Motivation letter for the AI Act Scientific Panel at the European Commission

Jakub Growiec

SGH Warsaw School of Economics, Poland

Submitted in August 2025

I believe that general-purpose AI (GPAI) is a forthcoming transformative technology which has the potential to end humanity. Extrapolating the rapid exponential growth in broad-ranging AI capabilities forward suggests that, in a business-as-usual policy scenario, superhuman GPAI may emerge within 2-10 years. My central catastrophic scenario is the one where misaligned GPAI rapidly proliferates across the available compute, resists switching off, and takes over our civilization's networked infrastructure and key decision processes. The ensuing human disempowerment can then either take a violent form, resulting in rapid human extinction, or a nonviolent form in which humans "only" gradually lose access to our key life-sustaining resources.

Humanity is unprepared for superhuman GPAI even in the scenario where the catastrophe is averted, and in my perception the positive scenario is unlikely to materialize given the insufficient regulation of the AI ecosystem, fierce competition among the AI labs, and the ensuing geopolitical race dynamic.

I believe that development of frontier GPAI capabilities should be paused as soon as possible in order for AI safety and alignment research, as well as AI governance, to catch up. Research on GPAI alignment ought to ensure that the superhuman GPAI will be dedicated to *selflessly care for the long-term flourishing of humanity*. Before any such model is released or even trained, we need to be certain that it will not bring about existential risk. Ramping up investment in AI safety and alignment research, in ways which do not inadvertently increase AI capabilities, should in particular help address the question if GPAI alignment under the current machine learning paradigm, with very little mechanistic interpretability (most of the AI's cognition and action happens in a "black box"), is at all feasible.

Unfortunately, in the current state of affairs, we have no indication of GPAI alignment, and no guardrails in place against reckless, racing AI labs which are training and releasing increasingly capable and agentic, but blatantly misaligned AI models.

I am strongly driven to contribute to improving AI governance. The European Union, while not being the leader in the global race towards superhuman GPAI, could still act as a positive example and spearhead prudent AI governance, which would go beyond privacy and intellectual property rights protection and into the realms of governance of dangerous GPAI capabilities and existential risk reduction. I hope EU actions could eventually facilitate the much needed global policy coordination on advanced GPAI—and ideally a globally coordinated pause.

I am a Professor of Economics at SGH Warsaw School of Economics, Poland. My core research topic is the drivers and mechanisms of long-run economic growth. Several years ago I realized that there is a fundamental qualitative difference between industrial-era technologies which

mechanize physical action but augment human cognitive work, and the emerging digital technologies which automate human cognitive work away. The advantages of digital computation compared to human brains (in terms of speed of computation, speed and accuracy of data transmission, replicability, coordination, etc.) suggest that once these technologies become sufficiently capable and gain control of sufficient computing power, they would overcome the critical growth bottleneck, owing to which in the early 21st century the pace of economic growth (~2-3% per annum) has not been keeping up with the pace of digital expansion (~20-30% per annum, the Moore's Law). Growth may accelerate when AI *fully* automates large parts of the economy, pushing human decision-making and supervision out of the loop (Growiec, 2022; Growiec, Jabłońska and Parteka, 2024). Unfortunately, the prospect of fully automating large parts of the economy also implies rapid concentration of power in the hands of AIs and AI companies, and—most importantly—loss of control and AI takeover.

Although systematic economic growth tremendously increased human prosperity and well-being since the Industrial Revolution, it has been the case because decentralized human control ensured that growth benefitted the people, and the scarcity of human cognitive work helped distribute these benefits widely (we cannot accumulate brains in a way we accumulate physical capital, which limits concentration of wealth and power). In a world where GPAI replaces human cognitive work, however, these conditions will likely no longer hold. GPAI scales proportionally to accumulable digital compute instead of non-accumulable human brains, and digital software can be (almost) instantaneously and costlessly copied. Hence, GPAI-driven economic growth may no longer benefit humanity, and GPAI will likely bring existential risks (Growiec, 2024). Under any realistic degree of risk aversion on behalf of the human policymakers, reducing this risk should be a top policy priority (Growiec and Prettner, 2025).