GLOBAL FORUM FOR AI AND THE SOCIAL SCIENCES

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

State of AI and Society

AI's Quantitative Impacts on Labour Markets

London School of Economics and Political Science
Houghton Street
London WC2A 2AE

This report has been prepared as a contribution to the 2026 edition of the Global Forum for AI and the Social Sciences. The research has been made possible thanks to the generous support of the John D. and Catherine T. MacArthur Foundation.

GlobalAIForum@lse.ac.uk

Lead Author:

Cosmina Dorobantu, Professor in Practice, and Co-Director, Data Science Institute, LSE

Co-authors (in order of contribution):

Hemawathy Balarama, PhD candidate, Department of Economics, LSE
Akash Uppal, PhD candidate, Department of Economics, LSE
Aristeidis Grivokostopoulos, PhD candidate, Department of Economic History, LSE
Maximilian Goehmann, PhD in Management, Department of Management, LSE
Helen Margetts, Professor of Political Science and Public Policy, Department of Government, and Director, Data Science Institute, LSE

We benefited greatly from comments and suggestions from:

Sarah Bana, Assistant Professor of Management, Department of Management, LSE
Alan Manning, Professor of Economics, Department of Economics, and Director of the Labour Markets Programme, Centre for Economic Performance, LSE
Judy Wajcman, Emeritus Professor, Department of Sociology, LSE

Published by The London School of Economics and Political Science

To cite this report:

Dorobantu, C., Balarama, H., Uppal, A., Grivokostopoulos, A., Goehmann, M. and Margetts, H. (2026) *State of AI and society: AI's quantitative impacts on labour markets.* London: London School of Economics and Political Science. Available at: https://doi.org/10.21953/researchonline.lse.ac.uk.00140781

Contents

1. Introduction 4

2. AI definition, types and deployment mechanisms 5
- 2.1 What do we mean by AI? 5
- 2.2 Two types of AI: virtual and physical 5
- 2.3 Three deployment mechanisms: automation, augmentation and commoditisation 6

3. Understanding the literature on AI's labour market impacts 7
- 3.1 Potential versus actual evidence 7
- 3.2 Conceptual underpinnings: how technology reshapes labour demand and supply 8
- 3.3 Methodological approaches 9
  - Research methods underlying potential evidence 10
  - Research methods underlying actual evidence 11
- 3.4 Interpreting the two evidence bases together 13

4. Examining the literature and its empirical findings 14
- 4.1 Introducing two resources that enable a holistic reading of the literature 14
  - The interactive literature tree 14
  - The evidence database 16
- 4.2 Where is the evidence concentrated? 16
- 4.3 What can we tell about the direction and the magnitude of the effects? 21

5. What the evidence shows, disputes and does not yet address 25
- 5.1 What the evidence shows 26
  - Aggregate effect is as of yet muted 26
  - Hiring at the bottom of the career ladder has slowed 26
- 5.2 What the evidence disputes 27
  - Do firm-level results show AI creating jobs or redistributing them? 27
  - Does AI move wages and can we determine why? 28
  - Who is even using AI? 28
- 5.3 What the evidence does not yet address 29
  - What do we know about AI's labour market effects in developing countries? 29
  - How can we study informal work? 29
  - How strong will the reinstatement effects be? 29
  - What happens to workers displaced by AI? 30
  - What topics are ahead of the evidence frontier? 30

6. Conclusion 30

References 32

Appendices 37
- Appendix A. Constructing the interactive literature tree 37
- Appendix B. Constructing the evidence database 39

1. Introduction

Predictions about the relationship between artificial intelligence (AI) and work are running ahead of the evidence. Headlines describe autonomous systems replacing large parts of white-collar work, while the measured effects on employment and wages remain smaller, more concentrated and considerably more uncertain than the headlines imply. In this report, we resist the temptation to endorse either side of the debate on AI's labour market impacts. Instead, we outline the boundaries of our knowledge and establish what claims the evidence currently can and cannot support.

Part of the difficulty in understanding AI's labour market impacts is that "AI" does not describe a single technology. Different systems perform different functions, and their labour market effects depend not only on the underlying technologies, but also on how firms adopt them, which tasks they change and what workers are affected. A large language model (LLM) that drafts text, a decision-support tool that triages cases and a robot moving containers around a warehouse are different economic objects; treating them as one is how we sometimes end up with a single dramatic estimate standing in for a set of narrower, more conditional results.

We begin here by introducing key distinctions that allow us to organise the existing evidence base. The first section outlines our definition of AI and proposes a classification that distinguishes between different types of AI and different deployment mechanisms. These distinctions are important not only for allowing us to compare the findings of different studies, but also for helping us understand where the gaps in our knowledge base are.

Evaluating the evidence for AI's labour market impacts requires an understanding of how economists conceptualise the avenues of impact and of the methodologies that researchers working on the topic employ in their studies. The second section provides this. It also introduces another important distinction, between studies that ask about AI's potential to change work and those that ask how it has already changed work. We refer to these two categories of studies as providing potential or actual evidence on AI's labour market impacts.

The third section turns to what we can learn from a holistic reading of the literature. There is enormous heterogeneity when it comes to the technologies that the existing publications study, the deployment mechanisms they focus on, the time period, geographies and worker populations covered by their data, the methodologies they employ and the results they obtain. To cut through the heterogeneity, we develop two resources that make it possible to examine the literature as a whole: an interactive literature tree that holds all the publications that we identified on AI's labour market impacts, and an evidence database that holds the narrower set of publications taking a position on job quantity or wages.

What emerges from examining the interactive literature tree and the evidence database is neither the wave of displacement promoted by many headlines nor any clean reassurance against it. The literature is far from reaching consensus on the overall direction and magnitude of AI's labour market effects. We find that it is concentrated in a narrow area, mainly focusing on generative AI, white-collar work and developed economies. It also only observes the early years of diffusion, meaning that it is better placed to capture the forces that reduce labour demand than those that offset them. Many of the areas that would determine the broader net effect of AI on the labour market remain understudied.

In the report's final section, we outline what the evidence shows, disputes and cannot yet see. Only two findings command some agreement: the aggregate effect of AI on employment and productivity remains muted, and hiring at the bottom of the career ladder has slowed in some developed economies. We show that neither finding is as settled as it first appears. We also set out the questions the literature actively disputes, including whether firm-level results show AI creating jobs or redistributing them, whether AI moves wages and through which channel, and how adoption should be measured. We close with the questions the literature has yet to reach, from the impacts of AI on developing economies and informal work to the strength of the reinstatement effect, the fate of workers displaced by AI, embodied and agentic systems, and the speed at which labour markets adjust.

Before we proceed, we provide a final note on how the report should be read. What follows is an assessment of the state of the evidence rather than an assessment of how AI impacts labour markets. Where we note that the literature has not established an effect, we make a statement about what has been studied and measured to date. We caution the reader against extrapolating from that statement that no effect exists in reality.

2. AI definition, types and deployment mechanisms

Examining the literature that studies AI's labour market impacts requires first defining terms. This section outlines our definition of AI and introduces the way we classify AI systems and their deployment mechanisms.

2.1 What do we mean by AI?

Following the Organisation for Economic Co-operation and Development's (OECD's) updated definition (2024), we define an AI system as:

> a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments.

AI systems differ in their ability to operate and adapt independently of human input.

Like earlier industrial revolutions, AI is a general-purpose technology that expands capabilities across large sectors of the economy. Nonetheless, this current technological cycle is distinguished by at least three features: AI systems act on non-routine cognitive work that earlier technologies largely left untouched; the technology itself improves at a very high pace; and AI is often adopted by workers directly, rather than only through firm-level investments. These differences mean that the labour market impacts of AI warrant careful and distinct analysis, rather than a straightforward application of the lessons learned from earlier waves of industrialisation.

2.2 Two types of AI: virtual and physical

To organise the evidence, we distinguish between two types of AI: systems designed to operate in physical environments and systems designed to operate in virtual environments.

AI systems designed to operate in physical environments interact directly with the physical world. Examples include autonomous warehouse robots and driverless cars.

AI systems designed to operate in virtual environments are further classified following Bright et al. (2025), who distinguish decision-support AI, generative AI and perceptive AI:

- **Decision-support AI.** Systems that predict, recommend or prioritise outputs to inform human decisions. Examples include systems that help triage hospital patients according to their presenting symptoms.
- **Generative AI.** Systems that create content such as text, images, video, or sounds, often on the basis of a prompt. Examples include using LLMs to draft, summarise, or respond to emails.
- **Perceptive AI.** Systems that process sensory information such as visual or auditory inputs. Examples include computer-vision systems used for identity verification or quality control.

These categories are not mutually exclusive. An AI system designed to facilitate operating in a physical environment may combine perceptive and decision-support capabilities, while a generative AI system may form part of a wider decision-support tool. They are therefore an organising framework rather than a set of perfectly delineated technical classes.

2.3 Three deployment mechanisms: automation, augmentation and commoditisation

Employment effects vary with how the technology is deployed and which tasks it changes. We capture that variation through three deployment mechanisms:

- **Automation.** AI substitutes for labour on a specific task, reducing the labour needed to produce a given level of output. Whether a job disappears depends on how central the task is to the role, how many other tasks remain and whether demand for the output expands enough to offset the fall in labour required. Jobs built around a small number of core tasks (data entry, routine coding, document processing) face stronger displacement pressure than jobs whose value comes from a wide range of activities.
- **Augmentation.** AI complements labour, raising productivity on the tasks the worker performs. The worker produces the same output faster, or the same effort yields larger-scale results. The employment prediction is ambiguous: productivity gains can raise output enough to increase demand for labour or reduce the hours needed to produce the same output. Empirically, augmentation can show up as larger gains for less experienced workers, which flattens the within-occupation performance distribution.
- **Commoditisation.** AI simplifies tasks without automating them, reducing the level of skill required to perform them to an acceptable standard. The worker keeps their job but becomes more substitutable, which weakens their bargaining position and compresses wages. The effects of this mechanism show up in wages and in the erosion of skill premia rather than in employment counts. It differs from augmentation in lowering the skill threshold for acceptable performance, which expands the pool of workers who can do the job and reduces the bargaining position of incumbents, regardless of what happens to their productivity. It is the most recently formalised of the three mechanisms and the least tested empirically, in part due to the difficulty of observing evolving skill requirements in occupations.

On paper, the mechanisms outlined above are distinct and the boundaries are clear-cut. In practice, however, they are not, and two qualifications are worth stating before we proceed.

The first qualification concerns the distinction between automation and augmentation. The two mechanisms are conventionally treated as opposite channels, and the empirical literature we review is organised around that opposition. However, not all researchers accept it. On an alternative reading, a technology that a worker uses productively always raises output per worker, and by the same token always reduces the labour needed for a given level of output, so the two labels describe one productivity change viewed from different directions. We retain the distinction between automation and augmentation because it structures how the evidence is organised, but treat it as a description of how AI is deployed rather than as a prediction about employment.

The second qualification is that AI deployment can involve all three mechanisms simultaneously. The same technology can automate one worker's tasks, augment another's and commoditise a third's, depending on how it is deployed and in what institutional setting. The distinction is often unclear in the empirical literature and many studies do not identify which mechanism is active. This is partly due to a data problem: the task databases underlying most exposure measures were built before AI and score tasks through an automation lens, so augmentation and commoditisation are harder to observe than automation. We return to this in section '3.3 Methodological approaches.'

3. Understanding the literature on AI's labour market impacts

Evaluating the evidence base for AI's labour market impacts requires an understanding of what the evidence shows, how economists conceptualise these impacts and the methodologies the literature employs. This section describes the two ways in which we can classify the existing evidence, how each category of evidence is produced and what each methodological design licenses a reader to conclude.

3.1 Potential versus actual evidence

Research on AI's impacts in the labour market can be decomposed into studies asking about AI's potential to change work and those asking how it has changed work (del Rio-Chanona et al., 2025). We refer to these two categories of studies as providing potential or actual evidence, respectively. We define these categories as follows.

Potential evidence refers to studies that estimate the potential, rather than the actual, impact of AI on labour markets. The central variable in these studies is an exposure score, an automation probability or a simulated counterfactual. The studies in this category shed light on the structural capacity of AI to impact the labour market by affecting worker output and the organisation of labour. They can establish that certain occupations have high task overlap with AI, that some sectors face greater structural exposure than others, and that these gradients are systematic. However, they cannot establish that workers have actually been displaced, that wages have fallen, or that employment has declined.

Actual evidence refers to studies that measure the realised impact of AI on labour markets. The central variable is a realised labour market outcome such as employment, hours worked, or wages. The studies in this category involve analysing administrative, firm, survey, or experimental data and an identification strategy that supports a causal interpretation. They can establish that the introduction of AI systems caused employment or wage changes in identified settings, that particular worker groups were differentially affected, and that effects operated through specific channels. However, they cannot easily establish credible aggregate effects due to the fact that they are narrow in coverage (e.g. one firm, one country, one occupation).

Keeping the two categories of studies distinct is a precondition for reading the evidence base accurately. Both categories of studies are valuable, but they address different questions and operate over different time horizons. Influential studies comprising potential evidence project sizable labour market disruption, while actual evidence has not yet detected a corresponding relationship between AI exposure and realised employment outcomes. Since adoption is ongoing and measured over a short window, some long-term projections are unfalsifiable at this time, just as early null effects cannot rule out future impacts. The Yale Budget Lab, which tracks occupational and industry composition against AI exposure and adoption measures on a rolling basis, reports that these measures currently show no sign of being related to changes in employment or wages, while emphasising a need for better data (Gimbel et al., 2025).

3.2 Conceptual underpinnings: how technology reshapes labour demand and supply

Before turning to the evidence, it is important to outline the conceptual apparatus through which economists analyse how technologies like AI reshape employment. Establishing the conceptual foundations at the outset makes it possible to interpret the empirical findings that follow. It also clarifies why the question of whether AI creates or destroys jobs rarely admits of a simple answer.

At the aggregate level, employment is derived from production. In standard growth accounting, AI's capability gains are represented as higher total factor productivity and, where AI is embodied in equipment, as capital deepening. However, the level of an economy's productivity does not itself pin down how many workers it employs; that depends on how production is organised and how markets respond to lower costs. Thus, whether AI raises or lowers job quantity is ambiguous.

The literature has a deep theoretical spine that long predates generative AI. The task-based framework of Zeira (1998), Autor, Levy and Murnane (2003), and Acemoglu and Autor (2011), extended by Acemoglu and Restrepo (2018) into an explicit contest between the displacement of labour by automation and its reinstatement through newly created tasks, remains the engine of almost every quantity-side argument. In this task-based framework, producing a good or service requires completing a range of distinct tasks, each of which can be carried out by labour or by capital, including AI. This shifts the discussion from automating or augmenting jobs to automating or augmenting tasks.

Notably, within this framework, the real wage may rise as a result of new technologies, even as the labour share of income falls, provided sufficient growth. We thus focus not only on employment, but also on wages. How work is organised within a firm stays fixed in the task-based framework, while in reality, the same technology can raise or lower labour demand based on how firms redesign roles, teams and production processes around the technology. Organisational change is an underappreciated driver of measured effects that go beyond adoption.

With this caveat in mind, a new technological capability sets three forces in motion. Their balance determines labour demand. The first is a displacement effect: as AI takes over tasks previously done by workers, direct demand for labour in those tasks falls. The second is a productivity, or scale, effect: by lowering the cost of production, automation can expand output, incomes and demand, which raises labour demand in the tasks and sectors that are not automated. The third is a reinstatement effect: technologies also create entirely new tasks, and often new occupations, in which labour holds a comparative advantage. Displacement pushes employment down, the scale and reinstatement effects push labour demand up, and the technology's capability alone cannot determine which dominates.

The productivity effect is worth drawing out further, since it often helps decide an effect's sign. A new technology makes workers more productive, such that a given amount of labour yields additional output. Whether that extra output is desired, however, remains a separate question. Since firms adopt the technology if it lowers costs, prices may fall and stimulate demand, allowing wider income effects to shift demand further as the economy grows richer or poorer in aggregate. Employment in an activity then rises or falls according to whether this demand response offsets the reduction in labour needed per unit of output, which may depend on the occupation in question. For some, the demand response dominates and employment expands; in others, employment contracts.

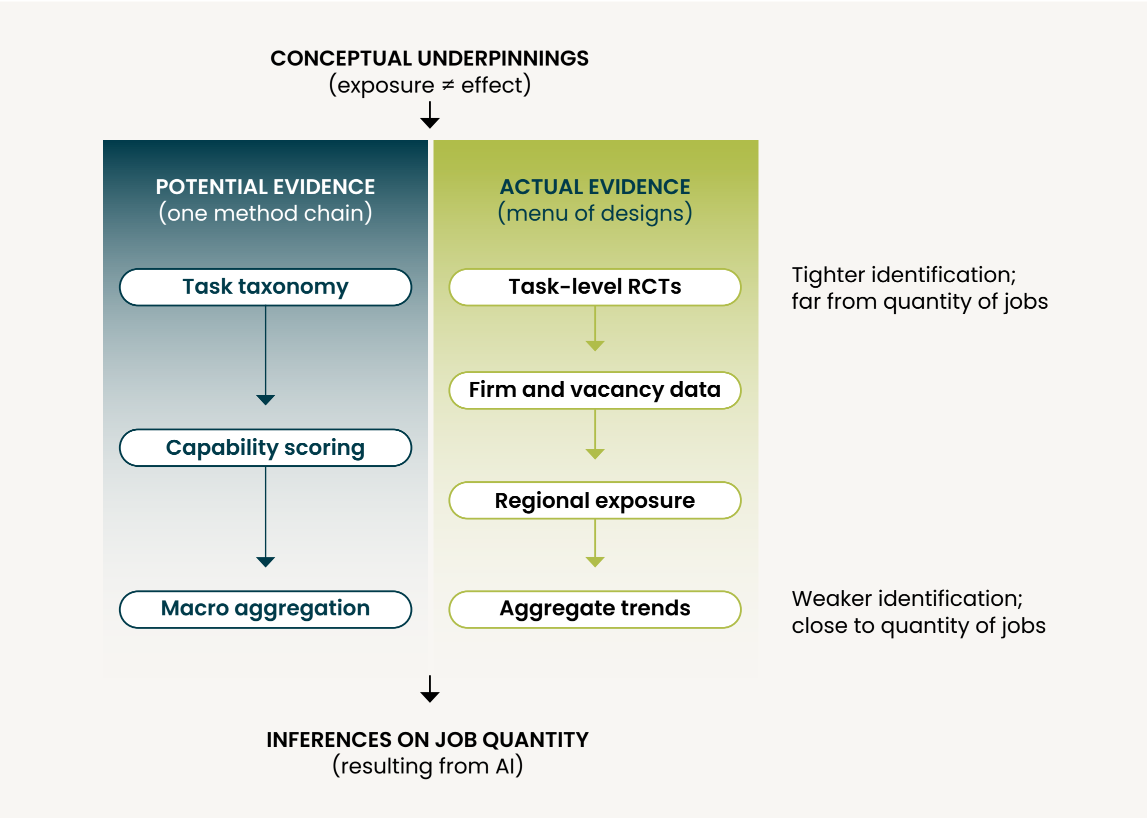
Labour supply may shift along several margins in turn as workers respond to shifting relative labour market returns, including through participation and hours, movement across tasks, occupations and sectors, and investments in new skills. The reinstatement effect, in particular, raises employment only insofar as workers can actually move into and are equipped for the new tasks. Moreover, whether a fall in demand for a task surfaces as fewer jobs or as lower pay depends on how elastically labour is supplied and how flexibly wages adjust. This partly explains why AI's measured effects can appear as changes in headcount in one setting and in wages in another.

The net effect on the quantity of jobs is thus the balance of these demand- and supply-side forces. What determines the sign is a set of measurable parameters that the evidence attempts to speak towards. While this framework is the workhorse model in this literature, it treats tasks as separable. Where production is joint, relying on teams and interdependence among tasks, the neat decomposition breaks down, and richer models are required to capture how AI reshapes work that cannot be unbundled. Moreover, when two occupations jointly produce output and only one is amenable to automation, demand for the non-amenable occupation can rise as it becomes the bottleneck whose complementarity to the now-cheaper task raises its value. This can lead to propagation effects and cross-occupation spillovers that have historically been sizable. Furthermore, mechanisms vary widely across time horizons: displacement can be rapid, but the scale and reinstatement effects work through investment, firm entry, the creation of new tasks and the reallocation of workers, which operate with lags and adjustment frictions. These should be accounted for when interpreting estimates that both evidence types present.

3.3 Methodological approaches

Having outlined the literature's conceptual underpinnings, this section introduces the methodological approaches of studies within the potential and actual evidence categories. Figure 1 summarises the methodological approaches that we cover for each evidence category.

Figure 1. Framework for potential and actual evidence



Research methods underlying potential evidence

The studies that fall within the potential evidence category rely on at least one of the methodologies in a three-step pipeline, with assumptions passed along:

1. **Task taxonomy.** Work is mapped into tasks, almost always via O*NET, a US database that renders work as a structured list of activities that can be classified and weighted, often using time use data to reflect how much of a job each task actually occupies (Autor, Levy and Murnane, 2003). Notably, the occupational structure may not apply to lower-income settings and faces difficulties in updating to reflect the new tasks that drive reinstatement. Broadly defined tasks like "data analysis" that span everything from routine tabulation to analytical interpretation, with an attached susceptibility score, may mask improvements in AI capability.

2. **Capability scoring.** Tasks are scored for AI susceptibility by expert judgement extrapolated with a classifier (Frey and Osborne, 2017), matching patent text to task text (Webb, 2019), or using human and LLM raters applying a shared rubric (Eloundou et al., 2024). One limitation is that each method is read by default through an automation lens, even where the underlying capability may augment labour. Moreover, when a task lumps together components that AI can do and components it cannot, a single score does not convey these more granular differences within tasks. Design decisions may heavily influence derived estimates. For instance, scoring whole occupations for automatability put nearly 47% of US jobs at high risk (Frey and Osborne, 2017), but scoring tasks within them using individual-level survey data on what workers actually do in their jobs reduced the OECD figure to only 9% (Arntz, Gregory and Zierahn, 2016). It is worth stressing that this is a limitation of using exposure scores downstream, since they identify where AI has the potential to affect tasks, rather than necessarily providing declarative evidence of where displacement will occur.

3. **Macro aggregation.** Scores are aggregated into projections. Acemoglu (2025) serves as an example: pairing the Eloundou et al. (2024) exposure estimates with task-level cost savings drawn from several AI productivity experiments, then aggregating through the task-based framework, yields an upper bound on the economy-wide productivity gain of half a percentage point of total factor productivity over a decade. This exercise suggests gains are unlikely to translate into sizable wage increases, that labour income inequality is unlikely to reduce, and that some groups may experience small real wage declines. Other exercises, including scenario analyses and computable general equilibrium models produced by consultancies and international organisations, generate larger headline figures. Estimates occupy a wide range: projections of AI's aggregate effect span from the negligible to the transformative.

Centrally, the outcomes of potential evidence studies in this category are conditional forecasts, not measurements.

A related strand of recent literature analyses usage data generated by AI providers themselves, mapped onto standard occupational taxonomies. Consider, for instance, the Anthropic Economic Index (Handa et al., 2025; Massenkoff and McCrory, 2026), which uses privacy-preserving techniques to document AI usage in occupations over time, including the rate of augmentation in tasks mapped to O*NET, learning curves in adoption of Claude and adoption of AI systems across countries. Similar efforts include Google's AI & Economy ATLAS (Iscenko et al., 2026), which maps Gemini interactions to Bureau of Labor Statistics (BLS) occupation codes and O*NET tasks, and Chatterji et al. (2026), who link ChatGPT Enterprise account records to worker roles, task classifications and the financial statements of publicly listed adopters. While these are among the most promising efforts to observe AI usage at scale, they represent largely descriptive evidence of adoption and use, rather than a clear identification of AI's impact on employment or wages. Each study also observes a single provider's users. Studies based on data from Google and Anthropic do not have visibility into non-users. Chatterji et al. (2026) compare adopting with non-adopting public companies, but observe workers only inside adopting firms. Measured task composition and automation shares, therefore, reflect the user base of the company supplying the data and the classification scheme it uses. They are not mirror images of the underlying structure of the economy.

Research methods underlying actual evidence

The studies that fall under the actual evidence category rely on a menu of methodological designs. In Figure 1, we order them by the level at which each methodological approach observes outcomes, since effects at one level need not carry over to another. As the outcome moves closer to the object of interest – net employment – the ability to attribute the effect to AI weakens due to sparse methods for rigorous causal identification.

Task-level randomised controlled trials (RCTs) use productivity shocks for clean measurement of changes in output per worker on a narrow band of tasks, but struggle in generalising to employment more broadly. Identification comes either from a controlled experiment in which the researcher randomises access to and training for an AI tool, or from a natural experiment that exploits quasi-random variation in access to AI systems, such as a staggered enterprise rollout or the sudden release of an LLM. In both cases, as-good-as-random assignment strips away the market response. That is, whether a task completed twice as quickly means half as many workers or twice as much output depends on the elasticity of demand for output, which is held constant. The evidence in this area concentrates on tasks like writing (Noy and Zhang, 2023), coding (Peng et al., 2023), consulting (Dell'Acqua et al., 2026) and customer support (Brynjolfsson, Li and Raymond, 2025). Its external validity is bounded as a result of its clean identification: a single task, a short horizon, an early tool and a setting that excludes the firm entry, reallocation and price adjustment through which the employment effects of AI tools may operate.

Firm and vacancy data studies move down a rung to hiring, separations, headcount and, importantly, the latent demand revealed by job vacancies. The appeal is that vacancies shift before headcounts do, offering a credible early quantity signal of AI's labour market impacts. Three study designs recur: near-universe vacancy data linking an establishment's AI exposure to the composition of its hiring (Acemoglu et al., 2022); firm-level adoption panels, drawing on sources such as the US Census Annual Business Survey, which link measured adoption to subsequent employment and growth (Acemoglu et al., 2022; Babina et al., 2024); and natural experiments that use the sudden release of ChatGPT as a datable shock in online labour markets (Hui, Reshef and Zhou, 2024; Teutloff et al., 2025). While this approach pairs a sharp, datable shock with a directly quantity-relevant outcome, adopting firms and platform workers are selected and unrepresentative, and estimates are partial equilibrium in the sense that jobs lost at an exposed firm or platform do not account for whether those workers are reabsorbed elsewhere. Moreover, firm-level growth among adopters may reflect market share taken from competitors rather than net job creation, so an expanding adopter is not necessarily evidence of expanding aggregate employment. The November 2022 release of ChatGPT was furthermore not a clean natural experiment. It coincided with other shocks and does not hold the wider environment fixed, limiting the ability to assume that other factors remain unchanged. Recent evidence suggests that AI-exposed jobs began deteriorating before this shock (Frank et al., 2026).

Regional intensity designs climb higher still, comparing local labour markets that differ in their exposure to AI, with exposure constructed from each region's mix of industries and occupations in the manner of shift-share or Bartik instruments (Acemoglu and Restrepo, 2020). Because they observe many local markets at once, they are the first rung at which reallocation of activity across places becomes visible. The AI-specific evidence at this level remains thinner and more mixed than the mature industrial automation literature, partly because AI exposure is harder to localise than a physical machine bolted to a factory floor. The binding limitation is conceptual rather than statistical: by comparing regions, these designs difference out anything common to the whole economy, so they recover AI's effect on a region's employment relative to other regions, rather than its aggregate effect. A finding that exposed areas lost jobs relative to unexposed ones is therefore consistent with a national total that rises, falls or holds, depending on the general equilibrium adjustment the cross-regional comparison is built to remove.

Aggregate trends measure economy-wide employment and productivity through macroeconomic time series and trend decomposition. This rung is closest to the target variable of interest, but weakest on attribution. The central observation so far is an absence: despite rapid advances in capability, aggregate productivity growth in the major economies shows no unambiguous acceleration. This may be because diffusion is still too early, gains are mismeasured or offset by adjustment costs, or realised effects are modest to date. Nor can movements in aggregate employment be attributed to AI specifically, since it is only one of many forces acting on the series alongside the pandemic's labour market aftermath, monetary tightening, demographic change and the long tail of pre-AI automation. Studies in this group attempt credible identification using real data, but often cannot isolate AI's contribution to the variable of interest directly; when this is the case, conclusions are correspondingly closer to informed inference than identified estimates and therefore within our classification system end up as potential evidence.

3.4 Interpreting the two evidence bases together

The potential and actual evidence branches provide complementary, rather than competing, answers as to how AI impacts labour markets. Potential evidence has the coverage that actual evidence lacks: it can speak to every occupation in the economy at once, including occupations where exposure has barely begun, and it is the main tool available for looking ahead. Actual evidence has the opposite profile. It examines the past to measure what has already happened where AI was deployed. As one climbs the ladder towards the object of interest (quantity of jobs), the rungs of causal identification and attribution of findings solely to AI systems are weaker.

Three cross-cutting considerations should be applied to any inference drawn from the evidence base of AI's labour market impact:

- **Horizon.** Every piece of actual evidence samples an early, partial phase of a technology still in diffusion. The scale and reinstatement effects that theory identifies as the main offsets to displacement operate through investment, firm entry, new-task creation and worker reallocation, unfolding with long lags. Current evidence is therefore weighted towards the displacement margin, which materialises first, and is least able to observe the reinstatement margin, which materialises last. Absence of a large realised effect to date is not evidence that a large effect will not eventually arrive, in either direction. Economic history suggests that new technologies take years or, at times, even decades to diffuse, be adapted and generate new tasks that contribute to labour absorption (Autor, 2015), providing reason to view early evidence as a snapshot of an unfinished process.

- **Deployment.** Whether AI substitutes for a worker or augments them is decisive for the quantity of jobs, yet the methods illuminate it unevenly. Exposure indices are, by default, read through an automation lens and tend to be silent on augmentation; productivity RCTs typically study AI in the worker's hands, and so speak more to augmentation and complementarity; firm and vacancy data capture net demand but rarely reveal which mechanism drove it. An inference about job quantity is only as credible as its account of which deployment channel the underlying studies observed, and sweeping conclusions that generalise from automation-weighted exposure to the whole of AI's labour market effect may overreach. Not only that, but the relevant counterfactual is itself fast moving: as AI becomes mainstream, the comparison group of no-AI becomes contaminated, such that an identical study run earlier versus later is not estimating effects against the same comparison group. Study designs that exploit loss of access rather than adoption of new tools will therefore identify different quantities.

- **Coverage.** The evidence base is overwhelmingly drawn from high-income economies. This is a consequence of where the task taxonomies are built, where the technology is developed, where the trials are run, and where rich vacancy and administrative data exist. Task content, the cost of labour relative to AI, the pace of adoption and the capacity to create new tasks all differ across the income distribution. Exposure scores and realised estimates calibrated to developed labour markets may not transfer to the economies in which most of the world's workers work, limiting the external validity of existing evidence.

This section outlined the ways in which the literature approaches the question of how AI will impact labour markets. In doing so, it provided the reader with an understanding of the distinction between actual and potential evidence, of how economists conceptualise the impact of AI on labour markets and of the methodologies that the literature employs. Armed with this understanding, we are now ready to turn to what the literature reveals about the effects of AI on labour markets.

4. Examining the literature and its empirical findings

In this section, we map the literature on AI's labour market impacts and outline where the blind spots in our collective knowledge lie. We also set out to answer a focused question, **which is what does existing research say about AI's effect on the quantity side of the labour market (employment and wages)**, as distinct from the related but different questions of its effects on productivity, job descriptions, or worker sentiment?

When this question arises, public discussion often gravitates towards a single headline figure chosen for its impact: a share of jobs at risk, or a wage effect attributed to AI. In this section, we offer a comprehensive account of why the literature does not converge on such a figure.

There is enormous heterogeneity among the publications that examine AI's labour market impacts. They study LLMs and coding assistants, industrial robots, algorithmic decision-making tools and general-purpose AI treated as a single category. They attribute effects to automation, augmentation, commoditisation, or to all three. They report job losses, job growth, a mixed effect, or no effect at all. They look at actual evidence such as realised outcomes at real firms, or potential evidence such as a projected exposure without an observed outcome or a theoretical prediction. They cover different countries, worker populations and outcome definitions. They rely on data collected years ago or in the past few months. Each of these differences narrows how far a given result can be generalised beyond the setting it studies.

We build two complementary resources to cut through this heterogeneity and enable a holistic reading of the literature: an interactive literature tree that holds all the publications that we identified on AI's labour market impacts and an evidence database that holds the narrower set of publications taking a position on job quantity or wages. In the subsections that follow, we introduce each resource and then outline the insights that emerge when we examine them together.

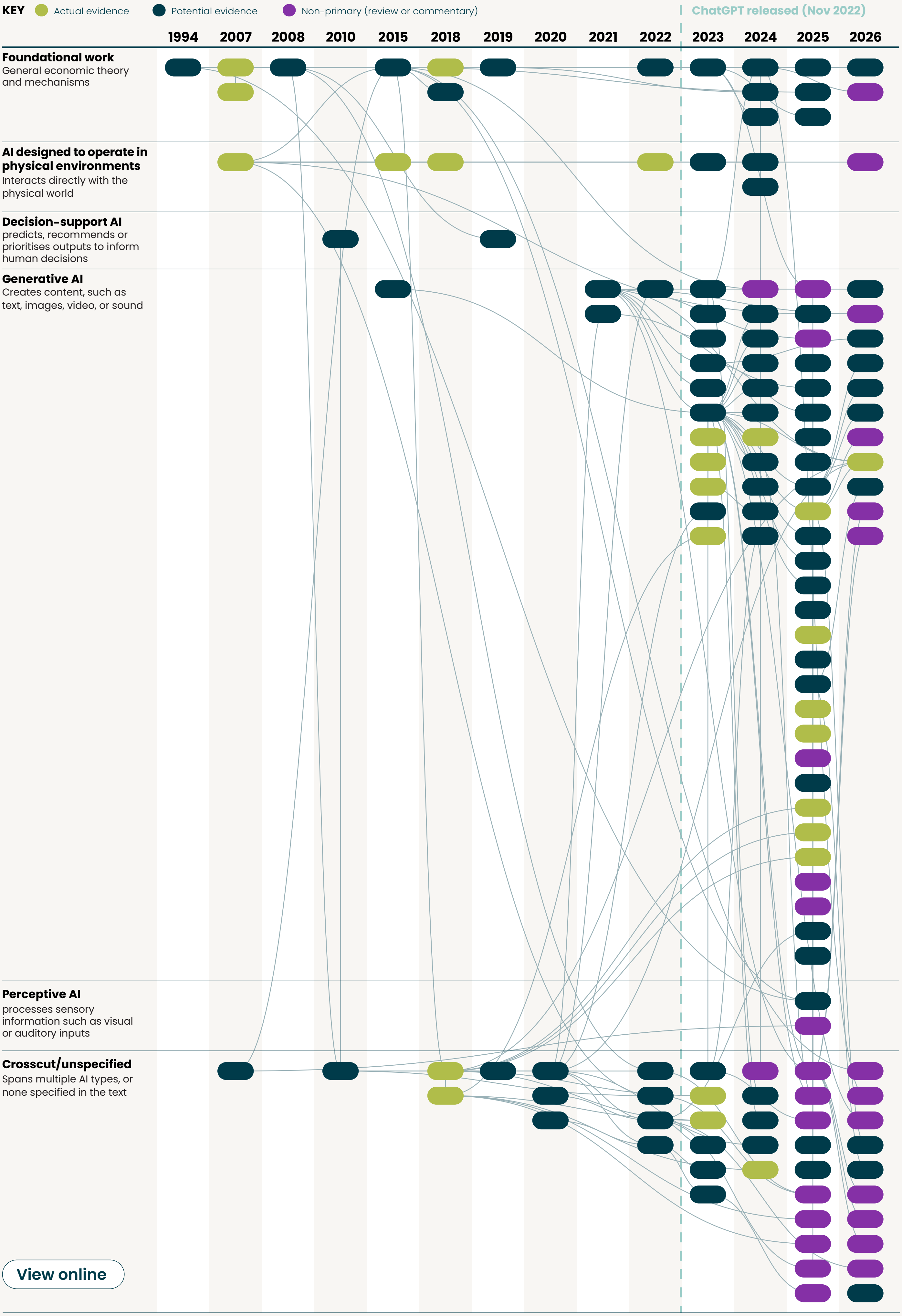
4.1 Introducing two resources that enable a holistic reading of the literature

The interactive literature tree

We identify 138 publications that examine AI's impact on labour markets, following a thorough thematic search. The interactive literature tree draws from a Zotero library that we created based on Google Scholar and SSRN searches, citation lineage snowballing around key works, and targeted retrieval of work by international organisations, research labs and industry. Appendix A sets out the full search strategy, the academic disciplines represented and the date at which we stopped adding publications.

The interactive literature tree (Figure 2) organises the 138 publications along three dimensions: the type of AI they study, the vintage of the underlying data and whether the evidence is potential or actual. Appendix A defines each dimension in full and provides instructions on how to read the tree. Clicking on a study in the interactive tree reveals its academic lineage. The tree also labels each publication by the geography it covers and the AI deployment mechanism at play.

Figure 2. The interactive literature tree



The evidence database

The interactive literature tree holds every publication we identified that relates to AI and labour markets, from theoretical papers to empirical studies, literature reviews and policy reports. To answer the question of how much AI impacts employment and wages, we rely on a narrower selection of the literature. We apply 3 successive filters to the 138 publications in the interactive literature tree to remove non-research publications, reviews and syntheses, as well as studies that take no position on the direction or magnitude of AI's effect on job quantity or wages. This leaves a final set of 74 publications. Interested readers can explore individual and collective findings from this set in our evidence database. Appendix B describes the three filters in full, reports the number of publications each filter removes, and sets out what information we collate and how we label each publication.

Examining the interactive literature tree and the evidence database together reveals what the literature has focused on and what it has not. As the tree maps the full set of publications that we identified, its shape is itself evidence about the state of our collective knowledge, and about where that knowledge is missing. The evidence database then allows us to put numbers behind our collective blind spots for the slice of the literature that takes a position on the direction or magnitude of AI's effects on employment and wages. Each publication in the database is coded along key dimensions, so counting the publications within each dimension produces a coverage map that shows where the literature's attention is distributed. In the subsection that follows, we outline where the evidence is concentrated and what coverage patterns emerge when we are able to examine the literature holistically.

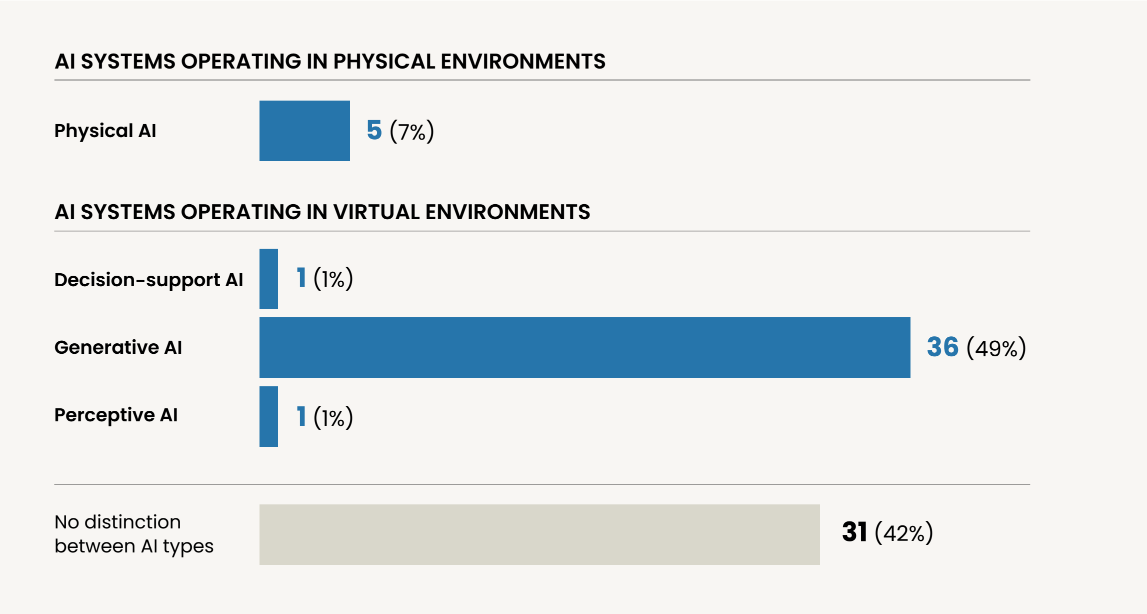
4.2 Where is the evidence concentrated?

Most studies focus on generative AI. A first look at the interactive literature tree reveals that the overwhelming majority of the literature about AI's labour market impacts focuses on generative AI. Decision-support AI appears in the literature through a couple of firm-value and firm-adoption studies. The category for perceptive AI is also almost empty of dedicated quantity-side work.

The literature on AI systems designed to operate in physical environments forms a separate and much older branch, which uses different instruments to measure labour market effects and different technology vintages. As deployment shifts towards multimodal, agentic and embodied AI systems, the branches of the interactive literature tree that are currently sparse are likely to matter substantially to labour market outcomes.

The evidence database puts numbers behind this pattern for the publications that take a position on the direction or magnitude of AI's impacts on employment or wages. Of the 74 publications in the database, 31 (42%) do not distinguish between the type of AI they study. Of the publications that identify the type of AI they examine, 36 focus on generative AI. Out of these 36 publications, agentic AI, which comprises systems that plan and execute multistep tasks with limited human oversight, is the primary subject of exactly one paper. A total of five publications study AI systems that are designed to operate in physical environments. Figure 3, from which these counts are derived, reveals that the evidence base is yet to catch up with the capability frontier.

Figure 3. Type of AI system studied



The type of AI studied has implications for the type of worker studied. The focus on generative AI is a focus on white-collar work. A careful reading of the literature reveals that the white-collar versus blue-collar split maps almost exactly onto the interactive literature tree's division between generative AI and AI systems designed to operate in physical environments. The existing literature measures the effects of AI on these two categories with different instruments and technology vintages.

The actual evidence on the labour market impacts of generative AI is overwhelmingly about cognitive, desk-based work: customer-support agents (Brynjolfsson, Li and Raymond, 2025), professionals doing writing tasks (Noy and Zhang, 2023), management consultants (Dell'Acqua et al., 2026), software developers (Peng et al., 2023; Cui et al., 2026), and early-career and junior workers in exposed white-collar occupations (Brynjolfsson, Chandar and Chen, 2025; Klein Teeselink, 2025).

Blue-collar and manual work enters the tree almost entirely through the older, separate robotics literature: Acemoglu and Restrepo (2020) on industrial robots, with more recent activity-based and meta-analytic evidence (Caselli et al., 2025; Guarascio, Piccirillo and Reljic, 2025), and a forward-looking cluster of manufacturing strategy work that is still largely about potential rather than actual impacts (Make UK, The Manufacturers' Organisation, 2025; Manufacturing Technology Centre, 2026).

The practical consequence is that the literature cannot speak about the effect of AI on blue- and white-collar work in the same evidentiary voice. The former rests on a mature robotics literature measuring a pre-generative technology, the latter on a young generative literature measuring a more recent technology. Since the choice of which technologies to study affects the evidence uncovered, the labour market questions specific to robotics, autonomous vehicles and embodied AI more generally, which apply to much blue-collar work, remain comparatively unanswered.

Potential evidence dominates. The literature that looks at potential evidence is far denser and better developed than the literature that looks at actual evidence. We know a great deal about the tasks that AI could affect and much less about the jobs that it has affected.

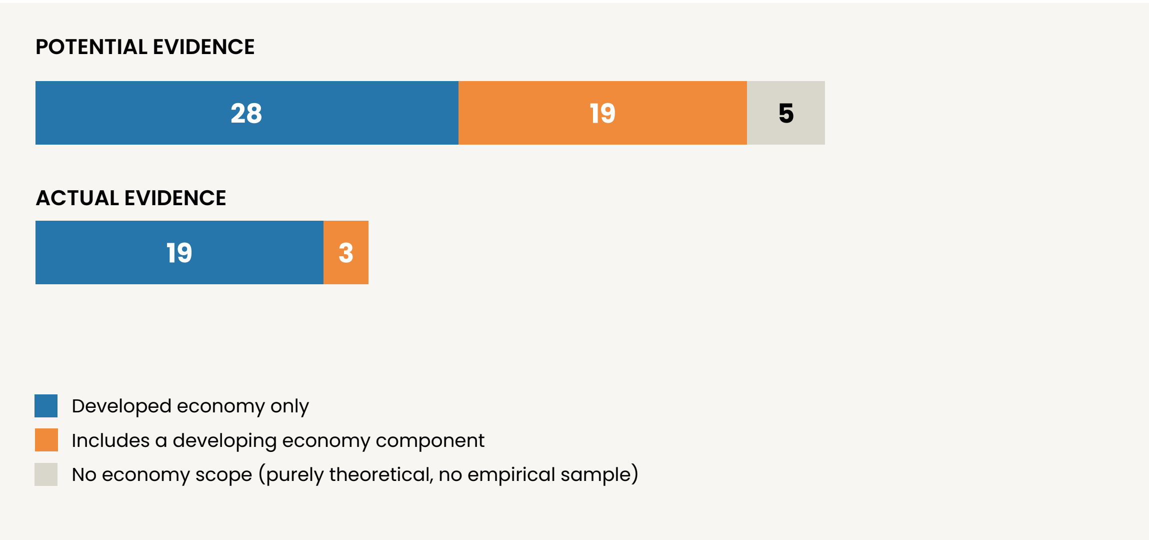
The imbalance is most pronounced where the stakes are highest, such as identifying the labour market impacts of the newest capabilities or pinpointing labour market effects within developing economies.

The evidence database confirms the imbalance for the slice of the literature that takes a position on the direction or magnitude of AI's impacts on job quantity or wages. Of the 74 publications in the evidence database, 52 (70%) classify as potential evidence and 22 (30%) as actual evidence.

Actual evidence is scarcest for developing economies. In the interactive literature tree, the developed versus developing economy split is evident when it comes to both potential and actual evidence. The literature on actual evidence is drawn almost entirely from developed economies: mostly the US, with further contributions from the UK, the European Union (EU), Canada, Denmark, Italy and France. Some developing economy coverage exists, but it sits overwhelmingly in the potential evidence category: the International Monetary Fund (IMF) exposure and preparedness work (Pizzinelli et al., 2023; Cazzaniga et al., 2024; Cerutti et al., 2025), the Latin American simulations (Ciaschi et al., 2025; Egana-delSol and Bravo-Ortega, 2025) and the between-country divergence literature (UNDP, 2025). The countries most likely to face constrained adoption and weak measurement are therefore those for which actual evidence is scarcest. Any claim for the impact of AI on labour markets in developing economies currently rests on prediction rather than actual observation.

The same gap appears in the evidence database, where the two evidence categories are not distributed evenly across the world (Figure 4). Only 3 of the 22 actual evidence studies that take a position on the direction or magnitude of AI's impacts on employment or wages include a developing economy component. None of them is a dedicated developing economy study. Hui, Reshef and Zhou (2024) study a global online labour platform whose freelancers span both developed and developing economies; Guarascio, Piccirillo and Reljic (2025) pool robot employment estimates across studies that include Brazil, Colombia, Mexico and China alongside a mostly developed economy sample; and Klein Teeselink and Carey (2026) analyse job postings across 39 countries spanning both OECD and Brazil, Russia, India, China and South Africa (BRICS) membership. Nineteen of the 52 potential evidence publications include a developing economy component.

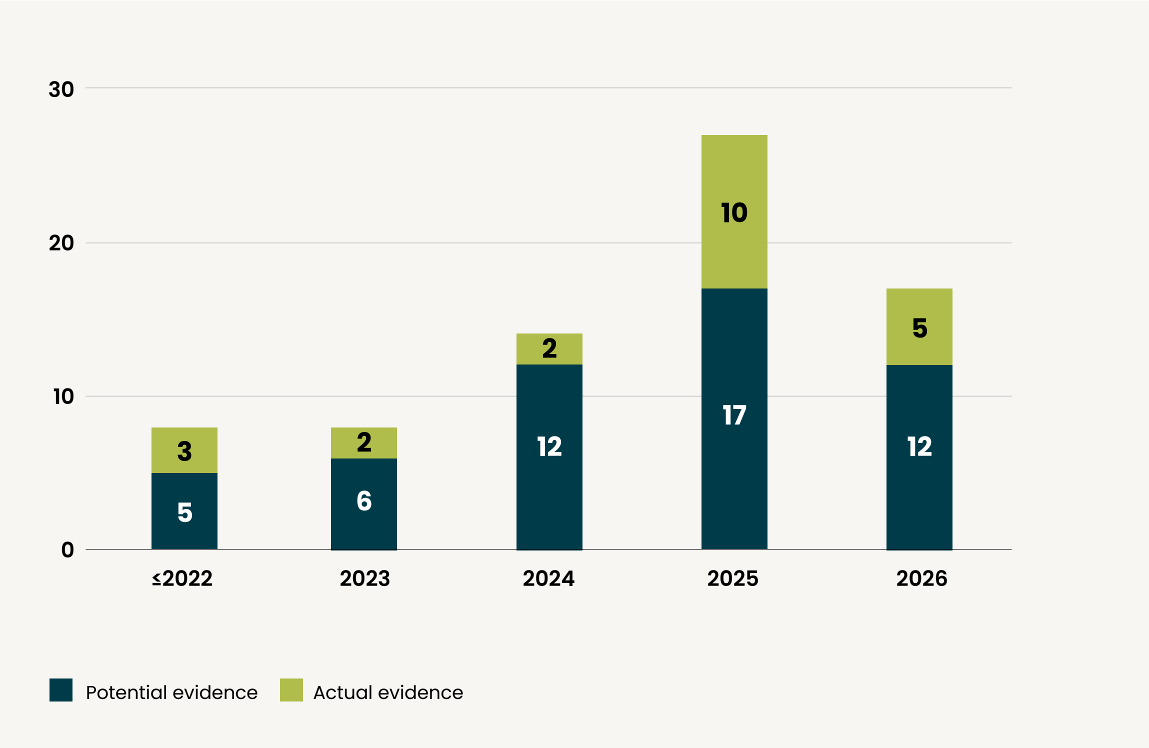
Figure 4. Type of evidence available for developed and developing economies
All potential and actual evidence publications by economy scope



Most studies are recent. The launch of ChatGPT in November 2022 led to a boom in the number of studies looking at AI's labour market impacts. The overwhelming majority of studies use data from 2023 onwards, which means that most of the existing evidence base is young and short panelled. The material that precedes 2018 is almost entirely the theoretical spine and the literature on robotics.

Within the evidence database, actual evidence is also concentrated in a narrower and more recent time window. Potential evidence publications span 2017–2026 and actual evidence publications span 2020–2026, with 15 of the 22 publications (68%) published in 2025 or later (Figure 5). The concentration has a specific cause: most identification strategies in this literature exploit ChatGPT's November 2022 release as a datable shock. This sets a floor on how early a causal design can appear. The pattern that we identify here reinforces the horizon consideration we highlighted in '3.4 Interpreting the two evidence bases together.' Even the publications that claim to establish causality observe only the very first years of diffusion. This is the window in which theory predicts that displacement materialises, while reinstatement has not yet had time to emerge.

Figure 5. Publications generating potential and actual evidence, by year

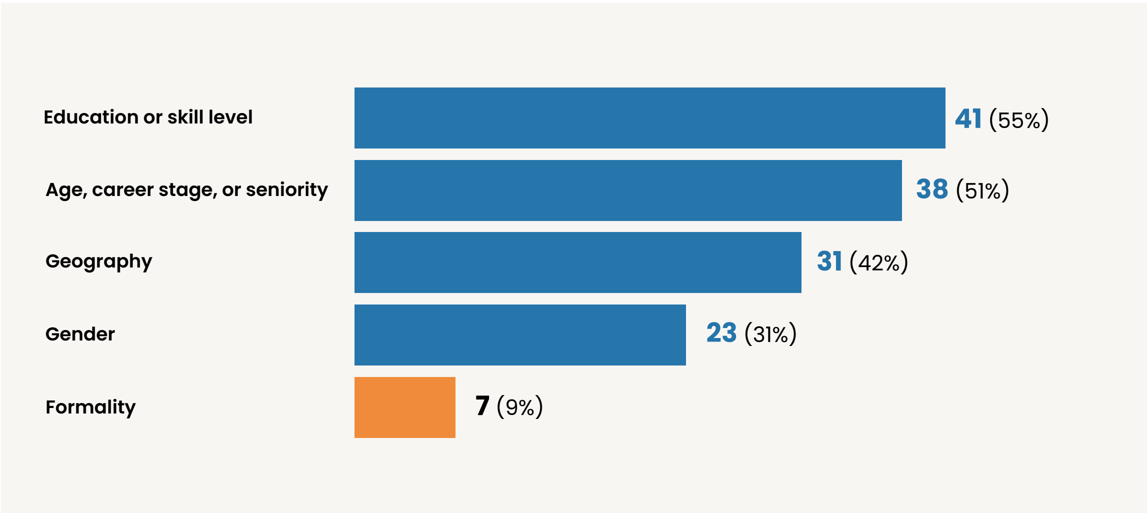


Note: actual evidence clusters after ChatGPT's November 2022 release.

Informal work is almost invisible in the literature. The formal versus informal split is the starkest absence in the interactive literature tree. Almost every study that looks at actual evidence measures formal sector, paid employment. This is because the exposure and outcome measures that the literature relies on are built on formal task taxonomies, stable job titles and payroll or administrative records. Informal work (such as own-account activity, household enterprise, casual and platform-mediated labour) is almost invisible in the interactive literature tree, and where it appears it does so through global exposure modelling rather than measured, actual outcomes. This is both a conceptual and a measurement problem, when one considers the fact that the unit the literature measures – the formal occupation – is not the unit in which much of the world's work is organised.

The evidence database records this absence directly, in the cleavages along which the publications that we track disaggregate their findings. Some level of distributional analysis is common, but almost never covers informality. Fifty-nine of the 74 publications (80%) report at least one distributional finding. As a share of all 74 publications, the most common cleavages are education or skill level (55%) and age, career stage, or seniority (51%). Geography follows at 42%, and gender at 31%. Formality, whether effects differ between formal- and informal-sector work, appears in seven papers (9%), the lowest of the five cleavages we track (Figure 6).

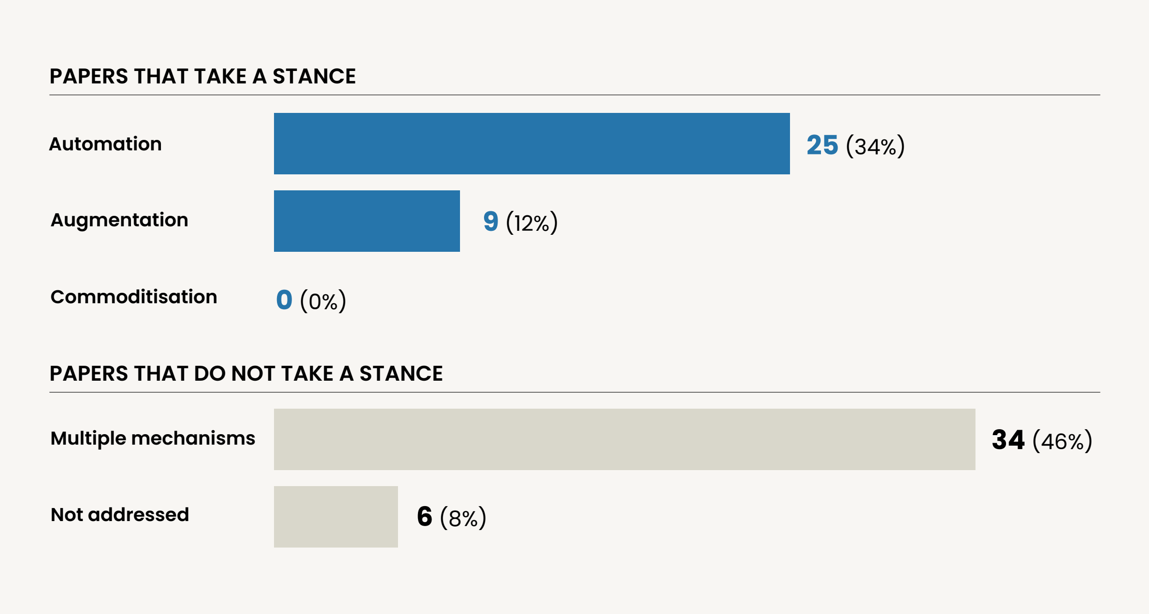
Figure 6. Distributional cleavages studied



Note: cleavages are not mutually exclusive – a publication may report more than one, so counts do not sum to 74.

Commoditisation is modelled but scarcely measured. The three AI deployment mechanisms we set out earlier in the report identify commoditisation as one mechanism through which AI simplifies tasks without automating them. This reduces the level of skill required to perform them to an acceptable standard. Both the interactive literature tree and the evidence database show that it is the least studied AI deployment mechanism by a wide margin. No publication in the evidence database is coded to commoditisation as its primary deployment mechanism (Figure 7). The channel is not entirely absent from the literature: Althoff and Reichardt (2026) model it under the label of simplification, and Fukui et al. (2026) theorise a related standardisation channel. However, it is only present in theory and simulations; to our knowledge, no study measures a realised wage or employment outcome and attributes it directly to commoditisation. Claims that AI is already compressing fees and flattening career ladders in law, consulting and other professional services are circulating in industry commentary, but the current evidence base has a limited basis on which to confirm or refute them.

Figure 7. Publications by AI deployment mechanism



Read together, the interactive literature tree and the evidence database describe a literature whose coverage is concentrated in a narrow region. Existing research focuses mostly on generative AI, on potential evidence rather than actual evidence, and on the very first years of technology diffusion. A close read of the literature reveals four further gaps in our collective knowledge: the evidence is far thicker for white-collar versus blue-collar work, for formal versus informal work, for developed versus developing economies and for automation and augmentation versus commoditisation. Together, these patterns show where our knowledge remains thin. What the literature has is a narrow beam that sheds light on a small part of the global labour market. Much of what we would want to see sits outside it, and it is there that claims about AI's effects on employment and wages should be treated with the most caution.

4.3 What can we tell about the direction and the magnitude of the effects?

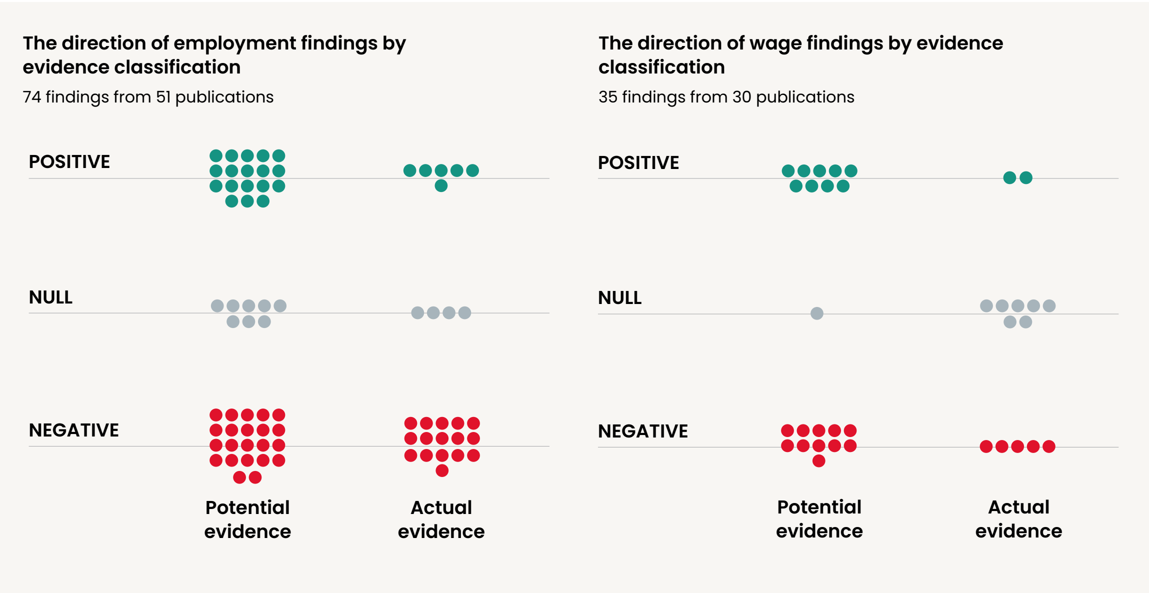
At the beginning of the section, we set out to answer a focused question – what does existing research say about AI's effect on the quantity side of the labour market (employment and wages)? The answer is that there is no consensus in the literature on the direction and magnitude of these effects.

In an ideal world, we would want to put all the estimates that the literature produced onto one graph to show the variety of results in the publications we examined. Unfortunately, that is not feasible. Aside from the differences between the studies that we identified above, the 74 publications in the evidence database report job quantity and wage effects in a variety of units, such as percentage changes, percentage points, raw headcounts, elasticities and correlations.

We find two ways to work around the difficulty of forcing all estimates onto a common scale. The first is to produce figures that only compare the direction of the finding (positive, null, negative), which deals with the difficulty of harmonising units of measurement. Of the 74 publications in the evidence database, 60 (81%) report a directional finding: 51 for employment and 30 for wages, with 21 publications contributing to both. The second is to compare the estimated magnitudes directly but restrict our attention to the subset of estimates whose unit is specifically a percentage change. This magnitude comparison rests on a smaller base: of the 60 publications with a directional finding, only 28 express it as a percentage change. We report the resulting figures and findings below, acknowledging that this approach ignores precision, sample size and study quality.

A higher number of findings lean towards a negative finding for both employment and wages, whether we look at actual or potential evidence. Figure 8 shows that for both actual and potential evidence, the findings reported by the 74 publications in the evidence database are split between positive, null and negative. While neither branch of the literature reaches a consensus on the direction of AI's effects for either employment or wages, both branches produce a higher number of negative findings compared to positive findings, with null findings only being ahead of negative ones for actual wage evidence. At this point in time, the negative lean in the findings is not surprising. As we noted earlier in this report, in the years that immediately follow the introduction of a disruptive technology, estimates are more likely to capture the displacement margin, which materialises first, rather than the reinstatement margin, which materialises much later.

Figure 8. Direction of employment and wage effects findings by evidence classification

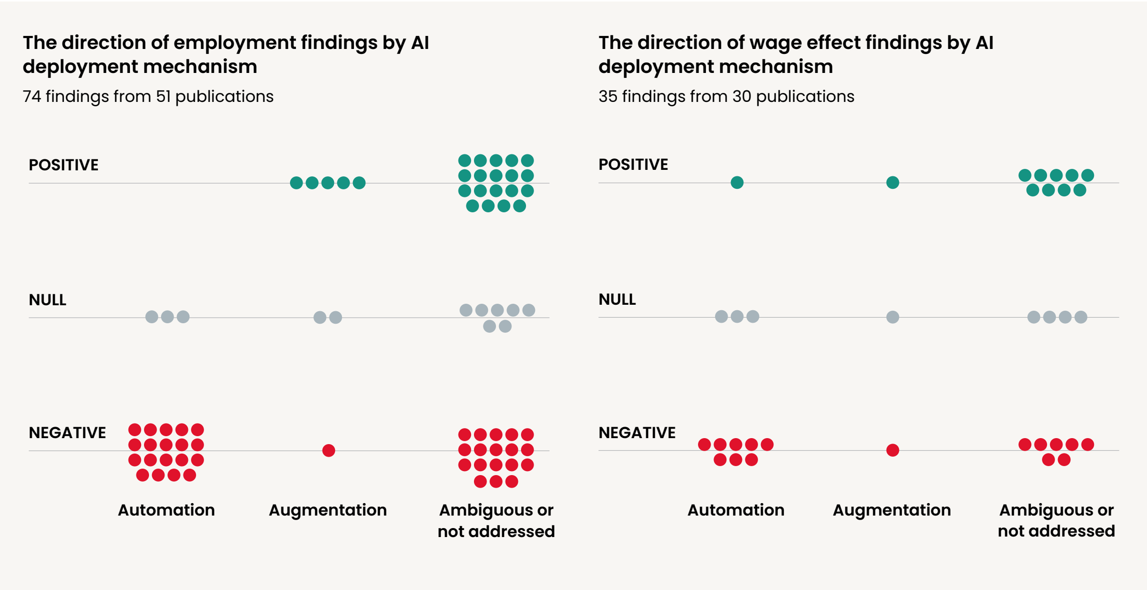


Note: each dot is one finding, not one publication. A publication contributing more than one distinct finding for the same outcome variable (e.g. due to employing two identification strategies) appears as multiple dots in the graph, which is why the finding count in a cluster can exceed the number of publications behind it. Null groups zero or no-change findings together with estimates reported as statistically insignificant. This category is heterogeneous in how much it tells us, since some nulls are precise and report economically meaningful effects, while others are less so. Our coding does not currently distinguish between the two, meaning the null column should be read as uniform evidence of an AI effect's absence. Evidence type reflects each publication's own study design, which we coded using the definitions laid out in section 3.1 'Potential versus actual evidence'. Some publications report directional findings for both employment and wages.

Automation and augmentation findings cluster in opposite directions. Of the 22 automation findings for employment, 19 are negative and not even 1 is positive (Figure 9). Augmentation reverses this: of eight findings, five are positive. When it comes to wage effects, the pattern for AI deployment mechanism carries over, albeit not quite as strongly. Automation estimates still skew negative, but less sharply than for job quantity. Augmentation rests on only three wage findings, too few to infer a direction.

These results are also in line with our expectations. Automation substitutes for labour on a specific task, while augmentation complements it, so estimates pointing in opposite directions are an unsurprising find. The takeaway from this set of results, however, is that the AI deployment mechanism that the literature engages with makes a difference to the direction of the results that follow. At this point in time, a higher proportion of the literature engages with automation rather than augmentation as a deployment mechanism, which means that we end up with higher numbers of negative estimates. We therefore have a second potential explanation for the negative lean we identified in Figure 8: it may also be a consequence of what the literature has chosen to study. A third explanation exists in what is published, which may be a biased sample of the distribution of produced estimates. Furthermore, since each finding is a dot, publications reporting several specifications will contribute more weight.

Figure 9. Direction of employment and wage effect findings by AI deployment mechanism

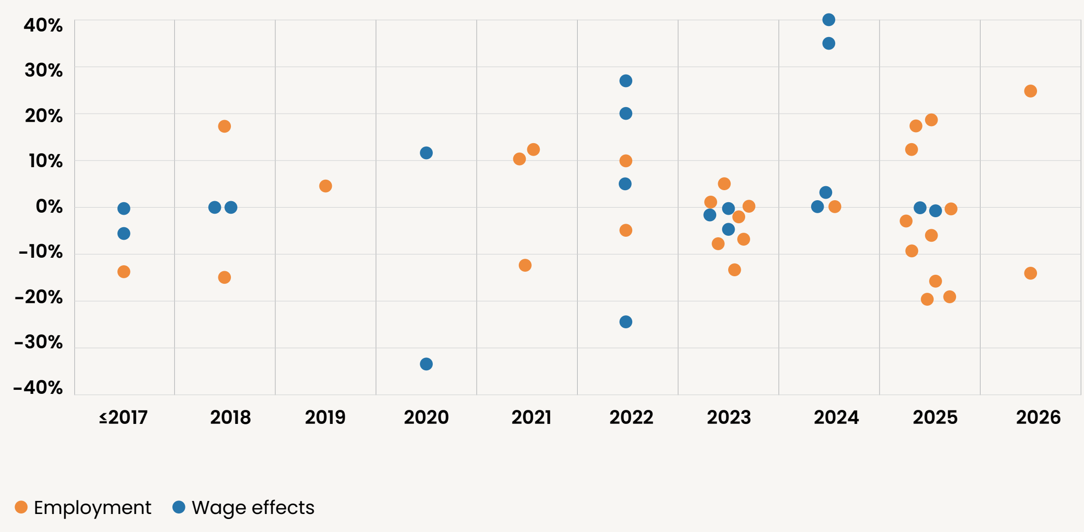


Note: each dot is one finding, not one publication. A publication contributing more than one distinct finding for the same outcome variable (e.g. due to employing two identification strategies) appears as multiple dots in the graph, which is why the finding count in a cluster can exceed the number of publications behind it. Null groups zero or no change findings together with estimates reported as statistically insignificant. This category is heterogeneous in how much it tells us, since some nulls are precise and report economically meaningful effects, while others are less so. Our coding does not currently distinguish between the two, meaning the null column should be read as uniform evidence of an AI effect's absence. Mechanism reflects how each publication frames the way AI changes work, coded from the publication's own description using the definitions laid out in section 2.3 'Three deployment mechanisms: automation, augmentation and commoditisation.' The "Ambiguous or not addressed" label is assigned to findings where either multiple deployment mechanisms are considered or no mechanisms are considered. Some publications report directional findings for both employment and wages.

The magnitude of estimated effects for both employment and wages varies widely. Discussions about AI's impact on labour markets often lead with a favourite estimate drawn from the literature, quoted at times as though it was a consensus finding. Figure 10 shows why no such single estimate exists. What the literature offers is a wide range of estimates.

Figure 10 plots all the estimates in the evidence database that are expressed as a percentage change. The horizontal axis records the year of the most recent data point on which each analysis relies. Putting these estimates on a single graph carries the caveats we set out above. The numbers are not directly comparable, because the publications that report them rely on different methodologies and study different technologies, geographies, deployment mechanisms, time frames, workers and settings. Importantly, the percentages have different denominators, where some describe employment in a single occupation at adopting firms, while others capture the relative position of one age group within an exposed quintile or even projected aggregate employment under simulation. We plot them together regardless, because the spread is illuminating. Anyone looking for a figure on AI's effect on employment can choose from estimates ranging from -20% to +25%. For wages, the range runs from -34% to +40%. Selecting a single headline number from a range this wide is a decision about which story to tell about AI's labour market impacts rather than an attempt to summarise what the evidence shows.

Figure 10. Percentage change estimates for AI's employment and wage effects by data vintage
28 job quantity and 19 wage estimates whose unit is a percentage change, from 29 distinct publications



Note: each point is one percentage change estimate, not one publication; only estimates whose unit is a percentage change are included here, so estimates expressed as percentage points, counts, elasticities, correlations and shares are excluded. Data vintage is the most recent year that a publication's underlying data covers, parsed from the dataset's own vintage field. For example, an estimate derived from a dataset that covers 2018–2022 is placed at 2022. Every vintage at or before 2017 is grouped into one ≤2017 category. Each later year gets its own category. Points within an x-axis category are spread out horizontally just far enough to avoid overlapping. Their exact horizontal position within the category they are assigned to carries no meaning of its own.

This section had two aims. The first was to map the existing literature on AI's labour market impacts and outline where the blind spots in our collective knowledge lie. The second was to answer a focused question, which is what existing research says about AI's effect on the quantity side of the labour market (employment and wages).

On the first aim, the interactive literature tree and the evidence database, read together, describe a literature whose coverage is concentrated in a narrow region. Most studies focus on generative AI, which translates into a focus on white-collar work. As such, the literature cannot speak about white-collar and blue-collar work in the same evidentiary voice: the former rests on a young literature measuring the current technology, the latter on an older robotics literature measuring a pre-generative one. Potential evidence is much more prominent than actual evidence, and the imbalance is sharpest for developing economies, where claims about AI's labour market impacts rely almost entirely on prediction. Most studies are recent, which means that even the most credible causal designs observe only the first years of technology diffusion. Informal work is almost invisible and commoditisation is the least studied of the three AI deployment mechanisms.

On the second aim, the answer is that there is no consensus in the literature on either the direction or the magnitude of AI's effects on employment and wages. More findings lean negative for both employment and wages in the actual and the potential evidence branches alike. We find that the direction of each finding depends heavily on the AI deployment mechanism a publication chooses to study, with automation estimates skewing negative and augmentation estimates positive. There are considerably more findings that engage with automation than with augmentation, providing a second explanation for the negative lean, alongside the fact that at this point in time findings capture the displacement rather than the reinstatement margin. The magnitude of the estimated effects varies widely: for employment, the estimates run from -20% to +25%, and for wages from -34% to +40%.

The overarching conclusion of this section is that the literature sheds light on a small and recent part of the global labour market, and it does not speak with one voice even there. Choosing a single figure from the literature says more about the motivation of the person selecting it than about fidelity to the evidence. Claims that fall outside the narrow region that the literature covers should be treated with utmost caution.

5. What the evidence shows, disputes and does not yet address

The previous section showed that coverage in the literature is uneven and estimates are variable. This section asks what, despite that, we can conclude. It accordingly sorts open questions by how far the evidence goes in settling policy-relevant debates, ranging from what has been established to what remains disputed or goes unstudied.

An important caveat is that every result in the literature is in relation to a particular vintage of AI, rather than AI in general. These include the GPT-4-era exposure scores of Eloundou et al. (2024), the copilot interfaces of Brynjolfsson, Li and Raymond (2025), and the pre-ChatGPT tools of earlier adoption studies. The technological frontier in 2026 already differs from what most published studies measure, with the 2026 Stanford AI Index reporting that capability has not plateaued, but risen further. Frontier models match or exceed human baselines on several advanced benchmarks and perform considerably better on real computer use tasks (Sajadieh et al., 2026). Benchmark performance is itself a potentially problematic proxy for capacity with economic relevance, given saturation and contamination concerns (Eriksson et al., 2025), suggesting caution in translating benchmark gains to greater labour market effects. A central uncertainty is whether the dominant deployment pattern remains LLM-based or shifts towards agentic and multimodal systems that plan and carry out multistep tasks with limited supervision, framed by Hadfield and Koh (2025) as an economy of AI agents. Every finding should be read with this caveat in mind.

5.1 What the evidence shows

The quantitative literature tends to agree on two findings, both of which relate to initial AI diffusion. The first is that aggregate effects remain muted and the second is that hiring at the bottom of the career ladder has slowed in some developed economies. Despite broad agreement on these findings, there is considerable disagreement on what explains these empirical observations and what can be drawn from either regarding AI's labour market impacts.

Aggregate effect is as of yet muted

At the level of economy-wide total employment, the Yale Budget Lab's work with the Current Population Survey shows no unusual change in the occupational mix, with the broader literature reporting limited evidence of economy-wide job loss or wage decline (Gimbel et al., 2025). After several years of rapid capability gains, aggregate employment and productivity series show no unambiguous break.

Some view this finding as evidence that AI's aggregate effect on the quantity of jobs is small and task-level gains are absorbed without much net displacement. An alternative view is that the true aggregate effect has yet to surface. In this reading of the evidence, the lag between adoption and measured gains while firms redesign workflows, integrate data and invest in complements is described by the productivity J-curve (Brynjolfsson, Rock and Syverson, 2021). This can explain why the large gains in productivity or quality in experiments surrounding customer support, writing, coding and professional services have yet to pass through to gross domestic product (GDP), hours, or headcount. Demand elasticity can also add to the difficulties of identifying effects at the aggregate level: where AI lowers the price of a service and demand is elastic, employment can rise even as tasks are automated (Bessen, 2019). Where demand is inelastic, the same productivity gain can reduce headcount. A null aggregate effect, therefore, can mask movements that are happening underneath and that offset each other. The same muted aggregate employment result is consistent with a benign steady state and with the quiet before an adjustment. While productivity experiments measure gains on tasks assisted by AI, the intensive margin of use remains small: between one and five per cent of all work hours are currently supported by generative AI, saving only 1.4% of total work hours (Bick, Blandin and Deming, 2026). The gap between experimental and macroeconomic predictions may therefore also reflect the extensive margin of use, rather than the failure of experimental results to generalise, pointing to the importance of diffusion depth as a metric to examine in addition to offsetting forces that may continue to materialise.

Hiring at the bottom of the career ladder has slowed

Employment among US workers aged 22–25 in the occupations most exposed to generative AI has fallen relative to older workers in the same occupations (Brynjolfsson, Chandar and Chen, 2025). Revised versions of the paper show a widening gap, from roughly 13% as of July 2025 to 19% in the most recent vintage, attributed to reduced hiring instead of elevated separations. Klein Teeselink (2025) offers a UK analogue using detailed labour market data that cover millions of firms and job postings between 2021 and 2025, showing reductions concentrated in junior and in technical and creative high-wage roles. In short, entry-level hiring is currently weak.

Several studies try to understand the mechanisms behind these trends. Wang, Wei and Wang (2026) find junior job postings adjusting through both hiring reallocation and task redesign in response to generative AI, suggesting firms may be reshaping the bottom rung rather than removing it. Hosseini and Lichtinger (2025) examine CV and job-posting data and find that junior employment falls relative to senior employment after firms advertise for generative AI integrator roles, mainly because junior hiring slows rather than because job separations rise.

Other studies caution against reading junior employment declines as a direct consequence of AI adoption. Humlum and Vestergaard (2025), linking large-scale adoption surveys to Danish administrative records, estimate precise nulls on earnings and recorded hours across 11 heavily exposed occupations and further rule out firms' chatbot adoption as a driver of early-career trends they are able to replicate. Cui et al. (2026), pooling 3 field experiments with nearly 5,000 software developers, find that access to an AI coding assistant raised completed tasks by around 26%, with less-experienced software developers gaining more, and more readily adopting the Copilot tool. While this suggests that junior employees may complement AI where firms supply the tools and workflows to utilise the technology, the effect on junior labour demand is ambiguous. Productivity experiments cannot adjudicate this question insofar as they hold firm hiring decisions fixed by design.

The reality is that we do not have sufficient evidence to conclude whether AI is behind weak entry-level hiring. On the one hand, the bottom rung of the career ladder is the fastest adjusting and most visible margin in the labour market, since hiring can be paused far more quickly than a workforce can be restructured. As such, patterns in junior employment are where we might expect to see early signals. On the other hand, whether we can attribute these signals to AI rather than alternative factors is highly contested. Graduates and young workers tend to fare worse when the economy slows, and the labour market of the mid-2020s has been weak in ways unrelated to AI, not least in that the extraordinary post-pandemic surge in vacancies has been unwinding, with many junior positions created during that period now being cut. Distinguishing AI's arrival from ordinary cyclical exposure of young workers is empirically challenging.

5.2 What the evidence disputes

As we have seen above, the difficulty of attribution is central to the separate disputes within the literature: separating the effect of AI from everything else moving through the labour market at the same time. Consider the COVID-19 pandemic's reallocation of remote work, the post-2022 monetary tightening and the long tail of pre-AI automation (Acemoglu and Restrepo, 2018). Coincident shifts in occupational demand, the presence of pre-trends and the absence of clean treatment timing make attribution difficult, even when strategies attempt to partial out the effects of AI.

Do firm-level results show AI creating jobs or redistributing them?

Studies that follow firms as they adopt AI find that strong task-level substitution can coexist with modest effects on total employment, since adopting firms often expand and reallocate labour contemporaneously. Hampole et al. (2025) show exactly this, and are explicit that whether AI displaces an exposed occupation depends on the relative strength of direct substitution, indirect effects and firm expansion, so that what occurs in aggregate remains theoretically ambiguous. A growing adopter, for example, may not create jobs on net if its growth is market share taken from competitors who shed workers, or if it draws labour from suppliers, contractors, or offshore providers that studies do not measure. Identification rests on data from a small number of providers, such as Revelio Labs (Hampole et al., 2025) or Cognism Inc and Burning Glass (Babina et al., 2024), and on closely related instruments, since both studies exploit firms' historical links to the universities supplying AI graduates. External validity to small firms, the public sector and economies outside the US is largely untested, along with outsourced work and broader displacement across industries that firm-level data may be unable to capture.

Does AI move wages and can we determine why?

The quantity literature has focused on whether jobs exist rather than on pay, and the direct wage evidence is thin and mixed. Acemoglu et al. (2022) find no detectable wage effects from AI exposure on US vacancy data that predate generative AI. Brynjolfsson, Chandar and Chen (2025) report that adjustment is more visible in employment than in compensation, including for early-career workers. Klein Teeselink (2025) finds UK effects concentrated in high-wage technical and creative roles and reads this as evidence that LLMs compress wage inequality.

Theory points firmly both ways at once: Fukui, Nakamura and Steinsson (2026) predict that task standardisation widens wage markdowns through commoditisation, while Althoff and Reichardt (2026) predict compression through simplification and comparative advantage. Their moderate scenario yields a 28% average wage rise and a 24% fall in the 90:10 wage ratio. However, once skill-dependent adoption costs are introduced, the average wage gain reduces to 0.3%, while the inequality compression is roughly halved. This emphasises the centrality of assumptions in simulation-based estimates. The sign has not yet confidently been adjudicated by the empirical literature.

Folded into this question is the commoditisation channel, in which AI raises the performance of less-skilled workers towards the best and thus compresses differences within an occupation without automating anyone. The closest empirical counterparts measure productivity and quality, not pay. Brynjolfsson, Li and Raymond (2025) and Dell'Acqua et al. (2026) find AI flattening within-occupation performance distributions, but on small populations of customer support agents and consultants, each within a single firm and with results for output and quality rather than realised pay or headcount. Whether simplification replicates beyond those samples, scales to entire occupations and dominates or is dominated by the automation channel in net wage terms remains open. Disentangling any wage movement from concurrent shocks that lead to wage adjustments is once again a binding empirical constraint.

Who is even using AI?

The literature has strong exposure measures and much weaker adoption measures. One difficulty is definitional. Adoption can mean that a worker tries a chatbot once, uses one weekly, is encouraged to use one by management, has a copilot in the development environment, or works in a firm that has embedded an application programming interface (API) into production systems. These treatments have different consequences: occasional individual use may raise personal productivity, while enterprise integration can reshape workflows, monitoring, customer interaction and hiring.

Measured adoption varies sharply by source, and each proxy has a blind spot. Bick, Blandin and Deming (2026) find very rapid population-level uptake of generative AI, spreading faster overall than personal computers or the internet at comparable stages, while worker surveys show use concentrated among younger, college-educated and higher-earning workers. The US Census measure cited by Chandar (2025) is closer to a tenth, partly because it asks about AI in the production of goods and services and may miss back office, experimental or informal use. Provider usage data give fine-grained task detail, but cover a single provider and user base. Adoption inferred from CVs and job postings connects usage to firm outcomes, but can miss outsourced development, internal pilots and off-platform experimentation. Business surveys often rely on one respondent to describe an entire firm. No single adoption number is therefore interpretable without the invaluable context of defined terms.

5.3 What the evidence does not yet address

This is the group of questions where evidence is thinnest, either because the topics have not yet received sufficient academic interest, or because the technologies at play are advancing at a pace that outstrips the evidence base. Our earlier review of the literature exposed many of the collective gaps in our knowledge.

What do we know about AI's labour market effects in developing countries?

As discussed earlier in this report, for developing countries, where most of the world's workers live, we have exposure modelling, but little actual evidence of AI's effects on employment, wages, or hours (del Rio-Chanona et al., 2025). There are reasons to believe that the external validity of existing results for these settings may be limited. First, the same occupation title does not imply the same task bundle across countries. Recent work on the digital divide (Gmyrek, Viollaz and Winkler, 2026) shows that conventional exposure measures overstate impact in developing economies by assuming that workers perform the same non-routine tasks as those near the frontier, so imported O*NET-style scores mislead unless they are adapted to the local context. Second, the digital divide might create "small buffer, big bottleneck" dynamics, where automatable, connected services roles are exposed to displacement even in poorer countries, while workers whose tasks could be augmented lack the broadband, electricity, devices, or firm-level complements to realise the upside. Trade may matter as much as domestic adoption, insofar as the automation of call centres and back office work that would traditionally have been outsourced to countries like India and the Philippines leads to fewer service imports, rather than layoffs within the outsourcing country's accounts. What is needed is country-level outcome data combining standardised metrics for adoption, connectivity, local-language access and sectoral or trade exposure.

How can we study informal work?

In line with what was discussed earlier, the relevant object of interest for informal work may be the person, stall, or customer network rather than a particular occupation within developing countries. AI may reach these workers through channels formal-employment studies never observe, and the benefits may not be evenly distributed. One such warning comes from Otis et al. (2026), who show that adoption of a WhatsApp AI assistant among Kenyan entrepreneurs only raised outcomes for high-performing entrepreneurs who implemented the advice; meanwhile, the average treatment effect was a precise null and the lowest performers did roughly 8% worse than the control group. Importantly, null average effects do not necessarily imply an absence of an effect. Underneath this overall finding, some groups may benefit, while others may be harmed. Fielding AI use modules, conducting phone panels and collecting platform data to cover household enterprises may offer fresh insights.

How strong will the reinstatement effects be?

The forces that offset displacement, reinstating labour through the creation of new tasks and occupations in which workers hold a comparative advantage – thereby raising labour demand indirectly by lowering costs – are underexplored. Partly this is a matter of timing, since displacement precedes reinstatement, but partly it is a matter of academic priority: the literature engages more with automation than augmentation, let alone commoditisation.

Yet, there is a deeper reason offsetting forces are hard to see: prices. When AI lowers the cost of producing a good or service, competition pushes prices down. Lower prices raise the real incomes of those who purchase the product, such that the purchasing power generates demand (and hence labour demand) across the whole economy. This includes sectors that AI has never affected. The offset to displacement in an exposed occupation may thus surface as new jobs in an entirely unrelated occupation, reached through the price level rather than a channel typically studied. Paying greater attention to the channel of price adjustment could help witness reinstatement effects better.

What happens to workers displaced by AI?

If AI is disruptive in displacing workers, there are some nearly unanswered questions: how quickly do they find new work, are their new roles comparable in pay and stability to their previous ones, and how persistent will any earnings losses prove to be? Although a substantial literature on adjustment after job losses from trade, automation and plant closure exists, it concerns other shocks that may differ in important ways from those brought about by AI. Distinguishing between outright displacement, reduced hiring and career stagnation through linked longitudinal records following displaced workers across employers would yield valuable information on an employee's life cycle.

What topics are ahead of the evidence frontier?

When AI acts in the physical world through autonomous vehicles, warehouses, farms and clinics, the labour market effects are largely unmeasured. The canonical reference of Acemoglu and Restrepo (2020) on industrial robotics predates the current wave of technology and studies dedicated machines instead of AI-enabled ones. This gap is also pointed out by Rahwan and Muthukrishna (2025), along with the World Development Report (2026). This gap will be challenging to close retrospectively, since it requires the collection of relevant baseline data.

While every result in this report is affected by how fast AI's effects materialise, the pace at which capability converts into adoption, adoption into workflow change, and workflow change into realised employment and wage effects determines whether the displacement margin is a transitional shock that reinstatement and falling prices will offset, or a persistent one. Remaining vigilant as the adjustment unfolds and investing in real-time measurement systems that consistently measure progress along key metrics will be hugely relevant for policy.

6. Conclusion

In this report we set out to separate what AI might do to work from what it has demonstrably done and evaluate how the evidence base compares to common claims. The predictions that dominate public debate describe a technology already widely adopted and operating at full scope. The evidence describes something narrower: successive iterations of a technology observed early, in limited slices of the economy, with effects that are smaller and more unevenly distributed than most headlines suggest. Though studies have made substantial progress, especially on the theoretical front, AI's actual net effects in the labour market remain highly uncertain.

Evidence gaps are not evenly distributed and point towards a necessary research and policy agenda. Where the evidence is largely absent – such as for developing economies, informal work, deployment mechanisms apart from automation, and worker outcomes – researchers, funders, and policymakers must work alongside industry to create new datasets and methodologies to fill the collective gaps in our knowledge. Where evidence is contested, the binding constraint is nearly always attribution, pointing towards a need for research designs that can separate AI from the many other forces affecting patterns in time series. Where the technology is running ahead of any possible evidence, as is the case with embodied and agentic systems, there are valuable insights to be drawn in building the measurement infrastructure that will allow these effects to be documented as they arrive, rather than reconstructed later from data that were not designed to capture them.

The deeper lesson of this report is one of humility about understanding a technology and its impacts in real time. Every empirical estimate in this report is a measurement of a past version of AI, taken during the early phases of diffusion, through instruments lagging behind an evolving world of work. That is not a reason to ignore the evidence, but rather to hold it lightly, and resist the pull of any single figure, regardless of how decisively it is declared. An honest account of AI's effect on work resists the temptation to issue a verdict at this moment in time, opting instead to chart the questions and research agenda that will help us understand what our uncertain future holds.

References

Acemoglu, D. (2025) 'The simple macroeconomics of AI', *Economic Policy*, 40(121), pp. 13–58. Available at: https://doi.org/10.1093/epolic/eiae042.

Acemoglu, D. and Autor, D. (2011) 'Skills, tasks and technologies: implications for employment and earnings', in O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics*. Amsterdam: Elsevier, pp. 1043–1171. Available at: https://doi.org/10.1016/S0169-7218(11)02410-5.

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022) 'Artificial intelligence and jobs: evidence from online vacancies', *Journal of Labor Economics*, 40(S1), pp. S293–S340. Available at: https://doi.org/10.1086/718327.

Acemoglu, D. and Restrepo, P. (2018) 'The race between man and machine: implications of technology for growth, factor shares, and employment', *American Economic Review*, 108(6), pp. 1488–1542. Available at: https://doi.org/10.1257/aer.20160696.

Acemoglu, D. and Restrepo, P. (2020) 'Robots and jobs: evidence from US labor markets', *Journal of Political Economy*, 128(6), pp. 2188–2244. Available at: https://doi.org/10.1086/705716.

Althoff, L. and Reichardt, H. (2026) *Task-specific technical change and comparative advantage*. Cambridge, MA: National Bureau of Economic Research. Available at: https://doi.org/10.3386/w35353.

Arntz, M., Gregory, T. and Zierahn, U. (2016) *The risk of automation for jobs in OECD countries: a comparative analysis*. OECD Social, Employment and Migration Working Papers No. 189. Paris: Organisation for Economic Co-operation and Development. Available at: https://doi.org/10.1787/5jlz9h56dvq7-en.

Autor, D.H. (2015) 'Why are there still so many jobs? The history and future of workplace automation', *Journal of Economic Perspectives*, 29(3), pp. 3–30. Available at: https://doi.org/10.1257/jep.29.3.3.

Autor, D.H., Levy, F. and Murnane, R.J. (2003) 'The skill content of recent technological change: an empirical exploration', *The Quarterly Journal of Economics*, 118(4), pp. 1279–1333. Available at: https://doi.org/10.1162/003355303322552801.

Babina, T., Fedyk, A., He, A. and Hodson, J. (2024) 'Artificial intelligence, firm growth, and product innovation', *Journal of Financial Economics*, 151, no. 103745. Available at: https://doi.org/10.1016/j.jfineco.2023.103745.

Bessen, J.E. (2019) 'AI and Jobs', in A. Agrawal, J. Gans and A. Goldfarb (eds) *The Economics of Artificial Intelligence: An Agenda*. Chicago, IL: University of Chicago Press, pp. 291–308. Available at: https://doi.org/10.7208/chicago/9780226613475.003.0010.

Bick, A., Blandin, A. and Deming, D.J. (2026) 'The rapid adoption of generative AI', *Management Science* [Preprint]. Available at: https://doi.org/10.1287/mnsc.2025.02523.

Bright, J., Enock, F., Esnaashari, S., Francis, J., Hashem, Y. and Morgan, D. (2025) 'Generative AI is already widespread in the public sector: evidence from a survey of UK public sector professionals', *Digital Government: Research and Practice*, 6(1), pp. 1–13. Available at: https://doi.org/10.1145/3700140.

Brynjolfsson, E., Chandar, B. and Chen, R. (2025) *Canaries in the Coal Mine? Six Facts About the Recent Employment Effects of Artificial Intelligence*. Stanford, CA: Stanford Digital Economy Lab. Available at: https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/ (Accessed: 19 August 2026).

Brynjolfsson, E., Li, D. and Raymond, L. (2025) 'Generative AI at work', *The Quarterly Journal of Economics*, 140(2), pp. 889–942. Available at: https://doi.org/10.1093/qje/qjae044.

Brynjolfsson, E., Rock, D. and Syverson, C. (2021) 'The productivity J-curve: how intangibles complement general purpose technologies', *American Economic Journal: Macroeconomics*, 13(1), pp. 333–372. Available at: https://doi.org/10.1257/mac.20180386.

Caselli, M., Fracasso, A., Scicchitano, S., Traverso, S. and Tundis, E. (2025) 'What workers and robots do: an activity-based analysis of the impact of robotization on changes in local employment', *Research Policy*, 54(1), no. 105135. Available at: https://doi.org/10.1016/j.respol.2024.105135.

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A.J., Pizzinelli, C., Rockall, E. and Tavares, M.M. (2024) *Gen-AI: Artificial Intelligence and the Future of Work*. SDN/2024/001. Washington, DC: International Monetary Fund. Available at: https://doi.org/10.5089/9798400262548.006.

Cerutti, E., Garcia Pascual, A., Kido, Y., Li, L., Melina, G., Tavares, M.M. and Wingender, P. (2025) *The Global Impact of AI: Mind the Gap*. WP/2025/076. Washington, DC: International Monetary Fund. Available at: https://doi.org/10.5089/9798229008570.001.

Chandar, B. (2025) *AI and Labor Markets: What We Know and Don't Know*. Stanford, CA: Stanford Digital Economy Lab. Available at: https://digitaleconomy.stanford.edu/news/ai-and-labor-markets-what-we-know-and-dont-know/ (Accessed: 23 August 2026).

Chatterji, A., Holtz, D., Rakholia, N., Tambe, P. and Weeratunga, G. (2026) 'How organizations use AI: evidence from ChatGPT'. arXiv. Available at: https://doi.org/10.48550/arXiv.2608.12236.

Ciaschi, M., Falcone, G., Garganta, S., Gasparini, L., Bertín, O. and Ramírez-Leira, L. (2025) *The Potential Distributive Impact of AI-driven Labor Changes in Latin America*. IDB-WP-14253. Washington, DC: Inter-American Development Bank. Available at: https://doi.org/10.18235/0013677.

Cui, Z.K., Demirer, M., Jaffe, S., Musolff, L., Peng, S. and Salz, T. (2026) 'The effects of generative AI on high-skilled work: evidence from three field experiments with software developers', *Management Science*, 72(5). Available at: https://doi.org/10.1287/mnsc.2025.00535.

Del Rio-Chanona, R.M., Ernst, E., Merola, R., Samaan, D. and Teutloff, O. (2025) 'AI and jobs. A review of theory, estimates, and evidence'. arXiv. Available at: https://doi.org/10.48550/arXiv.2509.15265.

Dell'Acqua, F., McFowland III, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K.R. (2026) 'Navigating the jagged technological frontier: field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality', *Organization Science*, 37(2), pp. 403–423. Available at: https://doi.org/10.1287/orsc.2025.21838.

Egana-delSol, P. and Bravo-Ortega, C. (2025) *Artificial Intelligence and Labor Market Transformations in Latin America*. DP 17746. Bonn: IZA Institute of Labor Economics. Available at: https://doi.org/10.2139/ssrn.5167711.

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2024) 'GPTs are GPTs: labor market impact potential of LLMs', *Science*, 384(6702), pp. 1306–1308. Available at: https://doi.org/10.1126/science.adj0998.

Eriksson, M., Purificato, E., Noroozian, A., Vinagre, J., Chaslot, G., Gomez, E. and Fernandez-Llorca, D. (2025) 'Can we trust AI benchmarks? An interdisciplinary review of current issues in AI evaluation'. arXiv. Available at: https://arxiv.org/abs/2502.06559.

Frank, M.R., Sabet, A.J., Simon, L., Bana, S.H. and Yu, R. (2026) 'AI-exposed jobs deteriorated before ChatGPT'. arXiv. Available at: https://doi.org/10.48550/arXiv.2601.02554.

Frey, C.B. and Osborne, M.A. (2017) 'The future of employment: how susceptible are jobs to computerisation?', *Technological Forecasting and Social Change*, 114, pp. 254–80. Available at: https://doi.org/10.1016/j.techfore.2016.08.019.

Fukui, M., Nakamura, E. and Steinsson, J. (2026) 'The commoditization of labor'. Authors' website. Available at: https://jonsteinsson.com/papers/Commoditization.pdf.

Gimbel, M., Kinder, M., Kendall, J. and Lee, M. (2025) *Evaluating the Impact of AI on the Labor Market: Current State of Affairs*. New Haven, CT: The Budget Lab at Yale. Available at: https://budgetlab.yale.edu/research/evaluating-impact-ai-labor-market-current-state-affairs (Accessed: 19 August 2026).

Gmyrek, P., Viollaz, M. and Winkler, H. (2026) *Disruption Without Dividend? - How the Digital Divide and Task Differences Split GenAI's Global Impact*. Geneva: International Labour Organization. Available at: https://www.ilo.org/publications/disruption-without-dividend-how-digital-divide-and-task-differences-split (Accessed: 30 April 2026).

Guarascio, D., Piccirillo, A. and Reljic, J. (2025) 'Robots vs workers: evidence from a meta-analysis', *Journal of Economic Surveys*, 39(5), pp. 2254–2271. Available at: https://doi.org/10.1111/joes.12699.

Hadfield, G.K. and Koh, A. (2025) 'An economy of AI agents'. arXiv. Available at: https://arxiv.org/abs/2509.01063.

Hampole, M., Papanikolaou, D., Schmidt, L.D.W. and Seegmiller, B. (2025) *Artificial Intelligence and the Labor Market*. Cambridge, MA: National Bureau of Economic Research. Available at: https://doi.org/10.3386/w33509.

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., Mueller, J., Hong, J., Ritchie, S., Belonax, T., Troy, K.K., Amodei, D., Kaplan, J., Clark, J. and Ganguli, D. (2025) *Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations*. Report series. Anthropic. Available at: https://doi.org/10.48550/arXiv.2503.04761.

Hosseini Maasoum, S.M. and Lichtinger, G. (2025) 'Generative AI as seniority-biased technological change: evidence from U.S. résumé and job posting data'. SSRN. Available at: https://doi.org/10.2139/ssrn.5425555.

Hui, X., Reshef, O. and Zhou, L. (2024) 'The short-term effects of generative artificial intelligence on employment: evidence from an online labor market', *Organization Science*, 35(6), pp. 1977–1989. Available at: https://doi.org/10.1287/orsc.2023.18441.

Humlum, A. and Vestergaard, E. (2025) *Still Waters, Rapid Currents: Early Labor Market Transformation Under Generative AI*. Cambridge, MA: National Bureau of Economic Research. Available at: https://doi.org/10.3386/w33777.

Iscenko, Z., Strand, S., Chen, Y., Aimard, G., Codreanu, M., Sampathkumar, V., Imas, A., Jacobs, J., Muiruri, E., Mateos-Garcia, J., Ng, J.J., Javed, S., Martin, J., Ajmeri, O., Calin, D., Kim, A., Millet, F.C. and Manyika, J. (2026) 'Google's AI & Economy ATLAS v1.0: mapping Gemini usage in the economy'. arXiv. Available at: https://doi.org/10.48550/arXiv.2608.00038.

Klein Teeselink, B. (2025) 'Generative AI and labor market outcomes: evidence from the United Kingdom'. SSRN. Available at: https://doi.org/10.2139/ssrn.5516798.

Klein Teeselink, B. and Carey, D. (2026) 'AI, automation, and expertise'. SSRN. Available at: https://doi.org/10.2139/ssrn.6134506.

Make UK, The Manufacturers' Organisation (2025) *Making It Smarter: Global Lessons for Accelerating Automation and Digital Adoption in UK Manufacturing*. London: Make UK, The Manufacturers' Organisation. Available at: https://www.makeuk.org/insights/reports/making-it-smarter-global-lessons-accelerating-automation-digital-adoption-uk (Accessed: 19 August 2026).

Manufacturing Technology Centre (2026) *Global Robotics Clusters Study: A Blueprint to Accelerate the UK Adoption of Robotics and Autonomous Systems*. Whitepaper. Coventry: Manufacturing Technology Centre. Available at: https://www.the-mtc.org/insights/global-robotics-clusters-study-blueprint-accelerate-uk-adoption-robotics-autonomous (Accessed: 19 August 2026).

Massenkoff, M. and McCrory, P. (2026) *Labor Market Impacts of AI: A New Measure and Early Evidence*. Research note. Anthropic. Available at: https://www.anthropic.com/research/labor-market-impacts (Accessed: 19 August 2026).

Noy, S. and Zhang, W. (2023) 'Experimental evidence on the productivity effects of generative artificial intelligence', *Science*, 381(6654), pp. 187–192. Available at: https://doi.org/10.1126/science.adh2586.

Organisation for Economic Co-operation and Development (2024) *Explanatory Memorandum on the Updated OECD Definition of an AI System*. OECD Artificial Intelligence Papers No. 8. Paris: Organisation for Economic Co-operation and Development. Available at: https://doi.org/10.1787/623da898-en.

Otis, N.G., Clarke, R., Delecourt, S., Holtz, D. and Koning, R. (2026) 'The uneven impact of generative AI on entrepreneurial performance: Evidence from a field experiment in Kenya', *Management Science* [Preprint]. Available at: https://doi.org/10.1287/mnsc.2024.06909.

Papagiannaki, E. and Geisler, J. (2026) *Recent Methodologies on AI and Labour: A Desk Review*. London: Institute for the Future of Work. Available at: https://doi.org/10.5281/zenodo.18801200.

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023) 'The impact of AI on developer productivity: evidence from GitHub Copilot'. arXiv. Available at: https://doi.org/10.48550/arXiv.2302.06590.

Pizzinelli, C., Panton, A., Tavares, M.M., Cazzaniga, M. and Li, L. (2023) *Labor Market Exposure to AI: Cross-country Differences and Distributional Implications*. WP/23/216. Washington, DC: International Monetary Fund. Available at: https://doi.org/10.5089/9798400254802.001.

Rahwan, I. and Muthukrishna, M. (2025) 'AI for productivity and empowerment in agriculture, health, education, and transport', in *The Next Great Divergence*: *Why AI May Widen Inequality Between Countries*. New York, NY: United Nations Development Programme, pp. 83–119. Available at: https://pure.mpg.de/view/item_3687807.

Sajadieh, S., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Santarlasci, L., Pava, J., Maslej, N., Altman, R., Brynjolfsson, E., Brodley, C., Clark, J., Dignum, V., Kumar, V., Landay, J., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Tabassi, E., Wald, R., Walsh, T. and Weld, D. (2026) 'Artificial intelligence index report 2026'. arXiv. Available at: https://arxiv.org/abs/2606.15708.

Teutloff, O., Einsiedler, J., Kässi, O., Braesemann, F., Mishkin, P. and del Rio-Chanona, R.M. (2025) 'Winners and losers of generative AI: early evidence of shifts in freelancer demand', *Journal of Economic Behavior & Organization*, 235, no. 106845. Available at: https://doi.org/10.1016/j.jebo.2024.106845.

United Nations Development Programme (2025) *The Next Great Divergence*: *Why AI May Deepen Inequality Between Countries*. New York, NY: United Nations Development Programme. Available at: https://www.undp.org/asia-pacific/next-great-divergence (Accessed: 19 August 2026).

Wang, F., Wei, Z. and Wang, Y. (2026) 'Generative AI and the reorganization of labor demand'. arXiv. Available at: https://arxiv.org/abs/2605.23159.

Webb, M. (2019) 'The impact of artificial intelligence on the labor market'. SSRN. Available at: https://doi.org/10.2139/ssrn.3482150.

World Bank (2026) *World Development Report 2026*: *Artificial Intelligence for Development*. Washington, DC: World Bank. Available at: https://www.worldbank.org/en/publication/wdr2026.

Zeira, J. (1998) 'Workers, machines, and economic growth', *The Quarterly Journal of Economics*, 113(4), pp. 1091–1117. Available at: https://doi.org/10.1162/003355398555847.

Appendices

Appendix A. Constructing the interactive literature tree

How we identified the publications. The tree is the product of a thorough thematic review. An initial reading list for the report was seeded by searching Google and Google Scholar for academic publications across three themes (the quantitative labour market impacts of AI, qualitative impacts and policy implications), and by targeted searches of publications from international organisations (including the International Labour Organization (ILO) and the OECD), think tanks (including Brookings), research labs (including Anthropic and OpenAI) and industry (including LinkedIn). We then carried out thorough searches to add academic and grey literature on the quantitative impacts of AI on labour markets to a shared Zotero library, using Google Scholar and SSRN keyword searches and snowballing from key works. We organised the resulting set as a citation lineage tree. We also searched specifically for evidence from developing and developed economies, because the empirical base otherwise defaults to developed country settings. Later additions to the set included robotics publications identified by consulting with experts in this field. We stopped adding publications to the working list on 10 July 2026. A paper published after that date is unlikely to appear in the tree.

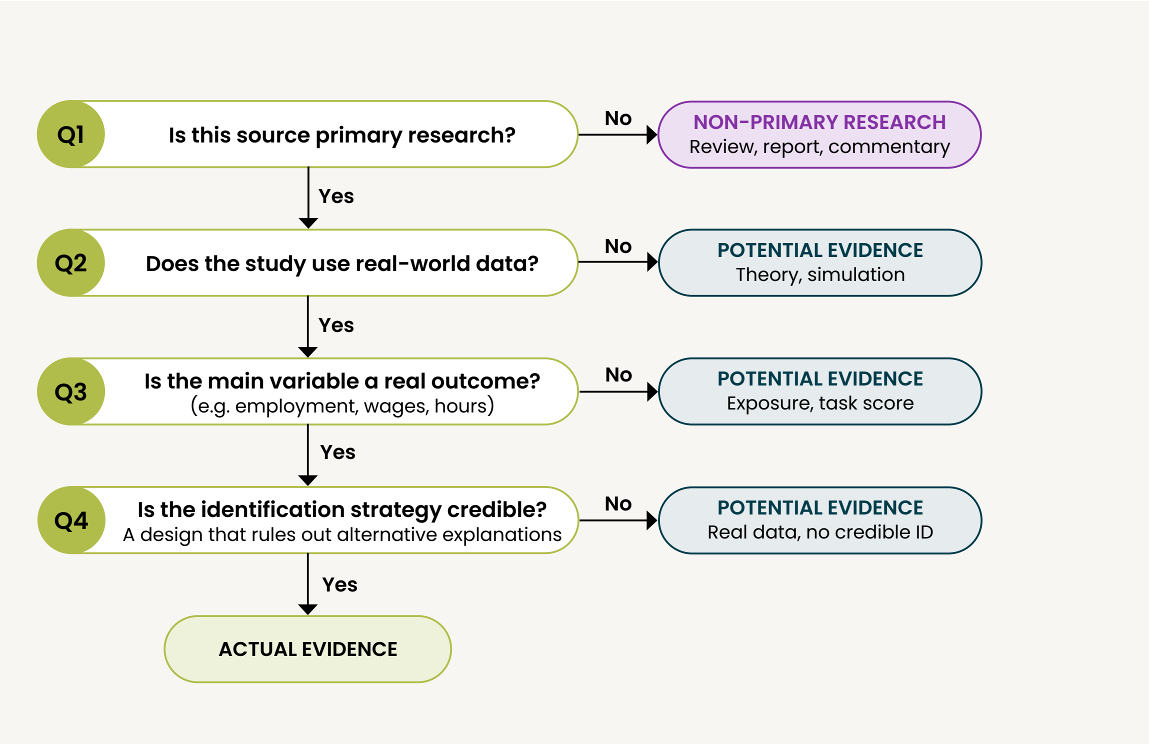
What the set contains. We bounded the search by topic rather than academic discipline, so we aimed to include publications about AI's labour market impacts from any field. The resulting collection is concentrated in economics, and also includes management, some sociology, computer science and human–computer interaction, and reports from international organisations, think tanks and industry. The tree itself includes theoretical papers, empirical studies, literature reviews and policy reports.

How we classify the publications. The interactive literature tree maps the publications we identified on AI's labour market impacts along the following three dimensions:

- **AI classification.** We distinguish between the form of AI that each publication examines. We follow the classification we outlined in Section 2.2, and distinguish between publications that look at AI systems designed to operate in physical environments, as well as decision-support AI, generative AI and perceptive AI. The AI classification distinction appears in the vertical axis of our interactive literature tree. We add two more categories to the vertical axis for publications that we cannot classify under an AI classification: (1) a foundational lane of publications that outline theory and conceptual mechanisms, and (2) a cross-cutting/unspecified lane for publications that either span multiple technology classes (e.g. literature reviews) or do not specify a technology class.
- **Year when the evidence was collected.** We also distinguish between the year when the data that each publication relies on were collected. We make this distinction because publications using data from 2017–2020, for example, describe a very different technology than publications using data from 2024–2026. The period when the underlying data were collected appears in the horizontal axis of our interactive literature tree. A paper published in 2025 on data from 2018–2020 appears on the left of a paper that reports the findings of a 2023 field experiment, because it describes an older technology (Papagiannaki and Geisler, 2026). Reading the interactive literature tree left to right is therefore reading forward through the capability frontier, from the pre-LLM task automation literature, through the exposure measures that followed the release of ChatGPT in late 2022, to the realised firm- and worker-level evidence of 2024–2026.

- **Evidence classification.** We also distinguish between the type of evidence that each publication produces. We do not classify non-primary sources (reviews, institutional reports, commentary) that synthesise or comment on primary research rather than generate new evidence. For primary research, we follow the classification we outlined in Section 3.1 and distinguish between publications that estimate the potential impact of AI on labour markets and publications that measure the actual impact of AI on labour markets. We classify theory and simulation papers as potential evidence: they estimate where AI has the structural capacity to alter work but cannot establish that any displacement or wage change has occurred. We classify empirical studies whose main outcome is a realised labour market indicator (employment, wages, hours), identified with credible causal identification strategies (e.g. difference-in-differences, instrumental variables, randomised control trials, regression discontinuity, etc.) as actual evidence. Studies using real data without a credible identification strategy (job posting indices, résumé records, administrative panels without a quasi-experimental design) document actual labour market activity but do not support causal inference, so we treat them as potential evidence. We flag this as the most consequential classification decision we make, since it groups forward-looking estimates of structural capacity with realised measurements that lack a design that supports causal interpretation. The colour of each bubble in the interactive literature tree carries the distinction between actual and potential evidence: lime green bubbles represent actual evidence and teal bubbles represent potential evidence. Purple bubbles represent non-primary research. Figure A1 summarises the way in which we classify the evidence.

Figure A1. Evidence classification



How to read the tree. In the interactive literature tree, links run from a study to the earlier work upon which it builds. The links trace the lineage of each strand of the literature and how the evidence has accumulated over time. They also show which branches of the literature hit a dead end and which expanded over time.

We encourage interested readers to explore the interactive literature tree. Clicking on each study reveals its academic lineage. We also label each publication according to the geography it covers and the AI deployment mechanism at play.

Appendix B. Constructing the evidence database

How we selected the publications. We need to rely on a narrower base of the literature compared to the interactive literature tree in order to answer the question of how much AI impacts employment and wages. We apply 3 successive filters to the 138 publications in the interactive literature tree to establish that base. Each filter removes a specific, well-defined category, leaving only primary research publications that speak directly to the quantity question we posed:

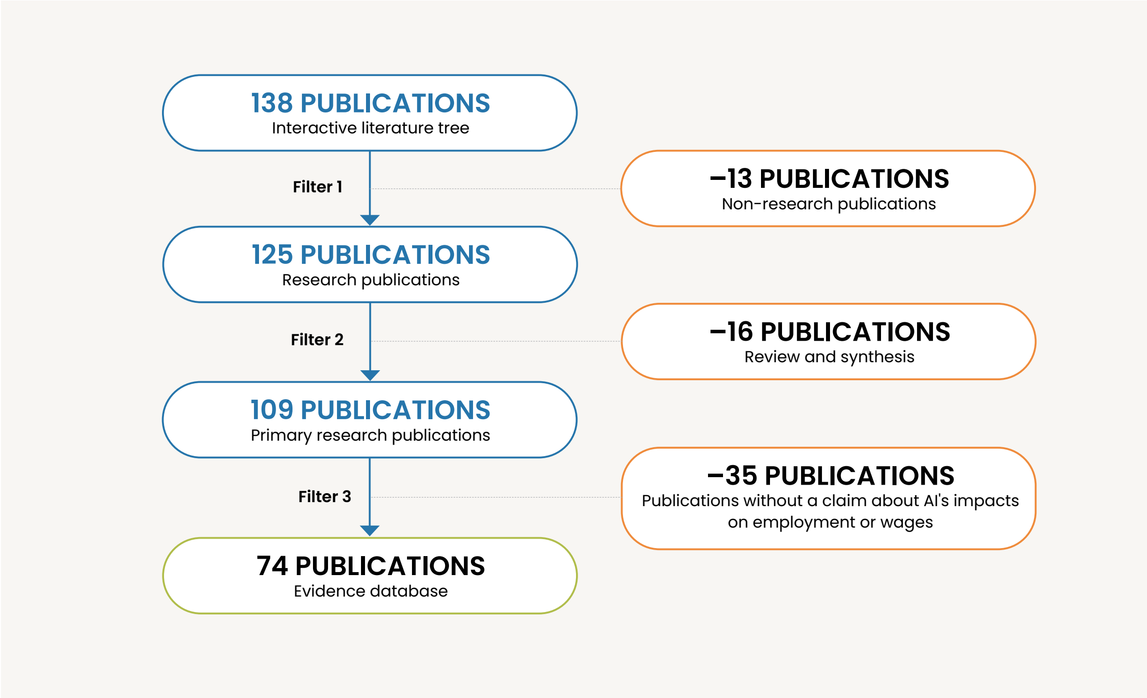
1. **Non-research publications.** We removed material published outside the outlets in which original research conventionally appears. In particular, we excluded webpages, blog posts, news articles and magazine articles. We kept reports, journal articles, preprints and books or book chapters. This removed 13 items, leaving 125 publications of working paper grade or better, though not yet confirmed as primary research.

2. **Reviews and syntheses.** We removed pieces that summarise or comment on others' work rather than presenting new evidence or analysis. In particular, we excluded government evidence reviews that synthesise other institutions' estimates, comparative policy reports built around "lessons from" case studies and commentary pieces critiquing others' findings rather than producing their own. Because this corpus is heavy on exposure indices, one classification rule is worth stating: we classify a publication that constructs its own index as primary research, while we classify a publication that only compares or critiques other people's indices without building one as a synthesis. This filter removes 16 items, leaving 109 publications that are primary research in the strict sense: each runs its own empirical analysis or builds its own model, using data that the authors collected or constructed themselves.

3. **Excluding publications without a claim about AI's impacts on employment or wages or an AI focus.** Stage 3 removed primary AI and labour market research that does not give a direction or a magnitude for job quantities or wages, and general automation or trade theory precursors with no AI content. We kept a publication if it makes a directional or magnitude claim by any route (e.g. measurements from observed data, forecasts, projections from exposure measures, insights derived from a model) because the question at this stage is whether the publication takes a position on the quantity question at all, not how strong its evidence for that position is. We excluded publications whose only outcome is a productivity or task completion measure, a capability benchmark against human performance, an adoption or usage pattern, or a gross exposure share with no net direction attached, along with publications where a job or wage figure is merely a calibration input to an unrelated model (e.g. a labour share assumption feeding a GDP projection) rather than a reported finding. We also excluded general automation or trade theory studies that precede AI-specific technology and make no claim about AI itself. This removed 35 items, leaving 74 publications that state a direction or a magnitude for the effect of AI on employment or wages.

Figure B1 provides a visual summary of the process outlined above.

Figure B1. The process to identify the literature that speaks to the quantitative effects of AI on employment and wages



How we code the publications. We code each publication based on the type of AI study, the AI deployment mechanism at play, and the sector, workers and economies covered by the data. Further fields separate potential evidence from actual evidence and delineate the type of study (observational, exposure index, simulation, theory, meta-analysis) and the data used. Other fields capture the finding of each publication and its direction, the statistic behind it, the unit of measurement and the population to which it applies. The database also flags whether an effect varies by gender; age, career, or seniority; education and skill level; formality; and geography.

How to use the database. The evidence database is built as an interactive tool rather than a static table. Each of the 74 publications is coded along the dimensions outlined above, and readers can filter, search and sort the full set by any of these fields. Clicking on a publication expands its row to show the full underlying detail, including its citation, research question and methodology. The database also offers a summary view, showing the distribution of publications across each coded dimension, and a matrix view, allowing any two dimensions to be cross-tabulated against each other.

GLOBAL FORUM FOR AI AND THE SOCIAL SCIENCES

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE