

Overall, with strong conceptual arguments but no public empirical evidence, it seems plausible but unproven that some AI systems will seek power.

---

when there are systems which are more capable, they'll probably be at least somewhat more goal-directed and then once we have goal-directedness, we can more convincingly argue that power-seeking is going to be a thing because we have theory and so on, but there's a lot of uncertainty about it because we don't know how much systems will become more goal-directed." [54:35] ([AI Impacts, 2023c](#))