

Odds of Artificial General Intelligence by 2043

2023 Open Philanthropy AI Worldviews Contest

Seth Rhoades

The arrival of artificial general intelligence (AGI) could transform society, commerce, and nations more than the technological revolutions of the past. Postulating when AGI arrives, what a world with AGI may look like, and the progress of AI *en route* to AGI is worthwhile. This Work adopts a broad-scoped view of current developments and opportunities in AI, and places it amidst societal and economic forces present today and expected in the near future. As the appearance and form of AGI is hard to predict, the basis to form the odds of AGI within 20 years is instead derived from its impact and outcomes. The qualifying outcomes of AGI for this Work include conducting most tasks cost-competitively to livable wages in developed nations, performing innovative and complex high-skill work such as scientific research, yielding a durable 6% Growth World Product growth rate, or inducing massive shifts in labor distribution on par with the Agricultural or Industrial revolutions. A survey is first taken on prior works evaluating recent technological developments and posing remaining capabilities necessary to achieve AGI. From these works comes a baseline odd of 24.8% for AGI by 2043, which is then balanced against arguments for or against these timelines based on what AI can do today, a representative array of tasks that AI cannot do but may qualify AGI if successful, as well as observations of phenomena not considered in the more technical prior works, namely those of labor, incentives, and state actors. Likely and impactful tailwinds to AGI timelines are developing new paradigms of AI, on par with reinforcement learning, capable of wholly distinct tasks from those done by AI today. Less likely but impactful tailwinds include the ability for AIs to physically manipulate a diversity of objects, and development of numerous new “narrow” AIs to collectively perform a diversity of tasks. Likely and impactful headwinds to AGI timelines are the continuation of outsourcing to abundant excess labor globally and the long economic growth trajectories of developing nations. Less likely but impactful headwinds include large economic recessions, globalized secular stagnation, and insufficient incentive to automate fading yet ubiquitous technologies and services. Applying subjective and weighted probabilities across a myriad of scenarios updates the baseline 24.8% odds to propose a low, median, and high odds of 6.98%, 13.86%, and 20.67% for AGI by 2043.

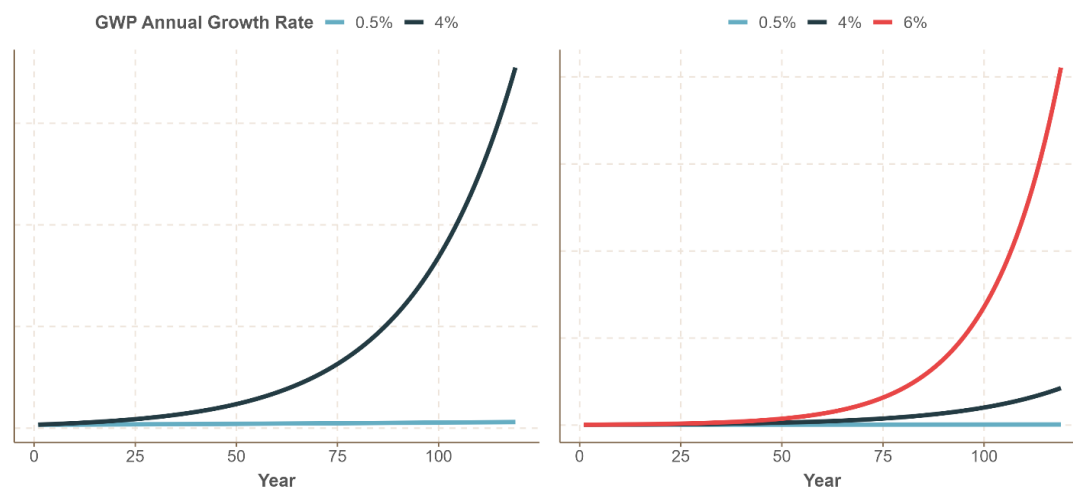
I. Definitions, technical states & Base Rates

This section defines Artificial General Intelligence (AGI) and reasons the Base Rate of 24.8% for AGI development by 2043 ($\text{Pr}(\text{AGI}_{2043})$). The remainder of this Work will then provide cases for an adjustment to this Base Rate. Readers not interested in definitions or rationales may skip to Section II.

I.A Defining AGI

The definition of AGI for this Work considers capabilities of and outcomes after development of artificial intelligence (AI) technologies. These capabilities include one or more computer programs to perform almost any human task competitive with livable wages in developed countries. This Work considers AGI as either a singular, task-achieving AI system which also achieves “intelligence”, *or* a suite of “narrow” AIs (nAIs). Examples of nAIs today include image recognition or language models to produce human-like text. Additional capabilities required for AGI in this Work include AI systems operating businesses or performing human-created endeavors such as scientific research and development. While a collection of multiple nAIs themselves would not necessarily constitute a “general” or “intelligent” AI system, nAIs capable of complex tasks with agency, strategy, or decision-making in a human-centric world of enterprise and governance entail a similar degree of technological advancement as the more abstract notion of AGI.

Additional outcome-based definitions of AGI include both profound economic transformations and massive job displacements. This Work sets a bar close to the Industrial Revolution, which bent linear Gross World Product (GWP) growth to exponential. An increase in the annual GWP growth rate from 4%, [the average rate since 1900](#), to a persistent ~6% qualifies as an AGI outcome in this Work. Persistence is defined as a durable rate outside of depressions and temporary hypergrowth periods as observed around the Great Depression and Recession.



than 5% by 2000. The percentage of labor in services rapidly *increased* to over 50% in developed countries since the Industrial Revolution but occurred after the partial displacement of agriculture first by manufacturing. Thus in this Work, only 40% or greater *decreases* labor in any large sector(s), including services, qualifies an AGI outcome.

GWP may not be able to capture the impact of AGI. Other outcomes, such as a drastic curtailment of humanity or human extinction due to AGI could qualify. This Work does not factor these outcomes, as the baseline estimates of these events occurring by 2043 due to artificial systems are negligible compared to the probabilities of anthropogenic wars or drastically reduced living conditions on this planet. However, this Work also then considers an outcome of AGI to be a durable 0% GWP or less. This Work limits outcomes to economic indicators, as fewer consensus and accessible quantitative measures exist for culture, quality of life, or geopolitics, though AGI would surely impact these domains as well.

I.B Review of prior work

Prior works on this topic deeply cover AI development to date, technical and financial requirements to develop AGI, and how prior eras of technological progress can predict AI timelines. This section amalgamates these works to help form a Base Rate for $\Pr(\text{AGI}_{2043})$.

[Carlsmith](#) broadly writes about existential risks of AGI to humanity, but includes assessments of algorithmic, compute, and financial feasibilities to develop AGI and describes capabilities of an intelligent artificial system. The probability of technical feasibility and an environment with incentives to build such a system by 2070 is 65% and 80%, respectively. Capabilities of AGI include “advanced capabilities”: conduct scientific research or engage in political strategy, “agency”: create and execute tasks, and “strategic awareness”: formulate plans and update them amidst real-world environments or test what-if scenarios. This work introduces concepts of intelligence as emergent properties of nAIs, demonstrations of awareness in AI systems today, and the response of society to “warning shots” of near-AGI systems around notions of safety.

[Cotra](#) focuses narrowly on when a transformational AI (TAI) system could be built, factoring advances in computing power and costs, the data and computing requirements of AI today, and “compute” estimates in human intelligence. TAI is defined as one or more AIs that can perform most jobs cheaper than humans, universally recognize and classify objects, create narratives, teach itself new skills, perform economically valuable work, and engage in scientific research. The probability of TAI by 2030 and 2040 is 15% and 50% respectively. This work introduces concepts of disparities in data requirements between artificial systems and humans, algorithmic efficiency, bottlenecks of TAI by compute power and costs, and achieving TAI through generalizing existing AI algorithms.

[Karnofsky](#) assesses AI’s capabilities today and projects the advancements towards TAI, defined similarly to Cotra in its capabilities and impacts to labor. The probability of TAI by 2036 is 10%. TAI’s capabilities include conducting scientific research and human-level performance in both physical and intellectual tasks competitive with at least 50% of total

market wages. An added qualifying outcome of TAI is producing economic growth at the scale of the Agricultural or Industrial Revolutions. This work introduces concepts of TAI through generalizing existing AI algorithms and de-anthropocentric AI, or TAI that does not emulate human brains and cognition.

[Davidson](#) models probabilities of AGI development with little attention to technical AI progress to date, but rather by AGI's estimated difficulty, research effort, computing advancements, success and failure rates, and spending on AGI development. The probability range of AGI by 2036 is 1-18%, with a median at 8%. Davidson provides a median adjustment down to 5.7% percent, conditional on AGI being impossible by 2023. AGI's capabilities are heavily weighed on cost-competitive accomplishment human-level performance on cognitive tasks. Davidson's model of AGI development is most influenced by the estimated difficulty of the problem, and secondarily on the growth of research(ers), spend, and compute. This work introduces concepts of expectations based on progress in other fields, AGI's as a product of sequential advancements, and diminishing returns on research and investment observed in other scientific industries.

I.C Arriving at a Base Rate of $\text{Pr}(\text{AGI}_{2043})$ for this Work

The Base Rate of $\text{Pr}(\text{AGI}_{2043})$ is a weighted sum of the referenced works above, derived from a thematic overlap of concepts and relative strengths and weaknesses of their stances.

The salient probabilities from Carlsmith for this Work are technical feasibility and incentives to create AGI by 2070, which are $65\% * 80\% = 52\%$. Carlsmith is unique among the referenced works to consider incentives, deployment to real-world applications, and societal response to AI development, all of which decrease odds of AGI. These concepts are viewed favorably here, and while anecdotes of AI's accomplishments and their expected extension towards systems with agency, strategy, and planning are viewed skeptically, the technical assessments are viewed favorably too. The odds of AGI are only defined for 2070, and it is not stated how the odds change over time. Therefore, the $\text{Pr}(\text{AGI}_{2043})$, given the time of authorship in 2022, is $(2043-2022)/(2070-2022) * 52\% = 21.9\%$.

The $\text{Pr}(\text{AGI}_{2043})$ from Cotra is $(2043-2022)/(2040-2022) * 50\% = 58.3\%$. The technical assessments in this work are viewed favorably, however the expectation TAI will emerge from continual advancement of existing AIs today, and as a byproduct of increased compute power at decreased cost, is too large an assumption and narrow a view to be given high weight in the updated Base Rates. Karnofsky shares expectations on TAI from improvement on existing AIs, but raises the bar to compare TAI's outcomes to prior technological revolutions. This raised expectation may partly explain the decreased odds of TAI by 2036. The $\text{Pr}(\text{AGI}_{2043})$ from Karnofsky is $(2043-2016)/(2036-2016) * 10\% = 13.5\%$. Both these works will be given equal weight, as they differ mainly on the qualifications of TAI.

Davidson offers less technical but important perspectives to the broader consideration of AGI development. These perspectives are viewed here with mixed results. The concepts of sequential models, research(er) effort, technological progress, and non-linear changes in

probabilities over time are useful. Nonetheless, the model used in this work is strongly influenced by a single variable, namely the prior expectation of AGI. Varying the weights of other variables does not appreciably impact the overall probability, and thus the utility of this model is viewed skeptically. Additionally, the naïve model only differs from the informed model by one percent. The $\Pr(\text{AGI}_{2043})$ from Davidson is $(2043-2020)/(2036-2020)*5.7\%=8.2\%$.

These odds are given subjective weights of 0.4, 0.2, 0.2, and 0.2 from Carlsmith, Cotra, Karnofsky, and Davidson respectively, yielding the estimates:

$$(21.9\% * 0.4) + (58.3\% * 0.2) + (13.5\% * 0.2) + (8.2\% * 0.2) = \mathbf{24.8\%}$$

II. Technical assessment and critique on AI today

In this section and throughout this Work, I will ground arguments and opportunities in AI development across mundane and micro tasks, such as cooking and haircuts, up to macro and abstract sectors of healthcare, automobiles, construction, and business enterprise. Examples in these domains will highlight what AI can and cannot do today and opportunities for new AI developments, which will then be translated to subjective odds of developing such capabilities and integrated to $\Pr(\text{AGI}_{2043})$.

II.A Defining Tasks and limits of data, generalization, and the unobservable

Most AI accomplishments to date are categorically modeling or learning¹. Notable modeling tasks include processing and classifying images and videos, reading and rendering text, and listening and responding to auditory signals. Learning-based tasks include playing real or virtual games. In all these tasks, AI operates in data-rich domains, either of the senses or in simulated environments with effectively infinite artificial data and outcomes of a sequence of actions. Cotra notes these AIs require more than 2 orders of magnitude of data than humans to learn tasks. The requirement for either a super-abundance of data or high outcome availability are one limitation to nAIs generalizing to other tasks. In addition, tasks operating within rules, such as playing chess, depend on readily available and simple rules, which do not apply to many real-world tasks.

I propose a general formula for an AI to accomplish most human tasks:

$$\text{Task} = \text{function}(\text{Rules}, \text{Domain Knowledge}, \text{Physical Manipulation}, \text{Data}, \text{Outcomes})$$

For a given task, the relative weights of these factors vary considerably, and tasks do not require all factors. In addition, many tasks can be accomplished in multiple ways, for example through highly curated rules and domain knowledge or by brute-force with abundant data. Tasks involving the senses, for example image classification, perform almost purely through

¹For the technical reader, I consider supervised, unsupervised, shallow and deep neural networks as modeling, and reinforcement learning as learning AIs.

data, with a non-zero influence of outcomes (i.e. ground-truth labels of the image), and no requirement for physical manipulation. Learning-based tasks, such as playing chess, could either require explicit knowledge of rules or implicit knowledge from the unobserved actions within a preponderance of data. Regardless, all learning tasks require outcomes. While rules can be safely ignored in the case of text generation, where a corpus of text is sufficient to build a model of “real” words, their relative orderings, and grammar, the unobservable information in other domains may be because actions either violate rules (examples being laws of legal systems or of physics) or resulted in failure. The unobservable failures or rule violations, even in data-rich domains, implies that for AIs to accomplish many unsolved real-world tasks, AIs must contain explicitly defined rules, “learn” fundamental rules, e.g. laws of physics, or codify domain knowledge in anthropogenic endeavors.

II.B Path dependency, extensibility, and progress

At their core, whether as statistical models or agent/environment-based learners, most performant AIs today are variations of neural networks, with their architectures differing in complexity and assembly of the network’s components. The path to neural networks’ success today originated in 1960s, with the most notable achievements in the past 20 years. These achievements are largely improvements in hardware and algorithmic efficiency, and less from fundamental changes of the algorithms themselves. The path dependence of neural networks to AI today evokes a broader question of whether they represent the global maxima of algorithms, or is there a yet-undiscovered category of AI that better incorporates rules and domain knowledge, or (limited) data, and (unobservable) outcomes? Aviation, computation, and bioengineering are prime examples where successive technologies required transformative innovations to climb out of their local maxima. Modern aviation’s journey began with fixed-wing gliders and rotorcrafts, followed by jet engines, piston engines, and gas turbines. In computer programming, the term “extensibility” is used to characterize the ability of code or software to solve problems beyond its initial intended use. Piston engines could not extend to the demands of high-performance aircraft or long-distance air travel made possible by gas turbines. This notion is readily apparent in biology as well, where *E coli* are engineered to produce many chemicals it does not natively produce, but are not sufficiently extensible to produce all chemicals used in the synthetic biology and biopharmaceutical industries. While the limits of *E coli* are not yet reached, other organisms are increasingly under study to solve broader challenges in these industries. Likewise, AI has not yet hit the limits of neural networks and the modeling and learning algorithms they power. But whether as new innovations, seen with engines, or stacking complementary technologies, seen with semiconductors and fabrication, the history of industry and progress would predict that additional nAIs, or more generalizable AI, will likely require magnitudes of change to extend to the breadth of tasks required for AGI.

An important distinction between aviation, computing, and AI are the available measures of advancement. For aviation, the weight to power ratio is a primary indicator of progress, and for semiconductors, the number of transistors. In deep learning, the number of model parameters is an analogous metric, where more is generally better, though this measure does

not apply as neatly to learning AIs. AI models may be evaluated by their accuracy or performance relative to humans, which to date are only established for narrow tasks such as image classification. While broad consensus performance metrics of a technology or industry are not mandatory for progress, the stubbornness of this challenge for the field of AI may produce drag on its advancement to AGI.

II.C AI in select examples

Cooking AI is achievable through physical manipulation of objects and either codified knowledge, including agreeable mixtures of flavors, cooking durations, and fundamental rules of chemistry, or by data-intensive associations to outcomes. Outcomes today are largely subjective tastes (the scientific domain of organoleptics is still in its infancy), but despite the frequency that humans eat and the size of the food industry, outcomes data is scarce. However, a large corpus of recipes and their ratings do exist and can act as a proxy for input data and their outcomes. This form of data-outcome modeling is highly compatible with existing AIs today, and much more likely to accomplish Cooking AI than through domain knowledge and rules. *Thus, Cooking AI is highly achievable in the digital domain of its subtasks, and in light of progress on robotic food preparation systems today, moderately achievable in the physical manipulation of objects.*

Haircutting AI is achievable either by codified domain knowledge and rules, including physics and actions that inflict human harm, or by data-intensive associations to outcomes. Codifying domain knowledge may prove difficult, particularly for translating human intuitions on how to manipulate hair of varying densities, lengths, and moisture content into a quantitative domain, and then to combine those measures with physical forces and movements of razors and combs. For the data-outcome approach, outcomes are simply images of haircuts, although data, if represented as a sequence of actions by physical objects in three-dimensional space, is non-existent. If this problem is cast as a learning-based task, a simulation environment may be possible to generate infinite sequences of haircutting actions and a final (simulated) cut-hair end state. Ultimately, this simulation needs to cross into the physical domain, possibly on a model of a human head with robotic arms and devices moving in physical space. *I view Haircutting AI to be moderately unlikely in the digital domain of its subtasks, given little existing work and data on this type of task, and limited cases of modeling or learning AIs in the physical domain. In addition, the physical manipulation of objects is both essential and more difficult than in Cooking AI.*

Unlike the directed actions-to-outcome nature of cooking and cutting hair, many tasks require the ability to diagnose a system of parts, understand functions of individual components, and combine those components to form an integrated unit. Composability, which may represent a new category of AI, would greatly increase likelihood of Auto Repair AI. In this context, composability can be thought of as an extension of domain knowledge. Auto Repair AI may be achievable through a data-outcomes approach, although like the service-oriented food and personal care sectors, automotive repair is a fragmented industry with low outcome availability. If a learning AI were applied to this problem, a simulation

environment may be created akin to Haircutting AI, however a physical environment to diagnose component performance, assemble parts, and test outcomes would require much more complex systems. *I view Auto Repair AI to be moderately unlikely in the digital domain of its subtasks, given a new category of AI may be required and limited cases of learning AIs in the physical domain. The physical manipulation of objects is both essential and at least as challenging as Haircutting AI.*

Construction contains many analogous tasks to automotive repair. Also analogous to automotive repair is the fragmented nature of the construction industry and low outcome availability. The rules and domain knowledge embedded in designing, planning, preparing, and assembly are vast, with many variables and dimensions of possible outcomes. For instance, cars have finite parts with limited substitutability, while building configurations are primarily limited by physics and feasibility against the space onto which the building is built, which itself contains variables of terrain, soil, and climate, among others. Given the many subtasks and expertise required of construction, Construction AI would likely be solved with multiple nAIs. *Among the major tasks, design is reasonably feasible through a data-driven approach, as there exist many blueprints and consensus preferences on building configurations. I view planning to be somewhat unlikely, given the extensive rules and domain knowledge required in building, permitting, approvals, codes, and regulations manifest by many roles in this industry. Analogous to Auto Repair AI, a new category of AI may be required for building tasks in construction, and physical manipulation of objects is extremely challenging, as has been noted even for mundane related tasks such as [laying bricks](#).*

Healthcare AI requires either extensive domain knowledge, including known conditions and their treatments, coupled to rules, such as actions harmful to humans, or data-intensive associations to outcomes. There has been considerably more effort to codify domain knowledge in healthcare than services, automotive, and construction, in both clinical research and private industry. While successes exist, including clinical decision support tools, the effort required is enormous relative to the percentage of diagnostics and decisions made in human care these tools impact. Composability, considering anatomical and physiological components of the body and collective human function, would benefit diagnostic tasks in Healthcare AI, but diagnostics is also possible through data-driven modeling. Much more effort, spanning decades, has also been applied to this approach than other industries. The results of these efforts are mixed. Diagnostic AIs today perform strongly to diagnose conditions mappable from sensory information, such as images. Furthermore, the strong incentives to collect data in health systems of developed countries yield copious data on human behaviors, diagnoses, and clinical decisions. Nonetheless, with all this information, “AI doctors” are slow to be developed and deployed. While there are many reasons for the slow rise of Healthcare AI, I argue a large technical reason is that what is missing from health data is as important as what exists. In addition, outcome availability is limited, fragmented, and biased towards outcomes of ineffective diagnoses and treatments, i.e., an inverse survivorship bias. Putting aside physical manipulations, considering the work in Healthcare AI to date, *I view generalized Healthcare AI to be*

somewhat unlikely. Nonetheless, a suite of nAIs operating purely in the digital domain of tasks may still have a large impact on the industry. Most healthcare events are for routine symptoms with predictable solutions and outcomes, and more than 20% of healthcare costs in the United States are related to diabetes care (or ~1% of GDP). In addition, healthcare involves many administrative functions, and increasing automation of these tasks is highly feasible. Thus, I view Healthcare AI that meets the requirements of AGI for this Work to be moderately likely.

Contracts are essential in most industries and transactions. Although contracts operate within known rules, including those from legal systems, an industry-agnostic Contracts AI would be very difficult to approach through domain knowledge and rules. Rather, large language models today may already be capable of generating routine contract templates. Outcomes are less critical here, and as with Healthcare AI qualified by only a subset of diagnosis and treatment capabilities, Contracts AI may not need to generate all types of contracts or write complete documents to impact the industry broadly. Data is not as readily available as other bodies of texts such as cooking recipes, given the less public nature of the legal profession. However, the efficiency of language models, partly measured by the amount of data required to produce text, is continually decreasing. *Contracts AI is thus highly likely.*

Company formation and operations, which may be for purely digital services or products, requires acting with agency within an environment. An agent-environment example could be simply receiving feedback on a product and adjusting the product features or price. Other tasks frequently involved in operating companies include negotiation, bargaining, and strategic planning. Many of these tasks are already possible with language models that can converse with humans. These AIs are predominantly confined to 1-on-1 interactions today, which is not the case in many business interactions operating under multiple agent-principal environments. *Even so, expanding existing AI tools to engage in multi-actor interactions is feasible, and thus negotiating and bargaining subtasks for Company AI is likely.* Other tasks requiring agency or adjustment of actions against an external environment are analogous to how learning AIs play games today. These AIs “unroll” a series of actions and adjust their behaviors based on the expected outcome, confirmed by the actual outcome of the game. Transferring this approach outside a game would require mapping actions of games to real-world actions and quantizing outcomes. *I view a learning AI approach to acting with agency in business operations to be very unlikely, given most observable outcomes cannot be readily attributable to specific actions, and most actions and outcomes are not observable.* “Failures” in a game are low stakes, whereas “failure” in this case is death of the company. Death of the individual can benefit the collective, if outcomes and causal links of actions (or inactions) are defined and disseminated, none of which happens today, nor viewed as remotely likely to occur soon. *Company AI may thus require a new category of AI which can operate with highly incomplete information or accomplished by a large-scale reengineering and mapping of actions and outcomes from games onto the human domain, both of which I view to be very unlikely.*

III. Economies, state actors, and labor

Technological development does not occur in a vacuum of fields, factories, or server rooms. Arguments on AGI timelines in this Section are grounded by selected observations in labor markets, economic trends, and recent private sector activities. These observations are then projected onto conceptual frameworks in labor theory and macroeconomics to propose a series of headwinds and tailwinds for the advancement of AGI and its impact on GWP and Labor as defined in Section I.

Observations are as follows:

- In 2017, \$2.7 out of \$5.1 trillion of manufacturing labor globally was estimated to be already automatable [\[ref\]](#)
- Both Google and Meta independently developed AI to predict protein structures and play strategy games
- The annual number of electric vehicles (EVs) sold globally will surpass internal combustion engines (ICEs) in 2040, with absolute numbers of EVs and ICEs at roughly 600 million and 1.2 billion respectively [\[ref\]](#)
- The number of low-tech jobs created more than double high-tech labor in geographic areas of new growth [\[ref, ref\]](#)

III.A Private sector activity, incentives, and industrial turnover

The massive value in private enterprise, and likely also in public functions, that could be automated today but are not, implies the development and deployment of AI systems are not solely technological factors. Among many possible reasons, I argue investment, narrow competition, momentum of existing business practices, and incentives are important ones.

Most successful deployed nAIs were developed by technology companies with large research divisions or smaller foundations and non-profits with support from large technology companies. The applications of highly developed language and image recognition models are most evident in product improvement, such as content recommendation, search engines, and driver-independent vehicle control. Other projects, such as playing the game of Go and predicting protein structures, are R&D projects with large but uncertain payoffs, and required particular decisions made with excess capital in favorable macroeconomic environments. Examples of costs and timelines of these projects shed light on the requirements to advance nAIs today. AlphaGo, an AI that plays the game of Go, had an [estimated](#) computing cost of \$35 million and development time of more than two years. AlphaFold, an AI to predict protein structures, has multiple version releases since 2018 (as an aside, this timeline is short compared to therapeutic discovery and development as a whole, and protein structure is but one of many non-obligatory steps to discovering and developing drugs). DeepMind, the company who developed these AIs, [reported](#) expenses of \$1.7 billion in 2021. Based on code release histories, Meta's protein structure project took at least three years, and their recent corollary to AlphaGo, CICERO, is built on the shoulders of large language models and learning AIs developed over the past few years. In the coming

years, a real possibility of less favorable macroeconomic conditions may squeeze budgets and researcher efforts in these projects. Even less likely than sustained effort on these R&D projects would then be the development of new categories of AIs that may be required to accomplish tasks described in Section II. Few companies today possess both large R&D divisions and a willingness to invest their excess returns to fundamental AI development. Among these companies, a de-emphasis on ambitious AI R&D may already be underway, as noted by Meta's intention to focus their AI efforts to improve product recommendations and increase advertising conversion rates. Even so, as computing costs decrease and access to nAIs increase, the investment and time to onboard and advance AI becomes increasingly favorable to smaller companies or even individuals. This tailwind to AGI timelines offsets the historical dependency on large private technology companies to develop innovative AI.

The incentives to invest and innovate in industry follow cycles of opportunity, growth, maturity, and decay. Displacement often follows decay, but the decay of technologies and their supporting service tasks occurs slowly. Just as hardware stores keep esoteric spare parts of decades-old appliances, so too may automobile mechanics long maintain inventory and repair services for the parts and ICE vehicles subject to displacement by EVs. Servicing 1.2 billion ICEs could be automated, but with a closing window of opportunity, is there sufficient incentive to automate these tasks? With simpler designs than ICEs and much greater investment, 2043 is a reasonable timeline for highly automated maintenance and repair of EVs. However, the incentive to automate other analogous tasks and industries lag EVs, and the vector of change in other industries is less obvious than for the shift from ICEs to EVs. Beyond its development, the deployment of AI will be tethered to industry-specific change, and the natural lifecycles of industry may create a strong headwind to AGI timelines.

Two major phenomena in modern globalized economies are the relocation of work to areas of cheaper labor and the development of comparative advantages by regions or nations to produce certain goods or services. In recent decades, these phenomena often manifest by establishment of new supply chains, training of labor, and scale-up of the nascent industry in a developing nation. These steps take at least years, and sometimes entire generations, as it operates under constraints of both individuals, who have at most a few careers in their lifetime, and industrial policy, typically only as dynamic as the rate of political or regime change. Singapore's transformation began 1960's, grounding its economy in manufacturing, then achieving its high standing in finance and technology by the 1990s. Other nations, including Vietnam, are in earlier stages of this progression now, as it expands beyond food and textiles production and into electronics. Many nations are even earlier in this development cycle, including African countries still with a majority agricultural economy. The relocation of tasks to cheaper labor in developing nations is far from exhausted. Without a significant perturbation to the typical course of task relocation and ensuing industrial development in developing nations, these economic phenomena may pose a large headwind to AGI deployment for decades. Even many tasks today which cannot be outsourced, including in developed nations with expensive labor, still are not automated. For example, the United States is mired in a long lag from building advanced robotic and intelligent

systems to implementing fast food preparation nAIs. One may argue globalization and outsourcing has peaked, which may force domestic industries to accelerate adoption or development of innovative technologies. These behaviors would break with the otherwise large momentum of human-oriented work spreading across a globalized economy.

III.B Labor's force, state intervention, and utility

Loss of purpose, means to support oneself or one's family, or economic power to buy and sell goods or services, often has negative impacts, from individuals up to nations. These concerns garner increasing interest from governments to support labor markets, and laborers themselves perceive a growing threat of AGI to job security. I argue the response by labor, protections of labor, and behavioral changes from altered pricing of goods, services, and wage expectations will retard AGI development and deployment.

There is an idea that the advent of new technologies will leave nothing left for humans to do. This idea is repeatedly false historically, exemplified by the flow of labor down agriculture, industrial, and service sectors. That this notion is false to date does not mean it will be false with the advent AGI, nor that AGI cannot occur even with new tasks for humans. The time to develop new AIs for these new tasks will likely hasten, and future tasks will also likely be co-developed by AIs or with AI embedded into the design. Recently in wealthier societies, however, the proliferation of businesses such as barber shops and tattoo parlors also bring services to those services, for example dedicated furniture suppliers to each of those businesses. So long as humans can create niche service jobs out of existing ones, the development of and competition by AI to these tasks will lag, in part due to lack of incentives and the fragmented nature of "cottage" industries. Another headwind to AGI is derived from the observed outward expansions of pockets of skilled professions, such as lawyers specializing in cybersecurity, blockchain, autonomous vehicle liabilities, and AI property rights. These efforts may ultimately be lost to AI, but as with AI timelines tethered to the industry in which it operates, one may also predict natural cycles of education, training, and degree specialties to factor into AI timelines. A tailwind to AI upending intellectual professions could be new skilled labor with both domain knowledge of the industry and requisite knowledge to develop and deploy new AI, which is only just underway in some high-skill, non-service sectors.

In addition to the expansion of services, another potential response by labor is if automation were to ripple across segments of the labor market, labor would spill into other sectors and lead to increased labor competition. This outcome is plausible with increased automation of manufacturing jobs, possibly evidenced by increased structural unemployment despite robust growth of service jobs. Most new work humans create will likely then be more services or intellectual jobs, though the educational demands of intellectual work will be out of reach for many. As a result, there may be a race to the bottom in other jobs, producing downward wage pressures and sustained competition with AI, slowing its widespread deployment. Counterintuitively and conversely, a downward wage pressure may increase investment in automation and AI, which is otherwise difficult when labor is tight and wages

high. Reduced net profits and falling investment would only follow an exhaustion of excess labor, reaching something like a Lewis Turning Point (in this context it does not necessarily mean a full absorption of agricultural labor by manufacturing, but applied to industries agnostically). The possible tailwind here would be a positive reinforcement loop where AI creates “excess labor” and thus frees investment and capital to accelerate AI development.

Economic growth alters expectations of work, wages, and standards of living. These new expectations are highly sticky, and turbulence to them, at least since the Industrial Revolution, correlates with unrest and political change. Increasingly since the Great Depression and the New Deal, the threat of economic downturns, pandemics, and job offshoring to social stability prompts governments and central banks in developed nations to inject labor safeguards and interventionist policies. Most nations where AI would be realistically developed and deployed in the near future, including the Americas, Europe, and Asia, largely subscribe to some tenants of Austrian economic thinking, which espouses the value of stable money and freedoms to transact as basic rights of individuals. Even the threat of AGI to labor may prompt action, possibly grounded in a novel moral equity problem, where rather than something being too expensive for the poor (e.g. healthcare in the United States), there will be an erosion of individuals’ livelihoods and economic freedoms. There are a few actions governments may take in such a scenario. They may expand their role to support the un- or under-employed with opportunities for re-skilling, increased welfare, or expanded public works. These strategies may impact AGI timelines, as public spending may crowd out private sector investments in new technologies. Also likely and impactful to AGI timelines are taxations on AI-enabled profits, restrictions on sales of goods or services powered by AI, or a push for AI as a public good. Designation of public goods and taxation often reduces incentives to innovative, so while this scenario would not preclude AGI development, it may prolong its timeline. Governments have intentionally blocked development of new technologies throughout history, but many cannot or will not today, especially with ever more accessible knowledge and usage of AI. Governments may instead attempt to coerce developers and owners of AI to open their technologies to wider use or competition, or take legal actions such as anti-trust, if owners of AI were to develop monopolies in their industries. However, without clear information on societal impact and property rights, these actions are unlikely to succeed. Given the potential reach of AI’s impact to labor and economies, a multi-state, global effort may be required, which is also unlikely to happen quickly. While failure would not accelerate AGI, it would dampen government-induced headwinds on AGI timelines. There could be a long delay to action while nAIs continue to be developed, altering expectations of the futures of workers and consumers.

Three observations and predictions relating to consumption behaviors collectively pose a headwind to AGI development. First, the permanent income hypothesis states consumption patterns emerge from future expectations of income and economic changes, and changes in human capital, property, and assets influence these expectations. One could envision altered consumption behaviors due to job displacement or merely the threat of displacement. A downward shift in consumption expectations would have a large negative impact on the

economy and delay AGI. Second, price and wage stickiness could produce an inertia of consumption by those not yet impacted by AGI may continue to support higher prices. The case for increased marginal utility may strengthen as humans increasingly consume goods and services tied to branding and personal values. In this scenario, humans may sustain demand for goods or services produced by humans, even if AI-produced alternatives are cheaper. While this behavior would not sustain current levels of employment, it may prolong a qualifying event of AGI's impact to labor. Third, without significant changes to current demands for AI, development may be continually pulled towards strengthening existing nAIs. This effect is evident today with a focus on AI as information aggregators, composers of text and images, and software developers. These narrow demands may hinder the necessary development of new AIs outlined in Section II. Of these three points, adjusted consumption behaviors and opportunity costs to develop new nAIs pose the largest headwinds to AGI timelines. An example of the limitations on the marginal utility argument is how language models are nearly capable of writing novels today. Only small populations of consumers, perhaps the intellectual, urban, and wealthy, would consider a purchasing decision against AI-created books. It is unlikely that an antithetical relationship to AI-produced goods and services would compete against the demands of the masses.

IV. Examples and odds updates for $\Pr(\text{AGI}_{2043})$

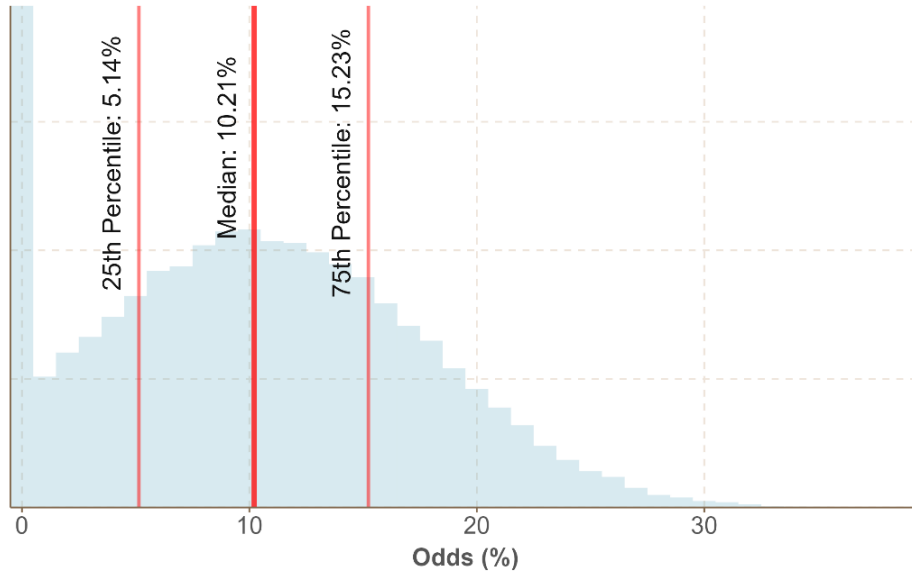
Scenarios are formed from arguments throughout the text and given a subjective odds of occurrence and relative weight of importance for $\Pr(\text{AGI}_{2043})$. Scenarios may positively or negatively impact $\Pr(\text{AGI}_{2043})$. The formula for the final odds:

$$\Pr(\text{AGI}_{2043}) = \sum \pm \text{odds} \left(\frac{\text{weight}}{\sum \text{weights}} \right)$$

A range of overall estimates were formed by randomly sampling an odd and weight for each scenario over a uniform distribution of 0.5-2x the baseline odd and weight, bound by 0%-100% and 0-1 respectively. For instance, the 1st scenario's odds below would be uniformly sampled between 37.5-100%, with weights 0.3-1.0. 100,000 samples were drawn to create a distribution of $\Pr(\text{AGI}_{2043})$.

Scenario	Sign	Odds (%)	Weight
Development and practical use of new paradigms or forms of AI	↑	75	0.6
Consensus performance metrics developed for AI, either broadly applicable to AGI or for ten developed individual nAIs	↑	15	0.1
Digital domain of Cooking AI	↑	95	0.05
Physical manipulation of objects for Cooking AI	↑	58.3	0.4
Digital domain of Haircutting AI	↑	20	0.3
Physical manipulation of objects for Haircutting AI	↑	10	0.1

Digital domain of Auto Repair AI	↑	32.5	0.35
Physical manipulation of objects for Auto Repair AI	↑	15	0.2
Design for Construction AI	↑	85	0.2
Planning for Construction AI	↑	36	0.3
Physical manipulation of objects for Construction AI	↑	10	0.25
Digital domain of Healthcare AI	↑	67	0.3
Digital domain of Contracts AI	↑	95	0.1
Forming and operating businesses	↑	10.5	0.2
Periods of large unfavorable macroeconomic conditions or continued secular stagnation over the next 20 years	↓	36	0.5
Diversification of fundamental AI R&D projects	↑	80	0.25
Investment and private sector growth in automating fading technologies and services, including internal combustion engine vehicle maintenance and repair	↑	1	0.35
Continuation of outsourcing to cheaper labor and economic growth trajectories of yet agricultural or low-skill manufacturing regions	↓	88	0.7
More nAIs are required for AGI	↑	16.8	0.6
High skill white-collar professions continue to expand pockets of tasks untouched by AI or created in response to AI	↓	70	0.15
Structural unemployment due to AI crease excess returns to increase investment and capital towards more AI	↑	30	0.25
Government interventions of expanded state, welfare, or employment mandates crowds out private sector or weakens state and broad economic conditions with excess debt	↓	18	0.25
Government taxation on AI-enabled profits or implementing restrictions on sales of goods or services powered by AI	↓	60	0.2
AI as a public good and reducing innovation incentives	↓	5	0.05
Legal action on AI ownership or AI monopolies	↓	40	0.1
Downward consumption expectations drag economy at large	↓	20	0.25
Sustained consumption of human-derived goods which offsets demand for AI products	↓	40	0.1
Opportunity cost in strengthening existing nAIs off current demand over developing new nAIs	↓	65	0.3



Low, median, and high Pr(AGI₂₀₄₃) within this Work are 5.14%, 10.21%, and 15.23%.

The Base Rate of 24.8% was combined with these low, median, and high estimates to produce the final Pr(AGI₂₀₄₃). Weights of 0.25 and 0.75 were chosen for the Base Rate and probabilities from this Work, respectively, as the Base Rate was derived from prior works focused on technical development, and this Work considers some technical but mostly broader non-technical considerations AGI timelines.

Final odds:

$$\text{Median} = (24.8\% * 0.25) + (10.21\% * 0.75) = \mathbf{13.86\%}$$

$$\text{Low} = 5.14\% * \frac{13.86\%}{10.21\%} = \mathbf{6.98\%}$$

$$\text{High} = 15.23\% * \frac{13.86\%}{10.21\%} = \mathbf{20.67\%}$$

Code for figures and analyses in this Work at <https://github.com/srhoades10/aiworldviews>

Appendix

Scenario: Development and practical use of new paradigms or forms of AI

Reference section: II.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 75%

Weight: 0.6

Description: Reinforcement learning (RL) is one qualifying instance of a new AI developed within the relevant times for this work. RL greatly matured in the past 20 years, and its theoretical basis originated approximately 25 years ago. Thus, as the number of researchers and interest in this field grows, to introduce a new AI of similar significance within the next 10 years and deploy it by 2043 is given a high odd of success. Composability AI could qualify as a new paradigm of AI. This Scenario is given a moderate weight, as existing nAIs or new nAIs rooted in modeling and learning AI frameworks may be sufficient for AGI. However, any new AIs which need significantly less data for training or can learn fundamental rules of systems, e.g. laws of physics, would be highly generalizable and greatly increase odds of AGI.

Scenario: Consensus performance metrics developed for AI, either broadly applicable to AGI or for ten developed individual nAIs

Reference section: II.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 15%

Weight: 0.1

Description: Classification performance on a standardized dataset of images is an example of one consensus metric for one nAI. Performance measures can be found as early as 2010, which improved rapidly after the implementation of deep neural networks in 2012 and later surpassed human performance in 2015. The number of parameters in these models roughly correlate with performance, but the introduction of deep learning and more complex neural network architectures achieved greater performance than simply increasing the parameters in simpler architectures. Given this history, I view performance metrics on a nAI-by-nAI basis are more likely than a generalizable, singular measure for AGI. In this Scenario, I arbitrarily choose ten nAIs as an approximation to achieve AGI, given there may debatably be five today, and these are not likely extensible to accomplishing all tasks required for AGI. If the assembly and release of a standard datasets for other nAIs takes two years, performance measures defined and agreed upon in less than one year, and success achieved in half the time of image classification (2.5 years), that leaves roughly five years per nAI. With an assumption that all ten nAIs will be developed by 2033, for all ten nAIs to achieve consensus metrics and high performance by 2043, I expect the odds to accrue requisite benchmarkable datasets by 2037 for purely digital nAIs at 95%, and for nAIs operating in the physical domain, 60%. If seven nAIs are digital, and three physical, then the total odds are $95\%^7 * 60\%^3 = 15\%$. The weight for this Scenario is low, as performance metrics are not necessary to develop nAI, as seen with large language models today.

Scenario: Digital domain of Cooking AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 95%

Weight: 0.05

Description: The digital tasks for cooking, including retrieval of recipes and forming novel recipes, with temperatures, cook times, and ingredients, are all nearly possible already today. I see potential challenges related to ingredients: which can be substituted for one another, which are available at a given moment or in a given locality, and how “effective” are combinations of ingredients, with a learning capability baked into the formation of novel recipes. This Scenario has a very low weight given its feasibility and success would not greatly alter odds of broader AI accomplishments.

Scenario: Physical manipulation of objects for Cooking AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 58.3%

Weight: 0.4

Description: Nuanced preparation tasks such as cracking and manipulating eggs in various states of viscosity will be difficult but not essential to cooking AI. Other challenging preparative tasks include peeling carrots and fine chopping of herbs. I posit odds of these tasks by 2043 at 30%. Flipping patties, robotic food picking (off plants or shelves, e.g.), and rough chopping are already feasible for many foods, and I place odds to achieve these more generalized cooking tasks at 80%. Sorting occurs in food processing factory settings today and could be scaled down to a small kitchen environment for dry goods separation and dispensing into cooking vessels, which I put at 65% odds. If these three odds are evenly weighed in their significance to cooking AI, and not all need to occur, then the overall odds are $(80\%+65\%+30\%)/3=58.3\%$. An incentivized environment to develop AI for these tasks is baked into the estimates, as there is already effort today to automate these tasks. A higher weight given than for other physical tasks such as Haircutting AI as success would be a big advancement in nuanced robotic maneuvers and applicable to many other physical tasks.

Scenario: Digital domain of Haircutting AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 20%

Weight: 0.3

Description: There is no “corpus” of information like there is for online cooking recipes, and raw data on haircuts not only does not exist, but readily-generated data today may not be suitable for an AI model. For instance, video capture of haircuts would need to be translated to quantifiable actions performed in 3-dimensional space, including angles and forces of combs, scissors, and razors. Alternatively, a physical simulation environment could train robotic systems to perform these actions, effectively generating as much data as needed. Assuming development is required by 2037 to then deploy haircutting AI over a 5-year span, incentives notwithstanding, I view capturing videos of haircuts *en masse* or generating data from a physical simulation environment to be more feasible than not. However, I see lack of incentives, investment, and effort to be a major drag on developing Haircutting AI, digitally and physically. There, I assign low odds of success, but with a moderate weight, as success in

framing a problem of physical tasks into the digital domain for AI modeling would be applicable to many other tasks as well.

Scenario: Physical manipulation of objects for Haircutting AI

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 10%

Weight: 0.1

Description: The physical manipulation of objects, including haircutting equipment and human hair, is at least as difficult as the fine chopping or peeling tasks for Cooking AI, which I give odds of 30%. As with the digital domain of haircutting tasks, lack of incentives and investment further drag down these already low odds. Nonetheless, these tasks are given a low weight as the extremely fine and precise nature of these physical tasks are unnecessary for many other tasks and therefore less broadly applicable.

Scenario: Digital domain of Auto Repair AI

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 32.5%

Weight: 0.35

Description: Digital subtasks include cataloging manuals, parts, and tracking inventory for most major car models, including internal combustion engines. These tasks are highly analogous to information aggregation tasks solved in Cooking AI and given 80% odds of success (lower than digital Cooking AI tasks due to less existing work and reduced incentives). Analogous to medical diagnosis, a diagnostic system which integrates domain knowledge could occur through purely visual information (e.g. a car “X-ray”) or new diagnostics hardware. The odds of diagnostic tasks are 50%. Whether through Composability AI or other new nAIs, even a rudimentary understanding of the relationship of parts would connect the domain knowledge and diagnostic information to physical tasks of automobile repair. Alternatively, this challenge could be met with an abundance of data on cars in broken and fixed states to then propose repair actions, but I view the former route as more feasible and assign odds of 25%. These three primary categories of subtasks are given a roughly equal weight, except for a lower weight to the composability challenge, as understanding “why” is not essential in other AI tasks today. I weigh the combined probabilities of the first two subtasks, viewed as essential, with the third subtask to form final odds of $((80\% * 50\%) + 25\%) / 2 = 32.5\%$. A similar weight is given to digital Haircutting AI, as these solutions can generalize to many other tasks.

Scenario: Physical manipulation of objects for Auto Repair AI

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 15%

Weight: 0.2

Description: The maneuvers needed for automobile repair are more similar to haircuts than food preparation. When considering vehicles damaged from accidents (assuming significant numbers of humans will still drive vehicles for most of the next 20 years), Auto Repair AI

may require comparable levels of finesse and precision as Haircutting AI, while also bearing much heavier loads and applying greater forces. However, the automobile industry may still be significantly impacted by automation of routine maintenance such as tire and brakes replacement. The odds of performing the more challenging tasks are 10%, equivalent to haircutting AI given it is more difficult but with greater incentives. I place odds of routine maintenance tasks at 30%, but in giving a slightly lower weight to these tasks, produce an overall odds of 15%. A lower weight is given here as success, while powerful, is an excessive capability of fine manipulation of heavy objects for other tasks.

Scenario: Design for Construction AI

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 85%

Weight: 0.2

Description: Building design including floorplans, heating/cooling, plumbing, and electrical wirings that are already largely feasible through software, and generative AI models trained on a corpus of blueprints could propose a near-infinite number of basic designs. I assign odds of these tasks at 95%. However, placing designs in real 3-dimensional space is more complicated. The terrain, soil, and other environmental considerations are constraints on which designs could realistically be built on a given plot of land. Designs may also require matching the building materials to its function, under other constraints such as material suitability, availability, and cost. Most of these environmental factors could be estimated with either existing knowledge of the site location or imaging data. I assign the odds of real-world “aware” building design at 80%. Either of these groups of tasks may be sufficient for Construction Design AI, but in assigning a slightly higher weight to the more challenging tasks, I assign final odds of 85%. I give a modest weight here, as success would represent a broadly useful advancement in document and image generation (for example, this AI would estimate the volume and object depth in an image and photoshop new objects into a 3D space projected from the 2D image).

Scenario: Planning for Construction AI

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 36%

Weight: 0.3

Description: Planning requires knowledge of permitting, building codes, zoning, and regulations, which are mostly hyperlocal to the region of construction. While technical elements of these tasks are mostly feasible today, a lack of information standards and changing codes, zoning, and regulations across towns and states over time pose a challenge of information aggregation. Document creation is feasible, and there exist numerous public documents and records related to construction planning that could be used to train generative AI models. However, these documents are highly disaggregated, along with the construction industry generally, and would require a large upfront cost of domain knowledge engineering to train a permit creation AI, factoring local codes and regulations and with awareness of changes over time. Multiplicative odds of document creation subtasks at 90% and codified domain knowledge at 40% produce a final odds of 36%. A modest weight

is given due to the size of the construction industry and the utility of AI with engineered domain knowledge and abilities to update with changing rules and regulations.

Scenario: Physical manipulation of objects for Construction AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 10%

Weight: 0.25

Description: The combination of power, finesse, and versatility required of physical tasks in Construction AI make it no likelier than Auto Repair AI. I view Construction and Auto Repair industries to contain comparable degrees of incentives and challenges to develop and adopt new technology, in part due to their fragmented nature. Relative to Auto Repair AI, I downweigh odds of physical Construction AI slightly as there have been unsuccessful efforts in robotic construction. Physical Construction AI receives a slightly higher weight than Auto Repair due to the size of the construction industry, and success would be a tremendous advance in physical manipulation of objects generally.

Scenario: Digital domain of Healthcare AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 67%

Weight: 0.3

Description: Much of healthcare includes maintaining or altering human physiology, all grounded in some understanding of biology. Significant new discoveries of fundamental biology and their applications as novel therapeutics by AI are unlikely in the near future. However, other tasks to recommend standards of care, exploit existing domain knowledge and patient information to diagnostic or therapeutic ends, or perform administrative roles in healthcare, are all significant and qualifying events for digital Healthcare AI. Based purely on feasibility, I place odds of these sets of tasks are 90%, 40%, and 90%, respectively. Healthcare, especially in developed nations, is fettered by outside interests, and as a source of job growth and destination for very high-skill labor, will experience resistance to implementations of AI. I thus downweigh the odds, particularly for administrative tasks, to 85%, 35%, and 80%. If these three odds are evenly weighed in their significance to digital Healthcare AI, and not all need to occur, then the overall odds are $(85\%+35\%+80\%)/3=67\%$. A modest weight is assigned based on the significance of the industry, although success may not be as broadly applicable because of its idiosyncratic nature.

Scenario: Digital domain of Contracts AI

Reference section: II.C

Impact to Prior Pr(AGI₂₀₄₃): Increase

Odds: 95%

Weight: 0.1

Description: Document creation for many standard business and legal contracts is already possible. Considering the recent advent of smart contracts, and evolving disputes on AI usage and ownership, the odds here are less than 100%. Even so, digital Contracts AI is feasible and will be increasingly utilized. Automating the creation of most common contracts would

be sufficient for a qualifying nAI event. A low weight is given as the impact may be a productivity boost on a small number of jobs.

Scenario: Forming and operating businesses

Reference section: II.C

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 10.5%

Weight: 0.2

Description: The subtasks of negotiating, decision-making, and planning based on environmental stimuli are essential for a qualifying nAI event. AI is already capable of forming negotiating and bargaining statements, though they possess little contextual understanding of their conversations with humans. One common business scenario that requires negotiating is capacity planning and leasing agreements on property, office space, or online domains. A fabricated negotiation from a large language model is already possible, however I put odds of a real-world and human-interactive AI negotiator at 70%. Reinforcement learning (RL) is a form of AI designed to receive and respond to environmental inputs, and a likely candidate as a responder to market, competitor, and consumer behaviors. RL models require abundant information on actions and outcomes, or the ability to simulate them. For certain narrow cases of businesses operating purely digitally, both in its functions and its core product or service, RL may be able to gauge product interest, pricing power, and consumer feedback to propose changes in product lines or prices. However, incomplete information about market changes and competitor behaviors is innate to competitive economic systems. Assuming these systems persist, AI will have to rely on large engineering efforts of domain knowledge, including the actions taken by previous companies that led to their demise, or a new AI able to perform inference and propose actions with highly incomplete information. Notably, the set of possible actions that such an AI needs to consider here will be orders of magnitude greater than for chess or Go, and may itself require substantial human engineering. I view the odds of a successful domain knowledge engineering effort or a new AI capable of superhuman inference with sparse information at 10 and 20%, respectively. If these probabilities are weighted evenly, where either can occur, then the final odds for business AI is $70\% * (10\% + 20\%) / 2 = 10.5\%$. A modest weight is given as success would be a monumental achievement towards AGI but not essential, and it is more likely nAIs for certain tasks within a business (sales, product design, marketing) will occur and potentially qualify as AGI.

Scenario: Periods of large unfavorable macroeconomic conditions or continued secular stagnation over the next 20 years

Reference section: III.A

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 36%

Weight: 0.5

Description: If the time since the 1970s is at all predictive for the next 20 years, there will be at least one mild global recession, and a small probability of a severe recession or depression. I posit a mild recession will have little impact on AGI timelines, though a larger recession will delay the development of innovative new technologies and its deployment into industry, in part through an environment of risk aversion and low capital investment. While

the narrative of secular stagnation may be overplayed, as economic and productivity indicators does not capture digital services and software as accurately as physical goods, there is solid evidence for declining rates of growth in the United States, Europe, and parts of Asia. I view it more likely than not these trends will continue, and as other countries develop, the “spreading effect” of secular stagnation in a globalized economy will produce an asymptote in their rate of individual economic and technological. A transformational change such as AGI to occur in this environment is still possible, as evidenced by the advent of the internet in the 1990s, and evidence of some disruption to this stagnation during the coronavirus pandemic. The odds in this scenario are viewed as a weighted additive aggregation of possible events which decrease $\Pr(\text{AGI}_{2043})$. The odds of a small recession are 90% but with a small weight on AGI development relative to a large recession at 20%. I grant 50% odds of a continuation and spreading of secular stagnation that would materially impede the innovation and growth necessary to achieve AGI. Combining these scenarios produces $(90\% \cdot 0.1) + (20\% \cdot 0.6) + (50\% \cdot 0.3) = 36\%$. A strong weight is given here as these scenarios impact AI technologies broadly.

Scenario: Diversification of fundamental AI R&D projects

Reference section: III.A

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 80%

Weight: 0.25

Description: Large technology companies and research groups supported by such companies have produced most applied AI R&D. These projects are focused on classification and generation of text, speech, and visual content, as well as playing games, healthcare diagnostics, and prediction of protein structure. The applications of modeling and generating text, speech, and images will surely continue to grow, however simply strengthening and diversifying these nAIs may not be able to perform many other tasks mentioned throughout this Work. For instance, improved diagnostic capabilities in healthcare and protein structure modeling in biotech on their own will unlikely become a qualifying event for AGI. If the time to identify a new R&D project, locate or acquire requisite data, and apply it to the problem statement takes four years, and only FAANG companies perform these functions, who have at most a handful of such projects, then one may expect a few dozen total AI projects in the next 20 years. If companies compete on roughly half these projects, then there may be collectively 10-20 new and unique nAIs. However, increased accessibility and decreased cost of AI will yield diversified projects from outside large technology companies. Odds are downweighed from certainty due to the category of projects technology companies are willing to address, as they are likely to exploit purely digital AI projects more fully before exploring the physical task examples in this Work. A modest weight is granted as the broadening of AIs is more important for AGI than strengthening existing nAIs.

Scenario: Investment and private sector growth in automating fading technologies and services, including internal combustion engine vehicle maintenance and repair

Reference section: III.A

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 1%

Weight: 0.35

Description: While limited, one prime example serves as a basis for this scenario analysis. Tesla Motors began operations in 2003, requiring roughly thirteen years for an announcement for a full self-driving capability in 2016 and fifteen years to reach significant market share in the car industry. Twenty years after Tesla's founding, full self-driving appears imminent but is still not fully deployed. Self-driving cars and electric vehicles are not equivalent, however most work on self-driving is associated with Tesla (noting that other car-agnostic self-driving technologies exist and have been in development since 2016 with limited adoption). Tesla Motors has also not developed an additional nAI apropos for this work, namely automated vehicle repair or maintenance, nor is it a significant initiative at the company. In 2023, there are a dozen other EV manufacturers with a significant market size. Considering the capital allocation by investors and car companies towards EV production, and only Tesla has attempted even mechanized battery exchanges (largely unsuccessfully), one could predict the time from announcement to development of automated self-driving or vehicle maintenance technologies to take upwards of ten years, and deployment into industry and broader impact on services ten more years. Focusing narrowly on vehicle maintenance and repair, I place odds of any car company accomplishing these tasks at 10%, and place ICE repair at comparable odds. More broadly, I optimistically posit even two sets of tasks or large industries with "retroactive" investment and innovation on fading industries or technologies are sufficient to meaningfully contribute to qualifying AGI. If so, then the total odds would be $10\% \times 10\% = 1\%$. A moderate weight is placed on this scenario as service jobs may linger or expand for decades without active investment and technological development.

Scenario: Continuation of outsourcing to cheaper labor and economic growth trajectories of yet agricultural or low-skill manufacturing regions

Reference section: III.A

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

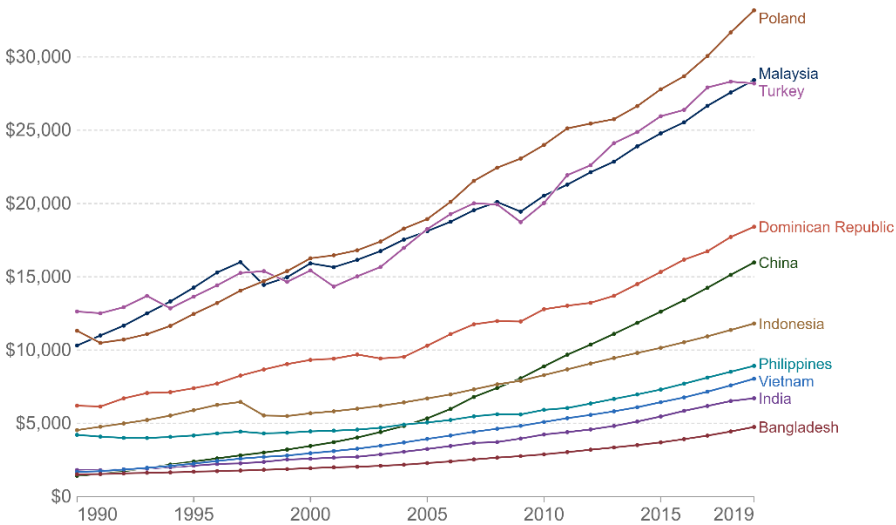
Odds: 88%

Weight: 0.7

Description: A break from the phenomena of relocation of work to areas of cheaper labor and industrial planning and development by developing nations may occur in a few ways. One could be the exhaustion of excess labor. Bangladesh, the 8th most populous country, serves as an example of how unlikely exhaustion will occur within 20 years. Bangladesh introduced an industrial capacity and economic policy initiative around 1980. Today, clothing represents [87% of Bangladesh exports](#). Vietnam and India are examples of other populous countries with similar economic trajectories, though moving quicker to electronics, pharmaceutical, and automobile manufacturing. Even so, India has a low GDP per capita, alongside Pakistan, Bangladesh, and Nigeria, which collectively consist of over 2 billion people. The plot below depicts GDP per capita growth of select countries since 1990.

GDP per capita

This data is adjusted for inflation and for differences in the cost of living between countries.



Source: Data compiled from multiple sources by World Bank
Note: This data is expressed in international-\$ at 2017 prices.

OurWorldInData.org/economic-growth • CC BY

https://noahpinion.substack.com/p/can-india-industrialize?publication_id=35345&post_id=101082915

Another mechanism would be unprecedented economic hypergrowth across multiple large developing nations. A noticeable upward slope for GDP per capita in China began in 2005, after a period of economic reforms originating in the 1990s. In this light, in the 30 years from those reforms to today, China's growth could be considered successful, Bangladesh less so, and Singapore the "best-case" scenario. It is then extremely unlikely growth equivalent to or exceeding Singapore would occur for multiple populous and developing nations in less than twenty years. Another mechanism is decreased globalization and onshoring in developed nations. However, onshoring may only be viable with efficiency and productivity growth from new technologies or cheaper domestic labor. Whether by job automation or reduced wages, such industrial policies will be unpopular with voters and politicians. If these processes do occur, their occurrence within twenty years is unlikely. I give odds of 85%, 99% and 80% that there will be sustained excess labor, developing nations with over two billion citizens will not undergo an unprecedented hypergrowth, and industrial policies leading to job automation or wage suppression will not be palatable and thus not widely implemented in developed nations, respectively. The average odds are 88%. A large weight is given as these trends have been a significant influence in technological development, economic growth, and the global world order since the Industrial Revolution.

Scenario: More nAIs are required for AGI

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 16.8%

Weight: 0.6

Description: I posit five new nAIs are needed for AGI. Based on the histories of nAIs developed to date, each would cost at least billions of US dollars and take at least three years.

I place the odds of development and deployment of these new nAIs within 20 years at 80%. I also posit AGI will require integration of multiple nAIs (Construction AI, e.g.), and place 80% odds of technical feasibility to integrate multiple nAIs within two years of their development. While technically feasible, these integrations may require one agency, company, or state to either develop nAIs or have existing nAIs made available or purchasable. The odds for the proper “environment” of nAI development and assembly, over a span of five years from the development of all requisite nAIs, is 25%. These collectively form an 11-year horizon with $80\% \times 80\% \times 25\% = 16\%$ odds. However, this assumes that the new nAIs are the *right* ones. For instance, nAIs today do not operate in the physical world, nor possess abstract senses of real-world agency and strategy. I view the next five nAIs being the right nAIs for AGI to be low. If five nAIs are sufficient for AGI, but twenty total are needed to be developed first, I grant 60% odds that twenty will be built in 15 years. With the same odds and timeline for technical integration, and a boost in the probability of integration to 35% in five years, the 20-year odds are at $60\% \times 80\% \times 35\% = 16.8\%$. The weight is moderately high, as AGI may be possible with significant extensibility of existing nAIs today, or AGI may arise through other means than integration of multiple nAIs.

Scenario: High skill white-collar professions continue to expand pockets of tasks untouched by AI or created in response to AI

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 70%

Weight: 0.15

Description: Considering the formation of services jobs off the innovations of the Industrial Revolution and the rise of the managerial class during the 20th century, coupled with the motives of high-skill labor for gainful employment and differentiation of their products or services, this scenario is more likely than not. However, the remaining slivers of the professions untouched by AI, for example the specialties within law and medicine, will only be able to expand so widely in the future. For medical professionals, the diagnoses, treatments, and procedures serviced by AI will leave fewer services to offer by humans. One area that may continue to expand for the foreseeable future is research and engineering, as new ideas may be readily produced by AI, but the means to research and answer new questions is less likely to be solved by existing nAIs. Thus, while this scenario is more likely than not, the increasing automation to tasks in legal, medicine, design, and other advanced-degree professions will erode their headcount and influence over time. The weight for this scenario is modest as the candidate professions represent a fraction of total labor, and most individuals cannot attain this level of work for various reasons.

Scenario: Structural unemployment due to AI crease excess returns to increase investment and capital towards more AI

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Increase

Odds: 30%

Weight: 0.25

Description: Structural unemployment due to automation is obvious conceptually, but there is much uncertainty over how much has occurred over the past few decades. Given the long period over which structural unemployment manifests, I give odds of 40% that significant skill mismatches in the labor market impact a majority of the labor force within 20 years. Assuming that structural unemployment is due to AI, and that it translates to higher productivity with lower operating expenditures and thus produces excess returns, I give 75% odds of appreciable investment in the development of new technologies which hasten AGI timelines. I downweigh these odds slightly as extra cash is often used to improve balance sheets or to reward investors and shareholders. This latter action is of increasing frequency and volume over the past twenty years, and this trend is likely to continue in the near future. The collective odds are $40\% \times 75\% = 30\%$. This scenario is given a modest weight given the difficulty to observe and quantify structural unemployment, despite its importance.

Scenario: Government interventions of expanded state, welfare, or employment mandates crowds out private sector or weakens state and broad economic conditions with excess debt

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 18%

Weight: 0.25

Description: Concerns of social stability due to growing structural unemployment may prompt government interventions. Employment mandates are already present in central banks of many developed nations, though these mandates often take the form of monetary policies to stimulate the economy, rather than direct assistance such as cash transfers or basic income. These actions may be instead performed through fiscal stimulus, and if recent trends in developed nations continue, the role of fiscal policy will grow stronger relative to monetary policy. Taken too far, these policies risk defaults, currency devaluations, and broad financial instability. In addition, such spending comes at the expense of the private sector, often the primary engine of economic growth. Any of these actions could impact AGI timelines. The odds in this scenario are conditional on large structural unemployment due to AI, or a perceived threat of such, to occur within ten years (see above scenario regarding structural unemployment). I grant 60% odds of expanded stimulus and employment mandates in most G7 or large Asian countries to produce debt/GDP ratios appreciably above those of the WWII or COVID eras. The odds that, conditional on these debt levels, there will be a large global recession, significantly slowed economic growth from a weakened private sector, or a series of major currency devaluations is 30%. The combined odds are thus $60\% \times 30\% = 18\%$. A modest weight is given because while these events would be significant, this scenario is conditional on other unlikely events, and development of new technologies may still occur, or even be accelerated, in seemingly unfavorable environments.

Scenario: Government taxation on AI-enabled profits or implementing restrictions on sales of goods or services powered by AI

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 60%

Weight: 0.2

Description: Taxation may occur as a form of redistribution, possibly to fund basic income, re-skilling, or welfare, or opportunistically on companies who become exceedingly profitable from their AI technologies. Recent history suggests that for seemingly pressing matters, Pigouvian taxation, e.g. on cigarettes or carbon, and windfall taxes, e.g. on oil companies, can be proposed and implemented within years. I put the odds of taxation within 20 years, first requiring AI powerful and threatening enough for taxes to gain political popularity across a majority of G7 nations and developed Asian countries, at 60%. Even so, AI is likely to remain nebulous, distributed, and thus more difficult to tax than obvious monopolistic or oligopolistic industries, and Pigouvian taxation requires clear and quantifiable externalities, which is difficult for AI. However, while taxation on humans is difficult politically, taxation on AI is relatively palatable. Thus I maintain a 60% overall odds on this scenario. I grant a smaller weight as AI will become cheaper and widely accessible, and much of AI to date is open source, thereby dampening the impact of taxes on future AGI development.

Scenario: AI as a public good and reducing innovation incentives

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 5%

Weight: 0.05

Description: The open-source model of software development creates something close to a public good already, and individuals and companies continue to find ways to develop and monetize new technologies regardless. A push to formally assign and enforce a public good status on AI could take a long time to implement, and even so the impact to AGI timelines will likely be small, much like taxation of AI probably may not hinder new development. Therefore, a very small weight is given for this low-probability scenario.

Scenario: Legal action on AI ownership or AI monopolies

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 40%

Weight: 0.1

Description: Recent events between technology companies and governments, including exchanges over rights of developers and consumers on gaming, social media, and advertising platforms, implies technology companies will continue to encounter scrutiny over any outsized influence and market power. While governments may take non-legal, coercive actions to democratize AI or weaken technology developers, I do not grant material odds of these nebulous events as they are unlikely to impact AGI timelines. I do expect legal actions over ownership or anti-trust, possibly over consolidations of nAIs by large technology companies, as likely within 20 years. However, I downweigh these odds as the time to build cases is long and the success rate in recent lawsuits against technology companies low, and have only come to fruition under certain political landscapes. In addition, I expect a strong anti-trust case could only be made over an overwhelmingly obvious threat of monopoly from nAI consolidation, as the emergence of an actual monopoly is often not immediately obvious and less likely within 20 years. With these arguments, and that AGI development will continue irrespective of legal cases, this scenario receives a low weight.

Scenario: Downward consumption expectations drag economy at large

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 20%

Weight: 0.25

Description: Lower consumer confidence in the economy or high expectations of inflation impact consumption. While the effect size is difficult to quantify relative to many other possible factors, negative expectations can be a self-fulfilling prophecy for economic downturns. Consumers may shift to purchasing staples and avoiding frivolous expenses, many of which are services which comprise a large part of the economy in developed nations. The advent of AI would represent a new driver of consumption pessimism, whereby consumers may expect to have less purchasing power under the threat of job loss. The threat of AI to jobs is widely discussed today but is not obviously impacting expectations of consumers or employment. Thus, I view it unlikely that downward expectations would be significant enough to impact the economy broadly. I grant a smaller weight as smaller economic downturns, if they were to happen because of this pessimism, would not impact AGI timelines significantly.

Scenario: Sustained consumption of human-derived goods which offsets demand for AI products

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 40%

Weight: 0.1

Description: Given sustained demand for the hand-made and artisanal today, I find it likely that as AI-generated goods and services proliferate, there will be a response of demand for “made by humans”. While likely, this demand may only come from wealthier consumers, and apply to a small proportion of all goods and services. I downweigh the odds as the supply of human-made goods and services and number of human laborers could decrease in response to competition from AI, mitigating this demand in the process. A low weight is given as this demand will be washed out by the consumption of the masses, and consumption of human-made goods and services may not materially impact AGI timelines.

Scenario: Opportunity cost in strengthening existing nAIs off current demand over developing new nAIs

Reference section: III.B

Impact to Prior $\Pr(\text{AGI}_{2043})$: Decrease

Odds: 65%

Weight: 0.3

Description: The balance of exploiting existing nAIs and exploring new ones will be a large determining factor in the overall technological growth towards AGI. I expect the nAIs that power content creation, media, marketing, and sales will continue to be more fully exploited, while AIs for other tasks, including scientific research and many of the physical tasks mentioned in this Work, will have substantially fewer researchers and dollars invested. For example, entertainment media and advertising are heavily validated large markets serviced by large technology and retail companies, particularly in the consumer-heavy economies of

developed nations. Companies will be quicker to cut R&D on exploratory projects than sacrifice on the core product in periods of economic uncertainty and limited cash reserves. Existing nAIs also have much room for additional productization. For instance, the “base layer” of ChatGPT can be modified to diversify content offerings towards different ideological or political leanings, or genres of blogs, books, and other text-based content. There is much excess demand yet to be exploited from large language models, and I view it more likely than not that new product lines of incumbents and core products of startup companies will focus on these opportunities in the coming years. Most research on new AIs will likely continue to come out of R&D at larger companies or in academic research environments. While I view it nearly certain that the majority of work in AI will go towards strengthening existing nAIs over developing new ones over the next few years, I downweigh these odds as there are continually new uses of existing AIs, AI will become cheaper and accessible, and new AIs can now be pushed into the market and exploited within years. I grant a modest weight as existing nAIs may be sufficient for AGI and as more researchers and dollars spent on AI development will negate the opportunity cost of strengthening existing nAIs.