

The AGI Race and Existential Risk

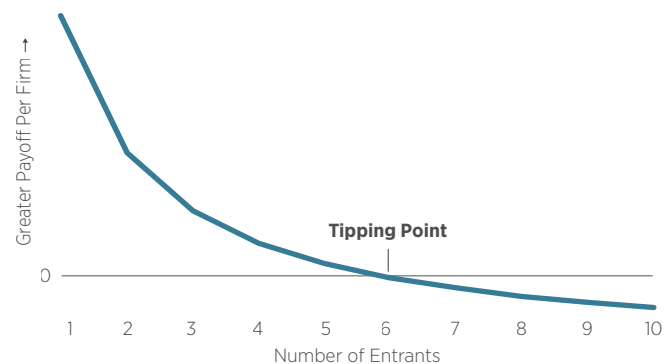
Based on BFI Working Paper No. 2026-71, *“The AGI Race and Existential Risk,”* by Ethan Bueno de Mesquita and Wioletta Dziuda, University of Chicago; and Mattias Polborn, Vanderbilt University

Competition among firms developing artificial general intelligence can encourage them to prioritize speed over safety, increasing the risk of a catastrophic outcome. Industry consolidation, regulation facilitating commitment to slow down, and governments entering the race in a safe way improves welfare, while restricting resources may backfire.

Artificial general intelligence (AGI) — systems capable of performing a wide range of economically and strategically valuable cognitive tasks at or beyond the human level — could generate enormous benefits. The first firm to develop AGI could also gain a substantial competitive advantage. Meanwhile, developing highly capable AI without adequate safeguards could carry serious risks: if a powerful system cannot be controlled adequately, is deployed too soon, or falls into the hands of malicious actors, AGI could cause catastrophic harm rather than broad social benefits. This outcome is referred to commonly in frontier-AI discussions as “doom.”

The race to AGI therefore poses a tradeoff to firms in deciding whether to allocate their limited resources — including computing power, engineering talent, energy, capital, and time — to speed or safety. In this paper, the authors consider this tradeoff, and how it is impacted by factors including market structure, resources, and policy. The authors study these forces using a model of the AGI race in which competing firms

Figure 1 • Payoff Decreases with Number of Entrants in AGI Race



Note: This graph shows how the expected payoff for each firm falls with each additional entrant in the AGI race, until hitting a critical number of entrants where there is no expected payoff for each individual firm.

select between allocating resources to safety or speed depending on the likelihood of arriving first and doom risk. The model reveals the following:

- **Race for AGI is too fast.** Both the society and the firms prefer slower but safer race, but each firm individually has an incentive to devote excessive resources to speed to increase its own chances of winning.

- **More competition can make the AGI race faster and riskier.** As the number of firms increases, each firm becomes less likely to develop AGI first and therefore has less incentive to make its own arrival safe.
- **Firms may continue racing even when achieving AGI has negative expected value.** The authors identify a critical number of competitors. Below this threshold, the expected benefits of achieving AGI exceed the expected costs for each firm. Above it, the combination of a smaller chance of winning and greater catastrophic risk makes achieving AGI negative in expected value. Firms may nevertheless keep racing because withdrawing does not protect them from risks created by competitors.
- **More resources always accelerate AGI development; however, they can make it either safer or riskier.** When relatively few firms compete and achieving AGI has positive expected value, firms devote relatively more additional resources to safety, and as a result AGI arrives sooner and safer. When many firms compete and achieving AGI has negative expected value, firms devote relatively more new resources to speed, making development both faster and riskier.
- **Regulation can help firms commit to moving more slowly.** A firm that slows down on its own risks falling behind. At the same time, the authors show that when one firm can credibly commit to greater caution, its competitors also have an incentive to slow down. A small commitment to reduced speed can leave all firms better off. This may help explain why AI firms themselves sometimes favor binding regulation: Common constraints can reduce the competitive cost of exercising caution.

- **A cautious public AGI developer can also improve outcomes.** A public entrant that prioritizes safety can change private firms' incentives by making defeat less risky. Under some conditions, this encourages private firms to slow down as well.

The findings suggest that policies aimed at reducing catastrophic AGI risk should focus not only on technical safeguards, but also on the competitive incentives facing developers. Measures that allow firms to credibly commit to slower development could improve safety without putting individual firms at a competitive disadvantage. The analysis also cautions against assuming that limiting resources will necessarily make AGI development safer: depending on the level of competition, resource constraints could push firms to devote a greater share of what they have to speed rather than safety. More broadly, effective policy will depend on how regulation, market structure, and resources interact to shape firms' incentives to prioritize safety.

READ THE WORKING PAPER

NO. 2026-71 · MAY 2026

The AGI Race and Existential Risk

bfi.uchicago.edu/working-papers/the-agi-race-and-existential-risk

ABOUT OUR SCHOLARS



Ethan Bueno de Mesquita

Dean and Sydney Stein Professor,
Harris School of Public Policy



Wioletta Dziuda

Associate Professor and Deputy Dean for Faculty
and Research, Harris School of Public Policy

