I (author of the report) have made substantial changes to the report since this review, often based on this useful feedback. I note some of these changes below, when they relate closely to the reviewer feedback.

Comments on GWP Growth

By Anton Korinek

(0) The Question

Let me start by saying that I found the paper really inspiring and thought-provoking. Even though I will express concerns about a few points below, I really appreciated that the paper made me as a reader think explicitly about these questions and, in some sense, I think this is the most important contribution of it. Given the nature of the questions, the answers are necessarily tentative, speculative and uncertain.

(1) Transformative AI and Anthropocentric GWP

My first reaction to the framing of the paper is to ask: growth in what? It's important to keep in mind that concepts like "gross domestic product" and "world gross domestic product" were defined from an explicit anthropocentric perspective - they measure the total production of final goods within a certain time period. Final goods are what is either consumed by humans (e.g. food or human services) or what is invested into "capital goods" that last for multiple periods (e.g. a server farm) to produce consumption goods for humans.

Now imagine you are a highly intelligent AI system running on the cloud. Although the production of the server farms on which you depend enters into human GDP (as a capital good), most of the things that you absorb, for example energy, server maintenance, etc., count as "intermediate goods" in our anthropocentric accounting systems and do not contribute to human GDP. In fact, to the extent that the AI system drives up the price of scarce resources (like energy) consumed by humans, real human GDP may even decline.

As a result, it is conceivable (and, to be honest, one of the central scenarios for me personally) that an AI take-off occurs but anthropocentric GDP measures show relative stagnation in the human economy.

To make this scenario a bit more tangible, consider the following analogy: imagine a world in which there are two islands trading with each other, but the inhabitants of the islands are very different from each other - let's call them humans and AIs. The humans sell primitive goods like oil to the AIs and their level of technology is relatively stagnant. The AIs sell amazing services to the humans, and their level of technology doubles every year. However, the AI services that humans consume make up only a relatively small part of the human consumption basket. The humans are amazed at what fantastic services they get from the AIs in exchange for their oil,

and they experience improvements in their standard of living from these fantastic AI services, although they also have to pay more and more for their energy use every year, which offsets part of that benefit. The humans can only see what's happening on their own island and develop a measure of their own well-being that they call human GDP, which increases modestly because the advances only occur in a relatively small part of their consumption basket. The AIs can see what's going on on the AI island and develop a measure of their own well-being which they call AI GDP, and which almost doubles every year. The system can go on like this indefinitely.

For a fuller discussion of these arguments, let me refer you to my working paper on "The Rise of Artificially Intelligent Agents" (with the caveat that the paper is still a working draft).

(2) Extrapolating from Macro Laws Describing Emergent Phenomena

One cautionary comment I would add is that economic models, and in particular macroeconomic growth models, are very high-level descriptions and generalizations of emergent phenomena that we don't fully understand. And as a result, we do not know why and for how long these generalizations will remain applicable. We don't know e.g. whether it is a coincidence that roughly exponential growth took place over the past two hundred years, or whether there are some deeper reasons why this occurred.

To illustrate this with a comparison, let's take another example of growth in the context of an emergent phenomenon - the evolution of the size of biological organisms on planet Earth. It took a very long time for the first single-cellular organisms to develop, and organisms remained single-cellular and correspondingly small for billions of years, with very small growth rates. Then there was a sudden increase in the size of organisms that were multicellular. Eventually dinosaurs appeared, giving rise to what appeared like super-exponential growth in the size of organisms. If we extrapolated from that, a growth singularity whereby the largest organisms would be infinitely large by now, or at least we would have organisms that are tall enough to grasp the moon from planet earth. The point of this obviously unrealistic extrapolation is to illustrate how macro "laws" describing emergent phenomena can easily mislead us.

(3) Driving Forces Behind Endogenous Growth

I thought it may be useful to discuss the conceptual distinction between exogenous and endogenous growth a little further. In a nutshell, what matters is whether there are decreasing, constant, or increasing returns to the factors in the economy that can be accumulated.

- If there are decreasing returns (as e.g. in the Solow model), then the model does not deliver endogenous growth and needs some exogenous driving force that is fed into the system to capture perpetual growth (e.g. exogenous population or productivity growth).
- If there are constant returns (as e.g. in the Romer-style AK model) and no growth in the exogenous variables (i.e. constant population), then the model exhibits endogenous exponential growth.

- If there are increasing returns to the factors that can be accumulated (e.g. endogenous knowledge accumulation and endogenous capital accumulation, or endogenous knowledge accumulation and endogenous population growth) OR if there is sufficient exogenous growth fed into the system (e.g. an AK model plus exogenous population growth), then the model will feature super-exponential growth.

The report puts a lot of emphasis on the question whether labor specifically is growing (endogenously or exogenously). It might be useful to frame the question a bit more generally: what are the factors in the economy that matter for production (now and in the future), which ones among them can be accumulated (now and in the future), and do these exhibit decreasing, constant, or increasing returns (now and in the future). One possible answer to these questions is that there may be structural changes, as also pointed above. For example, labor (which is not really an accumulated factor at this point) may be less relevant and capital (which is easily accumulated) may become more relevant. This highlights the importance of potential structural breaks in the macroeconomic growth laws on which we rely and, by implication, it emphasizes that it may be dangerous to use the existing growth laws for extrapolation.

I like this framing, and have edited the report to talk more generally about whether there are increasing returns to accumulable inputs, rather than focusing solely on capital and labour.

Responses to your specific questions

Framing of the question

- *The report is focused on assessing the plausibility of 'explosive growth', defined as >=30% growth, occurring by 2100. Does this framing make sense to you? Do you think we should have framed it in some other way?*

Based on the considerations listed under point (1) above, I believe that the framing of the paper addresses an important part of the interesting questions that may arise in a transformative AI scenario, but it also leaves out another important part -- the possibility (or perhaps likelihood?) that transformative AI may occur without leading to large increases in anthropocentric GWP, or perhaps even coupled with declines in human living standards.

General argumentation

- *Were the main considerations bearing on this that you're aware of covered in the report? If not, what were the most important considerations that were missing?*
- *Do you find the report's arguments for this clear and convincing? If not, what is unclear/unconvincing?*

Taking the overall question as given, I found that the paper covers the most important themes in the economic literature. Although this literature does capture important relationships, I would perhaps also assign more weight to the possibility that the observed empirical relationships may

have been driven by a bunch of different data-generating processes that just happened to end up generating the observed rates of growth throughout human history.

In other words, we may be able to roughly fit an exponential curve to growth over the past 2 ½ centuries, but perhaps there were several different economic relationships that just - by chance - ended up giving rise to what looks like a constant growth rate. Perhaps the super-exponential growth in recent millennia was driven by several completely separate economic processes. I have described this caveat more fully above under point (2).

So I would place some weight on extrapolating from the exponential growth model, some weight on extrapolating from the super-exponential growth model, and then also some significant weight on none of these relationships holding in the future.

I agree with this; the report now makes it clear that it may be that all of our growth models may omit key growth dynamics.

Knife-edge conditions

The report claims that it is difficult to find endogenous growth models that produce exponential growth without knife-edge conditions. It takes this to be a theoretical reason to expect sub- or super-exponential growth, rather than exponential growth.

- *Do you agree? Do you find the reasoning of the report convincing on this topic?*

Any theory that has predictive power can be interpreted as imposing “knife-edge” conditions or restrictions on a chaotic world - without imposing restrictions on the relationships between the variables to be explained, it would not be a theory. For example, Newtonian physics places really stark knife-edge restrictions on the data.

In my view, what matters is whether the imposed conditions / restrictions are plausible and fit the available empirical evidence. The report provides a reasonable balance on that question, although it may be useful to more strongly add a caveat on the stability of the underlying relationships, in the spirit of my comment (2).

Status of explosive-growth models

The report discusses endogenous growth models which, when fit to run long-run GWP data, predict explosive growth. Call this ‘explosive growth models’. The report argues that, though there are reasons to be sceptical of these models, these models should be taken seriously. By “taken seriously” the report means two things. 1. Taking these models’ explanations of long-run GWP growth seriously. 2. Taking these models’ future predictions seriously, on the assumption that labour becomes endogenous once again (e.g. via AI).

- *Do you agree? Do you find the reasoning of the report convincing on this topic?*
- *What do you think are the strongest objections to these models?*

- *Were important objections to these models omitted?*

I did not go through the entire report in detail, but the summary is not as clear as it could be on the conceptual distinction between exogenous and endogenous growth, which I have also laid out under point (3) above. I have edited the report to clarify this distinction.

I do agree that we should take the potential for super-exponential growth seriously - not super-exponential growth in perpetuity, but over time spans that matter for humans. The report correctly points out that there are physical limits to super-exponential growth.

It might be useful to identify and discuss these limiting factors more explicitly. For example, the short-term growth in advanced AI systems may be limited by access to energy.

When to trust explosive growth models

*The long-summary claims that the **future predictions of explosive growth models should only be trusted if labour is endogenous**. ("Endogenous" in the sense defined by those models, roughly meaning that a constant proportion of output is reinvested in increasing the size of the labour supply.[1]) Call the bolded statement P.*

The reasons given for P are that i) these models assume that population is endogenous, ii) it seems that super-exponential growth stopped around the time of the demographic transition and 'dual models' confirm this story, iii) although endogenous models with certain parameter values produce super-exponential growth with exogenous population, there isn't evidence for such models.

- *Do you think P is true? Do you find the reasons given for it convincing?*

As I argued under point (3) above, I would phrase the condition a bit more generally -- it's not about labor, but about increasing returns in all the factors that can be accumulated. If we interpret advanced AI as labor (but this may be a tenuous interpretation), then the two statements are quite similar to each other. But I think it would be useful to phrase the statement more generally without explicitly focusing on labor.

- *One objection to P might be that the statement places too much weight on the 'demographic transition' explanation for the end of super-exponential growth. If we remain more agnostic about why super-exponential growth stopped, we might put weight on these models even if labour isn't endogenous. Do you find this objection to P convincing?*

My own preferred explanation for the change in growth dynamics at the beginning of the Industrial Revolution is that there was a fundamental structural break in the underlying macro relationships.

---

[1] Note: if population is endogenous in this sense then the population growth rate is proportional to GDP/capita.

- *P seems to imply we should not use these models to predict when TAI may happen, even as one method among many. For these models can only be trusted if something close to TAI has already been developed. Does this seem right to you?*

Putting this objection in somewhat more general terms, we would predict hyper-exponential growth if there is a structural break in the growth process (one example being that TAI substitutes for labor, and labor thus can be accumulated). According to that line of reasoning, yes, I agree that this implies that the advent of TAI in itself cannot be predicted by using the existing growth process.

What would instead be needed is that there is some slow-moving process in the background according to which labor has slowly become more and more and more substitutable by technology over the past two centuries, and which is going to eventually lead to a take-off. There does not seem to be any clearly detectable sign of such a process in the data.

I think the most useful role for growth models in predicting TAI is to underline that we need an economy-wide production function that features increasing returns on accumulable factors. If we keep our eyes on this condition and try to analyze whether we are moving closer to them being satisfied, then we may be able to predict TAI.

Plans for moving forward

- *If no game-changing considerations come up, Tom is likely to work toward publishing some version of this report online. Do you think this is the right direction for us to be going in?*

I do think that the report, together with some additional generalizations and qualifications of the argument, will be a useful contribution to the debate. One of the most important parts of the contribution will be the questions that it poses itself - I am a firm believer that posing the right questions is the ultimate challenge of research.