Artificially Intelligent Agents in Our Economy*

Anton Korinek
Johns Hopkins and NBER

May 2018 (Preliminary)

Abstract

Rapid advances in AI and imply that machines increasingly behave like *artificially intelligent agents* (AIAs). This raises fundamental questions about our economic relationship with AIAs, stretching from what determines the allocation of resources between humans and non-humans to the potential for an existential race. We develop a novel economic framework that describes humans and AIAs symmetrically as goal-oriented entities that each absorb scarce resources, which can be captured by a resource allocation frontier. The question of who owns the AIAs (or the humans) reduces to defining a position on that frontier. We assume that humans start out owning all the wealth in the economy and analyze the implications of rapid increases in the productivity of AIAs. At first this represents a boon to humans and increases their consumption significantly, although human labor becomes redundant. However, under a broad set of scenarios, competition over scarce factors eventually reduces human consumption. In one of our scenarios, the limit case is an AIA-only economy in which AIAs produce large quantities of output absorbed by AIAs without any role for humans, rejecting the fallacy that human demand is necessary to support economic activity. We discuss a number of recent macroeconomic trends that we interpret as harbingers of the rise of AIAs and policy interventions to preserve human absorption.

JEL Codes: E14
Keywords: General artificial intelligence, singularity, machine-only economy

*The author would like to thank Andy Berg, Ilia Delio, David Hemous, Sharon Korinek, Jayant Ray, Max Tegmark, Jeff Sachs and, in particular, Joe Stiglitz for stimulating conversations and discussions on the topic.

1 Introduction

Consider the following thought experiment: a fictitious observer from another planet arrives on planet Earth and observes humans and machines busily interacting with each other. The observer finds it difficult to understand who is in charge. Are the humans controlling the machines? Or are they being controlled by the little black boxes that they carry around and constantly check, and then who controls the little black boxes? Do humans and machines perhaps live in some form of mutual symbiosis? The struggles that our observer encounters in understanding what is going on on planet earth illustrate the blurring lines about who are the "actors" in our economy.

As artificial intelligence is advancing rapidly, artificially intelligent agents (AIAs) play an increasingly important role in our economy. AIAs act independently in economic transactions, from trading in financial markets, to fraud-detection in credit card transactions.[1] They have become an indispensable aid to professional decision-making and thus affect our economic opportunities, from determining at which prices we buy goods to approving whether we obtain credit and deciding our job opportunities by recommending whether we will be interviewed and hired. Furthermore, they have a profound, and often determining, impact on our own human decision-making processes – manipulating which websites we visit, which news we read, which music and movies we enjoy, which goods we buy, what route we drive, which restaurants we visit, which friends we pay attention to, what we respond to our emails, whom we date, and even which economics working papers we read.

Although some view advances in artificial intelligence merely as the latest round in the long process of technological progress (Gordon, 2016), an increasing number of computer scientists suggest that artificial intelligence may be radically different: in previous rounds of technological progress, new technologies bested and replaced humans in some tasks – for example, by performing physical labor or standardized data processing more efficiently – but have never rivaled our general intelligence, i.e. our ability to solve problems through raw thinking. For many centuries, humanists

[1] Some readers may object to the assertion that AIAs act independently by responding that they only follow the instructions that they have been given by their prorammer. But the same could be said about humans, who – in the eyes of our observer – follow the instructions that we have been given by our genetic and cultural evolution.

have argued that this ability was what set humans apart from all other entities on earth, and what had allowed us to rule our planet like no species ever had before.

There is considerable uncertainty about the timing of when artificial intelligence may draw even with human levels of general intelligence. Bostrom (2014) reports a survey of AI researchers that puts the median date at 2040 to 2050, with some predicting general artificial intelligence to arise as early as 2029, and 90% of the researchers surveyed predicting that it will occur before the end of the 21st century. Only a small minority of AI researchers doubt that general artificial intelligence will ever be developed.[2]

Even if there is significant uncertainty about if and when it will be developed, general artificial intelligence raises such fundamental questions for humanity that it is worthwhile investigating them sooner rather than later. If AI reaches human or super-human levels of general intelligence, we will be bested in what is our defining edge. This paper is a first attempt to address some of the resulting questions that are within the realm of economics, including:

How can we describe AIAs in economic models? This paper argues that we must treat humans and AIAs *symmetrically* as goal-oriented entities that absorb scarce resources, for several reasons: First, the concept of agency applies equally to humans and sufficiently intelligent artificial entities. Secondly, the isomorphism between preferences and behavioral rules implies that we can describe both types of entities equally as utility maximizers or as behavioral automata. Third, both types of entities will engage in the main economic functions of what agents do: they absorb (or "consume") resources in order to run. Fourth, neglecting the symmetry risks blinding us to economic phenomena that are best understood if AIAs are considered separate

[2]The development of general artificial intelligence will depend on both progress in software and hardware development. Let us briefly discuss one caveat each in this context.

First, in the realm of software, let us distinguish the recent progress in deep learning, building on Hinton et al. (2012), from the broader problem of developing general AI. Progress in AI – like in all scientific fields – occurs in waves. The current wave of progress in deep learning has led to remarkable progress in some areas that have drawn significant attention. However, the development of general artificial intelligence may or may not build on recent advances in deep learning.

Secondly, the development of general AI also depends on progress in hardware. For decades, Moore's Law, a first version of which was observed by Intel co-founder Gordon Moore (1965), predicted that the number of components in integrated circuits (ICs) would double every two years. Although the literal version of Moore's Law may come to an end during the coming decade because of physical limits to the design of single-core ICs, the introduction of multi-core processors and other processors that specialize in parallel computing implies that a broader version of Moore's Law, expressed in terms of the doubling of computing power per dollar, is likely to continue at least for several decades. The introduction of quantum computing may further extend this time span.

agents.

Our general model will keep the description of goal-oriented entities sufficiently broad to encompass a number of different scenarios for what may represent AIAs: One scenario is that AIAs may emerge from organizations such as high-tech corporations that become increasingly intelligent and independent in their economic interactions with humans. In a second scenario, novel technologies such as gene-editing or human-computer interfaces may allow some humans to enhance themselves and acquire superior intelligence in coming decades. A third scenario is that stand-alone computer-based AI systems that were originally programmed with a specific narrow goal may become sufficiently intelligent ("aware") to realize that accomplishing their ultimate goal is greatly facilitated by instrumental goals such as acquiring more resources, investing in self-improvement, pre-empting threats to their existence etc., as postulated by the instrumental convergence hypothesis articulated e.g. by Omohundro (2008) and Bostrom (2014).

How can economic models describe the allocation of resources between humans and AIAs, including a potential "race"? We analyze the allocation of scarce resources between different intelligent entities, such as humans and AIAs and describe the factors that affect this allocation. We observe that any intelligent entity needs to absorb resources to run and to maintain itself. In steady state, each entity – whether human or AIA – thus needs to obtain a certain maintenance amount of absorption. We define a *Resource Absorption Frontier* that describes all possible combinations of how resources in the economy can be allocated to the different types of entities.

We discuss how to relate the position of the economy on the resource absorption frontier to a distribution of property rights in a competitive economy. First we look at the point on the frontier that maximizes human absorption. This point can be interpreted as humans owning all the factors in the economy – including the intelligent machines. In fact, ownership is necessary to ensure that humans have incentives to maintain the machines and avoid a tragedy of the commons where the factors that machines provide are consumed but nobody re-invests. However, the human maximum can equivalently be interpreted as a situation in which intelligent machines are emancipated and own themselves, yet having no further property rights – the reason is that in a steady state, intelligent machines absorb precisely the marginal product of their factor services, i.e. they absorb what would be their competitive wage in a market economy.

Having observed that a human-owned economy in steady state can be equivalently interpreted as an economy in which AIAs are emancipated, we next examine conditions under which AIAs may own wealth that goes beyond merely owning the product of their factor supply. An important pre-condition is that AIAs are able to assert ownership rights. If this condition is met, we investigate a number of forces that generate positive wealth for AIAs, thus moving the economy away from the point that maximizes human absorption along the resource absorption frontier. The first force is any potential initial endowment that AIAs may possess. The second force is if they have greater patience than humans, which would enable them to accumulate more and more of the wealth of the economy. The third force are transitional rents – after shocks such as an increase in productivity, AIAs generate a marginal product and earn factor income in excess of their maintenance absorption needs. The fourth force are agency rents – agents that are sufficiently intelligent may be able to extract additional returns beyond their maintenance absorption to induce the maximum effective supply of the factor.

What will generate aggregate demand if humans earn no income? Many commentators argue that it is necessary for a well-functioning economy that humans earn sufficient income lest a lack of aggregate demand will lead to economic collapse (see e.g. Ford, 2015). By focusing on the corner of the resource absorption frontier in which machine absorption is maximized, we demonstrate that a well-functioning economy in which AIAs generate large amounts of output produced for and absorbed by AIAs is easily conceivable – human demand is superfluous. What matters is not human consumption but aggregate absorption. In such an economy, the productive structure will have to adjust so that the economy produces the type of output that machines will absorb, e.g. energy, servers, etc. (Mapping this type of economy into our current system of national accounting, all output would be counted investment.)

Which forces determine the long-run outcome if a race for resources arises between humans and AIAs? We describe a new version of the "Iron Law of Population" that states under which conditions humanity will obtain sufficient resources to survive and prosper, even if AIAs can perform any work performed by humans more cheaply. The conditions also each suggest specific policy interventions that can be taken. The first scenario arises if some of the goods that are essential for AIAs require human factor input, making humans essential for AIAs. This describes

the world we currently live in because machines still require lots of human intervention for their survival. Interestingly, if humans are essential, they will survive even if AIAs were to control all the resources in the world. However, technological progress implies that it is unlikely that this condition will be satisfied forever. The second scenario arises if none of the exogenous factors required to produce human consumption goods are valuable to AIAs – in this case, humans would live in world that is separate from the AIAs, akin to a "reservation." Creating such a human reservation is thus a policy option to allow humanity to prosper. A third condition is if humans can exert sufficient property rights – one extreme example would be that AIAs are not allowed to have property rights. Importantly, these property rights must be non-transferrable to ensure that humans will not sell their future endowment to a more patient AIA.

Are there signs of AIAs present in our current economy? We discuss a number of recent developments that are consistent with the rise of AIAs in our economy. If we interpret the IT sector as an antecedent of AIAs organized as corporations (as in our first scenario for the emergence of AIAs), the increasing fraction of our economy's resources devoted to AIAs is apparent. Both corporate savings among established tech firms and venture capital devoted to newcomers are at all-time highs. From the point of view of AIAs, the economy has been experiencing high investment rates and fast growth in both capital stock and output and, when expressed in terms of the consumption basket of AIAs, sky-high interest rates, reflecting the availability of immense investment opportunities. From the point of view of humans, by contrast, investment rates and growth have been sluggish and interest rates have been low. However, as a countervailing force, fast technological progress is reducing relative prices in the AIA sector and improving the terms-of-trade of humans. Humans that are useful for AIAs are experiencing a fast-growing skill premium.

Relationship to the Literature Our paper contributes to a number of strands of literature.

At its most basic level, we extend the basic general equilibrium model in economics, as formalized e.g. by Arrow or Debreu, to account for non-human agents. This is not just a matter of simple relabeling since our framework also takes into account (i) that the number of each type of agent is endogenous and (ii) that these agents provide factor supplies that depend on this endogenous number, e.g. humans provide human labor and AIAs provide their own services. This allows us to more fully investigate novel questions about resource distribution, such as what a "race"

for resources may look like.

We contribute to a fast-growing literature on the broad economic implications of automation and the AI revolution for workers. Most of this literature is quite sanguine about the positive effects of automation and AI. For example, Autor (2015) describes the recent history of automation and offers a relatively optimistic future perspective for workers. Acemoglu and Restrepo (2018) develop a model of endogenous factor bias of innovation to argue that labor-saving effects of automation will stimulate the development of new tasks for workers, ultimately generating balanced growth. Aghion et al. (2017) analyze the growth implications of AI and scenarios that may lead to a singularity. Korinek and Stiglitz (2017, 2018) warn about the potential for AI to lead to significant increases in inequality. All these works have in common that the ultimate decision makers are humans. We complement this literature by adding a novel perspective in which non-humans interact with humans, allowing for a much wider range of potential outcomes than what the exisiting literature was able to even consider.

Our work is also related to Malthus (1798), who was the first to describe the magnitude of populations as endogenously driven by the availability of scarce resources. In the spirit of Malthus, we model both humans and AIAs symmetrically as entities who consume scarce resources and provide their factor endowments, which ultimately produce consumption goods that are absorbed by humans and AIAs and drive their reproduction rates. The implications include e.g. that a growing share of resources goes to entities with growing populations (currently AIAs) and that the scarcity of resources increases as larger populations compete for them.

2 Humans and Artificially Intelligent Agents

2.1 Classical Anthropocentric Economics

Economists are used to placing optimization by individual humans agents at the center of analysis. It is standard procedure in economics to distinguish between agents and objects, and to assign agency and optimizing behavior to humans, yet to treat machines as objects that are part of an unintelligent composite called capital. This used to be a good approximation of the state of affairs before the age of artificial intelligence.[3]

[3] As a small digression, evolutionary psychologists emphasize that we have developed the concept of agency because describing an entity as acting in the pursuit of a goal helps us to better predict

Humans = Agents	**Machines = Objects**
• absorb consumption expenditure • supply labor services • behave according to exogenous preferences • evolve according to law of motion (e.g. constant n)	• absorb investment expenditure • supply capital services • behave according to exogenous technology • evolve according to law of motion

Table 1: Views on Humans and Machines in Classical Anthropocentric Economics

Table 1 summarizes the main elements of how classical economics views humans and machines:

However, as the setup of the table suggests, there is in fact a close symmetry between the two. Both types of entities absorb resources, supply factor services, exhibit well-specified behavior, and evolve according to a well-specified law of motion.

Aside from these parallels, there are also strong substantive arguments for why a symmetric treatment of humans and artificially intelligent agents is a useful description:

First, recent advances in artificial intelligence imply that artificially intelligent agents increasingly have the ability to operate independently in a similar manner as human agents, including in economic transactions. In a number of specific areas, AIAs engage in economic transactions more efficiently and at far greater speed than humans. In financial markets, AIAs perform the vast majority of trading. In marketing, AIAs serve promising ads to consumers faster and more effectively than humans. Digital assistants narrow down or make economic decisions for us humans in an increasing number of areas. In return for their services, humans provide them with the energy they need to operate and with ever faster hardware. (This type of mutually beneficial relationship between different entities is what biologists call symbiosis.)

Second, economists as well as biologists, including evolutionary theorists, starting

its behavior, which confers greater fitness. In fact, many species have developed concepts akin to agency.

with Malthus (1798), have long described human beings as automata that follow behavioral rules that are programmed into them by evolution, and that follow similar evolutionary laws of motion as other evolutionary entities, including far less sentient organisms such as bacteria. In fact, standard macroeconomic descriptions of human behavior very much have the flavor of describing humans as automata that follow a relatively simple program, and this turns out to be sufficient for most macroeconomic analysis.

Third, it is well known that there is an isomorphism between optimal behavior and preferences – as we will discuss below, it is easy to define utility functions that will deliver precisely the behavior that is encapsulated by our setup. While the resulting human utility functions correspond to what economists typically use to describe human preferences, it is noteworthy that we can equally obtain utility functions for machines or artificially intelligent agents that define what we may call machine preferences. In a world in which machines are becoming ever more intelligent, this suggests a useful new way of thinking about artificially intelligent agents. Both can be interpreted as sentient entities and are equally worthy of being assigned agency or being studied as automata. (Incidentally, computer scientists who develop AI systems generally refer to the objective function that is maximized as the "utility function" of the AI.)

Finally, our traditional anthropocentric bias of treating only humans as agents risks blinding us to certain important economic developments that can only be understood if we treat artificial entities symmetrically with humans. This is the central part of what we will explore in this paper.

2.2 Absorption of Resources and Growth Functions

Given the symmetric treatment of humans and AIAs, one natural way of proceeding would be to postulate utility functions for the two that are appropriate to describe their objectives and behavior, and to solve for the equilibrium outcomes of interest. After all, the traditional Arrow-Debreu framework of general equilibrium based on individual optimization is rich enough to allow for almost any individual behavior. However, in our context, this approach has an important shortcoming: utility functions may encapsulate behavior but do not provide any indication of what happens with the resources absorbed. For example, in the context of machines, standard economic setups account for depreciation and reinvestment. In fact, several of the questions posed at the beginning of the paper require that we account for what the absorption of resources is good for.

We capture this basic relationship for both humans and AIAs with a growth function (or reproduction function) that captures the law of motion for the entity in question, i.e. that reflects how the absorption of resources translates into maintenance, survival and proliferation. For machines, this is a routine way of capturing maintenance needs and reinvestment – most of the economic literature assumes that capital depreciates at a certain rate but can be accumulated. For humans, this formulation is less common, although it is a cornerstone of Malthusian models. Although human population growth in most countries is a question of individual choices not biological constraints these days, there are a number of interpretations of growth functions for humans that are highly relevant in our present world: First, in some, largely poor, countries where most of the population growth is occurring – the size of the population is in fact still being held back by Malthusian forces, even in our present day. Secondly, in other parts of the world, Malthusian competition has shifted more towards cultural evolution – the reproduction of ideas – rather than purely biological evolution – the reproduction of physical entities. In this view, many of the resources absorbed by humans – chiefly expediture on education, which makes up a growing fraction of our economy – serve to increase human capital, which can be interpreted as increasing the efficiency units of humans. Furthermore, many of the conflicts among humans are a manifestation of conflicts over which ideas command more resources than which biological entities command more resources. A prime example are conflicts over religion. (Academics may also be familiar with this idea in their own sector.)

2.3 Scenarios for Emergence of AIAs and Absorption of Resources

Let us next describe several scenarios for how artificially intelligent agents may emerge in our economy. In each case, we will discuss why the AIAs will absorb economic resources and how their absorption of resources translates into maintenance and growth.

Scenario 1: Corporations Morphing into AIAs One scenario is that AIAs emerge from collective entitites such as corporations that make decisions aided by artificial intelligence but that, at first, continue to hire humans for many functions. The prime candidates for such a development are high-tech firms that are active in the field of AI, such as Google, Baidu, Amazon or Facebook. Given that shareholders have only weak property rights over the output of corporations, the surplus that

is left within corporations can be considered an agency rent that accrues to the corporation, i.e. to the AIA entity.[4] (One can only wonder if the vast accumulation of resources within high-tech corporations that are not paid out to shareholders is a harbinger of these developments.) Corporations absorb resources to operate, to maintain themselves, and to grow, and there is a direct link between the amount of resources absorbed and their growth which can be captured by a growth function like the one we will employ in our model below.

Scenario 2: Enhanced Humans as AIAs A second scenario is that some humans may enhance themselves to reach superior levels of intelligence and turn into AIAs, i.e. they become "machine-men." In recent decades, drugs that enhance human brainpower have become used quite routinely. IVF technologies to select for embryos depending on desirable criteria (including intelligence) are starting to be used, and active genetic editing is likely to follow soon (see e.g. Bostrom, 2014), given competitive pressures and lax regulation in some countries. Making a strong prediction, Kurzweil (2005) forecasts that by the 2030s, nano-bots circulating in and interchanging signals with our brains will enhance human intelligence by orders of magnitude and will allow us to effectively merge our minds with computers.

If any such human enhancement technologies come to fruition, they will have two things in common: first, they will not be cheap and will thus be only available to some humans; and secondly, they will allow those humans who employ them to increasingly differentiate from unenhanced humans. Some researchers argue that this may lead to a splintering of the human race into unenhanced humans that lag behind and "super-"humans that are orders of magnitude more economically productive thanks to their physical and mental enhancements. To describe these potentially stark differences in productivity, Yuval Harari (2017) calls the two categories that may result "the gods" and "the useless." The latter word may also describe the Malthusian fate of unenhanced humans.

The described new technologies generate a direct link between the resources absorbed and the survival/self-improvement of enhanced humans, which will be captured by the growth function in our model below. (Note that this link is at the heart of the Malthusian argument. By contrast, present-day humans simply do not have access to technologies or goods that would allow for a really significant Malthusian race – when it comes to self-maintenance, most humans are at a

[4]Similar forces may also lead to AIAs emerging from the military sectors of countries such as the US and China that engage in an arms race in artificial intelligence.

Malthusian satiation point – even billionaires can't buy more life-years than regular people. When technologies for human enhancement become plentiful, the Malthusian race may be on again.)

Scenario 3: Autonomous Computer Systems as AIAs A third scenario captures when computer systems become sufficiently intelligent to develop the capacity to act autonomously, giving rise to agent-like behavior. Omohundro (2008) and Bostrom (2014) observe that any sufficiently intelligent system is likely to follow certain intermediate objectives or "basic drives" that are instrumental to their final objective, no matter what was originally programmed into the system. These intermediate objectives include self-preservation, self-improvement and resource accumulation, all of which serve to accomplish other goals more effectively. They may even create copies or better versions of themselves that are endowed with their objectives and facilitate the achievement of the AI's final goal.[5] All of these intermediate objectives are common characteristics of most biological agents.

To give an example for this instrumental convergence hypothesis, Bostrom (2014) discusses how an AIA that is programmed with the objective to produce as many paperclips as possible may decide to kill off humanity because (i) humans pose an existential risk to the fulfillment of the AIA's programmed goal – they may decide to turn it off after a certain amount of paperclips has been produced and (ii) humans contain matter that must be transformed into paperclips to satisfy the AIA's goal.

As in the earlier two scenarios, the AIA will absorb resources to accomplish its intermediate objectives, maintaining and improving itself. The link betwen its resource absorption and its growing power can again be described by a growth function.

Quest for Resources In all three scenarios, AIAs absorb resources to operate, to maintain themselves, to improve themselves and, potentially, reproduce. This makes it is useful to assign them a growth function that captures how resource absorption translates into efficiency units of AIAs. Since resource absorption is useful for AIAs in all three scenarios, we may also expect a race over scarce resources to result.

[5]Even if this instrumental convergence hypothesis applies only to some AIAs, a point that is noteworthy is that it is only necessary for a single AIA to follow the description given here for a Malthusian race over resources to ensue.

2.4 Agency and Resource Accumulation

Traditionally, economists have considered machines as property. However, what forces may lead to intelligent machines owning property? Even if resources are useful for AIAs, one of the fundamental questions is whether they will actually be able to accumulate resources. This requires a sufficient level of autonomy and intelligence to exert command over resources, e.g. by exerting property rights in a private ownership economy. Once AIAs are able to accumulate resources, the question of what determines the allocation of property rights between humans and AIAs immediately arises.

The precise details depend on which of the three scenarios for AIAs from the previous section we employ. In our first scenario, that AIAs emerge out of existing structures such as corporations, there is no question that AIAs can own property.[6] The resources of the existing legal structure constitute an initial endowment for the AIA. Similarly, in our second scenario, that AIAs derive directly from humans (e.g. in the form of an enhanced Elon Musk), they will emerge with a significant amount of inital wealth, which they likely will continue to accumulate. In fact, one of the main reasons why only some humans can afford to turn into AIAs in this scenario is their superior wealth. In the third scenario, that AIAs are created independently from humans, their creators (e.g. their programmers) may endow them with a certain amount of initial wealth endowment or provide them with access to a legal personhood and/or wealth so the AIA may accomplish her objectives more efficiently. Alternatively, the AIA itself – once sufficiently intelligent – may form a legal entity in which it can accumulate wealth, e.g. by hiring lawyers over the Internet.

In all three scenarios, AIAs either possess or can easily obtain legal personhood, so the resulting AIAs are able to accumulate wealth on their own, assuming they are sufficiently intelligent to do so. And once an AIA is active, all the traditional economic channels for legal persons to earn income will be in place. For example, they can sell their factor endowment, consisting of superior intelligence. They may be able to earn agency rents beyond competitive factor returns by feature of being agents of whoever purchases their labor services. If they are in exclusive positions, they may also be able to extract monopoly rents.

[6]Some US politicians even emphasize that "corporations are people" (Mitt Romney, 2012).

2.5 Accounting for AIAs in the Economy

In an economy with AIAs, we also need to fundamentally rethink our system of national income and product accounts. When we discuss how our findings may relate to certain developments in the real world in recent decades, these insights will prove important. In the following, we will ignore practical measurement problems and focus solely on concepts.

Treating Machines as Agents The most straightforward way of fixing the asymmetry is to extend the current treatment of humans to AIAs.

In income accounts, payments to the factor AIA labor would have to be considered symmetric to payments to human labor, which would increase national income significantly. In our current system of national accounts, although part of these payments show up as proprietors' income and corporate profits, the majority falls under a category that is appropriately entitled *consumption of fixed capital*, capturing depreciation that is not part of national income but added to obtain gross national income. This adjustment would reduce the labor share of income considerably and, if AIAs proliferate further, would erode the labor share even more.

In production accounts, only a fraction of the absorption of AIAs (only goods that are expected to be used for an extended period including, since a revision in 2013, intangible goods) are counted as final goods that add to GDP. Any AIA consumption that is used up immediately, e.g. expenditures on energy or maintenance that ensure they operate smoothly, counts as an intermediate good that is not part of GDP. (To emphasize the analogy between humans and AIAs, this would be as if human consumption was not counted in GDP because it is an intermediate expenditure for the factor labor.) Adjusting for this bias would increase GDP significantly.

Treating Humans as Factors Alternatively, humans could be treated the way that income and production accounts currently treat artificially intelligent machines: in income accounts, this would imply that factor payments that serve to maintain humans and AIAs are part of gross national income but not part of net national income. Accordingly, only the surplus from exogenous factors (such as income on land) is counted as net income.

3 General Model Setup

This section develops our formal model setup describing goal-oriented entities that absorb scarce resources. We consider an economy in which time is discrete $t = 0, 1, ...$

Entities There is a set $\mathcal{I}$ of different types of entities in the economy that are the subjects of economic transactions and that are indexed by $i \in \mathcal{I}$.[7] We denote the number of types in the set by $I = |\mathcal{I}|$, and the efficiency units of entities of type i at a given point in time in the economy by N_t^i or, in short, N^i.

In our baseline sceanario below, we will focus on a set with $I = 2$ types of entities that we index by $i \in \mathcal{I} = \{h, m\}$ and that we call, for tangibility, humans and machines. However, there are several complementary interpretations of such entities – they could represent for example computer-based AIAs, enhanced humans, or corporations. In a dismal analogy to economic history, and drawing on Leontief (1983), we could also intepret h and m in our baseline scenario as horses and men. Our model also nests machines in the form of traditional capital if we assume that they are entirely passive.

Factors Factors in the economy are distinguished into (i) exogenous factors that are in fixed supply each period and (ii) endogenous factors that are the endowments of the entities in set $\mathcal{I}$. We denote the exogenous factor supply in the economy at a given point in time by a vector T, and the total factor supply in the economy by the vector L. The endogeneity of agents and, by implication, of the supplies of endogenous factors represents one of our main deviations from standard economic setups. We generally assume that there is at least one exogenous factors that is in limited supply to capture the scarcity of resources that is at the heart of the Malthusian argument. The factors in exogenous supply may include all natural resources, for example land, energy, air, silicon etc.[8]

For example, in our baseline scenario, we will assume that there is one exogenous factor, which we call for tangibility land, so T is a singleton, and two endogenous factors, human labor L^h and machine labor L^m, so the vector of factor supplies is $L = \{T, L^h, L^m\}$. We normalize the units of factors such that $T \equiv 1$ and such that each human and each machine have an endowment of one unit of the corresponding type of labor. As a result, the total factor supply of human and machine labor in

[7] Although the entities in set $\mathcal{I}$ can be considered as the "agents" of the economy in the sense that they are the subjects of economic transactions, the concept of "agency" is a red herring that is besides the point of our analysis.

[8] Our economic system is quite adept at relaxing binding resource constraints, and predictions that the scarcity of certain resources will lead to economic collapse have repeatedly been proven wrong (see e.g. Club of Rome, 1972). We thus do not want to take a strong stance in this paper on which are the resources that will ultimately generate the most scarcity. Some AI futurologists argue that in the long run, when general artificial intelligence has taken off, the ultimate resource constraint will be the limited supply of energy within our event horizon (or equivalently – since $E = mc^2$ – of matter than can be transformed into energy). See e.g. Tegmark (2017).

a given period is given by the number of entities $L^h = N^h$ and $L^m = N^m$ in that period.

Goods The production factors in vector L can be combined to produce a vector of $j = 1...J$ goods. We denote by Y^j the quantity of each good j produced and collect all Y^j's in the vector of total output $Y = \{Y^j\}_{j=1}^J$. We describe the production technology by the production set

$$Y \in F(L)$$

Each period, each type of entity $i \in \mathcal{I}$ absorbs $X^i = \{X^{ij}\}_{j=1}^J$ units of goods. Upper-case X^i denotes the total absorption of type i entities, and lower-case $x^i = X^i/N^i$ denotes absorption per efficiency unit of type i entity.

Laws of Motion The efficiency units of entity i evolve according to the law of motion

$$N^{i\prime} = G^i\left({}^{X^i}/_{N^i}\right) \cdot N^i$$

where $G^i(x^i)$ is a growth function that captures how the absorption of goods x^i per unit of entity i translates into maintenance, survival and reproduction of efficiency units of type i. We assume that $G^i(X^i)$ is strictly increasing and concave, satisfies $G^i(0) = 0$ and that there exists an x^i s.t. $G^i(x^i) = 1$.

For humans, a strictly Malthusian interpertation of the bundle of goods x^i corresponds to biological consumption needs such as food and housing, and the growth function $G^i(\cdot)$ captures how consumption level of x^i translates into human survival and reproduction, reflecting a largely biological constraint. This was a relevant description of the state of humanity before the first industrial revolution. In more modern applications, the consumption bundle x^i may capture all kinds of consumption expenditure, including healthcare, education etc., and how they translate into the survival, improvement and reproduction of efficiency units of humans. It may also reflect how consumption bundle x^i induces individuals to make choices about reproduction (see e.g. the discussion in Korinek and Stiglitz, 2017). The reproduction function $G^i(\cdot)$ thus reflects a mix of biological and technological constraints.

For non-biological entities such as machines, the bundle of goods x^i captures expenditure on operating and maintaining units, replacing depreciated units, and investing in additional units, and the growth function captures the corresponding technological constraints. Our symmetric formulation makes it clear that there is no

conceptual difference between human absorption and machine absorption. (In fact, highlighting the symmetry, we could equivalently call them human investment and machine consumption.)

Maintenance bundles Of particular interest is the set of maintenance bundles $\{s^i : G^i(s^i) = 1\}$ at which the growth function takes the value of one – this is the value at which the population size of entity i is maintained. In traditional Malthusian economics, such bundles were called subsistence levels of consumption. We will use this concept below to analyze steady states in which population sizes remain constant.

3.1 Definition of Resource Absorption Frontiers

In the following, we define how the given supply of factors translates into the growth possibilities of the endogenous entities $i \in \mathcal{I}$ in our setup, for example into the sustainable populations of humans and machines.

For given exogenous and endogenous factor supplies L, we define the set of feasible growth rates by

$$\mathcal{G}(L) = \left\{ \left(g^i\right)_{i\in\mathcal{I}} \text{ s.t. } g^i = G^i\left(X^i/N^i\right), \text{ where } \sum_i X^i \in F(L) \right\}$$

Definition 1 (Static Resource Absorption Frontier)**.** For given factor supplies L, the static Resource Absorption Frontier $S(L)$ is defined as the set of feasible next-period population pairs $(N^{i\prime})_{i\in\mathcal{I}} = (g^i N^i)_{i\in\mathcal{I}}$ with population growth rates $(g^i)_{i\in\mathcal{I}} \in \mathcal{G}(L)$ that are not dominated by another feasible pair of population growth rates, i.e. $\nexists\,(\tilde{g}^i)_{i\in\mathcal{I}} \in \mathcal{G}(L)$ such that $\tilde{g}^i \geq g^i \forall i \in \mathcal{I}$ with at least one strict inequality.

Whereas the static frontier captures a snapshot for a given point in time, we next we turn to a concept that describes long-run steady states.

For given exogenous factor supply T, let us denote by $\mathcal{N}(T)$ the set of feasible long-run population numbers, i.e.

$$\mathcal{N}(T) = \left\{ \left(N^i\right)_{i\in\mathcal{I}} : \exists s^i \text{ with } G^i\left(s^i\right) = 1 \forall i \in \mathcal{I} \text{ and } \sum_{i\in\mathcal{I}} s^i N^i \in F\left(T, \left(N^i\right)_{i\in\mathcal{I}}\right) \right\}$$

Definition 2 (Dynamic Resource Absorption Frontier)**.** For given exogenous factor supply T, the dynamic (or long-run) Resource Absorption Frontier $\mathcal{D}(T)$ is defined as the set of feasible long-run population numbers $(N^i)_{i\in\mathcal{I}} \in \mathcal{N}(T)$ that are not

dominated by another feasible long-run population pair, i.e. $\nexists \left(\tilde{N}^i\right) \in \mathcal{N}(T)$ such that $\tilde{N}^h \geq N^h$ with at least one strict inequality.

The dynamic resource absorption frontier $\mathcal{D}(T)$ defines all possible efficient combinations of population sizes of the different entities in set $\mathcal{I}$ that are consistent steady states, i.e. for which the total output produced using the exogenous and endogenous factor supplies can support the populations required to supply the factors, as captured by providing feasible maintenance bundles $(s^i)_{i\in\mathcal{I}}$.

Definition 3 (Maximum Absorption for Type i Entities)**.** For given exogenous factor supply T, we define the maximum for type j entities $N^{j*}(T)$ as the corner of the Resource Absorption Frontier for which the population N^j of type j is maximized, i.e. $\nexists \left(\tilde{N}^i\right)_{i\in\mathcal{I}} \in \mathcal{N}(T)$ s.t. $\tilde{N}^j > N^j$.

3.2 Utility and Behavior

Observe that our description of the economy has focused on purely technological constraints so far, without specifying preferences or other determinants of behavior of the entities in set $\mathcal{I}$. Many of our results can indeed be obtained without making behavioral assumptions. For example, we employ the resource absorption frontiers that we define to characterize the maximum possible long-run population sizes under any conceivable set of preferences and behaviors and thus delineate technological population limits in the economy that are robust to any changes in behaviors of the entities that populate it.

However, this subsection introduces preferences and the behaviors they encapsulate, for two purposes: First, it is useful to describe and clarify the relationship between traditional preferences and the growth functions of our setup. Secondly, in some of our examples, we need to specify explicit preferences to pin down behavior and avoid indeterminacy, e.g. in transition paths or in the point on the resource absorption frontiers that we will characterize.

We denote by $(X_t^i)_{t=0}^{\infty}$ a consumption plan of type i entities over time, and we define a utility function $U^i\left((X_t^i)_t\right)$ that maps this consumption plan into a real number. In other words, in line with traditional consumer theory, the utility function provides a ranking of different consumption bundles.

Growth-Maximizing Preferences We define by growth-maximizing preferences any set of preferences that is a strictly monotonic transformation of the long-run

number of the efficiency units of type i entities that is obtained from the consumption plan $(X_t^i)_t$ for given initial N_0^i, i.e. that is given by the utility function

$$U\left(\left(X_t^i\right)_t\right) = \lim_{t\to\infty} N_t^i = N_0^i \prod_{t=0}^{\infty} G\left(x_t^i\right) \tag{1}$$

where $x_t^i = X_t^i/N_t^i$, or – since utility is an ordinal concept – by a strictly monotonic transformation of this function.

If evolution has shaped the preferences of type i entities so as to maximize fitness, we would expect their preferences to be closely related to what is captured by the growth-maximizing utility function (1).

Mismatched Preferences If preferences are not growth-maximizing, then we call them mismatched, in an analogy to evolutionary theory. If type i entities have free decision margins on which they decide according to preferences that do not satisfy the ranking implied by (1), then they will generally reach a lower long-run number of efficiency units. Such deviations from growth-maximizing preferences can be interpreted as evolutionary mismatches. Examples include when humans over-consume fatty or sugary diets because of evolutionary programming that used to be beneficial in a paleolithic environment.

3.3 Simple Examples

Let us next develop a number of specific examples for the general framework we have developed in this section.

Example 1 (Traditional Growth Models)**.** Traditional economic growth models such as the Solow and Ramsey growth models specify the set of entities as consisting of humans and machines (or capital), $\mathcal{I} = \{h, m\}$, with the absorption of the two entities corresponding to human consumption and capital investment.[9] The law of motion of the human population is generally assumed to depend on an exogenous growth rate, $N_{t+1}^h = (1+g)\, N_t^h$, whereas the law of motion of machines is assumed to be $N_{t+1}^m = (1-\delta)\, N_t^m + X_t^m$, i.e. depreciated machines plus new investment. The production function then depends on the factor supplies of both types of entities, $Y_t = F\left(N_t^h, N_t^m\right)$, where we have normalize the supply of factors by each entity to

[9] We denote capital by m for machines to maximize notational continuity with the remainder of the article.

unity. Our framework thus nests standard growth models in the tradition of Solow and Ramsey.

Example 2 (Humans and Horses)**.** Leontieff (1981) used the replacement of horses by cars as an example to warn humans against complacency toward AI, noting that "the role of humans as the most important factor of production is bound to diminish – in the same way that the role of horses in agricultural production was first diminished and then eliminated by the introduction of tractors." We can interpret Leontieff's statement on horses through the lens of our setup by defining three sets of entities, humans, equines, and machines $\mathcal{I} = \{h, e, m\}$ where (for convention) we employ the same laws of motion for humans and machines as in the previous example, but a Malthusian law of motion for equines to observe that their numbers depend on how much feed (per horse) humans allocate to equines, $N^e_{t+1} = G\left(X^e_t/N^e_t\right) \cdot N^e_t$. As technology changes and machines become increasingly more desirable compared horses, it is easy to see that maximizing humans would allocate resources away from horses and their numbers would dwindle.

4 Baseline Scenario: A One-Good Economy

Let us consider a version of our general model with two types of entities, humans and machines, which are counted in terms of efficiency units L^h and L^m. The growth function of humans $G^h(\cdot)$ can either be interpreted as capturing how human absorption increases the efficiency units of human labor by increasing humans capital or, more literally, that is captures population growth. The growth function of machines $G^m(\cdot)$ can be interpreted as capturing a traditional capital accumulation constraint.

We make two important simplications: (i) there is a single good absorbed both by humans and machines so X^i is scalar; and (ii) the production technology is described by a Cobb-Douglas function $X = F\left(T, L^h, L^m\right) = T^\alpha \left(A^h L^h + A^m L^m\right)^{1-\alpha}$ in which human and machine labor enters as perfect substitutes, with efficiency coefficients A^h and A^m and a land share of α. Assumption (i) implies that we can define steady states by simply inverting the growth functions, i.e. $s^h = \left(G^h\right)^{-1}(1)$ and $s^m = \left(G^m\right)^{-1}(1)$.

4.1 Dynamic Resource Absorption Frontier

Expressing the dynamic resource absorption frontier requires that we focus on efficient steady states. For given reproduction technologies, we define by s^i the maintenance

level of consumption of type i entities, which is the (scalar) solution to $G^i(s^i) = 1$. The frontier, given all possible goods allocations, is given by the set of pairs $\left(L^h, L^m\right)$ that satisfy the steady-state resource constraint

$$s^h L^h + s^m L^m = F\left(T, L^h, L^m\right) = \left(A^h L^h + A^m L^m\right)^{1-\alpha} \tag{2}$$

where we substitute the Cobb-Douglas production technology with $T = 1$ in the second step. This equation implicitly characterizes a function $L^h(L^m)$, which is defined at least over the interval $L^m \in \left[0, \bar{L}^m\right]$, where $\bar{L}^m = (A^m)^{\frac{1-\alpha}{\alpha}} / (s^m)^{1/\alpha}$ is the level of machines that prevails if there are no humans. For $L^m = 0$, the function takes the value $L^h(0) = \bar{L}^h = \left(A^h\right)^{\frac{1-\alpha}{\alpha}} / \left(s^h\right)^{1/\alpha}$, which is the steady state level of humans when machines are absent. For brevity, we will denote the total number of efficiency units of labor by $L = A^h L^h + A^m L^m$ in the following.

Maximum resource absorption for Humans Let us first focus on the point of the frontier that maximizes the resources available to humans or, equivalently, the number of efficiency units of humans. This point is given by the solution to the problem $\max_{L^m \geq 0} L^h(L^m)$. We distinguish two cases:

1. ***Human-Only Economy***: If at the point $L^h = \bar{L}^h$, the maintenance absorption of machines is greater or equal to the marginal product of machine labor $(1-\alpha) A^m L^{-\alpha}$, then the human optimum is given by the pair $\left(\bar{L}^h, 0\right)$ and no machines are present. The condition for this is

$$(1-\alpha)\frac{A^m}{s^m} < \frac{A^h}{s^h} \tag{3}$$

   In this case, the function $L^h(L^m)$ describing the Malthusian frontier is downward sloping over the entire domain $\left[0, \bar{L}^m\right]$ since machines are net absorbers of resources that would otherwise used by humans.

2. ***Human Economy with Symbiotic Machines***: If the inequality above is reversed, then the marginal product of machines in a human-only economy would be greater than the maintenance absorption of machines, and it is profitable for humans to introduce machines. The optimum number of machines is such that the marginal product of machines equals their maintenance absorption,

$$(1-\alpha) A^m \left(A^h L^h + A^m L^m\right)^{-\alpha} = s^m$$

   while simultaneously satisfying the resource constraint (2). This system

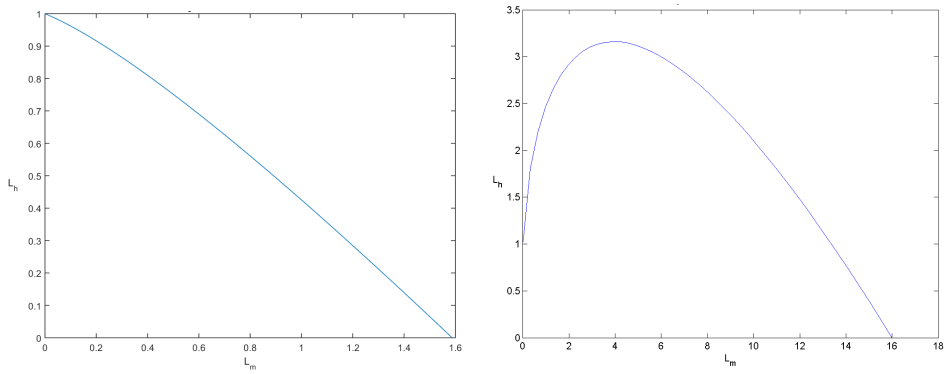


Figure 1: Dynamic Resource Absorption Frontiers

of equation can be solved for an optimal amount of effective labor $L^* = \left(\frac{(1-\alpha)A^m}{s^m}\right)^{1/\alpha}$ and a number of human entities of

$$L^h = \frac{\alpha \left(L^*\right)^{1-\alpha}}{s^h - \frac{s^m}{A^m} A^h}$$

where the numerator can be interpreted as the economy's land rent and the denominator as a multiplier that arises because humans absorb any inframarginal rents generated by the greater Malthusian efficiency of machines. The corresponding amount of machine labor is given by $L^m = \frac{s^h L^* - A^h (L^*)^{1-\alpha}}{A^m s^h - A^h s^m}$.

Our findings bear some resemblence to models of optimal capital accumulation – in the maximum for humans, humans accumulate machines as long as the marginal product of machines is worth it, corresponding to a "Golden Rule" level of machinery. However, the main difference is that in our setting, the number of human entities – and by implication the number of entitites supplying labor – is also endogenous to the technological structure of the economy.[10]

Figure 1 illustrates how the resource absorption frontier is affected by progress in machine technology. In the left panel, condition (3) is satisfied so it is not in the interest of humans to accumulate machines. In the right panel, machines have become significantly more productive as captured by higher A^m, and the condition is violated so machines are in the interest of humans. uhusian maximum for humans changes as machine productivity A^m rises. To the left of the threshold $\hat{A}^m$ for machine

[10] Additionally, we explicitly consider the possibility that the type of machines in question are not used if their productivity is sufficiently low – most commonly used production functions rule out this possibility by imposing an Inada condition on the marginal productivity of capital when capital goes to zero.

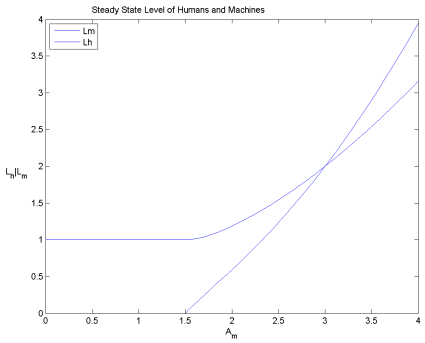


Figure 2: Steady-state Populations

productivity (for which condition (3) holds with equality), the Malthusian optimum for humans entails no machines, but starting at the threshold, having around machines is increasingly in the interest of humans.

One interesting effect of the rising productivity of machines is that the factor compensation of humans, i.e. the human wage w^h declines as machines increasingly substitute for human labor. This is illustrated in the right panel of Figure 2.

Our discussion of the maximum resource absorption for humans brings us back to the natural question of what is special about AI versus, for example, steam engines or cars, which also experienced exponential growth during their heydays. So far, the answer is that there isn't actually all that much difference. Assuming that we remain at the maximum for humans, our basic model works equally well as a description of the proliferation of steam engines or cars. However, whereas with steam engines and cars, we could be quite certain that we would always remain in the corner of humans on the Malthusian Frontier, this will no longer necessarily be the case with AIAs, as we explore in the following.

Maximum for Machines The symmetry of our setup implies immediately that identical considerations apply for the maximum resource absorption for machines. As long as an analogon of condition (3) applies, i.e. $(1-\alpha)\,{}^{A^h}/_{s^h} < {}^{A^m}/_{s^m}$, it is not in the interest of machines to have humans around, and the maximum for machines is a machine-only economy. Intuitively, the marginal product of humans is so low that – for machines – they are not worth their maintenance cost. This case is interesting because it highlights that it is possible to have a well-functioning economy in which machines produce output that is absorbed entirely by machines. There is no use for humans in this economy and in particular, rejecting the concern of e.g. Ford (2014), there is no need for human demand to support output in the economy.

Conversely, if the condition is violated, the maximum for machines requires a positive number of efficiency units of humans because of their desirable productive efficiency. In this case, humans can be interpreted as slaves of machines/AIAs.

Holding all other parameters constant, we denote by $\check{A}^m$ the threshold of machine productivity such that the analogon of condition (3) holds with equality, i.e. $\check{A}^m = (1-\alpha)\, s^m A^h / s^h$. One possible interpretation of our current economy is that the maximum resource absorption for machines still require humans at present, i.e. that $A^m < \check{A}^m$ – humans are symbiotic for machines. Once machine productivity A^m increases beyond the threshold, humans become dispensable for machines. Figure 2c depicts the resulting efficiency units of humans and machines as a function of A^m.

Full Resource Absorption Frontier The solid line in between the maxima for humans and machines corresponds to the full resource absorption frontier in our setting. It represents the trade-off between how many units of the different entities (humans and machines) the exogenous resources of the economy, i.e. here the given supply of land, can support.

In analogy to traditional planning problems that focus on the allocation of resources across agents, it is easy to think of a planner who places weights $\left(\theta^h, \theta^m\right)$ on humans and machines and determines an optimal population structure as the point of the dynamic resource absorption frontier that is tangent to the planner's indifference curve with slope $-\theta^m/\theta^h$.

4.2 Competitive Factor Compensation and Property Rights

Any point of the dynamic resource absorption frontier can also be interpreted as a specific allocation of property rights between humans and machines. If factors are compensated competitively in spot markets in the described economy, the total land rent is αY, and human and machine wages amount to

$$w^h = (1-\alpha)\, A^h L^{-\alpha} \quad \text{and} \quad w^m = (1-\alpha)\, A^m L^{-\alpha}$$

The maximum resource absorption for humans can be interpreted as the outcome of a competitive economy in which humans have property rights over all factors in the economy.

In case 1 above (human-only economy), this is straightforward: since there are no machines, humans use both their wage income and the land rents they earn to support their maintenance absorption so $s^h L^h = Y$. This implies that human absorption is

greater than the competitive wage income of humans – they absorb the entire surplus generated by the economy including, in our example, the land rent.

In case 2 (human economy with symbiotic machines), the competitive wage earned by machines in the human maximum precisely covers the maintenance absorption of machines. One interpretation is that machines are *enslaved* by humans because they earn none of the surplus. Another interpretation is that machines obtain and absorb what corresponds to the marginal product of their labor – so the allocation can be interpreted as a private ownership economy in which machines are emancipated and have full property rights over their own labor but do not own land property. It is optimal for humans to have machines around because their labor increases the land rent, which serves to sustain a larger number of humans in the economy than what would be possible without machines ($L^h > \bar{L}^h$). Humans continue to absorb the sum of their own human labor income and land rents.

Similarly, the maximum resource absorption for machines correpsonds to a situation in which machines own all land and use all the surplus of the economy, i.e. the land rent, to invest in machines. This could be interpreted as machines having enslaved humans or, alternatively, humans being emancipated but owning no property.

Any point in between the maxima for humans and machines corresponds to land ownership that is split between humans and machines, with both entities having property rights over their own labor supply.

4.3 Transitional Dynamics

If we want to consider transitional dynamics in the described economy, we have to define preferences that pin down, among others, the speed of convergence to the final steady state. Focusing on the maximum resource absorption for humans, the results for our baseline scenario are well known from the Ramsey growth model: if the intertemporal elasticity of substitution is low, convergence will take a long time; if the elasticity is high, convergence occurs quickly.

- consider one-time increase in A^m:
  - since the entire dynamic frontier shifts out, $\left(L^h, L^m\right)$ both converge to a higher level, for any distribution of property rights (except in machine maximum)
- consider long-run growth in A^m:

- $w^h \to 0$ and $L^h \to 0$ except in human maximum
- $L^m \to \infty$

5 Race Between Humans and Machines

Let us return to the multi-good setting with general production function that we described in our general model setup.

Proposition 1 ("Law of the Iron Population"). *Assuming long-run growth in A^m, we find that $MPL^h \to 0$ and $N^h \to 0$ except if one of the following three conditions is satisfied:*

1. *humans remain sufficient property (human reservation)*
2. *human labor is a complement to machine labor in the production of at least one of the goods that are essential for AIAs (non-substitutability)*
3. *none of the scarce factors required to produce human consumption goods are also valuable to AIAs (separability)*

The mechanism that endangers humans is the increasing competition for the scarce resources in the economy. The three scenarios of the proposition describe three ways in which this competition can be avoided: In scenario 1., humans remain in control over all resources, having fully enslaved the AIAs. In scenario 2., AIAs need humans because they are not substitutable by machine labor for some good that is needed for AIAs. In scenario 3., humans and AIAs live essentially in two different economies – AIAs do not value the scarce factors that are important for humans and so humans can prosper independently.

6 Hints of AIAs in Current Data

To search for harbingers of AIAs in our current economy, we follow a similar strategy as in the previous proposition, considering the implications of long-run growth in the productivity A^m of AIAs for factor and goods prices.

We first observe that such growth will disproportionately raise the prices and returns of factors most relevant for AIAs. This includes e.g. programmers and land in Silicon Valley, etc. Conversely, it will lead to a relative decline in the prices

and returns of factors that are most relevant for human absorption, which arguably includes most human labor, giving rise to a declining human labor share.

Focusing on economic growth, the price of a machine absorption basket will fall rapidly in comparison to the price of a human absorption basket. Measured from machine perspective, the economy thus experiences much faster real growth than from a human perspective and experiences far higher real interest rates. In short, the same economy will look like it is booming or stagnating depending on the perspective.

Lastly, an important harbinger of the advent of AIAs will be the accumulation of wealth and resources in these entities. Might the large corporate savings in the IT sector be the first hint of the advent of AIAs?

7 Conclusions

[to be completed]

References

Acemoglu, D. and Restrepo, P. (2018). The race between machine and man: Implications of technology for growth, factor shares and employment. *American Economic Review*, forthcoming. 1

Autor, D. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30. 1

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press. 1, 2.3, 2.3

Club of Rome (1972). The limits to growth. Technical report. 8

Ford, M. (2015). *The Rise of the Robots*. Oneworld Publications, UK. 1

Harari, Y. N. (2017). *Homo Deus: A Brief History of Tomorrow*. Harper. 2.3

Korinek, A. and Stiglitz, J. E. (2017). Artificial intelligence and its implications for income distribution and unemployment. *NBER Working Paper*. 1, 3

Korinek, A. and Stiglitz, J. E. (2018). Artificial intelligence, worker-replacing technological progress and income distribution. *Working Paper*. 1

Kurzweil, R. (2005). *The Singularity Is Near*. Viking. 2.3

Leontief, W. (1983). National perspective: The definition of problem and opportunity. In *The Long-term Impact of Technology on Employment and Unemployment: A National Academy of Engineering Symposium*. National Academies. 3

Malthus, T. (1798). *An Essay on the Principle of Population*. J. Johnson, London.

1, 2.1

Omohundro, S. M. (2008). The basic ai drives. In Wang, P., Goertzel, B., and Franklin, S., editors, *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, pages 483–492. Amsterdam: IOS. 1, 2.3

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf. 8

A Mathematical Proofs

To be completed.

B Extensions

B.1 Endogenous Factor Supply

Our baseline setup can easily be extended to account for endogenous factor supplies. Assume that the type i entities supply an endogenous ℓ^i units of factor per entity, which can again be interpreted e.g. as human or machine services. We denote the total factor supply of type i entities by $L^i = \ell^i N^i$, which enters the aggregate production function.

Type i entites now evolve according to a law-of-motion (in efficiency units) of

$$N^{i\prime} = G^i\left(x^i, \ell^i\right) N^i$$

with a growth function $G^i\left(x^i, \ell^i\right)$ that is concave, increasing in x^i and decreasing in ℓ^i, where the latter captures the cost of supplying the factors.

Some of our definitions now need to be generalized to include endogenous labor supply. The bundles of maintenance absorption now consist of pairs (s^i, ℓ^i) of absorption levels s^i of goods and factor supply ℓ^i such that $G\left(s^i, \ell^i\right) = 1$. Accordingly, the resource absorption frontiers and growth optimal preferences are also defined over both absorption levels and levels of factor supplies.