DIRECTORATE FOR EDUCATION AND SKILLS
CENTRE FOR EDUCATIONAL RESEARCH AND INNOVATION (CERI)
GOVERNING BOARD

Future of Skills: Understanding the Educational Implications of AI and Robotics

100th Session of the CERI Governing Board, 9-10 April 2019

This room document reports on the planning meeting for this project that was held on 27th March and sets out next steps for the work.

For further information, please contact Deborah Roseveare, Head, Innovation and Measuring Progress Division (Deborah.Roseveare@oecd.org) or Stuart Elliott, National Academies Scholar (SElliott@nas.edu).

JT03445906

Future of Skills: Understanding the Educational Implications of AI and Robotics

Background

1.   For the 2019-2020 Programme of Work and Budget, CERI approved a project on the Future of Skills to develop assessments of artificial intelligence (AI) and robotics. These assessments will support forward-looking educational policies by helping policymakers anticipate potential future changes in skill demand that could be caused by AI and robotics. The long-term goal of the project is to develop an approach for assessing AI and robotics capabilities that could be used in an ongoing assessment programme, providing regular updates as the capabilities evolve.

2.   To ensure a high level of scientific credibility in the assessments, the project proposal called for a collaboration with the U.S. National Academies of Sciences, Engineering, and Medicine (the U.S. National Academies), and with the national academies of science or equivalent bodies of other countries as well. To address this aspect of the proposal, CERI held a joint planning meeting on 27 March with the U.S. National Academies and the InterAcademy Partnership, a global network of over 130 academies in science, engineering and medicine. The meeting was held with partial support provided by the Alfred P. Sloan Foundation, along with support from CERI's approved budget for the project and internal programme development funds from the U.S. National Academies.

3.   The planning meeting was an opportunity to develop a shared understanding of the project's goals and approach across the three organisations, with opportunities to refine the initial plans about the project's activities and to build support for obtaining funding. In addition to including representatives of the CERI Secretariat and Governing Board, the meeting included staff and oversight group members for the U.S. National Academies and the InterAcademy Partnership, as well as a number of outside experts and stakeholders. The range of expertise included education research and policy, testing, survey methodology, computer science, cognitive science, engineering, economics, political science, and sociology. Participants came from Canada, Estonia, Finland, France, Germany, Norway, the United Kingdom, and the United States. The meeting was chaired by Andy Wyckoff, director of OECD's Directorate for Science, Technology and Innovation. Andy provided a connection to the other OECD efforts related to AI and robotics, as well as an understanding of the U.S. National Academies from his volunteer service in chairing a consensus study for them concerning science and technology indicators.

4.   The planning meeting was divided up into sections that addressed the overall goal of assessing AI and robotics capabilities, the basic approach for carrying out the assessment, and the roles for the three institutions. The day ended with a discussion of next steps, with implications for the project activities that would be carried out by CERI with the initial support approved for the current biennium. This summary covers these four topics in turn, with a few highlights of the discussion in each case.

The Overall Goal of Assessing AI and Robotics Capabilities

5.   The motivation for the Future of Skills project comes from a conviction that policymakers need to understand what AI and robotics can do—with respect to the skills people use at work and develop during education—as one key part of understanding how they are likely to affect work and how education should change in anticipation. The project will carry out a set of specific studies to develop a feasible way to assess AI and robotics capabilities across different human skill areas, with the hope that this approach will lead to an ongoing programme to provide regular updates as those AI and robotics capabilities continue to evolve.

6.   One of the central points of discussion in the first session at the planning meeting related to the idea that the Future of Skills project can provide a basic building block for analysing and implementing policies in several different spheres. The participants understood CERI's interest in the large-scale implications of AI and robotics for skill demand and education planning. At the same time, other participants noted the importance of using data on AI and robotics capabilities for other purposes, including near-term labour force policies and efforts to provide augmentation technologies to support the elderly. In each case, the work to draw out specific implications would likely be carried out by other groups that understand the different contexts, but the basic information on AI and robotics capabilities would provide a common starting point.

7.   A number of participants commented on the fact that a common understanding of AI and robotics capabilities is simply not available today for policy discussions of any type. As a result, many policy discussions related to the implications of AI and robotics start from a weak foundation, guided only by assumptions. The participants agreed that the overall goal of the project would be to correct this problem, providing a valid knowledge base about the capabilities of the technology itself.

8.   There was some disagreement in the discussion about the project's focus on using human skills as the framework for determining which AI and robotics capabilities to assess. Some participants understood this approach as providing a natural way to ensure that the project works on technologies that are likely to affect tasks that humans might otherwise carry out, with implications for changing skill demands. However, other participants thought it might be useful to assess AI and robotics capabilities of all types, both those that are similar to specific human skills and those that are quite different. The discussion at the meeting did not resolve this disagreement. However, one pragmatic response to this point for planning purposes might be to start with assessments focused on AI and robotics capabilities related to human skills, while leaving open the possibility of expanding to other types of AI and robotics capabilities in the future.

9.   Another point discussed during the session on the overall goal concerned the availability of information about the distribution of human proficiency in different skill areas. Since the goal is to provide a way of comparing AI and robotics capabilities to human capabilities, it is important to understand the relevant range of differences on the human side. This concern could be addressed by ensuring that the project chooses human tests that have results for appropriate reference groups. For many tests, the available comparison information would be for specific groups of people rather than the full population—such as workers in particular occupational groups for tests related to occupations, or students in particular education programs for tests related to education—but this limitation in the comparison population may well make the comparison more useful rather than less useful for considering potential policy implications.

The Basic Approach for Carrying out the Assessment of AI and Robotics Capabilities

10.   The project's basic approach for carrying out the assessment is to use existing human tests in different skill areas to assess the capabilities of AI and robotics. Human tests will be used because they provide a widely available tool for concretely specifying the capabilities of AI and robotics across the full range of human work skills, in a form that can be directly compared to results for humans.

11.   Pilot work for the project was carried out using OECD's Survey of Adult Skills (PIAAC) to compare AI and human capabilities related to literacy and numeracy. This work was supported by CERI in the 2015-2016 biennium, and resulted in the CERI report *Computers and the Future of Skill Demand* (Elliott, 2017). The Future of Skills project extends this approach across a broader range of skills and tests.

12.   As a preliminary step, it will be necessary to review available skill taxonomies and the available tests of different skills to identify a set of skills and tests—roughly spanning the full range of human skills used in the workplace—to assess and compare AI and robotics capabilities. Since the credibility of the project rests on a careful selection of a set of representative skills and valid tests, this preliminary step is particularly important.

13.   The discussion during the second session of the planning meeting included a number of observations about possible tests that could be used. The discussion started off with a comment by one of the computer scientists that many of the current assessments of AI capabilities that are reported in the press are misleading because they use tests that provide a very limited comparison with human capabilities. A number of participants made comments about the different types of human tests that could be used. For example, in the area of language, the available human tests range from tests of isolated language components (such as vocabulary or grammar), to tests of a single language modality (such as reading comprehension), to tests of a complex, integrated task (such as a work-related simulation requiring the test-taker to review multiple written documents and video clips to prepare for a simulated oral interview with a journalist). There was general agreement across the group that a more valid assessment of AI and robotics capabilities—for use in comparing with human capabilities—would probably be provided by using human tests composed of more integrated tasks that combine multiple skills into a more authentic and complete performance.

14.   Building on the pilot work using PIAAC, the primary assessment approach for the project is to ask experts to evaluate the ability of current AI and robotics techniques to answer different test questions, rather than to commission research teams to try to develop working computer systems to answer the questions or hold competitions. A number of participants expressed a desire to have empirical demonstrations as well, though they understood that it would be cheaper and faster to evaluate tests using expert judgment. A potential compromise solution emerged in the discussion, concerning the possibility that the project can indirectly encourage research groups to provide empirical comparisons for free simply by highlighting specific tests that would provide valid and particularly interesting comparisons with human capabilities. In this way, the project could prompt the field to adopt the identified tests as test beds for use by different research groups in evaluating the performance of their AI and robotics systems.

15.   With respect to the expert judgments that the project will use directly, a number of interesting questions were raised about how experts would be sampled, with respect to specialty, affiliation, and perspective. In particular, participants noted the importance of

including perspectives from researchers in the private sector and researchers in China—in both cases because of their large impact on the field of AI and robotics, and the possibility that the perspectives and specific knowledge of researchers in those settings may be substantially different than researchers in universities.

The Roles for the Three Organisations in Carrying Out the Work

16.   The Future of Work proposal approved by CERI for the current biennium calls for some initial preparatory work to identify skills and tests, followed by exploratory assessments in two major skill areas. The proposal further expresses the hope that exploratory studies carried out by CERI would be followed by complementary consensus studies carried out by the U.S. National Academies and similar bodies in other countries, with the goal of providing credible and validated assessments that reflect a broad consensus of experts in the field.

17.   The third session of the planning meeting discussed the complementary nature of the possible work by the three organisations. OECD is able to carry out relatively fast-paced studies that are led by the Secretariat and can be authored by individual staff officials. Such studies can be used to gather information about the range of views of AI and robotics capabilities with respect to a particular test, as was done in the pilot work using PIAAC. That range of views will indicate where consensus lies with respect to specific test questions that clearly can or cannot be addressed with current AI and robotics techniques. However, there will also be disagreements about some of the test questions, concerning capabilities that are at the research frontier, where individual experts will have different perspectives about the latest findings. In some cases, the disagreements will have practical policy implications with respect to the estimated number of people who may be affected by the technologies. In such cases, a more thorough expert consensus process will be needed to attempt to resolve the disagreements about exactly which tasks can currently be addressed by AI and robotics techniques. The U.S. National Academies and the InterAcademy Partnership carry out consensus studies of this kind.

18.   In addition to the specific work of developing the assessments of AI and robotics capabilities in different skill areas, the planning meeting also discussed two other roles that OECD might play in the overall development effort. First, as an organisation that brings together policymakers in different sectors and across different countries, OECD is well placed to help define the questions that experts should be trying to answer in order to allow policymakers to understand the implications of AI and robotics. One way to think about this process of defining the questions is in terms of the different human tests that could potentially be used to assess AI and robotics capabilities. The pilot test for the project used PIAAC, which assesses literacy and numeracy in adults across the full adult population using relatively practical tasks. That produced a comparison of AI capabilities with general literacy and numeracy proficiency across the full adult population, which provides insight in thinking about potential AI implications with respect to the entire education system and a few types of tasks in many different occupations and everyday activities. A different type of insight would be provided by using human tests that are linked to specific education programs or specific types of work. As the project proceeds, it will be useful to understand which types of comparison provide the most insight to the policy community.

19.   A second potential role for OECD to play in the overall development effort is by starting to consider how to establish an ongoing programme for carrying out the assessments on a regular basis. OECD is a more natural home for an ongoing programme— in contrast to the U.S. National Academies and the InterAcademy Partnership—because

OECD has a number of ongoing programs, whereas the other two organisations focus on one-time projects. Of course, an ongoing programme would go well beyond CERI's commitment to the Future of Work project, but it might be a possible addition to the larger stream of AI-related work being led by OECD's Directorate for Science, Technology and Innovation.

20.   With these considerations in mind, it may be useful to think about the collaboration in terms of the pilot project using PIAAC. The first step involved exploratory work carried out by CERI that indicated some clear areas of consensus about literacy and numeracy questions that current AI techniques could clearly answer or could clearly not answer. That first step indicated a consensus that current AI techniques are above level 1 and below level 4 in literacy and numeracy as measured by PIAAC. However, there was disagreement about whether the capabilities at the research frontier lie at level 2 or level 3—and that level of disagreement has big implications for how many adults have literacy and numeracy capabilities that are no better than computers. To resolve that disagreement, it would be helpful to carry out a more careful consensus study, along the lines of the studies carried out by the U.S. National Academies and the InterAcademy Partnership. In addition, to helping resolve the expert disagreements about the current level of AI capabilities in literacy and numeracy, that consensus study could also identify a set of key tasks to monitor in the research literature to determine when the next substantial increase in AI capabilities for literacy and numeracy occurs. That monitoring process could potentially become a part of the AI-related work being led by OECD's Directorate for Science, Technology and Innovation.

Next Steps

21.   The planning meeting concluded with a discussion of next steps to move the project forward. Much of the discussion concerned the importance of drafting a scoping paper that would describe the full project, expanding beyond the specific CERI role in the current biennium to outline all the steps that would be needed by the different organisations and how they fit together. The group recommended that the scoping paper specifically outline an incremental approach, with initial activities in particular areas that would help demonstrate the value of the approach and then lead to support for more comprehensive efforts at later stages. The group recommended describing potential use cases in the scoping paper, with the range of potential areas of policy insight and implications suggesting that several different potential use cases should be described. For practical purposes, the group also recommended that the scoping paper should map other related initiatives—to indicate pre-existing work the project might build on and where it can most contribute—and that it should include planning for the intended ongoing programme, even though that would not be created for a number of years.

22.   The participants had only a brief discussion of the funding requirements for the full development project, which have been previously outlined as being roughly 2 MEUR annually over six years, substantially greater than CERI's initial annual support of 400 KEUR. However, several participants noted that the continued and growing concerns related to the potentially disruptive role of AI and robotics have heighted policy interest in recent years. Many felt that it should be possible to find support for the full project, once a scoping paper is available.

23.   After the planning meeting, the OECD Secretariat met with the staff of the U.S. National Academies to outline the set of initial activities that would likely be included in the scoping document and targeted for immediate fundraising and work during the 2019-2020 biennium. On the CERI side, these activities would involve the following activities, if additional VC funding beyond the Part II funding is obtained:

- Exploratory assessment using an assessment instrument aligned with established professional competencies for a discipline such as engineering. The International Engineering Alliance has established Graduate Attributes and Professional Competencies and provides robust accreditation processes for education programmes to underpin mutual recognition of qualifications. Entry into the profession typically involves rigorous assessments, so test items should be reasonably straightforward to obtain.

- Exploratory assessment using an assessment instrument for different types of social interaction. Social interaction occurs in many diverse settings such as health and social care through to property management. Assessment instruments for these skills might be found in the assessment processes used in the World Skills Competition or in jurisdictions that include such assessments within their professional education and training qualifications.

- Small working group of CERI members to discuss the potential insights from assessments of AI and robotics capabilities using different types of tests.

- Literature review of existing comparisons of AI and robotics with humans.

24.   This OECD work would be complemented by the following work carried out jointly by the U.S. National Academies and the InterAcademy Partnership, if sufficient funding is obtained:

- Consensus study assessment of AI literacy and numeracy capabilities using PIAAC.

- Expert meeting to identify the range of human tests that could be used in the project.

- Expert meeting to discuss potential approaches for sampling and involving an appropriate range of experts in AI and robotics.

25.   If support is obtained, these initial activities would all be carried out over the current biennium, with the intention to hold a second joint planning meeting in mid- to late-2020 to discuss the results of these initial activities and refine the plans for the next stage.

26.   Obviously, this is an ambitious timeline, given the additional fundraising that would need to be done, along with the associated steps to establish activities at the different institutions. However, the pace and scale of the proposed activities for the coming 18 months responds to enthusiasm expressed at the planning meeting and the sense among the participants that it should be possible to find the required support, given the high level of current interest in questions related to AI and robotics capabilities.