# Economic Possibilities for Our Children: Artificial Intelligence and the Future of Work, Education, and Leisure

**Miles Brundage**
Consortium for Science, Policy, and Outcomes, Arizona State University, PO Box 875603, Tempe, AZ 85287-5603
miles.brundage@asu.edu

**Abstract**

Many experts believe that in the coming decades, artificial intelligence will change, and perhaps significantly reduce, the demand for human labor in the economy, but there remains much uncertainty about the accuracy of this claim and what to do about it. This paper identifies several ways in which the artificial intelligence community can help society to anticipate and shape such outcomes in a socially beneficial direction. First, different technical aspirations for the field of AI may be associated with different social outcomes, increasing the stakes of decisions made in the AI community. Second, the extent of researchers' efforts to apply AI to different social and economic domains will influence the distribution of cognition between humans and machines in those domains. Third, the AI community can play a key role in initiating a more nuanced and inclusive public discussion of the social and economic possibilities afforded by AI technologies. To pave the way for such dialogue, we suggest a line of research aimed at better understanding the nature, pace, and drivers of progress in AI in order to more effectively anticipate and shape AI's role in society.

## Introduction

Economist John Maynard Keynes predicted the future rate of economic growth impressively well when he wrote the 1930 essay, "Economic possibilities for our Grandchildren." (Keynes 1972; Skidelsky and Skidelsky 2012). However, Keynes was incorrect in predicting a major decline in working hours in developed countries along with a corresponding rise in leisure. Skidelsky and Skidelsky (2012) attribute Keynes's failed prediction to a combination of three factors: the enjoyment people derive from some forms of work, social and economic pressures to work, and the insatiability of human desires. Now, contemporary commentators suggest that this time will be different, i.e. that average working hours and/or the fraction of people who are economically employable will decline in the next few decades. Judging from a recent survey of AI researchers about the future of their field (Müller and Bostrom, forthcoming), this debate may be quite urgent: experts expect AI to progress substantially in the next half century, well within the expected lifetimes of many people alive today and almost all children. In this paper, we argue that the AI community has a critical role to play in understanding and managing these issues more deliberately in collaboration with researchers in other disciplines, policymakers, and the public at large.

The remainder of the paper is organized as follows. First, we explore ways in which the long-term research goals and design choices of AI researchers may influence the social outcomes associated with AI. Second, we look at the application space of AI and ask whether certain applications should be accelerated or slowed for ethical reasons. Third, we explore the question of public engagement with AI, asking how the AI community and the broader public can collaboratively envision possible futures. Finally, we suggest the broad outlines of a research program aimed at shedding new light on the nature, rate, and drivers of progress in AI in order to shape it more effectively, and then we conclude with a summary of the paper.

## Goals and Designs

Different conceptions of the goals of AI research entail different methods of evaluation—consider, e.g., the four-fold taxonomy of human-like thinking, human-like action, rational thinking, and rational action described in (Norvig and Russell 2009). This section makes a related claim, namely that different long-term research goals and short-term choices in the agent design space are likely to correspond to different social impacts of AI. The remainder of this section explores some of the possible connections between choices in the goal and design space of AI and the social outcomes associated with the resulting technologies, but these are merely intended to be illustrative (for elaboration of the reasoning in this paragraph and examples of

goal/design considerations beyond those below, see: Brundage, forthcoming).

That technology should augment human cognitive faculties rather than replace them is a common refrain in popular discussions of AI. For example, Brynjolfsson and McAfee's (2014) influential work on the economics of AI and robotics suggests governments should incentivize research and development on human-complementary technologies over human-like technologies. This, they suggest, would result in more socially equitable outcomes than a world in which a shrinking portion of the population participates in the workforce, which they associate with the development of human-like AI. A related argument comes from Carr's (2014) recent book *The Glass Cage: Automation and Us*, which argues that a more automated world tends to result in the atrophying of important skills. Carr bases his conclusion on analysis of the so-called "substitution myth," the false but persistent belief that cognitive responsibility can be shifted from a human to a machine in a certain task without altering the nature of that task. Since "higher" cognitive functions and holistic understanding typically draw on hands-on physical or social interaction with the world, it is folly to expect to automate only the "lower," unimportant cognitive aspects of a domain. Carr suggests human-centered automation (incorporating, e.g. dynamic switching of control between human and machine to encourage skill developments) as an alternative to unreflective automation. At the same time, others such as Nilsson (2005) argue for a renewed focus on creating broadly intelligent, human-like agents (at least in some respects) based on the reasoning that there are certain tasks we want done by machines which require such high levels of intelligence. Some pertinent research questions suggested by this discussion are: what would be the social and economic consequences of the AI community emphasizing and successfully making progress on different research paradigms? What role should the government play in influencing research trajectories in AI? Would the human-centered and human-like paradigms, assuming they are distinct in the first place, converge in their social implications over the long term because more intelligent computers will be more user-friendly and vice versa?

Another possible relationship between (explicit or implicit) choices in the AI community and their broader social contexts is the following: the accessibility, transparency, affordability, and usability of AI innovations may influence the extent to which they tend to empower disenfranchised people or to entrench existing inequalities. If AI innovations are largely patented and fiercely protected by corporate interests, incomprehensible to non-experts, and draw on data or other resources that are only in private hands, different social consequences may result than if all AI innovations are immediately available to everyone in the form of well-documented open-source code, binaries, IDEs, demos, and other forms for non-experts to apply freely to new domains (these extremes are not the only possibilities, of course). Not all of the dimensions of variation discussed here (e.g. high level research goals, patent strategies, usability, and the availability of open source alternatives) are necessarily correlated, nor are they all under the control of individual AI researchers, but they are illustrative of the sorts of social considerations that may be relevant to the AI community in reflecting on the broader context of their work.

## Applications

As suggested in the discussion of Carr's work above, the decision to automate (a part of) a given task is not value-neutral or made in a vacuum. Indeed, what we should and shouldn't do with our time has long been a focus of ethics, and visions of utopia often prominently figure changes in the nature or amount of work performed by humans (Sargent 2010). As discussed in (Brundage, forthcoming), moreover, many domains have already been suggested as either urgently in need of AI and robotics innovations (such as elder care, manufacturing, and sustainability) and others have been portrayed by many as areas to avoid automating (such as the decision to use deadly force in warfare). These examples hint at the hidden complexity of the seemingly simple question: what should humans do and what should machines do? In this section, we give illustrative examples of connections between decisions in the AI community to develop AI systems for one domain rather than other and their corresponding social implications.

One of the many ways in which AI innovations will affect the future distribution of work in society is indirect: by reducing the need for labor input to produce goods and services in a given domain, AI innovations will tend to reduce the cost of some goods and services, increasing consumers' and investors' disposable income and indirectly making possible new jobs or entire industries (given certain assumptions about humans still being employable after those AI innovations). More directly, decisions by AI researchers to preferentially develop certain applications could affect the type and quality of services in the realm of education and leisure. The AI community, as suggested by Kolodner et al. (2014), can play a critical role in re-envisioning the future of high-quality education for all. AI's role in improving education seems particularly ethically essential if people around the world, partially as a result of AI, begin spending more time on education rather than work and/or need education to transition to a new career. Novel means of entertainment could also rise in relative importance as a focus of AI researchers as basic needs are better satisfied and greater economic productivity makes possible (though doesn't guarantee) a reduction in working hours.

Additionally, AI may play a role in augmenting or reducing the socio-economic impact of intelligence and wealth in life depending on whether it is sufficiently accessible and usable to a wide population. As Gottfredson (1997) summarizes, there is now a large and robust body of evidence indicating that one's level of intelligence strongly influence's one's prospects in life, though it is far from the only factor. In addition, Gottfredson and others have noted the rising complexity of everyday life, and the consequent rise in the contribution of intelligence to life outcomes. As the infosphere (Floridi 2014a) becomes more suffused with computation and more complex, tools will be needed to help individuals cope with that complexity. However, whether such tools will exist and how they will be controlled is an open question. Information technologies such as AI can heighten existing inequalities (if they are monopolized by the rich and powerful), and they can simultaneously serve as an equalizing force by disproportionately benefiting those who are most cognitively burdened by scarcity of time and resources. Both of these seem possible, though not equally desirable. As with the goal and design space considerations discussed above, the AI research community is only one among many actors responsible for making certain technologies exist and be widely available in the market, but its responsibilities and affordances in this regard are nonetheless significant and worthy of further analysis.

## Engagement

AI could hardly be a much hotter topic in public discourse today, having sparked conversations on everything from the future of work, education, and leisure to human extinction. Despite all this discussion, much remains unclear about which options we face as a society with respect to the development and governance of AI as well as the related question of which of these options are most consistent with widely shared ethical values. The task of improving the quality of this discourse both within the AI community and in the broader society can be separated into two interrelated sub-tasks (Brundage, forthcoming).

First, the AI community needs to help inform individuals and groups about credible facts and perspectives on AI and its social context that are relevant to their lives. For example, people and organizations need to know about the likelihood that a job will be possible to automate over a given timeframe if they are to make appropriate decisions about, e.g. the education and jobs one pursues. However, the complex social context of AI implies that AI researchers don't (yet) have all the information and analysis the public needs —producing such knowledge is a task for interdisciplinary research aimed at illuminating the connections between technical, social, economic, and policy factors, in which the AI community should play a leading role.

Second, the AI community should strive to be responsive to public values and goals as they relate to the issues described above. In other words, public engagement and dialogue on AI should be bidirectional, and should reflect a serious commitment to democratizing science and technology. David Guston (2004) summarizes the nature of such democratization as follows:

> Democratizing science does not mean settling questions about Nature by plebiscite, any more than democratizing politics means setting the prime rate by referendum. What democratization does mean, in science as elsewhere, is creating institutions and practices that fully incorporate principles of accessibility, transparency, and accountability. It means considering the societal outcomes of research at least as attentively as the scientific and technological outputs. It means insisting that in addition to being rigorous, science be popular, relevant, and participatory.

No simple formula for such democratization exists, but illuminating examples from other scientific and technological domains abound, as do analyses of the goals (Stilgoe and Lock 2014) and processes (Wilsdon and Willis 2004; Guston 2014) appropriate to public engagement with science. In broad strokes, this literature tells us that there are many approaches to engaging the public about science, technology, and the future, with different characteristics such as the resources and time required, the portion of the public engaged by the method, the role of experts in the process, the outcomes that can be reasonably expected, and the degree of maturity of the approach. An illustrative example of a well-understood and tested method is the consensus conference, a process for bringing together lay citizens with experts to discuss and debate issues at the intersection of science, technology, policy, and values. Such conferences, which have been implemented for decades (especially by the Danish Board of Technology), are structured in a way that is intended to promote informed dialogue and the surfacing of important ethical questions, public values, and risks that weren't previously considered. Similar events have been held, with methodological and organizational innovations, in the United States in recent years (Guston 2014), allowing researchers and policy-makers to develop a better understanding of the role that such events can (and cannot) play in democratic governance of science and technology. In summary, then, there is a substantial body of research and practice to build on in thinking through why and how to engage the public, and many of these methods may be useful in the case of AI in order to stimulate rich and informed public dialogue about preferred futures.

## Anticipating Progress in AI

So far, we have discussed the future of AI in broad terms, glossing over the specific timeframes in question. This has been intentional, since the question of timeframes is fundamental in thinking through the social outcomes of AI and deserves a section of its own. In particular, we argue in this section that in order to effectively grapple with the relationship between AI, work, education, and leisure going forward, we need an improved understanding of the nature, pace, and drivers of progress in AI.

Many debates about the future social benefits and risks of AI can be fruitfully reinterpreted as debates about the nature, pace, and drivers of progress in AI. Consider two sets of arguments: those about which jobs are safe from automation and those about whether AI may one day pose an existential threat to humanity. While seemingly quite different, what these debates have in common is that participants in them often draw on different (implicit or explicit) models of what AI is, how quickly it is progressing, and what is causing it to progress. For example, if one thinks that AI's increased adoption in recent years is mostly the result of restructuring, digitizing, and simplifying environments rather than a breakthrough in the intelligence of agents themselves and that no breakthrough is on the horizon (Floridi 2014a), then one will tend to focus more on those aspects of task/environment structuring and dismiss existential risk concerns. If human-level (or otherwise highly advanced) AI is 20 versus 80 years away, different policies for education, welfare, and other domains may be called for. Furthermore, Danaher (2014) has argued that the question of technological unemployment can be framed in terms of the relationship between human skills and their rate of development on the one hand and AI skills and their rate of development on the other. If this framing has any truth to it, the AI progress question is intimately related to the technological unemployment question.

Some uncertainty about the future of AI is inevitable and even desirable – indeed, if there were no uncertainty, this would imply a lack of human agency over the future. However, in our current position, improved understanding of the nature, pace, and drivers of AI progress would be extremely beneficial for anticipating and consciously steering the future distribution of cognition between humans and machines in preferred directions. In particular, clarity is urgently needed with regard to the susceptibility of jobs to automation. Technical progress is only one driver of technology adoption in industry, but in the case of AI approaching or exceeding human levels of performance in a particular area or a wide category of areas, it is highly important to monitor and theorize the rate of development and what is causing it. Theories of the future distribution of jobs abound, but it isn't always clear whether these findings are consistent with the views of AI experts, let alone whether they're predictively useful. For example, some have emphasized the current difficulty of building high levels of performance in social communication, perception, and dexterity into robots, and then use this as a basis for predicting future job vulnerabilities (Frey and Osborne 2013); others emphasize the routine versus non-routine nature of the task in question (Levy and Murnane 2004); still others have made forecasts based on extrapolation of early results in the AI literature and assumptions about the rate of maturation and adoption of AI technologies (Elliot 2014). Major AI textbooks such as (Norvig and Russell 2009) also discuss characteristics of environments that pertain to the difficulty of designing agents to thrive in them, such as partial versus full observability and the existence of other agents, but these environment characteristics have not yet been compared in any systematic way with the theories of future agent capabilities/weaknesses discussed above.

An improved understanding of progress in AI and its plausible futures would provide a better foundation for the AI community to engage credibly and confidently with the public about how AI can (and cannot) contribute to meeting certain societal goals. In this respect, the goal of a theory of progress in AI is not to predict future progress per se but to identify options for influencing the future in preferred directions. The line of research suggested here would ideally result in a (meta-)model of progress in AI that would meet the following criteria: providing a basis for developing scenarios for the future of AI and society, capturing the diversity of research traditions in AI and thus the diversity of possible future societal outcomes associated with AI, and anticipating the implications of various policy options such as different distributions of government funding across AI research goals or domains. A model or ecology of models meeting these desiderata could serve as an input to economic projections, scenario planning approaches, plausible science fiction, and other means of engaging the public about progress in AI and the choices we face at various scales with respect to the changing economy.

## Conclusion

The opportunities and risks of automating a large fraction of existing cognition and labor and consequently changing the nature of work, education, and leisure raises far more questions than can be answered in a single paper, or by a single academic community. For example, Floridi (2014b) puts equal emphasis on the “resource problem,” the question of how work and its social role can be transformed by technology, and the “political problem” of avoiding a lazy, ignorant constituency placated by “bread and circuses” in a post-work world, the latter of which hasn't been discussed at all in this paper. Given the stakes involved, improving

the caliber of analysis, discussion, and action on these topics and incorporating the perspectives of a wider range of people is urgently needed. If it is ultimately successful in this regard, the AI community may one day be able to say that it played a key role in grappling with the challenges Keynes (1972) alluded to when he wrote:

> Thus for the first time since his creation man will be faced with his real, his permanent problem—how to use his freedom from pressing economic cares, how to occupy the leisure, which science and compound interest will have won for him, to live wisely and agreeably and well.

## Acknowledgments

The author thanks an anonymous reviewer for helpful feedback and suggestions. This work was supported by the National Science Foundation under award #1257246 through the Virtual Institute of Responsible Innovation (VIRI). The findings and observations contained in this paper are those of the author and do not necessarily reflect the views of the National Science Foundation.

## References

Brundage, M. Forthcoming. Artificial Intelligence and Responsible Innovation. *Fundamental Issues in Artificial Intelligence*, ed. Müller, V. Berlin: Springer.

Carr, N. 2014. *The Glass Cage: Automation and Us*. New York: W.W. Nortion & Company.

Danaher, J. 2014. Are we heading for technological unemployment? An Argument. *Disinformation*. Accessed online October 14, 2014 at http://disinfo.com/2014/09/heading-technological-unemployment-argument/#sthash.NX60aXhM.uxfs

Elliot, S. 2014. Anticipating a Luddite Revival. *Issues in Science and Technology* Winter 2014.

Floridi, L. 2014a. *The Fourth Revolution—How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Floridi, L. 2014b. Technological Unemployment, Leisure Occupation, and the Human Project. *Philosophy and Technology* 27: 143-150.

Frey, C., and Osborne, M. 2013. The Future of Employment: How Susceptible are Jobs to Computerisation? Accessed online October 14, 2014 at http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Guston, D. H. 2004. Forget Politicizing Science, Let's Democratize Science! *Issues in Science and Technology* Fall 2004.

Guston, D. H. 2014. Building the capacity for public engagement with science in the United States. *Public Understanding of Science* 23(1): 53-59.

Gottfredson, L. 1997. Why *g* matters: The Complexity of Everyday Life. *Intelligence* 24(1): 79-132.

Horvitz, E., and Selman, B. 2009. Interim Report from the AAAI Presidential Panel on Long-Term AI Futures. Association for the Advancement of Artificial Intelligence. Accessed online October 14, 2014 at http://www.aaai.org/Organization/Panel/panel-note.pdf

Keynes, J. M. (1972). Collected writings vol. 9: essays in persuasion (2nd ed.). London: Macmillan.

Kolodner, J. et al. 2014. AI Grand Challenges for Education. *AI Magazine* Winter 2014. Association for the Advancement of Artificial Intelligence.

Levy, F., and Murnane, R. J. 2004. *The New Division of Labor: How Computers are Creating the Next Job Market*. Princeton: Princeton University Press.

McAfee, A., and Brynjolfsson, E. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.

Müller, V., and Bostrom, N. Forthcoming. Future Progress in Artificial Intelligence: A Survey of Expert Opinion. *Fundamental Issues of Artificial Intelligence*, ed. Müller, V. Berlin: Springer.

Nilsson, N. 2005. Human-Level Artificial Intelligence? Be Serious! *AI Magazine* Winter 2005. Association for the Advancement of Artificial Intelligence.

Norvig, P., and Russell, S. 2009. *Artificial Intelligence: A Modern Approach*. Third edition. Upper Saddle River: Prentice Hall.

Sargent, L. T. 2010. *Utopianism: A Very Short Introduction*. Oxford: Oxford University Press.

Skidelsky, R., and Skidelsky, E. 2012. *How Much is Enough? Money and the Good Life*. New York: Other Press.

Stilgoe, J., and Lock, S. 2014. Why should we promote public engagement with science? *Public Understanding of Science* 23(1): 4-15.

Wilsdon, J., and Willis, R. 2004. *See-through Science: Why public engagement needs to move upstream*. Demos.