Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks (Extended Abstract)*

Songül Tolan[1*], **Annarosa Pesole**[1], **Fernando Martínez-Plumed**[1,2], **Enrique Fernández-Macías**[1], **José Hernández-Orallo**[2,3] and **Emilia Gómez**[1]
[1]European Commission, Joint Research Centre
[2]Universitat Politècnica de València
[3]Leverhulme Centre for the Future of Intelligence
{songul.tolan, annarosa.pesole, fernando.martinez-plumed, enrique.fernandez-macias}@ec.europe.eu,
jorallo@upv.es, emilia.gomez@ec.europa.eu

Abstract

We present a framework for analysing the impact of AI on occupations. This framework maps 59 generic tasks from different occupational datasets to 14 cognitive abilities and these to a comprehensive list of 328 AI benchmarks used to evaluate research intensity in AI. The use of cognitive abilities as an intermediate layer allows for an identification of potential AI exposure for tasks for which AI applications have not been explicitly programmed. We provide insights into the abilities through which AI is most likely to affect jobs, and we show how some of the abilities where AI research is currently very intense are linked to tasks with comparatively limited labour input in the labour markets of advanced economies.

1 Introduction

The latest advances in Artificial Intelligence (AI), driven by rapid progress in machine learning (ML) and its subfields, 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.*, 2018b]

Our perception of what AI is able to do is driven by the growing importance of benchmarks in AI [Hernández-Orallo *et al.*, 2017]. For instance, a decisive moment for deep learning really happened when it started to perform better than many other techniques in benchmarks such as ImageNet [Deng *et al.*, 2009]. In the end, breakthroughs in some particular challenges and benchmarks have been identified as landmarks of the field [Ferrucci, 2012; Brown *et al.*, 2020] and used as illustrations of what AI can do. Also, the activity around benchmarks is a good indicator of where the research effort in AI is focusing.

In this paper we present a framework[1] for analysing the potential occupational impact of AI (illustrated in Figure 1).

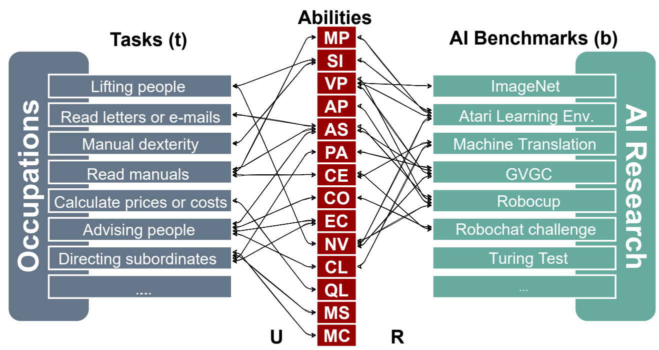

Figure 1: Illustrative example of the bidirectional and indirect mapping between occupations and Artificial Intelligence (abilities, tasks and benchmarks described in [Tolan *et al.*, 2021]). The 14 abilities are: Memory processes **(MP)**, Sensorimotor interaction **(SI)**, Visual processing **(VP)**, Auditory processing **(AP)**, Attention and search **(AS)**, Planning, sequential decision-making and acting **(PA)**, Comprehension and 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 notation we use is $\mathbf{t}$ for the tasks, $\mathbf{a}$ for the abilities and $\mathbf{b}$ for the benchmarks. The arrows are represented by correspondence matrices $\mathbf{U}$ (task-ability correspondence) and $\mathbf{R}$ (ability-benchmark correspondence).

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 evaluation tasks from benchmarking initiatives, challenges, competitions and scientific literature. These AI-related metrics reflect the intensity of current research and development in different AI techniques. This approach provides a sensible approximation to where AI

*This paper is an extended abstract of an article in the Journal of Artificial Intelligence Research [Tolan *et al.*, 2021]

[1]Code and data in https://github.com/nandomp/AIlabour

may have a bigger impact in the short and medium term, since we directly measure where more research effort is spent (see Figure 1).

The framework allows not only to define a single occupation-level AI exposure score, but also to identify the specific tasks affected and the different abilities that are most likely driving the implementation of AI in the workplace. Conversely, 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. In addition, connecting these benchmarks to work-related tasks allows to explore the question of the occupational impact of AI in the other direction, from occupational needs to specific AI benchmarks. That is, following the framework illustrated in Figure 1, we can identify occupations that are less exposed to AI, filter out the tasks that need to be performed in these occupations and specify which of the required abilities can be connected to corresponding benchmarks (and benchmark clusters) that would require increases in research activity for AI to have an impact on these occupations.

Matrix	Description	Unit
$\mathbf{t^o}$ $_{(59\times1)}$	tasks intensity	$\forall\ i^{\text{th}} \in \mathbf{t^o}, \mathbf{t_i} \in [0,1]$
$\mathbf{\Omega}$ $_{(59\times14)}$	task-ability annotation	$\forall \omega_{ij} \in \mathbf{\Omega}, \omega_{ij} \in \{0,...,6\}$
$\mathbf{U}$ $_{(59\times14)}$	task-ability correspondence	$\forall u_{ij} \in \mathbf{U}, u_{ij} \in \{0,1\}$
$\mathbf{\Psi^o}$ $_{(59\times14)}$	task intensity-ability	$\forall \psi_{ij} \in \mathbf{\Psi}, \psi_{ij} \in [0,1]$
$\mathbf{\Phi^o}$ $_{(14\times14)}$	task indices-ability	$\forall \phi_{ij} \in \mathbf{\Phi^o}, \phi_{ij} \in [0,1]$
$\mathbf{W}$ $_{(119\times14)}$	ability intensity - occupation	$\forall w_{ij} \in \mathbf{W}, w_{ij} \in [0,1]$
$\mathbf{R}$ $_{(328\times14)}$	ability-benchmark	$\forall r_{ij} \in \mathbf{R}, r_{ij} \in \{0,1\}$
$\mathbf{b}$ $_{(328\times1)}$	benchmark intensity	$\forall\ i^{\text{th}} \in \mathbf{b}, \mathbf{b_i} \in [0,1]$
$\mathbf{a}$ $_{(14\times1)}$	ability-specific AI intensity	$\forall\ i^{\text{th}} \in \mathbf{a}, \mathbf{a_i} \in [0,1]$
$\mathbf{V}$ $_{(119\times14)}$	occupation-ability AI impact	$\forall v_{ij} \in \mathbf{V}, v_{ij} \in [0,1]$

Table 1: Summary of notation.

2 Related Work

This paper contributes to the literature on the occupational impact of recent technological change [Nedelkoska and Quintini, 2018], but with a focus on identifying which occupations and types of tasks are more directly related to current developments in AI research, and therefore are more likely to be affected by AI in the future. We further complement the literature with a formal setting for measuring AI potential in cognitive abilities and in labour-related tasks and occupations. On the AI side, we perform this by relying on AI benchmarks, as used by researchers and industry to evaluate AI progress.This work contrasts with [Webb, 2020] and complements Brynjolfsson et al.'s measure of "suitability for machine learning" for labour-related tasks [Brynjolfsson *et al.*, 2018a], which draws upon particular technologies in machine learning only. Here, we use a more comprehensive list of AI tasks and benchmarks (extendable to future developments). Our approach relates most to [Felten *et al.*, 2018], who also link AI benchmarks to work-related abilities, although we draw from more sources for AI benchmarks and we measure AI intensity differently, to ensure better comparability between benchmarks with different scales.

3 Methodology

In this section we briefly explain the construction of the framework (we refer the reader to [Tolan *et al.*, 2021] for further information). We map between the three layers: (1) tasks (2) cognitive abilities, and (3) AI research.

Following the framework in Figure 1 from left to right, we construct an index for 119 standardised occupations (from [Fernández-Macías *et al.*, 2018]), using information on each occupation's intensity of 59 tasks ($\mathbf{t^o}(59 \times 1)$) we draw from the framework developed by [Fernández-Macías and Bisello, 2022]. These are in turn linked with 14 cognitive abilities (merging several categorisations in psychology, animal cognition and AI, originally introduced by [Hernández-Orallo and Vold, 2019]). Then, these 59 links per occupation and cognitive ability are sorted into the 14 task indices. We also link the 14 cognitive abilities to 328 AI benchmarks (based on previous analysis and annotation of AI papers [Hernández-Orallo, 2017; Martínez-Plumed *et al.*, 2018; Martinez-Plumed and Hernandez-Orallo, 2018; Martínez-Plumed *et al.*, 2020]. We analyse the activity level or *intensity* for a benchmark, measured in terms of the production (e.g., outputs such as research publications, news, blog-entries, etc) from the AI community ($\mathbf{b}(328 \times 1)$). In total, we construct the framework based on information for 119 standardised occupations, 59 tasks, 14 task indices, 14 cognitive abilities, and 328 AI benchmarks. We summarise the notation in Table 1. A more detailed explanation of this notation follows below.

Work Tasks to Cognitive Abilities

The mapping between labour-related tasks and cognitive abilities followed a multidisciplinary *Delphi Method* approach. Starting with matrix $\mathbf{\Omega}(59 \times 14)$ where the number of task variables is the row dimension and the number of cognitive abilities is the dimension in the columns, 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. To obtain matrices Ψ^o, the task intensity-ability matrices for every occupation $\mathbf{o}$, we take the Hadamard product of every vector $\mathbf{t^o}$ with the correspondence matrix $\mathbf{U}$.

Note that these 59 tasks are used to generate the left side of the task framework presented in [Fernández-Macías and Bisello, 2022] that consists of 14 task indices. That is, different sets of t^o contribute to the same concept of tasks which in turn require similar sets of abilities. To avoid that the ability scores are driven by data availability of tasks, we orthogonalise task information by averaging over task sets (rows of matrices $\mathbf{\Psi^o}(59 \times 14)$) that are assigned to the same task index. For each occupation $\mathbf{o}$ this reduces the rows of matrices $\mathbf{\Psi^o}(59 \times 14)$ from 59 (number of tasks) to 14 (number of task indices) which yields the task indices-ability matrix $\mathbf{\Phi^o}(14 \times 14)$.

In order to take into account the number of task indices that a cognitive ability is assigned to, we sum over all task indices linked to the same cognitive ability for each occupation. In addition, we take into account the additional complexity of combining multiple abilities in one occupation by normalising the ability-specific task intensities such that the sum of scores within each occupation is equal to one: $\sum_i \phi^{\mathbf{o}}_{\mathbf{i,j}} / \sum_i \sum_j \phi^{\mathbf{o}}_{\mathbf{i,j}} = \mathbf{w^o}$. Stacking each vector $\mathbf{w^o}$ yields matrix $\mathbf{W}(119 \times 14)$ which indicates the relative required intensity of each of the 14 cognitive abilities in each of the 119 occupations.

AI Benchmarks to Cognitive Abilities

Similar to the mapping between cognitive abilities and tasks, we link these 14 cognitive abilities to the data on AI benchmarks. 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 $\mathbf{b}$ of length $|\mathbf{b}| = 328$ and the 14 cognitive abilities, a 1 is put in an ability-benchmark correspondence (or mapping) matrix $\mathbf{R}$ (14×328) if an ability is inherently required, i.e., absolutely necessary to solve the respective benchmark. From here we can calculate the vector of relevance for each cognitive ability from the correspondence matrix $\mathbf{R}$ as $\sum_j \mathbf{r}_{ij}$ as row. We normalise the relevance by the total number of documents to obtain the ability-specific AI intensity vector $\mathbf{a}$: $\frac{\sum_j \mathbf{r}_{ij}}{\sum_i \sum_j \mathbf{r}_{ij}} = \mathbf{a}$

Combining Occupations and AI Through Abilities

We combine AI benchmarks to labour-market information using the common link to cognitive abilities. For this purpose we take the Hadamard product of matrix $\mathbf{W}$ with the respective AI research intensity vector $\mathbf{a}$. We obtain a single AI exposure score for each occupation by taking the sum over the rows of matrix $\mathbf{V}$, i.e. $\sum_j \mathbf{v}_{ij}$. The final score indicates which of the studied occupations are relatively more likely to be affected by AI research intensity (i.e. which occupations are more exposed to AI progress) in the analysed cognitive abilities. For illustrative purposes we normalise this score, which we call AI exposure score, to a $[0, 1]$ scale.

4 Results

Tasks and Cognitive Abilities

The ability-specific intensity matrix $\mathbf{W}(119 \times 14)$ shows, for every occupation, the relevance of each cognitive ability against the remaining cognitive abilities. Figure 2, which plots the elements in matrix $\mathbf{W}$, reveals two aspects about the relevance of cognitive abilities across occupations: variation and level. In terms of variations, the length of the bands reveals which abilities *vary* most in terms of their relevance across occupations. Many abilities have similar relevance variation where *quantitative reasoning* (QL), and *sensorimotor interaction* (SI) depict clear exceptions. Moreover, the variation in relevance for *communication* (CO) as well as *comprehension* (CE) is also noteworthy. This means that, from the perspective of work content, high AI research intensity in QL, SI, CO or CE exhibits the largest differences in terms of likelihood of AI impact across occupations, while

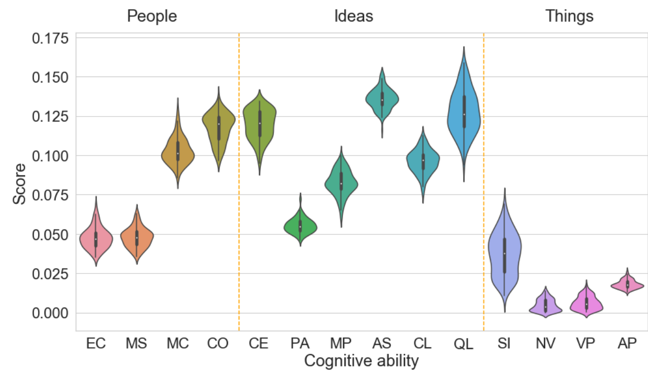

Figure 2: Distribution of ability-specific scores across occupations. We group the abilities into: (1) *people*, (2) *ideas*, and (3) *things*.

AI research intensity in the other abilities could potentially affect most occupations equally.

In terms of relevance levels, the figure shows that for most occupations, abilities of the categories *people* and *ideas* are more relevant at the workplace than *things* abilities.This implies that any AI that performs well on *ideas* or *people* abilities yields more occupational exposure than an AI that performs well on *things* abilities.

AI Research Intensity in Cognitive Abilities

Vector $\mathbf{a}(14 \times 1)$ indicates for each cognitive ability the relative AI research intensity. We illustrate this vector in Figure 3, which shows the computed AI research intensity for each cognitive ability from 2008 to 2018. The figure shows 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 those more ambient-related abilities belonging to the *things* category, namely, navigation and interaction with the environment.

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 as well as the challenge (today) of developing agents able to properly perform in social contexts. Finally,

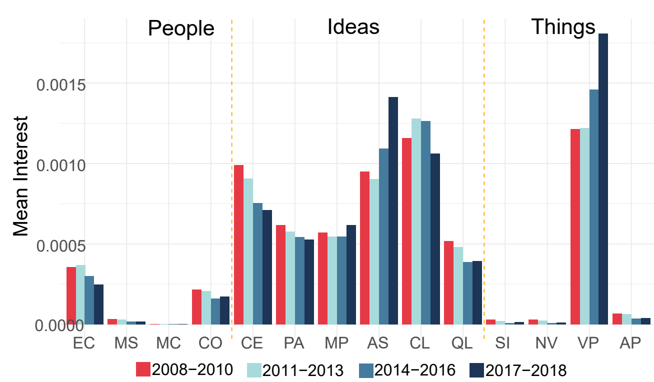

Figure 3: Relevance per cognitive ability weighted by (average) rate intensity for different periods of years over 2008-2018.

Figure 3 also shows a clear increasing (relative) trend in *visual processing* (VP) and *attention and search* (AS), while other abilities remain more or less constant or have a small progressive decline.

4.1 AI Exposure Score

Using the AI research intensity scores from 2018, we compute matrix $\mathbf{V}(119 \times 14)$, the ability-specific matrix of AI exposure scores. This score indicates which of the studied occupations are relatively more likely to be affected by AI research intensity through which cognitive ability.

Figure 4 depicts the computed AI exposure score differentiated by cognitive abilities and their categories for nine selected occupations. First, the figure shows that general office clerks, medical doctors and teachers are more exposed to AI research intensity than occupations that require comparatively lower skills such as cleaners, waiters or shop salespersons. This may be surprising, because these are occupations that were less affected by previous waves of automation. However, we would like to emphasise that our analysis does not focus on the automation potential of AI, but on what kinds of occupations are more likely to be affected by current developments of AI.

Second, Figure 4 shows that most of AI exposure is driven by its impact on tasks that require abilities that deal with *ideas*, such as *comprehension* (CE), *attention and search* (AS) as well as *conceptualisation* (CL). Compared to this, the exposure scores in the *things* category are negligibly small. However, our findings based on the tasks and occupation data indicate a relatively high need for *people* abilities in most occupations and a relatively low need for abilities dealing with *things*. Equivalently, the findings on AI research intensity suggest high activity in AI areas that contribute to abilities dealing with *things* but also to the abilities with the highest

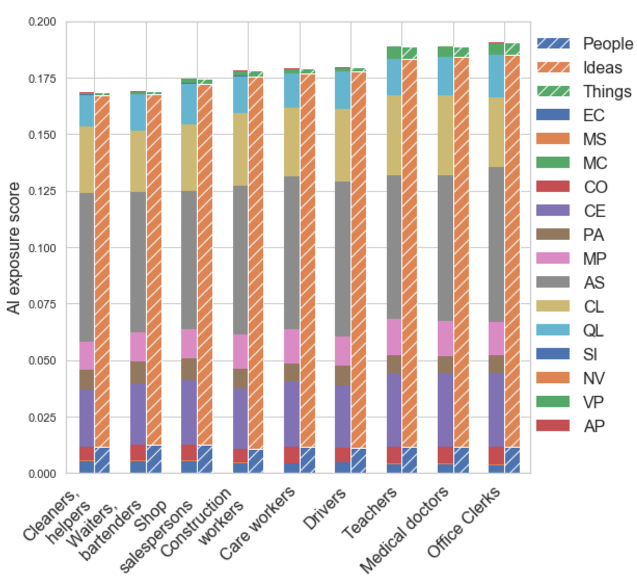

Figure 4: Ability-specific AI exposure scores for selected occupations. Right: grouped by people, ideas or things abilities. Left: detailed for the 14 abilities.

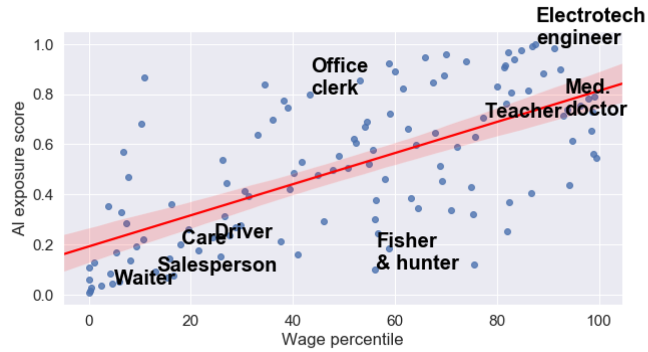

Figure 5: Scatterplot and best fit line, AI exposure score (percentiles) against wage percentiles

exposure score mentioned above, and low activity for abilities dealing with *people*.

Furthermore, in Figure 5 we plot the AI exposure score (in percentiles) against average wage percentiles of each studied occupation [Eurostat, 2014]. The figure shows a positive relationship between wages and AI exposure. That is, high-income occupations seem more likely to be affected by AI research intensity than low-income occupations. It is well possible that some very basic skills that are taken for granted for every human, such as naive physics (moving around and manipulating objects), language fundamentals (using language at the basic level, following orders) and naive psychology (understanding agency in other people), which are usually captured under the term common sense in AI [Davis and Marcus, 2015], are not fully represented in the descriptions that are used by the work intensities. But these may be required by all occupations and cover things, ideas and people. This can have different implications for occupational change (and consequently inequality) depending on the effect of AI on work. If the effect is labour-replacing, it could potentially lead to a reduction in income inequality [Webb, 2020]. If it is labour-enhancing, it could potentially lead to occupational upgrading effects and an expansion of income inequality.

5 Conclusion

The framework presented allows to accurately assess the technological potential 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 exposed to AI progress. We find that some jobs that were traditionally less affected by previous waves of automation may now be subject to relatively higher AI exposure. Moreover, we find that most of the AI exposure occurs through abilities that we use to deal with *ideas*. In light of the digital transformation and the rise of AI, these findings can help policymakers in directing their response in the form of education and (re-)training policies, and inform individuals in their career choice. In addition, breaking down the occupational effect to tasks and cognitive abilities can inform employers in the restructuring of occupations and tasks as the framework also informs about the particular capacities (abilities) within a task that may be supported by AI.

References

[Acemoglu and Restrepo, 2018] Daron Acemoglu and Pascual Restrepo. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *Ame. Economic Review*, 108(6):1488–1542, 2018.

[Agrawal *et al.*, 2015] Ajay Agrawal, John Horton, Nicola Lacetera, and Elizabeth Lyons. Digitization and the contract labor market: A research agenda. In *Economic Analysis of the Digital Economy*, pages 219–250. U. of Chicago Press, 2015.

[Autor and Dorn, 2013] David Autor and David Dorn. The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *Ame. Econ. Rev.*, 103(5):1553–97, 2013.

[Autor, 2015] David Autor. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3):3–30, 2015.

[Brown *et al.*, 2020] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language Models are Few-Shot Learners. *arXiv:2005.14165*, 2020.

[Brynjolfsson *et al.*, 2018a] Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. What Can Machines Learn, and What Does It Mean for Occupations and the Economy? In *AEA Papers and Proceedings*, volume 108, pages 43–47, 2018.

[Brynjolfsson *et al.*, 2018b] Erik Brynjolfsson, Daniel Rock, and Chad Syverson. 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, 2018.

[Davis and Marcus, 2015] Ernest Davis and Gary Marcus. Commonsense Reasoning and Commonsense Knowledge in AI. *Communications of the ACM*, 58(9):92–103, 2015.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A Large-Scale Hierarchical Image Database. In *CVPR*, pages 248–255, 2009.

[Eurostat, 2014] Eurostat. STRUCTURE OF EARNINGS SURVEY (SES). https://ec.europa.eu/eurostat/web/microdata/structure-of-earnings-survey, 2014.

[Felten *et al.*, 2018] Edward W Felten, Manav Raj, and Robert Seamans. A Method to Link Advances in Artificial Intelligence to Occupational Abilities. In *AEA Vol 108*, pages 54–57, 2018.

[Fernández-Macías and Bisello, 2022] Enrique Fernández-Macías and Martina Bisello. A comprehensive taxonomy of tasks for assessing the impact of new technologies on work. *Social Indicators Research*, 159(2):821–841, 2022.

[Fernández-Macías *et al.*, 2018] Enrique Fernández-Macías, Emilia Gómez, José Hernández-Orallo, Bao Sheng Loe, Bertin Martens, Fernando Martínez-Plumed, and Songül Tolan. A multidisciplinary task-based perspective for evaluating the impact of AI autonomy and generality on the future of work. *CoRR*, abs/1807.02416, 2018.

[Ferrucci, 2012] David A Ferrucci. Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3.4):1–1, 2012.

[Hernández-Orallo and Vold, 2019] José Hernández-Orallo and Karina Vold. AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. In *Proc. AAAI/ACM AIES*, pages 507–513, 2019.

[Hernández-Orallo *et al.*, 2017] José Hernández-Orallo, Marco Baroni, Jordi Bieger, Nader Chmait, David L Dowe, Katja Hofmann, Fernando Martínez-Plumed, Claes Strannegård, and Kristinn R Thórisson. A New AI Evaluation Cosmos: Ready to Play the Game? *AI Magazine*, 38(3), 2017.

[Hernández-Orallo, 2017] José Hernández-Orallo. Evaluation in Artificial Intelligence: From Task-Oriented to Ability-Oriented Measurement. *AI Review*, 48(3):397–447, 2017.

[Martinez-Plumed and Hernandez-Orallo, 2018] Fernando Martinez-Plumed and Jose Hernandez-Orallo. Dual indicators to analyze AI benchmarks: Difficulty, discrimination, ability, and generality. *IEEE Transactions on Games*, 12(2):121–131, 2018.

[Martínez-Plumed *et al.*, 2018] Fernando Martínez-Plumed, Shahar Avin, Miles Brundage, Allan Dafoe, Sean Ó hÉigeartaigh, and José Hernández-Orallo. Accounting for the Neglected Dimensions of AI Progress. *arXiv:1806.00610*, 2018.

[Martínez-Plumed *et al.*, 2020] Fernando Martínez-Plumed, Jose Hernández-Orallo, and Emilia Gómez. Tracking ai: The capability is (not) near. In *ECAI 2020*, pages 2915–2916. IOS Press, 2020.

[Nedelkoska and Quintini, 2018] Ljubica Nedelkoska and Glenda Quintini. Automation, Skills Use and Training. *OECD Social, Employment and Migration WP*, (202), 2018.

[Shoham *et al.*, 2018] Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, and Z Bauer. The AI Index 2018 Annual Report. *AI Index Steering Committee, Human-Centered AI Initiative, Stanford University.*, 202018, 2018.

[Tolan *et al.*, 2021] Songül Tolan, Annarosa Pesole, Fernando Martínez-Plumed, Enrique Fernández-Macías, José Hernández-Orallo, and Emilia Gómez. Measuring the occupational impact of ai: tasks, cognitive abilities and ai benchmarks. *JAIR*, 71:191–236, 2021.

[Webb, 2020] Michael Webb. The Impact of Artificial Intelligence on the Labor Market. *Department of Economics, Stanford University, Stanford, California*, 2020.