European Commission

JRC Technical Report

Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks

JRC Working Papers Series on
Labour, Education and Technology
2020/02

Songül Tolan, Annarosa Pesole, Fernando Martínez-Plumed,
Enrique Fernández-Macías, José Hernández-Orallo, Emilia Gómez

LABOUR EDUCATION TECHNOLOGY

Joint
Research
Centre

Contact information
Name: Songul Tolan
Address: Joint Research Centre, European Commission (Seville, Spain)
Email: Songul.TOLAN@ec.europa.eu@ec.europa.eu
Tel.: +34 9544-88354

EU Science Hub
https://ec.europa.eu/jrc

JRC119845

Seville: European Commission, 2020

How to cite this report: Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., Gómez, E. *Measuring the Occupational Impact of AI:Tasks, Cognitive Abilities and AI Benchmark*, Seville: European Commission, 2020, JRC119845.

Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks

Songül Tolan[1], Annarosa Pesole[1], Fernando Martínez-Plumed[1], Enrique Fernández-Macías[1], José Hernández-Orallo[2,3], and Emilia Gómez[1,4]

[1]Joint Research Centre, European Commission
[2]Universitat Politècnica de València
[3]Leverhulme Centre for the Future of Intelligence
[4]Universitat Pompeu Fabra
{songul.tolan, annarosa.pesole, fernando.martinez-plumed, enrique.fernandez-macias, emilia.gomez-gutierrez}@ec.europa.eu, jorallo@dsic.upv.es

Abstract

In this paper we develop a framework for analysing the impact of AI on occupations. Leaving aside the debates on robotisation, digitalisation and online platforms as well as workplace automation, we focus on the occupational impact of AI that is driven by rapid progress in machine learning. In our framework we map 59 generic tasks from several worker surveys and databases to 14 cognitive abilities (that we extract from the cognitive science literature) and these to a comprehensive list of 328 AI benchmarks used to evaluate progress in AI techniques. The use of these cognitive abilities as an intermediate mapping, instead of mapping task characteristics to AI tasks, allows for an analysis of AI's occupational impact that goes beyond automation. An application of our framework to occupational databases gives insights into the abilities through which AI is most likely to affect jobs and allows for a ranking of occupation with respect to AI impact. Moreover, we find that some jobs that were traditionally less affected by previous waves of automation may now be subject to relatively higher AI impact.

Keywords: artificial intelligence, occupations, tasks

Authors: Songül Tolan, Annarosa Pesole, Fernando Martínez-Plumed, Enrique Fernández-Macías, Emilia Gómez (Joint Research Centre, European Commission), José Hernández-Orallo (Universitat Politècnica de València)

Acknowledgments: We thank participants of the RENIR workshop 2019, as well as Vicky Charisi, Bertin Martens, and Cèsar Ferri for fruitful discussions on earlier drafts of this paper.

Joint Research Centre reference number: JRC119845

Contents

1 Introduction

There is a wide agreement that the latest advances in Artificial Intelligence (AI), driven by rapid progress in machine learning (ML), will have disruptive repercussions on the labour market (Shoham et al., 2018). Previous waves of technological progress have also had a sustained impact on labour markets (Autor and Dorn, 2013), yet the notion prevails that the impact of ML will be different (Brynjolfsson et al., 2018). An argument that supports this notion is that ML seems to circumvent the previously hard limit to automation known as Polanyi's Paradox (Polanyi, 1966), which states that we humans ``know more than we can tell''. While past technologies could only automate tasks that follow explicit, codifiable rules, ML technologies can infer rules automatically from the observation of inputs and corresponding outputs (Autor, 2014). This implies that ML may facilitate the automation of many more types of tasks than were feasible in previous waves of technological progress (Brynjolfsson et al., 2018).

In this paper we develop a framework for analysing the occupational impact of AI progress. The explicit focus on AI distinguishes this analysis from studies on robotisation (Acemoglu and Restrepo, 2018), digitalisation and online platforms (Agrawal et al., 2015) and the general occupational impact of technological progress (Autor, 2015). The framework links tasks to cognitive abilities, and these to indicators that measure performance in different AI fields. More precisely, we map 59 generic tasks from the worker surveys European Working Conditions Survey (EWCS) and Survey of Adult Skills (PIAAC) as well as the occupational database O*Net to 14 cognitive abilities (that we extract from the cognitive science literature) and these to a comprehensive list of 328 AI-related benchmarks which are metrics on publicly available datasets that indicate progress in AI techniques (see Figure 1).

Differently from previous approaches that have tried to link directly AI developments with task characteristics (Brynjolfsson et al., 2018), our framework adds an intermediate layer of cognitive abilities. With 14 distinct cognitive abilities, this layer is more detailed than the task charakteristics mentioned in the task-based approach as introduced in (Autor et al., 2003). In this model work tasks are defined by their routine, abstract, and manual content, all three characteristics of work organisation that point towards task automation (Autor and Handel, 2013). Although this approach has been very fruitful and inspired many studies (including this one), in our view these characteristics do not suffice to capture AI's potential to affect and transform work tasks that are not (yet) tailored to be performed (fully or partially) by a machine. Hence, we leave open the possibility that besides substituting an already standardised task, AI may cause workplaces to transform the way a task is performed by acquiring some of the abilities required for the task.

The ability perspective allows us to distinguish machines that, through ML, are empowered with the abilities of performing in a range of several tasks from machines that are constructed or programmed to perform a specific task. For instance, the ability of understanding human language (covered by the area of Natural Language Processing) can be applied in a variety of tasks (such as reading or writing e-mails, or advising costumers/clients). Abilities are therefore a better parameter to evaluate progress in AI (Hernández-Orallo, 2017a). We focus on abilities instead of skills because from a human perspective abilities are innate and primary. Instead, skills instead acquired through a combination of abilities, experience and knowledge (Fernández-Macías et al., 2018). Since knowledge and experience are not appropriate properties of AI, linking AI benchmarks to abilities (instead of skills) should be less prone to measurement error (Hernández-Orallo, 2017a).

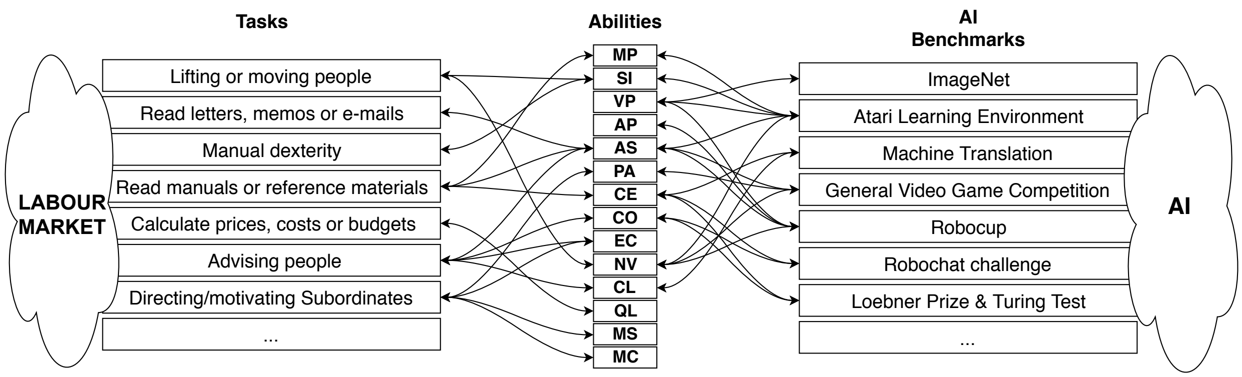


Figure 1: *Bidirectional and indirect mapping between job market and Artificial Intelligence (abilities described in Appendix A).*

Due to the intermediate layer of 14 different abilities, we also gain a broader understanding on the occupational impact of AI. That is, the framework allows us to not only define a single occupation-level AI exposure score but also lets us identify the different abilities that are most likely driving the implementation of AI in the workspace. In other words, we can identify which abilities are less likely to be performed by AI and are therefore less prone to changes in the way they are currently being performed.

Furthermore, we rely on a wide range of AI benchmarks to approximate the direction of AI progress. These benchmarks are performance indicating metrics (such as accuracy, or area under the receiver operating characteristics) on openly accessible datasets which are prominently promoted on online platforms where both AI researchers and industry players present their current performance in different AI domains. The collection of these benchmarks provides a thorough overview of the direction of AI progress. In many cases these benchmarks and the work on them exist before the explicit formalisation of its use at work. For instance, performing well in the game of ``Go'', which is recorded in a corresponding benchmark, is not explicitly mentioned in any work-related task. However, AI that performs well on these benchmarks needs to exhibit abilities in memory processing and planning. Both abilities are useful in the performance of some work-related tasks. Moreover, instead of looking at past progress of these benchmarks, we measure interest in specific AI domains through the prevalence of benchmarks in each category. This measure allows for the computation of future trends based on past developments in each category and can be easily updated for future years.

This paper contributes to the literature on measuring the occupational impact of AI (Frey and Osborne, 2017; Arntz et al., 2016; Nedelkoska and Quintini, 2018), although we distinguish between technological feasibility of AI and the full automation (substitution through machines) of a task. We further complement this literature by measuring AI potential in cognitive abilities using AI field benchmarks that are used as orientation by AI researchers and other AI industry players. This approach captures the entire AI research field more comprehensively than expert predictions on the future automatibility of occupations as in Frey and Osborne (2017) and subsequent studies.

This measure of AI progress complements Brynjolfsson et al. (2018)'s rubric to determine the suitability of tasks for ML since it can be easily updated to future developments in the already recorded benchmarks. In addition, some of Brynjolfsson et al. (2018)'s defined task properties are endogenous to the redefinition of an occupational task in which AI is already established. For instance, the property "task information is recorded or recordable by computer" emerges once the corresponding AI technology is specified to the task. That is, if AI performs well in one of the abilities required to perform the task, it requires a redefinition of the affected task around this ability in order to be considered a separate work task. In contrast, the ability perspective assesses AI progress one step ahead and does not require a redefinition of tasks.

Our approach relates most to Felten et al. (2018) who also link AI field benchmarks of to work-related abilities but there are some noteworthy differences. First, Felten et al. (2018)'s measure of AI progress relies on one platform only, the Electronic Frontier Foundation (EFF)[1], which is restricted to a limited set of AI benchmarks. The benchmarks in the present framework further rely on our own previous analysis and annotation of papers (Hernández-Orallo, 2017c; Martínez-Plumed et al., 2018; Martinez-Plumed and Hernandez-Orallo, 2018) as well as on open resources such as *Papers With Code*[2], which includes data and results from a more comprehensive set of AI benchmarks, competitions and tasks. This ensures a broad coverage of AI tasks, also providing insight into AI performance in cognitive abilities that go beyond perception, such as text summarisation, information retrieval, planning and automated deduction.

For better comparability across these benchmarks that come from a multitude of different AI domains,the measure of AI progress is also different. Felten et al. (2018) assess AI progress by computing linear trends in each benchmark. However, nonlinear performance jumps at different thresholds of each benchmark, impede comparability between different benchmarks. We address this issue by translating benchmarks to a measure of AI research activity that enables comparability across benchmarks from different AI fields.

In a more recent article, Webb (2020) measures AI's occupational impact by computing the overlap between O*NET job task descriptions and the text of patents. We complement this approach by measuring AI progress before it is formulated in patents.

The remainder of this paper is structured as follows. The following section provides background information on the construction of the layer of cognitive abilities in the framework. In Section 3 we describe the different data sources that we combine to construct the framework which is followed by Section 4 where we present the methodology used to construct the framework. We present the results of the application of our framework in Section 5. Section 6 concludes.

2 Background: cognitive abilities

A first glance at the tasks that are usually identified in the workplace and those that are usually set in AI as benchmarks (see Figure 1) reveals the difficulty of matching them directly, as the lists are very different. However, tasks and benchmarks have some latent factors in common, what we refer to as `cognitive abilities', which we can use to map them indirectly but at a level of aggregation that is more insightful. For this characterisation of abilities we look for an intermediate level of detail, excluding very specific abilities and skills (e.g., music skills, mathematical skills, hand dexterity, driving a car, etc.) but also excluding very general abilities or traits that would influence all the others (general intelligence, creativity, etc.). As we just cover cognitive abilities, we also exclude personality traits (e.g., the big five (Fiske, 1949): openness, conscientiousness, extraversion, agreeableness and neuroticism). Although we consider the latter essential for humans, their ranges can be simulated in machines by changing goals and objective functions.

[1] https://www.eff.org/es/ai/metrics

[2] https://paperswithcode.com/

At the intermediate level, we aim at a number and breadth similar to the “broad abilities” of the Cattell-Horn-Carroll hierarchical model (see Figure 2) (Carroll et al., 1993). However, some of them are very anthropocentric and are not really categorical, but orthogonal (such as processing speed or the distinction between short-term and long-term memory).

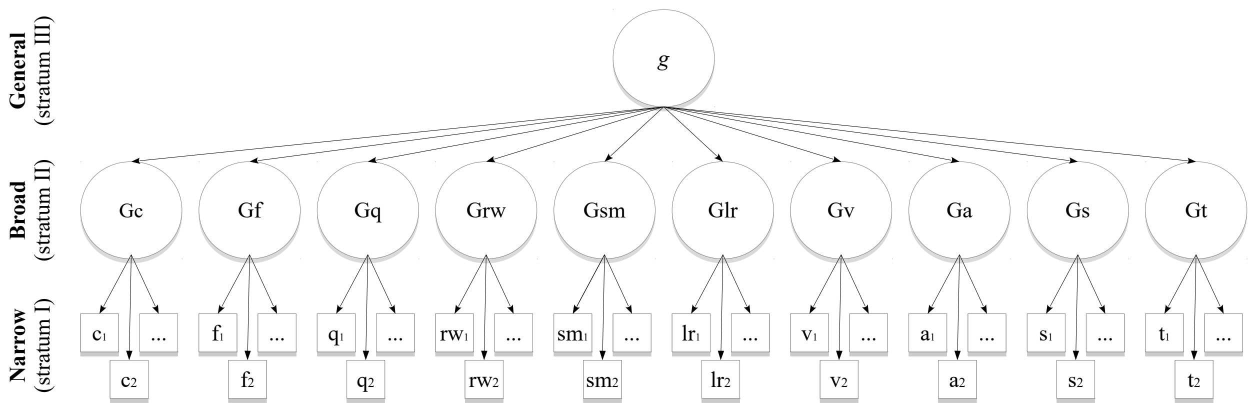


Figure 2: *Cattell-Horn-Carroll's three stratum model. The broad abilities are Crystallised Intelligence (Gc), Fluid Intelligence (Gf), Quantitative Reasoning (Gq), Reading and Writing Ability (Grw), Short-Term Memory (Gsm), Long-Term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Processing Speed (Gs) and Decision/Reaction Time/Speed (Gt).*

For our purposes we use 14 categories as the result of the integration of several tables and figures from Hernández-Orallo (2017c), originally collected from psychometrics, comparative psychology, cognitive science and artificial intelligence. The 14 categories are defined as follows:

- Memory processes (MP)
- Sensorimotor interaction (SI)
- Visual processing (VP)
- Auditory processing (AP)
- Attention and search (AS)
- Planning and sequential decision-making and acting (PA)
- Comprehension and compositional expression (CE)
- Communication (CO)
- Emotion and self-control (EC)
- Navigation (NV)
- Conceptualisation, learning and abstraction (CL)
- Quantitative and logical reasoning (QL)
- Mind modelling and social interaction (MS)

- Metacognition and confidence assessment (MC)

The hierarchical theories of intelligence in psychology, animal cognition and the textbooks in AI are generally consistent (at least partially) with this list of abilities, or in more general and simple terms, with this way of organising the vast space of cognition. The definition of cognitive abilities can be found in Appendix A, which also includes a *rubric* so that we can determine for each ability whether it is required for a particular task.

3 Data

For our analysis we rely on two different sources of data that provide information on task intensity in occupations (i.e. the relevance of and time spent on that task) on the one side and on AI research intensity on the other side. We start with a description of the data on tasks before moving to a description of the data on AI.

3.1 Tasks: work intensity

For the task dataset, we draw from the framework developed in Fernández-Macías and Bisello (2017). This data entails a list of tasks (presented in Table 3 in Appendix D) and their respective intensity (i.e. relevance and time spent) across occupations. The development of this dataset is described in detail in Fernández-Macías et al. (2016). In the following we provide a summary of the construction of this dataset.

We classify occupations according to the 3-digit International Standard Classification of Occupations (ISCO-3)[3]. Since there is no international data source that unifies information on all tasks required, we combine data from three different sources: the worker surveys the European Working Conditions Survey (EWCS)[4] and the OECD Survey of Adult Skills (PIAAC)[5] as well as the database the Occupational Information Network (O*NET)[6].

The data in the worker surveys are measured at the individual worker level based on replies to questions on what they do at work. Task intensity is derived as a measure of time spent on specific tasks. For instance, in the EWCS we derive the task *"Lifting or moving people"* from the survey question $q24b$ *"Does your main paid job involve lifting or moving people?"* and the corresponding 7-point scale answers ranging from *"All of the time"* to *"Never"*. Analogously, in the PIAAC we derive the task *"Read letters, memos or e-mails"* from the survey question $\mathrm{G\_Q01}b$ *"Do you read letters, memos or e-mails?"* and the corresponding 5-point scale answers ranging from *"Every day"* to *"Never"*. Due to the nature of survey data, we need to be aware of issues such as measurement error, high variation in responses across individuals and biased responses.

Similarly, the occupational database, O*NET is based on multiple waves of individual worker surveys but also on employer job postings, expert research and other sources. The data is curated by occupational experts and provided on a standardised occupational level. In this case, task intensity is derived from a variable that measures the extent to which the task is required to perform a job. For instance, the task *"Oral Comprehension"* is derived from the same variable and the corresponding level defined on a 7-point

[3] https://www.ilo.org/public/english/bureau/stat/isco/

[4] https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys

[5] https://www.oecd.org/skills/piaac/

[6] https://www.onetonline.org/

scale.

The O*NET is widely used in the literature on labour markets and technological change (Acemoglu and Autor, 2011; Frey and Osborne, 2017; Goos et al., 2009). Moreover, it covers a large share of the task list that we construct. However, the occupational level of the data precludes a further analysis into variation in task content within occupations. Moreover, much like the EWCS for Europe, the O*NET is based on US data only. Therefore, likely differences in the task content of occupations across countries due to institutional as well as socio-economic differences cannot be considered in the present analysis.

Finally, in order to make the measures of task intensity comparable across all three data sources, we equalise scales and levels of all variables. For this purpose, we rescale the variables to a $[0, 1]$ scale with 0 representing the lowest possible intensity and 1 representing the highest possible intensity of each variable. Moreover, we average scores measured on an individual level (i.e. all variables from PIAAC and EWCS) to the unified level of standardised 3-digit occupation classifications. The final database contains the intensity of 59 tasks across 119 different occupations.

To test the consistency of the variables that are derived from multiple datasources, Fernández-Macías et al. (2016) look at pairwise correlations and Cronbach's Alpha for multiple variables that measure similar concepts. Reassuringly, all tests yield high correlations and Cronbachs's Alpha values of between 0.8 and 0.9, suggesting consistency in the measurement of task intensity across the different data sources. Moreover, it is reasonable to doubt the comprehensiveness of the task framework. In fact, the results of the current study (see below) suggest that the following tasks are missing from the framework:

1. (Within Physical tasks) Navigation: moving objects or oneself in unstructured and changing spaces

2. (Within Intellectual - Information processing tasks) Visual and/or auditory processing of uncodified and unstructured information

3. (Within Intellectual - Problem Solving tasks) Information search and retrieval

4. (Within Intellectual - Problem Solving tasks) Planning

Since the collection of data on these additional task categories is yet to be conducted, the results of this paper will not include them.

3.2 Benchmarks: AI intensity

We consider a comprehensive set of AI benchmarks for our framework based on our own previous analysis and annotation of AI papers (Hernández-Orallo, 2017b; Martínez-Plumed et al., 2018; Martinez-Plumed and Hernandez-Orallo, 2018) as well as open resources such as *Papers With Code*[7] (the largest, up to date, free and open repository of machine learning code and results), which includes data from several repositories (e.g, EFF[8], NLP-progress[9], SQuAD[10], RedditSota[11], etc.). All these repositories draw on data from multiple (verified) sources, including academic literature, review articles and code platforms focused on machine learning and AI.

[7] https://paperswithcode.com/

[8] https://www.eff.org/es/ai/metrics

[9] https://github.com/sebastianruder/NLP-progress

[10] https://rajpurkar.github.io/SQuAD-explorer/

[11] https://github.com/RedditSota/state-of-the-art-result-for-machine-learning-problems

For the purposes of this study, from the aforementioned sources we track the reported evaluation results (when available or sufficient data is provided) on different metrics of AI performance across separate AI benchmarks (e.g., tasks, datasets, competitions, awards, etc.) from a number of AI domains, including (among others) computer vision, speech recognition, music analysis, machine translation, text summarisation, information retrieval, robotic navigation and interaction, automated vehicles, game playing, prediction, estimation, planning, automated deduction, etc. This ensures a broad coverage of AI tasks, also providing insight into AI performance in cognitive abilities that go beyond perception, such as the ability to plan and perform actions on such plans. Note that most of these benchmarks we are addressing are *specific*, implying that their goals are clear and concise, and that researchers can focus on developing specialised AI systems for solving these tasks. This does not mean researchers are not allowed to use more general-purpose components and techniques to solve many of these problems, but it may be easier or most cost-effective for the researchers to build a strongly specialised system for the task at hand. Specifically, our framework uses data from 328 different AI benchmarks for which there is enough information available to measure their progress for different evaluation metrics. Table 4 in Appendix E contains the details from the benchmarks used in our analysis.

When aiming at evaluating the progress in a specific (AI) discipline, we need to focus on objective evaluation tools to measure the elements and objects of study, assess the prototypes and artefacts that are being built and examine the discipline as a whole (Hernández-Orallo, 2017b). Depending on the discipline and task, there is usually a loose set of criteria about how a system is to be evaluated. See for instance Figure 3 showing the progress for various evaluation metrics of object recognition in the COCO (*Common Objects in COntext*) (Lin et al., 2014) benchmark. Several questions might arise regarding the latter: How can we compare results or progress between different metrics? How to compare between different benchmarks for the same task (e.g., COCO vs. MNIST (Bottou et al., 1994) vs. ImageNet (Deng et al., 2009)) or different tasks for the same benchmark? Or, even more challenging, how can we compare results from different tasks in the same domain or different domains? Actually, although there might be a general perception of progress due to the increasing trends of the metrics (or decreasing in case of error-based measures), it would be misleading to consider that the progress in AI should be analysed by the progress of specific systems solving specific tasks as there may be a complete lack of understanding of the relationships between different tasks. What does it mean, for instance, that one AI system demonstrates impressive (e.g., super-human) performance on a natural language processing task and another demonstrates impressive performance in a perception task (wrt. some evaluation metrics) if this does not imply that these developments can be integrated easily into a single agent in order to display more general perceptual or linguistic capabilities, or both at the same time (Brundage, 2016)? On the other hand, it is also hard to tell in many domains whether progress comes from better hardware, data, computing, software, and other resources, or better AI methods (Martínez-Plumed et al., 2018). Furthermore, the specialisation of many metrics to the domain, the evaluation overfitting (Whiteson et al., 2011), and the lack of continuity in some evaluation procedures can also be recognised as limitations and constraints (Hernández-Orallo, 2017b) when evaluating the progress of AI.

Given the above difficulties, instead of using the rate of progress, what we can analyse is the activity level around a specific benchmark, indicating the research intensity in a specific task in terms of the production (e.g., outputs such as research publications, news, blog-entries, etc.) from the AI community related to the above AI benchmarks. Benchmarks that have an increasing trend in their production rates indicate that more AI researchers and practitioners are working on them (i.e., there is a clear research effort and intensity). Note that this is not an indication of progress, although, presumably, effort may lead to some progress eventually. It is also worth considering that areas that usually gather more intensity are those where there is a general perception that breakthroughs are being made or about to be made. For instance, those problems that are already solved, where progress is expected to be minimal or those that are too

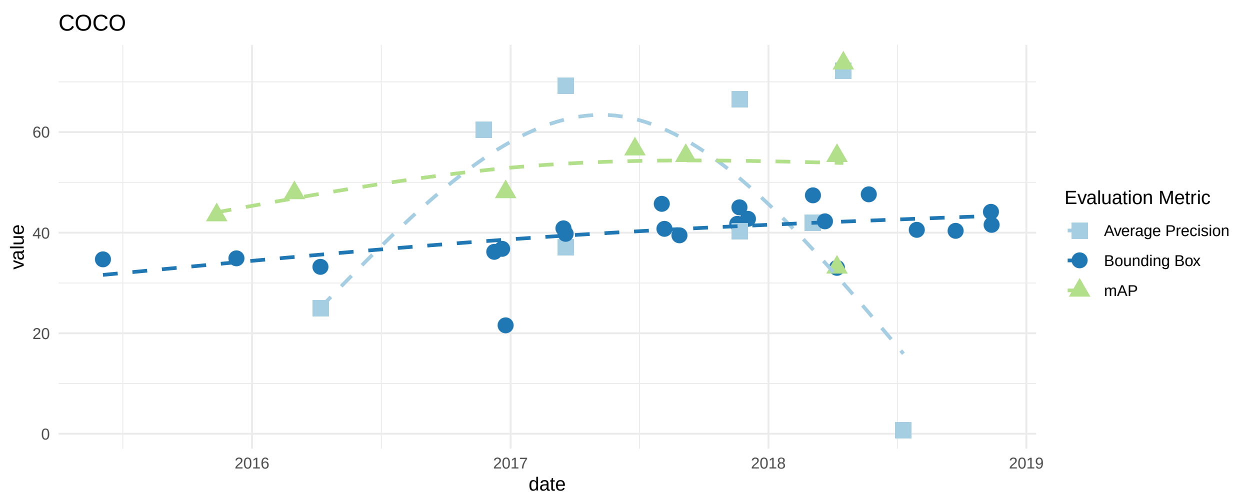


Figure 3: *Progress (trends represented with dashed coloured lines) across different evaluation metrics for COCO object recognition benchmark (Krizhevsky et al., 2009).*

challenging for the state of the art usually capture less attention.

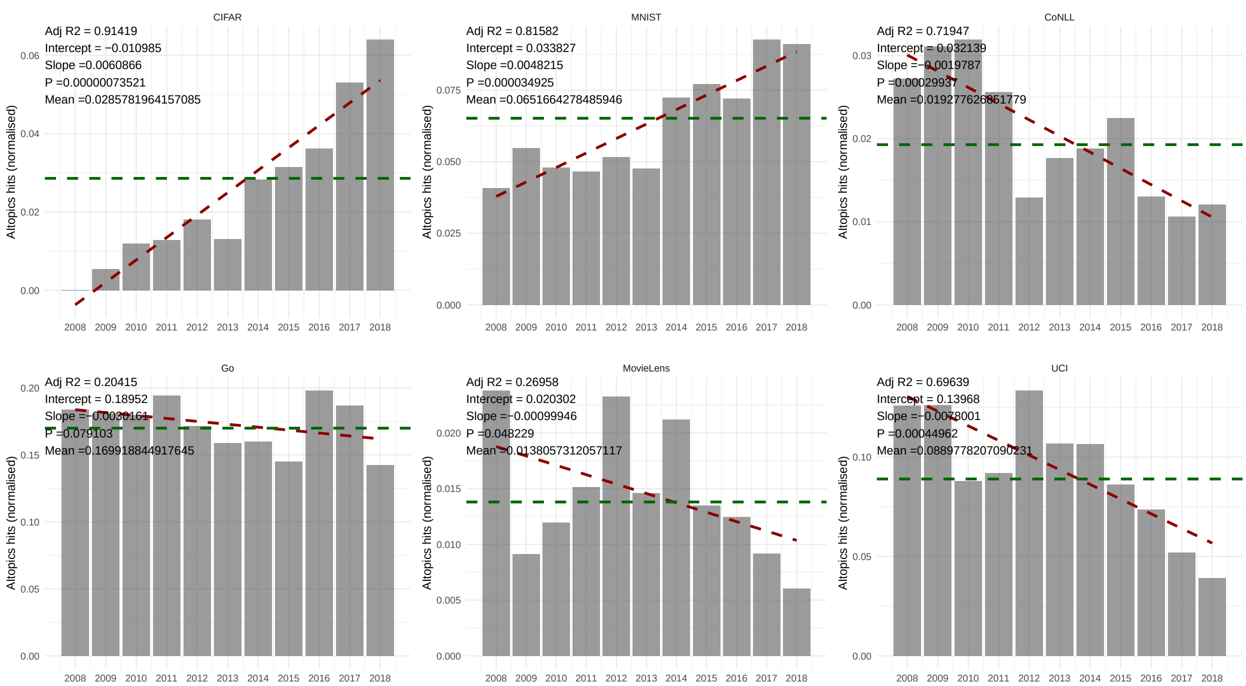


Figure 4: *Average rate of activity level or intensity (green dashed line) for a set of illustrative AI benchmark over the last decade (2008-2018).*

We can derive the activity level or *intensity* using some proxies. In particular, we performed a quantitative analysis using data obtained from *AI topics*[12], an archive kept by the Association for the Advancement of

[12] `https://aitopics.org`

Artificial Intelligence (AAAI)[13]. This platform contains a myriad of AI-related documents (e.g. news, blog entries, conferences, journals and other repositories from 1905 to 2019) that are collected automatically with NewsFinder (Buchanan et al., 2013). In this regard, in order to calculate the intensity in each particular benchmark, we average the (normalised[14]) number of hits (e.g., documents) obtained from *AI topics* per benchmark and year over a specific period of time (e.g, last year, lustrum or decade). This way we obtain a benchmark intensity vector (328 × 1) with values in $[0, 1]$, as they are counts divided by the total number of documents. Figure 4 presents the calculated relative intensity for a set of illustrative AI benchmarks over the last decade. Note that we make the assumption that a high relative intensity corresponds to breakthroughs or significant progress that can be translated to real applications in the short term.

4 Methodology

In this section we explain the construction of the framework. That is, we map between the three layers: (1) tasks (2) cognitive abilities, and (3) AI research.

4.1 Tasks to cognitive abilities

This section elaborates on the construction of the links between tasks (see Section 3.1) and cognitive abilities (see Section 2). To generate this mapping, a multidisciplinary group of researchers conducted an annotation exercise for each item of the task database. More precisely, in a cross-tabulation of the list of tasks (rows) and the cognitive abilities (columns) rubric (see Appendix A), each annotator was asked to put a 1 in a cell if an ability is inherently required, i.e. absolutely necessary to perform the respective task. In order to increase robustness in the annotations, we followed a *Delphi Method* approach (Dalkey and Helmer, 1963). That is, this process is repeated, in order to increase agreement among annotators. In the second round the annotators were asked to repeat the mapping exercise in light of the results of the first round and corresponding feedback. To increase robustness in the assignment of abilities to tasks, we define an ability as assigned to a task if at least two annotators assigned this ability. This makes the assignment less sensitive to outlier assignments.

Next, we adjust the $\{1, 0\}$ assignment of abilities to tasks by the intensity of the respective task in each occupation, where we obtain the measure of task intensity in each occupation from the dataset described in Section 3.1. The outcome of this reflects the occupational task intensity in the abilities assigned to the tasks in each occupation.

4.2 AI benchmarks to cognitive abilities

Similar to the mapping between cognitive abilities to tasks, we link these 14 cognitive abilities to the data on AI benchmarks (see Section 3.2). Specifically, a group of AI-specialised researchers was asked to consider how each AI benchmark is related to each cognitive ability: in a cross-tabulation of the vector of benchmarks b of length $|b| = 328$ and cognitive abilities a of length $|a| = 14$, a 1 is put in a ability-benchmark correspondence (or mapping) matrix (14 × 328) if an ability is inherently required, i.e. absolutely necessary to solve the respective benchmark. Note that since our correspondence or mapping matrix mean ``requires" (e.g, a particular benchmark needs a particular (set of) cognitive abilities to be addressed successfully), it makes sense to distribute the intensity values when a benchmark requires many abilities So, assuming that when more abilities are needed this requires more research effort or

[13] `https://www.aaai.org/`

[14] Document counts are normalised to sum up to 100% per year

work, we normalise the correspondence matrix through abilities. This means that columns are normalised to sum up 1 and values are thus in $[0, 1]$.

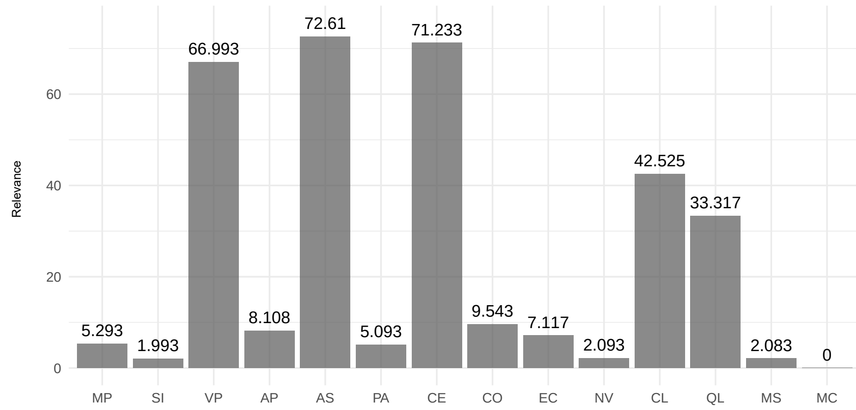


Figure 5: *Relevance (counting) of the cognitive abilities.*

From here we can calculate the vector of relevance for each cognitive ability from the correspondence matrix as row sums, thus obtaining the results in Figure 5. We see a clear dominance of *visual processing* (VP), *attention and search* (AS), *comprehension and compositional expression* (CE), *conceptualisation, learning and abstraction* (CL) and *quantitative and logical reasoning* (QL).

4.3 Combining occupations and AI through abilities

From the left side in Figure 1 we now have every occupation described in terms of 59 task intensities and the assignment (or non-assignment) of 14 abilities. That is, every cognitive ability appears in each occupation multiple times, depending on the number of tasks the ability has been assigned to. We simplify this mapping by summarising the information on the task layer in the abilities layer. In order to make sure that the ability scores are not driven by data availability of tasks, we first sort the task variables, to which the abilities have been assigned to, into the left side of the task framework presented in Fernández-Macías and Bisello (2017) and create task-ability indices by averaging within each task subcategory. In order to take into account the number of tasks that a cognitive ability is assigned to, we sum over all task indices linked to the same cognitive ability for each occupation. The final score indicates the total required intensity of each of the cognitive abilities for each of the 119 occupations.

Note that the differences in the intensities across different cognitive abilities are not linear, since the score of each cognitive ability derives from variables with highly varying scales. However, these scores take into account the number of tasks for which an ability is required weighted by the intensity of each task in each occupation. The scores therefore do allow for a ranking of the relevance of each ability within an occupation with a disregard for the distances. Similarly, the scores for the same cognitive ability across different occupations are measured on the same scale, which allows for a clear ranking of occupations along the same cognitive ability, again without interpretation of the distances between occupations.

For the computation of an AI impact score, we want the ability-specific scores to take into account the inter-connectivity of abilities that are required at the same time for the same occupation. That is, two very different occupations can have the same degree of intensity of one ability but can still be affected in very different ways by AI research intensity of this ability if the corresponding tasks require a different number of abilities at the same time. For instance, visual processing may be a very relevant ability for a person classifying offensive online content. Similarly, visual processing may be equally relevant for surgeons but also in combination with sensorimotor interaction. If we considered the intensity of each cognitive ability separately this would suggest that high AI intensity in visual processing but relatively low intensity in sensorimotor interaction would affect both occupations equally. However, in reality the surgeon would be affected less than the person classifying online content because the impact of high AI performance in visual processing would not be that high if performance in sensorimotor interaction would not also be high.

This is not an issue if we only compare the impact of AI on occupations through specific abilities. However, since AI research intensity is also connected with cognitive abilities, the ability-specific AI impact would be the same for every occupation. In order to construct an overall occupation-specific AI impact score that distinguishes occupations, we first establish a relation between the intensity scores of cognitive abilities for each occupation, which we denote relative ability-specific AI impact score. In detail, we transform the total score of each cognitive ability for each occupation such that the sum of scores within each occupation is equal to one. This transformed score entails a relationship between the different cognitive ability scores within each occupation.

Finally, we combine AI benchmarks (see Section 4.2) to labour market information using the common link to cognitive abilities. For this purpose we multiply the relative scores (described in the previous section) with the respective AI research intensity for each cognitive ability. Next, we take the sum over the products for each occupation. The final score indicates which of the studied occupations are relatively more likely to be affected by AI research intensity in the analysed cognitive abilities. For illustrative purposes we normalise this score, which we denote AI impact score, to a $[0, 1]$ scale.

5 Results

Before presenting results of the AI impact score, we illustrate the process of the development of the framework through intermediate results of the mapping of abilities to tasks and the mapping of AI benchmarks to abilities. More detailed results of the annotation exercise for the assignment of abilities to tasks are shown in Appendix B.

5.1 Tasks and cognitive abilities

In order to gain an overview of the task-ability mapping, we implement a principal component factor analysis on tasks and abilities. Principal component analysis (PCA) consists in the orthogonal transformation of a set of possibly correlated variables into components (values) that are linearly uncorrelated. That is, this transformation could be thought of as revealing the underlying structure in the data explaining the most of data variance.

PCA could provide us with two useful insights. First, it will tell us how many principal components we need to explain the most of the variance in the data. Second, it might help us gain a better insight into the structure of abilities and tasks. Let us clarify using the results reported in Figure 6.

The results of the PCA in Table 6 show that the first four components explain 66% of the variance in the data, two thirds of the original variance of all 14 cognitive abilities. The first component mostly summarises cognitive abilities that have a direct relationship with other abilities (such as *mind modelling and social interaction*, *planning* and *communication*) and could be interpreted as the latent variable measuring cognitive abilities in what we categorise as social tasks (i.e. tasks whose object is people). The second and the third components are mostly associated with tasks that necessitate a processing of information

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6
Memory Processes	0.1544	0.4479	0.1401	0.2434	-0.0459	-0.1987
Sensorimotor Interaction	-0.2008	-0.3758	0.1371	0.4184	-0.1691	-0.0398
Visual Processing	-0.1172	-0.0146	0.0877	0.2599	0.7654	0.303
Auditory Processing	-0.0621	0.0272	-0.3699	0.1038	-0.3489	0.7488
Attention and Search	-0.0602	0.4401	0.0604	0.2614	-0.1712	0.2538
Planning	0.429	-0.0629	0.2489	0.0152	-0.0817	0.1735
Comprehension and Expression	0.0736	0.2167	-0.6185	0.1818	0.2482	-0.053
Communication	0.4154	-0.0916	-0.326	0.0683	0.1295	-0.1136
Emotional Control	0.438	-0.2119	0.0288	0.0313	0.0877	0.1553
Navigation	-0.1562	-0.29	0.2012	0.5185	0.0683	0.068
Conceptualisation, learning, abstraction	0.0983	0.4498	0.1644	0.284	0.0846	-0.0994
Quantitative and logical reasoning	-0.0622	0.1593	0.3251	-0.4612	0.2797	0.3733
Mind modelling and social interaction	0.4607	-0.1937	-0.0112	0.0417	0.0924	0.0767
Metacognition	0.3326	0.1004	0.2976	0.1271	-0.2026	0.1108
Explained variance	27%	17%	12%	10%	8%	7%

Figure 6: *PCA on tasks and abilities*

streams without social interaction. The second component identifies particularly those tasks that involve the processing of encoded information (text and numbers) but requires a less extent of originality or problem-solving. Indeed the main cognitive abilities explaining the second factor are *memory processing, attention and search* and *conceptualisation, learning and abstraction*, which all resemble the abilities of recognising specific patterns/criteria in order to perform the task. On the other hand, the third component presents a broader variety of cognitive abilities which seems to point mostly to those tasks that require a certain degree of flexibility and problem solving capacity, without necessarily being demanding in terms of abstraction. The fourth component is clearly associated to more physical abilities since *navigation* and *sensorimotor interaction* are mostly required in more physical tasks. Finally, the fifth and sixth components are mostly explained respectively by *visual processing* and *auditory processing* associated to *quantitative and logical reasoning*.

In order to structure the discussion on cognitive abilities, we conduct a cluster analysis to categorise them. The detailed analysis and results can be found in Appendix C. Overall, the resulting clusters allow us to sort the cognitive abilities in more rough categories of social abilities (EC, MS, MC, CO), object oriented abilities (CE, PA, MP, AS, CL, QL), and physical abilities (SI, NV, VP, AP). This gives further insights into the nature of abilities and the corresponding occupations.

As mentioned in Section 4.3, we expect the impact of higher AI research intensity in a specific ability on a specific task and occupation to be lower, if the performance of this tasks requires a combination of multiple cognitive abilities. To further explore this idea, we analyse the likelihood of each cognitive ability to be assigned to a task in combination with multiple other cognitive abilities. For this purpose we compute the sum of assigned abilities per task and conduct a dominance analysis, an extension of multiple regression developed by Budescu (1993), of this sum on the 14 cognitive abilities.[15] The findings of the dominance analysis confirm the results of the cluster analysis. We can group the ranked abilities into social (rank 1-4), object oriented (rank 5-10) and physical abilities (rank 11-14).

5.2 AI research intensity in cognitive abilities

We can translate also the benchmark intensity vector (see Section 4.2) to cognitive abilities as a matrix-vector multiplication thus obtaining an ability intensity vector (14 × 1). This yields the **relative ability**

[15] More detailed results can be found in Appendix C.

indirect intensity, i.e. the relative AI research intensity for different periods of time. Figure 7 shows the computed AI research intensity for each cognitive ability for every two-year period from 2008 to 2018. The figure depicts that AI is currently having a larger relative intensity on those cognitive abilities that rely on memorisation, perception, planning and search, understanding, learning and problem solving, and even communication; smaller influence on physical-related abilities such as navigation or interaction with the environment. Since ``intensity'' depends on the the level of activity on *AI topics*, this would mean that there is a lower amount of documents related to those benchmarks for physical AI, but also, although to a lesser extent, due to a more limited number of robotics benchmarks, which are usually more difficult to build and maintain. Moreover, note that the focus of this paper is AI (i.e. rather *cognitive robotics*), which in many cases is distinct from robotics.

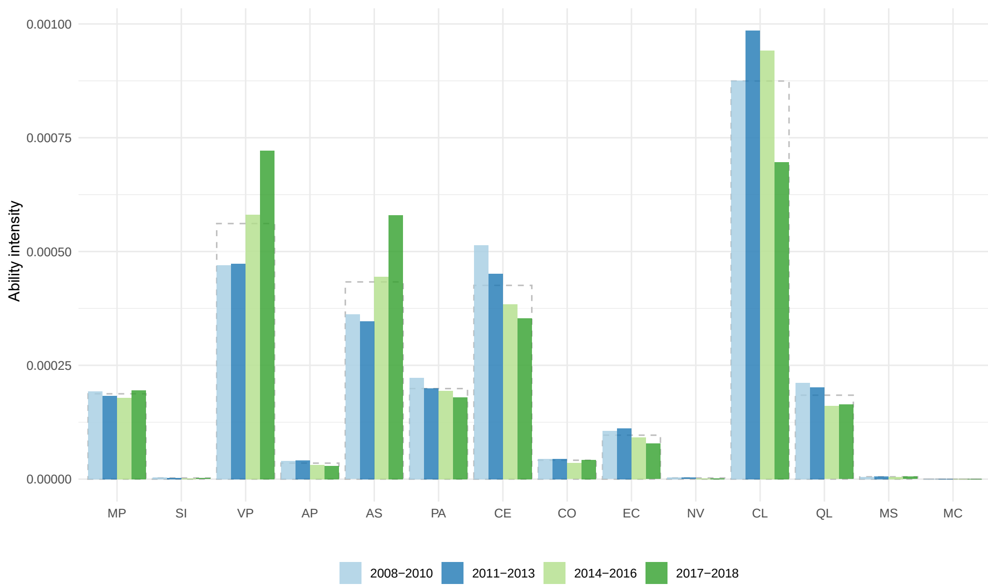


Figure 7: *Relevance per cognitive ability weighted by (average) rate intensity for different periods of years over the last decade (2008-2018). Empty grey dashed bars represent average values per ability for the whole period.*

We also see almost no research intensity on those abilities related to the development of *social interaction* (MS) and *metacognition* (MC). This may be due to the lack of suitable benchmarks to evaluate the interactions of agents (human and virtual) in social contexts; as well as the challenge (today) of developing agents able to properly perform in social contexts with other agents having beliefs, desires and intentions, coordination, leadership, etc. as well as being aware of their own capacities and limits.

Note that Figure 7 also shows trends over the years for each cognitive ability. There is a clear "increasing" trend in *visual processing* (VP) and *attention and search* (AS), while other abilities remain more or less constant (MP, SI, AP, CO, CL and MS) or have a small progressive decline (PA, CE, EC and QL). Note that these values are relative. For instance, PA, CE or QL have decreased in proportion to the rest. In absolute numbers, with an investment in AI research that is doubling every 1-2 years (Shoham et al., 2018), all of them are actually growing. Thus the figure shows that imbalances are becoming more extreme.

5.3 AI impact score

In this section we present the results from the the combination of all three layers of the framework: (1) tasks, (2) cognitive abilities, and (3) AI benchmarks in terms of occupations (see Section 4.3 for the corresponding methodology). We compute the AI impact score for each occupation using the AI research intensity scores from 2018. Before showing the final AI impact scores, we present the task-intensity of each cognitive ability and the ability-specific AI impact score in detail and focus on the following relevant selected occupations from ISCO-3 specifications: general office clerks; shop salespersons; domestic, hotel and office cleaners and helpers; medical doctors; personal care workers in health services; primary school and early childhood teachers; heavy truck and bus drivers; waiters and bartenders; building and related trades in construction.

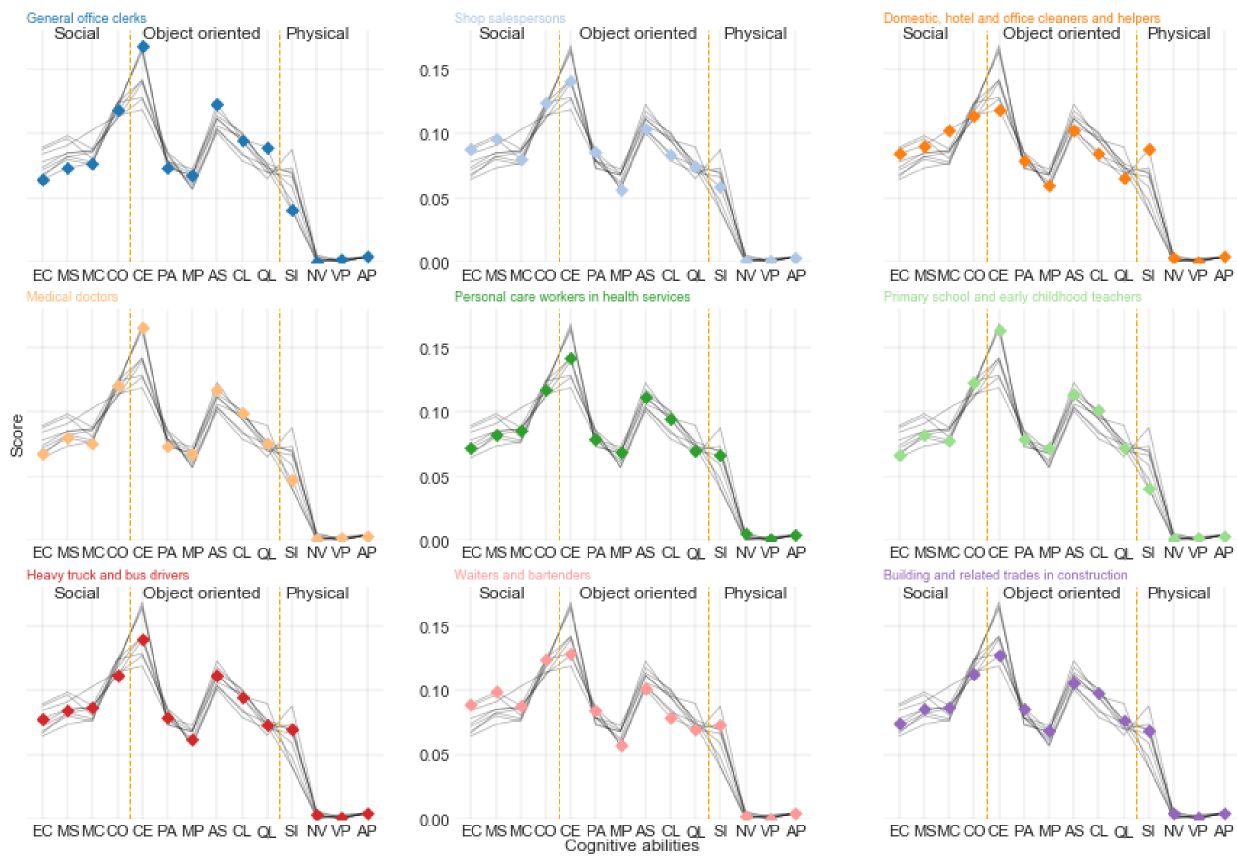


Figure 8: *Ability-specific scores of cognitive abilities for selected occupations*

Figure 8 depicts the relative ability-specific task intensity scores for the nine selected occupations mentioned above. That is, the figure shows for each of the nine selected occupations the relevance of each cognitive ability relative to the other cognitive abilities. In line with above findings, each subfigure is divided between social, object oriented and physical abilities. First, note that all occupations tend to exhibit similar profiles. On average, the most relevant ability is *comprehension* (CE), which is followed by *communication* (CO), and **?** (AS). Furthermore, we find a relatively high relevance for *conceptualisation* (CL) and *quantitative reasoning* (QL). Thus, all occupations tend to require social and object oriented abilities more than physical abilities.

Moreover, the figure shows that medical doctors, teachers and office clerks have high intensity scores for most cognitive abilities. These occupations also exhibit less pronounced scores for physical abilities. In contrast, heavy truck and bus drivers, waiters and bartenders as well as workers in building and related trades in construction have lower intensity levels for social and object-oriented abilities but higher intensity levels for the physical ability, *sensorimotor interaction* (SI). Finally, shop salespersons and waiters and bartenders have the highest levels for the social cognitive abilities, while these levels are very low for general office clerks. Overall, considering the nature of these occupations, the present scores depict reasonable ability profiles.

Figure 9 depicts the computed AI exposure score differentiated by cognitive abilities, for nine selected occupations. First, the figure shows that high-skill occupations such as medical doctors and teachers are more exposed to AI progress than comparatively low-skill occupations such as cleaners, waiters or shop salespersons. This is in line with the findings from Brynjolfsson et al. (2018) and Webb (2020). According to some studies, previous waves of technological progress led to more automation of mid-skill occupations, pushing mid-skill workers to either low- or high-skill occupations depending on education and skills, a phenomenon called technology driven labour market polarisation (Autor et al., 2003; **?**). This would contrast with the occupational impact of AI, which would be stronger in high-skilled occupations. If this effect is in fact a labour-replacement one, it would affect occupations that mostly remained unaffected by previous waves of automation, potentially leading to unpolarising effects and a reduction in income inequality (Webb, 2020). If this effect is a labour-enhancing one, it could imply a significant expansion of productivity for high-skilled occupations, potentially leading to occupational upgrading effects and an expansion of income inequality (very much like the traditional hypothesis of skills-biased technical change; see Acemoglu (2002)).

Second, Figure 9 shows that most of AI exposure is driven by its impact on tasks that require intellectual abilities, such as *comprehension*, *attention and search* as well as *conceptualisation*. On the other hand, not much AI impact can be expected through basic processing abilities, such as *visual* or *auditory processing*, nor through more social abilities, such as *mind modelling and social interaction*, or *communication*. However, our findings based on the task and occupation data indicate a relatively high need for social abilities in most occupations and a relatively low need for basic processing abilities. Equivalently, the findings on AI research intensity suggest high activity in AI areas that contribute to basic processing abilities but also to the abilities with the highest exposure score mentioned above, and low activity for social abilities.

This finding contributes to Deming (2017) who finds growing market returns to social skills in the 2000s as opposed to the 1980s and 1990s because social and *cognitive* skills (i.e. maths, statistics, engineering and science) are complements. In more detail, an increase in efficiency and quality due to automation of intellectual abilities could lead to increased demand for tasks that require intellectual abilities (Bessen, 2018). If these tasks also contain a high need for social abilities, of which we find that they are not likely to be automated in the near future, we can expect an increase in the returns to social abilities.

To complete this analysis, we present in Table 5 in Appendix F the overall AI impact score for all occupations. Note that this score does not represent a percentage but it can be used to infer a ranking between occupations in terms of AI impact. Regarding the nine selected occupations the table reflects the findings from the more detailed analysis. General office clerks have a relatively high AI impact score while we find relatively lower scores for shop salespersons. Surprisingly, the table suggests higher impact for many high skill occupations such as medical doctors or school teachers. These are occupations that were traditionally less affected by previous waves of automation. However, since we do not focus on the automation effect of AI but rather on the general impact, a lot of this impact can also be an indicator for a transformation

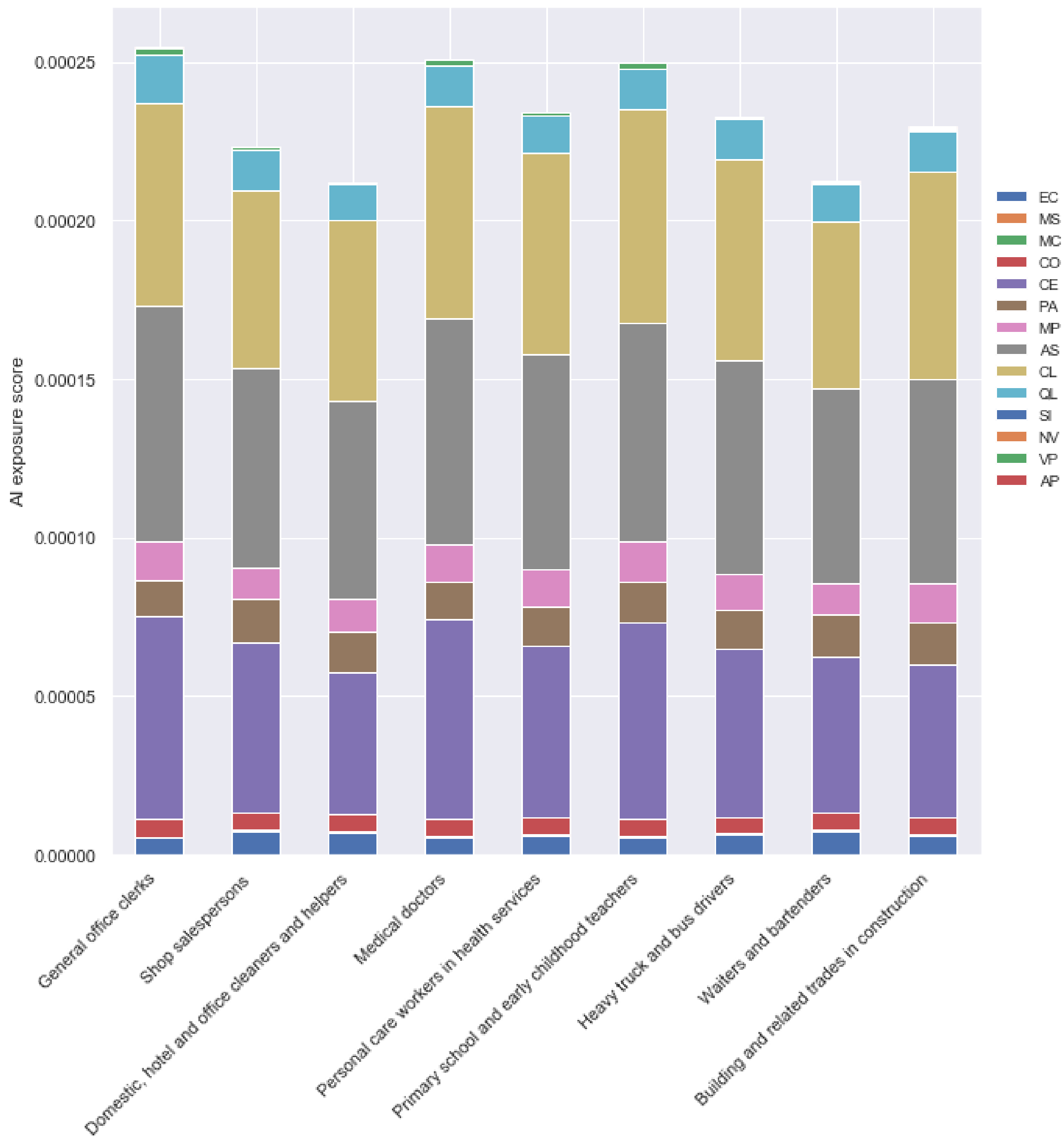


Figure 9: *Ability-specific AI impact scores for selected occupations*

of this occupation around the implementation of AI.

6 Conclusion

In this paper we developed a framework that allows for the analysis of the impact of Artificial Intelligence on the labour market. The framework combines occupations and tasks from the labour market with AI research intensity through an intermediate layer of cognitive abilities. This approach allows us to accurately assess the technological feasibility of AI in work related tasks and corresponding occupations. We use the framework to rank selected occupations by potential AI impact and to show the abilities that are most likely to drive the AI impact. Moreover, we find that some jobs that were traditionally less affected by previous waves of automation may now be subject to relatively higher AI impact.

The focus on abilities, rather than task characteristics, goes beyond measuring the substitution effect of AI. Most AI applications are built to perform certain abilities, rather than execute full work-related tasks and most tasks will require multiple abilities to be executed. Identifying the specific abilities that can be performed by AI gives a broader understanding on the impact of AI. Relying on AI field benchmarks that are used as orientation by AI researchers and other AI industry players makes the framework adoptable to future developments in AI research.

As mentioned above, AI exposure does not necessarily mean automation. Our findings do not imply that purely intellectual tasks will be automated, as other processes could occur when technology takes over some work that was previously performed by a human. For instance, better analytical predictions through AI could increase the value of *human judgement*, i.e. the ability to conduct meaningful inference and suggest appropriate actions (Agrawal et al., 2018). Overall, our findings show that most occupations will be significantly affected by AI but suggests that we should not fear an AI that is "taking over our jobs". For instance, most occupations involve a significant amount of social interaction tasks, and as previously mentioned progress in AI can in fact increase the value of social abilities and thus their demand in the future. We can be much more certain about the capacity of AI to transform jobs than about its capacity to destroy them.

In future work, this framework can be extended to integrate task characteristics of work organisation. This will allow us to measure and distinguish the impact of AI through newly acquired technical capabilities and the automation potential of tasks. Moreover, the measurement can be refined as more data on the relevance of specific work-related tasks as well as new benchmarks on AI progress arise. Overall, this framework presents an appropriate way to measure AI impact on labour markets as the connecting link, cognitive abilities, can capture general advances and advances in data collection well for both labour markets and AI research.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature 40*(1), 7--72.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, Volume 4, pp. 1043--1171. Elsevier.

Acemoglu, D. and P. Restrepo (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review 108*(6), 1488--1542.

Agrawal, A., J. Gans, and A. Goldfarb (2018). *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press.

Agrawal, A., J. Horton, N. Lacetera, and E. Lyons (2015). Digitization and the contract labor market: A research agenda. In *Economic analysis of the digital economy*, pp. 219--250. University of Chicago Press.

Arntz, M., T. Gregory, and U. Zierahn (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. *OECD Social, Employment and Migration Working Papers 2*(189), 47--54.

Autor, D. (2014). *Polanyi's paradox and the shape of employment growth*, Volume 20485. National Bureau of Economic Research Cambridge, MA.

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.

Autor, D. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553--97.

Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics 31*(S1), S59--S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279--1333.

Bessen, J. (2018). Ai and jobs: the role of demand. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Bottou, L., C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. D. Jackel, Y. LeCun, U. A. Müller, E. Säckinger, P. Y. Simard, et al. (1994). Comparison of classifier methods: a case study in handwritten digit recognition. In *International conference on pattern recognition*, pp. 77--77. IEEE Computer Society Press.

Brundage, M. (2016). Modeling progress in AI. In *AI, Ethics, and Society, Papers from the 2016 AAAI Workshop, Phoenix, Arizona, USA, February 13, 2016.*

Brynjolfsson, E., T. Mitchell, and D. Rock (2018). What can machines learn, and what does it mean for occupations and the economy? In *AEA Papers and Proceedings*, Volume 108, pp. 43--47.

Brynjolfsson, E., D. Rock, and C. Syverson (2018). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Buchanan, B. G., J. Eckroth, and R. Smith (2013). A virtual archive for the history of ai. *AI Magazine 34*(2), 86--86.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychological bulletin 114*(3), 542.

Carroll, J. B. et al. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Claveria, O., E. Monte, and S. Torra (2019). Economic uncertainty: a geometric indicator of discrepancy among experts' expectations. *Social Indicators Research*, 1--20.

Coxeter, H. S. M. (1961). Introduction to geometry.

Dalkey, N. and O. Helmer (1963). An experimental application of the delphi method to the use of experts. *Management science 9*(3), 458--467.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593--1640.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248--255. Ieee.

Felten, E. W., M. Raj, and R. Seamans (2018). A method to link advances in artificial intelligence to occupational abilities. In *AEA Papers and Proceedings*, Volume 108, pp. 54--57.

Fernández-Macías, E. and M. Bisello (2017). Measuring The Content and Methods of Work: a Comprehensive Task Framework. Technical report, European Foundation for the Improvement of Living and Working Conditions.

Fernández-Macías, E., M. Bisello, S. Sarkar, and S. Torrejón (2016). Methodology of the construction of task indices for the European Jobs Monitor. Technical report, European Foundation for the Improvement of Living and Working Conditions.

Fernández-Macías, E., E. Gómez, J. Hernández-Orallo, B. S. Loe, B. Martens, F. Martínez-Plumed, and S. Tolan (2018). A multidisciplinary task-based perspective for evaluating the impact of ai autonomy and generality on the future of work. *arXiv preprint arXiv:1807.02416*.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology 44*(3), 329.

Frey, C. B. and M. A. Osborne (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change 114*, 254--280.

Goos, M., A. Manning, and A. Salomons (2009). Job polarization in europe. *American economic review 99*(2), 58--63.

Hernández-Orallo, J. (2017a). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review 48*(3), 397--447.

Hernández-Orallo, J. (2017b, Oct). Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review 48*(3), 397--447.

Hernández-Orallo, J. (2017c). *The measure of all minds: evaluating natural and artificial intelligence*. Cambridge University Press.

Kodinariya, T. M. and P. R. Makwana (2013). Review on determining number of cluster in k-means clustering. *International Journal 1*(6), 90--95.

Krizhevsky, A., G. Hinton, et al. (2009). Learning multiple layers of features from tiny images. Technical report, Citeseer.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740--755. Springer.

Martínez-Plumed, F., S. Avin, M. Brundage, A. Dafoe, S. Ó. hÉigeartaigh, and J. Hernández-Orallo (2018). Accounting for the neglected dimensions of ai progress. *arXiv preprint arXiv:1806.00610*.

Martinez-Plumed, F. and J. Hernandez-Orallo (2018). Dual indicators to analyse ai benchmarks: Difficulty, discrimination, ability and generality. *IEEE Transactions on Games*, 1--1.

Nedelkoska, L. and G. Quintini (2018). OECD Social, Employment and Migration Working Papers No. 38. Technical Report 38.

Polanyi, M. (1966). The logic of tacit inference. *Philosophy 41*(155), 1--18.

Saari, D. G. (2008). Complexity and the geometry of voting. *Mathematical and Computer Modelling 48*(9-10), 1335--1356.

Shoham, Y., R. Perrault, E. Brynjolfsson, J. Clark, J. Manyika, J. C. Niebles, T. Lyons, J. Etchemendy, and Z. Bauer (2018). The ai index 2018 annual report.

Webb, M. (2020). The impact of artificial intelligence on the labor market.

Whiteson, S., B. Tanner, M. E. Taylor, and P. Stone (2011). Protecting against evaluation overfitting in empirical reinforcement learning. In *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pp. 120--127. IEEE.

7 Appendix

A Cognitive abilities rubric

We integrate several seminal psychometric models of intelligence to construct the following rubric of cognitive abilities.

MP: Memory processes: part of the information that is processed is stored in an appropriate medium to be recovered at will according to some keys, queries or mnemonics. This covers long-term memory and episodic memory, possibly using external devices such as books, spreadsheets, logs, databases, annotations, agendas and any other kind of analogical or digital recording and retrieval of data.

- Rubric question: Do all instances of this task inherently require that a robot or a human stores new memories to be recovered at a future time?
- Note: the ability is about creating new memories, not only recovering them. We exclude short-term and working memory, as almost any cognitive task requires them.

SI: Sensorimotor interaction: this deals with the perception of things, recognising patterns in different ways and manipulating them in physical or virtual environments with parts of the body (limbs) or other physical or virtual actuators, not only through various sensory and actuator modalities but in terms of mixing representations.

- Rubric question: Do all instances of this task inherently require that a robot or a human perceives the surrounding physical or virtual world, the body and the manipulation of objects with the physical properties of these objects?
- Note: this may be done through different modalities, e.g., blind people can do this well or a bat/robot using a radar.

VP: Visual processing: this deals with the processing of visual information, recognising objects and symbols in images and videos, movement and content in the image, with robustness to noise and different angles and transformations.

- Rubric question: Do all instances of this task inherently require that a robot or a human recognises static or moving elements in images or videos?
- Note: this processing excludes the assessment of the consistence of what is seen.

AP: Auditory processing: this deals with the processing of auditory information, such as speech and music, in noise environments and at different frequencies.

- Rubric question: Do all instances of this task inherently require that a robot or a human recognises specific sounds, signals, alarms, speech, melodies, rhythm, etc.?
- Note: in the case of speech, we exclude the full understanding of sentences or the subjective perception of harmony in music.

AS: Attention and search: this deals with focusing attention on the relevant parts of a stream of information in any kind of modality, by ignoring irrelevant objects, parts, patterns, etc. Similarly, it is the ability of seeking those elements that meet some criteria in the incoming information.

- Rubric question: Do all instances of this task inherently require that a robot or a human identifies, tracks or focuses on elements that meet some criteria, especially when surrounded by other elements not meeting the criteria?

- Note: criteria may be about any perceptual modality, and they can also be categories: for instance, focusing on the trajectory of straws in a stream of water or instruments in a symphony.

PA: Planning and sequential decision-making and acting: this deals with anticipating the consequences of actions, understanding causality and calculating the best course of actions given a situation.

- Rubric question: Do all instances of this task inherently require that a robot or a human evaluates the effects of different sequences of events, plan various courses of actions and make a decision accordingly?

- Note: this excludes complex reasoning processes about the world and assumes planning under mostly consistent information. Note also that we are not referring to simple actions or decisions, as almost any cognitive system makes actions; the task must involve sequences, time or other dependencies to be considered under planning.

CE: Comprehension and compositional expression: this deals with understanding natural language, other kinds of semantic representations in different modalities, extracting or summarising their meaning, as well as generating and expressing ideas, stories and positions.

- Rubric question: Do all instances of this task inherently require that a robot or a human understands text, stories and other representations of ideas in different formats, and the composition or transformation of similar texts, stories or narratives, summarising or expressing ideas?

- Note: this may be done through different modalities: text, auditory, drawings, etc. Note also that we are not referring to the processing of simple and predefined phrases or symbols; the task must involve the understanding or compositional use of elements that make a whole: sentences, stories, summaries, etc..

CO: Communication: this deals with exchanging information with peers, understanding what the content of the message must be in order to obtain a given effect, following different protocols and channels of informal and formal communication.

- Rubric question: Do all instances of this task inherently require that a robot or a human communicates information between peers or units, using different kinds of protocols and channels, at different registers, ensuring that the messages are sent, received and processed appropriately by all the interested peers?

- Note: this excludes the narratives that the messages may contain, focusing on the effective channels of information.

EC: Emotion and self-control: this deals with understanding the emotions of other agents, how they affect their behaviour and also recognising the own emotions and controlling them and other basic impulses depending on the situation.

- Rubric question: Do all instances of this task inherently require that a robot or a human understands emotions of others/themselves, when they are true or fake, expressing the right emotional reactions, controlling and using them in the appropriate context?

- Note: this excludes the complexities of social modelling and anticipation.

NV: Navigation: this deals with being able to move objects or oneself between different positions, through appropriate, safe routes and in the presence of other objects or agents, and changes in the routes.

- Rubric question: Do all instances of this task inherently require that a robot or a human transfers objects and oneself from one place to another at different scales (rooms, buildings, towns, landscape, roads, etc.), using basic concepts for locations and directions?

- Note: this may be done through different modalities, and approaches such as landmarking, geolocations, etc..

CL: Conceptualisation, learning and abstraction: this deals with being able to generalise from examples, receive instructions, learn from demonstrations, and accumulate knowledge at different levels of abstraction.

- Rubric question: Do all instances of this task inherently require that a robot or a human generate different levels of abstractions, provided by peers or self-generated, acquiring knowledge incrementally built upon previously acquired knowledge?

- Note: this ability to learn or to abstract must be present and happen to complete the task; in other words, the task is not limited to the use of abstractions or concepts or operations learnt in the past.

QL: Quantitative and logical reasoning: this deals with the representation of quantitative or logical information that is intrinsic to the task, and the inference of new information from them that solves the task, including probabilities, counterfactuals and other kinds of analytical reasoning.

- Rubric question: Do all instances of this task inherently require that a robot or a human produces new conclusions or facts from quantities, logical facts or rules given as inputs, detecting inconsistencies and fallacies?

- Note: this goes beyond the simple combination of rules or instructions, such as ordering a deck of cards. Note also that we are not referring to the internal processing of symbols or numbers that are not part of the task, such as the potentials of a neuron, the instructions of a programming language or the arithmetic of a CPU/GPU.

MS: Mind modelling and social interaction: this deals with the creation of models of other agents, so that their beliefs, desires and intentions can be understood, and anticipate the actions and interests of other agents.

- Rubric question: Do all instances of this task inherently require that a robot or a human successfully interacts in social contexts with other agents having beliefs, desires and intentions, the understanding of group dynamics, leadership and coordination?

- Note: this is not about sociability or agreeableness, i.e., how willing an agent is to social situations.

MC: Metacognition and confidence assessment: this deals with the evaluation of the own capabilities, reliability and limitations, self-assessing the probability of success, the effort and risks of own actions.

- Rubric question: Do all instances of this task inherently require that a robot or a human recognises accurately their own capabilities and limitations, when to assume responsibilities and when to delegate tasks and risks according to competences?

- Note: this goes beyond those cases covered by planning when considering the outcomes of several actions or no action. Note also that we are not referring to the mere selection of the action with highest probability or utility, as this is necessary for almost any task. This ability is about estimating and using the confidence of actions appropriately.

B Mapping abilities to tasks

In this section we summarise the results of the annotation of abilities to tasks. The annotations of each round are put together in a Matrix $M = T \times A$ with dimensions (62×14) for 62 tasks ($t \in T$) and 14 cognitive abilities ($a \in A$), where each cell in M ($c_{t,a}$) represents the sum over all annotations of a respective round. On the task level we describe our results by number of assigned abilities and consensus. An ability is assigned to a task if at least one annotation is equal to 1 for a respective task-ability cell. That is, for each task t, we define the number of assigned abilities as:

$$S(t) = \sum_{a} \left[c_{t,a} \geq 1 \right]$$

where $[P]$ are the *Iverson brackets* : $[P]$ is defined to be 1 if P is true, and 0 if it is false.

We also compute the level of consensus among respondents using a geometry-based disagreement measure following on from the work of Saari (2008); Claveria et al. (2019). Here, the authors define a framework to proxy economic uncertainty or to determine the likelihood of discrepancy among respondents. In our setting, we assume a dichotomous questionnaire with $N = 2$ reply options (e.g., ability is assigned or not to a task), and $R_{i,a}$ denoting the aggregate percentage of responses in category $i \in \{1, 0\}$ for a specific ability $a \in A$. As the sum of R adds to 100, a natural representation of the vector containing all the information from the respondents for a given ability a is as a point on a 1-dimensional (2 vertexes) simplex (Coxeter, 1961). Note that, while each of the N vertexes corresponds to a point of maximum consensus, if the point is near the barycenter, there would be a maximum discrepancy among the respondents. We can then compute the consensus between respondents as the relative weight of the distance of each point to the barycenter, formalised as:

$$C_a = \sqrt{\frac{\sum_{i=1}^{N} (R_{i,a} - \frac{100}{N})^2}{\frac{(N-1)}{N}}}$$

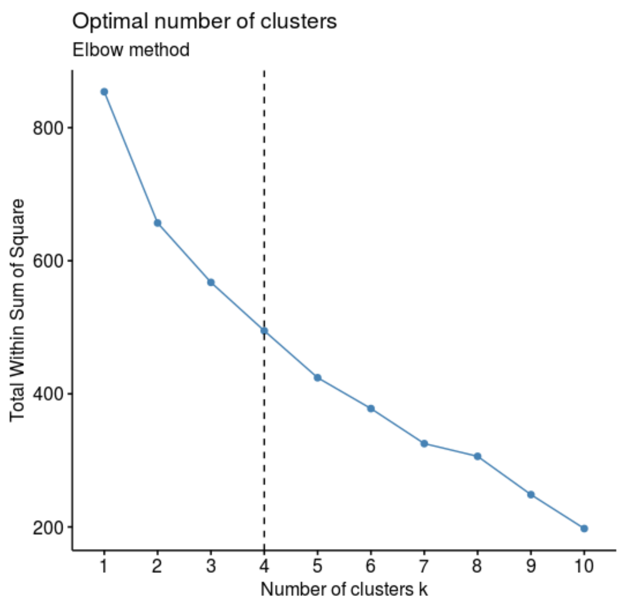
As can be seen in Table 1, we find that the annotators become stricter with their assignments of cognitive abilities to tasks in the second round. In addition, consensus in assignments increases from on average 80.65% to 87.6% from one round to the next.

Table 1: *Difference in annotations between round 1 and round 2*

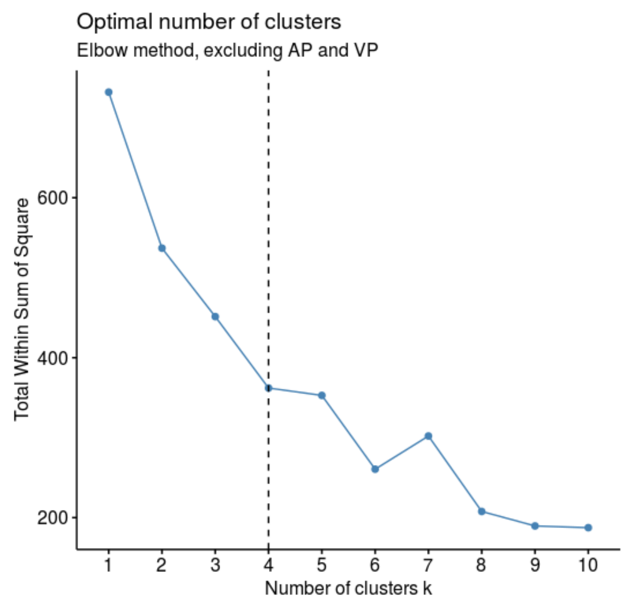
	S			C		
	round 1	round 2	diff.	round 1	round 2	diff.
Average	6.03	5.34	-0.69	80.65%	87.7%	7.05 p.p
Min	0	0	0	57.14%	69.05%	11.91 p.p
Max	13	10	-3	100.00%	100.00%	0

C Cluster analysis of tasks to abilities mapping and multi-ability tasks

The next step is to perform a cluster analysis to see how the tasks group together given the underlying structure of cognitive abilities. We perform a k-means cluster where the number of clusters is decided according to the elbow method (Kodinariya and Makwana, 2013) and following the results of the PCA. The elbow method minimizes the total within-cluster variance up to the point where adding an additional cluster does not increase the percentage of variance explained. Figure 10 shows the results of the elbow method including all the cognitive abilities (Panel 10a) and excluding vision and auditory processes (Panel 10b).



(a)



(b)

Figure 10: *Elbow method*

Looking at panel a in Fig 10, the elbow criterion seems to be reached somewhere between 4 and 6 clusters. Indeed, even if the data in panel a do not show a sharp decrease of the total within sum of squares it could be argued that after 4 the decline is less steep and the additional clusters do not contribute substantially to a reduction of the total within-cluster sum of squares. Following the PCA results, we compute the optimal number of clusters excluding auditory and visual processing from the set of cognitive abilities. The results shown in Panel 10b show how the data structure is better identified by four clusters, reinforcing the argument that visual and auditory processing might create some distortions in the underlying structure of the data. Fig 11 shows the projection of the tasks on the bi-dimensional plane identified by the two first principal components. The cluster to which social tasks belong to (the purple one) is clearly identified and separated from the rest. This cluster is characterised by tasks that involve human interactions. The blue cluster groups intellectual tasks that require a certain degree of information processing as well as the ability to understand natural language and identify a relevant stream of information. The green cluster includes mostly physical tasks. Finally, the red cluster shows an overlap with both the blue and the green ones suggesting the presence of multiple local optimums on the line of the complexity of the tasks. In a sort of polarised interpretation of complexity of the task, some simple tasks such as basic numeracy (i.e. calculate fractions, use of calculators, etc.) overlap with the cluster of physical tasks. On the upper end of complexity, *learning new things* overlaps with the blue cluster of intellectual task.

Figure 11: *K-means plot on the first two principal components*

Tasks with multiple abilities

We analyse the abilities in terms of their marginal contributions to R^2 (i.e., whether a predictor variable is dominant over another predictor) in a linear regression of the sum of abilities per task on the 14 cognitive abilities.

Table 2: *Ranking of abilities most likely to predict a task that requires multiple abilities combined*

Cognitive Ability	Dominance	Dominance (std)	Ranking
EC	0.1602	0.1703	1
MS	0.1390	0.1478	2
MC	0.1386	0.1474	3
CO	0.1099	0.1169	4
PA	0.1089	0.1158	5
MP	0.0685	0.0728	6
CL	0.0651	0.0692	7
AS	0.0510	0.0542	8
CE	0.0477	0.0507	9
QL	0.0212	0.0225	10
SI	0.0129	0.0137	11
NV	0.0080	0.0085	12
VP	0.0078	0.0083	13
AP	0.0017	0.0018	14

Ranking of cognitive abilities in terms of contribution to explaining the sum of assigned abilities in a task. The abbreviation "std" stands for "standardised. $R^2 = 0.9405$ *for a regression of the sum of assigned abilities per task on all cognitive abilities. Dominance represents average marginal contribution of cognitive ability towards* R^2 *over all potential model combinations.*

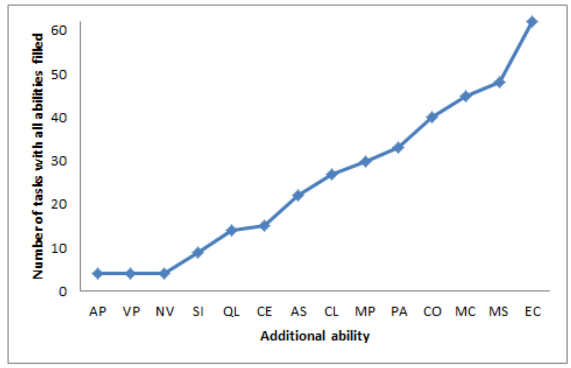


Figure 12: *Number of tasks with all abilities filled per additional ability*

Table 2 shows the ranking of the cognitive abilities in terms of their average contribution to explaining the variation in the sum of assigned abilities per task (Dominance). In other words, the higher the ranking of one ability the higher the sum of assigned abilities in the tasks that requires this particular ability.

Correspondingly, starting from the ability with the lowest complexity ranking (*auditory processing*) Figure shows the number of tasks that only require abilities up to higher ranks. For instance, there are four tasks that only require *auditory processing*, *navigation* and *visual processing* or a combination of these abilities but no abilities with a higher rank. Some additional abilities increase the number of "solved" tasks more than others; e.g. assuming that abilities are acquired from lowest to highest rank, the additional ability *comprehension* (CE) solves fewer additional tasks than the additional ability *attention and search* (AS) that is ranked one step higher. Similarly, *social interaction* (MS) does not enable as many additional abilities as the ability *emotion and self-control* (EC) that is ranked one step higher.

D List of tasks

Table 3: *Lists of Tasks used in Mapping*

1	Task involving tiring or painful positions
2	Lifting or moving people
3	Carrying or moving heavy loads
4	Standing
5	Static Strength
6	Dynamic Strength
7	Trunk Strength
8	Arm-Hand Steadiness
9	Manual Dexterity
10	Finger Dexterity
11	Oral Comprehension
12	Written Comprehension
13	Oral Expression
14	Written Expression
15	Read letters, memos or e-mails
16	Read bills, invoices, bank statements or other financial statements
17	Write letters, memos or e-mails
18	Read directions or instructions in your job
19	Read manuals or reference materials?
20	Read diagrams, maps or schematic in your job
21	Have to write reports
22	Have to fill in forms
23	Read articles in newspapers, magazines or newsletters
24	Read articles in professional journals or scholarly publications
25	Read books
26	Write articles for newspapers, magazines or newsletters
27	Mathematical Reasoning
28	Number Facility
29	Calculate prices, costs or budgets
30	Use or calculate fractions, decimals or percentages
31	Use a calculator either hand-held or computer based
32	Prepare charts, graphs or tables
33	Use simple algebra or formulas
34	Use more advanced math or statistics
35	Learning new things
36	Deductive Reasoning
37	Inductive Reasoning
38	Information Ordering
39	Solving unforeseen problems on your own
40	Apply your own ideas in your work
41	Originality
42	Performing for or Working Directly with the Public
43	Selling a product or selling a service
44	Advising people
45	Persuading or influencing people
46	Negotiating with people either inside or outside your firm or organisation
47	Persuasion
48	Negotiation
49	Selling or Influencing Others
50	Resolving Conflicts and Negotiating with Others
51	Instructing, training or teaching people
52	Making speeches or giving presentations in front of five or more people
53	Instructing
54	Training and Teaching Others
55	Coaching and Developing Others
56	Manage or supervise other employees
57	Planning the activities of others
58	Coordinating the Work and Activities of Others
59	Guiding, Directing, and Motivating Subordinates

E List of AI benchmarks

Table 4: *Set of AI benchmarks and their mean intensity calculated using AI topics.*

Benchmark	Mean intensity	Benchmark	Mean intensity	Benchmark	Mean intensity	Benchmark	Mean intensity
20NEWS	0.00498666	Event2Mind	0.000004	MR	0.042382	Shogi	0.00029975
300W	0.00064231	Fashion-MNIST	0.001135	MRR	0.004226	SighanNER	0.00000000
ACE 2004	0.00011392	FB15k	0.000759	MS COCO	0.001450	SimpleQuestions	0.00154478
ACE 2005	0.00063344	FB15k-237	0.000153	MS MARCO	0.000415	Sintel	0.00018307
ADE20K	0.00012735	FCE	0.000201	MSRA	0.002307	SK-LARGE	0.00000405
Aerial-to-Map	0.00000000	FDDB	0.000052	Multi-Domain Sentiment Dataset	0.000759	SLAM 2018	0.00000809
AEROBCOMP	0.00000000	FFHQ	0.000004	MultiMNIST	0.000124	SNLI	0.00047133
AFAD	0.00000000	FGNET	0.000557	MultiNLI	0.000150	Sogou News	0.00005982
AFLW	0.00011200	FGVC Aircraft	0.000057	MultiRC	0.000008	spider	0.00275807
AG News	0.00017801	fisher WER	0.000000	Mushroom	0.007158	SQuAD	0.00075022
AI2 Kaggle Dataset	0.00000000	FLIC	0.000180	Music domain	0.000850	SR11Deep	0.00000000
Amazon Review	0.00094449	Flixster	0.000882	MUV	0.000432	SST	0.00261894
ANGRY-BIRDS	0.00019157	Florence	0.003080	NABirds	0.000020	Stanford Cars	0.00006772
Annotated Faces in the Wild	0.00002672	Flowers-102	0.000252	NarrativeQA	0.000058	Stanford Dogs	0.00033553
Arcade Learning Environment	0.00088491	GENIA	0.001328	NELL	0.002441	STARE	0.00037779
bAbi	0.00004494	GigaWord	0.000357	NER	0.008085	Static Facial Expressions in the Wild	0.00000000
Bing News	0.00009736	GLUE	0.003006	Netflix	0.020838	STL-10	0.00191744
BIWI	0.00008235	Go	0.172822	New York Times Corpus	0.000721	Story Cloze Test	0.00004344
BlogCatalog	0.00084899	Google Dataset	0.001131	NewsQA	0.000139	STS	0.00359873
Bosch Small Traffic Lights	0.00000405	Google Street Images	0.000034	North American English	0.000014	SUBJ	0.00524186
BotPrize	0.00021779	GTA V	0.000045	Noun Phrase Canonicalization	0.000000	SUN-RGBD	0.00009389
BP4D	0.00005449	GTSRB	0.000375	NYU Depth v2	0.000311	SVHN	0.00392847
BPI challenge	0.00004494	GVGAI	0.000047	NYU Hands	0.000000	SVNH-to-MNIST	0.00000000
BRATS	0.00013988	HANDS 2017	0.000000	Occluded LINEMOD	0.000000	SWAG	0.00010718
BSD*	0.00267574	Helpdesk	0.000621	OCCLUSION	0.015062	Switchboard	0.00214692
BUCC	0.00003076	HIV dataset	0.000448	Ohsumed	0.003953	SYNTHIA	0.00011521
BUS 2017	0.00000000	HotpotQA	0.000000	OMNIGLOT	0.001333	T-LESS	0.00014327
CACD	0.00002522	Human3.6M	0.000198	One Billion Word	0.000602	TACRED	0.00001214
CACDVS	0.00001713	Hutter Prize	0.000429	OntoNotes	0.000543	TCIA Pancreas CT	0.00000000
CAFR	0.00003405	ICSI MRDA Corpus	0.000000	OpenML	0.000469	TempEval-3	0.00007928
Caltech	0.02095834	ICVL Hands	0.000000	Oxford 102 Flowers	0.000080	Text8	0.00151862
CamVid	0.00022812	IDHP	0.000000	Oxford IIIT Pets	0.000008	The ARRAU Corpus	0.00000959
Cats and Dogs	0.00148122	IEMOCAP	0.000122	PA-100K	0.000000	TimeBank	0.00053232
CCGBank	0.00002172	IJB	0.000530	Par6k	0.000000	TIMIT	0.00443222
CelebA	0.00244468	ILSVRC	0.006063	PASCAL VOC	0.008332	Tox21	0.00029954
ChaLearn	0.00059309	IMAGECLEF	0.000839	Pascal3D+	0.000069	ToxCast	0.00005048
CHALL	0.00022490	ImageNet	0.028748	PATHFINDMAZES	0.000000	Trading Agents Competition	0.00030921
Children's Book Test	0.00012210	IMDb	0.010094	Pavia University	0.000115	TREC	0.01721892
CHiME	0.00015276	iNaturalist	0.000089	PCBA	0.000113	TrecQA	0.00135855
Chinese Poems	0.00006816	Indian Pines	0.000211	Penn Treebank	0.009668	TriviaQA	0.00012031
CIFAR	0.02494334	iPinYou	0.000018	PETA	0.000678	Tsinghua-Tencent	0.00010600
CIHP	0.00000000	ISBI 2012 EM Segmentation	0.000078	PhC-U373	0.000000	Turing Test	0.00261238
Citeseer	0.02500602	iSEG 2017 Challenge	0.000004	Photo Art 50	0.000000	TuSimple	0.00002172
Cityscapes	0.00069756	ISIC 2018	0.000016	PLANNINGCOMP	0.000000	Twitter Dialogue	0.00008427
Click-Through Rate Prediction	0.00097511	ITOP	0.000078	PROMISE 2012	0.000000	Ubuntu Dialogue	0.00028614
CliCR	0.00000405	IWSLT	0.001082	Pubmed	0.006882	UCF CC 50	0.00000809
CMU-SE	0.00001363	JFLEG	0.000016	QAngaroo	0.000016	UCI	0.09358595
CNN / Daily Mail	0.00039284	JIGSAWS	0.000302	QASent	0.000017	UCI-KEEL	0.00000809
COCO	0.00412190	Kaggle Skin Lesion Segmentation	0.000000	QM9	0.000186	UD	0.00818373
Cohn-Kanade	0.00041558	KITTI	0.001659	QuAC	0.000016	Urban100	0.00005708
CompCars	0.00003809	Labeled Faces in the Wild	0.001891	Quasar	0.001982	UT Multi-view	0.00000000
COMPLEXQUESTIONS	0.00000000	Leeds Sports Poses	0.000058	Quora Question Pairs	0.000104	UTKFace	0.00000000
CoNLL	0.02031269	LexNorm	0.000000	R52	0.001746	V-SNLI	0.00000000
CoQA	0.00001618	LibriSpeech	0.000207	R8	0.014510	VggFace2	0.00001713
Cora	0.00826819	LineMOD	0.000027	RACE	0.019478	Vid4	0.00001363
CR	0.03195242	Loebner Prize	0.000045	RaFD	0.000014	Visual7W	0.00020654
Criteo	0.00067999	Long-tail emerging entities	0.000000	RAP	0.002529	VoxForge	0.00010058
CT-150	0.00000000	LSUN Bedroom 256 x 256	0.000000	Real-World Affective Faces	0.000008	WAF	0.00068041
CUB	0.00254865	LUNA	0.001589	RecipeQA	0.000000	WebFace	0.00010956
CUB-200-2011	0.00040947	MAFA	0.000071	RecSys	0.009449	WebNLG	0.00000809
CUFS	0.00005044	Mandarin Chinese	0.000233	Reuters-21578	0.004667	WebQuestions	0.00018390
CUFSF	0.00001214	Market 1501	0.000093	Reverb	0.000499	Weibo NER	0.00000000
CUHK	0.00562678	MCTest	0.000448	RLCOMP	0.000000	WikiBio	0.00000405
DailyDialog	0.00001618	MediaEval	0.000114	Robo chat challenge	0.000000	WikiHop	0.00002832
DARPAGC	0.00000000	Medical domain	0.003382	Robocup	0.004842	Wikipedia	0.05339900
DARPARESAVE	0.00000000	MegaFace	0.000164	RotoWire	0.000000	WikiQA	0.00019131
DARPAUC	0.00000000	METR-LA	0.000008	RT-GENE	0.000004	WikiSQL	0.00005259
DBpedia	0.00824343	MHP	0.000122	RumourEval	0.000008	WikiText-103	0.00005409
DCASE	0.00033334	Million Song Dataset	0.001785	RVL-CDIP	0.000000	WikiText-2	0.00018803
DensePose-COCO	0.00000809	MIMIC-III	0.000607	SBD	0.000265	Winograd Schema Challenge	0.00037346
Dianping	0.00014017	Mini-ImageNet	0.000245	Scan2CAD	0.000000	Wizard-of-Oz	0.00066401
DIC HeLa	0.00000000	MIREX	0.000934	ScanNet	0.000042	WMT	0.00329137
DISFA	0.00003736	MLDoc	0.000000	SciTail	0.000040	WN18	0.00049114
Disguised Faces in the Wild	0.00000000	MMI	0.001267	SCUT-FBP	0.000017	WOS	0.00023507
Douban	0.00058997	MNIST	0.063154	SearchQA	0.000114	WSJ	0.00565300
DRIVE	0.04997003	ModelNet40	0.000164	Second dialogue state tracking challenge	0.000008	XNLI	0.00001214
DUC 2004 Task 1	0.00000405	Monologue	0.000763	SemEval	0.004884	Yahoo! Answers	0.00375964
DukeMTMC-reID	0.00002427	MORPH	0.002163	SensEval	0.000159	YCB-Video	0.00000000
DuReader	0.00000405	MORPH Album2	0.000017	SentEval	0.000044	Yelp	0.00362348
ECCV HotOrNot	0.00000000	MOSI	0.000054	Sentihood	0.000004	YouTube Faces	0.00019026
EMNLP 2017	0.00062733	MovieLens	0.014568	Sequential MNIST	0.000247		
enwiki8	0.00000000	MPII	0.000567	ShanghaiTech	0.000115		
NULL	NULL	MPQA	0.002706	ShapeNet	0.000461		

F AI impact score for studied occupations

Table 5: *AI impact score for studied occupations*

ISCO code	Occupation	AI impact score
952	Street vendors (excluding food)	0.7703
912	Vehicle, window, laundry and other hand cleaning workers	0.7811
911	Domestic, hotel and office cleaners and helpers	0.7847
513	Waiters and bartenders	0.7856
961	Refuse workers	0.7884
921	Agricultural, forestry and fishery labourers	0.7895
941	Food preparation assistants	0.7907
523	Cashiers and ticket clerks	0.7969
521	Street and market salespersons	0.8185
933	Transport and storage labourers	0.8200
141	Hotel and restaurant managers	0.8220
522	Shop salespersons	0.8259
932	Manufacturing labourers	0.8288
962	Other elementary workers	0.8288
835	Ships' deck crews and related workers	0.8312
514	Hairdressers, beauticians and related workers	0.8365
622	Fishery workers, hunters and trappers	0.8372
751	Food processing and related trades workers	0.8405
524	Other sales workers	0.8457
515	Building and housekeeping supervisors	0.8465
834	Mobile plant operators	0.8486
711	Building and related trades in construction	0.8493
512	Cooks	0.8517
833	Heavy truck and bus drivers	0.8614
713	Painters, building struct. cleaners, related trades workers	0.8616
931	Mining and construction labourers	0.8621
324	Veterinary technicians and assistants	0.8646
832	Car, van and motorcycle drivers	0.8655
532	Personal care workers in health services	0.8663
224	Paramedical practitioners	0.8666
342	Sports and fitness workers	0.8675
815	Textile, fur and leather products machine operators	0.8705
753	Garment and related trades workers	0.8705
142	Retail and wholesale trade managers	0.8708
312	Mining, manufacturing and construction supervisors	0.8710
811	Mining and mineral processing plant operators	0.8721
754	Other craft and related workers	0.8722
814	Rubber, plastic and paper products machine operators	0.8729
222	Nursing and midwifery professionals	0.8743
516	Other personal services workers	0.8750
531	Child care workers and teachers' aides	0.8790
731	Handicraft workers	0.8801
821	Assemblers	0.8804
511	Travel attendants, conductors and guides	0.8821
143	Other services managers	0.8826
816	Food and related products machine operators	0.8838
322	Nursing and midwifery associate professionals	0.8875
712	Building finishers and related trades workers	0.8877
421	Tellers, money collectors and related clerks	0.8879
611	Market gardeners and crop growers	0.8893
265	Creative and performing artists	0.8900
541	Protective services workers	0.8904
818	Other stationary plant and machine operators	0.8916
225	Veterinarians	0.8936
422	Client information workers	0.8944
612	Animal producers	0.8959
226	Other health professionals	0.8970
721	Sheet and structural metal workers, related workers	0.8975
134	Professional services managers	0.8979
752	Wood treaters, cabinet-makers, related trades workers	0.8999
432	Material-recording and transport clerks	0.9008
112	Managing directors and chief executives	0.9009
817	Wood processing and papermaking plant operators	0.9030
722	Blacksmiths, toolmakers and related trades workers	0.9033
813	Chemical products plant and machine operators	0.9042
111	Legislators and senior officials	0.9068
335	Regulatory government associate professionals	0.9082
723	Machinery mechanics and repairers	0.9096
343	Artistic, cultural and culinary associate professionals	0.9099
332	Sales and purchasing agents and brokers	0.9101
235	Other teaching professionals	0.9105
262	Librarians, archivists and curators	0.9118
621	Forestry and related workers	0.9130
232	Vocational education teachers	0.9135
132	Manufacturing, mining, construct., distribution managers	0.9139
741	Electrical equipment installers and repairers	0.9144
321	Medical and pharmaceutical technicians	0.9144
341	Legal, social and religious associate professionals	0.9158
315	Ship and aircraft controllers and technicians	0.9182
333	Business services agents	0.9189
121	Business services and administration managers	0.9212
122	Sales, marketing and development managers	0.9217
812	Metal processing and finishing plant operators	0.9220
133	ICT service managers	0.9225
441	Other clerical support workers	0.9242
234	Primary school and early childhood teachers	0.9250
263	Social and religious professionals	0.9257
732	Printing trades workers	0.9280
221	Medical doctors	0.9284
831	Locomotive engine drivers and related workers	0.9297
412	Secretaries (general)	0.9315
742	Electronics and telecommunications installers, repairers	0.9356
216	Architects, planners, surveyors and designers	0.9390
334	Administrative and specialised secretaries	0.9390
243	Sales, marketing and public relations professionals	0.9409
613	Mixed crop and animal producers	0.9415
411	General office clerks	0.9421
431	Numerical clerks	0.9421
261	Legal professionals	0.9428
233	Secondary education teachers	0.9447
413	Keyboard operators	0.9487
352	Telecommunications and broadcasting technicians	0.9492
314	Life science technicians,related associate professionals	0.9500
242	Administration professionals	0.9503
231	University and higher education teachers	0.9510
211	Physical and earth science professionals	0.9522
264	Authors, journalists and linguists	0.9545
313	Process control technicians	0.9555
213	Life science professionals	0.9563
212	Mathematicians, actuaries and statisticians	0.9662
331	Financial and mathematical associate professionals	0.9700
311	Physical and engineering science technicians	0.9702
241	Finance professionals	0.9704
214	Engineering professionals (excluding electrotechnology)	0.9731
351	ICT operations and user support technicians	0.9829
251	Software and applications developers and analysts	0.9906
215	Electrotechnology engineers	0.9974
252	Database and network professionals	1.0000

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub