old that I had to share with 20 people in the lab at once. It's just a fundamentally different capital input that you need in industry.

John Thornhill (00:55:46):

And is that an argument for having a national research cloud, do you think?

Logan Graham (00:55:50):

Oh well James, what do you think over there? One of the main proponents.

Matt Clifford (00:55:54):

I mean I think one of the things that we, the UK government, is trying to do with the AI task force is to sort of bring together these two themes as one: just build state capacity, because it's certainly true that there are not enough people who work in government today who understand the technology even at a conceptual level, but certainly not at a technical level. And I think if we want to even get close to having the sort of regulatory regimes that have been hinted at on the panel, we just need more technical expertise in government. This is fundamentally a technical problem. And the second thing that the task force is trying to do is say: it's also just an empirical problem, and you know what Logan's already talked very eloquently about that, so I won't lay the point, but we just don't know so much about this technology and we don't know how it's gonna pan out. Saffron made that point before. I think, in general, although the sort of like paperclips metaphor is amusing and somewhat useful, I think that whole line of argument has been broadly unhelpful for getting towards like, actually good policy recommendations. And I think the good policy recommendations are gonna come from treating this as an engineering problem, where you learn what the models can do, you figure out which of those things you wanna let them do, and then you come up with a mix of policy and technical solutions to hopefully nudge these models in the direction that we want them to be.

John Thornhill (00:57:21):

Alright, we must end it there. Thank you very much to all of our panel for a fantastic discussion. And now we'll pass over to Manira for some closing comments.

Munira Mirza (00:57:38):

Um, thank you. That was a brilliant panel, really interesting discussion. I'm gonna say a few words in closing because we are at the end of the day sadly, and this is the end of the formal part of the day, but there is dinner and more alcohol and non-alcoholic drinks, so I hope that you can stay. For those of you obviously staying overnight, there is breakfast as well. But just very briefly, I wanted to say that we've covered a huge amount of ground today, clearly far too much for me to try and summarise, and I won't really attempt. I might ask ChatGPT to do a kind of summary of everything and then post it on the website. But I think there were two or three takeaways that I wanted to suggest.

(00:58:22):

One is from last night's discussion. I think it was very clear that despite disagreements there was a sense that the UK economy is in a bad state. And it's important that people do wake up to that regardless of all the fantastic new technologies that are coming and out and emerging. We have to recognise that much of the UK economy is very detached from that and it feels detached from people's lives. So, a kind of wake up call in that sense. But secondly, there is lots that can be done, lots that people are doing. So I think one thing that someone said earlier was: be a founder, start something, do something new, do something either inside the system and try and change it or start something outside the system. But rather than sitting around and complaining about it, which lots of us do, I'm guilty of it as well, there are things that can be done and we need to do them. So hopefully people will feel inspired, at the end of the day to, to do that. And then thirdly, lots of people who don't work in government think government is dysfunctional and will say they're not interested in government. Unfortunately, the government is very interested in you. So it's actually in everyone's interest that government is better and sometimes should stay out of people's lives and to allow entrepreneurs and businesses to do what they do well. Government has to be more aware of what they're doing in order to know when to hold back. And it might be that we don't necessarily want government to choose who will be developing technology or choose and make strategic bets, maybe we do, I don't know. That's a debate to be had. But certainly we might want government to take an interest in the cost of energy for businesses, and infrastructure, and there are certain things that even the most anti anti industrial strategy person might agree that it's important for government to take an interest in. And then, finally, I just want to say that there are people out there, and there are people in this room, who can give free policy advice to anyone who's interested and needs it. And I do this quite a lot informally, and I spend most of my time on emails doing that. So I'm very happy if people afterwards are doing things and they want to understand how government works. That's part of our mission at Civic Future to try and help to explain.

(01:00:47):

And, you know, we can put on events and bring people together to help them to do what they're trying to achieve. Very very briefly, we often go to these conferences, we have a great time, and then we think: what's next? How do we continue the conversation and make things happen? I hope that you're all inspired to go back out in the world, persuade people of the value of growth and tackling stagnation, talking to the general public, not just ourselves, which is a point that people have made. And also keep building connections with people in this room and people elsewhere. Hopefully some of you have met new people, and new things will start as a result. And please keep in touch with Civic Future and the events that we are doing.

(01:01:35):

We have an event planned on AI soon and many others besides. So I hope that you will continue to talk to us and do things with us. And then, very, very finally, I want to say a few thank yous to people who have helped make this event happen. First of all, I'd like to thank our speakers who have all been fantastic. Some of them have travelled a very long distance, so huge thanks to them for being part of this. Thanks to you, the audience, because the fact that you are here - you spent one and a half days long, slightly longer - being in this room away from your offices, or your labs, is really fantastic and I hope that you found it worthwhile. Big thanks to our advisory council, who I think are listed on our website, but they have been brilliant at providing input and really helping to shape this event. And I must pay a special thank you to Ben Yeoh for organising the Unconference, which passed off very smoothly and was great fun. Thank you to the staff here at the Mueller Institute and Churchill College, who have been brilliant to work with and really helped to facilitate this event. And were a great part of the team. A big thank you to the volunteers who have gone above and beyond, they are all individually brilliant people as well. They're not just volunteers at this event, they are out there doing important work. And, and some of them are experts in some of these areas. So, thank you very much for taking the time and helping to make this event happen.

(01:03:13):

A big thank you to my team at Civic Future. Thank you to Inaya, who has overseen the development of this conference and all the logistics going into it and the programming. And then a final huge thank you to Aria Babu and Nico MacDonald, who have curated this event, overseeing the programming, done lots of the work, planning, logistics, dealing with speakers, as well as the intellectual firepower behind it. So a big, big thank you to them. And so, please continue to talk and drink and make merry.