

Bottlenecking-related projects ideas

Will bottlenecks prevent AI from driving explosive growth?¹ If not, will they at least prevent fast takeoff?

This doc contains brief descriptions of some potential projects that are somewhat related to bottlenecks. Each could be a self-contained project, or part of a broader investigation into bottlenecks. The last two projects are particularly ambitious and poorly scoped.

I've roughly listed these projects from best to worst, according to what my preferences would be if I was choosing a project. But I don't feel strongly and would be potentially excited about any of them.

Explosive growth is more plausible for the things we care about than for human-measured GDP	1
Tell an inside-view story for how AGI causes explosive growth despite bottlenecks	2
Will AGIs give their controller a strong military advantage?	3
Talk to experts about bottlenecks in chip manufacturing	4
What % of software development tasks will super-codex automate?	5
Holden suggestion	5

Explosive growth is more plausible for the things we care about than for human-measured GDP

We can distinguish several things that AI may cause to grow explosively:

1. Physical metrics
 - a. Energy consumption, mass of manufactured items
2. Important domains that are hard to measure quantitatively
 - a. Military power
 - b. Technological capabilities
3. AI-inclusive GDP²
4. Human measured GDP

¹ This is a major objection that economists mention. This is highlight most explicitly in [Ben Jones' review](#) of the GWP report. It is also present in reviews from [Dietrich Vollrath](#), [Anton Korinek](#), and [Phil Trammell](#).

² See [Anton's review](#) of the GWP report.

The general project here is to argue that explosive growth in things like 1, 2, 3 is more plausible than for 4. Essentially, we try to meet economists' objections to explosive growth by granting they may be right about 4, and then explain why we think that explosive growth in 1, 2, 3 is still hugely important. The next three paragraphs contain ideas I'd try to flesh out if I was doing this project, but there may be other ways of making this basic point.

Some of the bottleneck-related arguments against explosive growth are somewhat convincing as arguments against (4), but much less convincing as arguments against 1, 2, or 3. E.g. the claim that people will have strong intrinsic preferences for human-produced goods and so spend ~all their \$ on such goods.

There is a good inside-view argument for why AGI would lead to explosive growth in 1, 2, and 3. In particular, it will be possible to make AGI-robot factories that reproduce in <1 year; there will be strong incentives for some actors to do this (gain relative power over others, make money by creating enough AGI researchers to cure aging); so there will be AGI-robot factories that reproduce in <1 year. Once the AGI-robot factories are collectively a sufficiently large fraction of the world, 1, 2, 3 grow explosively. Arguing that this world involves explosive growth in 4 is a further step, and less certain.

If I were doing this project, I'd look back over the specific objections leveled against explosive growth in [reviews](#) of my GWP report and identify the relative strength of these objections against each of 1, 2, 3, 4.

If I were doing this project, my first steps would be to make this case to a TAI-sympathetic economist and see if they're convinced. Then write up an explanation for why one the "strong intrinsic preference for human-produced services" argument might work against 4 but doesn't work against 1, 2, or 3.

[I just want to clarify that I'm imagining "Writing down the argument above as a hypothesis, then evenhandedly evaluating it" (rather than as "Setting out to argue for what's above and 'defeating' counter arguments"). This frame is generally implicit in Worldview Investigations, but I wanted to make sure there isn't confusion about it here. The aim to find the truth, not to defend some existing belief.]

Tell an inside-view story for how AGI causes explosive growth despite bottlenecks

The GWP report arguments are mostly based on growth modeling and historical analogies. A different approach is to make an inside view mechanistic argument for why sufficiently strong AGI would lead to explosive growth.

You start by assuming that disembodied AGI exists and can be run very cheaply. You then argue step by step for why this scenario leads to explosive growth. Throughout, consider possible bottlenecks and whether they would permanently prevent explosive growth.

Talk to economists about which parts of the story they don't buy. The (ambitious!) aim is to end up with a story they accept, at least granted the strong premises about AGI. This would allow us to say "economists object to explosive growth because they don't believe the AI tech will really be possible; they don't deny that AGI would drive explosive growth if it happened". I think this would be very valuable, as currently our position on AI and growth is seemingly in [strong tension](#) with the views of economists of AI.

(Of course, it's also possible that we end up thinking bottlenecks would prevent explosive growth. That would be valuable to learn!)

This is basically a more detailed version of a [paragraph](#) from the previous project suggestion; it might make most sense as an extension of that project.

My first steps here would be to speak to Carl. Challenge him to argue that AGI would lead to explosive growth while you play devil's advocate from the perspective of an economist that strongly expects bottlenecks. Then write up what seems to you like the most robust case and share with TAI-sympathetic economists, then other economists.

This project needs someone who groks the reasons why OP thinks advanced AI will drive explosive growth, but also understands where economists are coming from.

Will AGIs give their controller a strong military advantage?

The takeoff speeds project has, for me, highlighted the possibility that very fast hardware and software progress around AGI leads to us going from "no AGIs" to "enough compute to run billions of AGIs" in just a few years.

This raises questions like: How scary is that situation? How likely is it to create a DSA? Would having many more AGIs allow you to win wars despite having many fewer physical weapons?

One extreme view is that AGIs can very quickly and easily create nanobots and take over the world. I associate this view with Eliezer. A view on the opposite extreme is that no amount of cognitive labour can make up for having fewer physical weapons, and so you'd have to produce more weapons to gain a military advantage, which might take a long time.

I think Carl's view is less extreme than Eliezer, but he's fairly bullish on 'cognitive advantage → military advantage'. He explains many ways you could leverage the cognitive advantage to do assassinations, maneuver drones, optimise your factories, etc.

I personally feel confused about this topic, and would love to see someone evaluate the considerations that are already in the heads of people like Carl. I think the first step would be to chat to Carl, then write up your best-shot answer without research.

This topic feels somewhat related to discussion of bottlenecks. Some of the bottleneck arguments against explosive growth could potentially be used to argue that AGIs wouldn't immediately give a large military advantage (complementarity of labour and physical machinery, regulations, schlep). I like this setting for the bottlenecking discussion as it naturally dodges some of the less-relevant elements of bottleneck (e.g. intrinsic preferences for human workers).

Talk to experts about bottlenecks in chip manufacturing³

How quickly could you ramp-up production of AI chips if you had infinite money? What about if you also had infinite cognitive labour from disembodied AGIs?

In particular:

- What are the key bottlenecks in the chip production process?
 - Could they be removed with sufficient money?
 - What about with both money and remote work from AGIs?
- What are the key bottlenecks to building new fabs?
 - Could they be removed with sufficient money?
 - What about with both money and remote work from AGIs?
- What are the key bottlenecks to chip R&D?
 - Could they be removed with sufficient money?
 - What about with both money and remote work from AGIs?

I think these questions are very important for takeoff speeds. As we approach AGI there will be huge investments in AI chips. We'll use advanced AI to aid this investment, so there will also be lots of cognitive labour available for increasing AI chip production.

I would use expert networks heavily here. I'd focus on short-timelines scenarios, as it's harder to predict what things will look like decades out.

A big risk here is that the work is poorly scoped and spirals out of control. I'd just start by really understanding 1 important bottleneck and how much it could be removed.

³ H/t Leopold.

What % of software development tasks will super-codex automate?

Chat with software engineers (e.g. ppl in Lightcone) about what tasks they do and how much time they spend on each. Chat to ML people about what some iterated version of codex ("super-codex") would be able to do, and which of those tasks it could automate. Construct an estimate of how much time it would save software engineers.

This is a poorly scoped project with dozens of rabbit holes. A first step might be to identify the task(s) you expect to be hardest to automate and try to convince (e.g.) Carl that you're right. Or to identify the task(s) you're confident will be automated and try to convince some skeptic that you're right.

Why this is important:

- Automating software research is probably the most potent feedback loop in takeoff speeds. Tracking where we are in that process is a great way to track the run-up to TAI.
- If super-codex automates many software research tasks, it may trigger very large AI investments. This is relevant to timelines and takeoff speeds.
- This might shed some light on how likely GPT-X is to get us to AGI.

Holden suggestion

I think the project I am most excited about is basically this: "My first steps here would be to speak to Carl. Challenge him to argue that AGI would lead to explosive growth while you play devil's advocate from the perspective of an economist that strongly expects bottlenecks. Then write up what seems to you like the most robust case and share with EA-friendly economists, then other economists."

To spell out a bit more, I might:

1. Talk to Carl and make a list of the "craziest/most significant/most important stuff Carl thinks AI could do without getting bottlenecked." This could include robot factories, nanobots, explosive growth in available energy, etc.
2. Check in when you have the list, and pick a Carl-claim to investigate that seems both (a) tractable to get to a bottom line on (b) important if true.
3. Talk to Carl in more detail about that claim, then do an evenhanded analysis of it.