Organisation for Economic Co-operation and Development

DELSA/ELSA/WD/SEM(2022)7

Unclassified English text only

5 July 2022

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

Cancels & replaces the same document of 4 July 2022

Using Artificial Intelligence in the workplace: What are the main ethical risks?

OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS No. 273

JEL classification: J01, J08, J2, J7, O3.

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.

Angelica Salvi Del Pero (angelica.salvidelpero@oecd.org)
Peter Wyckoff (peter.wyckoff@oecd.org)
Ann Vourch
With contributions from: Karine Perset (karine.perset@oecd.org); Laura Galindo (laura.galindo@oecd.org).

JT03498941

OECD Social, Employment and Migration Working Papers

www.oecd.org/els/workingpapers

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to els.contact@oecd.org.

This series is designed to make available to a wider readership selected labour market, social policy and migration studies prepared for use within the OECD. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language – English or French – with a summary in the other.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

[1] Note by the Republic of Türkiye: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. The Republic of Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of the Republic of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus. The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Acknowledgements

This publication 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/work-innovation-productivity-skills and https://denkfabrik-bmas.de/.

Federal Ministry
of Labour and Social Affairs

This report would not have been possible without the input from experts on the ethics of workplace AI, including developers of AI tools, employers, trade unions, and academics. Many thanks to the participants of the two expert group meetings on the ethics of AI in the workplace, as well as the relevant panel discussions at the OECD AI-WIPS Conferences in 2021 and 2022: Jeremias Adams-Prassl (Oxford University), David Barnes (IBM corporation), Victor Bernhardtz (Unionen Sweden), Gabriel Burdin (University of Leeds), Birte Dedden (UNI Global), Valerio De Stefano (KU Leuven), Samuel Engblom (Sweden Ministry of Education and Research), Alex Engler (Brookings Institution), Lorraine Finlay (Australian Human Rights Commission), Joanna Goodey (EU Fundamental Rights Agency) William G Harris (ATP Global), Anke Hassel (Hertie School), Fabio Landini (University of Parma), Pauline Kim (Washington University in Saint Louis), Isaac Look (Malakoff Médéric Humanis), Phoebe Moore (University of Essex), Carolyn Nguyen (Microsoft), Hideaki Ozu (BIAC), Andrew Pakes (Prospect Union), Giles Pavey (Unilever), Katherine Platts (Unilever), Frida Polli (Pymetrics), Aída Ponce Del Castillo (ETUI), Oliver Roethig (UNI Europea), Calli Schroeder (EPIC), Keith Sonderling (US EEOC), William Spriggs (AFL-CIO), Filip Stefanovic (TUAC), Oliver Suchy (DGB), Mary Towers (TUC UK), Christo Wilson (Northeastern University).

Chapter 3 of the report benefitted from contributions from Karine Perset and Laura Galindo, from the Directorates of Science, Technology and Innovation, who also provided inputs and comments to the report overall alongside colleagues from the Directorate for Employment, Labour and Social Affairs (Stefano Scarpetta, Mark Pearson, Stijn Broecke, Sandrine Cazes, Glenda Quintini, Marguerita Lane, Clara Krämer, Morgan Williams and Daniela Jiménez Estrada), and colleagues from the Directorate for Education and Skills (Mila Staneva, Stuart Elliott, and Stephan Vincent-Lancrin).

Abstract

Artificial Intelligence (AI) systems are changing workplaces. AI systems have the potential to improve workplaces, but ensuring trustworthy use of AI in the workplace means addressing the ethical risks it can raise. This paper reviews possible risks in terms of human rights (privacy, fairness, agency and dignity); transparency and explainability; robustness, safety and security; and accountability. The paper also reviews ongoing policy action to promote trustworthy use of AI in the workplace. Existing legislation to ensure ethical workplaces must be enforced effectively, and serve as the foundation for new policy. Economy- and society-wide initiatives on AI, such as the EU AI Act and standard-setting, can also play a role. New workplace-specific measures and collective agreements can help fill remaining gaps.

Abrégé

Les systèmes d'intelligence artificielle (IA) sont en train de changer les lieux de travail. Les systèmes d'IA ont le potentiel d'améliorer les lieux de travail, mais assurer une utilisation fiable de l'IA sur le lieu de travail signifie aborder les risques éthiques qu'elle peut soulever. Cet article passe en revue les risques possibles en termes de droits de l'homme (vie privée, équité, agence et dignité), de transparence et d'explicabilité, de robustesse, de sûreté et de sécurité, et de responsabilité. Il passe également en revue les actions politiques en cours pour promouvoir une utilisation fiable de l'IA sur le lieu de travail. La législation existante visant à garantir des lieux de travail éthiques doit être appliquée efficacement, et servir de base pour la création de nouvelles politiques. Les initiatives à l'échelle de l'économie et de la société en matière d'IA, telles que la loi européenne sur l'IA et l'établissement de normes, peuvent également jouer un rôle. De nouvelles mesures et conventions collectives spécifiques au lieu de travail peuvent contribuer à combler les lacunes restantes.

Übersicht

Systeme der künstlichen Intelligenz (KI) verändern die Arbeitsplätze. KI-Systeme haben das Potenzial, Arbeitsplätze zu verbessern, aber um einen vertrauenswürdigen Einsatz von KI am Arbeitsplatz zu gewährleisten, muss man sich mit den ethischen Risiken auseinandersetzen, die damit verbunden sein können. In diesem Papier werden mögliche Risiken in Bezug auf die Menschenrechte (Privatsphäre, Fairness, Handlungsfähigkeit und Würde), Transparenz und Erklärbarkeit, Robustheit, Sicherheit und Verantwortlichkeit untersucht. Das Papier gibt zudem einen Überblick über laufende politische Maßnahmen zur Förderung des vertrauenswürdigen Einsatzes von KI am Arbeitsplatz. Bestehende Rechtsvorschriften zur Gewährleistung ethischer Arbeitsplätze müssen wirksam durchgesetzt werden und als Grundlage für neue politische Maßnahmen dienen. Wirtschafts- und gesellschaftsweite Initiativen zu KI, wie die KI-Verordnung der EU und die Festlegung von Standards, können ebenfalls eine Rolle spielen. Neue arbeitsplatzspezifische Maßnahmen und Tarifverträge können helfen, verbleibende Lücken zu schließen.

Table of contents

OECD Social, Employment and Migration Working Papers 2

Acknowledgements 3

Abstract 4

Abrégé 5

Übersicht 6

Executive summary 9

Résumé 11

Kurzfassung 14

1. Introduction 17

2. Emerging ethical issues in the use of AI in the workplace 22
- 2.1. Human rights: privacy, fairness, agency and dignity 24
- 2.2. Transparency and explainability 31
- 2.3. Robustness, security and safety 35
- 2.4. Accountability 36

3. Overview of policy measures to ensure the trustworthy use of AI in the workplace 40
- 3.1. Existing legislation being applied to ensure a trustworthy use of AI in the workplace 41
- 3.2. Policy efforts to promote trustworthy use of AI in economies and societies 45
- 3.3. Policies specific to the trustworthy use of AI systems in the workplace 49

4. Conclusion 52

References 53

FIGURES

Figure 1. Employees see risks in the use of AI at work, especially when already using such tools 18
Figure 2. VC Investments in AI have been increasing rapidly 19

Figure 3. The business processes and support services industry was the third largest industry for AI VC funding at the end of 2020 20
Figure 4. Different types of ethical concerns can arise, all the way through the different points of use of AI systems in the workplace 24
Figure 5. Nature Survey highlights significant discomfort over facial recognition technology 26
Figure 6. AI Research increasingly looking at interpretability and explainability of systems 33
Figure 7. Potential ethical concerns about AI use in the workplace are interlinked 39
Figure 8. Spectrum of approaches to promote the trustworthy use of AI in the workplace 41

BOXES

Box 1. OECD Principles for responsible stewardship of trustworthy AI 22
Box 2. Algorithmic management 25
Box 3. Fairness: Different definitions and measures 28
Box 4. Facial recognition systems 29
Box 5. Labour rights concerns in the development of AI applications 31
Box 6. AI and algorithm audits 38
Box 7. Ethics Washing 51

Executive summary

Artificial Intelligence (AI) systems can process vast and various amounts of data quickly and automatically, and provide decision-making assistance. These two abilities offer immense opportunities for improving workplaces – for example opportunities to improve efficiency, fairness and worker safety. The use of AI systems in the workplace, however, is also raising ethical concerns. The McKinsey State of AI survey in 2021 found that on the use of AI at work, 44% of respondents in advanced economies expressed concerns about explainability; 41% were concerned about privacy, and 30% about equity and fairness.

To help policymakers and other relevant stakeholders ensure an ethical use of AI in the workplace, this paper identifies the core principles of trustworthiness that need to be upheld by AI systems, applying the classification framework of the OECD AI Principles. **The paper takes a risk-management perspective aiming to identify and prevent potential harm**. The paper therefore focuses on cataloguing the potential risks rather than on the opportunities offered by AI in the workplace.

The paper also reviews ongoing policy action to promote trustworthy use of AI in the workplace. Existing legislation on anti-discrimination, data protection, deceptive practices and rights to due process must be enforced effectively when AI systems are used in the workplace, and should serve as the foundation for new policy. Economy- and society-wide initiatives on AI, such as the EU AI Act and standard-setting, can also contribute to trustworthy AI use in the workplace. Finally, the paper reviews new workplace-specific policies and collective agreements that are being used to promote trustworthy AI.

The paper benefitted from consultation with a number of key stakeholders and experts on ethical workplace AI, including AI developers, employers, unions, and academics.

Trustworthy use of workplace AI means recognizing and addressing the risks it can raise about human rights (including privacy, fairness, agency and dignity); transparency and explainability; robustness, safety and security; and accountability.

- AI's ability to make predictions and process unstructured data is transforming and extending **workplace monitoring.** The nature of the data that can be collected and processed also raises concerns, as it can link together sensitive physiological and social interaction data.
- Formalizing rules for management processes through AI systems can improve fairness in the workplace, but AI systems can multiply and systematize existing human **biases.** The collection and curation of high-quality data is a key element in assessing and potentially mitigating biases – but presents challenges for the respect of privacy.
- Systematically relying on AI-informed decision-making in the workplace can reduce workers' **autonomy and agency**. This may reduce creativity and innovation, especially if AI-based hiring also leads to a standardization of worker profiles. On the other hand, the use AI systems at work could free up time for more creative and interesting tasks.
- For **transparency and consent,** job applicants and workers may not be aware of AI system use, and even if they are may not be in a position to refuse its use.

- **Understandable explanations about employment decisions** that affect workers and employers are too often unavailable with workplace AI systems. Improved technical tools for transparency and explainability will help, although many system providers are reluctant to make proprietary source code or algorithm available. Yet enhanced transparency and explainability in workplace AI systems has the potential to provide more helpful explanations to workers than traditional systems.
- Workers can struggle to **rectify AI system outcomes** that affect them. This is linked to lack of explainability but also to lacking rights to access data used to make decisions, which makes them difficult to challenge. Contract- and gig-workers in particular can face such issues.
- AI systems present many opportunities to strengthen the **physical safety and well-being of workers**, but they also present some risks. Risks include heightened digital security risks and excessive pressure on workers. It can also be more difficult to anticipate the actions of AI-based robots due to their increased mobility and decision-making autonomy.
- Deciding who should be held **accountable** in case of system harm is not straightforward. Having a human "in the loop" may help with accountability, but it may be unclear which employment decisions require this level of oversight.
- **Audits** of workplace AI systems can improve accountability if done carefully. Possible requisites for audits include auditor independence; representative analysis; data, code and model access; and consideration of adversarial actions.

Enforcing and strengthening existing policy should be the foundation for policy action, even as society-wide and workplace-specific measures on AI help fill gaps.

- The reliance of workplace AI systems' on data can bring them into conflict with **existing data protection legislation**. For example, cases brought under Article 22 of the EU's General Data Protection Regulation (GDPR) have required companies to disclose data used in their AI systems, or to reinstate individuals dismissed solely based on algorithms.
- Employment **anti-discrimination** legislation is relevant to address some concerns about workplace AI bias**.**
- Legislation on **deceptive practices and consumer protection** is being used to require more transparency from companies about the functioning of workplace algorithms, and require developers to meet the ethical standards they advertise about their products.
- Workers' legal **rights to due process in employment decisions** can be used to require increased transparency and explainability**.**
- A number of OECD countries are considering **society-wide AI legislative proposals** that would also apply to the workplace**.** A notable example is the EU AI Act, which would classify some AI systems used in employment as "unacceptable risk" (e.g. those considered manipulative) and the rest as "high risk". This would subject them to legal requirements relating to data protection, transparency, human oversight and robustness, among others.
- National or international **standard-setting,** along with other **self-regulatory approaches,** can provide technical parameters for trustworthy AI systems, and notably for workplace use**.**
- **Regulatory efforts have also zeroed in on use of AI in the workplace.** In the US, Illinois and Maryland require applicant consent for the use of facial recognition tools in hiring. The New York City Council mandates annual algorithmic bias audits for "automated employment decision tools".
- Formalising an **agreement between unions and business associations**, legislation in Spain now mandates transparency for AI systems affecting working conditions or employment status. Indeed, social partners have proactively set out proposals on workplace AI use, and will be key stakeholders in developing new legislation.

Résumé

Les systèmes d'intelligence artificielle (IA) peuvent traiter rapidement et automatiquement des quantités importantes et variées de données, et fournir une aide à la décision. Ces deux capacités offrent d'immenses possibilités pour améliorer les lieux de travail – que ce soit par exemple en améliorant l'efficacité, l'équité et la sécurité des travailleurs. Toutefois, l'utilisation de systèmes d'IA dans les lieux de travail soulève également des préoccupations d'ordre éthique. L'enquête McKinsey sur l'état de l'IA en 2021 a révélé qu'en ce qui concerne l'utilisation de l'IA au travail, 44 % des personnes interrogées dans les économies avancées s'inquiètent par rapport à l'explicabilité, 41 % sont préoccupées par la protection de la vie privée et 30 % par l'équité et la justice.

Afin d'aider les décideurs et les autres parties prenantes à garantir une utilisation éthique de l'IA dans le monde du travail, ce document identifie les principes fondamentaux de fiabilité que doivent respecter les systèmes d'IA, en appliquant le cadre de classification des Principes sur l'IA de l'OCDE. **Le document adopte une perspective de gestion des risques visant à identifier et à prévenir les dommages potentiels**. Il se concentre donc sur le recensement de risques potentiels plutôt que sur les opportunités offertes par l'IA sur le lieu de travail.

Le document passe également en revue les actions politiques en cours pour promouvoir une utilisation digne de confiance de l'IA sur le lieu de travail. La législation existante en matière de lutte contre la discrimination, de protection des données, de pratiques trompeuses et de droit à des procédures régulières doit être appliquée efficacement lorsque des systèmes d'IA sont utilisés sur le lieu de travail, et devrait servir de base pour la création de nouvelles politiques. Les initiatives économiques et sociales générales en matière d'IA, telles que la loi européenne sur l'IA ou la définition au niveau national de normes, peuvent également contribuer à une utilisation fiable de l'IA sur le lieu de travail. Enfin, le document examine les nouvelles politiques et conventions collectives spécifiques au lieu de travail qui sont utilisées pour promouvoir l'IA fiable.

Le document a bénéficié de la consultation d'un nombre d'acteurs clés et d'experts en matière d'IA éthique dans le monde du travail, notamment des développeurs d'IA, des employeurs, des syndicats et des universitaires.

Une utilisation digne de confiance de l'IA dans le monde du travail implique de reconnaître et de traiter les risques qu'elle peut soulever en matière de droits de l'homme (y compris par rapport à la vie privée, à l'équité, à la représentation et à la dignité), de transparence et d'explicabilité, de robustesse, de sûreté et de sécurité, et de responsabilité.

- La capacité de l'IA à faire des prédictions et à traiter des données non structurées transforme et étend la **surveillance au travail**. La nature des données qui peuvent être collectées et traitées suscite également des inquiétudes, car elles peuvent associer des données physiologiques et d'interaction sociale sensibles.

- La formalisation des règles des processus de gestion par les systèmes d'IA peut améliorer l'équité, mais les systèmes d'IA peuvent multiplier et systématiser les **biais** humains existants. La collecte et la conservation de données de haute qualité est un élément clé pour évaluer et potentiellement atténuer les biais - mais présente des défis pour le respect de la vie privée.
- Le fait de s'appuyer systématiquement sur des décisions fondées sur l'IA sur le lieu de travail peut **réduire l'autonomie et l'agence des travailleurs**. Cela peut réduire la créativité et l'innovation, en particulier si l'embauche basée sur l'IA conduit également à une standardisation des profils des travailleurs. Cependant, l'utilisation de systèmes d'IA au travail pourrait libérer du temps pour des tâches plus créatives et intéressantes.
- En ce qui concerne la **transparence et le consentement**, les candidats à l'emploi et les travailleurs peuvent ne pas être au courant de l'utilisation des systèmes d'IA, et même s'ils le sont, ils peuvent ne pas être en mesure de refuser cette utilisation.
- Des **explications compréhensibles sur les décisions d'emploi** qui affectent les travailleurs et les employeurs sont trop souvent indisponibles avec les systèmes d'IA dans le monde de travail. Des outils techniques améliorés pour la transparence et l'explicabilité pourront aider, mais beaucoup de fournisseurs de systèmes sont réticents à mettre à disposition leur code source ou algorithme propriétaire. Ceci dit, l'amélioration de la transparence et de l'explicabilité de ces systèmes d'IA a le potentiel de fournir des explications plus utiles aux travailleurs que les systèmes traditionnels.
- Les travailleurs peuvent avoir du mal à **rectifier les résultats des systèmes d'IA** qui les affectent. Cela est lié au manque d'explicabilité mais aussi à l'absence de droits d'accès aux données utilisées pour prendre les décisions, ce qui les rend difficiles à contester. Les travailleurs sous contrat temporaire et les travailleurs indépendants, en particulier, peuvent être confrontés à ces problèmes.
- Les systèmes d'IA offrent de nombreuses possibilités pour renforcer la **sécurité physique et le bien-être des travailleurs**, mais ils présentent également certains risques. Les risques comprennent des risques plus importants en matière de sécurité numérique et une pression excessive sur les travailleurs. Il peut également être plus difficile d'anticiper les actions des robots basés sur l'IA en raison de leur mobilité accrue et de leur autonomie de décision.
- Il n'est pas simple de décider qui doit être tenu pour **responsable** en cas de problème induit par un système. La présence d'un humain "dans la boucle" peut contribuer à la responsabilisation, mais il n'est pas toujours évident de déterminer quelles décisions d'emploi nécessitent ce niveau de surveillance.
- Les **audits** des systèmes d'IA sur le lieu de travail peuvent améliorer la responsabilité s'ils sont réalisés avec soin. Les conditions requises pour les audits comprennent l'indépendance de l'auditeur, une analyse représentative, l'accès aux données, aux codes et aux modèles, et la prise en compte des actions contradictoires.

L'application et le renforcement de la politique existante devraient constituer le fondement de l'action politique, même si les mesures relatives à l'IA prises à l'échelle de la société et du lieu de travail permettent de combler les lacunes.

- La dépendance des systèmes d'IA sur le lieu de travail à l'égard des données peut les mettre en conflit avec la **législation existante en matière de protection des données**. Par exemple, des affaires portées devant les tribunaux en vertu de l'article 22 du règlement général sur la protection des données (RGPD) de l'UE ont obligé des entreprises à divulguer les données utilisées dans leurs systèmes d'IA ou à réintégrer des personnes licenciées uniquement sur la base d'algorithmes.

- La législation **anti-discrimination** en matière d'emploi est pertinente pour répondre à certaines préoccupations concernant les biais de l'IA sur le lieu de travail.
- La législation sur les **pratiques trompeuses et la protection des consommateurs** est en train d'être utilisée pour exiger plus de transparence de la part des entreprises sur le fonctionnement des algorithmes sur le lieu de travail, et obliger les développeurs à respecter les normes éthiques qu'ils proclament pour leurs produits.
- Les droits légaux des travailleurs à une **procédure régulière dans les décisions d'emploi** peuvent être utilisés pour exiger une transparence et une explicabilité accrues.
- Un certain nombre de pays de l'OCDE envisagent des **propositions législatives sur l'IA à l'échelle de la société** qui s'appliqueraient également au lieu de travail. Un exemple notable est la loi européenne sur l'IA, qui classerait certains systèmes d'IA utilisés dans l'emploi comme présentant un "risque inacceptable" (par exemple, ceux considérés comme manipulateurs) et les autres comme présentant un "risque élevé". Cela les soumettrait à des exigences légales en matière de protection des données, de transparence, de surveillance humaine et de robustesse, entre autres.
- L'établissement de **normes nationales ou internationales**, ainsi que d'autres approches d'autorégulation, peuvent fournir des paramètres techniques pour des systèmes d'IA dignes de confiance, notamment pour une utilisation sur le lieu de travail.
- **Les efforts de réglementation se sont également concentrés sur l'utilisation de l'IA sur le lieu de travail en particulier**. Aux États-Unis, l'Illinois et le Maryland exigent le consentement de candidats postulant pour un emploi avant l'utilisation d'outils de reconnaissance faciale. Le conseil municipal de New York impose des audits annuels par rapports aux biais algorithmiques pour les "outils automatisés de décision en matière d'emploi".
- Formalisant un accord **entre syndicats et associations d'entreprises**, la législation espagnole impose désormais la transparence des systèmes d'IA affectant les conditions de travail ou le statut d'emploi. En effet, les partenaires sociaux ont formulé de manière proactive des propositions sur l'utilisation de l'IA sur le lieu de travail, et seront des acteurs clés dans l'élaboration de la nouvelle législation.

Kurzfassung

Systeme der Künstlichen Intelligenz (KI) können große und vielfältige Datenmengen schnell und automatisch verarbeiten und Entscheidungshilfen geben. Diese beiden Fähigkeiten bieten immense Chancen für die Verbesserung von Arbeitsplätzen - zum Beispiel für die Verbesserung von Effizienz, Fairness und Arbeitssicherheit. Der Einsatz von KI-Systemen am Arbeitsplatz wirft jedoch auch ethische Bedenken auf. Die McKinsey-Umfrage zum Stand der KI im Jahr 2021 ergab, dass 44 % der Befragten in den Industrieländern Bedenken hinsichtlich der Erklärbarkeit des Einsatzes von KI am Arbeitsplatz äußerten. Weitere 41 % der Befragten waren besorgt über den Datenschutz und 30 % über Gerechtigkeit und Fairness.

Um politischen Entscheidungsträger und Entscheidungsträgerinnen und anderen relevanten Schlüsselakteuren dabei zu helfen, einen vertrauenswürdigen Einsatz von KI am Arbeitsplatz zu gewährleisten, werden in diesem Papier die Kernprinzipien der Vertrauenswürdigkeit identifiziert, die von KI-Systemen eingehalten werden müssen. Dabei wird der Klassifizierungsrahmen der OECD KI-Prinzipien angewendet. **Das Papier nimmt eine Risikomanagement-Perspektive ein, die darauf abzielt, potenziellen Schaden zu erkennen und zu verhindern**. Das Papier konzentriert sich daher auf die Katalogisierung der potenziellen Risiken und weniger auf die Chancen, die KI am Arbeitsplatz bietet.

Das Papier gibt darüber hinaus einen Überblick über laufende politische Maßnahmen zur Förderung des vertrauenswürdigen Einsatzes von KI am Arbeitsplatz. Bestehende Gesetze zur Antidiskriminierung, zum Datenschutz, zu irreführenden Praktiken und zum Recht auf ein ordnungsgemäßes Verfahren müssen wirksam durchgesetzt werden, wenn KI-Systeme am Arbeitsplatz eingesetzt werden, und sollten als Grundlage für neue politische Maßnahmen dienen. Wirtschafts- und gesellschaftsweite Initiativen zu KI, wie die KI-Verordnung der EU und die Festlegung von Standards, können ebenfalls zu einem vertrauenswürdigen Einsatz von KI am Arbeitsplatz beitragen. Schließlich gibt das Papier einen Überblick über neue arbeitsplatzspezifische Maßnahmen und Tarifverträge, die zur Förderung vertrauenswürdiger KI eingesetzt und vereinbart werden.

Das Papier profitierte von der Konsultation einer Reihe von wichtigen Interessenvertretern und Interessensvertreterinnen, sowie Experten und Expertinnen für ethische KI am Arbeitsplatz, darunter KI-Entwickler und Entwicklerinnen, Arbeitgeberverbände, Gewerkschaften und Akademiker und Akademikerinnen

Der vertrauenswürdige Einsatz von KI am Arbeitsplatz bedeutet, dass die damit verbundenen Risiken in Bezug auf Menschenrechte (einschließlich Privatsphäre, Fairness, Handlungsfähigkeit und Würde), Transparenz und Erklärbarkeit, Robustheit, Sicherheit und Verantwortlichkeit erkannt und angegangen werden müssen.

- Die Fähigkeit der KI, Vorhersagen zu treffen und unstrukturierte Daten zu verarbeiten, verändert und erweitert die Überwachung am Arbeitsplatz. Die Art der Daten, die gesammelt und verarbeitet

werden können, gibt ebenfalls Anlass zur Sorge, da sie sensible physiologische und soziale Interaktionsdaten miteinander verknüpfen können.

- Die Formalisierung von Regeln für Managementprozesse durch KI-Systeme kann die Fairness am Arbeitsplatz verbessern, gleichzeitig aber auch bestehende menschliche Voreingenommenheit vervielfachen und systematisieren. Die Sammlung und Aufbereitung qualitativ hochwertiger Daten ist ein Schlüsselelement für die Bewertung und potenzielle Abschwächung von Voreingenommenheiten, stellt jedoch eine Herausforderung für die Wahrung der Privatsphäre dar.
- Der systematische Rückgriff auf KI-gestützte Entscheidungen am Arbeitsplatz kann die Autonomie und Handlungsfähigkeit vonArbeitnehmern und Arbeitnehmerinnen einschränken. Kreativität und Innovationsfähigkeit könnten zum Beispiel eingeschränkt werden, wenn KI-gestützte Einstellungen auch zu einer Standardisierung von Arbeitnehmerprofilen führen. Andererseits könnte der Einsatz von KI-Systemen am Arbeitsplatz Zeit für kreativere und interessantere Aufgaben freisetzen.
- Im Hinblick auf Transparenz und Zustimmung sind sich Bewerber und Bewerberinnen, sowie Arbeitnehmer und Arbeitnehmerinnen möglicherweise nicht über die Verwendung von KI-Systemen bewusst, und selbst wenn sie es sind, sind sie möglicherweise nicht in der Lage, deren Verwendung abzulehnen.
- Verständliche Erklärungen zu Beschäftigungsentscheidungen, die sich auf Arbeitnehmer und Arbeitnehmerinnen, sowie auf Arbeitgeber und Arbeitgeberinnen auswirken, sind bei KI-Systemen am Arbeitsplatz allzu oft nicht verfügbar. Verbesserte technische Werkzeuge für Transparenz und Erklärbarkeit können dabei helfen, doch viele Systemanbieter zögern, urheberrechtlich geschützten Quellcode oder Algorithmen zur Verfügung zu stellen. Eine verbesserte Transparenz und Erklärbarkeit von KI-Systemen am Arbeitsplatz hat jedoch das Potenzial, Arbeitnehmern und Arbeitnehmerinnen hilfreichere Erklärungen zu liefern als herkömmliche Systeme.
- Arbeitnehmer und Arbeitnehmerinnen können Schwierigkeiten haben, sie selbst betreffende Ergebnisse von KI-Systemen anzufechten und zu korrigieren. Dies hängt mit der mangelnden Erklärbarkeit zusammen, aber auch mit dem fehlenden Recht auf Zugang zu den Daten, die für die Entscheidungsfindung verwendet werden. Vor allem Selbstständige, zum Beispiel in der Plattformökonmie,können mit solchen Problemen konfrontiert werden.
- KI-Systeme bieten viele Möglichkeiten, die physische Sicherheit und das Wohlbefinden von Arbeitnehmern und Arbeitnehmerinnen zu verbessern, bergen aber auch einige Risiken. Zu den Risiken gehören erhöhte digitale Sicherheitsrisiken und übermäßiger Druck am Arbeitsplatz. Zudem kann es schwieriger sein, die Handlungen von KI-gestützten Robotern vorherzusehen, da sie mobiler sind und autonomer entscheiden.
- Die Entscheidung darüber, wer im Falle eines Systemschadens von KI-Systemen zur Rechenschaft gezogen werden sollte, ist nicht einfach. Die Einbindung eines Menschen in KI-gestützten Entscheidungsfindungen kann bei der Rechenschaftspflicht zwar helfen, aber es kann unklar sein, welche Entscheidungen am Arbeitsplazt diese Aufsichtsebene erfordern.
- Prüfungen von KI-Systemen am Arbeitsplatz können die Rechenschaftspflicht verbessern, wenn sie sorgfältig durchgeführt werden. Zu den möglichen Voraussetzungen gehören die Unabhängigkeit der Prüfer und Prüferinnen, eine repräsentative Analyse, der Zugang zu Daten, Codes und Modellen, und die Berücksichtigung gegenteiliger Maßnahmen.

Die Durchsetzung und Stärkung vorhandener Richtlinien sollte die Grundlage für politische Maßnahmen bilden, auch wenn gesellschaftsweite und arbeitsplatzspezifische Maßnahmen zu KI dazu beitragen, Lücken zu schließen.

- Die Datenabhängigkeit von KI-Systemen am Arbeitsplatz kann sie in Konflikt mit bestehenden Datenschutzvorschriften bringen. In einigen Rechtsfällen wurden Unternehmen gemäß Artikel 22

der EU-Datenschutzgrundverordnung (DSGVO) beispielsweise aufgefordert, die in ihren KI-Systemen verwendeten Daten offenzulegen oder Personen, die allein aufgrund von Algorithmen entlassen wurden, wieder einzustellen.

- Die Antidiskriminierungsgesetze im Bereich der Beschäftigung sind relevant, um einige Bedenken hinsichtlich der Voreingenommenheit von KI am Arbeitsplatz auszuräumen.
- Die Gesetzgebung zu irreführenden Praktiken und zum Verbraucherschutz wird genutzt, um von Unternehmen mehr Transparenz über die Funktionsweise von Algorithmen am Arbeitsplatz zu fordern und von Entwicklern und Entwicklerinnen die Einhaltung ethischer Standards zu verlangen, mit denen sie für ihre Produkte werben.
- Die gesetzlichen Rechte von Arbeitnehmern auf ein ordnungsgemäßes Verfahren bei KI-gestützten Entscheidungen am Arbeitsplatz können genutzt werden, um mehr Transparenz und Erklärbarkeit zu fordern.
- Eine Reihe von OECD-Ländern erwägt gesellschaftsweite KI-Gesetzesvorschläge, die auch für den Arbeitsplatz gelten würden. Ein bemerkenswertes Beispiel ist die KI-Verordnung der EU, die einige KI-Systemeam Arbeitsplatz als "unannehmbares Risiko" (z. B. solche, die als manipulativ gelten) und die übrigen als "hohes Risiko" einstufen würde. Damit würden sie gesetzlichen Anforderungen unterliegen, die sich unter anderem auf Datenschutz, Transparenz, menschliche Aufsicht und Robustheit beziehen.
- Nationale oder internationale Normen sowie andere Ansätze der Selbstregulierung können technische Parameter für vertrauenswürdige KI-Systeme und insbesondere für den Einsatz am Arbeitsplatz liefern.
- Auch die Regulierungsbehörden haben sich auf den Einsatz von KI am Arbeitsplatz konzentriert. In den USA verlangen die Bundesstaaten Illinois und Maryland die Zustimmung von Bewerbern und Bewerberinnen für den Einsatz von Gesichtserkennungssystemen in Einstellungsverfahren. Der Stadtrat von New York schreibt zudem jährliche Prüfungen von Algorithmen für "automatisierte Instrumente zur Entscheidungsfindung bei der Einstellung" vor.
- Durch die Formalisierung einer Vereinbarung zwischen Gewerkschaften und Unternehmensverbänden schreibt die spanische Gesetzgebung nun die Transparenz von KI-Systemen vor, die sich auf die Arbeitsbedingungen oder den Beschäftigungsstatus auswirken. Die Sozialpartner haben proaktive Vorschläge zum Einsatz von KI am Arbeitsplatz unterbreitet und werden bei der Ausarbeitung neuer Rechtsvorschriften eine wichtige Rolle spielen.

1. Introduction

The use of Artificial Intelligence (AI) has the potential to transform the workplace because of two main characteristics of AI that are unparalleled in other digital technologies:

- AI systems[1] enable the automated processing of numerous types of data and often in vast amounts, producing outcomes and recommendations rapidly and at scale.
- AI technologies, notably through their ability to learn, allow for decision making assistance through predictions in tasks such as recognition, event detection, forecasting, personalisation, interaction support, goal-driven optimisation and reasoning with knowledge structures (OECD, forthcoming[1]).

AI systems hold great promise for the workplace. They can complement or augment human labour, and they can automate an expanding set of repetitive, demeaning or even hazardous tasks – increasing the amount of time that workers spend on safer and more fulfilling tasks, and increasing productivity. AI tools for human resource management can strengthen efficiency, improve workplace culture and reinforce the quality of the work environment.

At the same time, the use of AI systems at work involves risks. The risk of automation of certain tasks may be heightened by AI, especially for high-skilled workers (Georgieff and Hyee, 2021[2]). AI can also increase inequalities, if workers who are already at an advantage in the labour market are also those who stand to benefit the most from it (Georgieff and Hyee, 2021[2]). The use of AI systems in the workplace presents concerns about the respect of values such as privacy, fairness, agency, transparency and accountability. AI systems can also change the way employers interact with and hire workers as well as the way workers interact with each other (Lane and Saint-Martin, 2021[3]).

Some of these concerns may not be dramatically different from the concerns raised by other digital technologies, or traditional, non-digital, workplace management practices. Invasive workplace surveillance, for example, can exist without artificial intelligence, or digital technologies for that matter. (Lane and Saint-Martin, 2021[3]). Indeed, existing workplace practices and human decision making in the workplace also present ethical concerns. This paper does not assert that AI poses a larger ethical risk in the workplace than these practices and decisions. **But the use of AI systems can extend and systematize ethical failings as well as fundamentally change the relationship between workers and their managers.**

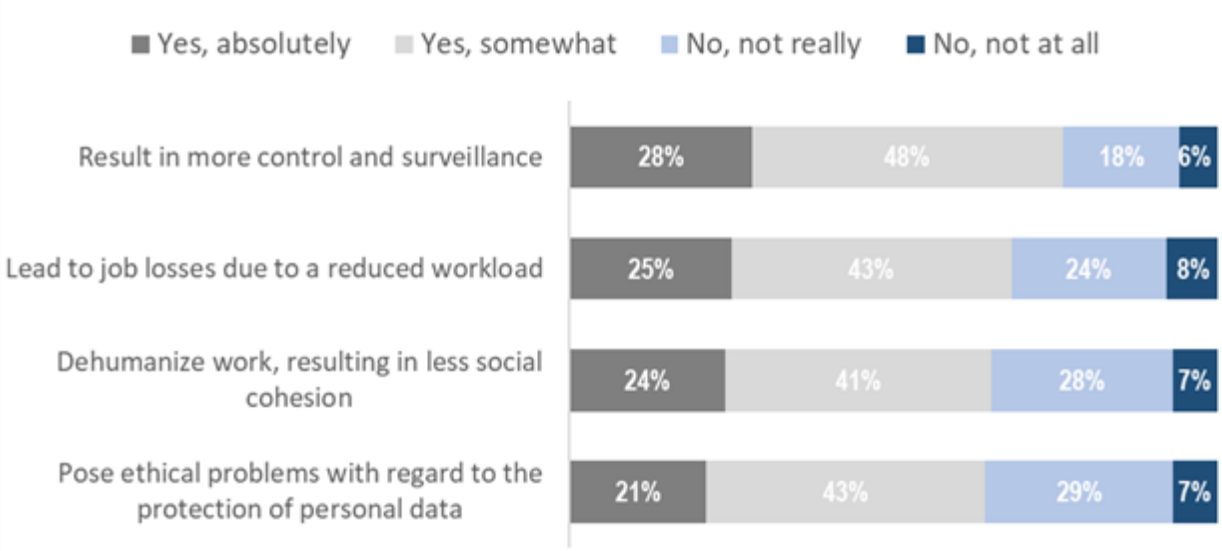
A number of surveys confirm that people see benefits and risks in the use of AI systems at work. According to a 2018 BCG GAMMA–Ipsos survey, workers share the view that AI and its applications in their workplace will help them work more efficiently, reducing the time they spend in tedious tasks, reducing the risk of errors and helping them meet deadlines (BGC GAMMA and IPSOS, 2018[4]). But workers also have concerns about the use of AI in the workplace. Among the surveyed workers, 76% see a danger that AI may result in more surveillance, 65% that it may dehumanize work and 21% that it may lead to unethical

[1] An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy. (OECD, 2019[283])

use of personal data – and these proportions are actually larger among users of AI systems at work than among non-users, potentially indicating that exposure to AI the workplace makes workers more aware of the connected ethical risks (Figure 4). Similarly, the McKinsey State of AI survey in 2021 found that 44% of respondents in advanced economies expressed concerns about explainability and the use of AI at work. 41% were concerned about privacy, and 30% about equity and fairness (McKinsey, 2021[5]). A 2019 survey of over 8,000 human resources leaders across 10 countries indicates that close to one third of them report security and privacy as the main barriers to adoption of AI systems in their workplace (Oracle & Future Workplace, 2019[6]). Still, the same survey noted that a number of respondents hope that changes to AI will actually help make their workplaces better – 43% were excited about how AI would affect the future of work, with 1 in 5 for example stating that they think it will allow for better or healthier work relationships.

Figure 1. Employees see risks in the use of AI at work, especially when already using such tools

Respondents reporting that in workplace there is a danger that the development of AI and its application may..., %



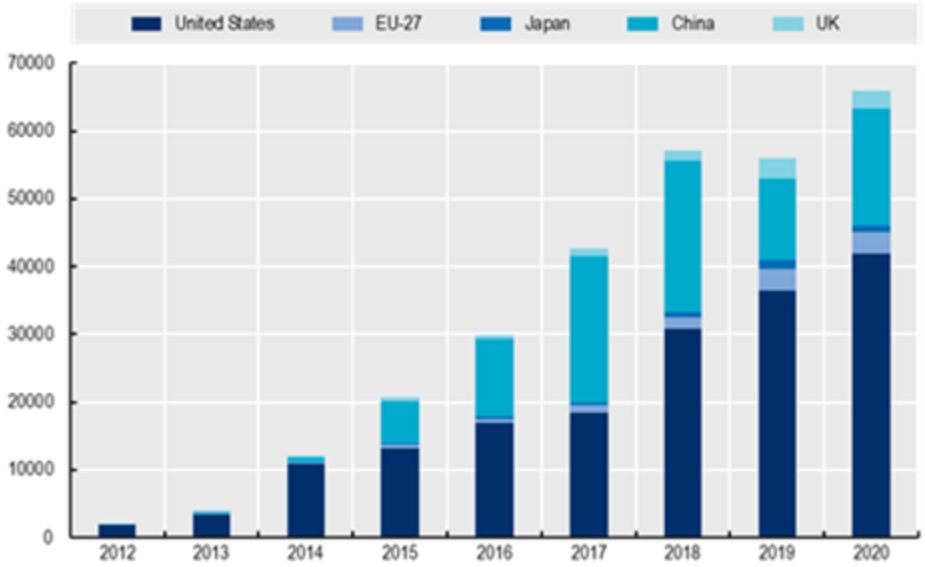
Note: Survey of a representative sample of the national active adult population in Canada, China, France, Germany, Spain, the UK, and the US.
Source: BCG GAMMA – Ipsos 2018 survey https://www.ipsos.com/en/revolution-ai-work.

Comparable evidence on the extent of the use of AI systems in the workplace is still scant. In the past decade, AI start-ups have increasingly and quickly gained venture capital funding (Figure 2), but diffusion seems[2] to still be low overall, and uneven. For example, ICT usage surveys in Canada, Denmark, France, and Korea found that between 2 and 11% of firms used AI technology in recent years, but large firms (250+ people) were 2 to 6 times more likely to do so (Montagnier and Ek, 2021[7]).

[2] Montagnier and Ek (2021[7]) highlight how existing analyses and surveys' differences in measurement definition and scope lead to different estimates of AI adoption in firms.

Figure 2. VC Investments in AI have been increasing rapidly

Venture Capital investments in AI by country, USD millions, 2012-2020



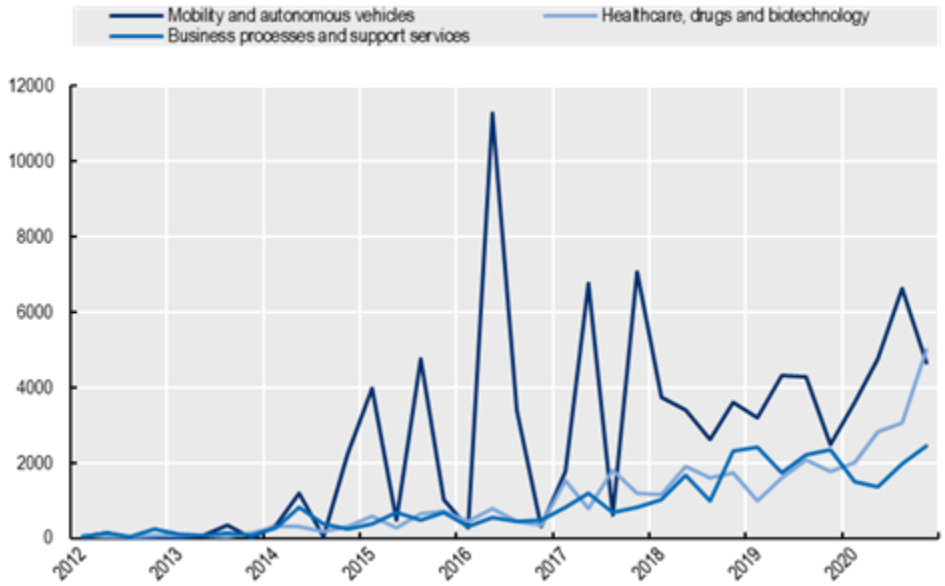
Source: OECD.AI (2021) using data from Preqin

Still, workplace AI system usage seems to be non-trivial, and growing, though estimates vary. In 2019, half of over 8000 workers surveyed in 10 countries reported using some form of AI systems at work (Bertallee, 2019[8]). In the PwC *AI Predictions 2021* survey of over 1000 US executives, a quarter of the participating companies reported widespread adoption of AI systems, up from 18% in 2020 (Likens et al., 2021[9]). The McKinsey *State of AI* surveys[3] suggest that 50% of respondents reported AI adoption in at least one function in 2020, rising to 56% in 2021. In 2021, 8% of respondents reported using AI for human resources (McKinsey, 2021[5]). The 2019 *Gartner Artificial Intelligence Survey* found that 17% of organization use AI solutions in their HR functions (Baker, 2020[10]). At the end of 2020, the business processes and support services industry was the third largest industry for AI VC funding (Figure 3).

[3] This online survey had 1843 respondents in May/June 2021. Weighting by a respondent's nation's GDP adjusted for differences in response rates worldwide.

Figure 3. The business processes and support services industry was the third largest industry for AI VC funding at the end of 2020

Venture Capital investments in AI by industry, USD millions, by quarter, 2012-2020



Note: Other industries were not included in this chart for simplicity
Source: OECD.AI (2021) using data from Preqin

As decisions about the adoption of AI in the workplace are being made, often at a pace faster than regulation, some unforeseen consequences and unfair outcomes have emerged (IBM, 2021[11]). While existing human-led workplace practices may not always be fully ethical, the introduction of AI technology in the workplace should not reinforce or add new layers of ethical concerns. Indeed, large numbers of reflections and guidelines around the ethics of AI are emerging. It is essential, however, to have a shared understanding of what an ethical use of AI in the workplace entails so that – in addition to identifying the associated issues – the needed policies, processes and safeguards can be put in place.

This paper contributes to the conversation by identifying core principles of trustworthiness that need to be upheld for an ethical use of AI in the workplace, using the OECD Principles for the Responsible Stewardship of Trustworthy AI.[4] The paper focuses on ethical risks posed by AI rather than on its opportunities, reflecting a risk-management perspective, and the need to peempt potential harm. In this sense, the paper does not present an assessment of the balance of risks and opportunities offered by AI in the workplace. Chapter 2 looks at ethical concersn in terms of human rights (privacy, fairness, agency, and dignity) (Section 2.1); transparency and explainability (Section 2.2); robustness, safety and security (Section 2.3); and accountability (Section 2.4).

Policy action is being taken to ensure the trustworthy use of AI in the workplace, whether through enforcement of existing legislation or the development of new policies. Lack of action to ensure trustworthy AI in the workplace may put brakes on AI diffusion and/or result in undesirable consequences for workers, employers, and society as a whole (OECD, 2019[12]). Chapter 3 first discusses how existing

[4] There is broad debate about the use of ethics to describe principles that should govern the use of AI in society and in the workplace. This report, as detailed in chapter 2, will follow the OECD AI principles's definition of trustworthiness, but it will also use the term ethical, in line with public debate, and particularly to discuss potential concerns that may emerge.

legislation (including legislation on data protection, anti-discrimination, deceptive practices and consumer protection, as well as rights to due process) can be applied to AI use in the workplace, and/or be used to provide a foundation for further policy evolution (Section 3.1). Section 3.2 then provides an overview of new society- and economy-wide policy efforts that have direct implications for trustworthy use of AI in the workplace. In particular, the chapter presents ambitious new AI legislation (such as the EU AI Act), that have important considerations and implications for the workplace. The chapter closes with how some jurisdictions are developing or proposing new workplace-specific policy on trustworthy AI, from municipal legislation to policy change via collective bargaining (Section 3.3).

2. Emerging ethical issues in the use of AI in the workplace

To help determine a shared understanding of what trustworthy use of AI in the workplace entails, this paper systematically identifies the ethical risks[5] that can emerge with the use of AI systems in the workplace, using the framework of the OECD "Principles for the Responsible Stewardship of Trustworthy AI" (OECD AI Principles). These principles, laid out in the Recommendation of the Council on Artificial Intelligence (OECD, 2019[12]) (see Box 1), have been developed to promote trustworthy use of AI in society and the economy in general, but can be directly applied to the use of AI systems in the workplace.

Box 1. OECD Principles for responsible stewardship of trustworthy AI

These Principles form Section 1 of the OECD Recommendation of the Council on Artificial Intelligence ("OECD AI Principles") (OECD, 2019[12]). The OECD AI Principles were adopted in May 2019 by the OECD member countries; since, other adherents include Argentina, Brazil, Egypt, Malta, Peru, Romania, Singapore, and Ukraine. In June 2019, the G20 adopted human-centred AI Principles that draw from the OECD AI Principles.

1. Inclusive growth, sustainable development and well-being

Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.

2. Human-centred values and fairness

a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.

b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.

[5] While also recognizing and mentioning some opportunities associated with the use of AI systems for a more ethical workplace, this paper is focused on potential risks.

3. Transparency and explainability

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

i. to foster a general understanding of AI systems,

ii. to make stakeholders aware of their interactions with AI systems, including in the workplace,

iii. to enable those affected by an AI system to understand the outcome, and,

iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

4. Robustness, security and safety

a) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.

b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.

c) AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

5. Accountability

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

The OECD, through its working group on Implementing Trustworthy AI within the OECD Network of Experts on AI, also developed a framework to compare the tools and practices that many stakeholders are developing to implement trustworthy AI systems – ranging from tools to check for bias and robustness of AI systems, to risk management guidelines, to educational material (OECD, 2021[13]).

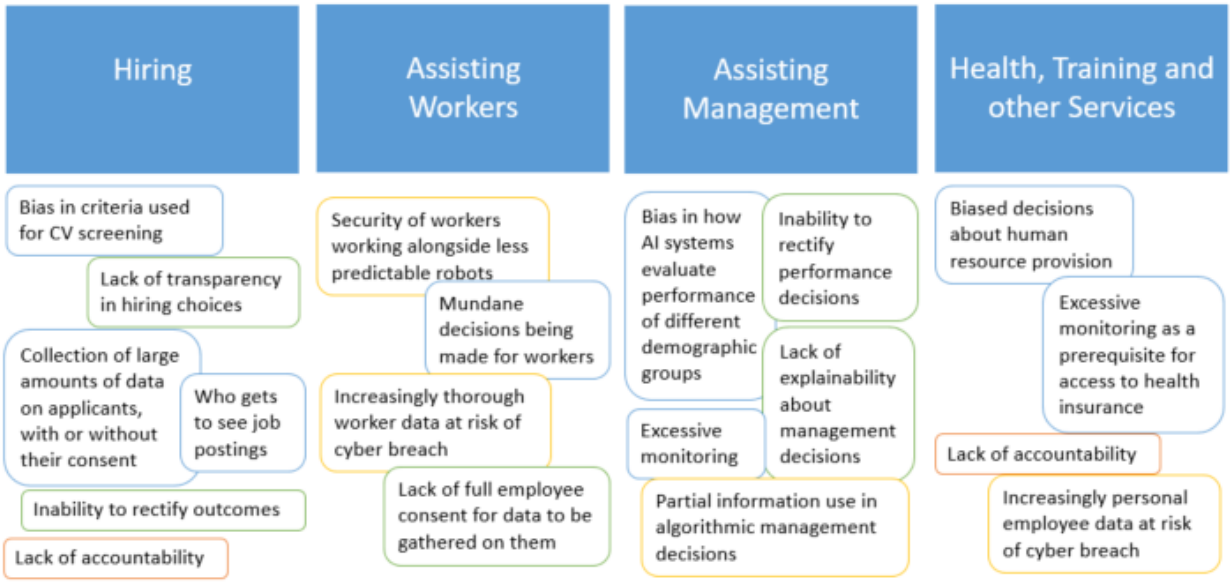
Source: OECD (2019), OECD AI Principles, http://Recommendation of the Council on Artificial Intelligence; OECD (2019), Recommendation of the Council on Artificial Intelligence, OECD, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

This chapter therefore discusses the considerations that need to be made when using AI in the workplace to uphold human rights; transparency and explainability; robustness, safety and security; and accountability. The report does not, on the other hand, discuss the inclusive growth principle: the possible challenges, and opportunities, presented by AI in terms of inclusivity are addressed in other OECD reports prepared for the OECD project AI and Work, Innovation, Productivity and Skills, including the reports on the impact of AI on job automation and on the labour market more broadly, on skills and training, or on the role of social dialogue (Verhagen, 2021[14]; Georgieff and Hyee, 2021[2]; OECD, 2022[15]).

This chapter focuses on ethical concerns as illustrated by how *existing* AI is and could be used in the workplace. The chapter grounds itself in examples of existing AI tools, and their abilities – with real and potential uses of these tools in the workplace. The chapter is organized by Principle, but it will discuss

examples of how ethical concerns emerge at various points of AI use. Points of use include use in the hiring and recruitment process, use in assisting or augmenting workers, assisting management, and finally use in providing human resource services, such as training or healthcare plans. Figure 4 illustrates a number of such examples and potential ethical concerns, grouped by Principle. As the technical capacities of AI tools evolve, some of the ethical concerns considered in this chapter may become more significant, or less relevant. For the latter, the chapter also provides brief mentions of how some technical developments could help assuage some of the ethical concerns mentioned.

Figure 4. Different types of ethical concerns can arise, all the way through the different points of use of AI systems in the workplace



Note: Different potential ethical concerns linked to different Principles of trustworthy use are presented in differently coloured boxes, according to this legend: Principle 2, Principle 3, Principle 4, and Principle 5 as defined in Box 1

2.1. Human rights: privacy, fairness, agency and dignity

The principle of human-centred values and fairness in the OECD AI Principle has a number of dimensions that are specifically relevant to ensuring a trustworthy use of AI systems in the workplace and in particular of algorithmic management systems (see Box 2): (1) whether AI systems breach workers' dignity and right to privacy; (2) whether AI systems uphold fairness, non-discrimination and avoid bias; and (3) whether they promote autonomy and agency.

Box 2. Algorithmic management

Algorithmic management can be broadly described as the full or partial automation of managerial functions, through AI systems and other digital tools, facilitated by the collection of data and the use of predictive modelling (Lee et al., 2015[16]; Nguyen and Mateescu, 2019[17]; Wood, 2021[18]). A common, related term is people analytics which describes the use of statistical tools, including AI systems, to measure, report and understand the workforce's performance in various dimensions (Briône, 2020[19]).

The expression "algorithmic management" was first coined for the platform economy – where it is often quite central – but algorithmic management is increasingly common in other employment settings as well, such as warehouses, retail and hospitality, and manufacturing (Wood, 2021[20]; Jarrahi et al., 2021[21]; Briône, 2020[19]). The types of functions for which algorithmic management is being used include direction of work (what needs to be done, how and when), evaluation, and discipline (Wood, 2021[20]; Kellogg, Valentine and Christin, 2020[22]).

These functions are being carried out with differing degrees of automation. Wood (2021[20]) finds that full automation in algorithmic management is rare. Algorithmic direction, for example, is common in platform work but, even there, workers generally retain the ability to ignore or override algorithmic directions - although they might have to provide justification of the choices made. Evaluation is found to be the type of function where the degree of automation is generally highest, with AI systems autonomously producing evaluations of workers – based for example on biometric information and customer satisfaction ratings – that are then interpreted by human managers. Forms of evaluation and discipline augmented by AI systems are common in platform work, where access to the best shifts is often decided based upon the rating made by the algorithmic system but human managers are generally able to review and overrule algorithmic decisions (Wood, 2021[20]).

Full automation in algorithmic management is rare notably because of the technical limitations inherent to modelling all the tasks and uncertainties that human managers have to take into account in their work. It is also rare because regulation in a number of countries – and notably the European Union's General Data Protection Regulation – provides a right to individuals to not to be subject to automated decisions that have significant effects (see Section on Data Protection). The ethical considerations discussed in this report should also be a deciding factor in the degree of automation that is chosen for algorithmic management in the workplace.

Overall, even if algorithmic management typically does not involve full automation and humans are involved in various ways, human decision-making is likely to be profoundly affected as algorithmic management encourages new ways of approaching, understanding, and acting upon the information that is provided to human managers involved in the process (Briône, 2020[19]). The adoption of algorithmic management systems is also bound to shape and be shaped by the power dynamics between managers and workers (Jarrahi et al., 2021[21]).

Privacy

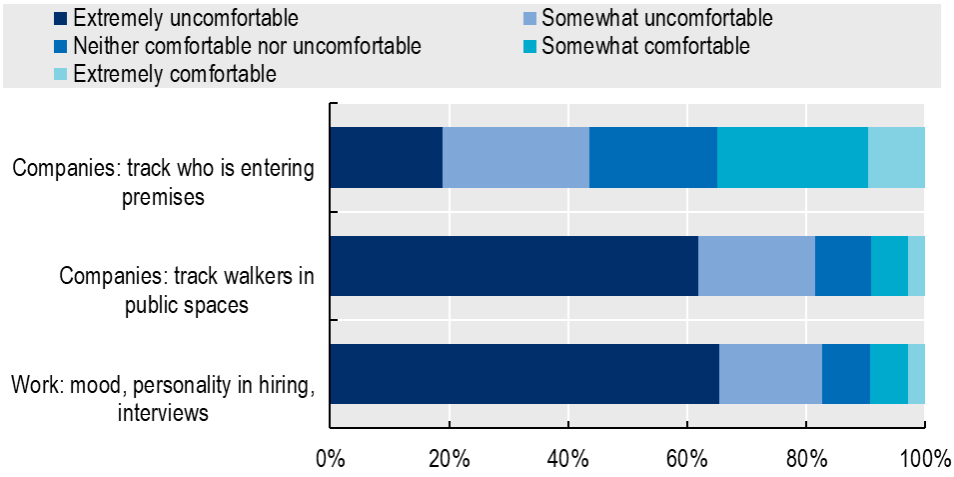
Monitoring of workers by their employers is not a new phenomenon. In fact, much of modern production strategies have hinged on the ability to monitor the execution of tasks at work to optimize workflows and productivity (Taylor, 1911[23]). Traditionally, monitoring and surveillance of workers relied on human control and were thus inherently limited. The deployment of predictive models and the processing of unstructured data (text, audio and video), together with the use of network records, phone apps, sensors, biometric tracking devices (such as wearable fitness trackers) and facial recognition systems, has enabled the development of multi-source datasets and induced a transformation in the nature of worker monitoring (Ajunwa, Crawford and Schultz, 2017[24]; Sánchez-Monedero and Dencik Sanchez-Monederoj, 2019[25]).

The extensive worker surveillance made possible by AI systems can lead to breaches of privacy, a right established as a human right by the UN Declaration of Human Rights (United Nations, 1948[26]). A survey conducted by Nature among researchers working in the field revealed, for example, broad discomfort with the use of facial recognition to monitor personality traits and the mood of candidates when hiring for a job (Figure 5). Around 65% of the 480 experts interviewed from around the world reported being extremely uncomfortable with this practice (Van Noorden, 2020[27]).

The use of remote surveillance software has exponentially increased as workers shifted to massive teleworking during the COVID-19 crisis. For example, *ActivTrak*, a major player in the industry, went from 50 client companies to 800 in March 2020 alone. *Teramind*, a leading provider of employee monitoring software solutions, reported a triple-digit percentage increase in new leads since the pandemic began (Morrison, 2020[28]). Some surveillance software include features leaving very little privacy to workers. For example, some software reportedly captured frequent live photos of workers through their company laptop webcam, displaying them on a digital shared space; others recorded workers' unsent emails or activated webcams and microphones on workers' devices (Gray, 2021[29]; Milne, 2021[30]). High levels of monitoring – where workers are not allowed to look away from a screen for example, or adjust their glasses – can make workers feel commoditized (Harwell, 2021[31]).

Figure 5. Nature Survey highlights significant discomfort over facial recognition technology

Attitudes on how comfortable expert respondents were with different uses of facial recognition technology, selected answers



Note: Questions and answers paraphrased for brevity.
Source: Nature Survey of nearly 500 researchers "who work in facial recognition, computer vision and artificial intelligence." Respondents were located around the world, but most were from Europe and North America. More information here: go.nature.com/2uwtzyh.

The nature of the data collected also raises concerns about possible privacy breaches and violation of human integrity or dignity. Wearable devices can capture sensitive physiological data on workers' health conditions, habits, and possibly the nature of their social interaction with other people. For example, analysis of heart-rate variability provides insights into emotional and physical endurance of employees; while this information can be collected and used to improve employees' health and safety, it can also be used by employers – even involuntarily – to inform consequential judgments (Maltseva, 2020[32]). The monitoring of workers' private emails, activity on social media, and location outside working hours – which have been heightened by the use of algorithmic systems – is a breach of privacy that can potentially also be used to "monitor and suppress collective action" as stated by the UK's Trade Union Congress (TUC, 2021[33]), and as observed for example for ride-hailing or delivery platforms (De Stefano, 2016[34]). Many

of these breaches may not be legal in OECD countries – see discussion of existing Data Protection legislation in Section 3.1 for example.

Fairness, bias and discrimination

Even without AI, discrimination and bias in the workplace – in recruitment, in management and promotion – are a longstanding and widespread phenomenon (Cahuc, Carcillo and Zylberberg, 2004[35]). A meta-analysis of 30 years of experiments in the United States found that white job applicants were 36% more likely to receive a callback than equally qualified African Americans, and 24% more likely than Latinos – with little significant evolution between 1989 and 2015 (Quillian et al., 2017[36]).

Formalizing rules that apply to management processes through AI can reduce bias and discrimination in these decisions, in addition to making them more efficient (Deshpande et al., 2021[37]; De Stefano, 2018[38]). Indeed AI systems used for HR processes often advertise their objectivity compared with traditional recruitment. The use of AI systems in hiring can also expand the pool of applicants considered for hiring, by allowing to expanding the number of applications that can be handled manually.

Yet AI systems struggle with bias, both at the system level and at the data or input level (Accessnow, 2018[39]; Executive Office of the President of the United States, 2016[40]). Bias at the input or data level is related to use of historical data that is biased, to the use of non-representative samples, or to the use of incomplete, incorrect or outdated data. Bias at the system level takes place through the choices of parameters and variables used in developing AI systems. The selection of the variables to be considered in an algorithmic system, the decision on how to measure them, the choice of data on which the system is trained, are all decisions made by humans based on their own priors. The lack of diversity in the AI tech industry raises concerns in this respect (West, Whittaker and Crawford, 2019[41]). Some features might be differently predictive for different groups. For example, in algorithms to recruit programmers, using the age at which the candidate started coding as a prediction variable might induce a bias towards male candidates as boys are historically more likely to start coding earlier than girls; other variables may matter to predict female capacity (Crawford et al., 2019[42]). Finally, omitting data about protected attributes (e.g. gender, race) is not a silver bullet – the presence of information that can indicate protected characteristics ("proxy characteristics") could lead to discrimination (Kim and Bodie, 2021[43]; EU FRA, 2020[44]; Prince and Schwarcz, 2020[45]).

Compared with biased human-led decision making, a systematic use of biased AI systems carries the risk of multiplying and systematizing the bias that they suffer from. While algorithms are not responsible for societal biases, they can replicate at scale, disseminate and standardize them (Institut Montaigne, 2020[46]), reinforcing historical patterns of disadvantage (Sánchez-Monedero, Dencik and Edwards, 2020[47]; Kim, 2019[48]). Besides, compared to traditional forms of discrimination, automated discrimination is more abstract and unintuitive, subtle, intangible, and difficult to detect, which challenges the legal protection offered by non-discrimination law (Wachter, Mittelstadt and Russel, 2020[49]).[6]

[6] Humans discriminate due to negative attitudes (e.g. stereotypes, prejudice) and unintentional biases (e.g. organisational practices or internalised stereotypes) which can act as a signal to victims that discrimination has occurred. Equivalent signalling mechanisms and agency do not exist in algorithmic systems (Wachter, Mittelstadt and Russel, 2020[49]).

Box 3. Fairness: Different definitions and measures

The need for fairness, and respect of human-centred values, are among of the principles that uphold trustworthy AI (OECD, 2019[12]). There is however no consensual definition and measurement of fairness in the field of machine learning and the choice about which definition to retain involves moral judgment as well as technical issues.

It is not easy to define what treating people fairly means and many definitions of fairness have been proposed. Fairness may refer to equality – by which everybody is treated in the same way – or to equity– by which everybody should have equal access to the same opportunities.

In the machine learning literature a number of fairness metrics have been proposed, including individual fairness, counterfactual fairness, unawareness, demographic parity, equalized odds, and predictive rate parity (Zhong, 2018[50]; Verma and Rubin, 2018[51]; Gajane and Pechenizkiy, 2018[52]). Individual fairness stipulates that similar individuals be classified similarly. Counterfactual fairness refers to whether a classifier produces the same result for two individuals that are identical, except for one sensitive attribute. Unawareness refers to not including the sensitive attribute as feature in the training data. Demographic parity consists in each considered group receiving positive outcomes at equal rates. Equalised odds refers to equal false positive and false negative rates across the considered groups. Predictive parity refers to whether the fraction of correct positive predictions is the same for each considered group, for a given classifier, the precision rates are equivalent for subgroups under consideration. The latter three metrics are based on the concept of group fairness, which ensures some form of statistical parity for members of different groups (Binns, 2020[53]). These statistical definitions of fairness are incompatible and non-universal; hence fairness requires a value judgment, and its application will vary according to cultures, political systems, and even possibly the field of application of the algorithm.

The lack of an objective, or at least an inter-subjective, measure of unfair bias or unfair indirect discrimination, makes it difficult to document unfairness, make it transparent, and thus mitigate it. Furthermore, whatever definition is retained, there will always be a tension between the objective of algorithms and the introduction of new constraints associated with that fairness definition. Indeed in most cases, the objective of algorithms is to maximise accuracy – i.e. the number of correct predictions made by the model. Ensuring fairness in algorithmic predictions usually consist in introducing additional constraints to the optimisation programme (Bertail et al., 2019[54]). This will generally reduce accuracy and potentially business profits (So, 2019[55]).

The risks of bias can emerge at all points of use of AI systems in the workplace.

Hiring practices are a first instance where – while offering great promise for improving matching between labour demand and supply – the use of AI systems can raise questions about fairness and discrimination (Broecke et al, 2022, forthcoming). Evidence points to bias affecting both who can see job postings and the selection of candidates when algorithms are used in hiring.

Platforms can use algorithms to select who will view a job posting based on factors such as age, ethnicity, gender, job seniority or connections to other companies – the so-called data-driven hiring funnel phenomenon (Sánchez-Monedero and Dencik, 2019[56]). For example, some companies in the US were brought to court for using algorithms that targeted potential candidates through Facebook job postings in a way that excluded older workers from seeing them (Kim, 2019[48]).

Bias and discrimination can also occur when algorithms are used to rank and select job applicants. For example, it emerged that an AI recruiting tool used by Amazon to hire software developers and other technical posts discriminated against women because it had been trained on hiring decisions made over

the previous ten years, when men comprised the vast majority of the workforce in the industry (Neff, McGrath and Prakash, 2020[57]). Predicting the success of future employees based on current employees indeed inherently skews the task towards finding candidates resembling those who have already been hired (Raghavan et al., 2019[58]).

Biases can also emerge in the use of AI systems in the workplace once someone is hired. A number of companies have implemented facial recognition systems (Box 4) to authenticate workers – though these systems continue to perform worse for people of colour (Harwell, 2021[31]). Others have used AI systems to evaluate performance in the workplace. Yet seemingly neutral automated decisions about performance evaluation can be affected by bias in the data on which systems are trained. Compared to hiring, managers also frequently have more extensive information about their staff, and can consider context including personal circumstances affecting the worker when making an evaluation. This kind of information would not typically be systematically coded in the data used by the algorithm.

Box 4. Facial recognition systems

Deborah Raji et al. (2020[59]) define facial processing technology (FPT) as a term encompassing tasks ranging from face detection, which involves locating a face within an image, to facial analysis, which determines an individual's facial characteristics, to face identification, which is the task of differentiating a single face from others.

Facial recognition technologies – which are sometimes used in the workplace to assess personality traits and emotions of job candidates during video interviews, or to recognize employees in the workplace – are being found to have much poorer accuracy for women, black people, and 18-30 year olds. In their seminal paper, Buolamwini and Gebru (2018[60]) found that male and light-skinned subjects were more accurately classified than female and dark-skinned subjects in automated facial analysis algorithms and datasets, largely because the systems were trained using mostly white, male-dominated data sets. Disabled people and all those suffering from conditions affecting their face or their voice are also at a disadvantage.

The scientific foundation of facial and especially affect recognition technologies is regularly questioned. Studying faces is unlikely to produce an objective reading of authentic interior states, as emotions are not fixed or universal, and depend on contextual, social, and cultural factors. Any simplistic mapping of a facial expression onto basic emotional categories through AI is likely to reproduce the errors of an outdated scientific paradigm. It also raises troubling ethical questions about locating the arbiter of someone's "real" character and emotions outside of the individual, and the potential abuse of power that can be justified based on these faulty claims (Whittaker et al., 2018[61]). Indeed a number of stakeholders are calling for a ban on the use of facial recognition technologies for certain applications (see Section 3.2 and Section 3.3).

If AI is only as good as the data it is trained on and AI-based decisions are only as fair as human-made decisions made in the past, then the promise of AI as a tool to promote fairness in the workplace needs to be supported by evidence of its trustworthiness (Kim and Routledge, 2021[62]).

To address bias, most vendors of hiring algorithms test their models.[7] A survey of English-language hiring algorithms vendors shows that most of them apply a de-biasing technique, called the four fifths rule, according to which applicants from a given demographic group must be selected at least 80% as often as those from any other complementary demographic group (Raghavan et al., 2019[58]). Applying this rule,

[7] Some AI software, such as for example *Applied*, are even specialised in assisting in the process of bias discovery and mitigation (Sánchez-Monedero, Dencik and Edwards, 2020[47]).

however, does not exhaust all sources of potential biases, and does not ensure that the algorithm will perform equally well across demographic groups. For example, if training data mostly comprises male employees, the algorithm might learn to associate success with traits often pertaining to men. The tool will thus be very good at predicting the top performing men and mediocre at figuring out the top performing women; if the hiring company then interviews an equal number of these selected top men and women, while the hiring practice will appear superficially unbiased, the job is still more likely to be given to a man (Rivero, 2020[63]). Beyond a focus on algorithmic de-biasing, the collection and curation of high-quality datasets used to assess and potentially mitigate fairness in algorithmic systems seems also needed (Deborah Raji et al., 2020[59]; Holstein et al., 2019[64]). More generally, the difficulty/impossibility to operationalise a consensual measure of fairness (Box 3) makes it very difficult to develop ways to reduce or eliminate the bias in algorithmic decision making in the workplace.

Autonomy, agency, and dignity

The use of AI systems can make jobs more interesting and free up time for creative tasks, but there is a risk that systematic management through AI systems and automated decision-making reduces space for workers' autonomy and agency to the point where workers are deprived of dignity in their work. Evidence on warehouse workers managed by AI systems, for example, shows that these workers are often denied the ability to make marginal decisions about how to execute their work, or even how to move their own limbs (Briône, 2020[65]). Devices used in some call centres give feedback to employees on the strength of their emotions to alert them of the need to calm down (Briône, 2020[65]). Other industries – including consultancy, banking, hotels and of course platforms work – are also adopting software that enables continuous real-time performance reviews either by managers or by clients (Wood, 2021[20]).

Algorithmic performance monitoring can push workers towards hitting the targets assessed by the algorithm, but systematically neglect performance factors that are not tracked, developing a compliance mind set, which may ultimately marginalise human-sense decision-making (Leicht-Deobald et al., 2019[66]).. For example, despite the many shortcomings that AI medical diagnosis systems still have, doctors often face pressure to follow the system advice to meet efficiency parameters used in healthcare systems (Neff, McGrath and Prakash, 2020[57]). This could also be the case for managers using AI systems to evaluate their staff. Though these practices are quite rare at present, and evidence on them is therefore still preliminary, if taken to the extreme, AI evaluation systems can generate a sense of alienation and decrease employees' engagement with work (Maltseva, 2020[67]; Fernández-Macias et al., 2018[68]). The lack of transparency and explainability of decisions (see section 2.2) based on AI systems also contributes to reducing workers' agency. For example, not providing explanations for decisions affecting them, does not enable workers to adapt their behaviour in ways to improve their performance (Loi, 2020[69]).

A reduction in autonomy and agency risks reducing workers' capacity to be creative and contribute to innovation. This raises an important societal concern, as creativity and the ability to think or sometimes act out of the box is certainly central in addressing new issues/challenges. Reductions (or further reductions) in autonomy and agency at work may also affect individuals' relationship with work; for many, work is an integral part of finding meaning and purpose in life (Saint-Martin, Inanc and Prinz, 2018[70]; Hegel, 1807[71]; OECD, 2014[72]; Bowie, 1998[73]).

More generally, it's also worth noting that many AI systems, including those used in the workplace, are developed using methods that themselves raise questions about human dignity, and that may be challenging labour rights (Box 5).

Box 5. Labour rights concerns in the development of AI applications

Reporting has highlighted that many of the tasks that underpin modern machine-learning and artificial intelligence systems are contracted out to a "ghost workforce" as coined by Gray and Suri (Gray and Suri, 2019[74]). This workforce completes the myriad of minute tasks that allow workplace AI systems, underpinning for example systems to monitor workers, or assess sections of CVs on a massive scale. The basic tasks include image and text classification or recording transcription, with workers getting paid on demand, per piece of data processed (Irani, 2016[75]). Co-founder of a data labelling factory in Jiaxian, China, Yi Yake affirmed in a 2018 New York Times article that "we're the construction workers in the digital world. Our job is to lay one brick after another. But we play an important role in A.I. Without us, they can't build the skyscrapers" (Yuan, 2018[76]).

While other gig workers such as Uber drivers or DoorDash delivers must be geographically proximate to clients, "ghost work" can be done from anywhere. Many workers reside in OECD countries, but reporting has highlighted how the AI tools provided by a number of major companies such as Amazon and Google significantly rely on the efforts of workers in a number of non-OECD countries, including Venezuela, India or Indonesia, with less stringent labour standards (Fang, 2019[77]; Gent, 2019[78]; Irani, 2016[75]).

This opens up ethical concerns that the AI software used in workplace settings may have been created through the exploitation of vulnerable workers (Milan and Trere, 2017[79]), with some calling for a "digital decolonisation" (Casilli, 2017[80]). This is compounded by the relative lack of input from lower-income countries to the development of international AI principles (Kak, 2020[81]; Garcia, 2021[82]).

Some companies have moved into this space with the express aim of avoiding the exploitation of workers. There is an opportunity in "ghost work" (notably due to the inherent geographic flexibility), to try to lift workers out of poverty. Sama, for example, is a company that provides data-training services for AI, but, unlike some other companies, it hires its employees outright, pays the local living wage, and provides training for employees (Gino and Staats, 2012[83]).

2.2. Transparency and explainability

There are three portions of the transparency and explainability principle (Box 1) that are particularly relevant in the workplace: (1) whether or not workers, managers and other stakeholders are aware of their interactions with AI systems; (2) whether or not those affected by an AI system can understand the outcome; and (3) whether those adversely affected by an AI system can challenge the outcome.

Awareness of interactions with AI systems in the workplace

Informing individuals of their interactions with AI systems in the workplace is a fundamental element of ensuring transparency in AI system use – yet workers are not always aware that they are being hired, monitored or managed via AI. Employees may not know, for example, the kinds of monitoring software – whether AI or not – placed on corporate computers (West, 2021[84]). The rise of remote work during the COVID-19 pandemic has led to a rise in at-distance monitoring, often facilitated by AI, and often unbeknownst to employees (Morrison, 2020[85]). Monitoring may be used at unexpected or not-agreed-upon moments: in a 2016 survey by Quickbooks, 45% of employees who had been tracked using GPS suspected being tracked 24 hours a day. 20% said that their tracking was switched on without warning (Quickbooks, 2016[86]). Similarly, job applicants are not necessarily informed that their data,

including their CVs, video applications, or even data taken from social media are used by AI hiring algorithms to screen their application (Heilweil, 2019[87]).

There is also a risk of "function creep" whereby data collected for one purpose is used for another. For example, data collected to measure productivity and provide advice to the employee for self-improvement could instead be used to decide on promotion or pay increase, or even firing or contract renewal. This practice would go against the data protection principles of 'purpose specification' and 'use limitation' (Moore, 2020[88]).

Although surveillance, data protection and consent are regulated in the EU, most AI tools are still developed by companies based in the United States or other countries with looser regulation. Even in countries where employers are supposed to obtain employees' consent for the use of their personal data, the imbalance in the employer/employee relationship means that workers may well find it difficult to deny their employer consent to data processing for fear of or real risk of detrimental effects (Data Protection Working Party, 2017[89]), thus questioning the very notion of free consent in the employment relationship. Consent to data collection by a worker is not equivalent to the consent to data collection from a customer. Indeed, 'freely given', can only exist in a situation where a data subject has an authentic say and a real choice (Moore, 2020[88]), which is difficult in an employment relationship. In the United States, for example, employees refusing to take part in AI-led wellness programmes sometimes face termination of their employer-provided health insurance (Ajunwa, Crawford and Schultz, 2017[24]). In one case, a woman sued her employer after allegedly being fired after she uninstalled the application that tracked her movements all day every day (Kravets, 2015[90]). Moore (2020[88]) also documents how workers are often asked to provide consent when they sign their contract (and there was no choice but to consent), with consent also often provided technically, i.e. when workers log in to the systems they use to work.

Ability to trust and understand the outcome of AI systems

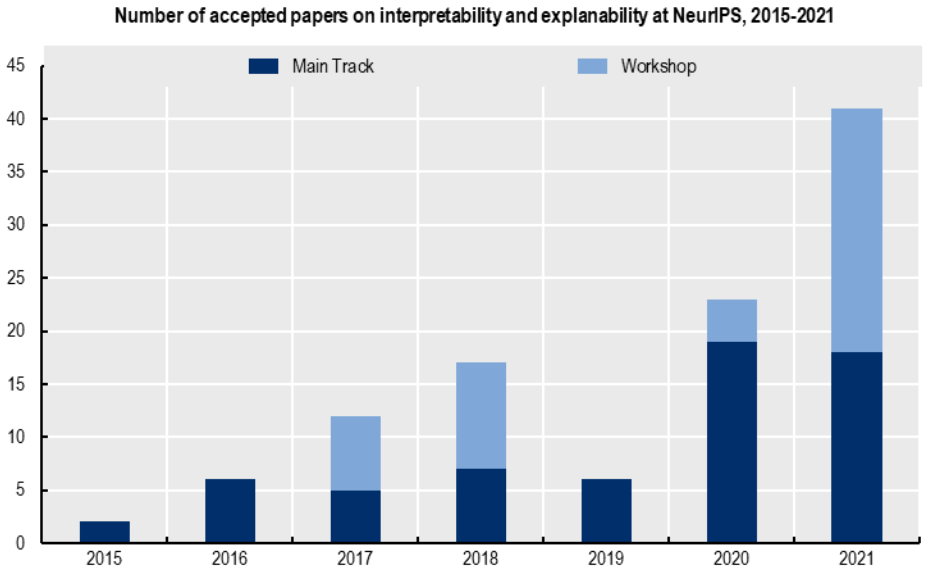
Many AI systems are complex and their outcomes are difficult to explain. It is difficult to get insight into how they work, especially those based on Machine Learning (Adadi and Berrada, 2018[91]). The more complex AI systems are, the more obscure they become. In some kinds of machine learning algorithms, the processes developed by the algorithm to generate certain results cannot even be explained by their developers (Jaume-Palasí and Spielkamp, 2017[92]). In some contexts – for example choices about booking trades in the stock exchange – limited to no explainability can be accepted and statistical probabilities sufficient. But in the workplace, to ensure that principles for trustworthy AI are upheld, both workers and employers should have understandable explanations as to why certain important decisions are being made, such as those that affect wellbeing, the working environment/conditions, or one's ability to make a living. This should apply to decisions based on AI systems of course, with the added factor that AI decisions that are not explainable are unlikely to be accepted by employees (Cappelli, Tambe and Yakubovich, 2019[93]). For example, without an explainability mechanism, workers and managers would not be able to understand why an AI system is recommending a pay rise or bonus for one worker but not another.

Transparency and explainability can be achieved in a number of ways (OECD, 2019[94]), **and are required in some jurisdictions** (Selbst and Powles, 2017[95]). The GDPR for example requires data subjects to be provided with "meaningful information about the logic involved" in automated decision-making processes (see section 3.1 on Data Protection). This often starts by providing "optimisation transparency," i.e. information about what the system's objectives are, and what it has been "optimised" to do. Beyond that, explainability means breaking down and avoiding black boxes (e.g. neural networks), using models that humans can understand, or using appropriate explanatory tools, such as a simple algorithm that approximates the behaviour of the AI system and thus provides approximate explanation. Transparency and explainability do not necessarily require an overview of the full decision-making process, but can be confined to either human-interpretable information about the main or determinant factors in an outcome, or information about what would happen in a counterfactual (Doshi-Velez et al., 2017[96]). One

example is quantifying the influence of different inputs into an HR decision: for example if an employee is refused a promotion based on an AI system recommendation, information can be given on what factors affected the decision, whether they affect it positively or negatively and what their respective weights are. Alternatively, counterfactual models would provide a list of the most important features that the employee would need to possess in order to obtain the desired outcome – e.g. "you would have obtained the position if you had had a better level of English and at least 3 additional years of experience in your present role" (Loi, 2020[69]). If explainability were sufficiently improved, in some contexts, AI systems could offer workers more systematic explanations about employment decisions than traditional workplace processes.

A number of tools have been developed to help improve explainability of Artificial Intelligence systems. Google Cloud's Explainable AI service, for example, helps developers understand outcomes of models they build, by providing predictions and scores that assess how much a factor affects a final outcome (Google, 2022[97]; OECD, 2021[13]). Other options are open-source, including IBM's AI explainability 360 Toolkit and Microsoft's Interpret ML, or unaffiliated libraries and systems such as Alibi Explain (which allows machine learning model inspection and interpretation), or RFEX (which aims to enhance explainability in the Random Forest method) (Klaise et al., 2021[98]; Barlaskar and Petkovic, 2018[99]; Petkovic, 2019[100]). Research into interpretability and explainability has grown significantly – as highlighted in the 2022 AI Index Report (Zhang et al., 2022[101]), accepted papers on those two topics at NeurIPS, a large AI conference, grew from 2 papers in 2015 to over 40 in 2021 (see Figure 6).

Figure 6. AI Research increasingly looking at interpretability and explainability of systems



Source: AI Index 2021, NeurIPS 2021

But while it may be feasible to resolve technical opacity, implementing transparency and explainability may remain complicated (Jarrahi et al., 2021[102]). In practice, many companies developing algorithms remain secretive about what their algorithms are doing, often on grounds that the source code is proprietary information (Briône, 2020[65]), and they are also reluctant to provide the information that would be necessary to assess how their model actually work and whether they provide biased results (Crawford et al., 2019[42]). Questions of intellectual property (IP) involving AI systems can indeed be complex, since they may apply to training data, AI algorithms, model-trained algorithms, as well as the outputs and insights produced (Anwer, 2021[103]). Still, existing case law suggests that there are a number of ways that explainability could be improved without completely dismantling IP rights for AI system

producers or users. Some employment-related issues may rise to be issues of public interest, where IP rights have been waived in the past. In other cases, disclosure could be required in a limited scope, or in full scope for a limited audience (Meyer, 2019[104]). Similarly to food manufacturers, there could also be a move in some cases to require disclosure of the "ingredients" as opposed to the "recipe" behind an AI system – the system would be required to disclose what factors are being considered, but not, for example, the relative weight of those factors (Chmielinski and Grennan, 2021[105]).

Small companies are lagging behind in terms of adoption of AI, and are more likely to use off-the-shelf AI systems (OECD, 2021[106]) which can offer even less control over the design, development or application of the AI, exposing users to an even higher barrier to accessing fully transparent and explainable AI (BCG Henderson Institute, 2018[107]).

Another challenge to explainability is that many managers and workers may only have limited experience with AI, and may not have the skills to understand what the AI applications are doing or how they are doing it. A recent survey found that 46% of respondents in the UK and 43% of respondents in the US admitted that "they have no idea what AI is about" (Sharma, 2017[108]). Equipping workers and managers alike with better knowledge and skills about AI would help facilitate explainability (OECD, 2019[94]), and would involve strengthening adult learning systems (Verhagen, 2021[14]; OECD, 2019[109]). Involvement of experts could help ensure employers and employees alike understand how AI systems work or would work in their workplace; German work councils can now call in an expert to evaluate AI systems (BMAS, 2021[110]).

Ability to rectify outcomes

Even when employees are aware that data on them are used by AI systems, they do not always have rights relative to the access or to the use of the data. For example, workers may not have the possibility to exclude others from seeing or using the data, or a recourse if the data is incorrect and used in adverse employment action. In some jurisdictions, employees do not have the right to contest the data used to make decisions, and the application of AI to these decisions may not necessarily grant the individual with the right to contest or rectify these outcomes.

Without being able to understand the logic of the argument about an AI's employment-related decision, it can be hard to rectify the outcomes of such decisions when needed. The 2019 AI Now Report documents the case of the Children's Hospital of Philadelphia (CHOP), which used an AI algorithm to set productivity rates for the contract workers hired to assemble and distribute supplies like syringes, gauze, and other essential equipment (Crawford et al., 2019[42]). The productivity rate was part of their contractual agreement and was enforced through the algorithm, instead of through on-site supervisors. Part of the challenge in rectifying outcomes was that it was never clear to contract workers how they were being assessed (Feliciano Reyes, 2019[111]), which made it difficult for workers to push back on some employment-related decisions. The increasing "fissuring" of the workplace and the turn towards contractors lead to an increased need for surveillance and may be spreading a no-recourse type of algorithmic management (OECD, 2021[112]; Weil, 2014[113]).

The recent rise in distanced work (in part due to the COVID-19 pandemic) may also be contributing to more automated AI-based employment decisions. In August 2021, Russia-based company Xsolla fired 150 employees by email following an AI-based productivity audit, with no prior warning, human intermediation or human contextualisation of individual decisions (Echarri, 2021[114]). The CEO's email announcing the decision highlighted that an AI system (of which staff were not aware) made the decision to fire 147 people whose productivity had dropped after a move to remote work. The decision was based off of data on how who was contributing more or less to virtual meetings, or spending time on email. Following media outcry, the CEO later stated that Xsolla would review complaints from dismissed staff, and that some may be retained (Batchelor, 2021[115]).

Non-standard forms of work, such as platform work, often leave even fewer options for recourse following AI management decisions. Amazon Flex for example, the gig-style delivery service Amazon uses as part of its delivery network, uses a broad range of algorithms to decide which drivers are given more deliveries, and which are removed from the platform. A number of these contractors have found the rating system impenetrable, if not downright whimsical – and official systems for recourse difficult to navigate. Seemingly genuine reasons for late deliveries are not accepted, and decisions about termination have gone unresolved (Soper, 2021[116]). Similar complaints have also been made against Uber Eats in the UK, over an AI-based facial identification software. Uber Eats workers in the UK are required to have their faces scanned and identified at the start of their shifts – yet many BAME (Black, Asian and Minority Ethnic) couriers have claimed that the face-scanning technology failed for them (see Section 2.1), leading to a dismissal from the application in less than 24 hours. Recourse options were limited, and did not allow for the possibility that the technology itself had made a mistake (Kersley, 2021[117]). These platforms' frequent use of AI to manage workers can therefore be combined with a less traditional work relationship and under-representation by unions and employers' organisations (OECD, 2019[118]), to make recourse for AI-generated decisions more difficult to contest or rectify (De Stefano, 2018[119]).

2.3. Robustness, security and safety

A trustworthy use of AI in the workplace calls for AI that protects the physical safety and the well-being of workers. Artificial Intelligence vastly expands the opportunities for automated systems to be present on the physical or digital work floor in a range of industries, which has significant potential to contribute to increased physical safety for workers. An EU OSHA policy brief from 2021 highlights that use of robots with embedded AI can "[remove] workers from hazardous situations" and improves "the quality of work by handing repetitive tasks to fast, accurate and tireless machines" (EU-OSHA, 2021[120]). To give one example, the rise of artificial intelligence and notably computer vision has allowed sophisticated trash-sorting robots to be deployed to recycling plants – at present, recycling workers face some of the highest risk of workplace incident (Nelson, 2018[121]). Indeed, in 2020, the Waste Management and Remediation Services industry had a fatal injury rate of 15 per 100,000 full-time-equivalent workers, or about 5 times the average rate (US BLS, 2021[122]).

AI systems could also help with more general occupational safety and health. Monitoring systems can help alert workers who may be at risk of stepping too close to dangerous equipment, for example, or that may not be following safety procedures (Wiggers, 2021[123]). Emotion AI systems are also being developed to detect non-verbal cues, including body language, facial expressions and tone of voice: these systems can be used in the workplace to detect workers who are overworked and those whose mental well-being is at risk (Condie and Dayton, 2020[124]). For example, the South China Morning Post reported that train drivers on the Beijing-Shanghai high-speed rail line were wearing "brain monitoring devices" created by Deayea, inserted into their caps. Deayea claims these devices measure different types of brain activities, including fatigue and attention loss, with an accuracy of more than 90%. If the driver falls asleep, the cap triggers an alarm (Chen, 2018[125]).

But if not implemented well, these AI systems can threaten the physical safety and the well-being of workers, thus presenting an ethical challenge. AI-led telematics systems used to monitor and manage delivery drivers are often introduced with the declared intention to increase drivers' safety, but put such pressure on drivers to "beat their time" that the resulting work intensification actually leads to the opposite result. For example, UPS verifies that drivers wear their seatbelt but, in reality, many drivers buckle their seatbelts behind them to maximize delivery speed (Kaplan, 2015[126]). Some systems penalized drivers for looking at a side mirror, or being cut off by another car in traffic (Kaori Gurley, 2021[127]). In some warehouses, wearable devices make possible the full implementation of lean logistics and just-in-time work practices, continuously communicating picking targets and scoring employees.

Combined with the threat of layoff, this can generate heightened stress and physical burnout (Moore, 2018[128]) to a point where they violate human dignity and safety, thus becoming an ethical concern.

Increased AI use in the workplace may also negatively affect human wellbeing due to increased human machine interaction (HMI), and decreased human contact. Some experiments suggest this could, for example, lead to more selfish behaviour on the part of workers involved, though it would depend on the design of the systems involved (Christakis, 2019[129]). Increased individualisation of workforce management, delegation of management and assistance functions to machines also risks reducing the social dimension of work and generate an isolation feeling among workers (Nguyen and Mateescu, 2019[17]).

The implementation of software – whether AI or other – raises the risk that software components may have bugs which could pose a serious threat in high-risk situations in the workplace. Particular risks can be created by machine learning if the situation of implementation is different enough from the situation in which the learning took place. In addition, there is the threat of adversarial data poisoning, whereby attacks on a training dataset threaten a model's integrity, and could lead to serious potential consequences for workers' safety (Müller, Kowatsch and Böttinger, 2020[130]). Robots' increased mobility and decision-making autonomy thanks to AI can make it more difficult for their human co-workers to anticipate the robots' actions, and keep themselves safe (Jansen et al., 2018[131]; Knight, 2021[132]). Overall, AI systems also change the scope of the potential impact on workers, since with their continuous evolution and widespread use, a design flaw may suddenly and simultaneously affect every one of a company's workers' safety and well-being.

The use of AI in the workplace can also contribute to heightened digital security risks. First, AI has expanded the kinds of systems connected to networks. For example, AI has enabled some mechanical arms to operate independently when connected to a cloud network. The connection to the cloud network however means that the consequences of a digital security issue may have significant implications for workers' physical security in proximity to the arm (Knight, 2021[132]).

Further, the wealth of data collected and generated (derived data) by AI tools also poses an additional security risk (Euronews, 2021[133]): it vastly expands the information on employees available to hackers in a cybersecurity breach (Deloitte, 2018[134]). With AI applications now able to collect and compile extensive browsing histories, geolocation data, and up-to-date health and wellness data (to name a few), the possible harm to employees due to a cybersecurity breach increases substantially (Strategic Foresight Unit, 2020[135]). During COVID-19, AI-powered applications extended to the collection of data in conventionally private spaces in the home (Euronews, 2021[133]). There are even concerns about AI tools implemented in workplaces in an attempt to heighten cybersecurity. For example, some AI workplace applications are now able to build a biometric profile of users based on the unique way users move their mouse (BehavioSec, 2020[136]), intended to strengthen identification requirements. Due to the significant monitoring and surveillance required to develop these profiles and continuously authenticate users, experts have highlighted the serious risks for employees should these biometric profiles themselves be hacked (Taddeo, 2019[137]). These risks are heightened by the fact that employees may have no legal recourse in response to a data breach at their place of work (Reid, 2021[138]; Cater and Heikkilä, 2021[139]), compounding concerns about data breaches (see Section 2.1).

2.4. Accountability

Accountability in the workplace refers to responsibility both that the AI system is designed and implemented correctly, and that outcomes of the AI system align with the other OECD AI principles (OECD, 2019[12]). Accountability relies on being able to tie a specific individual or organization to the responsible use of a specific AI system. In practice, AI systems pose challenges, since it is not inherently clear which actor linked to the AI system is responsible if something goes wrong – notably with complex

procurement arrangements. Is it the programmers, the developers or the firm/organisation using the AI algorithm? Programmers are natural persons (most likely with limited financial capacity), and as time goes on, would have less and less of an understanding of the machine learning algorithm is doing exactly. Besides, one important challenge to transparency and accountability for developers lies in the lack of reliability and reproducibility outside the lab (Loi, 2020[69]). Research shows that there is no guarantee that algorithms will achieve their intended goal when applied to new cases, in a new context, with new data (Neff, McGrath and Prakash, 2020[57]) (Heavens, 2020[140]). This is generally due to a mismatch between the data on which AI systems were trained and tested and the data they encounter in the world, a problem known as data shift. Another problem undermining the reliability of algorithms is "underspecification," a known issue in statistics, where observed effects can have many possible causes (Heavens, 2020[140]). Looking at a range of different AI applications, from image recognition to natural language processing (NLP) to disease prediction, Google researchers found that underspecification was to blame for poor performance in the real world in all of them, and the problem cannot be easily fixed. This is presumably the difficulty met in scaling AI developed in a given place of work across multiple sites of work (Neff, McGrath and Prakash, 2020[57]).

Without developers/designers involved, most responsibility in the workplace context would shift to employers, with court cases already arising in AI-based decisions about hiring (Maurer, 2021[141]; Engler, 2021[142]; Butler and White, 2021[143]) **and performance management** (Wisenberg Brin, 2021[144]). Indeed, existing occupational safety and health regulations in a number of countries may require employers to pre-emptively ensure that tools used in the workplace will not harm workers, including through rigorous pre-use risk assessments (ILO, 2011[145]). Accountability in any harm later incurred by workers could therefore fall with the employer, though there is some uncertainty about whether psychosocial risks posed by AI systems, for example, are appropriately covered by OSH regulations (Nurski, 2021[146]). Still, if the tendency towards employer responsibility continues, it could imply high litigation costs for businesses and increasing uncertainty and risks for employers (see Section 3.1). SMEs for example may simply not be able to support such costs.

Approaches to accountability in the context of automated decision-making processes often lie with having a human "in the loop" (e.g. they may have to approve a decision) or "on the loop" (e.g. they are able to view and check the decisions being made) who would take responsibility in the case of a poor decision (Enarsson, Enqvist and Naarttijärvi, 2021[147]). A significant question for workplaces is therefore deciding which kinds of decisions based on or made by AI systems will need a human involved to ensure an appropriate level of accountability.[8] Choices that have an important impact on individuals' lives are generally considered to need human involvement on top of an AI decision to be effectively accountable, including such choices in the workplace (OECD, 2019[94]).

An additional element of accountability lies in auditability. A number of firms are beginning to conduct audits to ensure that algorithms and AI systems are trustworthy. In the workplace, these audits have especially been concerned with discrimination (Wilson et al., 2021[148]), frequently seen as a pre-emptive move against legal action (Engler, 2021[142]; Brown, Davidovic and Hasan, 2021[149]), or in anticipation of regulation (O'Keefe, Moss and Martinez, 2020[150]). There are however a number of pre-requisites that AI audits need to satisfy in order to ensure accountability (Box 6). Furthermore, not all AI systems, however, are effectively auditable, especially if companies do not provide enough access and independence to auditors. Without a degree of auditability, responsibility for any errors becomes more difficult to ascertain.

[8] Discussions about legal personality for AI systems are ongoing in a number of countries. Australia and South Africa both recently designated AI systems as "inventors" in patent applications (Jones, 2021[282]). While legal personality is seemingly feasible in a number of countries (Allgrove, 2004[280]), it is unclear whether countries (or constituents) will seek to actually bestow such rights (Chesterman, 2020[281]), particularly when thinking about accountability.

Box 6. AI and algorithm audits

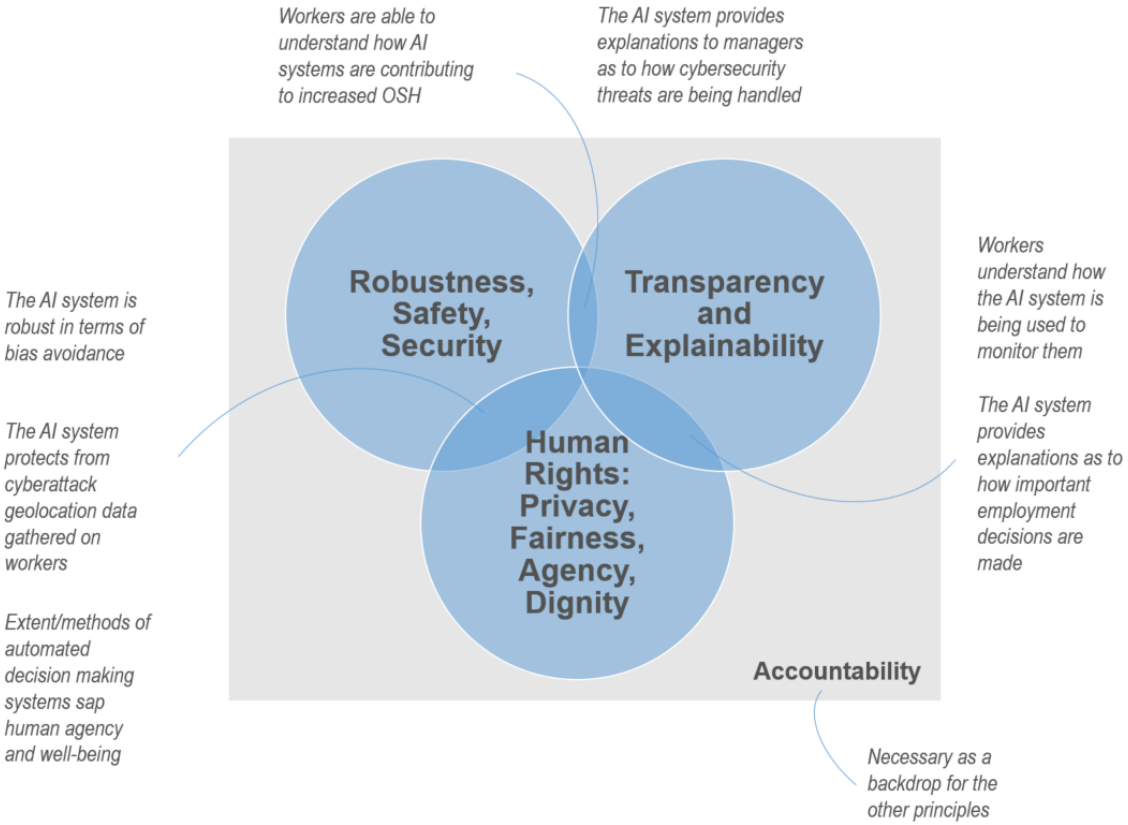
"AI auditing" or "algorithmic auditing" has become increasingly popular tool to assess AI systems, and ensure they follow the law and/or principles of trustworthiness. In 2016, the U.S. White House released "Big Data: A report on Algorithmic Systems, Opportunity, and Civil Rights." One of the recommendations of the report was to "Promote academic research and industry development of algorithmic auditing and external testing of big data systems to ensure that people are being treated fairly" (Executive Office of the President of the United States, 2016[40]). Additional actors in the United States have proposed or enacted legislation that would require external algorithm audits, including in New York City, in California, and at the federal level (Johnson, 2021[151]) (see Section 3.3).

How algorithmic audits should be conducted to ensure they contribute to trustworthy AI is still an area of active research (Ada Lovelace Instititute, 2020[152]; Brown, Davidovic and Hasan, 2021[153]). Generally speaking, in an algorithmic audit, a third-party assesses to what extent – and why – an algorithm, AI system and/or the context of their use aligns with ethical principles or regulation.

Key to making these audits trustworthy is ensuring that certain prerequisites are met, as detailed by Engler (2021[154]). According to a number of experts, prerequisites can include auditor independence; representative analysis; data, code and model access; and consideration of adversarial actions (verification should be conducted as much as possible that provided data has not been manipulated) (Wilson et al., 2021[148]; Engler, 2021[154]). Audits, particularly of algorithms used in the workplace, will need to be transparent, with reports accessible to workers if not the public, and with careful engagement of key stakeholders, such as unions (Colclough, 2021[155]). Narrow or weak standards for auditing run the risk of providing "checkbox certification," (Whittaker, 2020[156]). Audits therefore require important engineering involvement, both in terms of the careful process of interpreting concepts such as "fairness" into engineering practice (see discussion about the challenges in terms of Fairness, bias and discrimination in Section 2.1), as well as building algorithmic infrastructures that allow for auditing, particularly for high-risk applications (Koshiyama, Kazim and Treleaven, 2021[157]). As policy makers consider algorithmic auditing as a tool to promote fairness, transparency and accountability for AI systems, it will be important to agree on the prerequisites and standards that audits should meet..

Overall, accountability is important as a foundation for consideration of the other potential ethical concerns about the use of AI in the workplace. Without clear accountability, no actor may understand that they are responsible for upholding anti-discrimination principles, for example, or for ensuring that AI systems operate safely. If no one is responsible when AI systems do not work as they are reasonably expected to, transparency about the problems in the AI system will not necessarily translate into process improvements (Loi, 2020[69]). Providing clear accountability is therefore important to addressing other ethical concerns. While particularly true for accountability, all ethical concerns presented in this chapter are in fact interwoven, interdependent and inter-reinforcing (Figure 7).

Figure 7. Potential ethical concerns about AI use in the workplace are interlinked



Note: Schematic approximation of interactions between areas of potential concern identified. Annotations in the margin provide context or examples

3. Overview of policy measures to ensure the trustworthy use of AI in the workplace

Policy is important to ensure a trustworthy use of AI systems in the workplace, and to address the ethical concerns outlined in Chapter 2. Action is needed to facilitate the diffusion of AI and prevent undesirable consequences for workers, employers, and society as a whole (OECD, 2019[12]). Buy-in by workers, management, social partners and society as a whole will be essential to securing the opportunities AI systems present – but without resolving ethical concerns, trust will be difficultly earned.

This chapter reviews the main relevant policy measures being used, building on the EC-OECD database of national strategies and policies and on qualitative interviews conducted in 2021 with relevant contact points from national delegations, stakeholders and international organisations.[9] This chapter focuses on policy developments in OECD member countries, but relevant policy action is taking place in other countries as well.

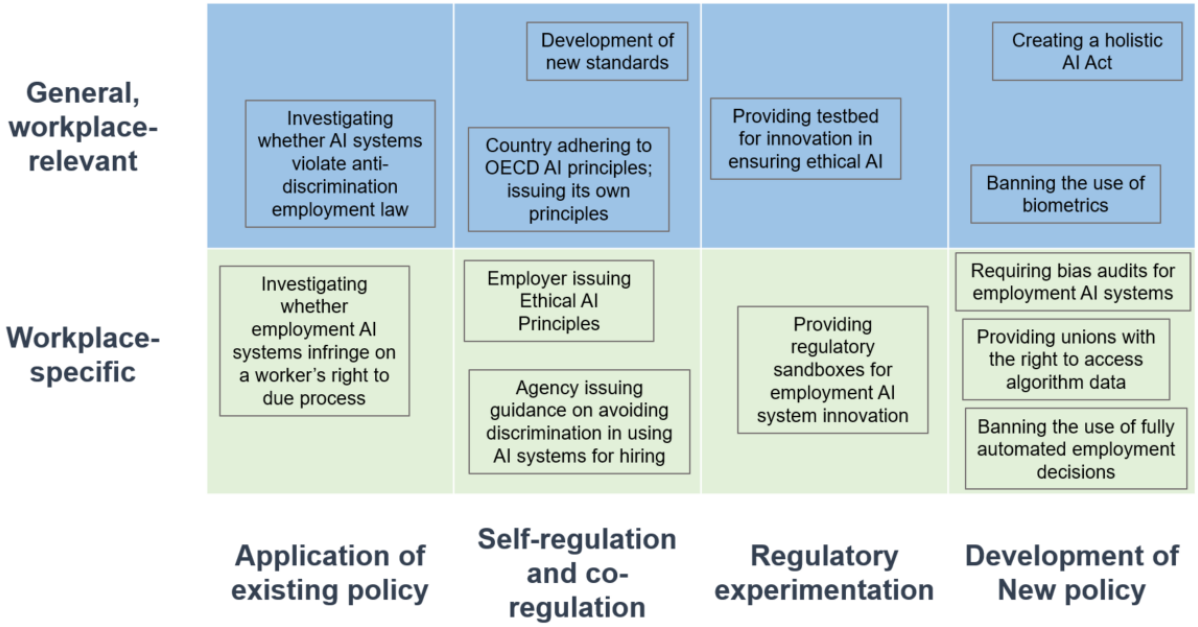
The range of policy approaches being used is quite broad (Figure 8). Potential measures range from the application of existing policies, to self- and co-regulation, to regulatory experimentation via sandboxes or test-beds, to the development of new policy (OECD, 2018[158]). A "multi-tier" approach, using different types of measures at the same time, can help address gaps in governmental regulation (Hunkenschroer and Luetge, 2022[159]). Relevant policy include measures specific to the workplace as well as general ones (OECD, 2021[160]) and can be targeted at one or more of the ethical concerns outlined in this report. The policy measures taken are typically aligned with the level of risk the relevant ethical concern represents (Clement-Jones, 2020[161]). The approaches also vary in terms of their coverage, from municipal, to national and international approaches. There are concerns that multiplication of standards and regulations may increase compliance costs for businesses (Candelon et al., 2021[162]). In addition, the application and development of policy areas may pose challenges for their enforcement if there is overlap, for example in terms of protecting privacy rights and anti-discrimination rights (Kim and Bodie, 2021[43]).

Existing legislation provides a strong foundation for addressing a number of ethical concerns about the use of AI in the workplace. Section 3.1 looks at how legislation on data protection, on anti-discrimination, on deceptive practices and consumer protection, and on rights to due process is increasingly being used to ensure an trustworthy use of AI in the workplace. In addition, a number of countries are in the process of developing society-wide policies for trustworthy AI. Section 3.2 provides an

[9] The EC-OECD database of national AI strategies and policies (available at: www.oecd.ai/dashboards) comprises more than 700 AI policy initiatives at various stages of implementation, building on the work of the OECD.AI Network of Experts' reports published in 2021 on the implementation of the OECD AI Principles: *Tools for trustworthy AI* (OECD, 2021[13]) and the State of Implementation of the OECD AI Principles: insights from national AI strategies and policies (OECD, 2021[160]).

overview of these initiatives - including, for example, the EU AI Act proposal – with discussion mainly on the proposed implications for workplace use of AI systems. Finally, Section 3.3 looks at policy measures specific to the use of AI in the workplace that are being considered in some countries, courts or legislatures.

Figure 8. Spectrum of approaches to promote the trustworthy use of AI in the workplace



Note: Boxed entries are only examples of what the specific proposed approach could resemble.

3.1. Existing legislation being applied to ensure a trustworthy use of AI in the workplace

Existing legislation not specifically focused on AI continues to be highly relevant to many of the concerns that arise from the use of AI in the workplace – and in society more broadly (OECD, 2019[12]). Policy and legislation on data protection and privacy, legislation against discrimination and deceptive practices, and policies protecting due process in employment-related decisions are all especially relevant to address the ethical concerns raised by the use of AI systems in the workplace, and have all been successfully applied to related cases in recent years. In addition, **labour law** (for example often regulating conditions about work time, or how employees are notified about firings) **or Occupational Safety and Health (OSH) policies often apply directly to AI use in the workplace**. Labour law often regulates conditions about work time or employee firing notification, for example, while OSH regulations can provide employees with a legal right requiring employers to protect their employees by avoiding risks to safety and health (Nurski, 2021[146]). Overall, as AI systems become more integrated in the workplace, legislation in these areas will likely need to adapt to effectively address concerns raised by the use AI (Jarota, 2021[163]; Kim and Bodie, 2021[164]).

Data Protection

Due to AI's reliance on data, data protection policies often apply to the use of AI in the workplace. The European Union's General Data Protection Regulation (GDPR) – which entered into force in 2018 — is perhaps the best known of such protection principles, enshrining a number of data

rights. The GDPR regulates entities that "process the personal data of EU citizens or residents" or "offer goods or services to such people" regardless of whether such entities are located within the EU (Official Journal of the European Union, 2016[165]). To encourage compliance, the GDPR allows each EU member state's data protection authority (DPA) — the "independent public authorities that supervise" GDPR application — to fine violators the greater of either EUR 20 million "or 4 per cent of the firm's worldwide annual revenue from the preceding financial year." Many of the data protection rights enshrined by GDPR apply to general data gathering and processing technologies, but have specific implications when extended to AI. This is particularly the case for rights to transparent information and communication, as well as rights of access (Art. 12, 13, 15), to rectification, to erasure and to restriction of processing (Art. 16-17). When enforced, these rights should help prevent concerns about workers' right to consent in the workplace for example, or about transparency. In addition, GDPR Article 22 also provides the right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her […]"; this significant threshold pretty **much rules out algorithmic management (see** Box 2**) that entails full automation of decisions in EU countries and that does not have 'meaningful' human input in such decisions** (Wood, 2021[20]).

In addition, in some countries, worker monitoring through digital technologies requires prior agreement with workers' representatives (Aloisi, 2021[166]). In Germany, the monitoring of employees in the workplace is also limited by the General Right of Personality which, coupled with the Data Protection Act, require a justification of the legitimate interest of the employer and permission to process employees' data on the basis of collective agreements. In France, workers have a right to private life and surveillance systems can only be installed with prior consultation of employees' representatives and notice to employees. In Italy, prior agreement with worker representatives is also needed to use monitoring instruments in the workplace unless the instruments are used by the worker to perform the working activity and to record access and presence on the premises.

In the United States, a number of states have laws in place requiring employers to notify employees of electronic monitoring. In May 2022, New York's such law came into effect, for monitoring of employees' phone calls, emails and/or internet usage, and requiring all new employees to acknowledge in writing their understanding of the employers' policy (O'Connor, 2022[167]), and is similar to Delaware's existing legislation (Delaware General Assembly, 2022[168]). Connecticut's 1998 law has a broader definition of electronic communications – requiring advance notice for the collection on any information at the workplace that goes beyond "direct observation" (Wiggin, 1999[169]).

Several national data protection authorities have issued guidance or specific reports on ethical challenges that AI systems raise for data protection regulation, notably for AI use in the workplace. For example, the UK's Information Commissioner's Office (ICO) published a "Guidance" on best practices for auditing AI systems and compliance with existing data protections. It includes sections on (1) accountability, (2) lawfulness, fairness and transparency, (3) data minimisation and security, and (4) data subject rights. As applied to workplace ethical issues, the guidance lays out a number of responsibilities for businesses, including ensuring the security and privacy of workers' data, avoiding discrimination, having meaningful human review of solely automated decision-making processes, and generally complying with individuals' right requests (ICO, 2020[170]). Recently, the UK's Information Commissioner's Office (ICO) also started a probe on a bank over allegations that the bank spied on its staff, using an AI-based app monitoring their computer usage to measure their productivity. In announcing the probe, the ICO restated that organisations "need to make employees aware of the nature, extent and reasons for any monitoring" (Singh, 2020[171]).The Spanish Data Protection Agency (AEPD), the Norwegian Data Protection Authority (Datatilsynet), and the French Data Protection Authority (CNIL) (ADEP, 2020[172]; Datatilsynet, 2018[173]; CNIL, 2017[174]) also produced guidance and reports. Some countries may also modify data protection or privacy laws to explicitly address AI systems, such as Bill C-11 (2020) in Canada, which would require companies, upon request, to provide individuals with explanation of decisions made by AI systems (House

of Commons of Canada, 2020[175]), or in California, where the California Privacy Protection Agency may be charged with creating a proposal on AI by 2023 (Gibson Dunn, 2022[176]).

A number of cases have been brought in EU Courts on the use of AI in the workplace, based on the legal rights accorded by the GDPR.

In March 2021, Amsterdam's District Court ruled that two ride-hailing companies, Uber and Ola, had to disclose the data used in their AI systems (Strauss and Venkataramakrishnan, 2021[177]). Uber's Real-Time ID system and Ola's Guardian systems are used to determining earnings, assign work and, in some cases, suspend drivers. Both companies were ordered to explain their use of driver surveillance systems. In the Ola's Guardian judgment, the Court required the company to explain the logic behind fully automated suspensions and wage penalties decisions and found that workers had been subject to automated decision-making in the context of article 22 of the GDPR, seemingly the first time a court reached such a conclusion.

Another case was brought against Uber by four drivers (3 in the UK and 1 in Portugal) over algorithm-based dismissal. Uber's algorithm deactivated their accounts due to a continued pattern of alleged inappropriate use. When the drivers requested an explanation from the platform, they were told that the algorithm could not explicitly explain what activities were suspicious, only that they were "fraudulent" in nature. The drivers were also unable to appeal the decision to deactivate their accounts. The lawsuit was filed in the Netherlands arguing a violation of article 22 of the GDPR. The Court's response was to order Uber to reinstate the drivers as its decisions were based solely on the automation of the process (Butler, 2021[178]).

European DPAs have also conducted joint operations under GDPR to investigate digital platforms using AI algorithms to manage platform workers across borders (e.g. Italy's GPDP and the Agencia Espanola de Protección de Datos "AEPD" in Spain). In June 2021, the Italian DPA, Garante per la Protezione dei dati personali "GPDP", fined the digital platform Foodinho over EUR 2.6 million for using algorithms to manage its food delivery riders in a manner deemed discriminatory. Among other violations, the company failed to supply transparent information about its reputational rating system for riders. The investigation revealed that the system generated discriminatory ratings that excluded riders from job opportunities (GPDP, 2021[179]). This was the first time that a DPA fined a company for GDPR breaches relating to its algorithmic processing of personal data. Despite these examples, case law interpreting relevant provisions of the GDPR as they pertain to AI systems remains limited.

Anti-Discrimination

A range of laws in OECD countries help protect workers from discrimination. They often place responsibility on employers to avoid discrimination, regardless of the technology (including AI) that is used.

In the United States, the Equal Employment Opportunity Commission (EEOC) enforces a number of laws protecting job applicants and employees from discrimination, and is increasingly examining whether AI systems used for hiring comply with equal opportunity laws. Reporting suggests that the EEOC has begun to investigate at least two cases about algorithmic discrimination in hiring, promotion and other job decisions (Opfer, Penn and Diaz, 2019[180]) and, in 2020, it was encouraged to take more action in this area of work by ten U.S. Senators (Bennet, 2020[181]). In October of 2021, the EEOC launched a formal initiative on AI and algorithmic fairness, which will include more coordinated agency-wide work and the issuance of technical assistance for guidance on algorithmic fairness (U.S. EEOC, 2021[182]). There are however some observed difficulties in applying existing anti-discrimination law to AI cases. For example, plaintiffs may find difficulties in proving "discriminatory motives or intent" of an algorithm and in overcoming an employer's "business needs" defense when the functioning of an algorithm remains unknown (O'Keefe et al., 2019[183]). In addition, a reliance on individual action in the United States for seeking redress could be challenging since many applicants and workers may not even know that AI is

being used to assess them, or may not have the skills and tools necessary to evaluate whether the AI is discriminating against them.

In 2021, an Italian court applied anti-discrimination laws to throw out an algorithm used by digital platform Deliveroo to assign shifts to riders. The case was brought by a trade union, which alleged that the system used by Deliveroo to determine access to work slots was discriminatory**.** The court found that Deliveroo gave priority access to work slots to workers using an algorithm – called "Frank" – which "scored" workers based on reliability and engagement. Reliability decreased if a worker did not log in to the application within 15 minutes of the start of an assigned shift; engagement increased if a worker served many periods during peak hours. The tribunal ruled that the algorithm, by using an unclear data processing method and no possible contextualisation for rankings, indirectly discriminated against workers who had booked a shift but could not work, including if due to personal emergencies, sickness or participation in a strike (Geiger, 2021[184]; Allen QC and Masters, 2021[185]; Tribunale Ordinario di Bologna, 2020[186]). The Tribunal highlighted the transparency problems with the algorithm, and that the algorithm needed to take into account context for the data used in its rankings. Deliveroo discontinued the algorithm in November 2020 but noted that the assessment of the algorithm was based on hypothetical cases and not on concrete examples (Tribunale Ordinario di Bologna, 2020[186]). Existing legislation may not be sufficient to protect workers from discrimination, particularly if enforcement does not adapt to specific challenges that arise from AI use.

Deceptive practices and consumer protection

Legislation against deceptive practices and for consumer protection is also being used to ensure ethical use of AI in the workplace. In the United States, an April 2021 post by the Federal Trade Commission (FTC) highlighted that some AI companies were using unfair or deceptive practices, notably through inaccurate descriptions of products, (possibly unintentional) discriminatory outcomes, and generally a lack of transparency about what their algorithms are doing (Federal Trade Commission, 2021[187]). The post noted that the FTC would use its authority, notably under the FTC Act, to pursue biased algorithms – and that companies were expected to use inclusive datasets, "test [their] algorithm" before and after use, expand transparency, and not exaggerate "what [their] algorithm can do." Following a complaint to the FTC by the Electronic Privacy Information Center, for example, "Hiretech" company HireVue announced in 2020 that it would stop analyzing facial expressions in videos to assess job candidates (Kahn, 2021[188]). The Center's complaint had called such practices 'deceptive,' alleging that while marketed as being more objective than decisions made by human resource managers, HireVue's AI systems' decisions were in fact more likely to be biased . The complaint also noted that, HireVue's AI system's practices "did not meet minimal standards for AI-based decision-making as set out in the OECD AI Principles" (EPIC, 2019[189]). In September 2021, the FTC approved new compulsory process resolutions in eight key enforcement areas, including bias in algorithms and biometrics, to enable more aggressive investigations of conduct and swifter action against companies in the US engaging in any conduct addressed by the resolutions (Khan et al., 2021[190]). Recent settlements have also highlighted that the FTC may be able to require companies to destroy algorithms and models that were trained on improperly collected data (e.g. an algorithm to screen resumes trained on the resumes of current and past employees without informed consent) (Gesser, Rubin and Gressel, 2022[191]).

Right to Due Process

Legal rights to due process are being be used to challenge the use of AI systems at work – especially uses related to decision making processes – when algorithms' opacity makes it difficult for workers to obtain clear explanations about employment decisions, either because of intellectual property or because of general lack of explainability,. For example, in 2017, a federal court in the United States found that the secrecy around the algorithm used to assess school teachers was denying them their right to complain

and challenge potential termination decisions. Between 2011 and 2015, teachers in Houston were evaluated through the Education Value-Added Assessment System (EVAAS), which took information on students' standardised test scores and, with the help of an algorithm, judged the effectiveness of teachers in their work, leading to the dismissal of some teachers (Webb, 2017[192]). The teachers never received an explanation as to how the algorithm worked, or how their scores could be improved. The Court ruled that the teachers' right to due process outweighed the algorithm developer's intellectual property rights, giving teachers the right to independently verify and challenge the algorithm's evaluation if they so deemed. However, the algorithm's developer refused to disclose its programming – as a result the algorithm is no longer used in teacher evaluation in Houston.

In New Zealand, the Employment Relations Act of 2000 (ERA) was used in 2013 to invalidate a decision to dismiss an employee, in part because the decision to dismiss was informed by the results of a psychometric test which the employer could not explain, or even seemingly understand. In Gilbert v Transfield Services Ltd, (Colgan, 2013[193]) a New Zealand court ruled that employees have the right to access information about adverse employment decisions that are being made about them and the right to "an opportunity to comment" before the decision is made (New Zealand Parliamentary Counsel Office, 2000[194]). Since the algorithm information (including whether it was AI per se or a less complex algorithm) was not available even to the employer, and therefore could not be available to the employee, the Court noted the inappropriateness of using the system in a process that requires openness and information exchange" (Colgan, 2013[193]).

Questions of ensuring transparency and explainability for AI systems used in employment decisions will continue to be an issue for due process rights going forward (Gavaghan, Knott and MacLaurin, 2021[195]).

3.2. Policy efforts to promote trustworthy use of AI in economies and societies

Policymakers from national and inter-governmental organisations are exploring different regulatory frameworks to ensure trustworthy AI systems in the economy and society (OECD, 2021[160]). While these frameworks are not specific to the use of AI in the workplace, they are generally relevant for the workplace. Alongside promoting the widespread adoption of AI, they call for policy measures to address concerns raised by AI applications often including the ethical concerns outlined in Chapter 2.

A number of commissions have been convened to propose potential legislative or regulatory changes, at a range of levels. In Germany, the Data Ethics Commission – composed mainly of data protection experts – released a report in 2019 highlighting the urgent need to strengthen the existing legislative framework in a number of areas including for example employee data protection (Daten Ethik Kommission, 2019[196]). A much broader report in 2020, produced by the German AI Inquiry Committee, highlighted the need to "ensur[e] that as social beings humans have the opportunity to interact socially with other humans at their place of work, receive human feedback and see themselves as part of a workforce" (Deutscher Bundestag Enquete-Kommission, 2020[197]). In November 2020, the US Office of Management and Budget (OMB), working with the Office of Science and Technology Policy, the Domestic Policy Council and the National Economic Council, issued guidance to federal agencies on when and how to regulate the private sector use of AI. The guidance focused on a risk-based, cost-benefit approach to AI regulation, prioritising non-regulatory approaches where possible and regulatory impact assessment (OMB, 2020[198]).

Overall, OECD countries' initiatives to promote trustworthy AI are still predominately self-regulatory (non-binding) approaches – or soft law. They include the development of ethical frameworks and guidelines, voluntary processes, technical standards, and codes of conduct. These guidelines are addressed to policy makers, businesses, research institutions and other AI actors. Examples include Australia's AI Ethics Framework; Colombia's Ethics Framework for Artificial Intelligence, Egypt's Charter on Responsible AI; Hungary's AI Ethical Guidelines; Japan's AI R&D Guidelines and AI Utilisation

Guidelines; and Scotland's AI explainability framework. These guidelines, which are being collected and classified in the OECD's database of tools for Trustworthy AI (see Box 1), are largely aligned with the OECD AI Principles (OECD, 2019[12]). In 2021, UNESCO also adopted a recommendation on the ethics of artificial intelligence, with an emphasis on protecting data, banning social scoring and mass surveillance, helping monitoring and evaluation, and protecting the environment (UNESCO, 2021[199])

On top of putting forward ethics-based principles, countries also often wrap measures to promote trustworthy use of AI into "soft legislation" in the form of AI strategies. Spain's National AI Strategy in December 2020 includes an ethics pillar, including notably an impetus for developing a trustworthy AI certification for AI practitioners (La Moncloa, 2020[200]). Germany's Artificial Intelligence Strategy, updated in December 2020, states that AI applications must augment and support human performance (OECD.AI Wonk, 2021[201]). The United Kingdom National AI Strategy, released in September 2021, highlights existing cross-sector legislation as well as regulators that already regulate the development and use of AI – including data protection (enforced by the Information Commissioner's Office, "ICO"), and human rights and equality (enforced by the Equality & Human Rights Commission) (HM Government, 2021[202]).

A number of countries have published strategies for the use of AI in the public sector – an initial mapping in 2019 identified 36 countries with such strategies (Berryhill et al., 2019[203]). For example, the government of New South Wales, Australia, released an AI Strategy to shape its own approach to AI including requirements that all automated decisions by the government would be subject to human review and intervention, and that accountability would always remain with NSW organisations and individuals (NSW Government, 2020[204]). The U.S. Government Accountability Office, an independent legislative branch agency in the US government, developed a framework to help assure accountability and responsible use of AI systems by the federal government. It defines the basic conditions for accountability throughout the entire AI life cycle, laying out the specific questions for leaders and organisations to ask, and the auditing procedures to use when assessing AI systems (GAO, 2021[205]). As a significant consumer of AI systems, governments' decisions about what a trustworthy workplace AI system looks like can help shape the market, as well as serve as a role model for other institutions (Pimentel, 2021[206]).

Countries and stakeholders are also developing standards to support the implementation of trustworthy AI. Australia (Standards Australia, 2020[207]), Germany (DKE, 2020[208]), and the United States (NIST, 2020[209]) standards authorities are all working towards such standards. The National Institute of Standards and Technology (NIST) in the United States, for example, is establishing benchmarks to evaluate AI technologies, as well as leading and participating in the development of technical AI standards. In March 2022, the NIST issued an initial draft of the AI Risk Management Framework, with a focus on technical characteristics (Accuracy, Reliability, Robustness, Resilience), socio-technical characteristics (explainability, interpretability, privacy, safety and managing bias), and guiding principles (fairness, accountability and transparency) (NIST, 2022[210]). The NIST also issued a publication focused on AI and bias, building on a 2021 proposal for identifying bias across the AI lifecycle (NIST, 2021[211]), but also noting the importance of addressing human and systemic biases (Schwartz et al., 2022[212]). The creation of this framework – which will take the form of a guidance document for voluntary use – was mandated by legislation (U.S. Congress, 2021[213]). AI standards to support trustworthy use have also been the focus of international cooperation, as set out in the EU-US Trade and Technology Council Inaugural Joint Statement (European Commission, 2021[214]), or an initiative by the UK via the Alan Turning Institute to establish global AI standards (Alan Turing Institute, 2022[215]). Several AI standards, both cross-sector and sector-specific, are under development or are being published, including those developed by organisations such as the International Organization for Standardization (ISO) and the Institute of Electrical and Electronics Engineers (IEEE).

Some institutions are calling for society-wide but use-specific bans of some AI technologies, in particular AI-powered facial processing technologies (see Box 4**)**. A 2021 report by the United Nations Human Rights Office called for a temporary ban on the use of facial recognition (UN Human Rights Council, 2021[216]). In its 2021 guidelines on how European countries should regulate the processing of biometric

data, the Council of Europe called on European countries to impose a strict ban on facial analysis tools that purport to "detect personality traits, inner feelings, mental health or workers' engagement from face images" (Council of Europe, 2021[217]).

Finally, a number of OECD countries are considering society-wide AI legislative proposals. In particular, European Union and the United States are considering large-scale legislation that would have significant impact on the use of AI, including in the workplace.

European Union AI Act and other EU proposals

The European Union's AI Act represents an ambitious contribution to the AI policy conversation, with concrete implications for use in the workplace. In April 2021, the European Commission ("EC") released an AI package including the EC review of the Coordinated Plan and the proposal for an AI Act to enable an "ecosystem of trust" on AI in Europe (European Commission, 2021[218]). The proposal may still face significant changes, as it makes its way through negotiations with the EU Parliament and Council. As of late 2021, the proposed AI Regulation would govern the "*development, placement on the market and use of AI systems"* in the EU following a horizontal and risk-based regulatory approach that differentiates between uses of AI that generate i) minimal risk; ii) low risk; iii) high risk, subject to specific safeguards; and iv) unacceptable risk, proposing a strict ban on the latter. It would have implications for the entire lifecycle of AI systems – from development to deployment and use – and would apply across both public and private sectors. The risk-based approach has drawn praise (Ebers et al., 2021[219]; DOT Europe, 2021[220]; Veale and Borgesius, 2021[221]), although there is debate about what should fall under each category (Johnson, 2021[222]). Critics also note that the risk categorisation of AI applications established in the proposal is not easily modifiable, which makes it difficult to adjust to developments in the field of AI (Lomas, 2021[223]; Circiumaru, 2021[224]).

Some AI systems that could be used in the workplace would be classified as "unacceptable risk" by the proposed AI Act, notably those that relate to manipulation, or distorting the behavior of a person based on "subliminal techniques" or by exploiting the vulnerabilities of a specific group of persons (European Commission, 2021[218]). Still, in a public letter, 114 civil society organizations called for the banning of a broader set of applications, asking for a shift in the threshold definition of manipulation as well as the inclusion of a broader set of vulnerabilities. In addition, the letter called for more practices to be defined as unacceptable risk, such as discriminatory biometric categorization or emotion recognition systems (European Digital Rights (EDRi) et al., 2021[225]), as also echoed by a joint opinion by the European Data Protection Board, the European Data Protection Supervisor and several national DPAs (EDPB-EDPS, 2021[226]).

In addition, the proposed AI Act classifies as "high risk" all AI systems used in "employment, workers management and access to self-employment," notably for recruitment, decisions about promotion, firing and task assignment, and monitoring of persons in work-related contractual relationships (European Commission, 2021[227]). The categorization in high-risk uses implies that AI systems used in the workplace would be subject to legal requirements relating to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security. The World Employment Confederation – Europe criticized the decision to include all employment-related decisions as high-risk, suggesting that solely hiring-related decisions should be subjected to that level of scrutiny, a broader grouping risks "stifling innovation," including potential innovative ways to reduce bias (WEC-Europe, 2021[228]).

Organizations that develop, sell or use AI systems would be subject to legal compliance requirements enforced with fines for non-compliance, based on a company's total worldwide annual turnover the preceding financial year, and varying based on the type of non-compliance. Still, most of the regulatory burden seems to be placed on "providers" of AI systems, in line with part of the Act's legal heritage from product safety regulation, as the Ada Lovelace Institute notes (Circiumaru, 2021[224]). This

somewhat eclipses the role of users – employers in particular – who are bound to the instructions provided by developers but little beyond (Circiumaru, 2021[224]). There are also concerns that the additional costs for legal counsel or certification will be an undue burden on smaller firms (Czarnocki, 2021[229]) and that it will be difficult to regulate pre-release a product like an AI system, which will "learn and evolve to create differential impacts post-release" (Circiumaru, 2021[224]).

In the draft in early 2022, enforcement of the EU AI Act would lie mainly with Member States as well as with market surveillance authorities, which could force non-compliant systems to be removed from the market. With a similar enforcement structure to the GDPR, itself unevenly enforced across Member States, the AI Act may struggle to get effective EU-wide implementation, particularly since self-assessment forms such a large part of the Act (Şimşek, 2021[230]; Ebers et al., 2021[219]; Veale and Borgesius, 2021[231]). There are additionally some concerns that the national authorities may not have the financial, technical or human resources necessary to effectively enforce the legislation (Benifei and Tudorache, 2022[232]; France Digitale, 2021[233]).

Like the GDPR, the proposal is intended to have an extraterritorial effect. Subject to specific exceptions, the proposed AI regulation would apply to i) *Providers* placing on the market or putting into service AI systems in the EU, regardless of where the providers are located; ii) *Users* of AI systems located within the EU, and iii) *Providers and users* of AI systems that are located outside the EU, of which the output is used in the EU. The proposal also encourages EU countries to **establish AI regulatory sandboxes** to facilitate the development and testing of innovative AI systems under strict regulatory oversight (European Commission, 2021[227]).

The AI Act legislative proposal was being discussed by the European Parliament and the Council of EU Member states in early 2022 and seemed to have a higher likelihood of adoption in some form in the near future, though likely with significant changes. Similar to the GDPR, the proposed AI Act could impact policy beyond the EU, even before full adoption (Gaumond, 2021[234]; Whitehead, 2021[235]).

There are a number of additional EU proposals that would also affect the use of AI systems in the workplace. A proposed directive on improving working conditions in platform work would require digital labour platforms to inform workers of the use and key features of automated monitoring and decision-making systems, while restricting the types of data that could be processed. It would also require human monitoring of automated systems, and review of key decisions made by such systems (European Commission, 2021[236]). In many ways, this clarifies the platform worker-relevant algorithm regulations that were uncertain in the wake of the Digital Services Act and Digital Markets Act being discussed by the European Parliament (ETUI, 2020[237]). These proposals would establish general rules applicable to all platforms on access to and processing of data, and would establish specific obligations that would apply to "core platform services" or "gatekeepers". In particular, they would require recommendation algorithms to be safe and transparent, while promoting fair competition and fostering innovation (European Commission, 2020[238]).

United States Algorithmic Accountability Act and other US proposals

While there have been various AI legislative proposals introduced in Congress, the US has not embraced a horizontal broad-based approach to AI regulation similar to the one proposed by the European Commission. Congress is considering an Algorithmic Accountability Act which would introduce mandatory impact assessments for AI use. The bill, first presented in 2019, has not made significant progress since being presented, but would direct the Federal Trade Commission (FTC) to develop regulations requiring large firms (defined as those with over USD 50 million in revenue or that have data about at least one million consumers or consumer devices) to conduct impact assessments for existing and new "high-risk automated decision systems". High-risk automated decision systems would include those that (1) may contribute to inaccuracy, bias, or discrimination; or (2) facilitate decision-making about sensitive aspects of consumers' lives by evaluating consumers' behaviour (wherein consumer means an

individual) (US Congress, 2019[239]). An update to the bill in 2022 shifts the focus from "high risk systems" to the use of algorithmic technology in "critical decisions," defined as a decision that has a significant effect on a consumer's life "relating to access to or the cost, terms, or availability of" nine different services, including "employment, workers management, or self-employment" (US Congress, 2022[240]).

A number of US states have proposed similar regulation, including California's Automated Decision Systems Accountability Act (California State Legislature, 2020[241]), New Jersey's Algorithmic Accountability Act (Zwicker, 2019[242]), Washington D.C.'s Stop Discrimination by Algorithms Act (Racine, 2021[243]) and Washington's SB 5527 (Hasegawa et al., 2019[244]).

3.3. Policies specific to the trustworthy use of AI systems in the workplace

Some jurisdictions are putting in place new policies to address policy gaps that are emerging specifically through the use of AI in the workplace. So far, enacted policy has focused on subsets of issues. Policy efforts have notably focused on potential discrimination and associated audits, on explainability, and on ensuring that applicants and employees provide informed consent. All sought to provide more accountability.

In August 2019, the State of Illinois was the first US state to address the deployment of AI systems for recruitment purposes, with the Artificial Intelligence Video Interview Act (ILCS, 2019[245]). The bill officially went into effect in January 2020 and applies to all employers that use an AI system to analyse video interviews of applicants for jobs based in Illinois, partly with the intention of providing regulatory clarity for companies interested in using such tools (Wisenberg Brin, 2019[246]). The law stipulates that, before the interview, employers need to inform candidates that AI is being used, explain to candidates how the AI being used works and which variables are under scrutiny, as well as obtain written informed consent from the individual. Following an applicant's request, employers will also need to limit the sharing of video interviews and destroy videos and copies of videos within 30 days. Wisenberg Brin (2019[246]) highlights that the law is not clear on what kind of explanations need to be given to candidates as well as the required level of algorithmic detail. The law also does not clarify what happens to the application of a candidate who refuses to be analyzed in this way. In addition, this law could conflict with other federal and state laws that require the preservation of evidence.

In May 2020, the State of Maryland also enacted a law that banned the use of facial recognition during applicants' interviews for employment (Box 4**), unless the interviewee signs a waiver** (Fisher et al., 2020[247]). Some critics have noted that the new law leaves broad gaps in terms of what will be recognized as "facial recognition services" and "facial templates", created by the facial recognition service, and will therefore require additional interpretation (Glasser, Forman and Lech, 2020[248]).

In November 2021, the New York City Council banned the use of "automated employment decision tools" without annual bias audits (Cumbo, 2021[249]). The bill also requires that the audit results be made publicly available, that candidates and employees be notified about use of such tools for hiring or promotion and about the job qualifications and characteristics used by the tool. The bill found substantial support, but many recognize that it is more of a starting point than final legislation on the topic (Simonite, 2021[250]). There are also concerns that there is potential for vendor-sponsored audits to "rubber-stamp" their own technology especially since there are few specifics in terms of what an audit looks like, who should be doing the audit, and what disclosure to the auditor and public look like (Turner Lee and Lai, 2021[251]). In addition, Scherer and Shetty (2021[252]) note that the enacted bill has a narrower scope than the original proposal, both in terms of which employment decisions are affected (decisions about compensation, scheduling or work conditions are not included), and in terms of what protected groups are included (solely race, ethnicity and sex).

Audit requirements similar to the requirement in the New York City law may become a more widespread. A similar bill in California on "Discrimination in employment: employment tests and selection procedures" would also require annual audits (California State Legislature, 2020[253]), as would some of the more comprehensive legislation such as the Washington, D.C. bill mentioned above (Racine, 2021[243]).

Spain passed legislation in August of 2021 making transparency mandatory for AI systems that make decisions about or influence either working conditions or employment status (Pérez del Prado, 2021[254]). The law followed a Supreme Court ruling in September 2020 that qualified digital delivery "riders" as employees, and is the formalisation of an agreement reached between unions and business associations in March 2021. The law modifies the Spain's Workers' Statute to make it mandatory for platforms to provide information to workers' representatives about the mathematical or algorithmic formulae used to determine working conditions, including individuals' access to, and maintenance of, employment and their profile (Aranguiz, 2021[255]). The law differs in two key ways from existing GDPR protections. In terms of consultation rights, which are now collective, at the union level, rather than individual as under the GDPR. And in the scope of application, which also concerns algorithms simply assisting human decision-making (Todolí-Signes, 2021[256]).

Going forward, a number of policies specific to the use of AI systems in the workplace are expected to generate further conversation, and potential for future reform. Data protection laws and emerging "rider laws", such as the one enacted by Spain, are expected to increase awareness and help mitigate risks associated with the transparency and explainability of AI systems for workers (De Stefano and Taes, 2021[257]). The Spanish law, the fruit of a collective bargaining agreement, also provides for a continued role for social dialogue. Unions now have the right to access considerable information about the functioning of digital platform's algorithms relative to working conditions. Further rounds of dialogue between the platforms and unions are therefore likely, with the possibility of further policy changes.

Indeed, social partners have also been quite active in advocating for policy responses to ensure trustworthy use of AI in the workplace (Global Deal and OECD, 2021[258]; OECD, 2021[259]). Collective bargaining and social dialogue can play an important role in supporting workers' and businesses in the AI transition, provided social partners tackle challenges such as reaching out members in AI exposed occupations and industries, and acquiring knowledge and capacity to initiate AI-related strategies (OECD, forthcoming[260]).

Employers have developed a number of reports and principles. Japan's Keidanren released in 2019 an AI "Utilization Strategy." It points out the need for AI to become well integrated in the workplace, by enshrining ethical standards such as fairness, accountability and transparency, as well as rules that ensure a balance between the use and protection of personal data, and guarantees for the safety and dependability of AI systems as a whole (経団連, 2019[261]). A report by Deloitte and the US Chamber of Commerce's Technology Engagement Center highlights a number of potential risks to workers from AI use and recommends the development of standards for AI trustworthiness (through NIST), the rapid implementation of an AI risk management framework, and the development of international partnerships and standards including by the OECD (Deloitte and U.S. Chamber of Commerce Technology Engagement Center, 2021[262]).

Many companies have also set out their own individual principles for AI use. A number of checklists or principles now offer guidance to employers on avoiding bias when using AI tools. For example, the Leadership Conference on Civil and Human Rights has promulgated Principles for Hiring Assessment Technologies (Kim and Bodie, 2021[164]). SAP committed to investigate new technical methods for mitigating biases, uphold quality and safety standards, clear communication on how, why, where and when data is used in AI software, as well as engagement with its AI Ethics Advisory Panel (SAP, 2018[263]). Some critics have however alleged that these corporate principles can be a form of "ethics washing" (see Box 7).

Box 7. Ethics Washing

Recent decades have seen the growth of Responsible Business Conduct (RBC) expectations for companies in OECD countries and beyond, as society recognizes a corporate responsibility to identify and address risks to people, the environment and society (OECD, 2018[264]). As companies increasingly advertise their efforts to be more responsible, there have also been critiques of a range of reputation "washing" attempts – such as green-washing, blue-washing, or rainbow-washing for example (Bernardino, 2021[265]) – whereby consumers and policymakers are presented with surface-level corporate measures, potentially to pre-empt the establishment of firmer mandated policy, rather than with thorough commitments (Wagner, 2019[266]; Nieuwenkamp, 2017[267]). Corporate codes of ethics have been found to be associated with less perceived ethics violations in organizations, but not with a higher likelihood to report wrongdoing (Somers, 2001[268]).

In the context of the development and adoption of AI, a focus has been placed on "ethics washing" – which consists in using voluntary, but often non-binding, corporate AI ethical codes or AI ethical boards. A number of these voluntary, internal AI ethics efforts have been found to have limited internal accountability, or effectiveness in changing behaviour (Whittaker et al., 2018[269]; McNamara, Smith and Murphy-Hill, 2018[270]). While effectively constructed self-regulatory approaches can contribute to building a more ethical workplace, a "Pick your own ethics" approach must not prevail in the face of existing legislation, or in the face of emerging fundamental concerns (Yeung, Howes and Pogrebna, 2020[271]).

Trade unions have prepared guidelines, ethical principles, as well as reports and policy briefs to highlight workers' concerns and potential policy solutions. Unions can serve as a mechanism for public participation in AI regulation, in the absence of broader governmental policy debate or action. For example, in a 2021 report, the UK's Trade Union Congress highlighted a number of ethical challenges, noted the value of a targeting high-risk systems, and made legislative recommendations to avoid AI-based discrimination, to safeguard privacy, and to establish a set of rights for workers (TUC, 2021[272]). Trade unions have also called for greater participation of workers and their representatives in the governance of AI at work. For example, there have been calls to adopt data governance models for data stewardship such as data trusts, data collectives and cooperatives (Allen and Masters, 2021[273]; ETUC, 2020[274]; Colclough, 2020[275]; Ada Lovelace Institute and UK AI Council, 2021[276]; British Academy for the Humanites and Social Sciences and The Royal Society, 2017[277]). When used in the workplace, these governance mechanisms would give workers give access to and rights over the collection, analysis, storage and off-boarding of data that concerns them (Colclough, 2020[275]), promoting a trustworthy and beneficial use of data.

Trade unions representing AI developers have also established their own principles for more trustworthy AI use, noting the need to strengthen transparency (including through open audit trails and real-time oversight), to develop technical standards and certifications to increase accountability, and to involve workers to a greater extent in decisions about adoption of AI in the workplace. They also provided policy recommendations, including the need for action on defining responsibility (notably beyond the engineering profession) and the need for frameworks about explainability (ANE and IT University of Copenhagen, 2018[278]; ANE et al., 2021[279]).

4. Conclusion

Using the framework of the OECD AI Principles, **this paper identified a number of ethical risks that need to be addressed when AI systems are used in the workplace**: human rights (including privacy, fairness, agency and dignity); transparency and explainability; robustness, safety and security; and accountability. Many of the principles for trustworthiness are linked – transparency is key to accountability, for example, and explainability can help ensure fairness. Ethical concerns can emerge wherever AI systems are implemented in the workplace, from recruitment and hiring, to worker or manager assistance, to the provision of human services. Nonetheless when ethical concerns are addressed, when AI use is trustworthy, AI systems can contribute to improving workplaces, for example by improving fairness and worker safety.

Policy action will be important in achieving trustworthy use of AI in the workplace. The paper discussed the broad range of policy approaches countries are putting in place. It found that existing legislation offers an important foundation for the regulation of AI systems. Novel, society-wide and AI-specific legislative proposals are being considered, for example in the EU and in the United States, with important implications for workplace AI. these proposals are still under discussion and may go through significant changes still. Amidst self-regulatory approaches, efforts to establish national or international standards for trustworthy AI use can provide important technical parameters, with direct relevance for workplace AI systems. The paper also discussed policies focused specifically on use of AI in the workplace; these efforts can build on or preface comprehensive legislative efforts, or serve as a complement to existing legislation. Tensions between some policy goals could emerge. For example, promoting fairness and combatting bias may require an amount and type of data that could raise risks for workers' privacy.

Overall, there is a need to build capacity for workers, employers, social partners and regulators and inform them on the ethics of workplace AI so that they will be able to understand the issues at stake and knowledgeably contribute to these considerations and decisions. Going forward, it will therefore be essential to continue collecting evidence on how AI is being used in the workplace, and how this use is coming in conflict with, or strengthening, the implementation of the OECD's AI Principles. As countries increasingly take policy action to ensure trustworthy use of AI in the workplace, rigorous, evidence-based and comparative assessments will be key to determining effective policy options . It will for example be important to understand the potential parameters needed to make accountability tools – such as algorithmic audits – effective in promoting trustworthy AI in the workplace.

Through the Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) programme, the OECD will continue to follow policy developments closely and promote discussion among policy makers to share lessons learned, and identify the most promising venues to promote an ethical use of AI systems in the workplace.

References

Accessnow (2018), *Human rights in the age of artificial intelligence*, Accessnow, https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf. [39]

Ada Lovelace Instititute (2020), *Examining Tools for assessing algorithmic systems*, Ada Lovelace Institute. [152]

Ada Lovelace Institute and UK AI Council (2021), *Exploring legal mechanisms for data stewardship*, https://understandingpatientdata.org.uk/news/accountability-transparency-. [276]

Adadi, A. and M. Berrada (2018), "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, Vol. 6, https://doi.org/10.1109/ACCESS.2018.2870052. [91]

ADEP (2020), *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial*, Agencia Española Protección datos. [172]

Ajunwa, I., K. Crawford and J. Schultz (2017), "Limitless Worker Surveillance", *California Law Review*, Vol. 105/735, pp. 735-776, https://www.researchgate.net/publication/319929548_Limitless_worker_surveillance (accessed on 24 November 2020). [24]

Alan Turing Institute (2022), *New UK initiative to shape global standards for Artificial Intelligence*, Alan Turing Institute, https://www.turing.ac.uk/news/new-uk-initiative-shape-global-standards-artificial-intelligence (accessed on 31 January 2022). [215]

Allen QC, R. and D. Masters (2021), *An Italian lesson for Deliveroo: Computer programmes do not always think of everything!*, ai-lawhub.com, https://ai-lawhub.com/2021/01/18/an-italian-lesson-for-deliveroo-computer-programmes-do-not-always-think-of-everything/ (accessed on 2 January 2022). [185]

Allen, R. and D. Masters (2021), *Technology Managing People-the legal implications*, Trade Union Congress. [273]

Allgrove, B. (2004), "Legal Personality for Artificial Intellects: Pragmatic Solution or Science Fiction?", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.926015. [280]

Aloisi, A. (2021), *Bringing the algorithm to court*, ETUI Workshop, https://www.slideshare.net/AntonioAloisi1/bringing-the-algorithm-to-court-etui-workshop (accessed on 7 January 2022). [166]

ANE et al. (2021), *Addressing Ethical Dilemmas in AI: Listening to Engineers*, Association of Nordic Engineers, https://nordicengineers.org/wp-content/uploads/2021/01/addressing-ethical-dilemmas-in-ai-listening-to-the-engineers.pdf (accessed on 26 November 2021). [279]

ANE and IT University of Copenhagen (2018), *Nordic engineers' stand on Artificial Intelligence and Ethics: Policy Recommendations and Guidelines*, Association of Nordic Engineers. [278]

Anwer, A. (2021), *How SMEs can use IP to secure success in the new data-fuelled AI paradigm*, https://www.iam-media.com/copyright/ip-opportunities-and-challenges-smes-in-the-new-data-fuelled-ai-paradigm (accessed on 4 January 2022). [103]

Aranguiz, A. (2021), *Spain's platform workers win algorithm transparency*, https://socialeurope.eu/spains-platform-workers-win-algorithm-transparency (accessed on 20 March). [255]

Baker, M. (2020), *AI Shows Value and Gains Traction in HR*, Gartner, https://www.gartner.com/smarterwithgartner/ai-shows-value-and-gains-traction-in-hr (accessed on 17 September 2021). [10]

Barlaskar, S. and D. Petkovic (2018), *GitHub Random-Forest-Explainability-Pipeline*, https://github.com/sabiha90/Random-Forest-Explainability-Pipeline (accessed on 31 August 2021). [99]

Batchelor, J. (2021), "Xsolla reportedly lays off up to 150 people based on big data", *GamesIndustry.biz*, https://www.gamesindustry.biz/articles/2021-08-09-xsolla-reportedly-lays-off-up-to-150-people-based-on-big-data (accessed on 17 October 2021). [115]

BCG Henderson Institute (2018), *The Build-or-Buy Dilemma in Artificial Intelligence*, https://www.bcg.com/publications/2018/build-buy-dilemma-artificial-intelligence (accessed on 3 November 2021). [107]

BehavioSec (2020), *Welcome to the next generation of behavioral authentication*, https://www.behaviosec.com/wp-content/uploads/2020/11/bhs-solution-brief-1.pdf (accessed on 4 August 2021). [136]

Benifei, B. and D. Tudorache (2022), *Press conference by Brando Benifei (S&D, IT) and Drago Tudorache (Renew, RO), co-rapporteurs on the Artificial Intelligence Act*, European Parliament, https://api.multimedia.europarl.europa.eu/documents/20125/21116147/1647877692210_I221920%5BSD-EN%5D.pdf (accessed on 25 March 2022). [232]

Bennet, M. (2020), *Bennet, Colleagues Call on EEOC to Clarify Authority to Investigate Bias in AI-Driven Hiring Technologies*, U.S. Senate, https://www.bennet.senate.gov/public/index.cfm/2020/12/bennet-colleagues-call-on-eeoc-to-clarify-authority-to-investigate-bias-in-ai-driven-hiring-technologies (accessed on 24 November 2021). [181]

Bernardino, P. (2021), "Responsible CSR Communications: Avoid "Washing" Your Corporate Social Responsibility (CSR) Reports and Messages", *Journal of Leadership, Accountability & Ethics*, Vol. 18/1, pp. 102-113. [265]

Berryhill, J. et al. (2019), "Hello, World: Artificial intelligence and its use in the public sector", *OECD Working Papers on Public Governance*, No. 36, OECD Publishing, Paris, https://dx.doi.org/10.1787/726fd39d-en. [203]

Bertail, P. et al. (2019), *Algorithmes : biais, discrimination et équité*, Telecom Paris Tech and Fondation Abeona, https://www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite (accessed on 9 January 2022). [54]

Bertallee, C. (2019), *New Study: 64% of People Trust a Robot More Than Their Manager*, Oracle, https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html (accessed on 17 September 2021). [8]

BGC GAMMA and IPSOS (2018), *Artificial Intelligence: Have No Fear The revolution of AI at work*. [4]

Binns, R. (2020), *On the Apparent Conflict Between Individual and Group Fairness*, ACM, New York, NY, USA, https://doi.org/10.1145/3351095.3372864. [53]

BMAS (2021), *Betriebsrätemodernisierungsgesetz*, Bundesministerium für Arbeit und Soziales, https://www.bmas.de/DE/Service/Gesetze-und-Gesetzesvorhaben/betriebsraetemodernisierungsgesetz.html (accessed on 1 April 2022). [110]

Bowie, N. (1998), "A Kantian Theory of Meaningful Work", *Journal of Business Ethics*, Vol. 17/9/10, pp. 1083-1092, https://www.jstor.org/stable/25073937?seq=1#metadata_info_tab_contents (accessed on 19 October 2021). [73]

Briône, P. (2020), *My boss the algorithm: an ethical look at algorithms in the workplace*. [65]

Briône, P. (2020), *My boss the algorithm: an ethical look at algorithms in the workplace*, https://www.acas.org.uk/my-boss-the-algorithm-an-ethical-look-at-algorithms-in-the-workplace/html (accessed on 21 October 2021). [19]

British Academy for the Humanites and Social Sciences and The Royal Society (2017), *Data management and use: Governance in the 21st century A joint report by the British Academy and the Royal Society*, https://www.thebritishacademy.ac.uk/documents/105/Data_management_and_use_-_Governance_in_the_21st_century.pdf (accessed on 5 January 2022). [277]

Brown, S., J. Davidovic and A. Hasan (2021), *The algorithm audit: Scoring the algorithms that score us*, SAGE Publications Ltd, https://doi.org/10.1177/2053951720983865. [149]

Brown, S., J. Davidovic and A. Hasan (2021), "The algorithm audit: Scoring the algorithms that score us:", *https://doi.org/10.1177/2053951720983865*, Vol. 8/1, https://doi.org/10.1177/2053951720983865. [153]

Buolamwini, J. and T. Gebru (2018), *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* *. [60]

Butler, D. and K. White (2021), *EEOC is Monitoring Use of Artificial Intelligence*, https://www.natlawreview.com/article/employers-beware-eeoc-monitoring-use-artificial-intelligence (accessed on 18 October 2021). [143]

Butler, S. (2021), *Court tells Uber to reinstate five UK drivers sacked by automated process*, https://www.theguardian.com/technology/2021/apr/14/court-tells-uber-to-reinstate-five-uk-drivers-sacked-by-automated-process (accessed on 23 November 2021). [178]

Cahuc, P., S. Carcillo and A. Zylberberg (2004), "LABOR ECONOMIC, SECOND EDITION", Vol. 67/6. [35]

California State Legislature (2020), *AB-13 Personal rights: automated decision systems*, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB13. [241]

California State Legislature (2020), *Bill Text - SB-1241 Discrimination in employment: employment tests and selection procedures*, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200SB1241 (accessed on 1 April 2022). [253]

Candelon, F. et al. (2021), *AI Regulation Is Coming*, https://hbr.org/2021/09/ai-regulation-is-coming (accessed on 17 March 2022). [162]

Cappelli, P., P. Tambe and V. Yakubovich (2019), "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward", *SSRN Electronic Journal*, https://doi.org/10.2139/SSRN.3263878. [93]

Casilli, A. (2017), "Global Digital Culture| Digital Labor Studies Go Global: Toward a Digital Decolonial Turn", *International Journal of Communication*, Vol. 11/0, p. 21, https://ijoc.org/index.php/ijoc/article/view/6349 (accessed on 10 January 2022). [80]

Cater, L. and M. Heikkilä (2021), *Your boss is watching: How AI-powered surveillance rules the workplace – POLITICO*, Politico, https://www.politico.eu/article/ai-workplace-surveillance-facial-recognition-software-gdpr-privacy/ (accessed on 4 August 2021). [139]

Chen, S. (2018), "'Forget the Facebook leak': China is mining data directly from workers' brains on an industrial scale", *South China Morning Post*, https://www.scmp.com/news/china/society/article/2143899/forget-facebook-leak-china-mining-data-directly-workers-brains?module=perpetual_scroll_0&pgtype=article&campaign=2143899 (accessed on 16 March 2022). [125]

Chesterman, S. (2020), "Artificial Intelligence and the Limits of Legal Personality", *International and Comparative Law Quarterly*, Vol. 69/4, pp. 819-844, https://doi.org/10.1017/S0020589320000366. [281]

Chmielinski, K. and L. Grennan (2021), *Responsible machine learning protects intellectual property*, World Economic Forum, https://www.weforum.org/agenda/2021/03/responsible-machine-learning-that-protects-intellectual-property/ (accessed on 4 January 2022). [105]

Christakis, N. (2019), *Will Robots Change Human Relationships?*, https://www.theatlantic.com/magazine/archive/2019/04/robots-human-relationships/583204/ (accessed on 18 October 2021). [129]

Circiumaru, A. (2021), *Three proposals to strengthen the EU Artificial Intelligence Act*, Ada Lovelace Institute, https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/ (accessed on 23 December 2021). [224]

Clement-Jones, L. (2020), *How the OECD's AI system classification work added to a year of progress in AI governance*, https://oecd.ai/wonk/oecd-ai-system-classification-year-of-progress-ai-governance. [161]

CNIL (2017), *Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle*, Commission Nationale de l'Informatique et des Libertés, https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de (accessed on 25 November 2021). [174]

Colclough, C. (2021), *Audits & Impact Assessments 2.0*, The Why Not Lab, https://www.thewhynotlab.com/post/audits-impact-assessments (accessed on 10 January 2022). [155]

Colclough, C. (2020), *Towards Workers' Data Collectives*, https://itforchange.net/digital-new-deal/2020/10/22/towards-workers-data-collectives/ (accessed on 21 October 2021). [275]

Colgan, G. (2013), *Gilbert v. Transfield Services (New Zealand) Ltd*, Employment Court Christchurch, https://www.pbarrett.net/tbv/2013-NZEmpC-71-Gilbert-v-Transfield-Services-NZ-Ltd.pdf (accessed on 20 December 2021). [193]

Condie, B. and L. Dayton (2020), "Four AI technologies that could transform the way we live and work", *Nature*, Vol. 588/7837, https://doi.org/10.1038/d41586-020-03413-y. [124]

Council of Europe (2021), *Guidelines on Facial Recognition*, The Consultative Committee of the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3. [217]

Crawford, K. et al. (2019), *AI Now 2019 Report*, https://ainowinstitute.org/AI_Now_2019_Report.pdf (accessed on 26 November 2020). [42]

Cumbo, L. (2021), *The New York City Council - File #: Int 1894-2020*, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID|Text|&Search=artificial+intelligence (accessed on 14 December 2021). [249]

Czarnocki, J. (2021), *Good intentions, unintended consequences? How the Proposal for the AI Act might kick innovation out of the EU*, CITIP Blog, https://www.law.kuleuven.be/citip/blog/good-intentions-unintended-consequences/ (accessed on 23 December 2021). [229]

Data Protection Working Party (2017), *Guidelines on Consent under Regulation 2016/679*. [89]

Datatilsynet (2018), *Artificial intelligence and privacy*, Norwegian Data Protection Authority. [173]

Daten Ethik Kommission (2019), *Opinion of the Data Ethics Commission*, https://datenethikkommission.de/wp-content/uploads/DEK_Gutachten_engl_bf_200121.pdf. [196]

De Stefano, V. (2018), ""Negotiating the algorithm": Automation, artificial intelligence and labour protection", *Employment*, No. 246, International Labour Organisation, http://www.ilo.org/publns (accessed on 18 October 2021). [119]

De Stefano, V. (2018), "Negotiating the Algorithmm: Automation, Artificial Intelligence and Labour Protection", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3178233. [38]

De Stefano, V. (2016), "The rise of the 'just-in-time workforce': On-demand work, crowdwork and labour protection in the 'gig-economy'", *Conditions of Work and Employment*, No. 71, ILO. [34]

De Stefano, V. and S. Taes (2021), *Algorithmic management and collective bargaining*, https://www.etui.org/sites/default/files/2021-05/Algorithmic%20management%20and%20collective%20bargaining-web-2021.pdf (accessed on May). [257]

Deborah Raji, I. et al. (2020), "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing", Vol. 7, https://doi.org/10.1145/3375627.3375820. [59]

Delaware General Assembly (2022), *Title 19 General Provisions Chapter 7 Employment Practices*, https://delcode.delaware.gov/title19/c007/sc01/index.html (accessed on 17 June 2022). [168]

Deloitte (2018), *People data: how far is too far?*, pp. 89-94, https://www2.deloitte.com/content/dam/insights/us/articles/HCTrends2018/2018-HCtrends_Rise-of-the-social-enterprise.pdf. [134]

Deloitte and U.S. Chamber of Commerce Technology Engagement Center (2021), *Investing in trustworthy AI*. [262]

Deshpande, A. et al. (2021), *Improving working conditions using Artificial Intelligence*, European Parliament. [37]

Deutscher Bundestag Enquete-Kommission (2020), *Artificial Intelligence – Social Responsibility and Economic, Social and Ecological Potential*, https://www.bundestag.de/resource/blob/804184/f31eb697deef36fc271c0587e85e5b19/Kurzfassung-des-Gesamtberichts-englische-Uebersetzung-data.pdf. [197]

DKE (2020), *German Standardization Roadmap on Artificial Intelligence*, https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf. [208]

Doshi-Velez, F. et al. (2017), "Accountability of AI Under the Law: The Role of Explanation", *SSRN Electronic Journal*, https://arxiv.org/abs/1711.01134v3 (accessed on 5 August 2021). [96]

DOT Europe (2021), *DOT Europe Position on the Artificial Intelligence Act*, DOT Europe. [220]

Ebers, M. et al. (2021), "The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)", *J*, Vol. 4/4, pp. 589-603, https://doi.org/10.3390/J4040043. [219]

Echarri, M. (2021), "Xsolla: One second, 150 dismissals: Inside the algorithms that decide who should lose their job", *El Pais*, https://english.elpais.com/usa/2021-10-14/one-second-150-dismissals-inside-the-algorithms-that-decide-who-should-lose-their-job.html (accessed on 17 October 2021). [114]

EDPB-EDPS (2021), *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down hamonised rules on artificial intelligence*, https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf. [226]

Enarsson, T., L. Enqvist and M. Naarttijärvi (2021), "Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts", *https://doi.org/10.1080/13600834.2021.1958860*, https://doi.org/10.1080/13600834.2021.1958860. [147]

Engler, A. (2021), *Auditing employment algorithms for discrimination*, Brookings, https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/ (accessed on 31 August 2021). [142]

Engler, A. (2021), *Auditing employment algorithms for discrimination*, Brookings, Washington, D.C., https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/ (accessed on 10 January 2022). [154]

EPIC (2019), *Complaint and Request for Investigation, Injunction, and Other Relief*, Electronic Privacy Information Center, https://epic.org/privacy/ftc/zoomEPIC-FTC-Complaint-In-re-Zoom-7-19.pdf; (accessed on 24 November 2021). [189]

ETUC (2020), *Resolution on the European strategies on artificial intelligence and data Introduction and Context*, https://www.etuc.org/sites/default/files/document/file/2020-07/Adopted%20-%20ETUC%20resolution%20on%20%20the%20European%20strategies%20on%20artificial%20intelligence%20and%20data%20-%20EN.pdf (accessed on 5 January 2022). [274]

ETUI (2020), *The digital services act package - Reflection on the EU Commission's policy options*, https://www.etui.org/sites/default/files/2020-09/The%20digital%20services%20act%20package.%20Reflections%20on%20the%20EU%20Commission%27s%20policy%20options-2-2020.pdf. [237]

EU FRA (2020), *Getting the future right – Artificial intelligence and fundamental rights*, European Union Agency for Fundamental Rights, https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights (accessed on 17 March 2022). [44]

EU-OSHA (2021), *Impact of Artificial Intelligence on Occupational Safety and Health*, EU-OSHA, https://osha.europa.eu/en/publications/osh-and-. [120]

Euronews (2021), "Companies used artificial intelligence to monitor workers during pandemic, says trade union", *Euronews*, https://www.euronews.com/2021/07/19/companies-used-artificial-intelligence-to-monitor-workers-during-pandemic-says-trade-union (accessed on 4 August 2021). [133]

European Commission (2021), *Directive of the European Parliament and of the Council on Improving Working Conditions in Platform Work*, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6605 (accessed on 23 December 2021). [236]

European Commission (2021), *EU-US Trade and Technology Council Inaugural Joint Statement*, https://ec.europa.eu/commission/presscorner/detail/en/statement_21_4951. [214]

European Commission (2021), *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (accessed on 24 November 2021). [218]

European Commission (2021), "Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence and amending certain Union Legislative Acts (Artificial Intelligence Act)", *COM(2021) 206*, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence. [227]

European Commission (2020), *Europe fit for the Digital Age: digital platforms*, https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347 (accessed on 23 December 2021). [238]

European Digital Rights (EDRi) et al. (2021), *An EU Artificial Intelligence Act for Fundamental Rights A Civil Society Statement*. [225]

Executive Office of the President of the United States (2016), "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights Executive Office of the President Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights", https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (accessed on 21 October 2021). [40]

Fang, L. (2019), *Google Hired Gig Economy Workers for Project Maven*, https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/ (accessed on 9 January 2022). [77]

Federal Trade Commission (2021), *Aiming for truth, fairness, and equity in your company's use of AI*, Federal Trade Commission, https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai (accessed on 24 November 2021). [187]

Feliciano Reyes, J. (2019), "In the basement of CHOP, warehouse workers say they're held to impossible quotas", *The Philadelphia Inquirer*, https://www.inquirer.com/news/warehouse-workers-quotas-rate-childrens-hospital-of-philadelphia-canon-20190422.html (accessed on 17 October 2021). [111]

Fernández-Macias, E. et al. (2018), *Game changing technologies: Exploring the impact on production processes and work*, Eurofound, https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/fomeef18001en.pdf. [68]

Fisher et al. (2020), *Bill Text: MD HB1202*, https://legiscan.com/MD/text/HB1202/id/2169556 (accessed on 14 December 2021). [247]

France Digitale (2021), *France Digitale is concerned about the impact of the EU's Artificial Intelligence (AI) Act on tech startups*, France Digitale, https://francedigitale.org/en/combat/artificial-intelligence-act/ (accessed on 25 March 2022). [233]

Gajane, P. and M. Pechenizkiy (2018), "On Formalizing Fairness in Prediction with Machine Learning", *arXiv*. [52]

GAO (2021), *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, https://www.gao.gov/assets/720/716110.pdf. [205]

Garcia, E. (2021), *The International Governance of AI: Where is the Global South? - The Good AI*, The Good AI, https://thegoodai.co/2021/01/28/the-international-governance-of-ai-where-is-the-global-south/ (accessed on 10 January 2022). [82]

Gaumond, E. (2021), *Artificial Intelligence Act: What Is the European Approach for AI? - Lawfare*, Lawfare, https://www.lawfareblog.com/artificial-intelligence-act-what-european-approach-ai (accessed on 3 January 2022). [234]

Gavaghan, C., A. Knott and J. MacLaurin (2021), *The Impact of Artificial Intelligence on Jobs and Work in New Zealand*, University of Otago, https://www.otago.ac.nz/caipp/otago828396.pdf (accessed on 20 December 2021). [195]

Geiger, G. (2021), *Court Rules Deliveroo Used 'Discriminatory' Algorithm*, Vice, https://www.vice.com/en/article/7k9e4e/court-rules-deliveroo-used-discriminatory-algorithm (accessed on 2 January 2022). [184]

Gent, E. (2019), "The 'ghost work' powering tech magic", *BBC*, https://www.bbc.com/worklife/article/20190829-the-ghost-work-powering-tech-magic (accessed on 9 January 2022). [78]

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://dx.doi.org/10.1787/c2c1d276-en. [2]

Gesser, A., P. Rubin and A. Gressel (2022), *Model Destruction – The FTC's Powerful New AI and Privacy Enforcement Tool*, https://www.debevoise.com/insights/publications/2022/03/model-destruction-the-ftcs-powerful-new (accessed on 30 March 2022). [191]

Gibson Dunn (2022), *2021 Artificial Intelligence and Automated Systems Annual Legal Review*, https://www.gibsondunn.com/2021-artificial-intelligence-and-automated-systems-annual-legal-review/ (accessed on 1 April 2022). [176]

Gino, F. and B. Staats (2012), *The Microwork Solution*, https://hbr.org/2012/12/the-microwork-solution (accessed on 10 January 2022). [83]

Glasser, N., A. Forman and C. Lech (2020), *New Maryland Law Requires Applicant Consent Prior to Using Facial Recognition Technology in Job Interviews*, https://www.natlawreview.com/article/new-maryland-law-requires-applicant-consent-prior-to-using-facial-recognition (accessed on 14 December 2021). [248]

Global Deal and OECD (2021), *The impact of Artificial Intelligence on the labour market and the workplace: What role for social dialogue?*, Global Deal and OECD, https://www.theglobaldeal.com/news/The-impact-of-artificial-intelligence-on-the-labour-market-and-the-workplace.pdf (accessed on 28 January 2022). [258]

Google (2022), *Explainable AI*, Google Cloud, https://cloud.google.com/explainable-ai (accessed on 17 March 2022). [97]

GPDP (2021), *Ordinanza ingiunzione nei confronti di Foodinho s.r.l. [9675440] - 10 giugno 2021*, https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9675440 (accessed on 30 July 2021). [179]

Gray, J. (2021), *The bossware boom is upon us: a look inside the employee monitoring software market | The Business of Business*, https://www.businessofbusiness.com/articles/employee-monitoring-software-productivity-activtrak-hubstaff-covid/ (accessed on 3 November 2021). [29]

Gray, M. and S. Suri (2019), *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*, Houghton Mifflin Harcourt, Boston. [74]

Harwell, D. (2021), “Contract lawyers face a growing invasion of surveillance programs that monitor their work”, *The Washington Post*, https://www.washingtonpost.com/technology/2021/11/11/lawyer-facial-recognition-monitoring/ (accessed on 10 January 2022). [31]

Hasegawa et al. (2019), *SB 5527 - 2019-20*, https://app.leg.wa.gov/billsummary?BillNumber=5527&Year=2019 (accessed on 14 December 2021). [244]

Heavens, W. (2020), *The way we train AI is fundamentally flawed*. [140]

Hegel, G. (1807), *The Phenomenology of Spirit*. [71]

Heilweil, R. (2019), “Artificial intelligence will help determine if you get your next job”, *Vox Recode*, https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen (accessed on 14 October 2021). [87]

HM Government (2021), *National AI Strategy*, Secretary of State for Digital, Culture, Media and Sport. [202]

Holstein, K. et al. (2019), *Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need*, ACM, https://doi.org/10.1145/3290605.3300830. [64]

House of Commons of Canada (2020), *Government Bill C-11 - First Reading*, House of Commons of Canada, https://www.parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading (accessed on 1 April 2022). [175]

Hunkenschroer, A. and C. Luetge (2022), “Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda”, *Journal of Business Ethics*, Vol. 1, pp. 1-31, https://doi.org/10.1007/S10551-022-05049-6/TABLES/5. [159]

IBM (2021), *AI Ethics - IBM*, https://www.ibm.com/cloud/learn/ai-ethics (accessed on 3 April 2022). [11]

ICO (2020), *Guidance on the AI auditing framework*, https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (accessed on 25 November 2021). [170]

ILCS (2019), *820 ILCS 42/ Artificial Intelligence Video Interview Act*, https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68 (accessed on 14 December 2021). [245]

ILO (2011), *OSH Management System: A tool for continual improvement*, International Labour Organization, https://www.ilo.org/wcmsp5/groups/public/@ed_protect/@protrav/@safework/documents/publication/wcms_153930.pdf (accessed on 6 January 2022). [145]

Institut Montaigne (2020), *Algorithms: Please Mind the Bias!*. [46]

Irani, L. (2016), “The hidden faces of automation”, *XRDS: Crossroads, The ACM Magazine for Students*, Vol. 23/2, pp. 34-37, https://doi.org/10.1145/3014390. [75]

Jansen, A. et al. (2018), *Emergent Risks to Workplace Safety; Working in the Same Space as a Cobot*. [131]

Jarota, M. (2021), "Artificial intelligence and robotisation in the EU - should we change OHS law?", *Journal of Occupational Medicine and Toxicology*, Vol. 16/1, https://doi.org/10.1186/s12995-021-00301-7. [163]

Jarrahi, M. et al. (2021), "Algorithmic management in a work context", *Big Data and Society*, Vol. 8/2, https://doi.org/10.1177/20539517211020332. [21]

Jarrahi, M. et al. (2021), "Algorithmic management in a work context:", *https://doi.org/10.1177/20539517211020332*, Vol. 8/2, https://doi.org/10.1177/20539517211020332. [102]

Jaume-Palasí, L. and M. Spielkamp (2017), "Ethics and algorithmic processes for decision making and decision support"*, AlgorithmWatch Working Paper*, No. 2, AlgorithmWatch, Berlin. [92]

Johnson, K. (2021), *The Fight to Define When AI Is 'High Risk'*, https://www.wired.com/story/fight-to-define-when-ai-is-high-risk/ (accessed on 1 February 2022). [222]

Johnson, K. (2021), *What algorithm auditing startups need to succeed*, VentureBeat, https://venturebeat.com/2021/01/30/what-algorithm-auditing-startups-need-to-succeed/ (accessed on 10 January 2022). [151]

Jones, A. (2021), "Artificial intelligence can now be recognised as an inventor after historic Australian court decision - ABC News", *ABC New Australia*, https://www.abc.net.au/news/2021-08-01/historic-decision-allows-ai-to-be-recognised-as-an-inventor/100339264 (accessed on 9 January 2022). [282]

Kahn, J. (2021), "HireVue stops using facial expressions to assess job candidates amid audit of its A.I. algorithms | Fortune", *Fortune*, https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/ (accessed on 24 November 2021). [188]

Kak, A. (2020), *The Global South is everywhere, but also always somewhere*, ACM, New York, NY, USA, https://doi.org/10.1145/3375627.3375859. [81]

Kaori Gurley, L. (2021), *Amazon's AI Cameras Are Punishing Drivers for Mistakes They Didn't Make*, https://www.vice.com/en/article/88npjv/amazons-ai-cameras-are-punishing-drivers-for-mistakes-they-didnt-make (accessed on 10 January 2022). [127]

Kaplan, E. (2015), *The Spy Who Fired Me*, https://harpers.org/archive/2015/03/the-spy-who-fired-me/2/ (accessed on 10 January 2022). [126]

Kellogg, K., M. Valentine and A. Christin (2020), "Algorithms at Work: The New Contested Terrain of Control", *https://doi.org/10.5465/annals.2018.0174*, Vol. 14/1, pp. 366-410, https://doi.org/10.5465/ANNALS.2018.0174. [22]

Kersley, A. (2021), "Couriers say Uber's 'racist' facial identification tech got them fired", *Wired UK*, https://www.wired.co.uk/article/uber-eats-couriers-facial-recognition (accessed on 18 October 2021). [117]

Khan, L. et al. (2021), *Resolution Directing Use of Compulsory Process Regarding Bias in Algorithms and Biometrics*, Federal Trade Commission. [190]

Kim, P. (2019), "Big Data and Artificial Intelligence: New Challenges for Workplace Equality", *University of Louisville Law Review*, Vol. 57, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3296521. [48]

Kim, P. and M. Bodie (2021), "Artificial Intelligence and the Challenges of Workplace Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy Discrimination and Privacy", *Journal of Labor and Employment Law*, Vol. 35/2, pp. 289-315, https://scholarship.law.slu.edu/faculty (accessed on 30 March 2022). [43]

Kim, P. and M. Bodie (2021), "Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy", *ABA Journal of Labor and Employment Law*, Vol. 289. [164]

Kim, T. and B. Routledge (2021), "Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach", *Business Ethics Quarterly*, pp. 1-28, https://doi.org/10.1017/BEQ.2021.3. [62]

Klaise, J. et al. (2021), "Alibi Explain: Algorithms for Explaining Machine Learning Models Alexandru Coca *", *Journal of Machine Learning Research*, Vol. 22, pp. 1-7, http://jmlr.org/papers/v22/21-0017.html. (accessed on 31 August 2021). [98]

Knight, W. (2021), *Ford's Ever-Smarter Robots Are Speeding Up the Assembly Line | WIRED*, https://www.wired.com/story/fords-smarter-robots-speeding-assembly-line/ (accessed on 31 August 2021). [132]

Koshiyama, A., E. Kazim and P. Treleaven (2021), *Familiar methods can help to ensure trustworthy AI as the algorithm auditing industry grows - OECD.AI*, The AI Wonk, https://oecd.ai/en/wonk/algorithm-auditing-trustworty-ai (accessed on 10 January 2022). [157]

Kravets, D. (2015), *Worker fired for disabling GPS app that tracked her 24 hours a day [Updated] | Ars Technica*, https://arstechnica.com/tech-policy/2015/05/worker-fired-for-disabling-gps-app-that-tracked-her-24-hours-a-day/ (accessed on 13 August 2021). [90]

La Moncloa (2020), *Pedro Sánchez presents National Artificial Intelligence Strategy with public investment of 600 million euros for period 2021-2023*, La Moncloa, https://www.lamoncloa.gob.es/lang/en/presidente/news/Paginas/2020/20201202_enia.aspx (accessed on 23 November 2021). [200]

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://dx.doi.org/10.1787/7c895724-en. [3]

Lee, M. et al. (2015), "Working with machines: The impact of algorithmic and data-driven management on human workers", *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2015-April, pp. 1603-1612, https://doi.org/10.1145/2702123.2702548. [16]

Leicht-Deobald, U. et al. (2019), "The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity", *Journal of Business Ethics*, Vol. 160/2, https://doi.org/10.1007/s10551-019-04204-w. [66]

Likens, S. et al. (2021), *AI Predictions 2021: How to navigate the top 5 AI trends facing your business: PwC*, PwC, https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html (accessed on 17 September 2021). [9]

Loi, M. (2020), *People Analytics must benefit the people - An ethical analysis of dtat-driven algorithmic systems in human resources management*, Algorithm Watch. [69]

Lomas, N. (2021), "Europe's AI Act falls far short on protecting fundamental rights, civil society groups warn", *TechCrunch*, https://techcrunch.com/2021/11/30/eu-ai-act-civil-society-recommendations/ (accessed on 23 December 2021). [223]

Maltseva, K. (2020), "Wearables in the workplace : the brave new world of employee engagement", *Business Horizons*, Vol. 63, pp. 493-505. [32]

Maltseva, K. (2020), "Wearables in the workplace: The brave new world of employee engagement", *Business Horizons*, Vol. 63/4, https://doi.org/10.1016/j.bushor.2020.03.007. [67]

Maurer, R. (2021), "Use of AI in the Workplace Raises Legal Concerns", *SHRM*, https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/use-of-ai-in-the-workplace-raises-legal-concerns.aspx (accessed on 18 October 2021). [141]

McKinsey (2021), *Global survey: The state of AI in 2021*, McKinsey Global Institute, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021 (accessed on 16 March 2022). [5]

McNamara, A., J. Smith and E. Murphy-Hill (2018), "Does ACM's Code of Ethics Change Ethical Decision Making in Software Develop-ment", *n Proceedings of the 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '18)*, https://doi.org/10.1145/3236024.3264833. [270]

Meyer, S. (2019), "A Looming AI War: Transparency v. IP Rights - Lexology", *Lexology*, https://www.lexology.com/library/detail.aspx?g=5ec0ae23-5a2c-401c-993e-9d807ba9745b (accessed on 4 January 2022). [104]

Milan, S. and E. Trere (2017), "Big Data from the South: The Beginning of a Conversation We Must Have", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3056958. [79]

Milne, S. (2021), *Bosses turn to 'tattleware' to keep tabs on employees working from home | Technology | The Guardian*, The Guardian, https://www.theguardian.com/us-news/2021/sep/05/covid-coronavirus-work-home-office-surveillance (accessed on 3 November 2021). [30]

Montagnier, P. and I. Ek (2021), "AI measurement in ICT usage surveys: A review", *OECD Digital Economy Papers*, No. 308, OECD Publishing, Paris, https://dx.doi.org/10.1787/72cce754-en. [7]

Moore, P. (2020), *Surveillance & monitoring: The future of work in the digital era*, STOA. [88]

Moore, P. (2018), *The Threat of Physical and Psychosocial Violence and Harassment in Digitalized Work*, ILO, Geneva, https://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---actrav/documents/publication/wcms_617062.pdf. [128]

Morrison, S. (2020), "Just because you're working from home doesn't mean your boss isn't watching you", *Vox*, https://www.vox.com/recode/2020/4/2/21195584/coronavirus-remote-work-from-home-employee-monitoring (accessed on 29 March 2022). [28]

Morrison, S. (2020), "Privacy: Coronavirus work-from-home policies create surge in employee tracking - Vox", *Recode*, https://www.vox.com/recode/2020/4/2/21195584/coronavirus-remote-work-from-home-employee-monitoring (accessed on 14 October 2021). [85]

Müller, N., D. Kowatsch and K. Böttinger (2020), "Data Poisoning Attacks on Regression Learning and Corresponding Defenses", *Proceedings of IEEE Pacific Rim International Symposium on Dependable Computing, PRDC*, Vol. 2020-December, pp. 80-89, https://arxiv.org/abs/2009.07008v1 (accessed on 31 August 2021). [130]

Neff, G., M. McGrath and N. Prakash (2020), *AI @ Work*, Oxford Internet Institute. [57]

Nelson, B. (2018), "How robots are reshaping one of the dirtiest, most dangerous jobs", *NBC News*, https://www.nbcnews.com/mach/science/how-robots-are-reshaping-one-dirtiest-most-dangerous-jobs-ncna866771 (accessed on 16 August 2021). [121]

New Zealand Parliamentary Counsel Office (2000), *Employment Relations Act 2000 No 24 (as at 26 November 2021)*, https://www.legislation.govt.nz/act/public/2000/0024/latest/DLM58317.html#DLM58327 (accessed on 20 December 2021). [194]

Nguyen, A. and A. Mateescu (2019), *Explainer: Algorithmic Management in the Workplace*, Data & Society Research Institute, https://doi.org/10.1145/2702123.2702548. [17]

Nieuwenkamp, R. (2017), *Ever heard of SDG washing? The urgency of SDG Due Diligence*, OECD Development Matters, https://oecd-development-matters.org/2017/09/25/ever-heard-of-sdg-washing-the-urgency-of-sdg-due-diligence/ (accessed on 31 January 2022). [267]

NIST (2022), *AI Risk Management Framework: Initial Draft*, National Institute of Standards and Technology. [210]

NIST (2021), *Draft NIST Special Publication 1270 - A Proposal for Identifying and Managing Bias in Artificial Intelligence*, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf?_sm_au_=iHVbf0FFbP1SMrKRFcVTvKQkcK8MG. [211]

NIST (2020), *Four Principles of Explainable Artificial Intelligence*, https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf. [209]

NSW Government (2020), *AI Strategy Overview*, New South Wales Government, https://www.digital.nsw.gov.au/policy/artificial-intelligence-ai/ai-strategy/strategy-overview (accessed on 26 November 2021). [204]

Nurski, L. (2021), *Algorithmic management is the past, not the future of work | Bruegel*, Bruegel, https://www.bruegel.org/2021/05/algorithmic-management-is-the-past-not-the-future-of-work/ (accessed on 6 January 2022). [146]

O'Connor, J. (2022), *New York Employers Must Notify Employees of Electronic Monitoring*, https://www.jdsupra.com/legalnews/new-york-employers-must-notify-8717062/ (accessed on 17 June 2022). [167]

OECD (2022), *Artificial intelligence in the labour market: What role for Social dialogue?*, OECD. [15]

OECD (2021), *OECD AI list of AI stakeholder initiatives*, OECD.AI, https://oecd.ai/en/countries-and-initiatives/stakeholder-initiatives (accessed on 26 November 2021). [259]

OECD (2021), "State of implementation of the OECD AI Principles: Insights from national AI policies", *OECD Digital Economy Papers*, No. 311, OECD Publishing, Paris, https://dx.doi.org/10.1787/1cd40c44-en. [160]

OECD (2021), *The Digital Transformation of SMEs*, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://dx.doi.org/10.1787/bdb9256a-en. [106]

OECD (2021), "The rise of domestic outsourcing and its implications for low-pay occupations", in *OECD Employment Outlook 2021: Navigating the COVID-19 Crisis and Recovery*, OECD Publishing, Paris, https://dx.doi.org/10.1787/937ad5bc-en. [112]

OECD (2021), "Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems", *OECD Digital Economy Papers*, No. 312, OECD Publishing, Paris, https://dx.doi.org/10.1787/008232ec-en. [13]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://dx.doi.org/10.1787/eedfee77-en. [94]

OECD (2019), "Left on your own? Social protection when labour markets are in flux", in *OECD Employment Outlook 2019: The Future of Work*, OECD Publishing, Paris, https://dx.doi.org/10.1787/bfb2fb55-en. [118]

OECD (2019), *OECD Skills Outlook 2019: Thriving in a Digital World*, OECD Publishing, Paris, https://dx.doi.org/10.1787/df80bc12-en. [109]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, OECD, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 31 August 2021). [12]

OECD (2019), "Scoping the OECD AI principles : Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", *OECD Digital Economy Papers*, No. 291, OECD Publishing, Paris, https://www.oecd-ilibrary.org/science-and-technology/scoping-the-oecd-ai-principles_d62f618a-en (accessed on 28 January 2022). [283]

OECD (2018), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264303072-en. [158]

OECD (2018), *Recommendation of the Council on the OECD Due Diligence Guidance for Responsible Business Conduct*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0443 (accessed on 31 January 2022). [264]

OECD (2014), *OECD Employment Outlook 2014*, OECD Publishing, Paris, https://dx.doi.org/10.1787/empl_outlook-2014-en. [72]

OECD (forthcoming), *Artificial intelligence in the labour market: What role for Social dialogue?*, OECD. [260]

OECD (forthcoming), *OECD Framework for Classifying AI systems*, OECD, http://oecd.ai/classification. [1]

OECD.AI Wonk (2021), *Germany's human-centred approach to AI is inclusive, evidence-based and capacity-building*, https://oecd.ai/wonk/germany-takes-an-inclusive-and-evidence-based-approach-for-capacity-building-and-a-human-centred-use-of-ai. [201]

Official Journal of the European Union (2016), *Regulation (EU) 2016/679 of the European Parliament and of the Council*. [165]

O'Keefe, J., D. Moss and T. Martinez (2020), *NYC Regulation on AI Bias in the Workplace*, https://www.natlawreview.com/article/mandatory-bias-audits-and-special-notices-to-job-candidates-new-york-city-aims-to (accessed on 31 August 2021). [150]

O'Keefe, J. et al. (2019), *AI Regulation and Risks to Employers*, Bloomberg Law, http://bna.com/copyright-permission-request/ (accessed on 22 December 2021). [183]

OMB (2020), *Guidance for Regulation of Artificial Intelligence Applications*, Office of Management and Budget, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/inforeg/inforeg/eo (accessed on 26 November 2021). [198]

Opfer, C., B. Penn and J. Diaz (2019), "Punching in: Workplace Bias Police Look at Hiring Algorithms", *Bloomberg Law*, https://news.bloomberglaw.com/daily-labor-report/punching-in-workplace-bias-police-look-at-hiring-algorithms (accessed on 24 November 2021). [180]

Oracle & Future Workplace (2019), *From Fear to Enthusiasm: Artificial Intelligence is Winning More Hearts and Minds in the Workplace*. [6]

Pérez del Prado, D. (2021), "THE LEGAL FRAMEWORK OF PLATFORM WORK IN SPAIN: THE NEW SPANISH "RIDERS' LAW"", *Comparative Labor Law and Policy Journal*, https://cllpj.law.illinois.edu/content/dispatches/2021/Dispatch-No.-36.pdf (accessed on 16 December 2021). [254]

Petkovic, D. (2019), *Application of improved random forest explainability (RFEX 2.0) on data from JCV Institute Lajolla, California*, SFSU. [100]

Pimentel, H. (2021), *Should the government play a role in reducing algorithmic bias?*, Brookings, https://www.brookings.edu/events/should-the-government-play-a-role-in-reducing-algorithmic-bias/ (accessed on 1 April 2022). [206]

Prince, A. and D. Schwarcz (2020), "Proxy Discrimination in the Age of Artificial Intelligence and Big Data", *Iowa Law Review*, Vol. 105/3, https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data (accessed on 18 March 2022). [45]

Quickbooks (2016), *What Do Workers Really Think About GPS Monitoring?*, Quickbooks, https://quickbooks.intuit.com/time-tracking/gps-survey/ (accessed on 14 October 2021). [86]

Quillian, L. et al. (2017), "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 114/41, pp. 10870-10875, https://doi.org/10.1073/pnas.1706255114. [36]

Racine, K. (2021), *AG Racine Introduces Legislation to Stop Discrimination In Automated Decision-Making Tools That Impact Individuals' Daily Lives*, https://oag.dc.gov/release/ag-racine-introduces-legislation-stop (accessed on 23 December 2021). [243]

Raghavan, M. et al. (2019), "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices", https://doi.org/10.1145/3351095.3372828. [58]

Reid, G. (2021), "No Standing and No Recourse: The Threat to Employee Data Under Current U.S. Cybersecurity Regulation", *Touro Law Review*, Vol. 36/4, pp. 1161-1201, https://digitalcommons.tourolaw.edu/lawreviewAvailableat:https://digitalcommons.tourolaw.edu/lawreview/vol36/iss4/18 (accessed on 4 August 2021). [138]

Rivero, N. (2020), *How to use AI hiring tools to reduce bias in recruiting*. [63]

Saint-Martin, A., H. Inanc and C. Prinz (2018), "Job Quality, Health and Productivity: An evidence-based framework for analysis"*, OECD Social, Employment and Migration Working Papers*, No. 221, OECD Publishing, Paris, https://dx.doi.org/10.1787/a8c84d91-en. [70]

Sánchez-Monedero, J. and L. Dencik Sanchez-Monederoj (2019), "The datafication of the workplace"*, Data Justice Project*, ERC. [25]

Sánchez-Monedero, J. and L. Dencik (2019), *The datafication of the workplace*, Cardiff University - Data Justice Lab. [56]

Sánchez-Monedero, J., L. Dencik and L. Edwards (2020), "What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems.". [47]

SAP (2018), *SAP's Guiding Principles for Artificial Intelligence*, https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/ (accessed on 26 November 2021). [263]

Scherer, M. and R. Shetty (2021), *NY City Council Rams Through Once-Promising but Deeply Flawed Bill on AI Hiring Tools - Center for Democracy and Technology*, Center for Democracy and Technology, https://cdt.org/insights/ny-city-council-rams-through-once-promising-but-deeply-flawed-bill-on-ai-hiring-tools/ (accessed on 14 December 2021). [252]

Schwartz, R. et al. (2022), *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence -- NIST Special Publication 1270*, National Institute of Standards and Technology, https://doi.org/10.6028/NIST.SP.1270. [212]

Selbst, A. and J. Powles (2017), "Meaningful information and the right to explanation", *International Data Privacy Law*, Vol. 7/4, pp. 233-242, https://doi.org/10.1093/IDPL/IPX022. [95]

Sharma, K. (2017), "Artificial Intelligence Is a Useful Tool for Business and Consumers", *Business Insider*, https://www.businessinsider.com/artificial-intelligence-is-a-useful-tool-for-business-and-consumers-2017-11?IR=T (accessed on 5 August 2021). [108]

Simonite, T. (2021), *New York City Proposes Regulating Algorithms Used in Hiring*, https://www.wired.com/story/new-york-city-proposes-regulating-algorithms-hiring/ (accessed on 14 December 2021). [250]

Şimşek, C. (2021), *Regulating Artificial Intelligence: Could the EU's "AI Act" lead the way forward?*, Sciences Po Chair Digital, Governance and Sovereignty, https://www.sciencespo.fr/public/chaire-numerique/en/2021/08/20/regulating-artificial-intelligence-could-the-eus-ai-act-lead-the-way-forward-2/ (accessed on 23 December 2021). [230]

Singh, K. (2020), "Barclays being probed by UK privacy watchdog on accusations of spying on staff", *Reuters*, https://www.reuters.com/article/us-barclays-surveillance-probe-privacy-idUSKCN25500P (accessed on 3 August 2021). [171]

So, K. (2019), *A primer on AI fairness. What it is and the tradeoffs to be made | by Kenn So | Towards Data Science*, https://towardsdatascience.com/artificial-intelligence-fairness-and-tradeoffs-ce11ac284b63 (accessed on 9 January 2022). [55]

Somers, M. (2001), , *Journal of Business Ethics*, Vol. 30/2, pp. 185-195, https://doi.org/10.1023/a:1006457810654. [268]

Soper, S. (2021), *Fired by Bot: Amazon Turns to Machine Managers And Workers Are Losing Out*, https://www.bloomberg.com/news/features/2021-06-28/fired-by-bot-amazon-turns-to-machine-managers-and-workers-are-losing-out (accessed on 17 October 2021). [116]

Standards Australia (2020), *An Artificial Intelligence Standards Roadmap: Making Australia's Voice Heard*, https://www.standards.org.au/getmedia/ede81912-55a2-4d8e-849f-9844993c3b9d/R_1515-An-Artificial-Intelligence-Standards-Roadmap-soft.pdf.aspx. [207]

Strategic Foresight Unit (2020), *Data subjects, digital surveillance, AI and the future of work*, European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656305/EPRS_STU(2020)656305_EN.pdf (accessed on 4 August 2021). [135]

Strauss, D. and S. Venkataramakrishnan (2021), "Dutch court rulings break new ground on gig worker data rights", *Financial Times*, https://www.ft.com/content/334d1ca5-26af-40c7-a9c5-c76e3e57fba1 (accessed on 23 November 2021). [177]

Taddeo, M. (2019), "Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity", *Minds and Machines*, Vol. 29/2, pp. 187-191, https://doi.org/10.1007/S11023-019-09504-8. [137]

Taylor, F. (1911), *The Principles of Scientific Management*, Harper & Brothers. [23]

Todolí-Signes, A. (2021), "Spanish riders law and the right to be informed about the algorithm", *European Labour Law Journal*, Vol. 12/3, pp. 399-402, https://doi.org/10.1177/20319525211038327. [256]

Tribunale Ordinario di Bologna (2020), *FILCAMS CGIL Bologna; NIDIL CGIL Bologna; FILT CGIL Bologna vs. Deliveroo Italia S.R.L.*, http://www.bollettinoadapt.it/wp-content/uploads/2021/01/Ordinanza-Bologna.pdf (accessed on 2 January 2022). [186]

TUC (2021), *Technology Managing People - the legal implications*, https://www.tuc.org.uk/sites/default/files/Technology_Managing_People_2021_Report_AW_0.pdf. [272]

TUC (2021), *When AI is the Boss*, Trade Union Congress. [33]

Turner Lee, N. and S. Lai (2021), *Why New York City is cracking down on AI in hiring*, https://www.brookings.edu/blog/techtank/2021/12/20/why-new-york-city-is-cracking-down-on-ai-in-hiring/ (accessed on 1 April 2022). [251]

U.S. Congress (2021), *William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021*, 116th Congress. [213]

U.S. EEOC (2021), *EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness*, U.S. Equal Employment Opportunity Commission, https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness (accessed on 24 November 2021). [182]

UN Human Rights Council (2021), *Report of the United Nations High Commissioner for Human Rights - The right to privacy in the digital age (A/HRC/48/31)*, https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session48/Documents/A_HRC_48_31_AdvanceEditedVersion.docx (accessed on 19 September 2021). [216]

UNESCO (2021), *UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence*, https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence (accessed on 23 December 2021). [199]

United Nations (1948), "Universal declaration of human rights", https://www.un.org/en/about-us/universal-declaration-of-human-rights (accessed on 20 October 2021). [26]

US BLS (2021), *2020 Census of Fatal Occupational Injuries (CFOI)*, U.S. Bureau of Labour Statistics, https://www.bls.gov/iif/oshcfoi1.htm (accessed on 16 March 2022). [122]

US Congress (2022), *H.R.6580 - Algorithmic Accountability Act of 2022*, https://www.congress.gov/bill/117th-congress/house-bill/6580/text?r=1&s=1 (accessed on 17 March 2022). [240]

US Congress (2019), *H.R.2231 - Algorithmic Accountability Act of 2019*, https://www.congress.gov/bill/116th-congress/house-bill/2231. [239]

Van Noorden, R. (2020), "The ethical questions that haunt facial-recognition research", *Nature*, Vol. 587/7834, https://doi.org/10.1038/d41586-020-03187-3. [27]

Veale, M. and F. Borgesius (2021), "Demystifying the Draft EU Artificial Intelligence Act", *Computer Law Review International*, Vol. 22/4, pp. 97-112, https://doi.org/10.9785/cri-2021-220402. [221]

Veale, M. and F. Borgesius (2021), "Demystifying the Draft EU Artificial Intelligence Act", *Computer Law Review International*, Vol. 22/4, pp. 97-112, https://doi.org/10.9785/cri-2021-220402. [231]

Verhagen, A. (2021), "Opportunities and drawbacks of using artificial intelligence for training", *OECD Social, Employment and Migration Working Papers*, No. 266, OECD Publishing, Paris, https://dx.doi.org/10.1787/22729bd6-en. [14]

Verma, S. and J. Rubin (2018), "Fairness Definitions Explained", *IEEE/ACM International Workshop on Software Fairness*, Vol. 18, https://doi.org/10.1145/3194770.3194776. [51]

Wachter, S., B. Mittelstadt and C. Russel (2020), *Why fairness cannot be automated : bridging the gap between EU non-discrimination law and AI*. [49]

Wagner, B. (2019), "Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?", in Emre Bayamlioglu, I. et al. (eds.), *Being Profiled: Cogitas Ergo Sum*, Amsterdam University Press, https://doi.org/10.1515/9789048550180-016/HTML. [266]

Webb, S. (2017), "Houston teachers to pursue lawsuit over secret evaluation system", *Houston Chronicle*, https://www.houstonchronicle.com/news/houston-texas/houston/article/Houston-teachers-to-pursue-lawsuit-over-secret-11139692.php (accessed on 24 November 2021). [192]

WEC-Europe (2021), *Concerns: The inclusive application of Artificial Intelligence on the European labour market*, World Employment Confederation - Europe, https://www.weceurope.org/uploads/2021/02/2020-02-04_Letter-EC-EP-AI-in-Recruitment-Services.pdf (accessed on 25 March 2022). [228]

Weil, D. (2014), *The fissured workplace : why work became so bad for so many and what can be done to improve it*, Harvard University Press. [113]

West, D. (2021), *How employers use technology to surveil employees*, Brookings, https://www.brookings.edu/blog/techtank/2021/01/05/how-employers-use-technology-to-surveil-employees/ (accessed on 14 October 2021). [84]

West, S., M. Whittaker and K. Crawford (2019), *Discriminating Systems: Gender, Race and Power in AI*, AI Now Institute, https://ainowinstitute.org/ discriminatingsystems.html. (accessed on 21 October 2021). [41]

Whitehead, D. (2021), *Why the EU's AI regulation is a groundbreaking proposal*, IAPP Blog post, https://iapp.org/news/a/why-the-eus-ai-regulation-is-a-ground-breaking-proposal/ (accessed on 3 January 2022). [235]

Whittaker, M. (2020), *United States House of Representatives Committee on Oversight and Reform "Facial Recognition Technology (Part III): Ensuring Commercial Transparency & Accuracy"*. [156]

Whittaker, M. et al. (2018), *AI Now Report 2018*, AI Now Institute, http://www.ainowinstitute.org. [61]

Whittaker, M. et al. (2018), *AI Now Report 2018*, AI Now Institute, http://www.ainowinstitute.org (accessed on 31 January 2022). [269]

Wiggers, K. (2021), *Computer vision-powered workplace safety systems could lead to bias and other harms*, https://venturebeat.com/2021/09/21/computer-vision-powered-workplace-safety-systems-could-lead-to-bias-and-other-harms/ (accessed on 4 November 2021). [123]

Wiggin (1999), *Electronic Monitoring of Employees*, Wiggin and Dana, https://www.wiggin.com/publication/electronic-monitoring-of-employees/ (accessed on 17 June 2022). [169]

Wilson, C. et al. (2021), *Building and Auditing Fair Algorithms*, ACM, New York, NY, USA, https://doi.org/10.1145/3442188.3445928. [148]

Wisenberg Brin, D. (2021), *AI's Potential Role in Employee Discipline Draws Attention in Europe*, https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/europe-ai-employee-discipline.aspx (accessed on 18 October 2021). [144]

Wisenberg Brin, D. (2019), *New Illinois Bill Sets Rules for Using AI with Video Interviews*, SHRM, https://www.shrm.org/resourcesandtools/legal-and-compliance/state-and-local-updates/pages/illinois-ai-video-interviews.aspx (accessed on 14 December 2021). [246]

Wood, A. (2021), *Algorithmic Management Consequences for Work Organisation and Working Conditions JRC Working Papers Series on Labour, Education and Technology 2021/07*, https://ec.europa.eu/jrc. [20]

Wood, A. (2021), *Algorithmic Management Consequences for Work Organisation and Working Conditions JRC Working Papers Series on Labour, Education and Technology 2021/07*, https://ec.europa.eu/jrc. [18]

Yeung, K., A. Howes and G. Pogrebna (2020), "AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing", in Dubber, M., F. Pasquale and S. Das (eds.), *The Oxford Handbook of AI Ethics,*, Oxford University Press. [271]

Yuan, L. (2018), "How Cheap Labor Drives China's A.I. Ambitions - The New York Times", *The New York Times*, https://www.nytimes.com/2018/11/25/business/china-artificial-intelligence-labeling.html (accessed on 9 January 2022). [76]

Zhang, D. et al. (2022), *AI Index Report 2022*, Stanford Institute for Human-Centered AI. [101]

Zhong, Z. (2018), *A Tutorial on Fairness in Machine Learning*, https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb (accessed on 24 March 2022). [50]

Zwicker, A. (2019), *Bill Text: NJ A5430*, https://legiscan.com/NJ/text/A5430/id/2023465 (accessed on 14 December 2021). [242]

経団連 (2019), *提言「Ａ Ｉ活用戦略～AI-Ready社会の実現に向けて」を公表*, 一般社団法人 日本経済団体連合会, https://www.keidanren.or.jp/journal/times/2019/0221_02.html (accessed on 26 November 2021). [261]