

Artificial Intelligence, Robots, and Work: Is This Time Different?

As technological innovation has eliminated many types of jobs over the past few centuries, economies have evolved to create new jobs that have kept workers well employed. Is there reason to worry that the future will be different?

Over the past few years, there has been increasing public discussion about the potential impact of artificial intelligence (AI) and robots on work. However, despite the attention given to the issue, there has not been much progress on understanding whether or not AI and robots deserve such special treatment. Specifically, are AI and robots just like past technologies, causing shifts in the workplace but leaving the fundamental structure of work in place? Or are they unique in some way that suggests this time is different?

On the technology side, the discussion often focuses on surprising examples of tasks that AI and robots can now carry out, but without putting those examples in perspective. How do they compare to the full range of tasks at work?

On the economics side, the discussion often focuses on analyses of past changes caused by technology, but without demonstrating that they apply to the question at hand. How do we know that AI and robots will affect work in the same way that technologies have in the past?

Neither of these approaches moves the argument beyond a simple repetition of opposing conclusions that these new technologies either will or will not cause big changes to work.

To make some progress on analyzing the issue, it helps to pay close attention to the way these discussions end, which is often with an enthusiastic description of the new

jobs that will come from AI and robots. Such descriptions tend to involve jobs requiring critical thinking, creativity, entrepreneurial initiative, and social interaction. Finally, there is usually a statement about the need to improve education to prepare people for these new jobs.

What I want to suggest is that the way to tell if AI and robots are different from previous technologies is by thinking more carefully about the jobs that will exist in the future and the education that will be needed to prepare for them.

In principle, as long as there are some types of work left for people to do, we can build an entire workforce and economic structure around them. Two centuries ago, roughly 80% of the workforce was involved in agriculture. Since then, a succession of new technologies mechanized many agricultural tasks, and employment on farms steadily decreased. Today, only a few percent of the workforce are in agriculture. There is no problem imagining a similar transition over the next few decades, in which AI and robots automate a large portion of current jobs, and the displaced workers—or their would-be replacements in the next generation—shift to other types of work.

With respect to the functioning of a work-based economy, it is irrelevant which jobs remain for people. It could be, as many analysts argue, that most jobs in a few decades will involve nonroutine tasks, with the routine tasks largely

automated. However, the economy could still function if the technology instead allowed the reverse, eliminating the nonroutine tasks and leaving the routine tasks. That seems odd to us now, though it is what occurred in the nineteenth century, when early mechanization eliminated craftwork, replacing it with more standardized tasks.

Whether technology implementation results in the upskilling or downskilling of work has large consequences for the difficulty of the transition for the workers themselves. From a macroeconomic perspective, it is equally easy to imagine an economy requiring either more nonroutine tasks or more routine tasks. However, with respect to education and training, it is likely to be much harder to upskill a workforce to carry out more nonroutine tasks than it is to downskill a workforce to carry out more routine tasks. Indeed, in some cases, the degree of upskilling required may simply not be feasible.

To understand whether AI and robots are likely to alter the fundamental structure of work, we need to know whether these new technologies will require changes in work skills that are feasible or not. If the changes are feasible, it is most likely that the overall effect of AI and robots will look like the changes we have seen with other technological innovations. However, if the necessary changes in work skills are not feasible, then it is most likely that this time will be different.

The skill of literacy

To illustrate the point, consider literacy, a basic skill that is widely used at work. Three-quarters of US workers use their literacy skills every day at work, reading materials such as emails, directions, or reference manuals. In addition, literacy is also a key foundational skill in many occupations for more advanced or specialized tasks involving reasoning or problem solving. This is particularly true of managerial, professional, and technical occupations that tend to involve extensive use of information.

Because of the importance of literacy in the economy, we collect data on the literacy proficiency of the workforce. A few decades ago, the United States collected these data using a national test of adult literacy. Today, the Organisation for Economic Cooperation and Development (OECD) conducts an international version of this test in its Programme for the International Assessment of Adult Competencies (PIAAC), which tests a representative sample of working-age adults in each participating country. The adults selected for the sample usually take the test at home, as part of a survey administered by a trained interviewer. The survey, which usually takes between one and two hours, includes a variety of background questions in addition to the test.

Unlike tests given to students, PIAAC specifically uses test questions that aim to represent tasks adults might encounter outside school, either in their personal lives or at work. The goal is to test adults using practical tasks that are similar

to some of the real tasks that adults need to carry out using literacy. To understand what the test is like, it is helpful to describe two examples of the literacy questions.

The first example relates to the task of placing an international telephone call. The text provides information on making an international call in the form of a website you might find if you conducted an online search for help in making such a call. The question asks when you need to use a particular calling code, based on the information on the website.

The second example relates to the task of finding a book. The text provides the results of a search for books about genetically modified foods, with each entry showing a short description of a book. The question asks for the book that asserts that there are problems with the arguments on both sides of this controversial topic. To answer the question, the respondent needs to figure out which of the book descriptions indicates a book that looks critically at the arguments for and against the use of genetically modified foods.

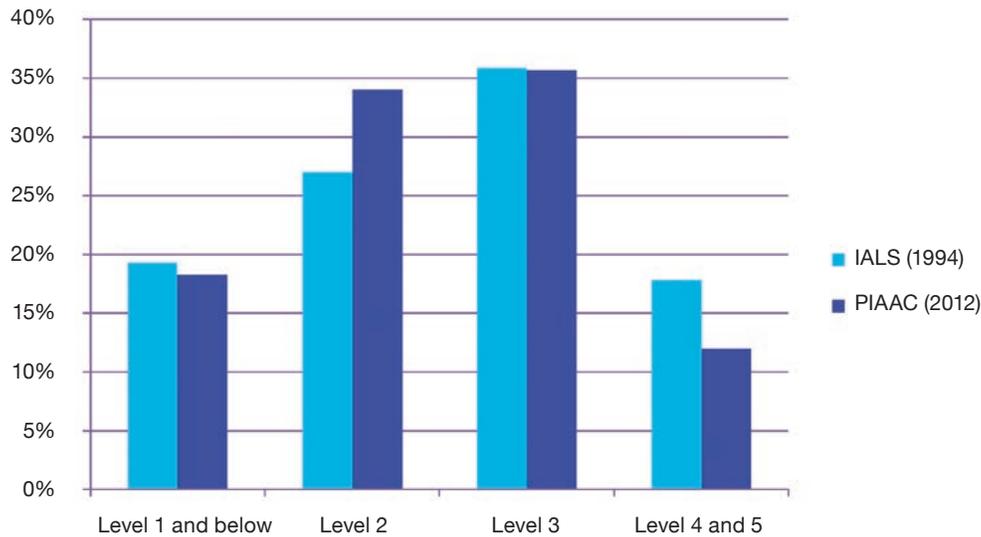
The scoring process for PIAAC grades each question in terms of five levels of difficulty, with the higher levels involving questions that are more difficult. The international call question is an example of a question at Level 3, the middle level of difficulty. The book search question is more difficult and is an example of a question at Level 4. PIAAC also describes the proficiency of adults using the same five levels. Adults score at the level where they can answer the questions correctly about two-thirds of the time. For example, adults at Level 3 will be able to answer the Level 3 questions correctly two-thirds of the time. They will be better at questions at Level 2, answering them correctly over 90% of the time. They will be worse with questions at Level 4, answering them correctly less than 30% of the time.

Using this scoring system, Figure 1 shows the PIAAC literacy results for adults in the United States in 2012, comparing them with the results from the earlier International Adult Literacy Survey (IALS) conducted in 1994.

Half of US working-age adults score at Level 2 or below on the PIAAC literacy test, meaning they cannot reliably answer either of the two example questions described above. The other half score at Level 3 or higher and can reliably answer questions such as the international call example. Only 12% of US adults score at Level 4 or Level 5 and can reliably answer questions such as the example about searching for books about genetically modified foods. These results are quite similar to the average results for other developed countries.

The comparison of the results for IALS and PIAAC shows that the literacy proficiency of US adults has declined over the past two decades. There are now fewer adults at Levels 4 and 5 and more adults at Level 2. Compared with the 1990s, a smaller percentage of US adults can now reliably answer questions such as the two examples

Figure 1. Literacy proficiency for working-age adults in the United States, 1994-2012



described above. Other developed countries show a similar pattern, though the shift in the United States is larger.

The PIAAC literacy results provide a way of thinking concretely about possible changes in work skills. If most jobs in a few decades will need lower levels of literacy—comparable to Levels 2 and 3 on the PIAAC scale—then there will be no problem for the workforce to adjust because most adults already have those skills. However, if most jobs will require Level 4 literacy, it is clear that most members of the workforce would face a challenge. The first hurdle would be stopping the decline in literacy that has been taking place over recent decades.

Computer capabilities

To think more about the skill change that could be required for the workforce in the next few decades, it is helpful to consider what capabilities computers themselves will be offering with respect to literacy. To do that, I worked with a group of 11 computer scientists to evaluate the PIAAC literacy test questions to determine whether or not computers could answer them using state-of-the-art techniques in AI.

To carry out the exercise, the participating computer scientists first decided on the ground rules for their judgments. Those judgments focused primarily on the capabilities of existing techniques, not on techniques that might be available in the future. In addition, their judgments did not address whether there are already computer applications that could answer the PIAAC literacy questions. Instead, the computer scientists focused simply on the power of current techniques as demonstrated in research settings.

The group did consider the possibility that adapting current computer techniques to the PIAAC tasks would require some further development. However, they limited such development to a modest level of effort, similar to a

year's worth of work by several people at a cost of no more than \$1 million. One can think of this as the development effort a large company might carry out to automate a set of literacy tasks performed frequently by its workers.

After agreeing on the ground rules, the computer scientists provided their judgments individually about the feasibility of computers answering each of the test questions and then they discussed their judgments as a group. Finally, we averaged the results across the group and compared the pattern of expected performance with the results for adults who take the test.

Overall, the computer scientists indicated that computer proficiency in literacy would be similar to adults at Level 2. In addition to analyzing the average responses across the group, we also analyzed the differences in judgments across the group. These additional analyses considered the effect of excluding particular test questions for an expert if the expert was unsure, focusing only on the test questions where the experts showed high levels of agreement, or excluding the experts with more extreme views. These variations all produced the same conclusion.

However, we also wanted a reasonable upper bound on current computer capabilities. To obtain that, we did two things. First, we tried an analysis coding each question as possible for current computers if at least three members of the group thought that it was. This approach allowed for the possibility of new techniques that only a few members of the group might know about. Second, three members of the group provided answers related to what computers will be able to do in 10 years, at the limit of any realistic planning horizon for current research programs. The two routes to describing an upper bound on current capabilities both suggested that computers are close to Level 3 performance on the literacy test.

Obviously, there are limits to this analysis, since it relies on the judgments of a group of experts rather than the

performance of a set of working computer systems. However, the computer scientists who participated in the exercise are prominent in the field and broadly familiar with the current literature on computer systems that work with natural language, so they should be in a good position to gauge the state of the art. One benefit of having the experts directly analyze the test questions came in their explanations of the nature of the difficulties that occurred with the harder questions.

The big contrast was between questions where keyword search techniques can produce the answer and questions that require fully understanding the language involved. The two PIAAC literacy examples described earlier illustrate this contrast in question difficulty. It is possible to get close to answering the international call example simply by searching the text for the calling code that appears in the question. In contrast, no keywords in the question for the book search example can identify the right book. Instead, finding the answer to the second example requires a full understanding of the meaning of the question and the descriptions of the individual books.

The two example questions show a key contrast that illustrates the limits of current techniques in computer science. Computers can answer literacy questions that rely primarily on search techniques, but questions that require full language understanding are still too difficult for current AI techniques.

Implications for work

The investigation of AI techniques for literacy shows that current computers are clearly limited. However, the PIAAC results for people show that many people have difficulty with the same questions that are hard for computers. What does this mean for the potential impact of AI on work?

Literacy is a skill that many adults use at work. In addition to its direct applications, literacy is also a foundation skill that is critical for many specialized reasoning and problem-solving skills in specific domains. This suggests that computers may be able to carry out many of the information-related tasks currently performed by workers with literacy proficiency at Levels 2 or 3. It also suggests that computers may not be able to perform many of the information-related tasks performed by workers with literacy proficiency at Levels 4 or 5. Of course, without substantial skill development, many workers with lower literacy levels may not be able to perform those tasks either.

As a society, we make large investments in providing everyone with education to develop his or her literacy skills, along with the associated reasoning and problem-solving capabilities. However, despite years of educational preparation, many adults achieve only limited literacy proficiency. Although the basic PIAAC results offer a rather pessimistic assessment of adult skills, further considerations

provide reason to hope that we can do better in preparing adults with higher-level proficiency in literacy and related information-processing skills.

First, PIAAC suggests, not surprisingly, that education makes a difference in literacy skill. Of US adults with a postsecondary degree (two-year degree and higher), 24% perform at Levels 4 or 5. Of course, this simple relationship alone does not prove causality, but other types of analyses support the common sense inference that education does indeed play a causal role in improving cognitive skills. This points the way to improving proficiency by increasing education.

Second, PIAAC suggests that the quality of education makes a difference in literacy, with some countries showing much better results than others do. For example, 22% of adults in Finland and 23% of adults in Japan are at Levels 4 or 5, compared with 12% in the United States. This points the way to improving proficiency by changing the education system to make it more effective.

Combining the effects of the quantity and quality of education suggests what may be possible at the limits of what we know how to do with large-scale education systems. For adults with postsecondary education (two-year degree and higher), 36% in Finland and 37% in Japan are at Levels 4 or 5.

These comparisons offer hope that the United States can do better, but the hope is modest with respect to the full labor force. Furthermore, we need to acknowledge that aggregate literacy levels have been moving in the wrong direction over the past few decades, despite efforts to improve education during this period. Recent history suggests that improvement in the literacy of the overall adult population is likely to be both difficult and slow.

These results raise cautions about any projections that have a large percentage of workers performing tasks involving critical thinking or creativity. Although the PIAAC literacy assessment is certainly not an assessment of those more advanced skills, it is hard to imagine adults being able to carry out meaningful levels of critical thinking or creativity without also having enough literacy to answer questions such as the book search example described above.

We know that three-quarters of US workers currently carry out tasks using literacy skills at work but that the literacy proficiency of most of these workers is at a level that computers are close to achieving. It is reasonable to expect that employers will automate many of these literacy-related tasks over the next few decades by applying the computer techniques that already exist to do so. However, our experience with educational improvement suggests that it is not reasonable to expect that the jobs of most of these workers will be able to shift toward tasks involving much higher levels of literacy and related skills, comparable to Levels 4 and 5 in PIAAC. Such a change would likely not be feasible to carry out over a few decades.

That suggests that many workers will need to switch instead toward other kinds of skills. In the abstract, this is a straightforward statement. However, whichever skills are suggested, we are faced immediately with two practical questions: how proficient are computers with respect to these other skills, and how many people are more proficient than computers are? The key issue is whether other skill areas look like literacy. For other skills, can most people do what computers cannot do, or do most people have trouble with those tasks as well?

Take social skills, for example. Occupations involving high levels of social interaction are often proposed as promising occupations for the future because people are assumed to be good at social interaction whereas machines not so much.

However, if we look closely at social skills, we are likely to find that we are overestimating the abilities of people and underestimating the abilities of computers. Most people are capable of simpler aspects of social interaction, such as facial recognition or responding to direct requests for information. However, computers also now have these capabilities. Of course, computers cannot perform more complex social interactions, such as conducting a sensitive negotiation or gaining the trust of an angry customer. However, experience suggests that those complex social interactions are also too difficult for many people.

The question for social skills—and for other major skill areas—is how computer capabilities compare with the distribution of human proficiency. Thinking in terms of literacy, computer performance at PIAAC Level 1 is not very threatening because few adults have literacy that low, but computer performance at Level 3 is quite threatening because few adults are better than that.

If computer proficiency in other major work skills looks more like PIAAC literacy at Level 1, then the adjustment to AI and robots should be similar to past technologies, at least over the next few decades. In that case, even if automation displaces many workers from some tasks, there will still be many tasks using other work skills that most people can perform.

However, if computer proficiency in other major work skills looks more like PIAAC literacy at Level 3, then AI and robots are likely to have a different effect than did past technologies. In that case, as automation displaces workers from some tasks, they are likely to find that the tasks in the jobs that remain are much more difficult to perform, and they may have trouble acquiring those skills.

So what about social skills: is computer proficiency with respect to social skills more like Level 1 or Level 3 in PIAAC literacy terms? Unfortunately, we have not yet looked carefully enough at human and computer capabilities to know the answer to that question, either for social skills or for many other major skill areas. Policy-makers should be directing researchers to answer that question.

Comparing computers and humans

In the coming years, we need to have a much better understanding of how the capabilities of computers and humans compare. In making this comparison, it is critical to consider the distribution of proficiency across the workforce for different skills, as well as the realistic potential for increasing the proficiency of the workforce for those skills where computers have already made substantial progress. It is not enough to say that *some* people have better skills than those provided by computers. If we are going to be able to continue a work-based economy, we need to know that *most* people can develop better skills than those provided by computers.

We know from the literature on the diffusion of technology that it often takes a substantial amount of time for industry to adopt and apply new technologies—time to learn about the technologies, refine them for particular applications, and invest in the technologies at scale. In many cases, widespread diffusion can take several decades. That means we have time to understand what computer capabilities currently exist and anticipate how they are likely to shift the skills needed by the workforce over the coming decade or two. However, we also know that improvements in education are often slow and difficult. So even a decade or two of warning may not be enough to develop the skills needed.

Much of the recent public conversation about AI, robots, and work has focused on observations by computer scientists about the technology and observations by economists about past changes in the labor force. However, these two perspectives are not sufficient for us to understand the likely effect of new computer capabilities on work and the way we need to respond. Crucially, we also need to hear from three other types of experts to understand whether AI and robots will cause a fundamental change in the nature of work and its role in the economy. First, we need to hear from psychologists to understand the capabilities that people have. Second, we need to hear from testing experts to understand the distribution of proficiency across the workforce for different types of skills. Finally, we need to hear from educators to understand what we know about improving human proficiency.

Only when we have assembled the insights from these three different perspectives on the capabilities of AI and robots—in addition to those from computer scientists and economists—will we be in a position to know whether this time is different.

Stuart W. Elliott is a scholar in the Division of Behavioral and Social Sciences and Education of the National Academies of Sciences, Engineering, and Medicine. He is the author of Computers and the Future of Skill Demand (OECD, 2017).