Will Artificial General Intelligence Bring Extinction or Cornucopia? Modeling the Economy at Technological Singularity

In recent years, we have observed a significant acceleration in the development of artificial intelligence algorithms. Artificial intelligence (AI) is no longer limited to performing narrow tasks such as finding the fastest route or playing chess, but can also handle an increasingly wide range of diverse tasks. It is becoming more and more general. This progress has been particularly evident since the rollout of the ChatGPT chatbot, especially in its significantly enhanced version, GPT-4, which has been available since March 2023. Today, Silicon Valley companies such as OpenAI – the company behind GPT-4 – as well as Google, Anthropic and Meta, are competing to develop ever stronger and more general AI, regularly demonstrating breakthrough achievements along the way. The median forecast of experts on metaculus.com indicates that general AI may exceed the level of human general intelligence as early as 2032; of course, the scale of uncertainty in such forecasts remains enormous.

However, when this happens, we are likely to face a technological singularity: through a cascade of self-improvements, AI will quickly raise its intelligence level far above that of humans. The next step may be to take control of significant decision-making processes and fully automate production. This will bring about an acceleration of economic growth, but also a decline in the labor share of income, leading to technological unemployment and rapid growth of inequality.

Above all, in the conditions of technological singularity, the key question is whether the goals of superhuman artificial general intelligence (AGI) will be aligned with the long-term well-being of humanity (the so-called *AGI alignment* problem). If so, AGI can bring us almost unlimited prosperity and the prospect of conquering space ("cornucopia"). If not, it will pose an extinction risk to humanity. In particular, the scenario of conflict over resources between humans and superhuman AGI seems very likely, and if it were to occur, humanity would certainly lose.

The project aims to consider – using mathematized economic models – three essential aspects of the economy at technological singularity: AGI as the subject making crucial decisions in the economy, labor market consequences of full automation, as well as possible mechanisms of income distribution.

In particular, we will try to answer how the existential threat to humanity, which can be expected if AGI were "unfriendly," can materialize? What will the probability of such a scenario depend on? We will also attempt to weigh this existential risk against the prospects of "cornucopia" that emerge if AGI were "friendly".

Regarding the labor market, we will try to answer under what circumstances people would still work even if all tasks could potentially be automated. In which tasks would human work have the greatest value and what would the wage be? We will also try to evaluate the acceleration of economic growth due to full automation of production.

We will also address the issue of income distribution: we will review possible mechanisms for income distribution other than wages, whose role will quickly decline under full automation.

The results obtained within the project will allow us to outline possible scenarios for the future in which superhuman AGI will emerge. They can lead to important conclusions for economic policy, preparing it for the possible future arrival of technological singularity.