AGI Vignettes

Disorientation

I want to write about this topic called the permanent underclass.

There is just this feeling of things moving so fast right now. Every day you see crazy new developments like inference time scaling and agents coding. These models are already highly capable. They are not just doing arbitrary technical stuff anymore. They are doing consequential project work.

I think the top labs consider it a realistic possibility that this is going to create massive disparities in wealth. It is going to create a very volatile time to be alive. But it is very hard to talk about it. It is not really politically okay.

People are basically operating as if no one knows what is going to happen. Everyone would rather position themselves to have the most options instead of doing the right thing. That makes sense, but it creates this massive vacuum. Who do you actually believe? What are their real intentions? Are they going to do something behind your back?

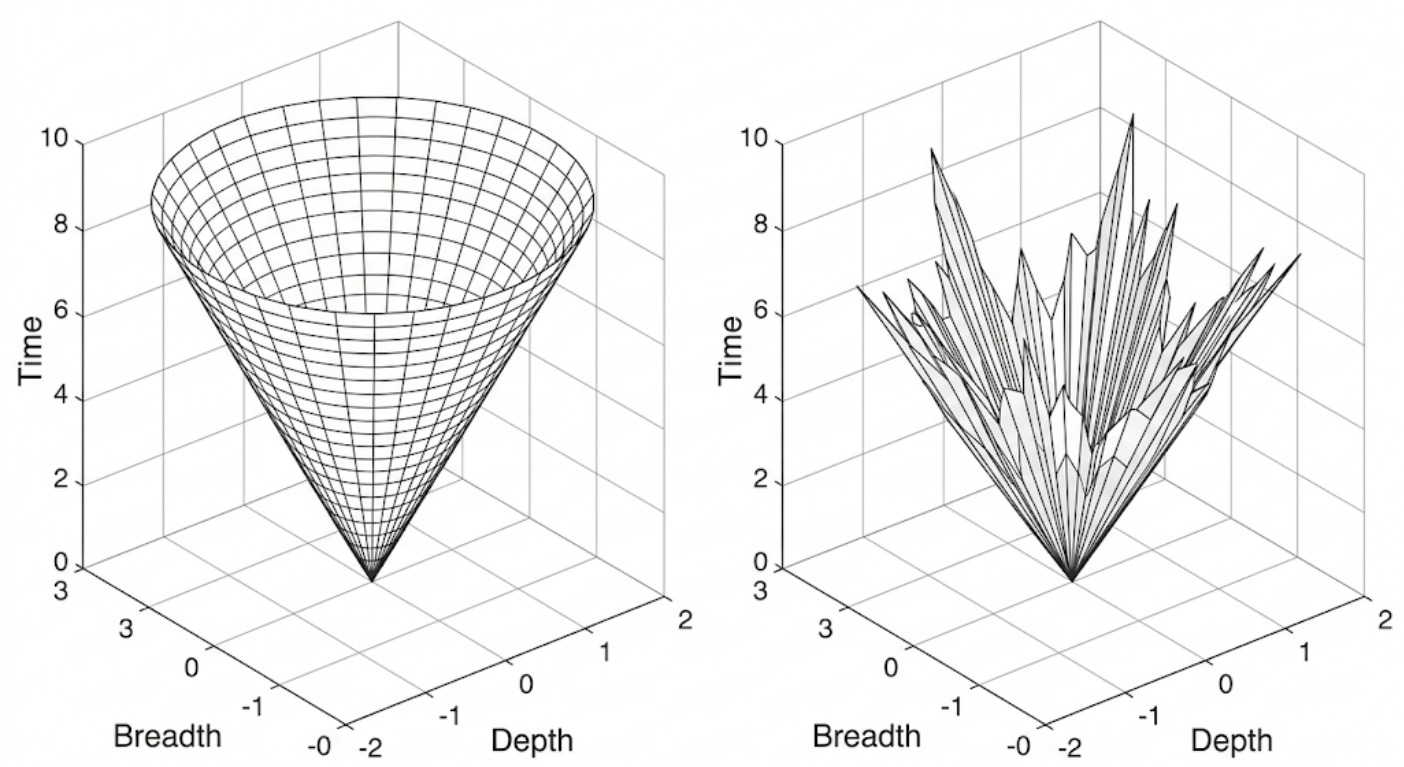
You just feel so helpless. You wonder if you should join the accelerationists, or if you should just give up and do something else. It is a tough space to be in. What do you do? Do you join the creators and lose your leverage anyway, or do you defect and try to gain as much capital as you can?

It is incredibly hard to elicit honest opinions from people on the inside because of this. Even if you actually believe things are going bad, you have no incentive to say so. Unless you can guarantee you are going to be on the winning side, your incentive is to not share that opinion. You would much rather be in the camp of people saying everything is great and not to worry. There are just huge incentives to keep building and accelerating.

Even if you see the end result, you are better off not being vocal about it until it is too late. What I am realizing is that most people inside these companies just do not have the slack to even worry about all this. They are just doing their jobs.

This is the whole notion of escaping the permanent underclass. Things are going to get tougher and tougher to get away from. It feels like getting stuck in a position where you are just not able to retain any control.

Shape of Progress (The Spiky Star)



I think sometimes it is better to take a step back and reflect on what I actually want out of this. Maybe you just want a simple life and you don't really care about all this stuff. But you can't ignore the way things are moving.

I see the progress happening on two axes right now. One is the timeline axis of how fast things are going. The other is the capability axis of how wide they can affect things.

If you map this out, people tend to think of progress as a cone. If you start plotting it, you assume it should be a cone because as time progresses, you are able to hit higher quality on a wider number of tasks.

But I think we are not really at a cone right now. It is more like a weird shape. If I can visualize it, it is more like a spiky star. Imagine you drop water on the floor from a height. You can just look at how it scatters and sprays around. Instead of a uniform cone, it is an expanding, spiky star.

We are hitting crazy advanced capabilities on very specific vectors, while other capabilities lag far behind. This is what makes it so disorienting to track how close we actually are to the end goal.

The Descent into the Permanent Underclass

I think most of the top labs do consider the realistic possibility that this is going to create massive disparities in wealth and just create a very volatile environment. But then, it's very hard to talk about it. It's not politically okay. I think people are operating as if, "Hey, no one knows what's going to happen, so I would rather position myself as having the most number of options rather than doing the right thing."

Which does make sense, but then it creates this vacuum. You start asking, "Who do you believe? What are their intentions? Are they going to do something behind your back?" It just feels so helpless.

It's hard to elicit honest opinions from people because there is this reality where, even if you actually believe things are going bad, unless you can be on the winning side, your incentives are not aligned to share that opinion. You would rather be the Pollyanna who says, "Everything is great, everything is fine, don't worry." There are obviously just huge incentives to keep building and accelerating. Even if you believe things are going to go bad, the world is set up in such a way that you would be better off not being vocal about it until it's too late. I think that is just such a bad place to be in.

What I'm realizing is that most people just don't have the slack to even worry about all this. It's a tough space to be in. What do you do? Do you join the protesters and lose your leverage more, or do you just defect and gain as much capital as you can?

That is the whole notion of escaping the permanent underclass. Things are going to get tougher and tougher to get away from this—from getting stuck in a position of not being able to have leverage. I just don't see a way out of this unless more people talk about it, express their concerns, and feel okay to discuss it. You just can't operate business as usual; it's going to be so disorienting and so confusing. It is a gradual descent into the permanent underclass, where people are going to lose their political, economic, and cultural leverage.

Individual Empowerment vs. Community Decay

I did talk about this whole new vector of how AI is going to empower individuals. There will be a local empowerment, but then there will be a global disempowerment. Usually, shared adversaries create a sense of community or collaboration. When you are empirically empowered economically or institutionally, people lose motivation in keeping relationships or communities, and communities start breaking away.

Running a community is tough because there is going to be a lot of friction. There are positive externalities, but it is exactly like a chemical reaction with a positive Gibbs free energy. In thermodynamics, Gibbs free energy determines if a reaction happens spontaneously or if it requires an outside energy source. If you put salt in water, it dissolves on its own because the reaction has a negative Gibbs free energy. But to convert water to ice, you have to put it in a freezer, which requires putting energy into the system. Even life itself operates this way. Maintaining a living organism requires a positive Gibbs free energy, which is why we have to constantly consume food and burn that energy just to make things happen and maintain order.

Building a community is the same. It does not happen spontaneously. You need continuous energy input, catalysts, or other compounds to keep the organism going. What AI is doing is essentially increasing the activation energy required to form those bonds. If a lot of people feel they are better off defecting and being on their own, they might just choose that.

This happens all the time; it is economically common. If I get a job and become independent, I can use that financial independence to meet all my needs, which disincentivizes me from putting energy into sustaining this social reaction. In the same way, if I can get most of my work done with the help of an AI, get my social needs met with an AI, or become economically empowered by an AI, why would I put up with all the messy nature of human relationships?

This fragmentation is happening at a time when we need humans to be closely knit more than ever, because there is going to be extreme power concentration and extreme uncertainty.

Corporate Facade: OpenAI, Anthropic, and Palantir

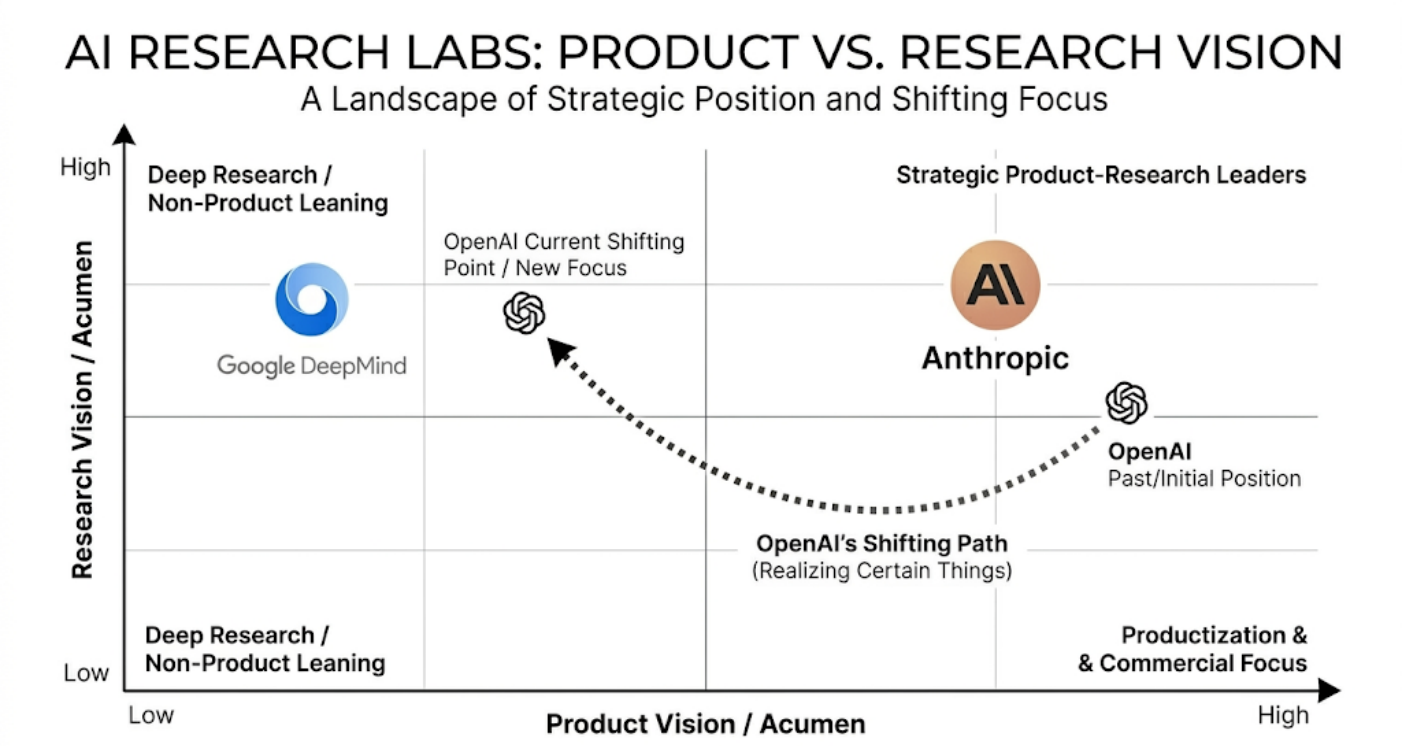
You look at startups left and right vertically integrating. They know enterprises are going to be the thing, because why wouldn't they be? But I just don't know what OpenAI is even doing. This kind of vision comes from some futuristic idea of where you see AGI going, and you're strategically

positioning yourself for that future. But then the narratives that you're sharing are totally different.

I think that is where I feel very uncomfortable. The things that you say are different from the moves you make behind closed doors—or, I mean, it's not really closed doors, it's just that they're not explicitly stated to the public. You set up a narrative or a worldview, but then you're simultaneously preparing for a gamut of worldviews, which is what strategically makes sense. But then you also maintain a notion of plausible deniability. You could play it easily every day saying, "Hey, this is just for business," or whatever. That just makes me very uncomfortable.

I could make a case that, hey, these moves you made are motivated by this worldview, but the narrative you're selling is totally different. Why would you do that if you mean something else? Most people just can't seem to see through all this veneer. I just hope it's not too late before people start questioning and holding people responsible. You have hundreds of billions in capex—building nuclear reactors, building energy stations just for this. I think everyone should read Critch's work on industrial dehumanization. Supply chains are getting shifted. The center of gravity of power is shifting just under our feet, and most people are not aware. I'm almost glad for the Iran war, because it's making a lot of things clear.

Take Anthropic and Palantir. I had written about Palantir a year back. What Palantir is doing is pretty much AGI, because you get operational data, make quick decisions, send those actions out, and you have a feedback loop. It's a cybernetic feedback loop that operates in the real world.



And Anthropic is right in the center of it. I'm really scared of Anthropic. If you look at it, there are two axes: product vision and research vision. Of the top three labs, Anthropic is in the top right for both product and research acumen. DeepMind is mostly focused on research; they are not really product-leaning and don't really interact well with product. OpenAI seemed like they were in the top right, or at least in the middle—good enough research but slightly product-leaning. But it seems like they are shifting now because they are realizing certain things.

I think Sam Altman is not really the driving force here, because most people are not "AGI-pilled" enough. Most people are more like, "Let's just see and we'll decide." But Anthropic is the most AGI-pilled lab, where they're like, "Oh shit, yeah, it's not like there are going to be any bottlenecks; it's just going to be as general as it gets and as powerful as it can get."

And if you really believe that, then what leverage would normal humans have? The most leverage is going to reside with the people with the capital. I'm not against capitalism. I think capitalism is awesome; it builds so much stuff. But you need certain checks and balances to govern it in the right way.

Navigating the Markov Chain

We are entering an unprecedented era of extreme power concentration and extreme uncertainty. A lot of times you kind of know the stable equilibrium, but you do not know the path.

This is an analogy from Markov chains. A Markov chain has states and transition probabilities. At every state, you can go to some other state with a certain probability, and the process keeps going. If you run this system for a while, you start seeing a probability distribution over the states where you will likely end up.[1] I feel a lot of predictions about the future are exactly like that. You kind of know the approximate distribution of states you are going to end up in. If you ask me exactly how that is going to happen, I have to ask if the exact path really matters. There are so many low level details that can happen in between that it is irrelevant. What matters are the aggregate macro level changes that are going to happen.

I am noticing a lot of predictions about these uncertainties, especially given the things happening right now. Most people seem to live in different timelines. If you are constantly on Twitter, you feel like so much crazy stuff is happening every single day. At the same time, the physical world is not that different yet from how it has always been.

Just worrying about it and panicking is not a solution. There are obviously a lot of good things that can come out of this technology, and you need to be able to discern and value that. So what do you do in this confusing state? I think it entirely depends on your current situation and what you want to do. If your basics are not sorted out and you have other personal things going on, you should obviously focus on that and solve your near term issues rather than dwelling on the macro picture. It is an important problem that needs to be fixed, but I have realized that focusing on the near term while maintaining a vision for the long term is much better than always worrying and trying to do too many things at once.

Keep an eye out on the progress. Everyone has to do their part in voicing their opinions and not just being complicit. The incentives are currently aligned in such a way that you are better off not being vocal about these risks until it is too late. That is simply a terrible place for us to be in.

---

[1] Note: This analogy specifically applies to stationary Markov chains, where the system eventually converges to a stable probability distribution regardless of the starting state.

Revision #2
Created 6 April 2026 08:09:04 by bhishma
Updated 6 April 2026 11:57:31 by bhishma