Working paper series
No 6 / October 2021

ARTIFICIAL INTELLIGENCE AND JOB AUTOMATION: AN EU ANALYSIS USING ONLINE JOB VACANCY DATA

Konstantinos Pouliakas

The **European Centre for the Development of Vocational Training** (Cedefop) is the European Union's reference centre for vocational education and training, skills and qualifications. We provide information, research, analyses and evidence on vocational education and training, skills and qualifications for policy-making in the EU Member States.
Cedefop was originally established in 1975 by Council Regulation (EEC) No 337/75. This decision was repealed in 2019 by Regulation (EU) 2019/128 establishing Cedefop as a Union Agency with a renewed mandate.

Europe 123, Thessaloniki (Pylea), GREECE
Postal: Cedefop service post, 570 01 Thermi, GREECE
Tel. +30 2310490111, Fax +30 2310490020
Email: info@cedefop.europa.eu
www.cedefop.europa.eu

Jürgen Siebel, *Executive Director*
Nadine Nerguisian, *Chair of the Management Board*

Please cite this publication as:
Pouliakas, K. (2021). *Artificial intelligence and job automation: an EU analysis using online job vacancy data*. Luxembourg: Publications Office of the European Union. Cedefop working paper; No 6. http://data.europa.eu/doi/10.2801/305373

Luxembourg: Publications Office of the European Union, 2021

PDF ISBN 978-92-896-3275-1
ISSN 1831-2403
doi:10.2801/305373
TI-BA-21-004-EN-N

Acknowledgements

The author is grateful to colleagues Jiri Branka, Vladimir Kvetan and Jasper van Loo for their invaluable support and to Giovanni Russo, Joanna Napierala and Matteo Sostero (JRC) for helpful comments. Much gratitude is also extended to Jiri Branka for the provision of the original Skills-OVATE database used in this paper and his continuous guidance. Special thanks are also extended to participants of the *New techniques and technologies for statistics* (NTTS 2021) conference, organised by Eurostat, for their useful comments. Finally, the author wishes to acknowledge the support of the editorial board of Cedefop's working paper series, especially Antonio Ranieri, Head of Department for VET and skills.

Contents

Non-technical summary

Not long before the coronavirus outbreak, popular fears about artificial intelligence (AI) algorithms and smart machines resulting in a jobless society were widespread; some cautioned that about half of all jobs in advanced economies may become extinct. These analyses were spurred by the will to understand the potential consequences of the so-called fourth (4th) industrial revolution for the future of work. Subsequent studies that have utilised a task decomposition approach to predicting the risk of machine automation have tended to dispel such fears of rampant job destruction. Analyses using Cedefop's European skills and jobs survey data have also illustrated that technological alarmism is unwarranted, as technological progress is typically associated with greater job-task variety and worker upskilling. Despite such rebuttals, concerns about accelerating automation in labour markets have resurfaced in the face of the COVID-19 crisis. The pandemic and associated social distancing measures have accelerated the incentives of companies and societies to adopt new digital and data-driven technologies, some of which have the intended purpose of replacing human labour.

This working paper focuses on identifying a set of workers' job tasks that are associated with higher occupational automation risk, as well as greater digital exposure to robots, computer software and AI technologies. Better understanding of which occupations, skills and job tasks may be displaced by AI and other digital technologies, especially within the context of the COVID-19 shock, is crucial for the formulation of preventive upskilling/reskilling and job design policies, including informing the European skills and digital agendas.

To achieve this, and in contrast to previous research that has used relatively broad individual survey data, the study utilises a novel big data set that contains information on the skills, work activities and technologies requested by EU employers. Specifically, the research utilises rich data collected by Cedefop via web sourcing of online job advertisements across EU countries: this is the so-called Skills OVATE database. Online job adverts have recently become a rich source of detailed information on skills and other job requirements, which are difficult to gather via traditional methods. They can provide granular and timely insights into labour market trends and possibly new and emerging jobs and skills. But there are also several significant constraints and biases in the use of such big data. Online job advertisement data can generally provide only a piecemeal and non-representative picture of required skills in labour markets. Posted vacancies online only represent employer's stated, and not necessarily their revealed, preferences for job applicants. In other words, we can only know what employers

say they want from potential job applicants, but we do not know if they eventually decide to fill the advertised position, what the underlying skills and qualities of selected job applicants are, and if these are a good match for the post or correspond to the advertised requirements.

Acknowledging such caveats, this paper utilises suitable machine learning methods to detect, using a deductive approach, the key work activities associated with higher automatability or digital exposure of occupations. By identifying which job tasks in different occupations are the best for predicting occupational automation risk, as well as higher digital exposure, the paper sets a scene to enable future policy design that may safeguard high-risk jobs or promote labour-augmenting AI technologies.

The analysis shows that core work activities associated with high risk of machine displacement occupations are those that rely on highly codifiable information retrieval and evaluation skills as well as routine, manual skills. Work activities that are instead relatively immutable to machine learning algorithms include those dependent on high socioemotional and interpersonal skills, managerial skills and problem-solving skills.

The paper further confirms that it is a common misnomer to associate AI technologies only with higher automation; it adds to better understanding of how AI may enhance worker productivity across multiple sectors. Work activities in jobs with greater exposure to computer software and industrial robots are loosely related to those that utilise AI technologies. Employers posting advertisements for occupations exposed to AI technologies are more likely to demand job applicants who can carry out high-skilled work activities, such as ‘evaluating information to determine compliance with standards’ or ‘thinking creatively’.

A key contribution of the paper is that it compares the performance of different predictive models that use information on occupational work activities from big data sets (such as Skills OVATE) with those obtained from occupational experts and incumbent workers (such as O*NET). It is shown that it is possible to predict correctly whether an occupation is automatable or not about 7 out of 10 times, using the big data set deployed in this paper and the best machine learning model. This prediction accuracy increases when information on work activities obtained from experts or incumbent work assessments is used instead. However, the analysis generally cautions that predicting the automatability of occupations is a challenging exercise, which depends on a complex interplay of many other factors in addition to the task content of occupations.

CHAPTER 1.
Introduction

Not long before the coronavirus outbreak, popular fears about artificial intelligence (AI) algorithms and smart machines resulting in a jobless society were widespread (Brynjolfsson and McAfee, 2014; Ford, 2015). A widely cited 2013 University of Oxford study cautioned that about half of all jobs in advanced economies may become extinct due to advancing machine learning methods (Frey and Osborne, 2013, 2017).

Subsequent studies that have utilised an approach of deconstructing jobs according to their task composition have tended to dispel such fears of rampant job destruction. This research has typically focused on estimation of automation risk in labour markets by relying on representative survey data from individual workers, such as the OECD's Survey of adult skills (Artnz et al., 2017; Nedelkoska and Quintini, 2018) and the Cedefop European skills and jobs survey (ESJS) (Pouliakas, 2018). The studies have sought to detect the latent relationship between the risk of automation and a limited set of broad skills and task groupings, enriching the original occupational estimates of Frey and Osborne (2017). Their main conclusions regarding the skills and tasks with higher probability of automation have been made at higher (2-digit) occupational level and for a relatively broad set of job tasks, due to data constraints associated with conventional labour market surveys. Brynjolfsson et al. (2018) is a notable exception, as these authors use data on workers from a popular crowdsourcing platform who were asked to rank jobs and direct work activities according to their susceptibility to machine learning replacement. However, such scores are obtained from a relatively limited number of respondents who are active in crowdsourcing platforms. They are dependent on workers' own assessments of the extent to which some tasks can be carried out in machine-readable format.

Concerns about accelerating automation in labour markets have resurfaced in the face of the COVID-19 crisis. The pandemic and associated social distancing measures have accelerated the incentives of companies and societies to adopt new digital and data-driven technologies (Coombs, 2020; IFOW, 2020). Most notable automation episodes in history have also tended to spike following major economic crises (Frey, 2019). Early predictions that COVID-19 will have a positive automation effect may, however, prove to be false. Occupations identified as high risk of COVID-19 exposure and social distancing disruption (Pouliakas and Branka, 2020) have been found to correlate weakly with those facing higher automation risk (Chernoff and Warman, 2020). Many of the occupations and sectors disproportionately affected by COVID-19 are typically in the service sector and heavily reliant on interpersonal skills (for instance, hospitality, leisure, retail), which are relatively less prone to replacement by automating technologies despite anecdotal evidence that they too are undergoing marked digital transformation.

The above findings highlight that in-depth understanding of which skills and job tasks may be displaced by AI and other digital technologies, especially within the context of the COVID-19 shock, and which occupations are more inclined to automation, is crucial for the formulation of preventive upskilling/reskilling and job design policies. The need to design effective training programmes that can enable individuals and firms to make the transition to a digital economy is a key aim, for instance, of the *European skills agenda* ([1]).

In contrast to previous research that has used representative, yet relatively broad individual survey data, this study utilises a novel big data set which contains information on the skills, work activities and technologies required by EU employers. Specifically, the research utilises rich data collected via web sourcing of online job advertisements (OJAs) in all EU countries by the European Centre for the Development of Vocational Training (Cedefop). This work has manifested in the development of the online vacancy analysis tool for Europe (Skills-OVATE) (Cedefop, 2019a, 2019b). This data enables in-depth exploration of the detailed task/skills profiles of narrow occupational groups that may be immutable or threatened by the proliferation of emerging digital technologies. It also allows for the creation of a predictive model of occupational automatability in the EU labour market based on a parsimonious and statistically significant feature set of work activities.

The aim of this paper is twofold. First, it applies deductive analysis, using suitable machine learning methods on the full Skills-OVATE data, to detect which key work activities are associated with higher automatability or digital exposure among occupations. Second, it leverages the sparse feature set from the first step and aims to detect a suitable machine or deep learning model with high predictive capacity that may be used with new input data in the future. Section 2 engages in a short review of the literature on the impact of technological innovation on employment and occupational automation risk. Section 3 describes the data, key variables and empirical methodology deployed. Section 4 presents key findings as extracted from online job advertisement data and examines their robustness by considering alternative metrics of work activity intensity within jobs and comparing with other external representative labour market sources. It also evaluates the predictive capability of several machine or deep learning models in terms of predicting occupations facing very high risk of automation. Section 5 concludes.

([1]) https://ec.europa.eu/social/main.jsp?catId=1223&langId=en

CHAPTER 2.
Literature review

Prominent economic theories highlight that technological innovation is associated with changing employment and skill needs. Theories of skill-biased technological change (SBTC) have noted that technological growth and associated skill demand outpacing skill supply is the main reason for underlying growing demand for higher educated workers and rising wage inequality, at the expense of the lower-skilled (Katz and Murphy, 1992; Katz and Autor, 1999). Job polarisation/routine-biased technological change theories imply, instead, growing demand for skills complementary to non-routine, analytical tasks, but also interpersonal skills, at the expense of those in routine and manual jobs (Autor et al, 2003; Goos, 2018).

Recent years have seen an upsurge in the number of studies and policy reports focusing on the impact of new digital technologies on jobs and skills. These analyses have generally been spurred by an increasing willingness to understand the potential consequences of the so-called fourth industrial revolution for the future of work (Brynjolfsson and McAfee, 2014; Bessen, 2015; Ford, 2015; World Economic Forum, 2016). Much of this literature has renewed interest in the old scientific question of whether innovation fosters technological unemployment and adverse distributional consequences for employment and wages. Characteristic of this literature has been the resurfacing of 'technological alarmism' (Autor, 2015; Mokyr et al., 2015) or the 'technological unemployment' hypothesis (Keynes, 1933): this brings widespread concerns that technological change, in the form of robotics and artificial intelligence (AI), will take over peoples' jobs and livelihoods.

Some estimates that bred such widespread attention indicated that close to half of all jobs in the US and UK are susceptible to replacement by machines (Frey and Osborne, 2013, 2017) and that technological progress and automation is a main driver of labour market polarisation (Autor et al., 2006; Goos et al., 2009). Most academic and policy attention has also focused in recent years on the need to support medium- and lower-skilled workers with appropriate reskilling policies, to secure their fast reintegration into the labour market and/or foster job mobility (World Economic Forum, 2018; McKinsey Global Institute, 2017). Even though more recent estimates of automatability, adopting a task-based approach, indicate a much lower risk of full job displacement by machines (Arntz et al., 2017; Nedelkoska and Quintini, 2018; Pouliakas, 2018), they too highlight that it is predominantly lower-educated workers who are most susceptible to job and incomes losses as a result of new automating technologies.

The methodology used in these automation risk papers has generally been criticised, as it extrapolates to all workers predictions based on an estimated automatability equation from a small training set of occupations. This latter training set is usually identified by the informed, albeit subjective, views of experts. Due to survey data constraints, the matching of the detailed group of occupations that comprise the training set to data sets that contain more specific information on occupational job tasks and skill needs has generally been done at a broad occupational level (2-digit), introducing measurement error. More important, assessments about future automation risk are static, as they are bound by the current task-set of occupations and fail to acknowledge adequately that automation may imply changing task content within jobs (Acemoglu and Restrepo, 2018).

By focusing narrowly on the automation properties of machines, this literature has thus generated and sustained, in the words of Acemoglu and Restrepo (2018), a ‘false dichotomy’ about the impact of technological progress on labour market outcomes. It has side-tracked the debate from a fuller understanding of the impact of technological change on labour and skill demand and its associated effect on labour productivity. As recently modelled by Acemoglu and Restrepo (2019), the history of automation and technological change in the 19th and 20th centuries is one of task (re)generation, whereby the task content of production typically expands as a result of new or a broader range of tasks and skill needs emerging. Obtaining a satisfactory understanding of the way technological progress affects labour and skill demand, and its impact on productivity growth, is hence dependent on whether such task reengineering – a so-called reinstatement effect – acts as a countervailing force to the displacement effect.

Focusing solely on the automating properties of new digital technologies also provides a one-sided story of their impact on skill needs. Fossen and Sorgner (2019), for instance, have investigated the heterogeneous effects of new digital technologies on individual-level employment and wage-dynamics in the U.S labour market. They find a significant impact of high computerisation risk on individuals’ labour market outcomes, such as deceleration in wage growth. They highlight, however, that advances in artificial intelligence (AI) software are likely to improve an individual’s job stability and wage growth, in contrast to computer technologies that aim at replacing routine job tasks. While automating digital technologies tend to disproportionately affect middle-skill occupations, facilitating job polarisation, digital technologies such as AI that increasingly crowd out cognitive tasks are found mostly to affect higher educated workers.

A recent analysis by Webb (2020) further strengthens this point by accounting for such heterogeneity in digital technologies. By exploiting the overlap between the text of job task descriptions and that used in patents, the author derives three distinct digital exposure indices measuring whether occupational tasks may be affected by industrial robots, computer software or AI technologies. Although the analysis finds that occupations highly exposed to automating technologies saw historical declines in employment and wages, it confirms that AI technologies are mostly directed at high-skilled tasks.

Cedefop (2020) has further highlighted that automation risk estimates tend to focus on the technical, not the economic, feasibility of automation. The latter should consider that firms' incentives to automate depends on complex and interrelated interactions between the true 'business case' for adopting new technologies, their cost, diffusion hurdles, relative supply and price of skill and labour, uncertainty ('animal spirits') in investment decisions and shifting social attitudes. In this sense, there is significant residual heterogeneity in firms' automation decisions unaccounted for by the task content of occupations. Considering (some of) such factors reveals that the effect of new technologies on future employment gains is mediated by companies' individual performance management schemes and a consultative environment between management and workforce.

But even if the net effect of technological innovation on jobs and skills is positive or neutral, the adjustment process of an economy to the introduction of new technologies is expected to be slow and mediated by the extent to which technological progress may render workers' skills obsolete and the degree to which a mismatch is created between the requirements of new technologies and skills. Using data extracted from online job vacancies as in this paper, the theoretical framework of Deming and Noray (2020) demonstrates that technological advancements can both erode and enhance demand for different skills in occupations. Using data from Cedefop's first European skills and jobs survey, McGuinness et al. (2021) have recently confirmed such transmission mechanisms of technological progress. They show that any skills-displacing effects of new technologies tend to be outweighed by new task creation and associated skill formation/upskilling taking place within occupations.

What the above studies highlight is that obtaining in-depth understanding of the specific skills and job tasks that may or may not be associated with jobs at risk of automation is imperative. Such knowledge is central to designing informed upskilling and reskilling policies that can enable individuals and companies to cope with the emerging skill demands posed by, among other drivers, new digital technologies.

CHAPTER 3.
Methodology

3.1. Use of online job advertisement database

To obtain deeper understanding of the core type of work activities and skills needs that are associated with higher occupational automatability, this paper utilises data from the new online vacancy analysis tool for Europe (Skills-OVATE) covering the period from July 2018 until July 2020 ([2]). The tool has been developed by the European Centre for the Development of Vocational Training (Cedefop) and aims to utilise information on skill demands available in online job advertisements (OJAs) to generate faster and more detailed EU skills intelligence. Skills-OVATE offers detailed information on jobs, and the skills employers want as stated in online job advertisements, with data collected from all European countries. Data from millions of OJAs coming from thousands of sources, including private job portals, public employment service portals, recruitment agencies, online newspapers and employer websites, are available in the tool ([3]).

OJAs have recently become a rich source of detailed information on skills and other job requirements, including analyses of the impact of digital technologies on skills (Acemoglu et al., 2020), which are difficult to gather via traditional methods. With the advent of greater digitalisation, jobseeker and employer behaviour is shifting towards a growing use of online recruiting and job search. Similarly, technological change, together with skill shortages faced by employers, is inducing them to rely more often on web-based channels (e.g., online platforms) as a means of attracting key professionals or employees in possession of specific skills and characteristics.

OJA analysis can provide granular and timely insights into labour market trends and possibly new and emerging jobs and skills. But it is crucial to acknowledge that meaningful analysis requires sound understanding of the online labour market in different countries and awareness of the key challenges in using OJAs for skills and labour market analysis. OJA data generally provide only a piecemeal and non-representative picture of emerging technologies and skills in

---

([2]) We use a pooled cross-section of the data as opposed to its time-series, given challenges of ensuring data consistency across time and series breaks.

([3]) The tool is continuously improved, and its data quality strengthened, while system functionalities are regularly expanded. An extensive description of the methodology used, including associated steps of data digestion by online job portals, data cleaning and pre-processing and information extraction using suitable machine learning classification methods, is available in Cedefop (2019a, b).

labour markets. The information extracted reflects the type of technologies, tools and skills that employers request from job applicants, but this knowledge is likely to be bounded by the technology they currently use and a heterogeneous group of employers have incentives to overrepresent their skill needs (Gambin et al., 2016). Further, data extracted from OJAs are often fraught with statistical and selection biases and may only be loosely related to the 'actual' skill needs in jobs ([4]).

It is also clear that the overall skills distribution within occupations as revealed by online job ads may represent a subset of their total skill requirements. Not all skills are listed in job ads, since job-specific skills may be taken for granted and transversal skills may be (over)emphasised instead. Online job postings also serve the function of a 'beauty contest', aiming to attract potential job applicants to the recruitment stage and to overcome the 'adverse selection' problem associated with their unobserved abilities (Cedefop, 2019b; Akerlof, 1970). It is possible, therefore, for online recruiting to encourage superfluous vacancy postings by employers and inferior skills matching outcomes, the latter an outcome of a large share of unsuitable applicants attracted per vacancy (Gürtzgen et al., 2021). Vacancies posted online only represent the employer's stated, and not necessarily their revealed, preferences for job applicants who are eventually hired by firms.

3.2. Measuring work activity intensity

While acknowledging that there are caveats, the analysis in this paper deploys a new version of Skills-OVATE that has been classified using the O*NET hierarchical structure. Specifically, all detailed skill-requirement terms collected by OJAs have been clustered according to the following O*NET taxonomy: abilities (originality, fluency of ideas, oral expression); knowledge (language, computers and electronics, sales and marketing, personnel and human resources, administration and management); technology (Office suite software, web platform development, object or component oriented development software); skills (complex problem solving, time management, programming, management of financial resources);

([4]) Online job portals do not cover most vacancies, which tend to be filled via word of mouth, while the availability of OJA is affected by socioeconomic context and the digital divide in countries. Representativeness also varies due to occupation-specific hiring strategies and coverage of different labour market segments varies by public or private job portals, with some restricting or regulating access of different groups. OJAs represent flow data, so it is expected that jobs with above-average turnover will be overrepresented in the samples. Some ads can represent multiple vacancies, or even no vacancy at all, given the low cost of posting online job ads and the phenomenon of some employers only posting jobs online to scan potential available candidates in the labour market (so-called 'ghost' vacancies) (Cedefop, 2019a, 2021).

work activities (interacting with computers, organising, planning and prioritising work, thinking creatively, assisting and caring for others, communicating with persons outside the organisation, handling and moving objects, controlling machines and processes); and work styles (adaptability, cooperation, initiative, dependability, etc.) ([5]).

Each skill-requirement term incorporated within the main areas described above is mapped to four-digit occupational groups as described by the International standard classification of occupations (ISCO). While Skills-OVATE contains over 2000 skill-requirement terms in total, the third level of the O*NET-based hierarchical structure is used, which maps these to 142 broader skill-requirement cluster families. We calculate for each skill-requirement cluster, s_i $(i=1 \ldots S)$, its relative frequency or yield (S^f) across all job advertisements, j, in an occupation, o, as follows:

$$S^f = \frac{\sum_{j=1}^{n} s_{ijo}}{\sum_{j=1}^{n} j_o} \qquad [1]$$

where the frequency is computed by the ratio of the total number of mentions of each skill-requirement in the job advertisements of a given 4-digit occupation over the total number of posted advertisements in the occupation.

For empirical but also theoretical tractability reasons we also derive and use the discrete occurrence of a given skill-requirement cluster within an occupation. We have created binary variables $\{s_{di} = 0{,}1\}$ denoting whether or not a given skill-requirement cluster is mentioned in the OJAs of an occupation. This is done because there is a high degree of skewness in the distribution of the skill-requirement frequency scores across occupations. Some are evenly represented in most occupations and others are heavily concentrated in few occupations and not mentioned at all in others, distorting the estimation due to insufficient occupational variation. On theoretical grounds it is also reasoned that the discrete requirement of a given skill or work activity by employers in occupations could constitute in itself a bottleneck to their replacement by machines.

To account for the fact that the occurrence of a given skill-requirement in occupations may be accompanied by a very low yield in some of them, we further corroborate the robustness of the findings by excluding instances with S^f below 1% in a given occupation. The analysis is replicated using an alternative measure of the importance of skill-requirements within occupations, the so-called revealed

([5]) Each broad 'skill-requirement cluster' is underlined by an extensive range of individual, more detailed, terms e.g. 'interacting with computers' comprises 'use a computer'/'use Microsoft Office'/'use office systems'/'use spreadsheets', etc. See Annex Table A.1.

comparative advantage (RCA). This measure ([6]) aims to correct for the inability of skills frequency scores to normalise for skills or other job characteristics that generalise across all occupations in high frequencies ([7]) (Dawson et al., 2020). The RCA measure, by contrast, weights the frequency of a skill term in an occupation by its total importance in the entire universe of job advertisements. Formally, the RCA for skill-requirement *s* and job ad *j* is:

$$RCA_{tj} = \frac{\sum_{j=1}^{n} s_{ijo} / \sum_{j=1}^{n} j_o}{\sum_{j\epsilon J, o\epsilon O} s_{ijo} / \sum_{j\varepsilon J, o\epsilon O} j} \qquad [2]$$

where *O* is the set of all distinct occupations and *J* the set of all job ads in the data set. The RCA adjusts for the biases that emerge from high-occurring terms in all jobs by weighting by the total share of demand for a given skill requirement across all occupations and job ads.

For the purposes of this paper, and in alignment with the task approach to analysing the extent to which some jobs comprise codifiable or non-programmable tasks (so called ‘engineering bottlenecks’), we focus mainly on the set of 39 ‘work activities’ in the database (see Annex Table A.2 for descriptive statistics). These variables reflect the more objective technical requirements of employers, as expressed in OJAs, corresponding to the implied structure of work within occupations. They provide a more deterministic description of the activities that workers are asked to carry out in their jobs, as opposed to the more subjective and generic descriptions related to knowledge, skills and work styles. It is also possible to engage in robustness analysis of such indicators, given that similar information exists from other representative labour market surveys (such as O*NET) ([8]).

([6]) The RCA is used in trade economics as a proxy for key export orientation of a country in relation to total world production.

([7]) For example, in the Skills-OVATE data set this would include terms such as ‘interacting with computers’, ‘communicating with persons outside the organisation’, ‘organising and planning work’ or ‘adaptability skills’, which are mentioned in between 30-70% of all job advertisements. For comparison, the skill-requirement ‘repairing and maintaining electronic equipment’ is mentioned in only 0.3% of all job adds during the given period, but accounts for as much as 14.4% of the total advertisements mentioned in the specific occupation ISCO 2522 ‘System administrators’.

([8]) Such an approach has also been recently espoused, for instance, by Bisello et al. (2021) in the creation of a comprehensive European database of task indices. It is nevertheless possible for the analysis to also explore the relationship between different skills and technologies and the automation propensity of occupations. We avoid engaging in such an exercise in this paper in the absence of a strong theoretical framework underpinning such analysis, in contrast to the well-documented task approach of labour economics.

3.3. Linking Skills-OVATE with digital exposure indices

The primary aim of this paper is to detect which job tasks are associated with occupations with higher predicted probability of automation or digital exposure, using a new big data set of OJAs in Europe. Given the possible inadequate representativeness of certain labour market segments in OJAs, as described in section 3.1, we do not engage in estimation of automatability risk using the Skills-OVATE data but match it with the externally derived automatability and digital exposure indices of Frey and Osborne (2017) and Webb (2020), respectively ([9]).

We first match the estimated scores of automatability as derived by Frey and Osborne (2017) at the four-digit occupational level, using an appropriate correspondence (cross-walk) matrix between the O*NET and ISCO occupational codes. These automation risk scores have been estimated after asking experts to assess the extent to which the tasks of 70 minor occupations (a training set) can be codified by machine learning algorithms. The focus was on task characteristics that may or may not constitute ‘engineering bottlenecks’ to their replacement by machine learning programmes: physical or manual dexterity, social or creative intelligence. After deriving a suitable machine learning classifier on the training set, the authors subsequently estimate automation risk scores for 702 O*NET occupations.

To account for heterogeneity in digital technologies and to consider that not all of them solely have automating properties, we also use the recent digital exposure indices developed by Webb (2020). These capture the susceptibility of tasks and occupations to different types of digital technologies: computer software, industrial robots and AI. Using the same classification matrix that allows mapping of O*NET occupations to ISCO, these digital indices have been matched to the corresponding 4-digit occupations available in the Skills-OVATE data set.

3.3.1. Empirical strategy

Following the matching of the Frey and Osborne and Webb digital exposure indices to the Skills-OVATE data, the merged database contains detailed information on

([9]) Such estimation of predicted automation risk is feasible and can be done after matching the Frey and Osborne training set of minor O*NET occupations, which were assessed by experts in terms of their potential susceptibility to machine learning replacement, with their respective 4-digit ISCO counterparts in Skills-OVATE. However, as 18 occupations among the original 70 of the Frey and Osborne training set cannot be effectively matched to the Skills OVATE data, and since the Frey and Osborne estimates have been corroborated and used widely in research, it was preferred to use the latter as the main dependent variable in this paper.

the automation risk, digital intensity (robots, software, AI) and specific skill-requirements of 379 detailed occupations in European labour markets ([10]).

To examine the link between the job tasks advertised by employers in their job postings and the technological exposure of occupations, we estimate the following multivariate linear regression model:

$$\boldsymbol{D_o} = \boldsymbol{\beta}'\boldsymbol{S_{io}} + \boldsymbol{u_o} \qquad [3]$$

where D is a vector of either occupational machine learning risk (Frey and Osborne, 2017) or digital intensity scores (Webb, 2020) for a sample of 4-digit occupational groups and $\boldsymbol{S}$ is a [379 x 39] matrix of occupational work activities (measured by their frequency, occurrence or RCA scores) as extracted from the OJAs of European employers. $\boldsymbol{\beta}$ is the coefficient vector to be estimated and **u** the unobserved effects assumed to be i.i.d ~ (0,1).

Given that estimation of equation (3) is affected by a high degree of collinearity among the independent variables ([11]), the estimation strategy deploys appropriate machine learning methods to detect a minimum set of least correlated and significant feature variables to be retained in the regression. Specifically, least absolute shrinkage and selection operator regressions (Lasso) with cross-validation lambda selection method are carried out ([12]) to eliminate the weights of the least important feature variables by minimising the following cost function:

$$C(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha\sum_{i=1}^{n}|\beta_i| \quad [4]$$

Following estimation of equation (4), a sparse model is retained that links the most important feature variables with the relevant automation or digital intensity indices. For the purposes of obtaining correct inferential statistics (especially standard

([10]) The original Skills-OVATE data set is not matched to twenty-five 4-digit occupations available in the Frey and Osborne and Webb data sets. Skilled agricultural workers (major ISCO code 6) as well as some craft and related trades workers (major ISCO code 7) and few high-skilled occupations are particularly not well-represented in Skills-OVATE.

([11]) For example, half of the work activity independent variables have variance inflation factors that exceed a value of two, once undertaking ordinary least squares estimation of equation (3).

([12]) As Lasso regressions can also behave erratically when several feature variables are strongly correlated, robustness checks are also performed using elastic net regressions, which balance the regularisation term the cost function (4). Elastic regressions have been run in all instances following Lasso ones to ensure agreement in the minimum retained set of feature variables and the selected parsimonious specifications.

errors of the selected coefficients), a double selection Lasso linear regression model is estimated as a final stage of the empirical procedure.

CHAPTER 4.
Empirical findings

4.1. Descriptive analysis

Before engaging in multivariate empirical analysis, Figure 1 depicts the raw difference in the mean occurrence of each work activity between occupations with a very high probability of automation against those with medium or very low risk. We follow the convention adopted in the literature, whereby occupations are deemed to have very high automation risk if the estimated probability is greater than 0.7. The mean difference is shown for all 39 work activities included in the Skills-OVATE database.

Figure 1. **Difference in mean occurrence of work activities between occupations with very high and medium or low automation risk**

NB: The occurrence of work activities in occupations is obtained from Cedefop's European database of online job advertisements (Skills OVATE); occupations are distinguished into highly automatable (below highly automatable) if the Frey and Osborne (2017) estimated probability of automation is greater than 0.7 (below 0.7).

Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores

From the figure it is evident that the job ads of highly automatable occupations are more likely to mention work activities related mainly to 'inspecting equipment, structures or materials', 'evaluating information to determine compliance with standards', 'operating vehicles, mechanised devices or equipment', 'maintaining and repairing mechanised equipment' or 'controlling machines and processes'.
By contrast, occupations characterised by lower automation risk have greater reliance on managerial and interpersonal tasks, such as 'guiding, directing and motivating subordinates', 'communicating with persons outside the organisation', 'communicating with supervisors, peers or subordinates', 'establishing and maintaining interpersonal relationships', or 'developing and building teams'. The same holds for leadership tasks such as 'coaching and developing others' or 'provide consultation and advice to others'. They are also more reliant on reasoning and cognitive tasks, such as 'thinking creatively', 'analysing data or information', 'judging the qualities of things, services or people' or 'making decisions and solving problems'.

Occupations with higher degree of interaction with computers are, perhaps contrary to expectations, found to be less likely to be associated with occupations prone to machine substitution. This potentially reflects the fact that most computer interaction demanded by employers tends to involve tasks that require a relatively moderate digital skill level, which complement, rather than eliminate, information processing tasks (Cedefop, 2018).

4.2. Machine-job displacement and required work activities

Analysis of EU employers' OJAs using a multivariate Lasso regression technique reveals a relatively parsimonious set of work activities associated in a statistically significant manner with higher occupational automation risk and digital exposure.
Table 1 shows that core work activities susceptible to machine learning displacement include 'inspecting equipment, structures or materials', 'evaluating information to determine compliance with standards', 'updating and using relevant knowledge', 'recording or documenting information' and 'operating vehicles, mechanised devices or equipment' ([13]). 'Assisting and caring', which mostly comprises customer assistance activities, is also positively related to higher

([13]) The findings do not deviate significantly when using the frequency of skill requirements variables (as derived in equation 1) in the explanatory set. Results of OLS regressions that include the full set of feature variables are available in the Annex Table A3.

automation risk (14). By contrast, work activities that are relatively insulated from higher risk of machine replacement include 'coaching and developing others', 'communicating with persons outside the organisation', 'communicating with supervisors, peers or subordinates', 'guiding, directing or motivating subordinates', 'training and teaching others', 'judging the qualities of things, services or people', 'thinking creatively' or 'repairing and maintaining electronic equipment'.

Table 1. **Occupational automation risk/digital technologies exposure and occurrence of work activities, Lasso regressions**

	(1)	(2)	(3)	(4)
Work activities	*Automation risk*	*AI*	*Software*	*Robots*
Identifying objects, actions and events	0.09			-0.03
	(0.073)			(0.070)
Inspecting equipment, structures or material	0.11**			0.18**
	(0.048)			(0.069)
Monitoring processes, materials or surroundings	0.04			
	(0.041)			
Assisting and caring for others	0.08**			
	(0.039)			
Coaching and developing others	-0.18***			
	(0.066)			
Communicating with persons outside organisation	-0.19***		-0.09**	-0.32***
	(0.044)		(0.035)	(0.084)
Communicating with supervisors, peers or subordinates	-0.15***	0.07		
	(0.053)	(0.053)		
Developing and building teams	-0.06			

(14) The work activity 'assisting and caring' is particularly prone to selection bias during the classification process of unstructured terms in online job ads to the work activities taxonomy. The reason is that highly relevant jobs such as nurses and other health or care workers tend to be subject to a high degree of occupational regulation in EU countries and would hence not be recruited via online channels. This may explain why this work activity is found to have a positive association with automation risk, as opposed to a negative one that would be expected *a priori* or as estimated with representative survey data sources.

Work activities	(1) Automation risk	(2) AI	(3) Software	(4) Robots
	(0.048)			
Establishing and maintaining interpersonal relationships	-0.06	0.05		-0.11
	(0.047)	(0.042)		(0.074)
Guiding, directing and motivating subordinates	-0.14***			-0.20***
	(0.055)			(0.071)
Monitoring and controlling resources	0.05			
	(0.047)			
Performing administrative activities	0.05			
	(0.045)			
Working directly with the public	-0.06	-0.02	-0.03	
	(0.050)	(0.036)	(0.031)	
Selling or influencing others	0.13	-0.10*	-0.13***	-0.18**
	(0.080)	(0.054)	(0.043)	(0.076)
Staffing organisational units	0.12			
	(0.148)			
Teaching and training others	-0.14*			
	(0.073)			
Evaluating information to determine compliance with standards	0.11**	0.09**	0.05	0.10
	(0.043)	(0.037)	(0.032)	(0.068)
Judging the qualities of things, services or people	-0.11**			
	(0.055)			
Thinking creatively	-0.11**	0.08**		-0.06
	(0.045)	(0.036)		(0.068)
Updating and using relevant knowledge	0.52***			
	(0.144)			
Controlling machines and processes	0.04			
	(0.044)			
Documenting/recording information	0.16***			-0.06

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
	(0.054)			(0.067)
Handling and moving objects	0.05			
	(0.047)			
Interacting with computers	-0.06	0.04		-0.14
	(0.054)	(0.042)		(0.103)
Operating vehicles, mechanised devices or equipment	0.08**		0.06**	0.11
	(0.037)		(0.027)	(0.072)
Performing general physical activities	0.03	-0.10***		0.14**
	(0.045)	(0.037)		(0.073)
Repairing and maintaining electronic equipment	-0.23**	0.08	0.08	
	(0.089)	(0.075)	(0.070)	
Maintaining and repairing mechanical equipment	0.06		0.05	0.12
	(0.079)		(0.065)	(0.142)
Resolving conflicts and negotiating with others		-0.08*	-0.10**	-0.18***
		(0.049)	(0.040)	(0.062)
Organising, planning and prioritising work		0.04	0.03	
		(0.036)	(0.036)	
Scheduling work and activities		-0.13***	-0.14***	
		(0.043)	(0.035)	
Observations	379	379	379	379

NB: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores and Webb (2020) indices of exposure to digital technologies.

Like the work activities associated with higher automation risk, job ads requiring that workers 'operate vehicles, mechanised devices or equipment' are more likely to be appear in occupations with greater reliance on computer software. The reverse holds for occupations that require 'communication with persons outside the organisation', 'resolving conflicts and negotiating with others', 'selling to or influencing others' and 'scheduling work and activities'.

Occupations with higher demand for industrial robot technologies are also in greater need of employees who can ‘inspect equipment, structures or materials’ and ‘perform general physical activities’ but not to ‘communicate with persons outside the organisation’, ‘guide, direct and motivate subordinates’, ‘resolve conflicts and negotiate with others’ or ‘sell to or influence others’.

Webb (2020) and other authors (e.g. Mitchel, 2019) have noted that AI technologies, which are based on machine learning methods, are not restricted to routine tasks. This contrasts with computer software and robotic technologies that mostly depend on expert ‘if-then’ systems or symbolic AI systems. For this reason, the occupations exposed to AI advancements differ from those exposed to robots and computer software. They mostly include high-skilled jobs involving visual and analytical work, reasoning and communication skills, but also low-skilled, production jobs involving inspection and quality control (Webb, 2020).

The analysis carried out with the Skills-OVATE data confirms that work activities in jobs with greater exposure to computer software and industrial robots are loosely related to those that utilise AI. Employers posting advertisements for occupations exposed to AI technologies are more likely to demand job applicants who can ‘evaluate information to determine compliance with standards’ or ‘think creatively’. AI technologies are, by contrast, less prevalent in occupations that require ‘resolving conflicts and negotiating with others’, ‘selling or influencing others’, ‘scheduling work or activities’ or ‘performing general physical activities’.

4.3. Robustness checks

4.3.1. Testing outlier sensitivity and relative importance

To ensure that the main findings are not influenced by the high skewness in the distribution of s_i in some occupations, we corroborate their robustness by excluding instances of work activities with $s_i < 1\%$ (see Table A.4 in Annex). We subsequently repeat the Lasso regression estimation procedure as described above on this new set of dummy feature variables, which now describe the occurrence $\{s_{ci} = 0,1\}$ of each work activity accounting for at least 1% of the total job ads in an occupation. By doing this, the variables ‘assisting and caring’ and ‘repairing and maintaining electronic equipment’ are found to be no longer statistically significant, while ‘training and teaching others’ becomes more statistically significant than in the original estimation; the tasks ‘coaching and developing others’ and ‘updating and using relevant knowledge’ are no longer selected within the sparse feature set, given their overall low skill-requirement yield. The two additional variables ‘interpreting the meaning of information for others’ (negative effect on automation) and ‘performing general physical activities’ (positive effect) emerge instead as statistically significant predictors of automation risk.

Similarly, the significance of the task variables ‘selling or influencing others’ and ‘scheduling work or activities’ as predictors of AI occupational exposure is lost as a result of their low yield in several occupations. However, it is confirmed that the work activities of ‘evaluating information to determine compliance with standards’, ‘thinking creatively’, ‘performing general physical activities’ (as well as ‘developing objectives and strategies’) are strongly associated with occupational exposure to AI technologies.

It is further confirmed that employers in occupations that rely more on computer software are less likely to demand job applicants who can engage in communication, selling or scheduling work activities but will seek those who can ‘get information’. In contrast to the original estimation, ‘operating vehicles, mechanised devices or equipment’ is not significantly associated with the use of computer software in occupations.

All significant associations reported earlier for the use of industrial robot technologies in occupations are also confirmed. However, some additional positive effects are found in relation to ‘getting information’, ‘making decisions and solving problems’ and ‘organising, planning and prioritising work’, while ‘teaching and training others’ and ‘working directly with the public’ are inversely associated with occupations that rely more heavily on the use of robots.

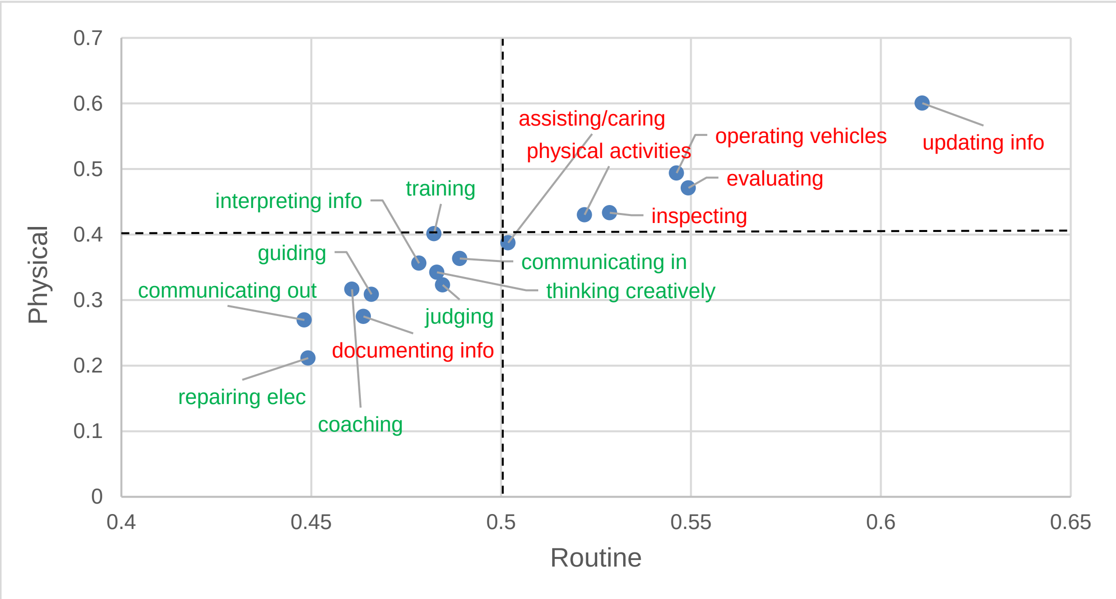
Repeating the estimation using as independent variables in the regression the RCA measures as derived in equation (2), instead of the frequency or occurrence of the respective tasks, further confirms the robustness of the main findings (see Annex Table A.5).

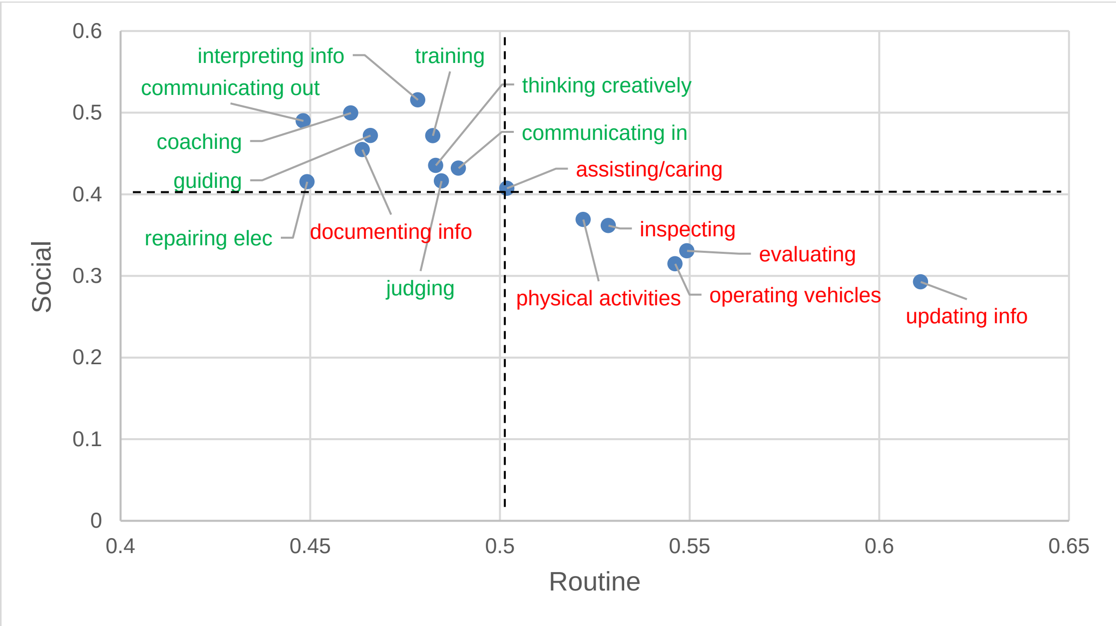
Figure 2 aims to summarise the main findings of the paper by building on the theoretical frameworks of Autor et al. (2003) and Deming (2017). The figure categorises the statistically significant work activities identified in the empirical analysis according to whether they involve routine, manual or social interaction elements. To obtain the respective values of the routine, physical and social task content of the most significant work activities, we first identify the 2-digit occupational groups that have a non-trivial frequency of each work activity ($s_{ci} > 0$). We subsequently calculate the mean value of work routine, physical and social task content associated with the subset of occupations containing each work activity, utilising the European database of task indices (Fernandez-Macias et al., 2016; Bisello et al., 2021) ([15]).

([15]) Although Figure 2 uses the new version (2.0) of the European task indices database, as developed by Bisello et al. (2021), we have confirmed that the same patterns hold with the first version (Fernandez-Macias et al., 2016) (available upon request). The task ‘documenting/recording information’ is a notable outlier.

From the figure it is evident that work activities associated with lower automation risk, such as those entailing higher reasoning tasks (thinking creatively, judging qualities), managerial tasks (coaching or guiding others) and social tasks (communication with clients outside the organisation or with supervisors and peers) are mostly found in non-manual, non-routine and social occupations. Job tasks that face the highest threat of machine learning replacement, such as those characterised by codifiable information retrieval tasks (evaluating information or updating knowledge) or physical tasks (operation or inspection of mechanised equipment and other physical activities at work) are mostly concentrated in manual and routine occupations with little social interaction.

Figure 2. **Automatable and non-automatable work activities by mean degree of routinisation, physical and social task content**





NB: The dots indicate the mean value of routine intensity and physical or social task content in the occupations in which each work activity is most demanded by employers in their OJAs. Work activities associated with higher predicted automation risk are indicated in red and those with lower predicted automation risk are in green; the dashed lines indicate mean values physical-social-routine task content for all occupations.

Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores and European task indices database (Bisello et al. 2021; Fernandez-Macias, 2016).

4.3.2. Comparing big data with representative surveys

While a valuable trait of the Skills-OVATE data set is that it provides highly granular information on work activities requested by employers, the unstructured and non-probabilistic nature of the data extraction process raises some concern about the robustness of the main associations reported above (Cedefop, 2019, 2021). To examine further the robustness of the data, we have sought to compare the findings with information on similar work activities obtained from other large-scale, representative surveys in which the relative importance of different job tasks has been assessed either by occupational experts and/or incumbent workers.

Specifically, relevant work activity indicators obtained from the US O*NET repository as well as the Italian occupational survey ICP-O*NET have been integrated into the data set ([16]). The main empirical specification (equation 3) has then been estimated using the Frey and Osborne automation probability scores as

([16]) The Italian survey on occupations, developed by Istat in 2004 and 2012, is structured according to the information content of the US Occupational Information Network (O*NET) survey. It describes how employed people carry out the 800 professional units that make up the elementary structure of the Italian Classification of Occupations (CP2011) connected with ISCO.

dependent variable and work activities that mirror as closely as possible those from Skills-OVATE ([17]).

The comparison of the estimated regression coefficients (see Annex Table A.6), drawn from three different data sets with information on the importance of selected occupational work activities, provides partial confirmation of the validity of the Skills-OVATE analysis employed in this paper, as well as highlighting some inconsistencies. It is strongly confirmed that occupations reliant on ‘guiding, directing and motivating subordinates’, ‘judging the qualities of thing, services or people’, ‘thinking creatively’ and ‘training and teaching others’ are significantly insulated from the risk of automation, while those involving the ‘operation of vehicles, mechanised devices or equipment’ are significantly prone to it.

However, the representative survey data sources fail to corroborate the statistical significance and (in some cases) the direction of the Skills-OVATE task descriptors ‘assisting and caring for others’, ‘coaching and developing others’, ‘communicating with persons outside organisation’, ‘communicating with supervisors, peers or subordinates’, ‘documenting/recording information’, ‘evaluating information to determine compliance with standards’, ‘inspecting equipment, structures and materials’, ‘repairing and maintain electronic equipment’ and ‘updating and using relevant knowledge’ ([18]).

It is not immediately clear if such differences arise because of differences in the nature of the data (online big data versus survey data) or their differential measurement of work activity intensity ([19]). More research is needed to understand clearly the nature of such discrepancies between different data sets and job task scales.

([17]) Specifically, while for the US O*NET we have information for all work activities available in Skills-OVATE, for the ICP-O*NET we use the subset of task indicators available in Sostero et al. (2020): ‘manual dexterity’, ‘finger dexterity’, ‘performing generally physical activities’, ‘handling and moving objects’, ‘inspecting equipment, structures or materials’, ‘operating vehicles, mechanised equipment or devices’, ‘selling or influencing others’, ‘training and teaching others’, ‘assisting and caring for others’, ‘working directly with the public’, and ‘coordinate the work and tasks of others’.

([18]) The robustness analysis has been done both by comparing the OLS regression coefficients of the variables as derived from the three different data sources and by performing Lasso regression analysis to select only the least correlated feature set. Both approaches converge in their findings.

([19]) The work activity variables in the Italian ICP-ONET measure the share of workers within occupations who carry out a particular task as part of their job. The ONET measures capture if the variables are important, namely if they have a score above 4 in an importance rating scale as assessed by occupational experts. The Skills-OVATE data captures if a specific work activity per occupation is requested by employers in online job adds.

4.4. Identifying a predictive machine learning model

As a last step in the analysis, the set of core work activities detected above is used to identify a predictive model of occupational automation risk. Leveraging appropriate machine and deep learning methods, the aim is to develop a model that may enable accurate inferences about the potential displacement of jobs based on a limited set of occupational task descriptors, when using future rounds of Skills-OVATE data.

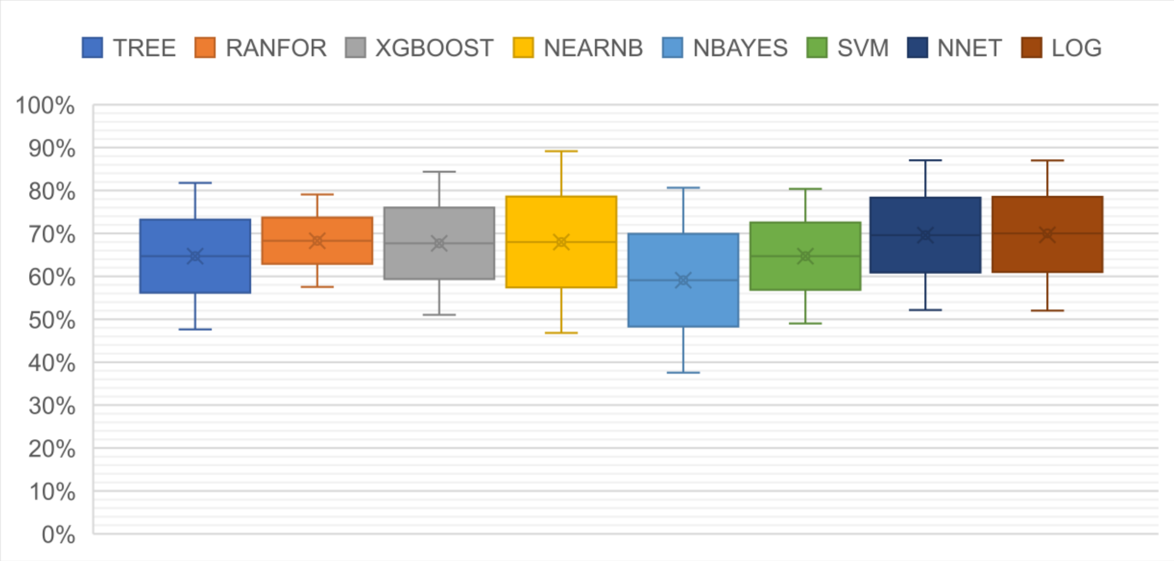
To carry out this exercise, the set of statistically significant work activities identified in Table 1 is used as feature set in a supervised machine learning framework, which aims to predict the incidence of occupations facing a high risk of automation. The target variable is labelled, as is customary in the literature, with a value equal to one for those occupations characterised by a 0.7 or higher probability of automation and zero for all others.

After shuffling and splitting the occupational database into appropriate training and testing data sets implementing an 80:20% rule, and following standardisation of the feature variables, several supervised machine learning classification models have been estimated. The methodology employed deploys a 10-fold cross-validation approach. As shown in Figure 3, these models comprise a logistic regression (LOG), a support vector machine (SVM) classifier, a decision tree (TREE), a random forest (RANFOR) and extreme gradient boosting (XGBOOST) decision tree algorithm, a nearest neighbour classifier (NEARNB), a Bernoulli naïve bayes estimator (NBAYES) and a multilayer perceptron neural network (NNET). The performance of these models is evaluated using standard machine learning performance measures: for instance, Figure 3 showcases the mean accuracy of the test set and associated confidence intervals as constructed based on the standard error of the test accuracy ([20]).

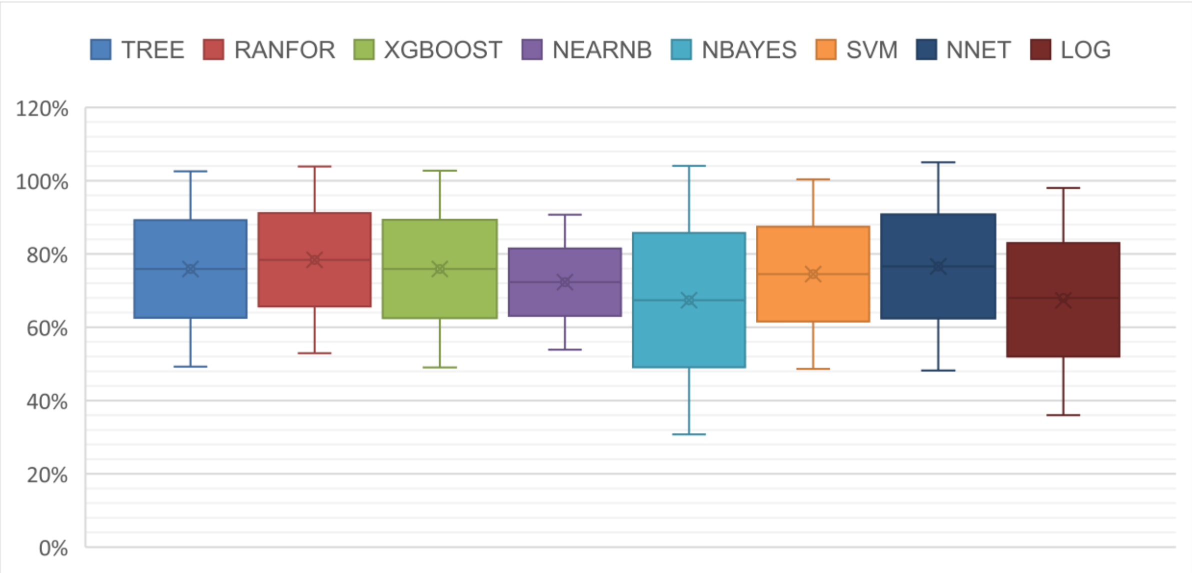
([20]) See Table A.7 in the Annex for the full set of training and test set mean accuracy scores and associated standard errors following the cross-validation estimation (Cerulli, 2020). Other performance measures, such as precision, recall, F1 and area under the receiving operating characteristics (ROC) curve (Geron, 2019) have also been consulted to assess the optimality of the different algorithms.

Figure 3. **Performance of different machine/deep learning predictive models of occupational automation risk**

Skills-OVATE



O*NET



NB: The figure shows the mean cross-validation test accuracy and associated confidence intervals of different machine and deep learning classification algorithms, following estimation of a model of occupational automation risk on a parsimonious set of significant work activity variables (Cerulli, 2020).

Source: Cedefop European database of online job advertisements (Skills-OVATE) and US O*NET matched to Frey and Osborne (2017) automation risk scores.

Figure 3 indicates that an optimal predictive model of occupational automation risk that relies on a sparse set of significant work activity features obtained from Skills-OVATE can reach a high of 70% accuracy score. The best predictive models (with highest optimal test accuracy scores) are obtained using a neural net classifier (with two layers and five neuros) as well as a logistic regression estimator. An ensemble random forest method also reaches a mean accuracy score of 68% with relatively greater precision (tighter confidence interval) in the estimation compared to other models.

For comparison purposes, the above procedure has been repeated and all machine and deep learning estimators have been run on the O*NET data set. Predictions of the occupational automation risk are obtained using the parsimonious set of statistically significant work activities as obtained in column (1) of Table A.6. In this case, random forest or neural net classification algorithms yield an optimal test accuracy score of about 77-78%.

Overall, the best estimated model is found to predict correctly whether an occupation is automatable or not about seven out of 10 times, with relatively high precision (low share of false positives) but lower recall (high share of false negatives) ([21]). While such prediction accuracy is encouraging, it also highlights that the automatability of occupations depends on the complex interrelation of many other factors other than their task content. It is also noted that superior predictive performance is obtained when using as predictor variables the relative importance of distinct work activities as assessed by occupational experts and incumbents, as opposed to them being mentioned in employers' OJAs. With continuing improvements in Skills-OVATE data quality, and the collection of a longer and more stable trend of work activities data, further model improvements may be possible.

([21]) This is concerning given that, from a training policy perspective, it can be relatively less costly (though still inefficient) erroneously to identify an occupation as automatable compared to falsely misclassifying it as non-automatable.

CHAPTER 5.
Conclusions

In-depth understanding of which skills and job tasks may be displaced by AI and other digital technologies is crucial for the formulation of preventive upskilling and reskilling policies; such actions can support individuals' and firms' twin transition to a digital and greener economy. Most of the available automation research to date has used representative survey data that contains information on workers' job tasks and skill needs. Adding to previous studies that have had own limitations, this paper utilises a novel big data set containing information on the work activities and skill needs required by EU employers in their online job advertisements.

Using suitable machine learning 'shrinkage' methods, the analysis has detected a minimum set of least correlated work activity variables associated with occupational digital exposure. Core work activities associated with occupations that have high risk of machine displacement are those that rely on highly codifiable information retrieval and evaluation skills as well as routine, manual skills. Work activities that are relatively immutable to machine learning algorithms include those dependent on high socioemotional and interpersonal skills, managerial skills and problem-solving skills.

Rigorous sensitivity analysis has been carried out to ensure that the findings are not influenced by boundary conditions regarding the intensity of work activities, or by the non-representativeness of OJA data. It is confirmed that occupations reliant on 'guiding, directing and motivating subordinates', 'judging the qualities of thing, services or people', 'thinking creatively' and 'training and teaching others' are significantly insulated from the risk of automation, while those involving the 'operation of vehicles, mechanised devices or equipment' are significantly prone to it. Occupations with higher exposure to industrial robots (and hence a higher automation risk) are also found to be in greater need of workers who can 'inspect equipment, structures or materials' and 'perform general physical activities'.

The research finds that it is a common misnomer to associate AI technologies only with higher automation; it adds to a better understanding of how AI may increase worker productivity across multiple sectors. Work activities with greater exposure to automating technologies are loosely related to jobs that rely more heavily on AI. Employers posting advertisements for occupations with higher exposure to AI technologies are generally more likely to demand applicants who can 'think creatively' or 'evaluate information to determine compliance with standards'.

It is further highlighted that suitable machine or deep learning models of automatability that use a sparse set of significant work activity variables from (future waves of) the Skills-OVATE data, could predict the automation risk of occupations with close to 70% accuracy. However, the analysis also cautions that, in addition to the task content of occupations, analysts need to consider many other complex and interrelated factors affecting occupational automation. The use of job task data obtained from OJAs may also be suboptimal as a potent measure of skill demand, compared to other representative data sources that rely on expert or worker assessments.

The paper's overall conclusions can potentially provide useful input for the effective design of upskilling and reskilling policies that can aid adjustment of individuals, firms and economies to the automation dynamics of new and emerging digital technologies. Future research could utilise the estimated models of this study and aim to identify suitable occupational upskilling and reskilling pathways for individuals at high risk of machine displacement, based on the complementarity of required skill sets between occupational groups with different automatability prospects.

References

[URLs accessed 12.5.2021]

Acemoglu, D. and Restrepo. P. (2018). *Artificial intelligence, automation and work.* NBER Working paper, No 24196.

Acemoglu, D. and Restrepo. P. (2019). Automation and new tasks: how technology displaces and reinstates labor. *Journal of Economic Perspectives*, Vol. 33, No 2, pp. 3-30.

Acemoglu, D. et al (2021). *AI and jobs: evidence from online vacancies.* NBER paper No 28257. https://www.nber.org/papers/w28257

Akerlof, G.A. (1970). The market for 'lemons': quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, Vol. 84, No 3, pp. 488-500.

Arntz, M.; Gregory, T. and Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, Vol. 159, pp. 157-160.

Autor, D.H.; Levy, F. and Murnane, R.J. (2003). The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics*, Vol.118, No 4, pp. 1279-1334.

Autor, D.H; Katz, L. and Kearney, M. (2006). The polarization of the US labor market. *American Economic Review*, Vol. 96, No 2, pp. 189-194.

Autor, D.H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, Vol. 29, No 3, pp. 3-30.

Bessen, J.E. (2015). *Learning by doing: the real connection between innovation, wages and wealth.* Yale: Yale University Press.

Bisello, M. et al (2021). *A comprehensive European database of tasks indices for socio-economic research.* Seville: European Commission, JRC 124124.

Brynjolfsson, E. and McAfee, A. (2014). *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. New York: WW Norton & Company.

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

Cedefop (2018). *Insights into skill shortages and skill mismatch: learning from Cedefop's European skills and jobs survey*. Luxembourg: Publications Office. Cedefop reference series; No 106. https://op.europa.eu/en/publication-detail/-/publication/8d1a7959-2284-11e8-ac73-01aa75ed71a1/language-en

Cedefop (2019a). *Online job vacancies and skills analysis: a Cedefop pan-European approach*. Luxembourg: Publications Office. https://op.europa.eu/en/publication-detail/-/publication/191bf3cb-59ac-11e9-a8ed-01aa75ed71a1/language-en

Cedefop (2019b). *The EU job vacancy market: driving forces and emerging trends*. Luxembourg: Publications Office. Cedefop research paper, No 72. https://op.europa.eu/en/publication-detail/-/publication/3ae2f003-b4d7-11e9-9d01-01aa75ed71a1/language-en

Cedefop (2020). *Assessing the employment impact of technological change and automation: the role of employer's practices*. Luxembourg: Publications Office. Cedefop research paper, No 79. https://op.europa.eu/en/publication-detail/-/publication/16e6c3e3-91cc-11eb-b85c-01aa75ed71a1/language-en

Cedefop (2021). *Understanding technological change and skill needs: big data and artificial intelligence methods. Cedefop practical guide 2*. Luxembourg: Publications Office. https://op.europa.eu/en/publication-detail/-/publication/bf485068-9d94-11eb-b85c-01aa75ed71a1/language-en

Cerulli, G. (2020). *C_ML_STATA: Stata module to implement machine learning classification in Stata*. Statistical Software Components, Boston College Department of Economics. Available at: https://econpapers.repec.org/software/bocbocode/s458830.htm.

Chernoff, A.W. and Warman, C. (2020). *COVID-19 and implications for automation.* NBER WP 27249. http://www.nber.org/papers/w27249

Coombs, C. (2020). Will COVID-19 be the tipping point for the intelligent automation of work? A review of the debate and implications for research. *International Journal of Information Management*, Vol. 55. https://doi.org/10.1016/j.ijinfomgt.2020.102182.

Dawson, N. et al. (2020). Predicting s*kill shortages in labor markets: a machine learning approach*. https://ssrn.com/abstract=3681132 or http://dx.doi.org/10.2139/ssrn.3681132

Deming, D.J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, Vol. 132, No 4, pp. 1593-1640.

Deming, D.J. and Noray, K.L. (2020). Earnings dynamics, changing job skills and STEM careers. *Quarterly Journal of Economics*, advance access: June16 2020.

European Commission (2020). *European Skills Agenda for sustainable competitiveness, social fairness and resilience.* https://ec.europa.eu/social/main.jsp?catId=1223

Fernández-Macías, E.; Hurley, J. and Bisello, M. (2016). *What do Europeans do at work? A task-based analysis: European Jobs Monitor 2016*. Luxembourg: Publications Office.

Ford, M. (2015). *Rise of the robots: technology and the threat of a jobless future.* New York: Basic books.

Fossen, F.M. and Sorgner, A. (2019). *New digital technologies and heterogeneous employment and wage dynamics in the United States: evidence from individual-level data.* IZA Discussion paper, No 12242.

Frey, C. (2019). *The technology trap: capital, labour and power in the age of automation*. Princeton: Princeton University Press.

Frey, C. and Osborne, M. (2013). *The future of employment: how susceptible are jobs to computerisation.* Oxford: University of Oxford.

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

Gambin, L. et al. (2016). *Research to understand the extent, nature and impact of skills mismatches in the economy*. London: Department for Business Innovation and Skills. BIS Research paper; No 265, May 2016.

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and Tensorflow.* Sebastopol: O'Reily media.

Goos, M.; Manning, A. and Salomons, A. (2009). Job polarization in Europe. *American Economic Review*, Vol. 99, No 2, pp. 58-63.

Goos, M. (2018). The impact of technological progress on labour markets: policy challenges. *Oxford Review of Economic Policy*, Vol. 34, No 3, pp. 362-375.

Gürtzgen, N. et al. (2021). *Does online search improve the match quality of new hires?* IZA Discussion paper; No 14031.

Institute for the future of work (IFOW) (2020). *The impact of automation on labour markets: interactions with COVID-19.* Future of work commission evidence review, London.

Katz, L. F. and Murphy, K.M. (1992). *Changes in relative wages, 1963-1987: supply and demand factors*. *Quarterly journal of economics*, Vol. 107, No 1, pp. 35-78.

Katz, L.F. and Autor, D.H. (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter and D. Card (eds.). *Handbook of Labor Economics*, Vol. 3A, pp. 1463-1555.

Keynes, J.M. (1930). *Economic possibilities for our grandchildren*. In *Essays in persuasion*. New York: Harcourt, Brace and Company (1932), pp. 358-373.

McGuinness, S.; Pouliakas, K. and Redmond, P. (2021). *Skills-displacing technological change and within-job reallocation effects on tasks and skills: challenging technological alarmism? Economics of Innovation and New Technology.* DOI: 10.1080/10438599.2021.1919517.

McKinsey Global Institute (2017). *Jobs lost, jobs gained: what the future of work will mean for jobs, skills, and wages*. [s.l]: McKinsey & Company.

Mitchel, M. (2019). *Artificial intelligence: a guide for thinking humans.* New York: Farrar, Straus and Giroux.

Mokyr, J.; Vickers, C and Ziebarth, N. (2015). The history of technological anxiety and the future of economic growth: is this time different? *Journal of Economic Perspectives*, Vol. 29, pp. 31-50.

Nedelkoska, L. and Quintini, G. (2018). Automation, skills use and training. *OECD Social, Employment and Migration Working Papers*, No 202. Paris: OECD Publishing.

Pouliakas, K. (2018). Automation risk in the EU labour market: a skill-needs approach. In Hogarth, T. (ed). *Economy, employment and skills: European, regional and global perspectives in an age of uncertainty.* Rome: Fondazione Giacomo Brodolini.

Pouliakas, K. and Branka, J. (2020). *EU jobs at highest risk of COVID-19 social distancing: is the pandemic exacerbating the labour market divide?* Luxembourg: Publications Office. Cedefop working paper; No 1. http://data.europa.eu/doi/10.2801/968483

Sostero, M. et al. (2020). *Teleworkability and the COVID-19 crisis: a new digital divide?* Seville: European Commission, JRC 121193.

Webb, M. (2018). *The impact of artificial intelligence on the labour market*. Stanford: Stanford university.

World Economic Forum (2016). *The future of jobs: employment, skills and workforce strategy for the fourth industrial revolution.* Geneva: World Economic Forum.

World Economic Forum (2018). *Towards a reskilling revolution: industry-led action for the future of work.* Geneva: World Economic Forum.

Annex

Table A 1. **List of 'work activities' clusters contained in Skills-OVATE**

	Work activity (WA)	Example of (most prominent) detailed terms included in broad WA
1.	Inspecting equipment, structures or materials	Examine merchandise, inspect machinery, visit places of work, etc.
2.	Evaluating information to determine compliance with standards	Apply quality standards, follow safety standards in industrial contexts, follow clinical guidelines, etc.
3.	Operating vehicles, mechanised devices or equipment	Drive vehicles, park vehicles, drive in urban areas, etc.
4.	Maintaining and repairing mechanical equipment	Maintain machinery, repair plumbing systems, perform maintenance on firm alarms, etc.
5.	Controlling machines and processes	Use online communication tools, tend CNC drilling machines, tend CNC laser cutting machines, etc.
6.	Scheduling work and activities	Administer appointments, fix meetings, schedule regular machine maintenance, etc.
7.	Staffing organisational units	Recruit members, hire human resources, organise auditions, etc.
8.	Updating and using relevant knowledge	Use learning strategies, stay up to date with social media, maintain updated professional knowledge, keep up to date on product knowledge, etc.
9.	Performing general physical activities	Maintain work area cleanliness, provide food and beverages, exercise sports etc.
10.	Handling and moving objects	Use food preparation techniques, use cooking techniques, manage clinical environments, perform warehousing operations, install machinery, etc.
11.	Estimating quantifiable characteristics of products	Use measurement instruments, conduct land surveys, keep time accurately, etc.
12.	Selling or influencing others	Sell products, sell services, apply social media marketing, etc.
13.	Monitoring and controlling resources	Apply procurement, order supplies, manage warehouse inventory, etc.
14.	Interpreting the meaning of information for others	Translate spoken language, interact with healthcare users, apply technical communication skills, interpret law, etc.
15.	Performing administrative activities	Execute administration, handle mail, process applications, etc.

	Work activity (WA)	Example of (most prominent) detailed terms included in broad WA
16.	Getting information	Use technical documentation, carry out internet research, gather data, etc.
17.	Training and teaching others	Adapt teaching to target, apply teaching strategies, provide training, etc.
18.	Resolving conflicts and negotiating with others	Conclude business agreements, manage contracts, enforce customers' debt repayment, etc.
19.	Developing objectives and strategies	Define quality standards, manage database, design campaign actions, etc.
20.	Processing information	Process data, pick orders for dispatching, process qualitative information, etc.
21.	Repairing and maintaining electronic equipment	Maintain ICT server, repair ICT devices, maintain electromechanical equipment, etc.
22.	Working directly with the public	Run errands on behalf of customers, entertain people, dance, etc.
23.	Monitor processes, materials or surroundings	Monitor customer service, follow manufacturing work schedule, receive goods, etc.
24.	Identifying objects, actions or events	Identify improvement actions, identify opportunities, prospect new customers, etc.
25.	Organising, planning and prioritising work	Adjust priorities, prioritise tasks, plan teamwork, etc.
26.	Assisting and caring for others	Assist customers, provide customer follow-up, babysitting
27.	Interacting with computers	Use a computer, use Microsoft Office, use office systems, use spreadsheets, etc.
28.	Coaching and developing others	Motivate others, mentor individuals, encourage teams for continuous improvement, etc.
29.	Documenting/recording information	Report analysis results, present reports, adapt teaching to student capabilities, provide technical documentation, etc.
30.	Provide consultation and advice to others	Direct customers to merchandise, give advice to others, use consulting techniques, etc.
31.	Developing and building teams	Team building, plan team building, work closely with new teams, etc.
32.	Making decisions and solving problems	Make decisions, implement sales strategies, apply transportation management concepts, implement marketing strategies, etc.

	Work activity (WA)	Example of (most prominent) detailed terms included in broad WA
33.	Judging the qualities of things, services or people	Set production KPI, execute feasibility study, manage project metrics, etc.
34.	Analysing data or information	Analyse software specific specifications, perform data analysis, analyse financial risk, etc.
35.	Communicating with supervisors, peers or subordinates	Coordinate communications activities within a team, supervise merchandise displays, communicate with nursing staff, manage project information, etc.
36.	Establishing and maintaining interpersonal relationships	Maintain working relationships, liaise with managers, build business relations, etc.
37.	Thinking creatively	Think creatively, use software design, design prototypes etc.
38.	Communicating with persons outside the organisation	Provide information, use communication techniques, communicate with customers
39.	Guiding, directing and motivating subordinates	Delegate activities, manage staff, manage guest support services etc.

Source: Cedefop European database of online job advertisements (Skills-OVATE).

Table A 2. **Descriptive statistics of Skills-OVATE variables used in analysis**

	Mean	Std. Dev	Min	Max
Dependent variables				
Automation score (Frey and Osborne)	0.483	0.374	0.003	0.990
Very high automation risk (Frey and Osborne)	0.391	0.489	0.000	1.000
AI score (Webb)	0.434	0.286	0.009	1.911
Software score (Webb)	0.455	0.245	0.064	1.483
Robot score (Webb)	0.511	0.553	0.008	2.842
Feature variables				
Frequency of work activities				
Estimating quantifiable characteristics of products	0.004	0.019	0.000	0.287
Getting information	0.003	0.010	0.000	0.103
Identifying objects, actions or events	0.002	0.006	0.000	0.066
Inspecting equipment, structures or materials	0.007	0.029	0.000	0.376
Monitoring processes, materials or surroundings	0.018	0.044	0.000	0.680
Assisting and caring for others	0.110	0.168	0.000	1.000
Coaching and developing others	0.001	0.003	0.000	0.036
Communicating with persons outside the organisation	0.087	0.113	0.000	1.000
Communicating with supervisors, peers or subordinates	0.011	0.030	0.000	0.214
Developing and building teams	0.040	0.105	0.000	0.460
Establishing and maintaining interpersonal relationships	0.026	0.064	0.000	0.847
Guiding, directing and motivating subordinates	0.023	0.057	0.000	0.810
Interpreting the meaning of information for others	0.001	0.004	0.000	0.056
Monitoring and controlling resources	0.013	0.037	0.000	0.339
Performing administrative activities	0.012	0.037	0.000	0.411
Working directly with the public	0.003	0.014	0.000	0.190
Provide consultation and advice to others	0.019	0.040	0.000	0.309
Resolving conflicts and negotiating with others	0.002	0.017	0.000	0.303
Selling or influencing others	0.002	0.017	0.000	0.314
Staffing organisational units	0.000	0.001	0.000	0.011
Training and teaching others	0.004	0.035	0.000	0.474
Analysing data or information	0.043	0.098	0.000	0.818

	Mean	Std. Dev	Min	Max
Developing objectives and strategies	0.024	0.063	0.000	0.739
Evaluating information to determine compliance with standards	0.023	0.044	0.000	0.358
Judging the qualities of things, services or people	0.009	0.046	0.000	0.808
Making decisions and solving problems	0.004	0.015	0.000	0.198
Organising, planning and prioritising work	0.125	0.166	0.000	0.789
Processing information	0.008	0.034	0.000	0.340
Scheduling work and activities	0.001	0.004	0.000	0.061
Thinking creatively	0.096	0.173	0.000	1.000
Updating and using relevant knowledge	0.000	0.001	0.000	0.020
Controlling machines and processes	0.047	0.094	0.000	1.000
Documenting/recording information	0.023	0.072	0.000	0.855
Handling and moving objects	0.038	0.111	0.000	1.294
Interacting with computers	0.359	0.359	0.000	1.000
Operating vehicles, mechanised devices or equipment	0.018	0.055	0.000	0.593
Performing general physical activities	0.020	0.066	0.000	1.000
Repairing and maintaining electronic equipment	0.001	0.008	0.000	0.144
Maintaining and repairing mechanical equipment	0.002	0.008	0.000	0.084
Occurrence of work activities				
Estimating quantifiable characteristics of products	0.290	0.454	0.000	1.000
Getting information	0.257	0.438	0.000	1.000
Identifying objects, actions or events	0.196	0.397	0.000	1.000
Inspecting equipment, structures or materials	0.275	0.447	0.000	1.000
Monitor processes, materials or surroundings	0.579	0.494	0.000	1.000
Assisting and caring for others	0.611	0.488	0.000	1.000
Coaching and developing others	0.158	0.366	0.000	1.000
Communicating with persons outside the organisation	0.653	0.476	0.000	1.000
Communicating with supervisors, peers or subordinates	0.272	0.446	0.000	1.000
Developing and building teams	0.210	0.408	0.000	1.000
Establishing and maintaining interpersonal relationships	0.480	0.500	0.000	1.000
Guiding, directing and motivating subordinates	0.423	0.495	0.000	1.000
Interpreting the meaning of information for others	0.158	0.366	0.000	1.000
Monitoring and controlling resources	0.292	0.455	0.000	1.000

	Mean	Std. Dev	Min	Max
Performing administrative activities	0.374	0.484	0.000	1.000
Working directly with the public	0.248	0.432	0.000	1.000
Provide consultation and advice to others	0.450	0.498	0.000	1.000
Resolving conflicts and negotiating with others	0.173	0.379	0.000	1.000
Selling to or influencing others	0.136	0.343	0.000	1.000
Staffing organisational units	0.077	0.266	0.000	1.000
Training and teaching others	0.151	0.358	0.000	1.000
Analysing data or information	0.517	0.500	0.000	1.000
Developing objectives and strategies	0.490	0.501	0.000	1.000
Evaluating information to determine compliance with standards	0.426	0.495	0.000	1.000
Judging the qualities of things, services or people	0.337	0.473	0.000	1.000
Making decisions and solving problems	0.240	0.428	0.000	1.000
Organising, planning and prioritising work	0.696	0.461	0.000	1.000
Processing information	0.210	0.408	0.000	1.000
Scheduling work and activities	0.126	0.333	0.000	1.000
Thinking creatively	0.611	0.488	0.000	1.000
Updating and using relevant knowledge	0.064	0.246	0.000	1.000
Controlling machines and processes	0.656	0.476	0.000	1.000
Documenting/recording information	0.322	0.468	0.000	1.000
Handling and moving objects	0.584	0.493	0.000	1.000
Interacting with computers	0.851	0.356	0.000	1.000
Operating vehicles, mechanised devices or equipment	0.416	0.493	0.000	1.000
Performing general physical activities	0.525	0.500	0.000	1.000
Repairing and maintaining electronic equipment	0.101	0.302	0.000	1.000
Maintaining and repairing mechanical equipment	0.124	0.330	0.000	1.000

Source: Cedefop European database of online job advertisements (Skills-OVATE).

Table A 3. **Automation/digital technologies exposure and frequency of work activities, OLS regressions**

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
Estimating quantifiable characteristics of products	0.54	0.21	-0.37	0.28
	(0.728)	(0.555)	(0.318)	(1.045)
Getting information	1.60	-1.71	-0.07	6.25**
	(1.517)	(1.412)	(1.367)	(2.727)
Identifying objects, actions or events	1.04	-3.20	0.68	-4.13
	(4.418)	(3.612)	(2.805)	(4.486)
Inspecting equipment, structures or materials	1.15*	0.36	1.01**	1.66*
	(0.618)	(0.560)	(0.434)	(0.954)
Monitoring processes, materials or surroundings	-0.41	0.78	0.07	-2.15**
	(0.521)	(0.472)	(0.379)	(0.899)
Assisting and caring for others	0.29**	-0.05	-0.00	0.21
	(0.142)	(0.108)	(0.089)	(0.172)
Coaching and developing others	-7.19	0.70	-4.45	-8.55
	(5.175)	(5.458)	(3.215)	(6.068)
Communicating with persons outside the organisation	-0.49**	-0.52***	-0.54***	-1.33***
	(0.239)	(0.167)	(0.140)	(0.293)
Communicating with supervisors, peers or subordinates	-0.49	-1.17*	-0.14	1.81
	(0.698)	(0.701)	(0.366)	(1.215)
Developing and building teams	-0.21	-0.02	0.10	0.07
	(0.188)	(0.157)	(0.114)	(0.220)
Establishing and maintaining interpersonal relationships	0.04	0.95	0.19	-0.24
	(0.595)	(0.808)	(0.497)	(0.617)
Guiding, directing and motivating subordinates	-1.69***	0.09	-0.04	-0.14
	(0.617)	(0.631)	(0.432)	(0.768)
Interpreting the meaning of information for others	-7.07***	2.88	0.62	-3.82*
	(2.456)	(2.777)	(1.898)	(2.075)
Monitoring and controlling resources	0.67	-0.40	0.19	0.68
	(0.421)	(0.376)	(0.377)	(1.051)
Performing administrative activities	0.02	0.10	0.26	-0.04
	(0.805)	(0.508)	(0.320)	(0.372)
Working directly with the public	-0.66	-2.51**	-2.99***	-3.83**
	(1.322)	(1.073)	(0.799)	(1.608)

Work activities	(1) Automation risk	(2) AI	(3) Software	(4) Robots
Provide consultation and advice to others	-0.37 (0.503)	-0.39 (0.517)	-0.68* (0.352)	-0.43 (0.605)
Resolving conflicts and negotiating with others	-0.24 (0.593)	-0.45 (0.709)	-0.70 (0.430)	-1.20*** (0.369)
Selling to or influencing others	-0.13 (0.462)	-0.96*** (0.323)	-1.22*** (0.315)	-1.25** (0.562)
Staffing organisational units	-29.12* (15.206)	-3.58 (8.369)	-6.89 (14.307)	-19.87 (16.723)
Training and teaching others	-1.04*** (0.378)	0.57*** (0.136)	0.10 (0.192)	-0.50 (0.580)
Analysing data or information	0.34 (0.266)	0.18 (0.265)	0.15 (0.204)	0.36 (0.326)
Developing objectives and strategies	0.03 (0.417)	-0.01 (0.369)	-0.23 (0.283)	-0.59 (0.385)
Evaluating information to determine compliance with standards	0.61 (0.451)	0.60 (0.395)	0.38 (0.292)	0.40 (0.485)
Judging the qualities of things, services or people	-0.05 (0.730)	-0.55 (0.627)	-0.28 (0.647)	0.07 (0.954)
Making decisions and solving problems	1.02 (1.248)	2.12 (1.465)	1.02 (1.087)	3.20** (1.399)
Organising, planning and prioritising work	-0.23 (0.174)	0.15 (0.138)	0.06 (0.104)	-0.24 (0.194)
Processing information	-0.89 (0.894)	-0.33 (1.093)	-0.05 (0.695)	0.67 (0.823)
Scheduling work and activities	5.10 (3.557)	-5.10* (2.598)	-4.05* (2.143)	-1.99 (4.169)
Thinking creatively	-0.38** (0.165)	0.26** (0.123)	0.19* (0.100)	0.05 (0.153)
Updating and using relevant knowledge	19.32*** (2.825)	-13.20*** (2.733)	-8.19*** (2.154)	-16.57*** (4.418)
Controlling machines and processes	0.23 (0.205)	0.10 (0.161)	0.30* (0.165)	0.29 (0.412)
Documenting/recording information	1.16** (0.523)	-0.04 (0.564)	0.26 (0.519)	0.54 (0.641)
Handling and moving objects	0.10 (0.210)	0.05 (0.115)	-0.12* (0.071)	-0.31* (0.163)
Interacting with computers	-0.11 (0.093)	0.07 (0.068)	-0.09 (0.057)	-0.47*** (0.104)

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
Operating vehicles, mechanised devices or equipment	0.18	-0.04	0.25	1.43*
	(0.305)	(0.281)	(0.365)	(0.854)
Performing general physical activities	0.42	-0.21	0.04	2.48***
	(0.351)	(0.325)	(0.250)	(0.906)
Repairing and maintaining electronic equipment	0.04	-0.54	1.15	3.58***
	(0.875)	(0.866)	(0.712)	(1.289)
Maintaining and repairing mechanical equipment	-0.14	-0.57	-1.27	5.00
	(1.855)	(1.624)	(1.755)	(4.614)
Constant	0.57***	0.39***	0.48***	0.70***
	(0.033)	(0.025)	(0.025)	(0.056)
Observations	379	379	379	379
R-squared	0.26	0.16	0.14	0.30

NB: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores and Webb (2020) indices of exposure to digital technologies

Table A 4. **Automation/digital technologies exposure and occurrence of work activities (>1%), Lasso regressions**

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
Estimating quantifiable characteristics of products	0.03 (0.066)			
Getting information	0.11 (0.075)		0.14** (0.057)	0.28*** (0.097)
Identifying objects, actions or events	-0.06 (0.083)			
Inspecting equipment, structures or materials	0.16** (0.063)			0.22** (0.093)
Assisting and caring for others	0.04 (0.037)			
Communicating with persons outside the organisation	-0.17*** (0.043)		-0.07** (0.033)	-0.29*** (0.074)
Communicating with supervisors, peers or subordinates	-0.14** (0.059)			-0.08 (0.068)
Developing and building teams	-0.04 (0.053)			
Guiding, directing and motivating subordinates	-0.24*** (0.060)			-0.19*** (0.064)
Interpreting the meaning of information for others	-0.20** (0.087)			
Monitoring and controlling resources	0.03 (0.048)			
Performing administrative activities	0.03 (0.054)			-0.07 (0.069)
Selling to or influencing others	0.20 (0.135)	-0.14 (0.088)	-0.17*** (0.056)	-0.18* (0.100)
Training and teaching others	-0.49*** (0.062)			-0.31*** (0.092)
Developing objectives and strategies	-0.04 (0.047)	0.09* (0.046)		-0.10 (0.063)
Evaluating information to determine compliance with standards	0.08* (0.043)	0.07** (0.037)	0.04 (0.029)	0.07 (0.063)
Making decisions and solving problems	-0.07 (0.083)	0.16 (0.096)		0.21** (0.106)

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
Organising, planning and prioritising work	0.05 (0.041)	0.05 (0.035)		0.16** (0.070)
Scheduling work and activities	0.13 (0.159)	-0.12 (0.106)	-0.17** (0.079)	-0.24* (0.128)
Thinking creatively	-0.10** (0.047)	0.07* (0.035)		-0.09 (0.066)
Controlling machines and processes	0.00 (0.037)			0.06 (0.060)
Documenting/recording information	0.22*** (0.050)			
Operating vehicles, mechanised devices or equipment	0.07* (0.039)		0.05 (0.032)	0.10 (0.078)
Performing general physical activities	0.07* (0.038)	-0.10*** (0.033)		0.22*** (0.077)
Repairing and maintaining electronic equipment	-0.07 (0.177)			0.10 (0.252)
Establishing and maintaining interpersonal relationships		0.07 (0.050)		-0.08 (0.056)
Working directly with the public			-0.08 (0.059)	-0.16* (0.087)
Processing information			0.02 (0.043)	0.06 (0.104)
Provide consultation and advice to others				-0.10 (0.074)
Resolving conflicts and negotiating with others				-0.42** (0.182)
Interacting with computers				-0.12 (0.088)
Observations	379	379	379	379

NB: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores and Webb (2020) indices of exposure to digital technologies

Table A 5. **Automation/digital technologies exposure and revealed comparative advantage (RCA) of work activities, OLS regressions**

Work activities	(1) *Automation risk*	(2) *AI*	(3) *Software*	(4) *Robots*
Estimating quantifiable characteristics of products	0.00 (0.003)	0.00 (0.003)	-0.00 (0.001)	0.00 (0.005)
Getting information	0.01 (0.009)	-0.01 (0.009)	0.00 (0.008)	0.04** (0.016)
Identifying objects, actions or events	0.00 (0.018)	-0.01 (0.013)	0.00 (0.011)	-0.01 (0.018)
Inspecting equipment, structures or materials	0.01* (0.005)	0.00 (0.004)	0.01 (0.004)	0.01 (0.008)
Monitoring processes, materials or surroundings	-0.01 (0.009)	0.01 (0.009)	0.00 (0.007)	-0.04** (0.016)
Assisting and caring for others	0.05* (0.026)	-0.02 (0.019)	0.01 (0.017)	0.02 (0.032)
Coaching and developing others	-0.01 (0.007)	0.00 (0.007)	-0.01 (0.004)	-0.01 (0.009)
Communicating with persons outside the organisation	-0.07** (0.032)	-0.05** (0.021)	-0.07*** (0.020)	-0.19*** (0.041)
Communicating with supervisors, peers or subordinates	-0.01 (0.016)	-0.03* (0.015)	-0.01 (0.010)	0.05 (0.029)
Developing and building teams	-0.02 (0.016)	0.00 (0.012)	0.01 (0.010)	0.00 (0.019)
Establishing and maintaining interpersonal relationships	0.00 (0.025)	0.04 (0.031)	0.01 (0.019)	-0.02 (0.028)
Guiding, directing and motivating subordinates	-0.07*** (0.024)	-0.00 (0.023)	-0.00 (0.017)	0.01 (0.029)
Interpreting the meaning of information for others	-0.01*** (0.002)	0.00 (0.003)	0.00 (0.002)	-0.00** (0.002)
Monitoring and controlling resources	0.01 (0.009)	-0.01 (0.007)	0.00 (0.007)	0.01 (0.021)
Performing administrative activities	-0.00 (0.016)	-0.00 (0.010)	0.00 (0.007)	-0.00 (0.008)
Working directly with the public	-0.00 (0.002)	-0.00** (0.002)	-0.00*** (0.001)	-0.01* (0.003)
Provide consultation and advice to others	-0.02 (0.019)	-0.00 (0.016)	-0.02 (0.013)	-0.00 (0.022)

Work activities	(1) Automation risk	(2) AI	(3) Software	(4) Robots
Resolving conflicts and negotiating with others	-0.00	-0.00	-0.00*	-0.00***
	(0.002)	(0.001)	(0.001)	(0.001)
Selling to or influencing others	-0.00	-0.00***	-0.00***	-0.00*
	(0.002)	(0.001)	(0.001)	(0.002)
Staffing organisational units	-0.00*	0.00	-0.00	-0.00
	(0.002)	(0.001)	(0.002)	(0.002)
Training and teaching others	-0.00**	0.00***	-0.00	-0.00
	(0.001)	(0.000)	(0.001)	(0.002)
Analysing data or information	0.04	-0.00	0.01	0.04
	(0.027)	(0.025)	(0.022)	(0.034)
Developing objectives and strategies	0.00	-0.02	-0.01	-0.03
	(0.022)	(0.019)	(0.015)	(0.020)
Evaluating information to determine compliance with standards	0.02	0.03**	0.01	0.01
	(0.014)	(0.011)	(0.009)	(0.015)
Judging the qualities of things, services or people	0.00	-0.00	-0.00	-0.00
	(0.013)	(0.011)	(0.012)	(0.018)
Making decisions and solving problems	0.01	0.00	0.01	0.02**
	(0.009)	(0.006)	(0.006)	(0.008)
Organising, planning and prioritising work	-0.06*	0.06**	0.01	-0.07*
	(0.037)	(0.028)	(0.021)	(0.043)
Processing information	-0.02	0.02	0.01	0.04**
	(0.022)	(0.023)	(0.015)	(0.020)
Scheduling work and activities	0.01	-0.00	-0.00**	-0.00
	(0.004)	(0.003)	(0.002)	(0.004)
Thinking creatively	-0.08**	0.06***	0.04**	0.02
	(0.038)	(0.021)	(0.021)	(0.030)
Updating and using relevant knowledge	0.00***	-0.00***	-0.00***	-0.00***
	(0.000)	(0.000)	(0.000)	(0.000)
Controlling machines and processes	0.01	0.00	0.02	0.01
	(0.013)	(0.010)	(0.012)	(0.026)
Documenting/recording information	0.06**	-0.01	0.01	0.03
	(0.027)	(0.028)	(0.026)	(0.034)
Handling and moving objects	0.01	0.00	-0.01*	-0.02*
	(0.012)	(0.006)	(0.004)	(0.009)
Interacting with computers	-0.04	-0.01	-0.07*	-0.29***
	(0.065)	(0.047)	(0.040)	(0.070)
Operating vehicles, mechanised devices or equipment	0.00	-0.00	0.00	0.02*
	(0.005)	(0.004)	(0.006)	(0.013)

Work activities	(1) Automation risk	(2) AI	(3) Software	(4) Robots
Performing general physical activities	0.01	-0.00	0.00	0.05***
	(0.006)	(0.005)	(0.004)	(0.015)
Repairing and maintaining electronic equipment	0.00	-0.00	0.01**	0.03***
	(0.005)	(0.004)	(0.003)	(0.006)
Maintaining and repairing mechanical equipment	-0.00	-0.00	-0.00	0.01
	(0.005)	(0.005)	(0.005)	(0.014)
Constant	0.56***	0.38***	0.48***	0.68***
	(0.032)	(0.025)	(0.025)	(0.056)
Observations	377	377	377	377
R-squared	0.26	0.16	0.13	0.29

NB: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Source: Cedefop European database of online job advertisements (Skills-OVATE) matched to Frey and Osborne (2017) automation risk scores and Webb (2020) indices of exposure to digital technologies.

Table A 6. **Occupational automation risk and work activities, comparing representative surveys and online job ads, OLS regressions**

Work activities	(1) *ONET*	(2) *Skills-OVATE*	(3) *ICP-ONET*
Analysing data or information	-0.19*** (0.071)	0.04 (0.045)	
Assisting and caring for others	-0.24*** (0.061)	0.08** (0.039)	-0.21** (0.108)
Coaching and developing others	0.30*** (0.101)	-0.18*** (0.066)	
Communicating with persons outside organisation	-0.01 (0.066)	-0.19*** (0.044)	
Communicating with supervisors, peers, or subordinates	-0.03 (0.041)	-0.15*** (0.053)	
Controlling machines and processes	0.11** (0.042)	0.04 (0.044)	
Coordinating the work and activities of others	0.01 (0.076)		-0.72*** (0.178)
Developing objectives and strategies	-0.10 (0.079)	0.00 (0.042)	
Developing and building teams	-0.04 (0.099)	-0.06 (0.048)	
Documenting/recording information	-0.03 (0.045)	0.16*** (0.054)	
Drafting, laying out, and specifying technical devices, parts, and equipment	-0.05 (0.108)		
Establishing and maintaining interpersonal relationships	-0.09 (0.053)	-0.06 (0.047)	
Estimating the quantifiable characteristics of products, events, or information	0.17** (0.080)	-0.07 (0.046)	
Evaluating information to determine compliance with standards	-0.01 (0.049)	0.11** (0.043)	
Getting information	-0.05 (0.040)	0.03 (0.051)	
Guiding, directing, and motivating subordinates	-0.18** (0.086)	-0.14*** (0.055)	
Handling and moving objects	0.15*** (0.045)	0.05 (0.047)	1.03*** (0.327)
Identifying objects, actions, and events	0.00 (0.042)	0.09 (0.073)	
Inspecting equipment, structures, or material	-0.05 (0.045)	0.11** (0.048)	0.18 (0.184)
Interacting with computers	0.06 (0.053)	-0.06 (0.054)	

Work activities	(1) *ONET*	(2) *Skills-OVATE*	(3) *ICP-ONET*
Interpreting the meaning of information for others	-0.06 (0.059)	-0.03 (0.063)	
Judging the qualities of things, services, or people	-0.25*** (0.095)	-0.11* (0.055)	
Making decisions and solving problems	-0.15*** (0.047)	0.08 (0.060)	
Monitoring processes, materials, or surroundings	-0.00 (0.047)	0.04 (0.041)	
Monitoring and controlling resources	0.52*** (0.161)	0.05 (0.047)	
Operating vehicles, Mechanised devices, or equipment	0.10* (0.055)	0.08** (0.037)	0.01 (0.119)
Organising, planning, and prioritising Work	0.05 (0.054)	0.02 (0.045)	
Performing administrative activities	-0.04 (0.097)	0.05 (0.045)	
Performing general physical activities	-0.04 (0.051)	0.03 (0.045)	-0.35 (0.231)
Performing for or working directly with the public	0.05 (0.054)	-0.06 (0.050)	0.11 (0.137)
Processing information	0.17*** (0.059)	-0.00 (0.060)	
Provide consultation and advice to others	-0.03 (0.101)	-0.04 (0.047)	
Repairing and maintaining electronic equipment	-0.11 (0.187)	-0.23** (0.089)	
Repairing and maintaining mechanical equipment	-0.01 (0.090)	0.06 (0.079)	
Resolving conflicts and negotiating with others	-0.15* (0.088)	-0.01 (0.067)	
Scheduling work and activities	-0.09 (0.095)	-0.02 (0.083)	-0.20 (0.132)
Selling or influencing others	0.17** (0.073)	0.13 (0.080)	
Staffing organisational units	0.66*** (0.228)	0.12 (0.148)	
Thinking creatively	-0.31*** (0.051)	-0.11** (0.045)	
Training and teaching others	-0.18** (0.072)	-0.14* (0.073)	-0.57*** (0.146)
Updating and using relevant knowledge	-0.05 (0.055)	0.52*** (0.144)	

Work activities	(1) *ONET*	(2) *Skills-OVATE*	(3) *ICP-ONET*
Manual dexterity			0.16
			(0.439)
Finger dexterity			-1.06***
			(0.398)
Constant	0.71***	0.56***	0.96***
	(0.031)	(0.053)	(0.104)
Observations	352	379	378
R-squared	0.56	0.33	0.43

NB: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Col (1) captures whether a given work activity has been assessed as of high importance for an occupation by occupational experts or incumbents (score above 4 on 1-5 scale); Col (2) measures the occurrence of a given work activity in online job advertisements of employers; Col (3) refers to the importance of a given task (0-100 scale) at 3-digit ISCO level as assessed by workers.

Source: Cedefop European database of online job advertisements (Skills-OVATE), US O*NET and Italian ICP-O*NET matched to Frey and Osborne (2017) automation risk scores.

Table A 7 **Performance of different machine/deep learning predictive models of occupational automation risk**

	Skills-OVATE			O*NET		
	Mean train accuracy	Mean test accuracy	S.E. test accuracy	Mean train accuracy	Mean test accuracy	S.E. test accuracy
TREE	0.667	0.647	0.087	0.823	0.759	0.136
RANFOR	0.714	0.683	0.055	0.826	0.784	0.130
BOOST	0.695	0.677	0.085	0.760	0.759	0.137
NEARNB	0.685	0.680	0.108	0.742	0.723	0.094
LOG	0.729	0.697	0.090	0.775	0.676	0.16
NBAYES	0.610	0.591	0.110	0.726	0.674	0.187
SVM	0.869	0.647	0.080	0.853	0.745	0.132
NNET	0.789	0.696	0.089	0.826	0.766	0.145

NB: The figure shows the mean cross-validation train and test accuracy and associated confidence intervals of different machine and deep learning classification algorithms, following estimation of a model of occupational automation risk on a parsimonious set of significant work activity variables (Cerulli, 2020).

Source: Cedefop European database of online job advertisements (Skills-OVATE) and US O*NET matched to Frey and Osborne (2017) automation risk scores.

ARTIFICIAL INTELLIGENCE AND JOB AUTOMATION: AN EU ANALYSIS USING ONLINE JOB VACANCY DATA

Not long before the coronavirus outbreak, fears about artificial intelligence (AI) algorithms and machines resulting in a jobless society were widespread. Concerns have resurfaced in light of the COVID-19 crisis potentially accentuating automation. This study utilises a novel big data set based on online job advertisements – Cedefop's Skills OVATE – with information on the skills and work activities required by EU employers. The data provide insight into the task profiles of detailed occupations faced with higher automation risk or those relying on alternative digital technologies (robots, computer software, AI). The paper explores suitable machine and deep learning models to test how well a parsimonious set of task indicators can predict occupational automatability. Work activities associated with greater occupational automation risk and robot exposure (e.g. inspecting equipment, performing physical activities), typically concentrated in routine or manual jobs, differ from those prominent in occupations with higher AI exposure (e.g. thinking creatively, evaluating standards).

6206 EN - TI-BA-21-004-EN-N - doi:10.2801/305373

CEDEFOP
European Centre for the Development of Vocational Training

Europe 123, Thessaloniki (Pylea), GREECE
Postal: Cedefop service post, 570 01 Thermi, GREECE
Tel. +30 2310490111, Fax +30 2310490020
Email: info@cedefop.europa.eu

www.cedefop.europa.eu

Publications Office of the European Union

ISBN 978-92-896-3275-1

9 789289 632751