OECD Social, Employment and Migration Working Papers
No. 287

The supply, demand and characteristics of the AI workforce across OECD countries

Andrew Green,

Lucas Lamby

https://dx.doi.org/10.1787/bb17314a-en

OECD

Organisation for Economic Co-operation and Development

DELSA/ELSA/WD/SEM(2023)4

Unclassified **English - Or. English**

16 February 2023

DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS
EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS COMMITTEE

The Supply, Demand, and Characteristics of the AI Workforce across OECD countries

OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS No. 285

JEL classification: J21, J23, J24, J31, J44

Authorised for publication by Stefano Scarpetta, Director, Directorate for Employment, Labour and Social Affairs.

All Social, Employment and Migration Working Papers are now available through the OECD website at www.oecd.org/els/workingpapers .

Andrew Green, Andrew.Green@oecd.org
Lucas Lamby, Harvard Kennedy School

JT03512468

OECD Social, Employment and Migration Working Papers

www.oecd.org/els/workingpapers

Acknowledgements

This working paper contributes to the OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) programme, which provides policymakers with new evidence and analysis to keep abreast of the fast-evolving changes in AI capabilities and diffusion and their implications for the world of work. The programme aims to help ensure that adoption of AI in the world of work is effective, beneficial to all, people-centred and accepted by the population at large. AI-WIPS is supported by the German Federal Ministry of Labour and Social Affairs (BMAS) and will complement the work of the German AI Observatory in the Ministry's Policy Lab Digital, Work & Society. For more information, visit https://oecd.ai/workinnovation-productivity-skills and https://denkfabrik-bmas.de/.

Federal Ministry
of Labour and Social Affairs

The work was carried out under the supervision of Stijn Broecke (Team Lead for the Future of Work) and Mark Keese (Head of the Skills and Employability Division). The report also benefitted from helpful comments from colleagues from the Directorate for Employment, Labour and Social Affairs specifically Stijn Broecke, Glenda Quintini, Julie Lassébie, Marguerita Lane, and Fabio Manca, as well as from colleagues from the Directorate for Science, Technology and Innovation. Erika Xiomara Chaparro Pérez provided invaluable statistical assistance.

Abstract

This report provides representative, cross-country estimates of the artificial intelligence (AI) workforce across OECD countries. The AI workforce is defined as the subset of workers with skills in statistics, computer science and machine learning who could actively develop and maintain AI systems. For countries that wish to be at the forefront of AI development, understanding the AI workforce is crucial to building and nurturing a talent pipeline, and ensuring that those who create AI reflect the diversity of society. This report uses data from online job vacancies to measure the within-occupation intensity of AI skill demand. The within-occupation AI intensity is then weighted to employment by occupation in labour force surveys to provide estimates of the size and growth of the AI workforce over time. The report finds that the AI workforce in OECD countries is still relatively small – less than 0.3% of employment – but growing rapidly. Workers with AI skills are not representative of the overall employed population in OECD societies: they tend to be disproportionately male and tertiary-educated. The report also finds that demand for AI workers is strong, and there is some evidence that the supply of workers with AI skills may be keeping up with demand in many OECD countries.

Résumé

Ce rapport fournit des estimations représentatives et transnationales de la main-d'œuvre en intelligence artificielle (IA) dans les pays de l'OCDE. La main-d'œuvre de l'IA est définie comme le sous-ensemble des travailleurs ayant des compétences en statistiques, en informatique et en apprentissage automatique qui pourraient développer et maintenir activement des systèmes d'IA. Pour les pays qui souhaitent être à l'avant-garde du développement de l'IA, il est essentiel de comprendre la main-d'œuvre de l'IA pour créer et entretenir un vivier de talents, et pour s'assurer que ceux qui créent l'IA reflètent la diversité de la société. Ce rapport utilise des données provenant d'offres d'emploi en ligne pour mesurer l'intensité de la demande de compétences en IA au sein d'une même profession. L'intensité de l'IA au sein d'une profession est ensuite pondérée par l'emploi par profession dans les enquêtes sur la population active afin de fournir des estimations de la taille et de la croissance de la main-d'œuvre de l'IA au fil du temps. Le rapport constate que la main-d'œuvre de l'IA dans les pays de l'OCDE est encore relativement faible - moins de 0,3 % de l'emploi - mais qu'elle croît rapidement. Les travailleurs possédant des compétences en IA ne sont pas représentatifs de l'ensemble de la population active des pays de l'OCDE : ils ont tendance à être, de manière disproportionnée, des hommes et à avoir une formation supérieure. Le rapport constate également que la demande de travailleurs en IA est forte et que certains éléments indiquent que l'offre de travailleurs ayant des compétences en IA pourrait suivre la demande dans de nombreux pays de l'OCDE.

Übersicht

Dieser Bericht enthält repräsentative, länderübergreifende Schätzungen der Arbeitskräfte im Bereich der künstlichen Intelligenz (KI) in den OECD-Ländern. Die KI-Arbeitskräfte sind definiert als die Untergruppe der Arbeitnehmer mit Kenntnissen in Statistik, Informatik und maschinellem Lernen, die aktiv KI-Systeme entwickeln und instandhalten könnten. Für Länder, die an der Spitze der KI-Entwicklung stehen wollen, ist das Verständnis der KI-Arbeitskräfte entscheidend für den Aufbau und die Pflege einer Talentpipeline und um sicherzustellen, dass diejenigen, die KI entwickeln, die Vielfalt der Gesellschaft widerspiegeln. In diesem Bericht werden Daten aus Online-Stellenangeboten verwendet, um die Intensität der Nachfrage nach KI-Fachkräften innerhalb eines Berufsfeldes zu messen. Die berufsinterne KI-Intensität wird dann mit der Beschäftigung nach Berufen in Arbeitskräfteerhebungen gewichtet, um Schätzungen der Größe und des Wachstums der KI-Arbeitskräfte im Laufe der Zeit zu erhalten. Der Bericht kommt zu dem Ergebnis, dass die Zahl der KI-Beschäftigten in den OECD-Ländern immer noch relativ gering ist - weniger als 0,3 % der Beschäftigten -, aber schnell wächst. Arbeitnehmer mit KI-Fähigkeiten sind nicht repräsentativ für die gesamte erwerbstätige Bevölkerung in den OECD-Gesellschaften: Sie sind tendenziell überproportional männlich und verfügen häufiger über einen Hochschulabschluss. Der Bericht stellt auch fest, dass die Nachfrage nach KI-Arbeitskräften groß ist, und es Anzeichen dafür gibt, dass das Angebot an Arbeitskräften mit KI-Fähigkeiten in vielen OECD-Ländern mit der Nachfrage Schritt halten könnte.

Table of contents

Acknowledgements 3

Abstract 4

Résumé 5

Übersicht 6

Executive Summary 9

Principaux résultats 11

Kurzfassung 13

1 Introduction 15

2 Measuring the AI workforce 19

Using online job vacancies to estimate the size of the artificial intelligence workforce in OECD labour markets 19
The share of AI workers is low, but rising 22

3 The demographic profile of the AI workforce 27

The AI workforce is highly educated and primarily male 27
Adult learning does not appear to be a key channel for AI skills acquisition 32

4 The Demand for the AI Workforce 36

The demand for workers with AI skills is generally strong, though wages suggest supply may be keeping up with demand 36

References 41

Annex A. Additional Figures and Tables 44

Annex B. Estimates of the share of AI vacancies using a Naïve Bayes classifier 49

The Naive Bayes classifier applied to Lightcast data using skills as features 49
The Naive Bayes classifier is a generalisation of list-based approaches to classification 51
The naïve Bayes estimator provides similar estimates to list-based classifiers without the need for a list of AI skills 53

Tables

Table A.1. Methods for identifying AI–related jobs and jobs with AI exposure from the literature 44

Figures

Figure 2.1. Share of employment with AI skills in selected OECD countries, 2019 23
Figure 2.2. The AI workforce is growing rapidly 24
Figure 2.3. The employment shares of the top 10 AI occupations, by country 2019 26
Figure 3.1. Educational attainment of the AI workforce in OECD countries 28
Figure 3.2. Women are underrepresented in the AI workforce 30
Figure 3.3. The AI workforce is not younger, on average 31
Figure 3.4. With the exception of a few countries, the AI workforce is not disproportionately foreign-born 32
Figure 3.5. The AI workforce is as likely to train than those employed with a tertiary education 33
Figure 3.6. AI workforce is not more likely to train for work-related purposes 34
Figure 3.7. Only a small share of training is for ICT 35
Figure 4.1. Employment growth for AI workforce exceeds employment growth overall 37
Figure 4.2. Hours growth has been much stronger for the AI workforce than for workers overall 38
Figure 4.3. There is more heterogeneity in wage growth between the AI workforce and the labour market overall 39
Figure 4.4. The AI workforce is comparatively well-paid 40

Figure A.1. Share of AI job postings according to different methodologies 46
Figure A.2. Job postings classified as demanding AI skills in USA, GBR & DEU according to AI skill list, 2012 - 2019 46
Figure A.3. Share of job postings classified as demanding AI skills using a Naive Bayes classifier, 2019 47
Figure A.4. Decomposition of the change in AI employment shares, 2011 and 2019 47
Figure A.5. Top 10 AI intensive occupations 48
Figure A.6. Top 10 AI intensive occupations predominantly have a tertiary degree 48

Boxes

Box 3.1. Demographic analysis of the AI workforce in labour force surveys 29

Executive Summary

Countries looking to position themselves to maximally benefit from AI will want to be at the forefront of artificial intelligence (AI) development. The past few years have seen growing advances and interest in AI. As with previous technological advances, AI is likely to bring benefits to workers and firms, and it should lead to higher growth overall.

Cultivating a robust, diverse, AI talent pipeline will be crucial for successful AI development and deployment. A skilled and diverse workforce is a key input to the development of AI. The successful outputs of AI are mostly software applications and their advancement will rely on the skills and human capital of its developers. The diversity of the AI workforce is just as important as its size, or level of skill. AI systems partially reflect the data available, and model choices of its developers. A diverse AI workforce will contribute to an equitable AI that works for all of society.

This report provides cross-country evidence on the supply of, and demand for, the AI workforce. Despite its importance to AI development, little is known about the AI workforce—defined here as the workers who have the set of skills to develop and maintain AI systems. First order questions such as who is developing or using AI directly in their jobs are not yet well covered by official statistics.

The AI workforce currently represents only a small share of overall employment in OECD countries. On average, the AI workforce represents a little over 0.3% of employment in 2019. However, the AI workforce has grown considerably, almost tripling as a share of employment from less than a decade before. As one might expect from the types of skills required to develop AI systems – principally machine learning methods – the AI workforce is tightly clustered in a small set of high-skill technical occupations, including software engineers and statisticians.

The AI workforce differs from the overall population employed in OECD countries. Over 60% of the AI workforce has at least a tertiary degree, on average, and that figure approaches 80% when one zooms in on the top 10 occupations most demanding AI skills. Less than 40% of the AI workforce are women compared to well over 50% of the employed population with a tertiary degree across OECD countries. In addition, almost 50% of the AI workforce earns above the 80th percentile in the earnings distribution. In contrast, the AI workforce is just as likely to be young or foreign-born compared to the employed population with a tertiary degree.

Compared to other workers, the AI workforce does not appear to rely disproportionately on adult learning as a channel for skill acquisition. The AI workforce is just as likely as workers with a tertiary degree to have trained in the last month. When the AI workforce does train, it is no more likely to train for job-related purposes and training is not especially likely to concentrate on digital or technical subjects that would signal building on their AI skills.

Finally, the evidence suggests strong demand for AI workers. Using various indicators to capture the relative demand for the AI workforce, the report finds that demand for the AI workforce is generally stronger than labour demand for the economy as a whole. In particular, employment and average weekly hours growth for the AI workforce far exceeds that of the overall workforce. Wages of the AI workforce have grown roughly in line with that of the overall workforce, suggesting that countries are perhaps doing a decent job of meeting their demand for AI skills.

Principaux résultats

Les pays qui cherchent à se positionner pour bénéficier au maximum de l'IA voudront être à la pointe du développement de l'intelligence artificielle (IA). Ces dernières années, l'IA a connu des avancées et un intérêt croissants. Comme pour les avancées technologiques précédentes, l'IA est susceptible d'apporter des avantages aux travailleurs et aux entreprises, et elle devrait conduire à une croissance globale plus élevée.

Cultiver un vivier de talents en IA, robuste et diversifié, sera crucial pour réussir le développement et le déploiement de l'IA. Une main-d'œuvre qualifiée et diversifiée est un élément clé du développement de l'IA. Les résultats de l'IA sont principalement des applications logicielles et leur développement dépendra des compétences et du capital humain de ses développeurs. La diversité de la main-d'œuvre de l'IA est tout aussi importante que sa taille ou son niveau de compétence. Les systèmes d'IA reflètent en partie les données disponibles et les choix de modèles de leurs développeurs. Une main-d'œuvre diversifiée dans le domaine de l'IA contribuera à une IA équitable qui fonctionnera pour l'ensemble de la société.

Ce rapport fournit des données transnationales sur l'offre et la demande de main-d'œuvre en IA. Malgré son importance pour le développement de l'IA, on sait peu de choses sur la main-d'œuvre de l'IA - définie ici comme les travailleurs qui possèdent l'ensemble des compétences nécessaires pour développer et maintenir les systèmes d'IA. Les questions de premier ordre, telles que celles de savoir qui développe ou utilise directement l'IA dans son travail, ne sont pas encore bien couvertes par les statistiques officielles.

La main-d'œuvre de l'IA ne représente actuellement qu'une faible part de l'emploi global dans les pays de l'OCDE. En moyenne, la main-d'œuvre de l'IA représente un peu plus de 0,3 % de l'emploi en 2019. Cependant, la main-d'œuvre de l'IA a considérablement augmenté, triplant presque sa part de l'emploi par rapport à moins d'une décennie auparavant. Comme on pouvait s'y attendre au vu des types de compétences requises pour développer des systèmes d'IA - principalement des méthodes d'apprentissage automatique - la main-d'œuvre de l'IA est étroitement regroupée dans un petit ensemble de professions techniques hautement qualifiées, notamment les ingénieurs logiciels et les statisticiens.

La main-d'œuvre de l'IA diffère de l'ensemble de la population active des pays de l'OCDE. Plus de 60 % de la main-d'œuvre de l'IA possède au moins un diplôme de l'enseignement supérieur, en moyenne, et ce chiffre approche les 80 % lorsque l'on se concentre sur les dix professions les plus exigeantes en matière de compétences en IA. Moins de 40 % de la main-d'œuvre de l'IA sont des femmes, alors que plus de 50 % de la population active est diplômée de l'enseignement supérieur dans les pays de l'OCDE. En outre, près de 50 % de la main-d'œuvre de l'IA gagne plus que le 80e percentile dans la distribution des revenus. En revanche, la main-d'œuvre de l'IA est tout aussi susceptible d'être jeune ou née à l'étranger que la population active occupée diplômée de l'enseignement supérieur.

Par rapport aux autres travailleurs, la main-d'œuvre de l'IA ne semble pas dépendre de façon disproportionnée de la formation des adultes comme moyen d'acquisition de compétences. Les travailleurs de l'IA sont tout aussi susceptibles que les travailleurs titulaires d'un diplôme d'études supérieures d'avoir suivi une formation au cours du dernier mois. Lorsqu'ils se forment, les travailleurs de l'IA ne sont pas plus susceptibles de le faire à des fins professionnelles et la formation n'est pas particulièrement susceptible de se concentrer sur des sujets numériques ou techniques qui leur permettraient de renforcer leurs compétences en IA.

Enfin, les données suggèrent une forte demande de travailleurs de l'IA. En utilisant divers indicateurs pour saisir la demande relative en main-d'œuvre de l'IA, le rapport constate que la demande pour cette main-d'œuvre est généralement plus forte que la demande de main-d'œuvre générale. En particulier, la croissance de l'emploi et des heures hebdomadaires moyennes de la main-d'œuvre de l'IA dépasse de loin celle de l'ensemble de la main-d'œuvre. Les salaires de la main-d'œuvre de l'IA ont augmenté à peu près comme ceux de l'ensemble de la main-d'œuvre, ce qui donne à penser que les pays parviennent peut-être assez bien à répondre à la demande en compétences en IA.

Kurzfassung

Länder, die sich in eine Position bringen wollen, in der sie maximal von der KI profitieren können, werden bei der Entwicklung der künstlichen Intelligenz (KI) eine Vorreiterrolle einnehmen wollen. In den letzten Jahren haben die Fortschritte und das Interesse an KI zugenommen. Wie bei früheren technologischen Fortschritten wird die KI wahrscheinlich Vorteile für Arbeitnehmer und Unternehmen bringen und dürfte insgesamt zu einem höheren Wachstum führen.

Der Aufbau einer robusten, vielfältigen KI-Talentpipeline wird für eine erfolgreiche KI-Entwicklung und -Einführung entscheidend sein. Eine qualifizierte und vielfältige Belegschaft ist eine wichtige Voraussetzung für die Entwicklung von KI. Die erfolgreichen Ergebnisse der KI sind zumeist Softwareanwendungen, deren Entwicklung von den Fähigkeiten und dem Humankapital der KI-Entwickler abhängen wird. Die Vielfalt der KI-Mitarbeiter ist ebenso wichtig wie ihre Größe oder ihr Qualifikationsniveau. KI-Systeme spiegeln teilweise die verfügbaren Daten und die Modellauswahl ihrer Entwickler wider. Eine vielfältige KI-Belegschaft wird zu einer gerechten KI beitragen, die für die gesamte Gesellschaft funktioniert.

Dieser Bericht liefert länderübergreifende Erkenntnisse über das Angebot an und die Nachfrage nach KI-Arbeitskräften. Trotz ihrer Bedeutung für die KI-Entwicklung ist nur wenig über die KI-Arbeitskräfte bekannt - hier definiert als die Arbeitnehmer, die über die Fähigkeiten zur Entwicklung und Instandhaltung von KI-Systemen verfügen. Fragen erster Ordnung, wie z.B. wer KI entwickelt oder direkt an seinem Arbeitsplatz einsetzt, werden von den amtlichen Statistiken noch nicht gut erfasst.

Die KI-Beschäftigten machen derzeit nur einen kleinen Teil der Gesamtbeschäftigung in den OECD-Ländern aus. Im Durchschnitt machen die KI-Beschäftigten etwas mehr als 0,3 % aller Beschäftigten im Jahr 2019 aus. Die Zahl der KI-Beschäftigten ist jedoch beträchtlich gestiegen, ihr Anteil an der Gesamtbeschäftigung hat sich seit weniger als einem Jahrzehnt fast verdreifacht. Wie man angesichts der für die Entwicklung von KI-Systemen erforderlichen Qualifikationen - hauptsächlich Methoden des maschinellen Lernens - erwarten könnte, konzentrieren sich die KI-Beschäftigten auf eine kleine Gruppe hochqualifizierter technischer Berufe, darunter Softwareingenieure und Statistiker.

Die KI-Beschäftigten unterscheiden sich von der Gesamtbevölkerung in den OECD-Ländern. Über 60 % der KI-Beschäftigten haben im Durchschnitt mindestens einen Hochschulabschluss, und diese Zahl nähert sich 80 %, wenn man die 10 Berufe mit den höchsten Anforderungen an KI-Fähigkeiten näher betrachtet. Weniger als 40 % der KI-Beschäftigten sind Frauen, verglichen mit weit über 50 % der erwerbstätigen Bevölkerung mit Hochschulabschluss in den OECD-Ländern. Außerdem verdienen fast 50 % der KI-Beschäftigten über dem 80. Perzentil der Einkommensverteilung. Im Gegensatz dazu ist die Wahrscheinlichkeit, dass die AI-Erwerbstätigen jung oder im Ausland geboren sind, genauso hoch wie bei der erwerbstätigen Bevölkerung mit Hochschulabschluss.

Im Vergleich zu anderen Arbeitnehmern scheinen die KI-Beschäftigten nicht unverhältnismäßig stark auf die Erwachsenenbildung als Kanal für den Erwerb von Qualifikationen angewiesen zu sein. Die KI-Erwerbstätigen haben sich im letzten Monat genauso häufig weitergebildet wie die Erwerbstätigen mit einem tertiären Abschluss. Wenn sie sich weiterbilden, dann nicht häufiger zu beruflichen Zwecken, und es ist nicht besonders wahrscheinlich, dass sich die Weiterbildung auf digitale oder technische Themen konzentriert, die einen Ausbau ihrer KI-Fähigkeiten signalisieren würden.

Schließlich deuten die Daten auf eine starke Nachfrage nach KI-Arbeitskräften hin. Unter Verwendung verschiedener Indikatoren zur Erfassung der relativen Nachfrage nach KI-Arbeitskräften stellt der Bericht fest, dass die Nachfrage nach KI-Arbeitskräften im Allgemeinen stärker ist als die Nachfrage nach Arbeitskräften in der Wirtschaft insgesamt. Insbesondere das Wachstum der Beschäftigten und der durchschnittlichen Wochenstunden für die KI-Arbeitskräfte übersteigt bei weitem das der Gesamtbelegschaft. Die Löhne der KI-Arbeitskräfte sind in etwa im gleichen Maße gestiegen wie die Löhne der Arbeitskräfte insgesamt, was darauf hindeutet, dass die Länder die Nachfrage nach KI-Fachkräften eventuell ganz gut befriedigen können.

1 Introduction

1. The past few years have seen growing interest in artificial intelligence (AI). Spurred by new statistical approaches, such as machine learning, along with increasingly large datasets and better computing power, modern AI systems can take human-defined objectives and make predictions to formulate recommendations for action (OECD, 2019[1]). AI uses software and machines to replace tasks previously performed by human labour across a diverse set of sectors (Acemoglu and Restrepo, 2019[2]). It therefore represents a general purpose technology (GPT) and, if history is any guide, it will likely lead to increases in productivity, enhancements in science and innovation, as well as changes to employment and inequality (Agrawal, Gans and Goldfarb, 2019[3]).

2. As with previous general purpose technologies, artificial intelligence is likely to bring benefits to some workers and firms, but it may leave others in a more precarious position. AI will automate some tasks rendering certain occupations or even firms redundant. At the same time, other workers or firms will benefit from the enhanced productivity, and better quality goods and services AI delivers. This should increase wages, profits and/or market share. Winners and laggards will also emerge at the national, regional or institutional level. For example, digital technology companies – themselves leading developers of AI – have tended to cluster in tight regional agglomerations such as Seattle (Microsoft, Amazon) and Silicon Valley (Google, Apple and Facebook). Such clustering can produce productivity spill overs and contribute to an outsized share of innovation at the national level (Moretti, 2021[4]). To avoid falling behind, national governments have formulated AI strategies in an attempt to shape and develop AI within their borders.[1]

3. A skilled workforce is one of the pillars to the development of artificial intelligence. The current wave of AI development is concerned with prediction and aiding decision-making using machine learning which itself is a branch of computational statistics (Agrawal, Gans and Goldfarb, 2019[5]). In contrast to previous waves of GPT innovations, and much like the internet revolution, the commercially successful outputs of AI development are mostly software applications and their development does not require an abundance of physical space or capital to develop.[2] Future advances in AI development will rely mostly on the human capital of its developers as well as on a dynamic labour market for AI talent.

4. The diversity of the AI workforce is just as important as its size, or skill level. AI systems partially reflect the data available, and model choices of its developers, as well as the institutions where they are implemented (Salvi del Pero, Wyckoff and Vourc'h, 2022[6]). This can be one source of bias in AI decision-making, and to avoid such bias, it is important that the AI workforce is representative of society at large. For example, if an AI system is trained on internal firm data in order to determine who to promote, but this firm has a discriminatory history of promoting only men, AI will likely replicate these policies and continue to mostly promote men unless developers notice this problem and adjust the models accordingly (Lane and Saint-Martin, 2021[7]). Decision-making bodies – and the tools that aid them, such as AI -- will tend to

[1] See, for example, Van Roy et al., (2021[27]) for an overview of recent policy strategies in European Union countries. The United States established a National AI initiative in 2021 with the passage of the National AI Initiative Act of 2020. The OECD itself has also been active in coordinating and sharing AI policies among its members including recommendations to member countries (OECD, 2022[28]).

[2] Robotics is an exception though it depends on how one chooses to classify robots within the classification of AI.

take actions that reflect the preferences and lived experiences of its members.[3] A diverse AI workforce, therefore, is more likely to recognise these biases and lead to a more equitable AI.

5. Despite its importance to AI development, little is known about the AI workforce. First order questions such as who is developing or using artificial intelligence directly in their jobs are not yet well covered by official statistics. This problem often comes back to the lack of a clear, nationally representative classification of AI employment. Existing occupational classifications used in labour force surveys are not detailed enough to capture the niche skills required to develop AI systems. For example, if one would like to measure the size of the AI workforce, one could concentrate on a few occupations, such as software engineers. However, measuring employment at the occupational level will confound those software engineers who have AI skills with those who do not.[4]

6. In addition to problems of measurement, the lack of a clear taxonomy for classifying AI workers results in a deficit of basic knowledge about their demographic characteristics and educational attainment. What is the gender and age profile of workers with AI skills? And how do these workers gain their skills—through initial education, or through on-the-job skill acquisition and/or lifelong learning?

7. This report helps fill this knowledge gap by providing representative estimates of the size of the AI workforce across OECD countries. The focus is on the size, profile and demand for the AI workforce, which are defined as workers with AI skills: workers who possess the skills and competencies to develop AI applications directly, such as knowledge of specific programming languages, software packages and/or statistical models.[5] These workers may not be developing AI systems in their current jobs, but they form the potential talent pool for firms hiring for positions developing AI.[6]

8. To measure the AI workforce, this report calculates the share of job vacancies within an occupation that demand AI skills (where these AI skills are defined using a pre-defined list).[7] This is done using a dataset of the quasi-universe of online job vacancies, which contain descriptions of the key skills and competencies required in posted job vacancies. The share of vacancies within an occupation that require AI skills is then used to build an occupation-based measure for the intensity of demand for AI skills. The final step is to weight this within-occupation AI intensity measure to the employment by occupation distribution in labour force surveys.

9. The report finds that the current share of workers with AI skills is small. Across OECD countries, the average share of employment is just above 0.3%, ranging from 0.5% in the United Kingdom to 0.2% in Greece. Due to concerns about the representativeness of the vacancy data in some countries, and in order to cover as many OECD countries as possible, the preferred estimates for the size of the AI workforce use a common within-occupation intensity measure from the United States, where the share of vacancies

[3] See for example Chattopadhyay and Duflo, (2004[30]) for gender, and Anwar, Bayer and Hjalmarsson, (2012[31]) for race.

[4] This is the approach taken by Gehlhaus and Mutis (2020[8]). One can view this approach as measuring the *potential* AI workforce, and as such, it represents an upper bound on the size of the workforce taking into account those workers who could potentially work on developing AI.

[5] This report will use the terms “AI workforce’, “AI employment” and “AI workers” synonymously to denote the employed population with AI skills.

[6] This is in contrast to workers *exposed* to artificial intelligence. AI exposure is a broader term which includes workers who complement or interact with AI on a daily basis, but likely have no involvement with the development or maintenance of the underlying technology. Workers exposed to AI likely have had some tasks in their jobs automated by AI, but they themselves play no role in its development nor possess the underlying skills to do so (Brynjolfsson, Mitchell and Rock, 2018[26]; Felten, Raj and Seamans, 2018[25]; Georgieff and Hyee, 2021[32]).

[7] “Machine Learning”, “Natural Language Processing” and “Neural Networks” are a few such examples.

demanding AI skills is large. The estimates in this report are, therefore, likely an upper bound on the share of workers with AI skills.

10. Although small, the AI workforce has grown rapidly from less than 0.1% in 2012 to above 0.3% in 2019. This is the equivalent of an occupation growing from the size of all manicurists in the United States, to the size of the childcare workforce.[8] The growth is almost entirely due to increasing demand for AI skills within occupations, rather than to the growth in occupations demanding AI skills.

11. The analysis also reveals that the AI workforce is confined to a narrow demographic segment of the population. The majority of the AI workforce has at least a tertiary degree. When one focuses on the top 10 occupations with the largest shares of workers with AI skills, more than 75% of workers have at least a tertiary degree. In addition, the AI workforce is disproportionately male. Only 35% of the AI workforce is female, compared to over 50% of the entire population with a tertiary degree. Finally, and with some notable country exceptions, the share of workers with AI skills who are foreign-born is no different than the overall employed population with a tertiary degree, nor is there any evidence that the AI workforce is disproportionately younger.

12. There is no indication that the AI workforce relies heavily on adult learning for skill acquisition. The AI workforce is just as likely as other workers with a tertiary degree to have trained in the last month. When they do train, the AI workforce is no more likely to train for job-related purposes and training is not especially likely to concentrate on digital or technical subjects that would signal building on their AI skills. This suggests that policy to encourage greater training within firms should be privileged over investments in general education.

13. The report also finds evidence of strong demand for AI skills across most OECD countries in the sample, but with some heterogeneity in the indicators. While employment growth and increasing average weekly hours suggest strong demand for AI workers, wage growth among the AI workforce has not been higher than for other occupations. Regardless, the AI workforce is a high-pay profession: across European OECD countries in the sample, almost 50% of the AI workforce earns above the 80th percentile in the earnings distribution providing further evidence that the AI workforce is a small, unrepresentative slice of the population in OECD countries.

14. A few other studies have examined the AI workforce. Gehlhaus and Mutis, (2020$_{[8]}$) use a variety of datasets including online job postings to study the AI workforce in the United States. Their methodology differs from this report as they concentrate on a set of occupations which have AI skills and the potential to work on AI. In contrast, the present report uses all occupations, but allows the share of workers with AI skills to vary within occupations. The overall estimate for the AI workforce obtained by Gehlhaus and Mutis, (2020$_{[8]}$) is therefore higher than in this report. However, Gehlhaus and Mutis, (2020$_{[8]}$) do find a similar demographic profile of AI workers. In another study, Linkedin (2019$_{[9]}$) examines AI talent in Europe by using a statistical learning procedure to identify Linkedin profiles with AI skills. The analysis is only valid for their social network and it only presents AI talent relative to other countries, and relative to a country's population shares within the European Union. However, their ranking of countries by the share of workers with AI skills is similar to the AI employment shares presented in this report.

15. The analysis in this report is also closely related to the burgeoning academic literature on the demand for AI skills. Alekseeva et al. (2021$_{[10]}$) use job vacancies to study the wage premium for AI skills, and Acemoglu et al. (2022$_{[11]}$) use the same dataset of job vacancies to study the changing skill demands of establishments that hire workers with AI skills. This report uses the same dataset and skill lists to define AI vacancies, but focuses primarily on employment estimates and demographics, as opposed to changing demand for vacancies, skills demanded or wage premiums associated with AI skills. Finally, Squicciarini

[8] In May 2021, there were 120,540 workers employed as manicurist and pedicurists in the United States compared to 438,530 childcare workers.

and Nachtigall (2021[12]) use a list of AI skills drawn from Baruffaldi et al., (2020[13]) combined with a more complex decision rule to classify AI job postings. Their estimated share of vacancies demanding AI skills is similar, although slightly smaller, than what is found in this report.

16. This report is organised as follows. Section 2 explains the methodology for estimating the AI workforce and provides estimates of its size and growth over the past decade. Section 3 describes the demographic profile of the AI workforce including how it differs by education, gender, age, and country of birth. It also provides insights into the incidence of adult learning for the AI workforce. Section 4characterises the demand for AI talent using a variety of relevant labour market indicators.

2 Measuring the AI workforce

17. This section provides cross-country, comparable, estimates of the size of the artificial intelligence (AI) workforce in OECD countries. AI workers are defined as workers who have the skills to develop and maintain AI systems. These workers are identified by linking AI skills demanded in job vacancies by occupation to the distribution of employment by occupation in labour force surveys. AI skills in vacancies are identified using a pre-specified list. The share of vacancies demanding AI skills is then aggregated at the occupation level to create a within-occupation AI intensity measure. This intensity measure is used to reweight the occupation-employment distribution in each country using labour force surveys. The final output is a country-level measure of the AI workforce which combines the within-occupation AI intensity measure from online vacancy data, and the country-level occupation distribution of employment. In addition to measuring the size of the AI workforce, this section provides further details on how it has grown over time, and what occupations have the highest demand for AI skills.

Using online job vacancies to estimate the size of the artificial intelligence workforce in OECD labour markets

Data: Job vacancies allow for a finer identification of AI skills

18. Existing occupational classifications and labour force surveys do not neatly capture the incidence of artificial intelligence skills. There is currently no occupation that identifies AI workers. Instead, AI workers appear in various occupations to different extents, and they are classified with other similar, technical workers, who may not be actively developing AI or possess the requisite skills to do so. For example, the ISCO-08 (212) occupation Mathematicians, Actuaries and Statisticians includes the sub-occupation Statisticians, many of whom will have AI skills, although some will not. This occupation classification therefore gives rise to two problems. First, Statisticians are grouped with pure actuaries, for example, who are engaged in running a pension fund, and who would be unlikely to have AI skills. Second, even if Statisticians were grouped separately, some of them would not possess AI skills. The problem is that the specific skills and knowledge which define jobs requiring AI skills are not neatly bundled into one or more occupations, but appear in jobs dispersed across a number of occupations.

19. In order to overcome this problem, researchers and policy makers have turned to unofficial sources of information, such as commercial databases of vacancies or social networks, in order to estimate the size of the AI labour force. These datasets provide insights not available in labour force surveys. For the present study, one advantage of vacancy data is that they provide an up-to-date snapshot of the skills demanded or used in the labour market at a more granular level than is available in conventional labour force surveys, or even in government administrative data.

20. This report uses a dataset of the universe of online job vacancies from Lightcast to measure the size of the AI workforce.[9] Lightcast collects job postings from thousands of online job boards and company

[9] Burning Glass Technologies merged with Emsi and then became known as “Emsi Burning Glass”. In the summer of 2022 they rebranded, and are now called “Lightcast”.

websites, deduplicates, standardises and then disseminates the job postings in machine-readable form. The database includes information on location, sector, occupation, required skills and education of job postings across an expanding list of countries.[10] The use of Lightcast data for academic research has proliferated in recent years, and at least for the United States, has become a widely accepted and representative dataset for the analysis of various aspects of labour demand (Modestino, Shoag and Ballance, 2016[14]; Deming and Kahn, 2018[15]). The advantage of the Lightcast data for this report is that, for each vacancy, Lightcast lists the skills demanded for the job, which are pulled directly from the vacancy listings with Lightcast standardising the spelling and concepts. This goes beyond standard occupational classifications and skill descriptions which is paramount in a rapidly evolving field such as AI that spreads across occupations quickly and manifests itself in different and sometimes niche skills demands.[11]

21. Lightcast data does, however, have some limitations. One of these is that not all jobs are posted online. The data will therefore miss jobs that are advertised word to mouth, as well as occupations and industries that rely on more informal networks for hiring. Lightcast data will therefore somewhat oversample high-skilled jobs that have a higher probability of being posted online (Carnevale, Jayasundera and Repnikov, 2014[16]). This will likely result in a slight overestimate of the share of vacancies demanding AI skills, as such vacancies tend to be overwhelmingly high-pay, high-skilled and posted online. However, previous research indicates that the Lightcast data is representative of aggregate variables at the occupational level in the United States, a wider set of Anglophone countries, and many OECD European Union countries (Hershbein and Kahn, 2018[17]; Cammeraat and Squicciarini, 2021[18]; Araki et al., 2022[19]).[12]

Methodology: within-occupation demand for AI skills

22. This report uses the share of vacancies within each occupation demanding AI skills as a proxy for the within-occupation share of the AI workforce. The skills demanded in a vacancy are used to classify vacancies as demanding AI skills, or not. Following the literature, the classification procedure uses a pre-specified list of skills. If at least one of the skills on this list is demanded in a vacancy, the vacancy is classified as demanding AI skills. The share of vacancies classified as demanding AI skills within an occupation is the measure of within-occupation AI intensity.

23. The within-occupation AI intensity measure is then merged to labour force surveys at the occupational level to obtain an estimate of total employment with AI skills. The within-occupation demand for AI skills and the distribution of national occupation shares from the labour force surveys combine to provide an estimate of AI employment.[13] To be explicit, the AI employment estimate is constructed as follows:

---

[10] The United States, Canada, the United Kingdom, Australia, New Zealand and Singapore were the original six Anglophone countries. Lightcast has now expanded to covering many European Union countries.

[11] See OECD, (2021[34]) for a further discussion of the advantages of using vacancy posting data with skills demanded compared to more traditional sources of skills in occupations such as O*NET and ESCO.

[12] Vacancies may not report some skills because they are implicit or captured in the job title such as cooks and cooking skills. Firms may also report skills that are not needed for the position. For example, firms may at times demand more and rarer skills in job postings as rationing mechanism in slack labour markets rather than because the skills are required in the job (Modestino, Shoag and Ballance, 2020[35]).

[13] The key assumption for moving from vacancy data measuring demand for AI skills (a flow), and employment with AI skills (a stock), is that the distribution of skills by occupation is the same. This has never been verified, however, Hershbein and Kahn, (2018[17]) find that the average education requirements found in Lightcast data, another measure of skill, align well with the education levels of employed workers at the occupation level for the United States. Just as with other assumptions made in this report, to the extent this assumption introduces bias into the analysis, this is likely to lead to an overestimate of the size of the AI workforce.

$$AW(l)_{c,t} = \sum_{o} A(l)_{c,t,o} \cdot E_{c,t,o} \qquad (1)$$

24. $AW(l)_{c,t}$ measures total employment of workers with AI skills in country *c,* at time *t*; $E_{c,t,o}$ is total employment in country *c,* at time *t,* in occupation *o*; and $A(l)_{c,t,o}$ is the within-occupation AI intensity measure for country *c* in time *t,* for occupation *o*. All occupations referenced in the report follow the 3-digit ISCO-08 occupation classification. The with-occupation AI intensity measures are computed at the SOC 6-digit level and then converted to ISCO using an employment-weighted crosswalk. Total AI employment is obtained by multiplying employment in an occupation by the share of vacancies demanding AI skills in that occupation, and then summing across all occupations.[14]

25. The choice of skill list, l, will have a non-trivial impact on the final estimate of the AI workforce. The use of skill lists for classifying vacancies follows the economics literature, which has mostly converged on pre-specified skill lists to classify vacancies, but there is no consensus in the literature on the AI skill list to use. This report relies on the skill list specified in Alekseeva et al. (2021[10]) *(Alekseeva)* for the core results, which is the AI skill taxonomy developed by Lightcast for artificial intelligence on data from the United States.[15] The *Alekseeva* list has a few advantages. First, it is clear, transparent and comprehensive. Second, the authors follow the Lightcast skill taxonomy, which allows for easy replication. Finally, the estimates of the AI labour force obtained with this list accord well with other list-based estimates obtained in the literature (Acemoglu et al., 2022[11]) (*Acemoglu)* (Annex Table A.1 provides an overview).[16]

26. There are other possible choices from the literature for classifying vacancies that do not rely on pre-specified lists. These include skill lists derived from Lightcast which follow a data-driven approach (Babina et al., 2020[20]), or selecting a set of occupations most likely to demand AI skills (Gehlhaus and Mutis, 2020[8]). Data-driven approaches to classification allow a better classification of vacancies over time, because fixed lists may not capture new, emerging, skills and competencies. Data-driven approaches are also more portable across countries where language differences may call into question the use of a single list across countries. In addition to the pre-defined list approach, this report also makes use of a data-driven approach to classification, which does not require a pre-specified list 4Annex B). Data driven approaches are dependent on the statistical model one uses, and there is no consensus in the literature on the best model. For this reason, and because of doubts about the representativeness of the Lightcast data in some European countries (see4Annex B), this report uses a pre-defined list for the main results. Finally, approaches that do not classify vacancies, but measure similarity between artificial intelligence and occupations (or other aggregations) both capture too many workers who do not have AI skills, while neglecting others who appear in occupations not classified as demanding AI skills (see above).[17]

[14] The distribution of firms posting vacancies will oversample expanding firms and will therefore represent a biased sample of the underlying firm distribution. This will likely overstate total AI employment if expanding, and likely newer firms, disproportionately hire workers with AI skills.

[15] The full list of skills and keywords is enumerated in Table A1 (p. 17) in Alekseeva et al. (2021[10]).

[16] Squicciarini and Nachtigall, (2021[12]) use a skill list developed by experts (Baruffaldi et al., 2020[13]) augmented with software keywords from UK BEIS. In addition, the authors group skills into categories (“Generic AI keywords”, “AI approaches”, “AI applications”, and “AI software and libraries”) and require at least one skill from two different categories. Despite the different approach, their published estimates of the share of vacancies demanding AI skills are similar to those found in this report.

[17] Dawson, Williams and Rizoiu, (2021[22]) and Manca, (forthcoming[39]) use revealed comparative advantage and vector space models, respectively, to characterise the similarity of different occupations to the skill “artificial intelligence”. They do not classify vacancies as demanding AI skills; without additional assumptions, these models can only say which occupations are most similar to the skill or bundles of skills characterising artificial intelligence.

27. In addition to the choice of list, this report makes the assumption that the within-occupation intensity measure is common to all countries and is estimated from the United States. The list of skills used to classify AI vacancies and generate the within-occupation AI intensity measure is in English, and derived from data from the United States. Applying this list to other OECD Anglophone countries is reasonable, but extending its use to non-Anglophone countries risks biasing the estimates of the set of AI vacancies demanding AI skills.[18] For this reason, and because the representativeness of Lightcast data has only been validated for a subset of non-Anglophone European OECD countries (Cammeraat and Squicciarini, 2021[18]; Araki et al., 2022[19]), this analysis relies on data from the United States for the within-occupation AI intensity measure, $A(l)_{c,t,o}$ which is then applied to all countries. Cross-country estimates of the size of the AI workforce are therefore driven by differences in the occupation distribution across countries. In short, this report gains greater cross-country coverage by introducing some bias to the within-occupation intensity measure. Moreover, compared to applying the list of AI skills derived from the United States to all countries, the bias has a clear direction: the results of this report represent an upper bound on the size of the AI workforce (see below).

The share of AI workers is low, but rising

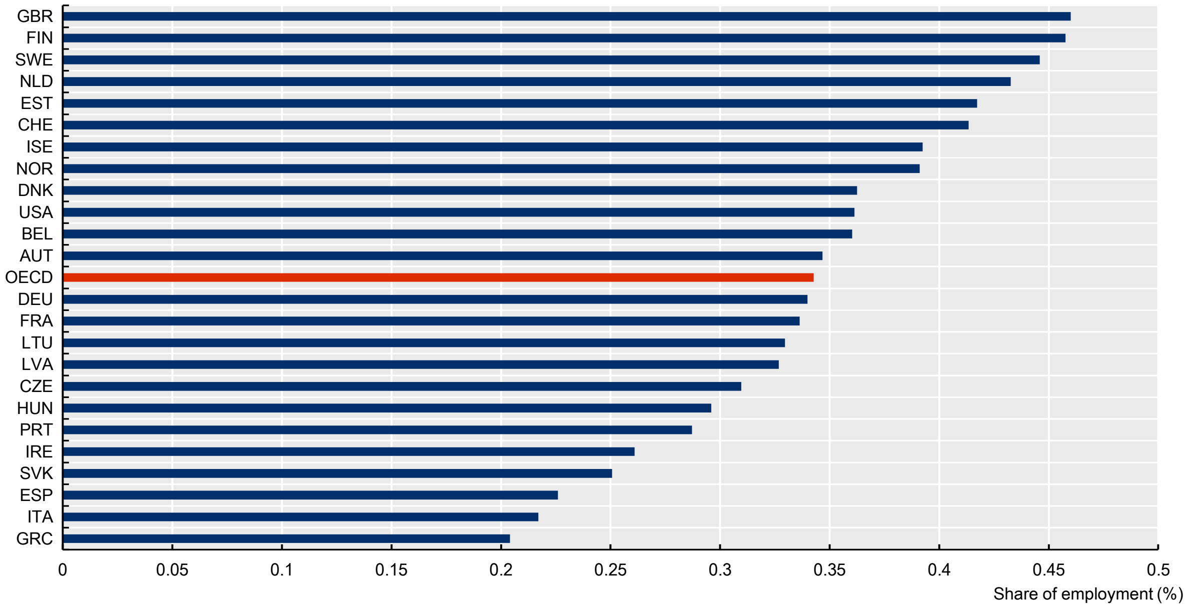
28. The share of workers with AI skills is low – less than half of one percent of employment. Figure 2.1 shows the share of workers with AI skills as a share of total employment as estimated according to equation (1). Across OECD countries in the sample, 0.34% of employment has AI skills. To put this number into perspective, for a hypothetical OECD country with a population of 100 million (and assuming an average employment-to-population ratio of 68.8%) this would amount to an AI workforce of a little over 230,000 workers.[19] This is about one third the total employment in agriculture in France, a country with a total population of approximately 67 million.

29. There is significant variation in the size of the AI workforce across countries. The largest share of employment with AI skills is found in Northern European countries and the United Kingdom (around 0.45%), while Southern European countries such as Greece, Italy, Spain and Portugal have the smallest shares (<0.27%).

[18] When one applies the *Alekseeva* skill list to Lightcast data from the United Kingdom and Germany, the estimates of the overall within-occupation AI intensity are lower than in the United States (Annex Figure A.2). Of course, this could reflect a real effect: demand for AI skills may simply be lower in the United Kingdom and Germany compared to the United States. On the other hand, this may be because this skill list – designed from Lightcast data from the United States – fails to pick up key skills in these countries due to language differences, or differences in skill usage. For example, when applying the *Alekseeva* list to Canada – a similar Anglophone country – there are AI skills which do not appear in the Canadian data at all.

[19] This is likely still an overestimation. The 68.8% employment-to-population ratio is defined over the working-age population rather than the entire population. Applying the 68.8% figure to the smaller working-age population would result in an even lower hypothetical estimate.

Figure 2.1. Share of employment with AI skills in selected OECD countries, 2019



Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).
Source: OECD analysis of Lightcast, European Labour Force Survey (EU-LFS) and Current Population Survey (CPS) data.

30. These estimates are comparable to those obtained in some previous studies. Unsurprisingly (given that the same, or similar, skill lists are used), these results are closest to those of *Alekseeva* and *Acemoglu* for the United States (see Annex Figure A.1). The sum of the within-occupation AI intensity measure across occupations is equal to the overall share of vacancies requiring AI. [20] The estimates of the overall share of vacancies demanding AI skills obtained from the sum of the within-occupation intensity measure in this report are almost identical to those of *Alekseeva* and *Acemoglu.*[21]

31. It is also possible to relate AI employment estimates in this report to those from Linkedin, (2019[9]) which provide employment estimates using a different methodology. The concepts and numbers are different, but the relative ranking of countries within the EU provides a useful reference point. The LinkedIn report finds that OECD Nordic countries have the highest shares of AI workforce, while the lowest shares are found in Eastern European countries. Southern European countries are in the middle or towards the lower end of the ranking.[22]

32. Although the estimates in this report show the AI workforce to be a small share of overall employment, there is reason to believe that the true shares may be even smaller. Estimates of the size of

---

[20] From equation (1), if one takes a vacancy-weighted average of $A(l)_{c,t,o}$ across all occupations, one will recover the overall share of AI vacancies. *Alekseeva* and *Acemoglu* do not estimate the overall size of employment with AI skills, only the share of vacancies demanding AI skills.

[21] An update of Squicciarini and Nachtigall, (2021[12]) estimates through 2019 is available in Samek, Squicciarini and Cammeraat, (2021[40]), and found AI-related jobs accounted for 0.6% of vacancies in the United States in 2019, which is not far from the within-occupation AI intensity used in this report. See Annex Figure A.1 for a comparison of this report's replication results.

[22] OECD.AI, (2022[38]) also provides country-level estimates of workers with AI skills. Due to different data and methodology – the frame is Linkedin users -- the employment shares are larger than in this report and the results are not comparable.

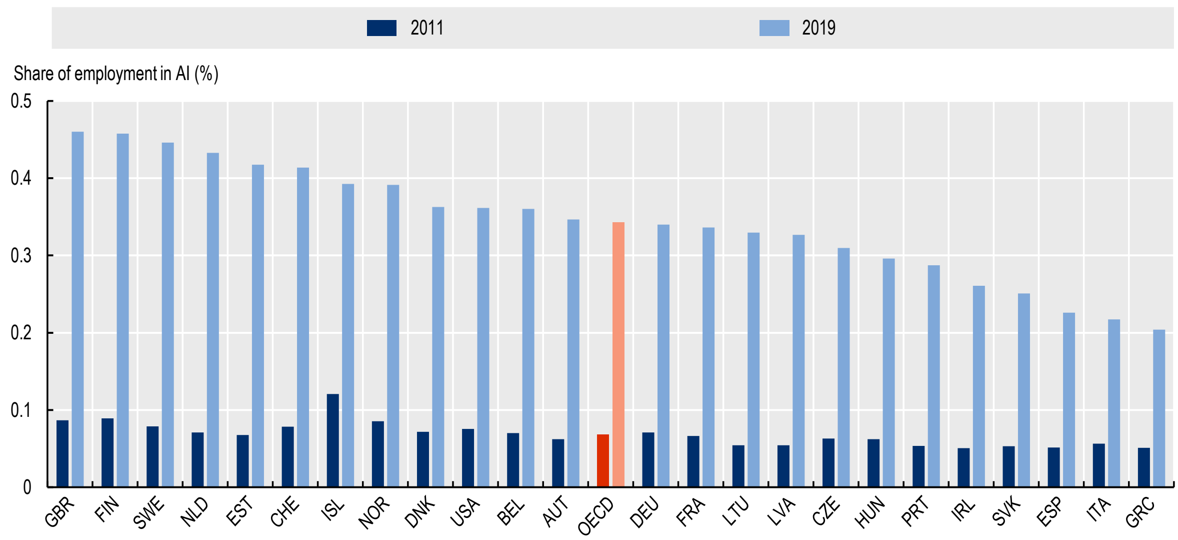
the AI workforce are few, and they largely concentrate on one or a few countries. However, the existing estimates indicate that the United States generally has a larger share of AI vacancies (Linkedin, 2019[9]; Squicciarini and Nachtigall, 2021[12]). To verify this finding, this report additionally uses a Naïve Bayes (NB) classifier to classify vacancies demanding AI skills, which does not require a pre-selected list of AI skills, and uses a common model and one shared skill ("Machine Learning") common across all countries to train the model (4Annex B). Using this classifier, estimates of the share of vacancies demanding AI skills, which are then allowed to vary across countries, provide additional evidence that the within-occupation AI intensities are lower outside of the United States (Annex Figure A.3).

Although small, the AI workforce is growing rapidly

33. The share of workers with AI skills has grown rapidly over the past decade. Figure 2.2 shows the share of workers with AI skills between 2011 and 2019. The share has grown from 0.07% to 0.34% on average for OECD countries in the sample. In general, the countries with the highest share of employment with AI skills in 2019 also had the highest shares in 2011. Notable exceptions are Norway and Iceland, which had some of the highest shares in 2011, but their AI workforce has not grown as fast as in some other countries.

Figure 2.2. The AI workforce is growing rapidly

The Share of employment with AI skills, 2011 & 2019



Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).
Source: OECD analysis of European Labour Force Survey (EU-LFS) for European countries, the Current Population Survey (CPS) for the United States, and Lightcast data.

34. The growth in workers with AI skills is due primarily to an increase in AI skills within occupations, and not to the shifting share of employment across occupations. As noted above, the within-occupation AI intensity measure for all countries is based on the measure from the United States and does not vary across countries. Cross-country variation at a point in time is determined by the occupation distribution across countries. Changes over time in the employment share of AI skills demanded can, therefore, be due either to changes in the occupation distribution, or to a change in the within-occupation AI intensity measure. Employing a shift-share measure to decompose changes in the shares in Figure 2.2 over time

confirms that it is primarily the increase in the intensity of AI within occupations that drives the growth in the AI workforce, and not changes in the occupation distribution (Annex Figure A.4).[23] In sum, across almost all occupations, firms are demanding more workers with AI skills.

AI employment is mostly concentrated in a few high-skill occupations

35. Occupations are the essential link between the within-occupation AI intensity measure and the between-occupation employment shares in labour force surveys. Occupations with high within-occupation AI intensity combined with high shares of overall employment will be responsible for generating higher shares of employment with AI skills. The size of the employment shares in these key occupations is what is responsible for the cross-country variation in the overall size of the AI workforce.

36. The occupations with the highest share of vacancies demanding AI skills are high-skill, technical ones: Mathematicians, Actuaries and Statisticians (5.2% demanding AI skills); Software and Application Developers (4.9%); Information and Communication Technology managers (4.3%); Database and Network Professionals (3.6%) and Electrotechnology Engineers (3.2%).[24] Annex Figure A.5 lists these occupations and the rest of the top 10. The top five occupations all have within-occupation shares of over 3% which is over four times larger than the average rate across all occupations.[25]

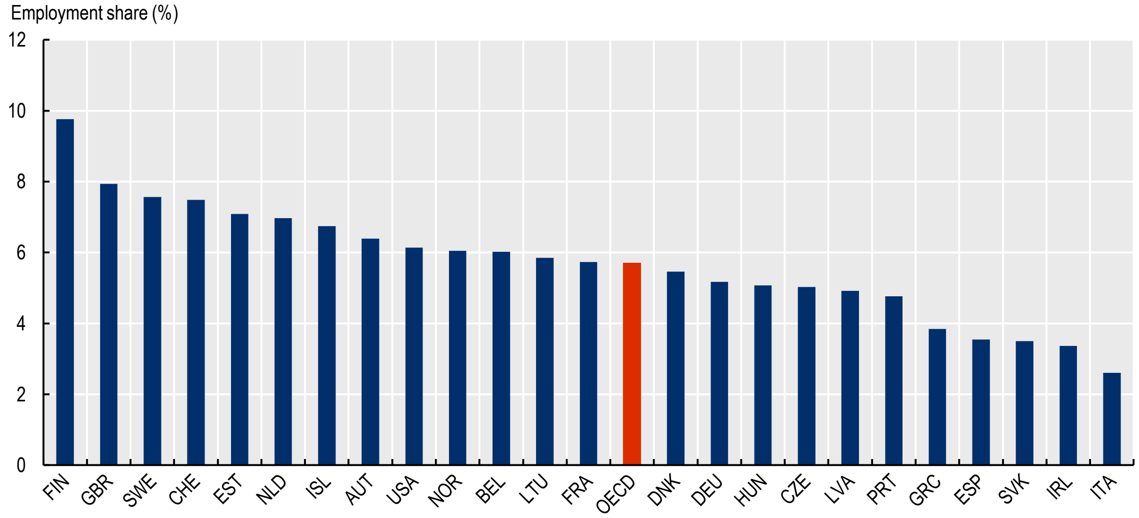
37. The countries with largest AI Workforces are also the countries with the largest employment shares in the top 10 occupations by within-occupation AI intensity. This shows that the size of the AI workforce is driven by employment shares in the occupations most demanding AI skills rather than differences across all possible occupations including those that do not really demand AI skills. Finland, the United Kingdom and Sweden have the highest shares of employment in the top 10 occupations, while Italy, Ireland and the Slovak Republic have the lowest (Figure 2.3).

[23] The shift-share is defined to compose the change in the share of the AI workforce between time *t* and *t+i*: $AW(l)_{c,t+i} - AW(l)_{c,t} = \sum_o \big(A(l)_{c,t+i,o}\big[E_{c,t+i,o} - E_{c,t,o}\big] + E_{c,t,o}\big[A(l)_{c,t+i,o} - A(l)_{c,t,o}\big]\big)$. The first term in the summation is the “between” contribution of changing occupational employment shares, and the second term is the “within” contribution of changing within-occupation demand for AI skills.

[24] All occupations referenced in the report follow the 3-digit ISCO-08 occupation classification.

[25] One occupation, Animal Producers, does not, at first glance, appear to fit with the occupations in the top 10. This is at least partly due to the fact that Lightcast data under samples agriculture workers (Cammeraat and Squicciarini, 2021[18]), and the types of jobs from this occupation that do get posted online skew towards positions demanding AI skills. On the other hand, examples of AI in agriculture are ubiquitous in the press (Page, 2022[41]), and recent work finds that agriculture occupations in general (Lassébie and Quintini, 2022[42]), and Animal Producers specifically (Tolan et al., 2021[43]), may be some of the most at risk of automation from newer developments in AI.

Figure 2.3. The employment shares of the top 10 AI occupations, by country 2019



Note: The top 10 AI occupations reflect the cross-country average of ISCO-08, 3-digit, within-occupation share of online vacancies demanding AI skills determined based on the skill list from Alekseeva et al. (2021[10]). They are (in descending order): Mathematicians, Actuaries and Statisticians - 212; Software and Applications Developers and Analysts - 251; Information and Communications Technology Service managers - 133; Database and Network Professionals - 252; Electrotechnology Engineers - 215; Physical and Earth Science Professionals - 211; Animal Producers - 612; Life Science Professionals - 213; Sales, Marketing and Development Managers - 122; and Engineering Professionals - 214. An employment weighted SOC - ISCO crosswalk has been used to transform the results for the United States.
Source: The European Labour Force Survey (EU-LFS) for European countries, the United Kingdom, Iceland, Norway and Switzerland, and the Current Population Survey (USA).

3 The demographic profile of the AI workforce

38. This section uses the estimates of the within-occupation AI intensity to build the demographic profile of the AI workforce. More specifically, the analysis shows how the AI workforce differs from the employed population with a tertiary degree on dimensions such as: educational attainment, gender, age and country of birth. In addition, the section sheds light on the incidence and types of training undertaken by the AI workforce.

The AI workforce is highly educated and primarily male

39. To provide a demographic profile of the AI workforce, this section builds on the methodology of section 2, which used the within-occupation AI intensity measure from online vacancy data to estimate the share of employment with AI skills. This section slightly tweaks this methodology to reweight the employed population to the demographic characteristic of interest as well as to the occupation distribution. Everything else – including the within-occupation AI intensity – stays the same. Box 3.1 provides details.

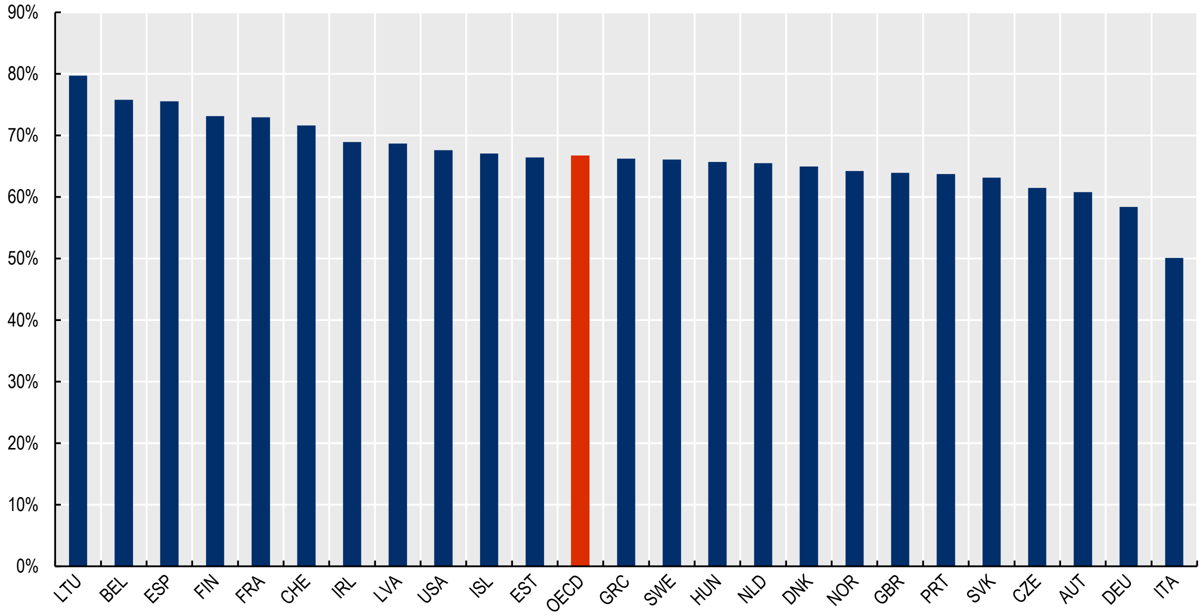
The AI workforce is overwhelmingly composed of workers with a tertiary degree

40. The share of the AI workforce with a tertiary degree is relatively high across OECD countries in the sample. Two thirds of the AI workforce has a tertiary degree, on average, across countries in the sample.[26] In all but four countries, 60% or more of the AI workforce has a tertiary degree (Figure 3.1). This is not a surprise when one considers that the occupations with the highest within-occupation AI intensity are generally high-skill occupations requiring more education on average (above). Lithuania and Belgium have the highest shares with over 75% of the AI workforce holding a tertiary degree. The lowest share is found in Italy – just over 50%. These estimates are in close accordance with findings in the literature. For example, this report finds that slightly less than 68% of the US AI workforce holds a tertiary degree. Using a different methodology and dataset, Gehlhaus and Mutis, (2020[8]) estimate this share to be slightly over 67%.

[26] For comparison, the share of the employed population in the sample in 2019 with a tertiary degree was 38%.

Figure 3.1. Educational attainment of the AI workforce in OECD countries

Share of AI employment with a tertiary degree



Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-education-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 - 2019 for all European countries and 2018 - 2020 for the USA.
Source: OECD analysis of the European Union Labour Force Survey (EU-LFS), Current Population Survey (USA) and Lightcast data.

41. The share of the AI workforce with a tertiary degree is calculated as a weighted average of occupations with the weights computed from the within-occupation AI intensity (Box 3.1). The average therefore gives some weight to most occupations, even ones with little role in developing AI. This partially explains why in some countries the share of workers with AI skills who have a tertiary degree is somewhat low. If one zooms in on the top 10 occupations by within-occupation AI intensity (above), the share with a tertiary degree grows significantly in every country to over 75%, on average (Figure A.6). Given the strong association between tertiary education and the occupations most demanding AI skills, the employed population with a tertiary degree will be used as a within-country comparison group for workers with AI skills.[27]

42. The high share of the AI workforce with a tertiary degree suggests that the employed population with a tertiary degree makes a suitable comparison group for the AI workforce. Presenting the attributes of the AI workforce within the proper context requires a suitable comparison group. Whether the share of the AI workforce that is male is "high" ultimately depends on the question: "compared to what"? The choice of control group is important as it acts as a guide for policy choices. One choice could simply be the employed population, but this risks confusing policies specific to AI with other policies related to attributes correlated with the AI workforce. For example, if one finds the share of the AI workforce that is male is higher compared to the overall employed population, one might be tempted to conclude that policy should work to make AI careers more gender inclusive. However, if this is simply due to the fact that those with a tertiary

[27] Due to data limitations, this report only focuses on those with a tertiary degree. This does not reflect the varying levels within tertiary education. Linkedin (2019[9]) finds that in Europe, much of the AI workforce holds masters and doctorates.

education are more likely to be male (with no difference compared to the AI workforce), then the correct policy is more likely to address gender divides in access to tertiary education more generally. The high share of the AI workforce with a tertiary education is one of its defining characteristics, and the broader group of the employed with a tertiary degree will serve as a comparison group for the rest of this section.

Box 3.1. Demographic analysis of the AI workforce in labour force surveys

The methodology to identify the AI workforce in equation (1) can be tweaked slightly to provide demographic information on the AI workforce. Recall from equation (1) that the within-occupation AI intensity share measures the share of job vacancies in an occupation demanding AI skills. This within-occupation AI share is denoted by $A(l)_{c,t,o}: 0 \leq A(l)_{c,t,o} \leq 1$ in country *c* at time *t* in occupation *o*, and is dependent on the choice of skill list *l.* As explained in the previous section, the analysis uses the United States to derive the within-occupation share for all countries, and the index *c* will henceforth be suppressed for the within-occupations shares. By applying these AI shares to the share of employment in each occupation group within a country at time *t* ($E_{c,t,o}: 0 \leq E_{c,t,o} \leq 1$), estimates of the overall AI workforce can be derived as outlined in section 0.

The demographic analysis of the AI workforce makes use of the demographic information that is available at the occupational level in labour force surveys. The demographic element will be introduced by using employment shares defined over occupation and the demographic variable of interest. Given a demographic variable $x \in X$, employment shares are defined over cells composed of the crossing of occupation by the demographic of interest: $s_{c,t,o,x}: 0 \leq s_{c,t,o,x} \leq 1$. The occupation by demographic cells are mutually exclusive and collectively exhaustive and therefore sum to one. Note as well that these are employment shares and still vary by country, *c*.

The only difference to the methodology outlined for the overall estimates is that the employment shares take on an extra margin. For example, if the demographic of interest is gender ($x \in \{M, F\}$), the employment shares would be defined over cells of occupations and male or female. Taking into account the new margin, the AI talent measure is defined as follows:

$$AW(l)_{c,t} = \sum\nolimits_{x\in\{M,F\}} s_{c,t,x} = \sum\nolimits_{x}\sum\nolimits_{o} s_{c,t,o,x}\, A(l)_{t,o}$$

The term in the middle of the two equalities, $s_{c,t,x}$, is the employment share of the AI workforce that is male (M) or female (F). This results by taking only the first sum over occupation in the last term in the above equation after applying the within-occupation AI intensity. The demographic is fully driven by the occupational AI shares identified in the Lightcast data and the demographic distributions within occupations.

Finally, one can also normalise the within-occupation AI intensity measures in equation (1) to compute country-wide averages as well as sums. This is most salient for section 4 For example, if one replaces $E_{c,t,o}$ (total employment), with $H_{c,t,o}$ (average weekly hours) and normalises the within-occupation AI intensity measure so it sums to one, equation (1) would give the average weekly hours for workers engaged in AI development. measure so it sums to one, equation (1) would give the average weekly hours for workers engaged in AI development. The intuition for this interpretation is that the within-occupation AI shares can be interpreted as weights for the probability of having AI skills in a given occupation. The outcome of interest is then a weighted average of occupations with the weights determined by how likely a vacancy within an occupation is to demand AI skills.

AI workforce is more likely to be male than the employed population with a tertiary degree

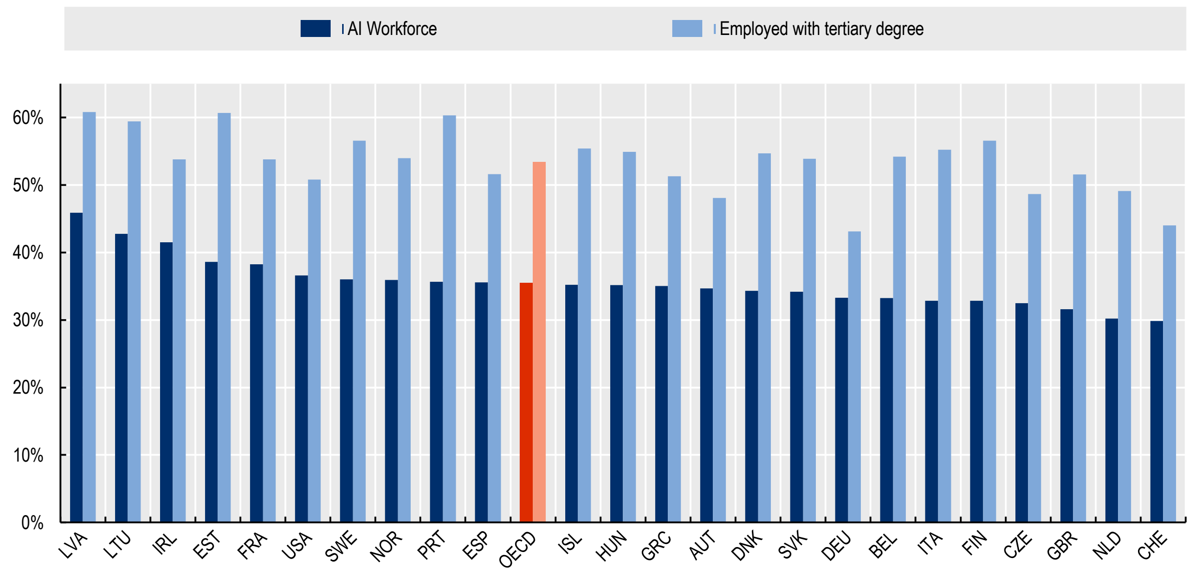
43. Across OECD countries, the AI Workforce is disproportionately male, even compared to the employed population with a tertiary degree. Figure 3.2 shows the sex distribution within the AI workforce across countries. On average, 35% of the AI workforce is female compared to 53% of those employed with a tertiary degree. The share of women in the AI workforce varies between 46% (Latvia) and 30% (Netherlands). There are no countries where women are the majority of the AI workforce, or even are represented in the same proportions as the employed population with a tertiary degree.

44. It is notable that the countries with the largest AI workforces – as measured by employment shares – also have some of the lowest female representation. Finland, the United Kingdom, Switzerland and the Netherlands are all in the bottom six countries with respect to female AI employment shares, and in the top six of overall AI employment.

45. The findings by gender, again, accord well with the literature. Gehlhaus & Mutis (2020[8]) find that 36% of the US AI workforce is female, which is close to the estimates in this report despite using different methodologies and data. In contrast, Linkedin (2019[9]) finds that less than 20% of AI workers in the US are women.

Figure 3.2. Women are underrepresented in the AI workforce

Share of women



Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-gender-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 - 2019 for all European countries and 2018 - 2020 for the USA.
Source: OECD analysis of the European Union Labour Force Survey (EU-LFS), Current Population Survey (USA) and Lightcast data.

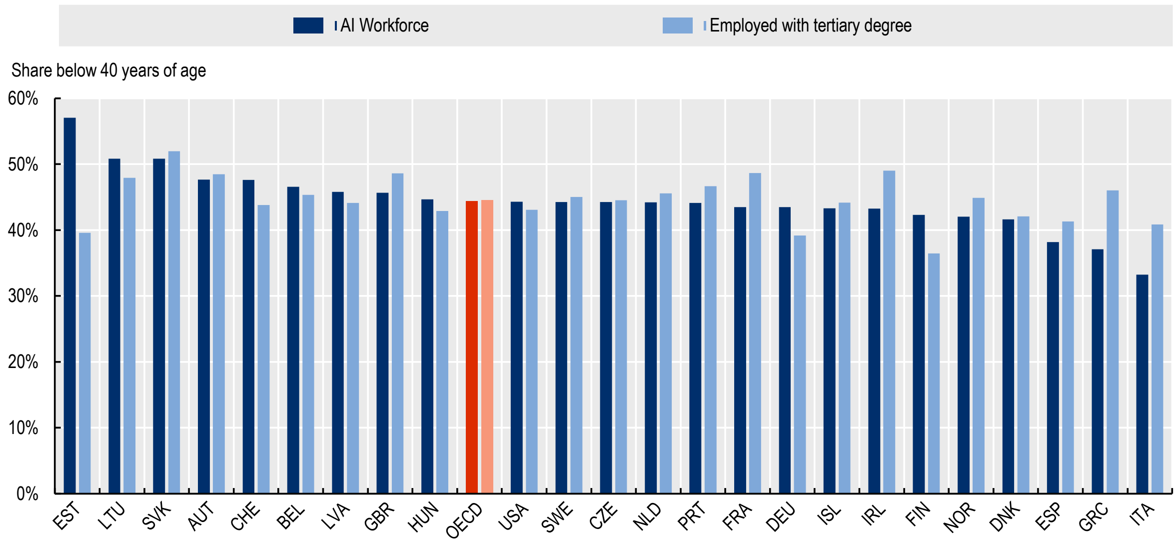
The AI workforce is not noticeably younger than the employed population with a tertiary degree

46. The AI workforce is no younger than the overall employed population with a tertiary degree. Across OECD countries in the sample, 44% of the AI workforce is under 40 years of age compared to 45% of the

overall employed population with a tertiary degree (Figure 3.3). There is some country heterogeneity, with Lithuania and especially Estonia having younger AI workforces. In Greece and Italy, the share of workers under 40 in the AI workforce is significantly less than the share under 40 with a tertiary degree overall. Estonia, Finland and Switzerland all have relatively large AI workforces and they appear to be younger, on average, while the Netherlands and the United Kingdom also have large AI workforces but they do not appear to be any younger than the overall employed population with a tertiary degree. In general, there is no clear pattern relating the age of the AI workforce and the overall size of this AI workforce.

Figure 3.3. The AI workforce is not younger, on average

Share of employment below 40 years of age



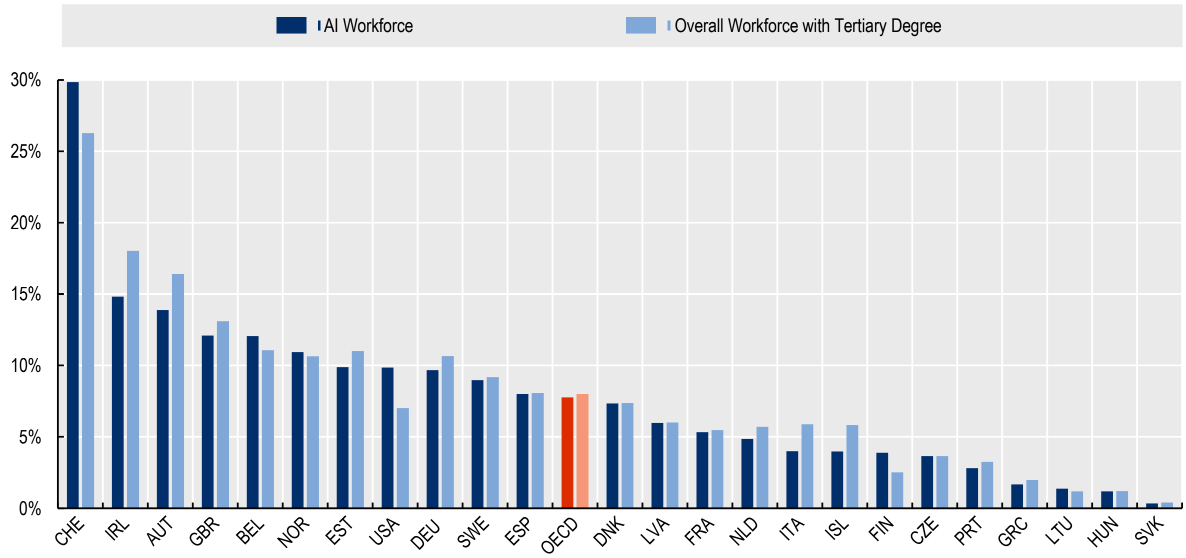
Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-age-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 - 2019 for all European countries and 2018 - 2020 for the USA.
Source: OECD analysis of the European Union Labour Force Survey (EU-LFS), Current Population Survey (USA) and Lightcast data.

The AI workforce is not primarily foreign-born

47. The share of the AI workforce that is foreign-born displays heterogeneity across countries, but on average, it is around 8%, which is similar to the overall employed population with a tertiary degree (Figure 3.4). There are some exceptions, with certain countries relying on a larger foreign-born population for the AI workforce. Switzerland, Ireland, the United Kingdom and Austria have the highest proportion of foreign-born in the AI workforce. Switzerland is an outlier with 30% of its AI workforce foreign-born, which is greater than that of the employed population with a tertiary degree (26%). The United States is another example of a large divergence between the size of its AI workforce that is foreign-born and that of the overall workforce with a tertiary degree (10% vs. 7%). The Slovak Republic and Hungary have the lowest shares, but these shares are still comparable to that of the overall workforce with a tertiary degree. In general, the foreign-born population with AI skills appears to reflect general immigration patterns of a country rather than an exceptional pattern of sourcing workers with these skills from abroad.

Figure 3.4. With the exception of a few countries, the AI workforce is not disproportionately foreign-born

Share which is foreign-born



Note: Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-place of bith-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 - 2019 for all European countries and 2018 - 2020 for the USA.
Source: OECD analysis of the European Union Labour Force Survey (EU-LFS), Current Population Survey (USA) and Lightcast data.

Adult learning does not appear to be a key channel for AI skills acquisition

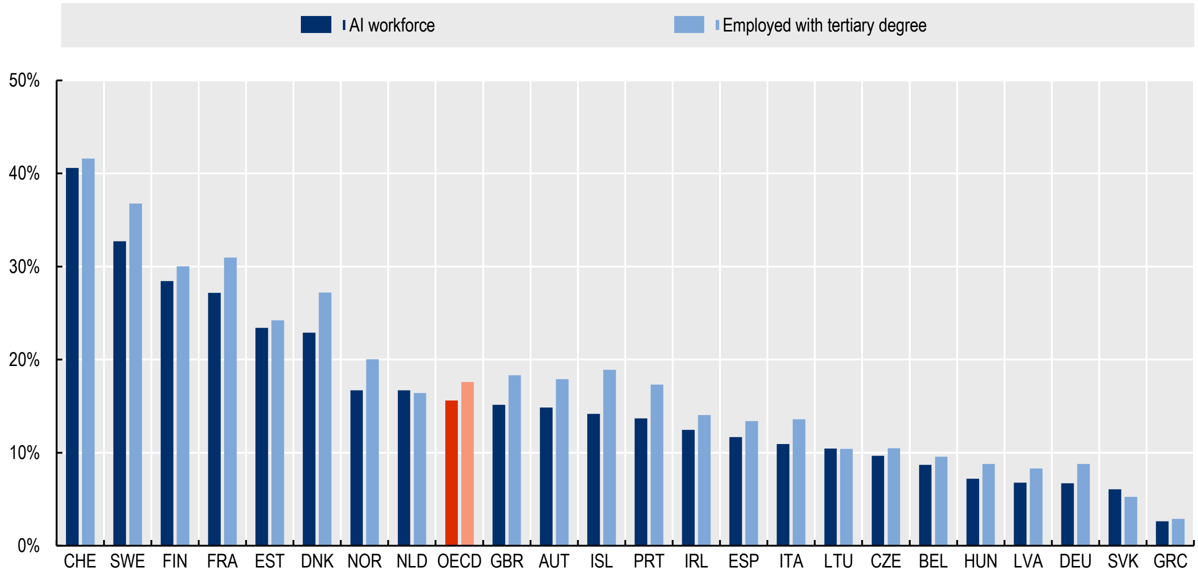
48. The AI workforce is characterised mainly by highly educated workers concentrated in high-skill occupations. An important policy question regarding the AI workforce is how AI talent is developed and what channels firms use to hire workers with AI skills. Are AI skills mainly developed within universities as part of a worker’s initial education, or are AI skills built up later through adult learning? This section takes up the question of whether adult learning is an important channel for the acquisition of AI skills.

The AI workforce is just as likely to participate in adult learning as all workers with a tertiary degree

49. The AI workforce is no more likely to participate in adult learning than employed workers with a tertiary degree overall. The share of workers who report having participated in some sort of training in the last four weeks is 16% for the AI workforce and 18% for the employed population with a tertiary degree. Participation in training varies between countries (Figure 3.5). Over 40% of those employed in both groups in Switzerland have trained in the last four weeks, compared to less than 5% in Greece and the Slovak Republic.

Figure 3.5. The AI workforce is as likely to train than those employed with a tertiary education

Share that has undertaken some sort of training in the past four weeks, 2019



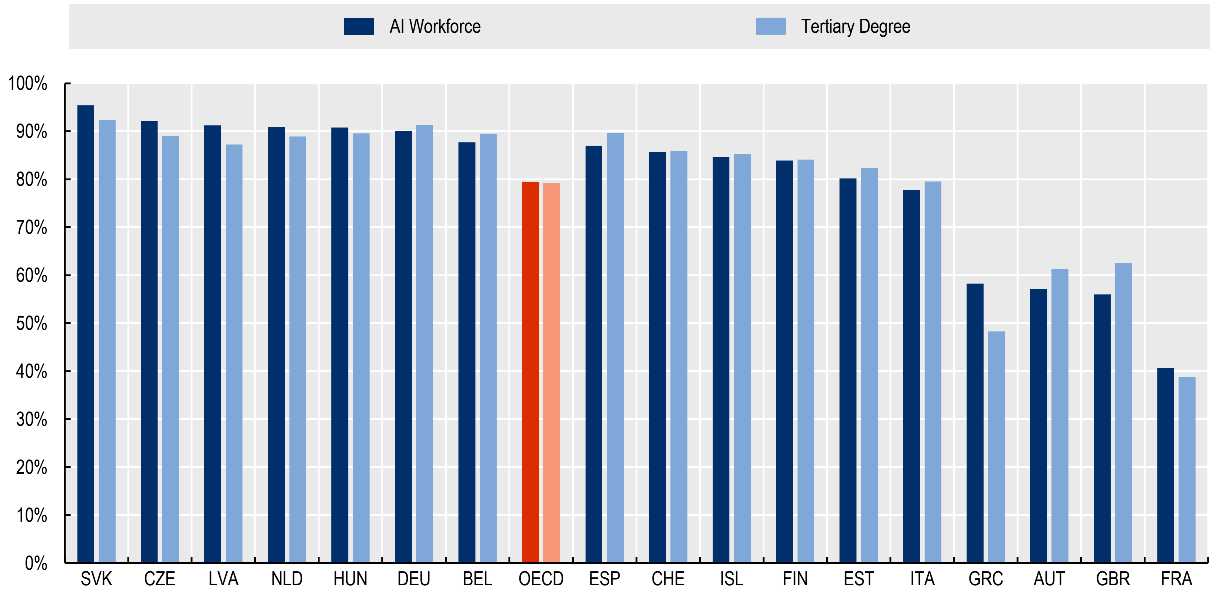
Note: Training is defined as having attended any courses, seminars, conferences or received private lessons or instructions outside the regular education system within the last four weeks. Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-training-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 – 2019.
Source: OECD analysis of European Labour Force Survey (EU-LFS) and Lightcast data.

50. This measure of adult learning, however, confounds job-related training, and training undertaken for other reasons. These other reasons might include passion projects not related to one's current job, or training that is not explicitly tied to current duties. Even though the AI workforce is no more likely to train compared to the employed population with a tertiary degree, the AI workforce may be more likely to engage in job-related training.

51. Just as with training overall, the AI workforce undertakes job-related training at about the same rate as the employed population with a tertiary degree. In both groups, the vast majority of training is for work-related reasons (Figure 3.6). Conditional on participating in any type of training, almost 80% of training by both groups is undertaken for job-related purposes. The highest rates for the AI workforce are in the Slovak Republic, Czech Republic and Latvia with job-related training rates that are generally higher than for the employed population with a tertiary degree in these countries. These are all countries with a below average share of employment with AI skills, and where the AI workforce trains little overall.

Figure 3.6. AI workforce is not more likely to train for work-related purposes

Share of training that was done for job related or for professional purposes, 2019

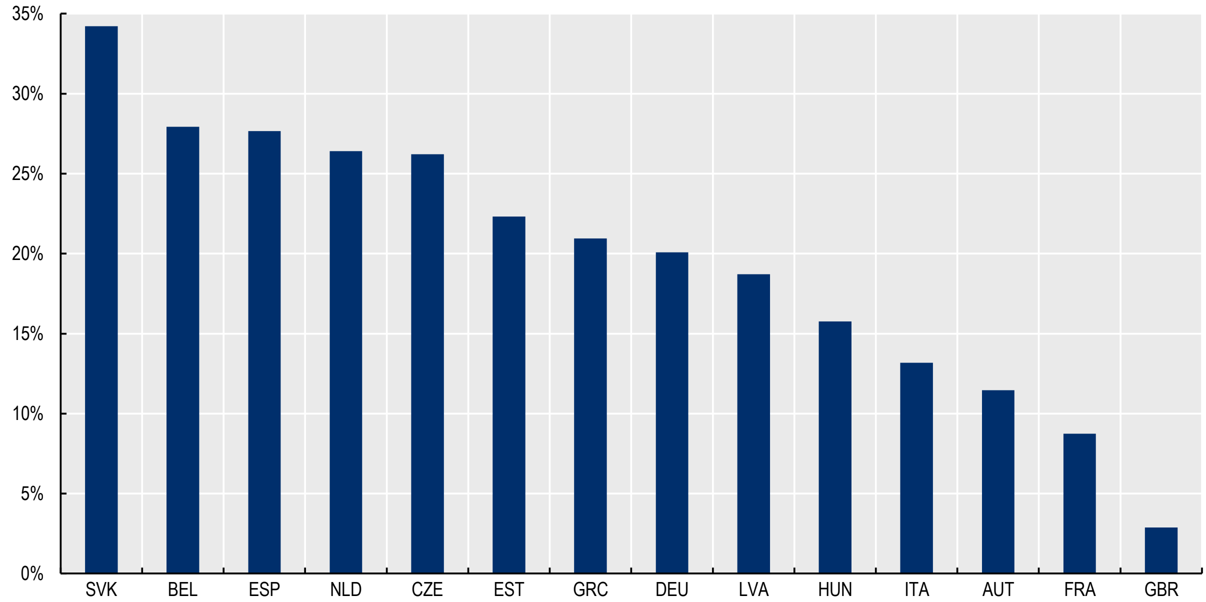


Note: Figures are the share of workers who have undertaken some sort of non-formal training in the last four weeks who report the training was mostly job related with a professional purpose. Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-training-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 – 2019.
Source: OECD analysis of European Labour Force Survey (EU-LFS) and Lightcast data.

52. The majority of training that the AI workforce undertakes is not technical in nature. Information technology services (ICT) can be considered a rough proxy for more technical knowledge required to develop AI. On average, only 20% of all training of the AI workforce is in the field of ICT in those countries for which this data is available (Figure 3.7). This is the largest field of training followed by “Business, Administration and Law”, “Services” and “Arts and Humanities”. There is significant variation across countries with 34% of the AI workforce in the Slovak Republic’s, and 28% in Belgium and Spain, having undergone training in the field of ICT compared to only 11% in Austria and 9% in France. This low share of technical training among the AI workforce casts doubt on the theory that adult learning is an important source of AI skill acquisition.

Figure 3.7. Only a small share of training is for ICT

Share of AI workforce that has undertaken training in information and communication technologies in the past four weeks



Note: The denominator is the AI workforce that has undertaken some sort of non-formal education/training in the past four weeks. This information is not provided by all countries. Estimates obtained by summing the product of within-occupation shares of AI skill demand and employment-training-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).The analysis is pooled over the years 2018 – 2019.
Source: OECD analysis of European labour force survey (EU-LFS) and Lightcast data.

53. After looking at the incidence of training in various ways, there is no strong evidence that adult learning is an important channel for acquiring AI skills. This conclusion is not iron-clad and should be viewed in light of other considerations. Many firms actively developing AI will also be involved in research. Workers who have skills closely related to AI skills may acquire more explicit AI skills simply by being part of a research team or the AI development process within their firms. This will be invaluable for skill acquisition and life-long learning, but would likely not be considered formal work-related training. More research is needed on the detailed mechanics of acquiring AI skills within firms.

4 The Demand for the AI Workforce

54. This section reviews a number of labour market indicators which together paint a picture of the demand for workers with AI skills. The justification for this approach is that there is not one single labour market indicator that provides a clear signal of the underlying labour demand for a particular type of worker. By reviewing a larger number of labour market indicators, one can extract a more reliable signal of labour market demand than with any one indicator individually. The indicators used here are: growth in employment; growth in average weekly hours; and wage growth. The methodology is the same as in the previous sections: one can interpret the AI workforce as a weighted average of occupations with the weights derived from the within-occupation intensity of demand for AI skills.

The demand for workers with AI skills is generally strong, though wages suggest supply may be keeping up with demand

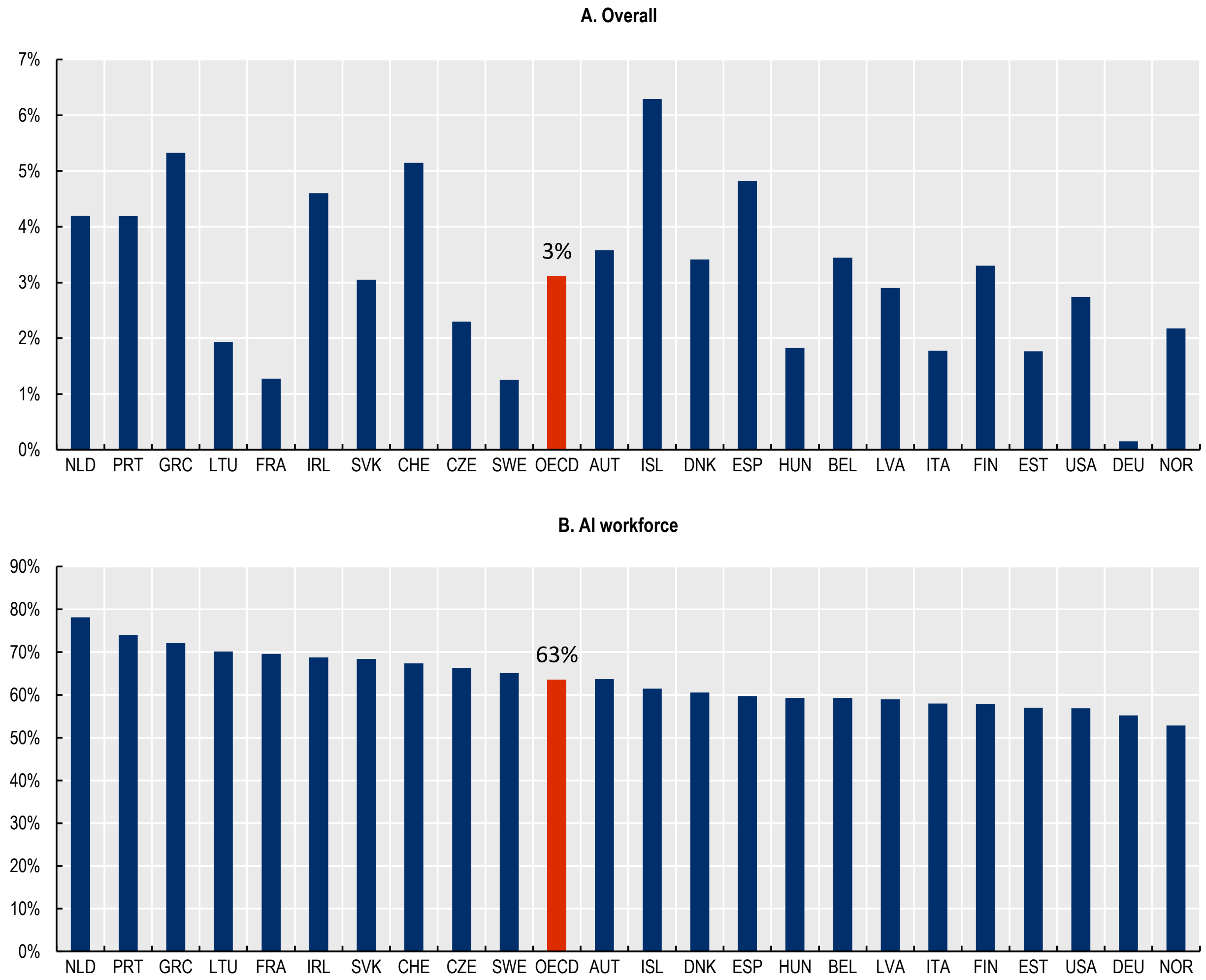
There is strong growth in the employment for workers with AI skills

55. Across all OECD countries in the sample, employment growth for the AI workforce is strong. Figure 4.1 shows employment growth for the AI workforce and for all workers overall. On average, employment growth was 63% for the AI workforce between 2017 and 2019, but just 3% for workers overall.[28] The 63% growth rate over the two years is equivalent to a 28% average annual growth rate and represents an acceleration in growth compared to earlier time periods in the sample (22% average annual growth from 2011-2017). The highest growth rates for the AI workforce were found in the Netherlands, Portugal and Greece. The smallest growth rates were found in the United States, Germany and Norway. However, it is the difference between the growth rate of employment overall and the rate for the AI workforce which provides the clearest interpretation of labour demand, and on that metric, there is little difference between the countries.

[28] An alternate approach to measuring labour demand is to create an index by aggregating the various indicators of labour market demand into a single, standardised measure as in OECD, (2017[29]). Given the short time series and manageable number of indicators, it is clearer to simply present each indicator separately rather than reduce everything to one index. The analysis in this section is partially based on the use of an index, and in keeping with that convention, the comparison of growth rates with the average worker will be used as a reference.

Figure 4.1. Employment growth for AI workforce exceeds employment growth overall

Employment growth for the AI workforce and overall, 2017-2019



Note: Estimates of the AI workforce obtained by summing the product of within-occupation shares of AI skill demand and employment-occupation shares by occupation. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).
Source: European Labour Force Survey (EU-LFS), Current Population Survey (CPS), and Lightcast data.

Growth in average weekly hours also indicates strong demand for the AI workforce

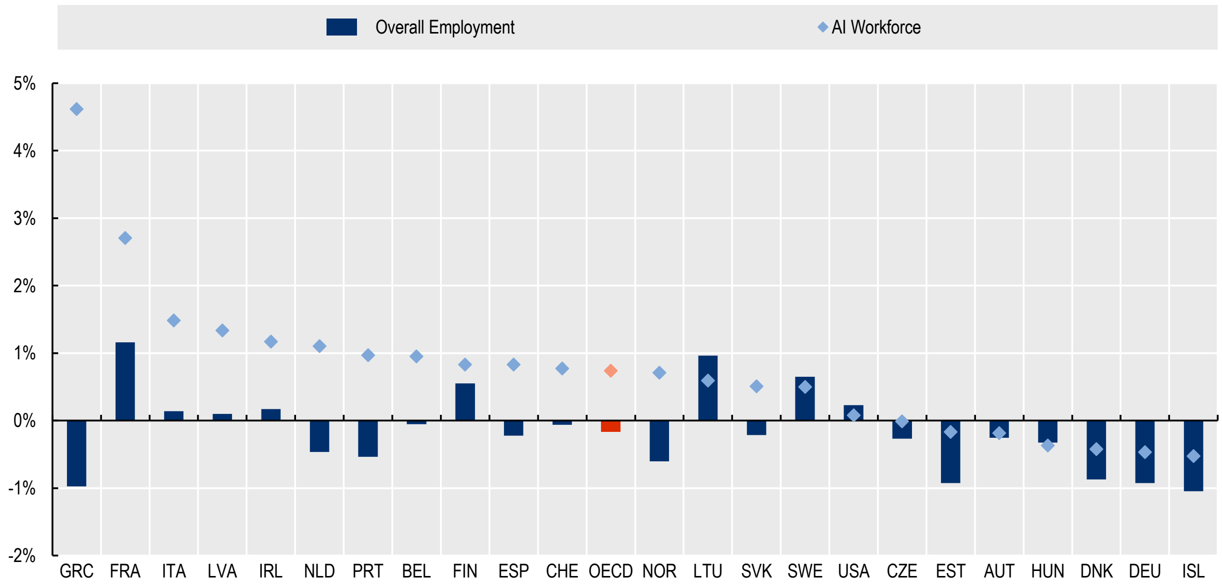
56. Employment growth can be a misleading indicator for underlying labour demand. Hiring is costly for employers, and it often entails certain fixed costs, which can discourage hiring at the margin. As demand begins to build for a particular occupation or segment of the labour market, firms will often first respond to this demand by asking existing workers to work more hours to absorb the increase in demand (Hart, 2017[21]).[29] This means that employment growth may not always fully capture the underlying state of labour demand, and that growth in average weekly hours may provide a better signal in the early stages of an increase in labour demand.

[29] This assumes fixed costs of hiring, declining productivity as hours worked exceeds a certain threshold and firms operating below the threshold at the time of the labour demand shock (Hamermesh, 1993[33]).

57. Just as with employment growth, growth in the average weekly hours of the AI workforce far exceeds that of the economy overall. The average weekly hours of work for the AI workforce has grown by 0.7% over the sample period compared to a *decline* in average weekly hours for the economy overall (Figure 4.2).[30] The largest increases in weekly hours worked for the AI workforce were in Greece, France and Italy. Italy and Greece have some of the lowest shares of overall AI employment. This is maybe an indicator that they are simply later to AI adoption and developing workers with AI skills than other OECD countries in the sample, but labour demand will increase for these countries as well.

Figure 4.2. Hours growth has been much stronger for the AI workforce than for workers overall

Average hours growth for the AI workforce and overall employment, 2017-2019



Note: Estimates of the AI workforce obtained by taking a weighted average of average weekly hours by occupation with the weights corresponding to the within-occupation shares of AI skill demand. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).
Source: European Labour Force Survey (EU-LFS), Current Population Survey (CPS), and Lightcast data.

More muted wage growth for the AI workforce suggests supply of workers may be keeping up with demand

58. Wage growth for the AI workforce does not show the same strong growth compared to employment and hours. Hourly wage growth represents an important perspective when considering overall labour demand. If demand for labour is high, one should expect that to show up in higher wage growth for affected workers. If not, it suggests that the supply of available workers is roughly in line with the strong demand. Where employment and hours pick-up changes in quantities, one can think of changes in wages as the corresponding changes in prices.

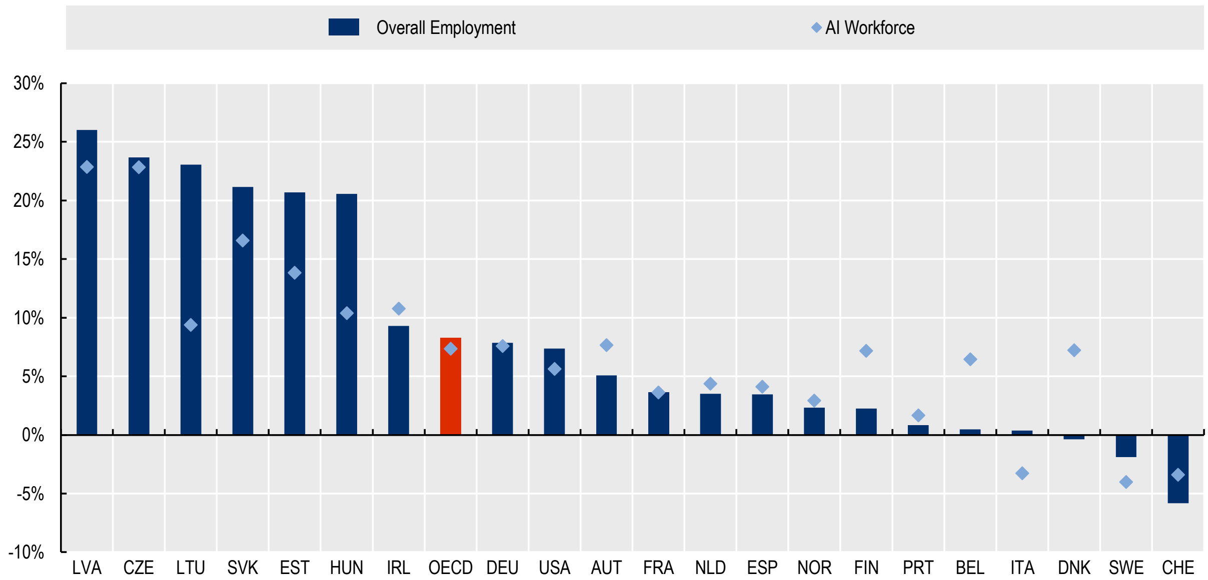
59. On average across countries in the sample, wages grew cumulatively by 7% for the AI workforce compared to 9% overall over the period 2017 to 2019. The largest increases for the AI workforce were in Latvia and the Czech republic, but these are countries which also saw similarly strong earnings growth for

[30] The results are robust to restricting the sample to only full-time workers.

the economy overall. Denmark and Finland, however, saw the largest gaps between hourly wage growth for those with AI skills and the economy overall.[31]

Figure 4.3. There is more heterogeneity in wage growth between the AI workforce and the labour market overall

Average hourly wage growth for the AI workforce and overall employment, 2017-2019



Note: The AI workforce is computed as a weighted average of occupations with the weights derived from observed skills demanded using Lightcast.
Source: European Survey of Income and Living Conditions (EU-SILC), Current Population Survey (CPS), and Lightcast data.

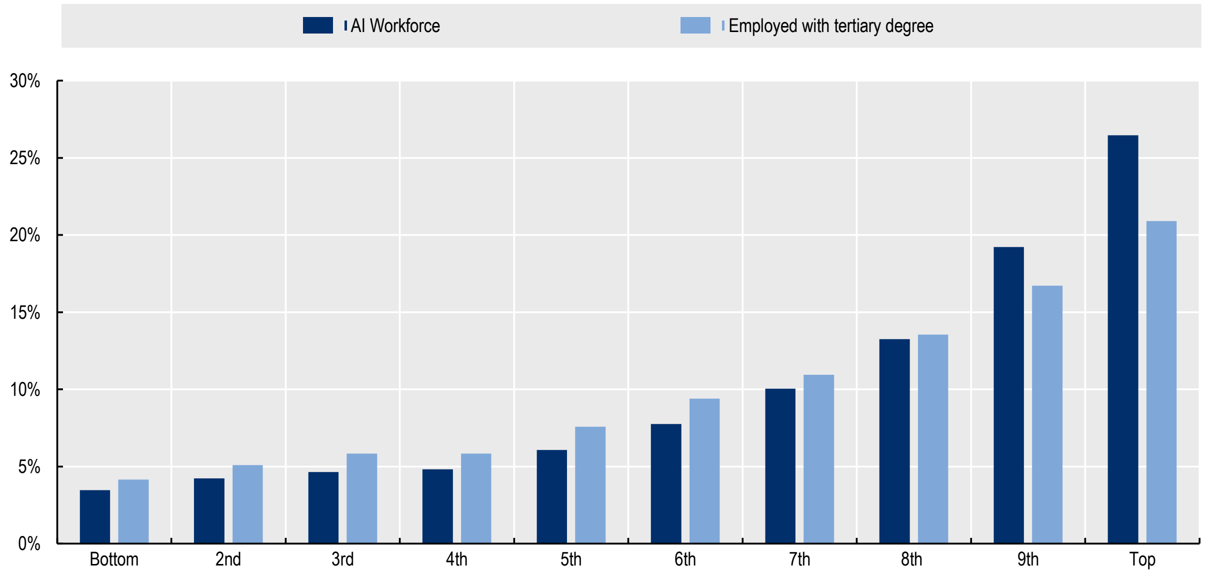
60. Despite slower earnings growth, the level of labour earnings are much higher for the AI workforce. Earnings growth appears to show supply keeping up with demand, but it is important to put this in the proper perspective. Figure 4.4 shows the monthly labour earnings distribution for the AI workforce and the overall workforce with a tertiary degree.[32] The data is pooled across European OECD countries and presented as within-country deciles. The AI workforce is less likely to earn below the 8th tercile (80th percentile) in the overall earnings distribution, and much more likely to earn in the 9th or 10th deciles. In short, the AI workforce is much more likely to earn more than an already high-earning group: those employed with a tertiary degree. This is consistent with Alekseeva (2021[10]), who find that, in the United States, there is a considerable skill premium to AI skills.

[31] The AI workforce is more likely to be compensated in shares of start-ups for which they work. This may lead earnings growth to be understated.

[32] If one would like to continue to compare the AI workforce to the overall employed distribution as in the rest of this section, note that this is equivalent to bars at each decile equalling 10% on the y-axis.

Figure 4.4. The AI workforce is comparatively well-paid

Labour earnings distribution for OECD European countries, 2019



Note: Estimates of the AI workforce obtained by taking a weighted average of decile shares by occupation with the weights corresponding to the within-occupation shares of AI skill demand. Within-occupation AI shares identified according to AI skill lists enumerated in Alekseeva et al. (2021[10]).
Source: European Labour Force Survey (EU-LFS) and Lightcast data.

References

Acemoglu, D. et al. (2022), "Artificial Intelligence and Jobs: Evidence from Online Vacancies", *Journal of Labor Economics*, Vol. 40/S1, pp. S293-S340, https://doi.org/10.1086/718327. [11]

Acemoglu, D. and P. Restrepo (2019), "Artificial Intelligence, Automation, and Work", in Agrawal, A., J. Gans and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, https://doi.org/10.7208/chicago/9780226613475.003.0008. [2]

Agrawal, A., J. Gans and A. Goldfarb (2019), "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction", *Journal of Economic Perspectives*, Vol. 33/2, pp. 31-50, https://doi.org/10.1257/jep.33.2.31. [5]

Agrawal, A., J. Gans and A. Goldfarb (eds.) (2019), *Introduction to The Economics of Artificial Intelligence: An agenda*, University of Chicago Press, https://doi.org/10.7208/chicago/9780226613475.001.0001. [3]

Alekseeva, L. et al. (2021), "The demand for AI skills in the labor market", *Labour Economics*, Vol. 71, p. 102002, https://doi.org/10.1016/j.labeco.2021.102002. [10]

Anwar, S., P. Bayer and R. Hjalmarsson (2012), "The Impact of Jury Race in Criminal Trials", *The Quarterly Journal of Economics*, Vol. 127/2, pp. 1017-1055, https://doi.org/10.1093/qje/qjs014. [31]

Araki, S. et al. (2022), "Monopsony and concentration in the labour market", in *OECD Employment Outlook 2022: Building Back More Inclusive Labour Markets*, OECD Publishing, Paris, https://doi.org/10.1787/0ecab874-en. [19]

Babina, T. et al. (2020), "Artificial Intelligence, Firm Growth, and Industry Concentration", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3651052. [20]

Baruffaldi, S. et al. (2020), "Identifying and measuring developments in artificial intelligence: Making the impossible possible", *OECD Science, Technology and Industry Working Papers*, No. 2020/05, OECD Publishing, Paris, https://doi.org/10.1787/5f65ff7e-en. [13]

Belkin, M. et al. (2019), "Reconciling modern machine-learning practice and the classical bias–variance trade-off", *Proceedings of the National Academy of Sciences*, Vol. 116/32, pp. 15849-15854, https://doi.org/10.1073/pnas.1903070116. [37]

Brynjolfsson, E., T. Mitchell and D. Rock (2018), "What Can Machines Learn and What Does It Mean for Occupations and the Economy?", *AEA Papers and Proceedings*, Vol. 108, pp. 43-47, https://doi.org/10.1257/pandp.20181019. [26]

Cammeraat, E. and M. Squicciarini (2021), "Burning Glass Technologies' data use in policy-relevant analysis: An occupation-level assessment", *OECD Science, Technology and Industry Working Papers*, No. 2021/05, OECD Publishing, Paris, https://doi.org/10.1787/cd75c3e7-en. [18]

Carnevale, A., T. Jayasundera and D. Repnikov (2014), *Understanding Online Job Ads Data: A Technical Report*, Georgetown University, Center on Education and the Workforce, https://cew.georgetown.edu/wp-content/uploads/2014/11/OCLM.Tech_.Web_.pdf. [16]

Chattopadhyay, R. and E. Duflo (2004), “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India”, *Econometrica*, Vol. 72/5, pp. 1409-1443, https://doi.org/10.1111/j.1468-0262.2004.00539.x. [30]

Deming, D. and L. Kahn (2018), “Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals”, *Journal of Labor Economics*, Vol. 36/S1, pp. S337-S369, https://doi.org/10.1086/694106. [15]

Felten, E., M. Raj and R. Seamans (2018), “A Method to Link Advances in Artificial Intelligence to Occupational Abilities”, *AEA Papers and Proceedings*, Vol. 108, pp. 54-57, https://doi.org/10.1257/pandp.20181021. [25]

Fu, S. (ed.) (2021), “Skill-driven recommendations for job transition pathways”, *PLOS ONE*, Vol. 16/8, p. e0254722, https://doi.org/10.1371/journal.pone.0254722. [22]

Gehlhaus, D. and S. Mutis (2020), *The U.S. AI Workforce: Understanding the Supply of AI Talent*, Center for Security and Emerging Technology, https://doi.org/10.51593/20200068. [8]

Georgieff, A. and R. Hyee (2021), “Artificial intelligence and employment : New cross-country evidence”*, OECD Social, Employment and Migration Working Papers*, No. 265, OECD Publishing, Paris, https://doi.org/10.1787/c2c1d276-en. [32]

Hamermesh, D. (1993), *Labor Demand*, Princeton University Press, Princeton, NJ. [33]

Hand, D. (2006), “Classifier Technology and the Illusion of Progress”, *Statistical Science*, Vol. 21/1, https://doi.org/10.1214/088342306000000060. [24]

Hand, D. and K. Yu (2001), “Idiot’s Bayes?Not So Stupid After All?”, *International Statistical Review*, Vol. 69/3, pp. 385-398, https://doi.org/10.1111/j.1751-5823.2001.tb00465.x. [23]

Hart, R. (2017), “Hours vs employment in response to demand shocks”, *IZA World of Labor*, https://doi.org/10.15185/izawol.393. [21]

Hastie, T., R. Tibshirani and J. Friedman (2009), *The Elements of Statistical Learning*, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-84858-7. [36]

Hershbein, B. and L. Kahn (2018), “Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings”, *American Economic Review*, Vol. 108/7, pp. 1737-1772, https://doi.org/10.1257/aer.20161570. [17]

Lane, M. and A. Saint-Martin (2021), “The impact of Artificial Intelligence on the labour market: What do we know so far?”*, OECD Social, Employment and Migration Working Papers*, No. 256, OECD Publishing, Paris, https://doi.org/10.1787/7c895724-en. [7]

Lassébie, J. and G. Quintini (2022), “What skills and abilities can automation technologies replicate and what does it mean for workers? New Evidence”*, OECD Social, Employment and Migration Working Papers*, OECD, Paris. [42]

Linkedin (2019), *AI Talent in the European Labour Market*, Linkedin Economic Graph, https://economicgraph.linkedin.com/research/ai-talent-in-the-eu. [9]

Manca, F. (forthcoming), "Six questions about the demand for artificial intelligence skills in labour markets", *OECD Social, Employment and Migration Working Papers*, OECD, Paris. [39]

Modestino, A., D. Shoag and J. Ballance (2020), "Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?", *The Review of Economics and Statistics*, Vol. 102/4, pp. 793-805, https://doi.org/10.1162/rest_a_00835. [35]

Modestino, A., D. Shoag and J. Ballance (2016), "Downskilling: changes in employer skill requirements over the business cycle", *Labour Economics*, Vol. 41, pp. 333-347, https://doi.org/10.1016/j.labeco.2016.05.010. [14]

Moretti, E. (2021), "The Effect of High-Tech Clusters on the Productivity of Top Inventors", *American Economic Review*, Vol. 111/10, pp. 3328-3375, https://doi.org/10.1257/aer.20191277. [4]

OECD (2022), *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449. [28]

OECD (2021), "Navigating skill demands in turbulent times", in *OECD Skills Outlook 2021: Learning for Life*, OECD Publishing, Paris, https://doi.org/10.1787/01a857c7-en. [34]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en. [1]

OECD (2017), "Skill needs and mismatch indicators: Methodology", in *Getting Skills Right: Skills for Jobs Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/9789264277878-4-en. [29]

OECD.AI (2022), "Visualisations powered by JSI using data from LinkedIn", OECD AI Observatory, Paris, http://www.oecd.ai/. [38]

Page, T. (2022), "Can an AI-powered insect trap solve a $220 billion pest problem?", *CNN Business*, https://edition.cnn.com/2022/11/24/business/trapview-ai-pest-management-spc-intl/index.html?utm_source=substack&utm_medium=email (accessed on 30 November 2022). [41]

Salvi del Pero, A., P. Wyckoff and A. Vourc'h (2022), "Using Artificial Intelligence in the workplace: What are the main ethical risks?", *OECD Social, Employment and Migration Working Papers*, No. 273, OECD Publishing, Paris, https://doi.org/10.1787/840a2d9f-en. [6]

Samek, L., M. Squicciarini and E. Cammeraat (2021), "The human capital behind AI: Jobs and skills demand from online job postings", *OECD Science, Technology and Industry Policy Papers*, No. 120, OECD Publishing, Paris, https://doi.org/10.1787/2e278150-en. [40]

Squicciarini, M. and H. Nachtigall (2021), "Demand for AI skills in jobs: Evidence from online job postings", *OECD Science, Technology and Industry Working Papers*, No. 2021/03, OECD Publishing, Paris, https://doi.org/10.1787/3ed32d94-en. [12]

Tolan, S. et al. (2021), "Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks", *Journal of Artificial Intelligence Research*, Vol. 71, pp. 191-236, https://doi.org/10.1613/jair.1.12647. [43]

Van Roy, V. et al. (2021), *AI Watch - National strategies on Artificial Intelligence: A European perspective, 2021 edition*, https://doi.org/10.2760/069178 (online). [27]

Annex A. Additional Figures and Tables

- Table A.1 lists various methodologies to classify vacancies as those demanding AI skills or similarity measure for skills and occupations "close" to the skill artificial intelligence. This list does not claim to be a comprehensive review of the overall literature on the topic. It is rather a selection of approaches – primarily from labour economics -- that have to some degree informed methodological choices made in this paper.

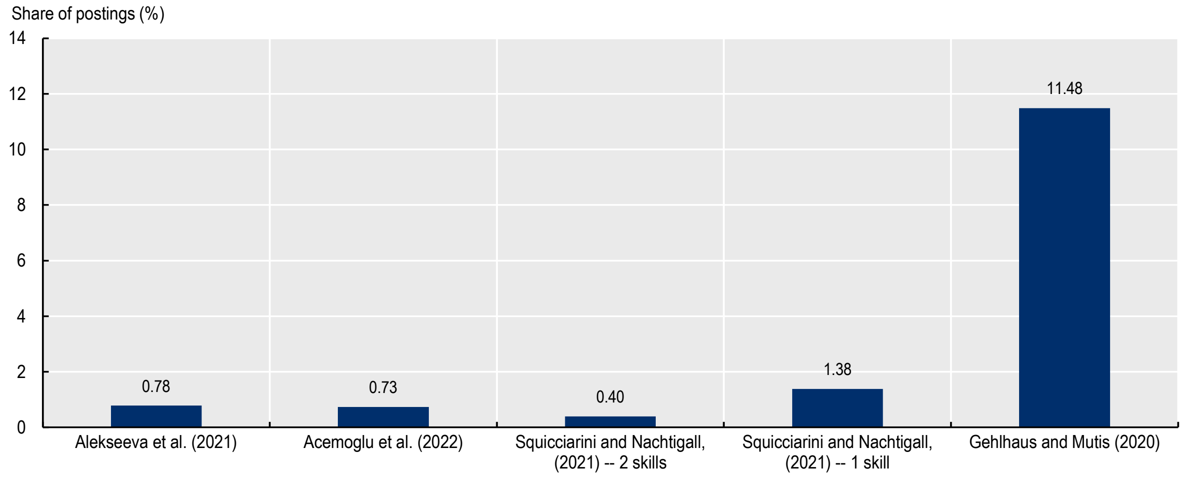
Table A.1. Methods for identifying AI–related jobs and jobs with AI exposure from the literature

Publication	Identification Aim	Data Sources Used	Methodology	Occupation Classification
Acemoglu et al., (2022[11])	AI-related jobs & exposure	Lightcast List of AI-related skills compiled by the authors	The authors create a list of AI-related skills (footnote 14) and apply those keywords to the Lightcast data in order to identify job postings that seek or require AI-related skills. This allows to construct the share of AI-related job postings that is posted by every firm. It can also be used to identify the share of AI-related job postings by occupations as 95% of job-vacancies can be assigned to six-digit (SOC) occupation codes within the US data.	O*NET SOC classification (6-digit)
Alekseeva et al. (2021[10])	AI-related jobs	Lightcast	The authors use a list of keywords that includes AI-related skills and software/packages. This is derived from the Lightcast skill taxonomy. They apply this keyword list to the Lightcast data to analyze job postings that require these AI-related skills. An "AI vacancy" is defined by them as a vacancy that requires at least one AI-related skill from this list.	O*NET SOC classification (2-digit & 6-digit)
Dawson, Rizoiu and Williams (2021[22])	AI-related skills	Lightcast Quarterly Detailed Labor Force (Australian Bureau of Statistics) Household, Income and Labour Dynamics in Australia (HILDA)	The authors construct a measure for skill similarity between different skill sets. It is based on the importance of a skill for a particular job posting. This is measured using its frequency in other job postings and the amount of other skills required in the posting. They then measure the distance between two skill sets that depends on the similarity of their most important skills. Based on this the authors construct a list of AI-related keywords from Lightcast data. The list is constructed by analyzing which are the most important skills for a list of core AI "seekd skills" such as "Aritificial Intelligence, Machine Learning, Data Science etc." This allows the list to be dynmaic and to be updated every year. The top 100 skills are selected for every year that result from this process. The annual AI skill list can then be compared to the skill list of different industries and occupations. Using the skill distance measure the authors show to what extent industries have adpoted AI skills .	Australian and New Zealand Standard Classification of Occupations (ANZSCO) 6-digit

Gelhaus and Mutis (2020[8])	AI-related jobs	Lightcast Occupational Information Dataset: O*NET U.S. Census Bureau	The authors create a conceptual framework of how AI teams work together and how AI applications are being developed. They use this in order to review the results from the following skill and occupation matching steps: They identify a set of AI-related keywords that are provided by Lightcast and listed in Toney & Flagg (2020). They then scan the O*NET database with detailed occupation titles, work activities and tasks to analyze matches between the two. A manual crosschecking with the consistency with their own framework as well as additional checks with job titles from U.S. Census data yields the final results.	O*NET SOC classification & crosswalk to U.S. Census codes
Squicciarini and Nachtigall (2021[12])	AI-related jobs	Lightcast	The authors use a set of AI-related keywords provided by Barufafaldi et al. (2020[13]) and a list of AI-related software and repositiories validated by experts at the UK Department for Business, Energy and Industrial Strategy (BEIS). They apply these keywords and software names to the Lightcast dataset of online job postings. AI-related jobs are then identified as those postings that contain at least two AI-related skills belonging to different concepts or methodologies. Only one of those keywords may be a software-related skill.	ONET Codes for Canada; SOC/SSOC for UK, USA & Singapore Crosswalk to ISCO08 was used for Canada, UK & USA

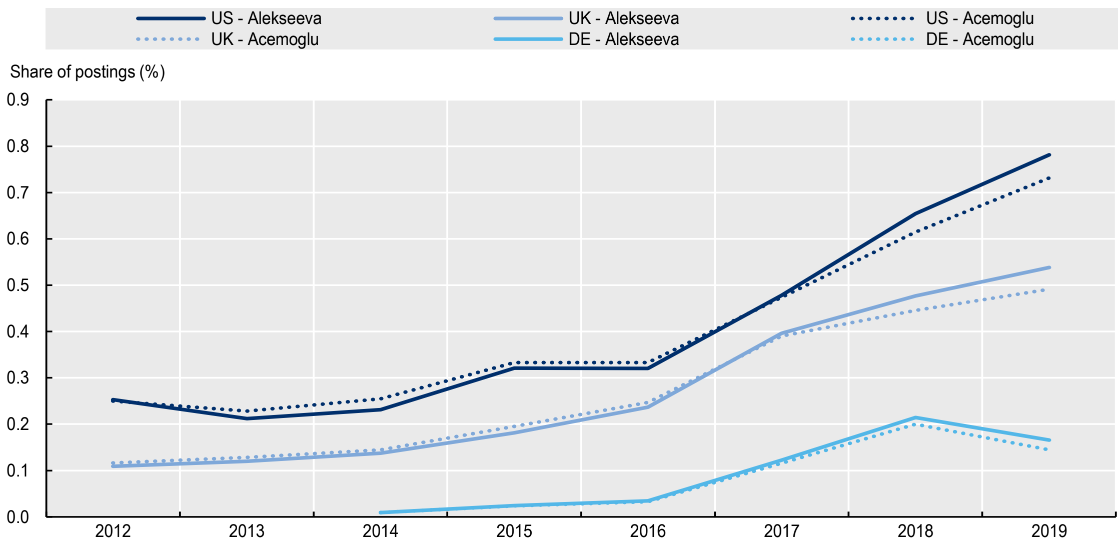
Figure A.1. Share of AI job postings according to different methodologies

Share of job postings classified as demanding AI skills in the United States, 2019



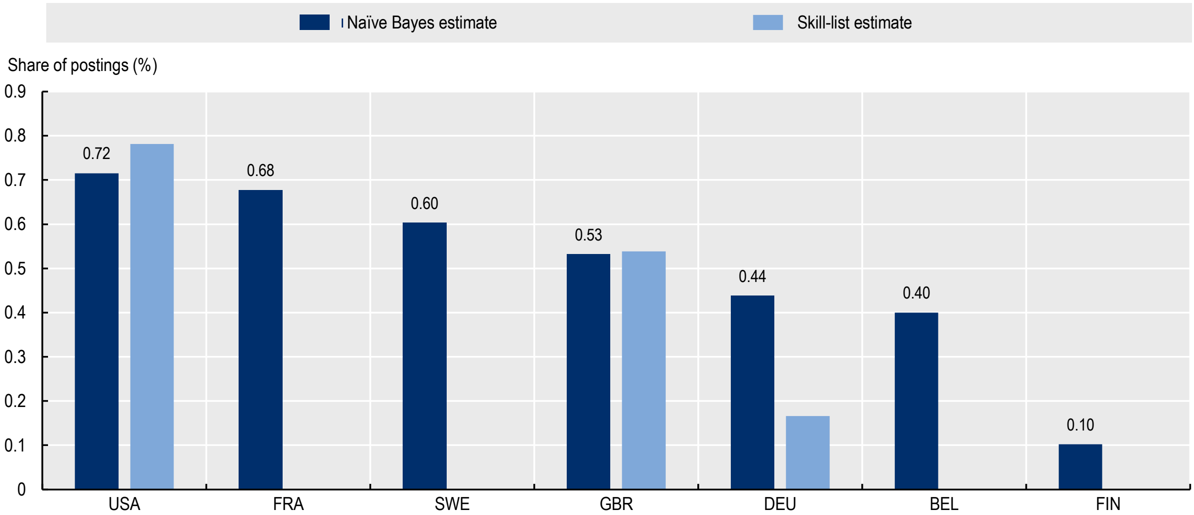
Notes: Replication results based on the methodologies as outlined in the various papers on data from 2019. Alekseeva et al. (2021[10]), Acemoglu et al., (2022[11]) and Squicciarini and Nachtigall, (2021[12]) (SN) use lists of skills to identify vacancies demanding AI skills. Replication of SN with two skills only requires two skills from their list, and not that each additionally comes from at least two separate sub-categories. Estimates are obtained by computing the share of vacancies demanding AI skills implied by each skill list. Gehlhaus and Mutis, (2020[8]) use the occupations which comprise the AI workforce. Estimates for this methodology are the share of vacancies in these pre-defined occupations.
Source: OECD analysis of Lightcast, European Labour Force Survey (EU-LFS) & and Current Population Survey (CPS) data.

Figure A.2. Job postings classified as demanding AI skills in USA, GBR & DEU according to AI skill list, 2012 - 2019



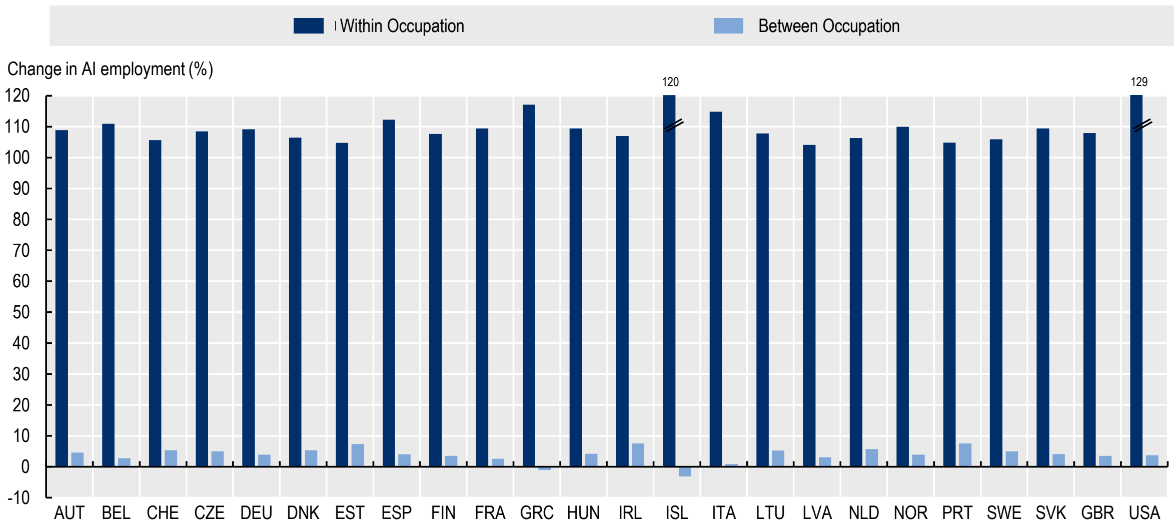
Note: AI skill lists from Alekseeva et al. (2021[10]) and Acemoglu et al., (2022[11]) applied to Lightcast data from the United Kingdom, Germany and the United States.
Source: OECD Analysis of Lightcast data.

Figure A.3. Share of job postings classified as demanding AI skills using a Naive Bayes classifier, 2019



Notes: : "Skill-list estimate" identifies AI job postings using a list of AI skills from Alekseeva et al. (2021[10]). "Naive Bayes estimate" uses a naïve bayes classifier trained on vacancies which contain the skill "Machine Learning", which uses skills associated with each job posting as features. The Naive Bayes classifier uses a Bernoulli likelihood and Laplace prior. The number of skills used for classification are selected using average conditional entropy. Vacancies are classified as demanding AI skills if the odds ratio for the class compared to "not AI" is at least 9. The final share of vacancies classified as demanding AI skills include those classified from the naïve bayes estimate as well as vacancies containing the skill "Machine Learning".
Source: OECD analysis of Lightcast data.

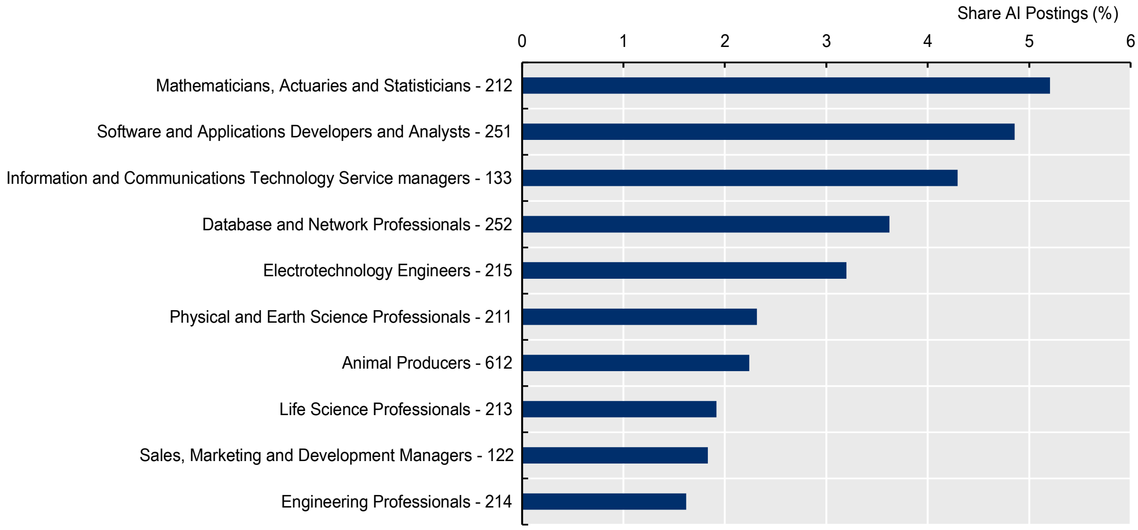
Figure A.4. Decomposition of the change in AI employment shares, 2011 and 2019



Notes: Occupational AI talent shares are identified according to AI skills lists as in Alekseeva et al. (2021). The change over time of the overall AI share in a country is decomposed using a shift-share into the change of AI postings within occupations and the change in the occupational
Source: OECD analysis of Lightcast, European Union Labour Force Survey (EU-LFS) & Current Population Survey (CPS) data.

Figure A.5. Top 10 AI intensive occupations

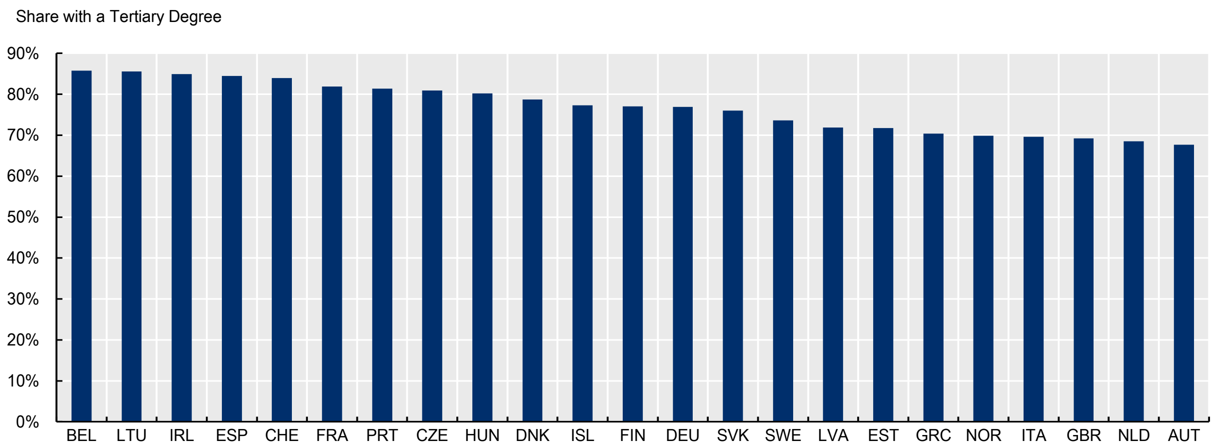
Share of job postings demanding AI skills by occupation in the United States, 2019



Notes: The results reflect the average share of AI vacancies by occupation across all countries in the sample. The occupations use ISCO-08 at the 3-digit level. An employment weighted SOC - ISCO crosswalk has been used to transform the results for the United States. Occupational AI talent shares are identified according to the keyword list used in Alekseeva (2020).
Source: Lightcast data.

Figure A.6. Top 10 AI intensive occupations predominantly have a tertiary degree

Share of the Top 10 most AI intensive occupations with a tertiary degree in 2019



Note: The ten most AI-intensive occupations are derived from the application of the Alekseeva et al. (2021) keyword method to the US data of Lightcast. They are (in descending order): Mathematicians, Actuaries and Statisticians - 212; Software and Applications Developers and Analysts - 251; Information and Communications Technology Service managers - 133; Database and Network Professionals - 252; Electrotechnology Engineers - 215; Physical and Earth Science Professionals - 211; Animal Producers - 612; Life Science Professionals - 213; Sales, Marketing and Development Managers - 122; and Engineering Professionals - 214. An employment weighted SOC - ISCO crosswalk has been used to transform the results for the United States.
Source: OECD analysis of the European Union Labour Force Survey (EU-LFS) and the Current Population Survey (CPS).

Annex B. Estimates of the share of AI vacancies using a Naïve Bayes classifier

61. This annex explains the Naïve Bayes (NB) classifier applied to Lightcast job vacancy data to classify vacancies demanding AI skills. The results of this analysis were previously presented in Figure A.3. The Naïve Bayes classifier does not require a pre-specified list of AI skills, which is the main criticism of the list-based approach to classification used in this paper, and other papers using Lightcast (Table A.1).[33] This Annex introduces the NB classifier and shows how it can be employed in a variety of applications. It shows that the NB classifier allows for credible classification of vacancies demanding AI skills in Anglophone countries, and reasonable results in the subset of non-Anglophone countries where Lightcast data was previously shown to be representative of employment. In addition, this annex argues that the list-based approach to classification can be viewed as a special case of the Naïve Bayes estimator; the NB classifier is simply a generalisation of the list-based classifier obtained by relaxing some key, but strong, assumptions. The section concludes by presenting the practical details of implementing the NB classifier for classifying vacancies demanding AI skills with the results providing further evidence for using the common within-occupation intensity measure from the United States.

The Naive Bayes classifier applied to Lightcast data using skills as features

62. Before introducing the NB estimator, it is important to define the Lightcast data, and the general problem, in a parsimonious manner. The Lightcast database consists of a set of job vacancies, V, with each individual vacancy denoted $v \in V$. For each vacancy in the database, there exists a set of skills demanded in the job posting, $s \in S_v$, where the total set of skills in the database is simply the union of all sets of skills demanded in each vacancy in the entire database, $S \equiv \cup_{v \in V} S_v$.[34] The goal of the NB estimator, but also the list-based classifier, is to partition the set of vacancies into mutually exclusive and collectively exhaustive sets, or classes, $C \in \{A, \neg A\}$. The class A contains vacancies demanding AI skills, and $\neg A$, contain the job postings which do not require AI skills.[35] It is also useful to define a pre-specified set of AI skills, $L \subset S$, which for a list-based classifier, classifies a vacancy as demanding AI skills if the vacancy demands a skill on the list.[36]

63. The set-up focuses on skills, and specifically using skills to classify vacancies as those demanding AI skills, but it generalises to other applications using Lightcast data. One can simply reimagine the classification set, A, as "Green Jobs", "Teacher" skills or "Tax Preparer" skills, for example, and the set of pre-specified skills, L, would correspond to each of those groupings. The rest of this annex will continue

[33] Babina et al., (2020[20]) employs a similar approach to the NB classifier described here. They use an ad-hoc decision rule to classify AI vacancies which amounts to taking the average conditional probability of all skills listed in a vacancy. This is quite similar to, but not exactly a naïve bayes classifier which takes logs of the likelihood function.

[34] There are a small set of vacancies for which no skills are demanded, or they are missing. These cases are not explicitly modelled, and are classified below as not demanding AI skills.

[35] Formally, the sets obey the properties: $A \cap \neg A = \emptyset \; and \; A \cup \neg A = V$.

[36] The classification rule is formally defined: $\forall\, v \in V,\; S_v \cap L \neq \emptyset \; \rightarrow v \; \in A$, else, $S_v \cap L = \emptyset \; \rightarrow v \in \neg A$

with the more concrete case of AI skills, but the set-up easily applies to a range of applications that use skills to classify vacancies in Lightcast data.

64. The NB classifier uses Bayes' law and the skills associated with a vacancy to estimate the probabilities that a given vacancy demands AI skills, or not, conditional on observed skills demanded in the vacancy. The definition is:

$$P(C|S_v) = \frac{P(S_v \mid C)P(C)}{P(S_v)} = \frac{P(C)\prod_{s\in S_v} P(s|C)}{\prod_{s\in S_v} P(s)} \qquad (1B)$$

The equation is computed for both classes, $\{A, \neg A\}$.[37] The first two terms on either side of the first equality are simply Bayes' law.[38] The last term, after the second equality, makes the simplifying – naïve – assumption that all skills demanded in a vacancy enter the likelihood function independently, and they can therefore be multiplied together without modelling a joint probability distribution. This is the defining characteristic of the NB classifier. It is almost certainly wrong; the skills demanded for any given vacancy are usually highly correlated. The NB classifier assumes, however, that the skills demanded in a vacancy are independent draws from S.

65. The NB classifier classifies vacancies into the class that maximizes the probability from equation (1B):

$$\underset{C\in\{A,\neg A\}}{\text{argmax}} \quad P(C \mid S_v) \propto P(C)\prod_{s\in S_v} P(s|C) \qquad (2B)$$

The objective is to allocate a vacancy to the class which produces the highest probability (according to equation 2B) that the vacancy belongs to that class conditional on the set of skills demanded. The denominator from equation (1B) is omitted because the data is fixed and does not vary across classes. The resulting probabilities are therefore proportional up to an integrating constant.

66. The assumption of independence of skills demanded in a vacancy is strong, and creates some weaknesses for the NB estimator. First and foremost, it induces bias in the estimator: for any fixed realisation of the data, the NB classifier is unlikely to be the best at classifying vacancies compared to the set of feasible classifiers.[39] In addition, and likely due to the simplifying assumption of independence, the predicted probabilities produced by the NB classifier are unlikely to be the best possible estimates. If one wishes to use the predicted probabilities beyond classification, the NB classifier may not be the best choice.

67. Despite its weaknesses, the NB estimator is widely used in a variety of commercial and academic applications. Researchers in both academia and industry have studied the performance of the NB estimator, and it has found success – including compared to much more sophisticated methods – in medicine, fraud detection and email spam filters, among others (Hand and Yu, 2001[23]).[40] The success of the NB classifier is likely due to its low variance. For any fixed realisation of the data, the NB classifier may not produce the best fit of the data, but for future draws of the data it will approximately maintain its performance where other estimators will perform poorly and likely require a reworking of the model. In

[37] One is not limited to only two classes as in this application to AI. One can specify more classes depending on the problem at hand.

[38] Despite the name, the NB classifier is not a Bayesian method of inference, rather it is an application of Bayes law.

[39] Where "best" is defined as minimising some loss function such as a measure of the classification error, and the set of feasible classifiers would be any supervised learning models as found in Hastie, Tibshirani and Friedman, (2009[36]), for example.

[40] The model is still clearly in use in current commercial applications. Casual exploration of recent Lightcast job postings that demand the skills "artificial intelligence" or "machine learning" will sometimes also demand the skill "naïve bayes".

other words, it trades some bias for low variance. In many classification applications, simple models produce most of the classification power and more sophisticated models deliver marginal performance gains but at the cost of much greater complexity, computational resources or model variance (Hand, 2006[24]).[41]

68. The advantages of the NB estimator are particularly suited to the needs of large organisations, such as the OECD, which are not primarily, or even usually, engaged in classification problems using large datasets. The low variance of the NB estimator means that one can easily use the model on data from a variety of countries with minimal resource costs for retuning. Indeed, one feature of the model is that it is fast (amounting to taking sums – see below) and requires fitting only one or two parameters. In the present application, one can also discard the use of pre-specified lists of skills and train the model on only one or a few skills which are ubiquitous across countries (perhaps with some minimal translation).

69. The NB estimator is also a natural choice for relaxing the assumption of a pre-specified list of skills. The next section shows that list-based approaches can be viewed as a special case of the NB estimator. When one relaxes the need for a pre-specified list within this framework, one continues to use the same estimator, but rather than assuming the associated probabilities and producing a deterministic decision rule, one estimates the probabilities from the data.

The Naive Bayes classifier is a generalisation of list-based approaches to classification

70. First, note that list-based approaches make the same impendence assumption on features as the NB estimator. The NB estimator, therefore, shares a key theoretical link and modelling assumption with list-based approaches.

71. List-based approaches to classification can be viewed as an NB classifier where one assumes the probabilities rather than estimating them from the data. In classic applications of the NB estimator, $P(s|C)$ and $P(C)$, respectively, would be estimated from microdata data or pulled from official published sources. For example, one might estimate the $P(s|C)$ with the share of vacancies in class *C* that demand skill *s.* The assumption that moves one from the NB estimator to list-based approaches is to assume these probabilities rather than estimate them. The most natural justification for this assumption is that one estimates the associated probabilities, $P(s|C)$ for example, using a Bayesian probability model with a "strong" prior on the individual probabilities. This strong prior could take the form of a degenerate prior with zero variance or infinite precision. Such a prior would be improper, and one would never use such a prior for statistical inference -- a degenerate prior would imply there is nothing to estimate rendering the problem deterministic rather than probabilistic.[42] However, list-based approaches make implicit assumptions on the associated probabilities that produce a deterministic decision rule. The assumption of a strong prior and

[41] Recent advances in deep learning may be breaking this link delivering greater performance even as model complexity and parameters grow (Belkin et al., 2019[37]).

[42] One need not assume a degenerate or improper prior. If the past evidence or data underlying the prior is sufficiently large relative to the new realisation of the data, new data will change the posterior only marginally, and one could reasonably rationalise simply using the prior estimates. For example, if one continues to assume skills follow a Bernoulli distribution and ones imposes a Beta prior (the conjugate prior for Bernoulli), the resulting posterior estimate of the mean will be a Beta distribution with the mean a weighted average the maximum likelihood estimate and the prior, with the weights determined by the relative sample sizes of data, and past realisations of data captured by the prior. In the limit, as the sample size of the prior tends to infinity, the posterior will converge to the mean of the prior. In such cases where the accumulated prior information is so large (for example, the measure of the velocity of gravity), one can reasonably take the prior as a constant without the need for new estimates.

reducing the problem to a deterministic decision rule, therefore, is a necessary assumption to produce the special case of the list-based approach.

72. Once one imposes the assumption of assumed probabilities on the NB estimator, the final step is specifying the probabilities. One advantage of viewing a list-based classifier as a special case of the NB classifier is that it imposes structure on the implicit assumptions one makes. One cannot pick probabilities at random without regard for how they relate to each other and the problem at hand. Even if one is content with a list-based approach to classification, the following assumptions on the probabilities should give a deeper understanding of the assumptions one makes when one uses a pre-specified list of skills.

73. Starting with an NB classifier, one needs to impose the following assumptions to recover a list-based classification rule as used in the literature. First, one needs to assume the probabilities in the NB classifier (one's beliefs) rather than estimate them. Next, one needs to provide a pre-specified list of skills indicating artificial intelligence (L). Finally, one needs to make the following assumptions[43] on the assumed probabilities[44]:

1. $\forall s \notin L,\ P^*(s|A) = P^*(s|\neg A)$
2. $P^*(\neg A) > P^*(A)$
3. $\forall s \in L,\ P^*(s|A) > P^*(s|\neg A)$ with $\frac{P^*(s|A)}{P^*(s|\neg A)} > \frac{P^*(\neg A)}{P^*(A)}$

74. The first assumption imposes equality of probability between the two classes for skills that do not appear on the list. This assumption intuitively says that skills that do not appear on the list provide no information for determining the class of a vacancy. This assumption makes explicit the fact that with list-based classifiers, skills that do not appear on the list are not considered, and provide no information for distinguishing the class.

75. The second assumption says that the prior probability that a vacancy is classified as not demanding AI skills is more likely than vacancies that do require AI skills. In other words, if one were to draw a vacancy at random, and without seeing any other information, one would assume the vacancy did not demand AI skills. This is quite reasonable and intuitive, although it often goes unstated with list-based classifiers. The first two assumptions together with equation (1B) provide half of the list-based classifier: vacancies with no skills on the list are classified as not requiring AI skills.

76. The final assumption is that skills on the list are more likely to appear in vacancies demanding AI skills. This is straightforward and intuitive. Unfortunately, it is not by itself sufficient, and one needs to put a minimum distance between the probabilities that a skill appears in a class for each skill. In assumption three, that is represented by the odds ratio, which needs to be larger than the odds ratio of the priors. The intuitive explanation for this minimum odds ratio is the following: skills on the list provide information that one should classify a vacancy demanding this skill as demanding AI skills in general. However, the assumption on the priors (assumption 2) works in the opposite direction. The skills on the list therefore need to have sufficient information -- as represented by their odds ratio or relative probabilities -- to move one off the prior, and classify the vacancy as demanding AI skills.

77. Assumption 3 represents the minimum information necessary for a list-based classification rule to require only one skill, but it generalises to rules that require more than one skill on the list to classify a vacancy as demanding AI skills. Squicciarini and Nachtigall, (2021[12]), for example, require two skills from a pre-specified list to classify a vacancy as demanding AI skills. Within this framework, and for a given

---

43 These are technically a set of sufficient conditions to produce a list-based classifier from a naïve bayes framework. Depending on the problem at hand, they may not be necessary.

44 The notation P^* is adopted to distinguish probabilities that are assumed rather than estimated. A practical consideration when implementing a naïve Bayes classifier is that all probabilities are strictly positive. This assumption is maintained in this more theoretical discussion as well.

odds ratio on the prior, the authors are implicitly assuming that the odds ratio of the probabilities of skills on the list is less than that of the odds ratio of the priors as stated in assumptions two and three. However, they are still implicitly imposing a minimum bound on the odds ratio for skills on the list such that they require only two skills and not three, four or more.[45]

78. In summary, starting with a naïve Bayes classifier, if one assumes the associated probabilities rather than estimating them, with the probabilities dependent upon whether or not skills appear on a pre-specified list, one can recover the list-based classifier used in the literature. In other words, these list-based classifiers can be viewed as a special case of the naïve Bayes classifier. The final section shows that by keeping the naïve Bayes classifier, but estimating the probabilities directly from the Lightcast data rather than assuming them, one can relax the need for a pre-specified list of skills while keeping similar estimates of the share of vacancies demanding AI skills.

The naïve Bayes estimator provides similar estimates to list-based classifiers without the need for a list of AI skills

79. In order to relax the need of a pre-specified list of skills, one must designate a training set to fit the conditional probabilities for each skill. One needs to estimate the conditional probabilities $P(s|A)$ and $P(s|\neg A)$ for every skill. This assumes that one already knows – at least for a subset of the data –the classes of some vacancies. The classic way to do this to have one or a set of researchers take random samples of the data and review vacancies by hand classifying them as either demanding AI skills or not depending on the skills demanded in a vacancy and potentially other information included in the database. This is the preferred method, but it is resource intensive, and becomes restrictive for more than a few countries.

80. A less resource intensive training set is available if one is willing to assume that one or a few skills signal a class. In the present example, if a vacancy demands the skill “Machine Learning”, one can be reasonably sure that that vacancy is demanding AI skills. Machine learning represents the current wave of AI development, and it shows up with great frequency across countries in the Lightcast data. One can therefore define a training set as a sample – one for each country -- with all vacancies claiming to demand the skill “Machine Learning” the set demanding AI skills (A), with the rest of the vacancies making the set not demanding AI skills ($\neg A$).

81. From these training data, one then needs to estimate the probability that a vacancy should be classified in each class. To model the probability, this report assumes that skills that appear in vacancies follow a Bernoulli distribution. This is appropriate as skills demanded in each vacancy are Boolean: they only appear as demanded or not, and there is no associated count or intensity measure for each skill demanded. The model is:

$$P(C \mid S_v) \propto P(C)\prod_{d=1}^{D}\Big(\mathbf{1}(d \in S_v)P(d|C) + \big(1 - \mathbf{1}(d \in S_v)\big)\big(1 - P(d|C)\big)\Big) \qquad (3B)$$

There are two notable deviations in this model from the more general set-up in equation 2B. First, the Bernoulli model explicitly models not just the presence of a skill demanded in a vacancy but its absence as well. Second, the set of skills under consideration is not every skill demanded in a vacancy, or all skills in the database (*S*), but a subset of skills $D \in S$. There are over 10,000 skills in the database for the United States alone, and modelling the presence or absence of each one would be both unwieldly and inefficient:

[45] For a given odds ratio on the priors, $\frac{P^*(\neg A)}{P^*(A)}$, and an assumption on the odds ratio of skills on the list, $\frac{P^*(s|A)}{P^*(s|\neg A)} > 1$, with $\frac{P^*(s|A)}{P^*(s|\neg A)} <, \frac{P^*(\neg A)}{P^*(A)}$, which therefore necessitates more than one skill, the number of skills required, *n>1*, is determined by the set-valued sequence on the real line such that $\frac{P^*(s|A)}{P^*(s|\neg A)} \in \left[\left(\frac{P^*(\neg A)}{P^*(A)}\right)^{\frac{1}{n}}, \left(\frac{P^*(\neg A)}{P^*(A)}\right)^{\frac{1}{n-1}}\right]$.

the vast majority skills show up very infrequently, and provide little to no information for deciding how to classify a vacancy as demanding AI skills or not.

82. This report uses average mutual information to rank skills based on their importance for classifying vacancies as demanding AI skills and select the subset of skills, *D*. This is done by calculating the entropy of a class minus the entropy conditional on the absence or presence of a skill averaged across classes. In other words, the ranking asks how much information one would lose if a skill was removed the data representing each class, averaged across classes. Average mutual information is given by:

$$I(C,s) = \sum_{C \in \{A, \neg A\}} P(C,s) \log\left(\frac{P(C,s)}{P(C)P(s)}\right) \qquad (4B)$$

The probability $P(C)$ is the share of vacancies in the class divided by total vacancies, $P(s)$ is the share of vacancies demanding a particular skill divided by the total number of vacancies, and $P(C,s)$is the total number of vacancies *both* in the class and demanding a particular skill divided by the total number of vacancies. In practice, average mutual information ranks skills by how frequently they appear in both classes with the highest ranking going to those skills that appear frequently in each class. These are not skills that are most likely to signal demand for AI skills, rather they are skills that provide the most information for deciding in which class to classify a vacancy.

83. The final step is to estimate the probabilities and tune the model to pick the optimal number of skills to use, *D.* Estimation of the associated probabilities ($P(C), P(d|C)\ \forall\ d \in D$ in both classes) proceeds by simply taking the share of vacancies that correspond to each probability separately for each country. This is equivalent to the maximum likelihood estimate of a Bernoulli random variable.[46]

84. The final set of skills to use in the analysis, *D,* is a data-driven choice estimated on a training set by estimating the probabilities, classifying vacancies, and then finding the *D* that maximises the F1 score. The F1 score is the harmonic mean of precision, which is the share of correctly classified vacancies divided by vacancies classified as demanding AI, and recall, which is the share of correctly classified vacancies divided by all vacancies that should have been classified as demanding AI skills. There is an inherent tension between precision and recall, and the F1 score finds the set of skills that maximises the combination of the two. For all countries in the sample, the F1 score is maximised with less than eight skills which vary by country.

85. The results for a selection of countries are shown in Figure A.3, which provides three relevant points for this report. First, the naïve Bayes classifier provides evidence that list-based approaches to classifying AI skills do not perform as well in non-Anglophone countries. The NB classifies vacancies in Anglophone countries as demanding AI skills at similar rates compared to classification using a pre-specified list of skills. This is true for the United States and the United Kingdom where classification rates are similar to those when one uses the list of skills form *Alekseeva*. For Germany, the naïve Bayes classifier estimates a higher share of vacancies demanding AI skills suggesting a discord between the skills on lists derived from Anglophone countries and the skills correlated with “Machine Learning” in Germany, and possibly other non-Anglophone countries.

86. Second, the naïve Bayes classifier confirms that a higher share of vacancies demand AI skills in the United States than in other countries using Lightcast data. The next highest countries demanding AI skills in the subset of countries analysed are Sweden and France with over 0.6% of vacancies demanding AI skills compared to over 0.7% in the United States using either the naïve Bayes classifier or the list of

[46] The estimates are augmented with a Laplace prior, which is another way of saying that a pseudo-count of one is added to each of the probability estimates to avoid estimates of zero probability. From a practical perspective, estimates of zero probability in a naïve Bayes classifier severely bias the corresponding classification. Using average mutual information for feature selection greatly reduces the need for the pseudo-count, however, and the results in this report are not sensitive to the use of a pseudo-count.

skills from *Alekseeva*. This provides further evidence that the size of the AI workforce estimated in this report is likely an upper bound.

87. Finally, the small share of vacancies demanding AI skills in Finland provides some evidence that not all countries with Lightcast data are representative of the employment or hiring distribution in each country. Finland has highest share of employment in the 10 occupations most intensely demanding AI skills (Figure 2.3). It is possible that Finland has simply missed the boat on AI, and despite its high share of employment in occupations demanding AI skills, very few people in these occupations have AI skills. An alternative explanation is that the Lightcast data is not representative of employment in Finland, and misses many vacancies demanding AI skills. In their validation of European Lightcast data, Araki et al., (2022[19]) find evidence for the latter interpretation. This is not the final word on this question, but it seems likely that using a within-occupation intensity measure for Finland would not improve the final estimates of the AI workforce in Finland.