Institute *for the* Future of Work

Toolkit

Recent Methodologies on AI and Labour - a Desk Review

Eleni Papagiannaki and Joana Geisler

October 2026



Overview

This desk review toolkit brings together recent methodologies used to analyse AI's effects on employment, wages, and productivity.

It is structured into two tables:

Table 1 covers original works from 2003 to 2020

Table 2 focuses on academic, policy, and think tank works from 2021 to 2026. After 2023 especially, the focus is significantly on LLMs.

Each table categorises methodologies, high-level findings, and method-specific critiques to provide a comprehensive overview of the field.

Key points to note from most recent methodologies (Table 2)

Shift from Occupation-Level to Task-Level Analysis:

Recent methodologies increasingly focus on tasks rather than entire occupations, recognising the heterogeneity within jobs. For example, Felten et al. (2023) and Eloundou et al. (2023) use AI exposure indices to measure task-specific impacts.

Integration of Advanced Technologies:

Methods now incorporate Natural Language Processing (NLP) (e.g., BERT, LSTM) and LLMs (e.g., GPT-4) to analyse job descriptions and predict automation risks (Xu et al., 2025; Hampole et al., 2025).

Scenario Planning and Policy Focus:

Think tanks like TBI and IPPR, and organisations like IMF emphasise scenario-based modelling to inform policy, highlighting the need for reskilling and labour market reforms (TBI, 2024; IPPR, 2024, Korinek, 2023). They assume different 'initial conditions' in adoption to estimate different conclusions in employment, wages, productivity ect.

Studies consistently find that AI's impact varies by skill level, with low-skilled workers facing higher displacement risks and high-skilled workers benefiting from augmentation (Brynjolfsson et al., 2023; Chen et al., 2024).

Many methodologies struggle with static assumptions, lack of causal evidence, and overreliance on theoretical models. For instance, Acemoglu & Restrepo (2022) assume fixed comparative advantage, while Webb (2020) ignores adaptation.

As of year 2025, conclusions of some of the pre-2020 research is already redundant in terms of what professions will remain or not. The pre-2020 conclusion that creative and intellectual occupations will remain in high demand, has already been disproven.

Table 1 - Original Works 2003 - 2020

Methodology	High-level findings	Critique
Task-Based Approach (Autor et al., 2003) • Computer capital as i) substitute and ii) complement to workers. • Mapping of work tasks within occupations: Routine vs Non-routine. • Historical data analysis.	• Computerisation reduces labour input for routine tasks and increases it for non-routine cognitive tasks. • Explains 60% of the demand shift favouring college labour (1970-1998).	• Static view: Assumes tasks remain fixed over time. • Ignores task reconfiguration within jobs. • Binary classification (routine vs. nonroutine).
Occupation-Based Approach (Frey & Osborne, 2013) • Occupation level analysis: Routine vs Non-routine. • Expert assessments or automation scores: "computerisable" or "non-computerisable". • Gaussian Process Classifier (GPC) to estimate the probability of computerisation.	• 47% of US employment is in occupations at high risk of computerisation. • Jobs in transportation, logistics, and office support are most susceptible. • Occupations requiring creative intelligence, social intelligence, and fine motor skills are less susceptible to computerisation.	• Lacks granularity: Occupations contain heterogeneous tasks. • Static view: Ignores worker adaptation. • Technologically deterministic: automation exposure equates to automation adoption. • Occupations requiring creative intelligence as of year 2025 are more exposed due to LLMs.
Task- and Occupation-Based Approach with Social Dimensions (OECD) (Arntz et al., 2016) • Task-Based Automation Framework within occupations. • Firms and workers adaptability. • Tech adoption faces economic, legal, and societal barriers, slowing displacement. • Emergence of new jobs emerge from tech demand and productivity gains.	• Only 9% of jobs are at high risk of automation. • Workers adapt by shifting to non-automatable tasks. • Higher automation potential in countries with lower average skill levels. • Higher-educated workers face lower risk due to task complexity. • Countries with rigid task specialisation (e.g., Germany) show higher risk. • Nations with existing robotics/ IT infrastructure (e.g., South Korea) have fewer remaining automatable tasks.	• Focuses on current technology, missing future advancements. • Assumes task adaptability, which may not hold for low-skilled workers. • Although they consider wider socio-economic dimensions, they Ignore cost-benefit analysis of automation adoption (e.g., firms may retain workers if wages are low). • No discussion of AI-augmented roles. • Does not model job creation from new technologies.

Methodology	High-level findings	Critique
Task-Based, Ability-Centric Framework (Felten et al., 2019) • Distinction between "narrow AI" (current, domain-specific) and "general AI" (future, hypothetical). • Introduces the AI Occupational Impact (AIOI) measure: it captures complementarity and substitution.	• AI exposure linked to higher wages, especially in high-skill jobs. • AI's benefits more among high-skill, high-income, software-intensive jobs, potentially exacerbating labour market polarisation. • No significant job loss observed.	• Correlational, not causal. • Static snapshot of occupational structures. • Techno-deterministic: Ignores economic and regulatory factors. • It does not address the potential for displacement or wage suppression in lower-skill roles.
Expert Elicitation and Surveys (Grace et al., 2018) • Underlying premise: expert predictions more reliable forecast than relying solely on objective measures or individual predictions. • Survey to elicit predictions about AI timelines, capabilities, and societal impacts.	• 50% chance of AI achieving "High-Level Machine Intelligence" (HLMI) within 45 years. • Asian respondents expect HLMI sooner than North Americans. • HLMI to outperform humans in tasks like language translation by 2024, writing high-school essays by 2026, driving trucks by 2027, working in retail by 2031, writing a bestselling book by 2049, and working as a surgeon by 2053.	• Subjective expert bias, (researchers from two specific conferences were surveyed., overly optimistic or pessimistic due to their immersion in the field, not be fully aware of the limits of their knowledge). • Lacks empirical grounding. • Broad and ambiguous definitions. • Examples like writing high-school essays fell into the predicted timeline already as of year 2025. • Techno-deterministic: does not consider the economic, social & political factors.
Political Economy of Task-Based Approach with Job Displacement vs. Job Creation (Acemoglu & Restrepo, 2018) Theoretical, not empirical, work with 3 Countervailing Forces: • Productivity Effect. • Capital Accumulation. • Deepening of Automation.	• Productivity Effect: increasing demand & creating jobs. • Capital Accumulation: driving investment in capital & jobs. • Deepening of Automation, with labour demand falling leading to job displacement.	• Theoretical, lacks empirical testing. • Difficult to measure new task creation. • Limited policy implications.
Suitability for Machine Learning (SML) Rubric (Brynjolfsson et al., 2018) • Employs algorithms (e.g., regression, random forests, neural networks) to model & predict labour market trends using historical data & AI exposure indices. • Rubric-based task scoring for SML.	• Most occupations have some SML tasks; few have all. • Job redesign required to leverage ML potential. • SML correlation with wage percentile/wage bill is low: this wave of automation may affect a different part of the labour force than previous ones.	• Opaque "black box" models. • Subjectivity of SML Rubric. • Limited empirical validation. • Techno-deterministic: ignores economic and organisational factors.

Methodology	High-level findings	Critique
Scenario Analysis (OECD, 2019) • Eclectic approach including: i) Skills-Biased Technological Change (SBTC) ii) Task-Based Approach based on Autor et al. (2003) iii) Human Capital Theory iv) Institutional Economics on labour market outcomes v) Welfare Economics.	• 14% of jobs could disappear due to automation in the next 15-20 years within OECD countries • 32% of jobs may change radically. • Non-standard workers are more vulnerable.	• Highly dependent on assumptions. • Broad generalisations without country-specific nuances. • Techno-optimistic about job creation. • Less focus on innovation policy VS. the skills development & adaptation: employee-, firm- and market-level individualism.
Historical data for Task Within Occupation Approach (Das et al., 2020) • "Task-share" for each task within each occupation. • Time series of task-shares to predict future task demands (ARIMA model).	• Big data & AI have risen significantly among high-wage occupations. • There are 15 tasks that are mentioned in more than 300 occupations, 3,976 tasks that occur in fewer than 10 occupations, some of the tasks that appear in only 1 occupation. • Healthcare and IT industries involve more tasks.	• Job postings may not reflect actual work. • Limited scope (2010-2017 data). • ARIMA model may not capture labour market complexities. • It documents change in task demands but does not fully explore their underlying causes.
Historical data for Input-Output Economic Modelling (Acemoglu & Restrepo, 2020) • Model Inputs: industrial robot density, employment data, wage, demographic and economic controls. • Model Outputs: employment, wage, labour market polarisation, displacement & productivity effects.	• 1 robot per 1,000 workers reduces employment by 0.39% and average wages by 0.77% between 1990 and 2007. • Effects are most pronounced in manufacturing.	• Focuses on industrial robots, not service-sector AI. • Short time frame (1993-2007). • Does not account for long-term worker adaptation.

Methodology	High-level findings	Critique
Text-Based Approach on Job Task Descriptions & Patents Analysis (Webb, 2020) • Task-Based within Occupation Framework adapting Acemoglu and Restrepo (2018) • Patent-Text Analysis using verb-noun pairs extracted from job descriptions and patents to quantify the overlap, calculating an exposure score. • Originally applied to software and robots, estimating exposure scores & employment and wages (1980-2010), then applied to AI exposure.	• AI disproportionately affects high-skilled tasks, potentially reducing 90:10 wage inequality. • Higher education, lower exposure for robots (higher) & software (lower). • Higher education, higher exposure for AI. • Estimates sectoral impacts and job displacement.	• Assumes fixed technology adoption patterns. • Ignores worker and firm adaptation. • Relies on patent data, which may lag behind actual trends. A fragment of patents is adopted. • Techno-deterministic: ignores economic and organisational factors.
NBER Task Exposure Index (Acemoglu et al., 2020) • They classify establishments as "AI-exposed" based on the compatibility of their workers' tasks with current AI capabilities, using 3 measures (Felten et al., 2018, SML from Brynjolfsson et al., 2019 & Webb, 2020). • Also correlation between AI exposure and employment / wages	• 15% rise in AI vacancies for exposed sectors. • No aggregate employment or wage effects at industry/ occupation level. • 3 measures agree: Information, Professional services & Manufacturing more exposed, VS. retail trade that are less. • 3 measures disagree: Agriculture, Mining, Government. • SML shows smooth wage changes per occupation, Felten et al. staggering effects on occupation wage, & Webb shows diminishing returns.	• Limited by online vacancy data representativeness. • Measures exposure, not actual impact. • Short time frame (2010-2018). • It focuses on establishments using AI, rather than producing it. • Possible endogeneity as firms experiencing declining hiring may be more likely to adopt AI.

Table 2 - Recent Academic, Policy, and Think Tank Works (2021-2026)

Methodology	High-level findings	Critique
Political Economy of Dynamic Task Share Analysis & Historical data (Acemoglu & Restrepo, 2022) • Task-Based Framework, where automation displaces workers from tasks where they have comparative advantage, reducing their wages. • General Equilibrium Analysis.	• 50-70% of US wage structure changes (1980-2016) driven by automation. • Robots, software, and machinery adoption explain 45% of industry labour share declines (1987–2016). • Automation raised TFP by 3.4% but caused stagnant wages.	• Relies on routine tasks as a proxy, missing AI's impact on non-routine roles. • Static assumptions about worker adaptation. • Despite large distributional effects, modest productivity gains raise questions about automation's net benefits.
Technology Adoption Model (Firm-Based) (Acemoglu et al., 2022) • Task-Based Framework, where automation displaces workers from tasks where they have comparative advantage, reducing their wages. • It links technology adoption with firm characteristics.	• Adoption of these technologies remains low, especially for AI (3.2% of firms) and robotics (2% of firms). • Higher adoption rate for specialised software (40.2%) and cloud computing (34%). • Use of advanced technologies is associated with an 11.4% higher labour productivity, and lower labour shares.	• Descriptive, not causal. • Limited time frame (2016-2018). • Self-reported data by firms may be biased. • Skill heterogeneity among groups or demographic groups not dealt with.
Firm Case Study on Productivity Analysis (Brynjolfsson et al., 2023) • Quasi-Experimental Design and Difference-in-Differences (DiD) Approach touching upon the task-based framework of automation and explores AI effects of across different skill levels at a Fortune 500 firm selling business-process software.	• AI assistance increased productivity by 14-15%. • Less experienced workers benefited most (34-35% gains).	• Single-firm, single-task context. • Short-term effects only. • No wage or aggregate employment data.

Methodology	High-level findings	Critique
LLM-Driven Task- and Ability-based framework Job Impact (Felten et al., 2023) • It builds on a task- and ability-based framework AI Occupational Exposure (AIOE) from Felten et al. (2018, 2021). • Crowd-sourced survey to link 10 AI capabilities to 52 human abilities from O*NET. • Also construct a language-model and image-generation AIOE by focusing only on language model abilities.	• Telemarketers, teachers, and sociologists are most exposed to LLMs. • High-wage, creative, and high-educated jobs are more exposed. • Women, White, and Asian are more exposed.	• Exposure does not distinguish augmentation vs. displacement. • Static snapshot.
Scenario Planning for AGI Future (Korinek & Suh, 2024, Korinek 2023) • The paper models AGI's impact on output and wages, assuming work is made of tasks with: bounded vs. unbounded task complexity. • It shows wages depend on the balance between automation and capital-driven demand.	• (a) Business-as-usual scenario (b) Baseline AGI scenario (c) Aggressive AGI scenario (d) Mixed scenario • Bounded task complexity could lead to wage collapse; unbounded could sustain wage growth. • AGI may cause rapid growth but risks wage declines.	• Theoretical, lacks empirical validation. • Assumes exogenous automation index growth. • Simplifies labour and capital dynamics.
Task-Based within Occupation Exposure Analysis on LLMs (Eloundou et al., 2023) • Task exposure, acknowledging both potential augmentation and displacement effects. • Development of a new rubric to assess occupations based on their alignment with LLM capabilities. • Integration of human expertise and GPT-4 classifications to apply rubric to occupational data.	• 80% of US workers could have ≥10% of tasks affected by LLMs. • 19% may see ≥50% task impact.	• Techno-deterministic: findings rely on exposure, not adaptation. • Limited data (2023); too early for empirical validation. • Potential bias in human/GPT-4 annotations.

Methodology	High-level findings	Critique
Tony Blair Institute (TBI) – Scenario-Based Modelling (TBI, 2024) • AI's impact on labour markets through: Labour Demand, Labour Supply, and Task-based. • Scenario analysis: Tailwind, Jet Stream, Whirlwind, Breeze. • UK labour-market data to contextualise AI-induced displacements	• AI could displace 1-3 million UK jobs, peaking at 60,000-275,000 annually. • Potential GDP growth up to 6% by 2035. • 23% private-sector productivity gains, with admin roles most exposed to displacement. • TBI regional data complements GWM2025: areas excelling in GWM are more exposed to AI, while lower-performing regions face less exposure—possibly due to firms prioritising automation of high-wage, high-autonomy roles.	• Reliance on historical analogies, and assumption that AI's impact will mirror past technologies (e.g., ICT). • Assumes smooth transition; underestimates wage polarisation. • 'Time-saving' focus might underplay job displacement.
IPPR – Task-Based Exposure Analysis (Jung & Srinivasa Desikan, 2024) • Task-based assessment of GenAI. • GenAI on labour market depends on the i) design, ii) adoption, and iii) policy choices. • AI can be used for "augmentation" (boosting worker productivity) or "displacement" (reducing labour demand). • Scenario planning: Full Augmentation, Full Displacement, Central Scenario (augmentation and displacement)	• 11% of tasks in the UK (scaled by hours worked) are exposed to 'here and now' GenAI. • Full Augmentation: Widespread integration of generative AI leads to a 13% boost to GDP with zero job displacement. • Full Displacement: 8 million jobs are lost with no GDP gains. • Central Scenario (Augmentation and displacement): 4.4 million jobs are lost, but there are significant economic gains of about 6.4% of GDP. • Advocates for "job-centric industrial strategy" in the UK.	• Relies on three scenarios with uncertain AI adoption assumptions ('here and now GenAI' vs more integrated AI systems). • Heavy use of expert surveys, lacking labour market dynamics. • Qualitative metrics lack clarity and causal AI links. • Empirical analysis is limited; outcomes are exploratory, not predictive.
Harvard Business School (HBS) Augmentation Index (Chen et al., 2024) • GenAI on the labour market through displacement or complementation. • Argument: Automatable tasks will experience displacement, while those mixed occupations (automatable & non-automatable tasks) augmentation. • GenAI's target: cognitive tasks VS past manual/blue-collar work. • Synthetic Difference-in-Differences (DiD).	• 17% decrease in job postings for automatable roles. • 22% increase for augmentation-prone roles. • 24% decrease in GenAI-exposed skills among jobs in the top quartile of automation exposure. • 15% increase in GenAI-exposed skills for jobs most susceptible to augmentation.	• Limited longitudinal perspective. • DiD can address endogeneity better than simple correlations, but it is still difficult to establish causality. • ChatGPT introduces subjectivity.

Methodology	High-level findings	Critique
AI Startup Exposure (AISE) Index (Fenoaltea et al., 2024) • Index based on occupational descriptions & AI applications. • Critiques techno-determinist models, proposing analysis on dynamics & market realities. • Startups as indicative of venture capital investments, reflecting economic viability & societal desirability. • Comparing the AISE rankings with those of Felten et al. (2021)'s AI Occupational Exposure (AIOE) index.	• High-skilled roles are heterogeneously targeted by startups. • Roles with routine tasks show significant exposure. • Ethical or high-stakes jobs (e.g. judges, surgeons) see lower exposure. • AI impact is concentrated in knowledge hubs and service sectors.	• Focuses on startups, missing broader industry trends. • Semantic similarity ≠ actual impact. • Llama3 is a black box - it adds objectivity, but its internal logic and similarity scoring remain opaque.
Natural Language Processing (NLP): AI Adoption and Task-based Exposure (Hampole et al., 2025) • NLP/LLMs to measure firm-level AI adoption based on employee resumes (2010–2023). • AI applications matched to O*NET tasks based on semantic similarity to construct task-level AI exposure measure. • Occupation-level mean exposure and concentration of exposure measured. • Decompose total employment effect into substitution, reallocation, and firm productivity effects	• Firm-level AI adoption linked to greater revenue, productivity, profits, and employment. Non-adopting firms have lower labour demand due to slower growth. • Exposure to AI increases substitution but offset by reallocation & productivity gains. • Demand for routine skills declines, while AI-complementary skills remain stable. • Highest-paid occupations see greatest decrease in employment as exposure increases	• NLP models rely on quality data and may miss context around AI adoption. • Misses post-ChatGPT acceleration. • Relies on job descriptions / professional profiles, which may not reflect actual skill use or adaptation. • Productivity gains are modelled but not empirically validated at the macro level.
Deep Learning Models – LSTM for Labour Demand Forecasting (Kim, 2025) • Applies Long Short-Term Memory (LSTM) models (a type of recurrent neural network, RNN) to forecast Job Openings and Labour Turnover Survey (JOLTS) data. • Then compares performance with traditional autoregressive (including ARIMA, SARIMA, & Holt-Winters) models.	• LSTM outperforms traditional models in predicting JOLT job openings. • Captures complex temporal dependencies.	• Mainly methodological research. • High computational cost. • Limited interpretability. • Overstates LSTM advantages without acknowledging limitations.

Methodology	High-level findings	Critique
Task-Level Automatability Prediction Using BERT (Xu et al., 2025) • Skill-Biased Technological Change (SBTC) & Routine-Biased Technological Change (RBTC) hypotheses. • Utilises a BERT-based classifier (Bidirectional Encoder Representations from Transformers) to analyse & classify text to assess the automatability of individual tasks across occupations, offering a granular analysis.	• 25.1% of occupations at substantial automation risk. • BERT outperforms traditional ML models.	• Relies on quality of training data. • The expertise of the annotators is not fully detailed. • GPT-4 Reliance. • No real-world validation. It does not examine the economic feasibility of these tasks getting automated • Limited details on baseline models
Deloitte Employee Surveys (Cantrell et al., 2025) • Employee Value Proposition (EVP) as a unique set of benefits that an employee receives in return for their skills, capabilities, and experience. • Human-Machine Collaboration proposes a shift from task replacement to augmentation & ultimately to convergence.	• AI can automate 40–60% of routine tasks, but few complex ones. • 77% say AI has increased workload & reduced productivity; 61% expect more burnout. • 14% of EU workers face algorithmic management, which may lower care & accuracy. • 33% report less human interaction; 28% feel a loss of personal connection. • 54% worry about blurred lines between human and AI work. • 60% link increased turnover to AI-driven worker data use.	• Methodology details (e.g. sampling, analysis) are missing. • Potential bias between findings and company's services. • Data reflects business/HR leaders more than workers. • Terms like "convergence" & "human capabilities" lack clear definitions. • Focuses on AI's benefits & EVP adaptation, with little attention to inequality or worker challenges. • Minimal discussion of algorithmic bias, privacy, job displacement.
Use-based survey measure (Humlum & Vestergaard, 2025) • Difference-in-difference analysis linking administrate data on employment and survey data on employer initiatives of AI to estimate impacts on employment in 11 exposed occupation • Employee perceptions of employer policies towards AI - i.e. whether AI is encouraged, whether there is an enterprise chatbot, and training – is taken to proxy exposure	• Workplaces with encouraged AI use did not have different employment patterns to firms where AI use is not encouraged • AI-adopting employees more likely to have transition, translating into 4% more movement in full-time equivalent • No significant change in pay or hours worked	• Potential confounders with employer policies and employment, for example management practice • Limited to 11 occupations • Current AI use may not reflect future automation

Methodology	High-level findings	Critique
Global Index of Occupational Exposure to GenAI (Gmyrek et al., 2025) • Polish occupational classification system used to calculate GenAI exposure for each 4-digit ISCO • Based on mean exposure and variability in exposure of tasks, occupations divided into four exposure gradients, minimal exposure, and no exposure • Method included i) three algorithmic assessments of potential exposure of each task ii) survey of workers on how likely they think a sub-sample of tasks representative of their professional group can be automated and iii) validation from experts on labour (not on AI) of a sub-set of tasks	• Jobs with cognitive tasks more exposed than those with high physical share. Most exposed occupations are clerical (gradient 4) • Globally, 1 in 4 jobs sits within one of the exposure gradients (i.e. above minimal), with more women being (highly) exposed (28% for women vs 21% for men) • High-income countries are more exposed (34% of employment), low-income countries are least exposed (11%) • Europe and Central Asia most exposed region (32%)	• Poland's occupation classification system may not be representative, especially of other regions (e.g. Low income countries) • May focus too much on what AI can do relative to what it will be able to do in near future. Labour expert validation may be insufficient to judge potential for task automation. • Focus on substitution risk (e.g. in survey), rather than range of possible impacts from AI
AI Exposure Score by Joint Research Centre (European Commission, 2025) • Indirect matrix between AI research and work tasks • Measured using relative AI research intensity across AI benchmarks (expert assessment of which cognitive abilities are relevant) and relative required intensity of cognitive abilities for tasks (based on expert assessment of which cognitive abilities are necessary for each task) • This allows exposure score to be updated based on AI capabilities (proxied by AI research intensity across benchmarks)	• Most exposed occupations are electrotechnology engineers, software developers, and secretaries • Positive relationship between AI exposure and wages • AI benchmarks relating to images, texts, video, and audio have grown significantly, meaning that occupations deemed exposed has changed too	• Does not consider substitution vs other effects – as a result, teachers are deemed highly exposed, even though they are unlikely to be automated due to in-person element • Limited set of occupations considered • AI benchmarks may not reflect actual capabilities (see Kapoor et al., 2026)

Methodology	High-level findings	Critique
Occupational AI Applicability Score (Tomlinson et al., 2025) – Microsoft • AI Applicability score based on the extent to which AI usage completes activities that form the occupation (assessed by LLM and validated via user feedback). • Data from 200,000 anonymised Bing Copilot conversations, classified by an LLM into O*NET work activities. • Distinguish between AI assisting or performing a task – AI Applicability Score separated.	• Highest AI Applicability Score for Translators and interpreters, historians, authors, sales representatives, and programmers. • Activities where AI use leads to successful completion of activity concentrate in information creation, gathering, and dissemination. • Physical occupations tend to use AI to assist, whereas media and financial operations delegate disproportionally to AI.	• Biased sample of Microsoft Copilot usage, non-usage not recorded. • Completion measure foundational to AI Applicability score is binary, does not capture whether quality of output changes (for worse or better).
Task-based generative AI exposure Index – LLM task rating and worker-reported task data (Henseke et al., 2025) • Combines task-occupation ratings of AI exposure from nine LLM raters (building on Eloundou et al., 2023), weighted by worker-reported task importance to construct a job-level Generative AI Exposure Index. • Study how this index relates to workers' reported AI use.	• LLM ratings have relatively high agreement in ranking jobs in their exposure, but disagree about the level of exposure for individual jobs. • AI exposure has increased in the UK between 2017 and 2024, mainly because of the labour market composition. • More exposed roles have seen decrease in vacancies relative to less exposed roles. • AI use and exposure are positively correlated, and use has increased disproportionally in more exposed groups.	• Bias that is shared across LLM raters cannot be accounted for. • Assumes that task importance increases job-level exposure, but expertise of tasks that are exposed matters too (Autor & Thompson, 2025).
Firm-Level Adoption via Job Postings (Hosseini Maasoum & Lichtinger, 2025) • Like Hampole et al. (2025), uses job postings to identify AI adoption, but flags GenAI-specific "integrator" roles and does not combine adoption with task exposure. • Triple difference to measure difference in hiring by seniority and adoption	• 3.7% of firms posted a GenAI integrator role, but those made up 16% of total employment • Observe seniority-based technological change: roughly 9% decrease in junior relative to senior employment in adopting firms, compared to non-adopting firms • Negative effect on junior employment driven primarily by slower hiring than increased separations, in part with firms possibly adjusting because of expectations of productivity increases	• Adopting firms may differ in many ways from non-adopting firms beyond their approach to AI. • Posting AI integrator roles may not capture AI adoption in firms, either because of existing capabilities (not reflected in hiring) or silent adoption.

Methodology	High-level findings	Critique
GDPval: Expert-Graded Realistic Task Benchmark (Patwardhan et al., 2025) – OpenAI • Experiment comparing human vs AI performance at 1320 representative tasks across 44 occupations, created by experienced professionals. • Experts blind-grade model vs human outputs • Measure is whether AI output wins or ties against human output when paired and ranked by industry professional.	• Highest average win/tie rate is 47.6% • Highest win/tie rate is in retail trade sector (56%) • Highest win/tie rate is for sales managers (79%) • Win/tie rate increasing linearly with improved AI models over time • Marginally lower win/tie rate when less context provided	• Measured in lab setting – real-world context likely significantly messier in ways likely not accounted for by reduced context robustness check. • Focus on self-contained knowledge work, which may only reflect a limited sub-set of tasks across occupations.
Occupation-level expertise (Klein Teeselink & Carey, 2026) • Operationalise expertise (Autor & Thompson, 2025) to study AI's impacts on wages and employment across 39 countries. • Expertise elicited from LLMs for all O*NET task descriptions on a scale of 0-100, aggregated to the occupation. • Difference-in-difference comparing occupations based on their AI exposure and expertise change.	• Small increase in advertised wages where automation is likely to increase expertise, no significant effect on job postings. • Countries with more stringent labour regulations see a larger decreases in hiring in AI-exposed occupations. • Countries with more digital readiness see smaller decreases in hiring in AI-exposed occupations.	• Exposed tasks may be restructured rather than automated. • Predicted rather than realised changes in expertise. • Does not account for new task creation and impacts on expertise.
Observed Exposure (Massenkoff & McCrory, 2026) - Anthropic • Combining AI usage data from Claude products and API with AI exposure (Eloundou, 2023), i.e. whether the task can be theoretically done by an LLM or an LM with an additional tool • Differentiate between whether Claude is being used for automation or augmentation – greater exposure where Claude is used more for automation	• For all areas, Claude use is far below what theoretical exposure predicts (Eloundou et al., 2023). • Only a weak negative relationship between observed exposure and US Bureau of Labor Statistics employment predictions and job finding rate for exposed under-25s has decreased 14% • Women, older people, White people and Asians, higher earners, and higher educated are more exposed by this measure	• AI use is different from exposure, and non-use may still be coupled with high vulnerability for automation • Claude usage does not represent all AI use, and may be biased towards, for example, computer science and coding

Methodology	High-level findings	Critique
AI Jobs Transition Framework & Realised Exposure (Richmond, 2026) – OpenAI • Framework to understanding the possible outcomes from exposure based on whether AI can do a meaningful share of tasks, whether demand can absorb productivity increases, whether human worker remains central to delivery, and existing AI use for tasks. • Operationalised as realised exposure: combines share of task time with theoretical AI exposure (Eloundou, 2023) and observed work-related ChatGPT usage for that task.	• Defines four occupational groups: high automation risk (18%), reorganisation (24%), growing with AI (12%), less immediate change (46%). • Large gap between realised exposure and theoretical exposure, described as 'capability overhang'. • Jobs that grow with AI and the most at risk of automation have the highest gap between usage and theoretical capabilities.	• Only based on usage data from OpenAI products, meaning that usage data may be biased. • Treats theoretical exposure as 'real' AI capabilities, but capability at real-world tasks might be better captured by usage data than theoretical capabilities self-estimated by LLMs due to messiness at application layer.
Firm-Level Adoption via AI Spending (Kharazian, Simon, & Stevens, 2026, Ramp / Revelio Labs) • AI adoption measured by spending on AI vendors to divide firms into high vs low adopters in the US. • Linked to employment data collected by Revelio Labs. • Staggered event study comparing high / low adopters to not-yet high / low adopters.	• High-intensity adopters increase headcount by around 10% across types of roles - no change for low adopters. • Entry-level headcount increases by 12%, contrary to many other studies. • Information sector sees largest gains.	• Not necessarily a representative sample of firms – non-adopting firms not analysed. • Endogeneity: later adopters may systematically differ from early adopters, even if spending levels are similar (e.g., they are already growing faster). • Data not easily accessible and therefore not replicable by independent researchers.
Simulation-Based Exposure & Economic Benchmarks (EconEvals) (Wan et al., 2026) • Combine chatbot conversation data and synthetic data to devise a "whitebox" AI exposure measure. • Usage data covers 143 activities, remaining activities addressed with synthetic data generated by a language model which simulates a worker, breaks each task into steps, and estimates time saved per step, producing reasoning traces alongside the estimate.	• 47% of US occupations could save at least 25% of their time for at least half of their tasks. • Predicted exposure exceeds usage in all SOC groups. • Biggest factor preventing time savings is that task requires physical presence. • Biggest factor enabling time savings is when task output involves communicative text.	• Limited to AI chatbots and not full range of AI tools. • Possible bias in synthetic data, synthetic benchmarks correlate more with each other than with real data. • Time-saving is LLM-generated, not validated against real-world experiments.

Methodology	High-level findings	Critique
Task-level GenAI Adoption Index (Bick, Blandin, Deming, and Schumacher, 2026) • Large-scale representative survey in the US asking about whether AI is used for a task within each occupation. • Unlike other studies integrating adoption (i.e. from Anthropic and OpenAI), they also measure who does not use AI.	• AI is used across almost all occupations, but only less than half of workers tend to use it, showing great adoption variation within the same occupation. • Exposure scores capture some but not all the variation in AI use between occupations (5%-53%). • Gap between adoption and exposure is greatest and negative for receptionists and information clerks (adoption < exposure) and positive for construction equipment operators (adoption > exposure).	• Only asks about whether AI is used for a task, not how much. It may therefore miss differences in depth of AI use for a given task.

Figure 1: Methods on AI's impact on labour markets

https://public.flourish.studio/visualisation/22803877/

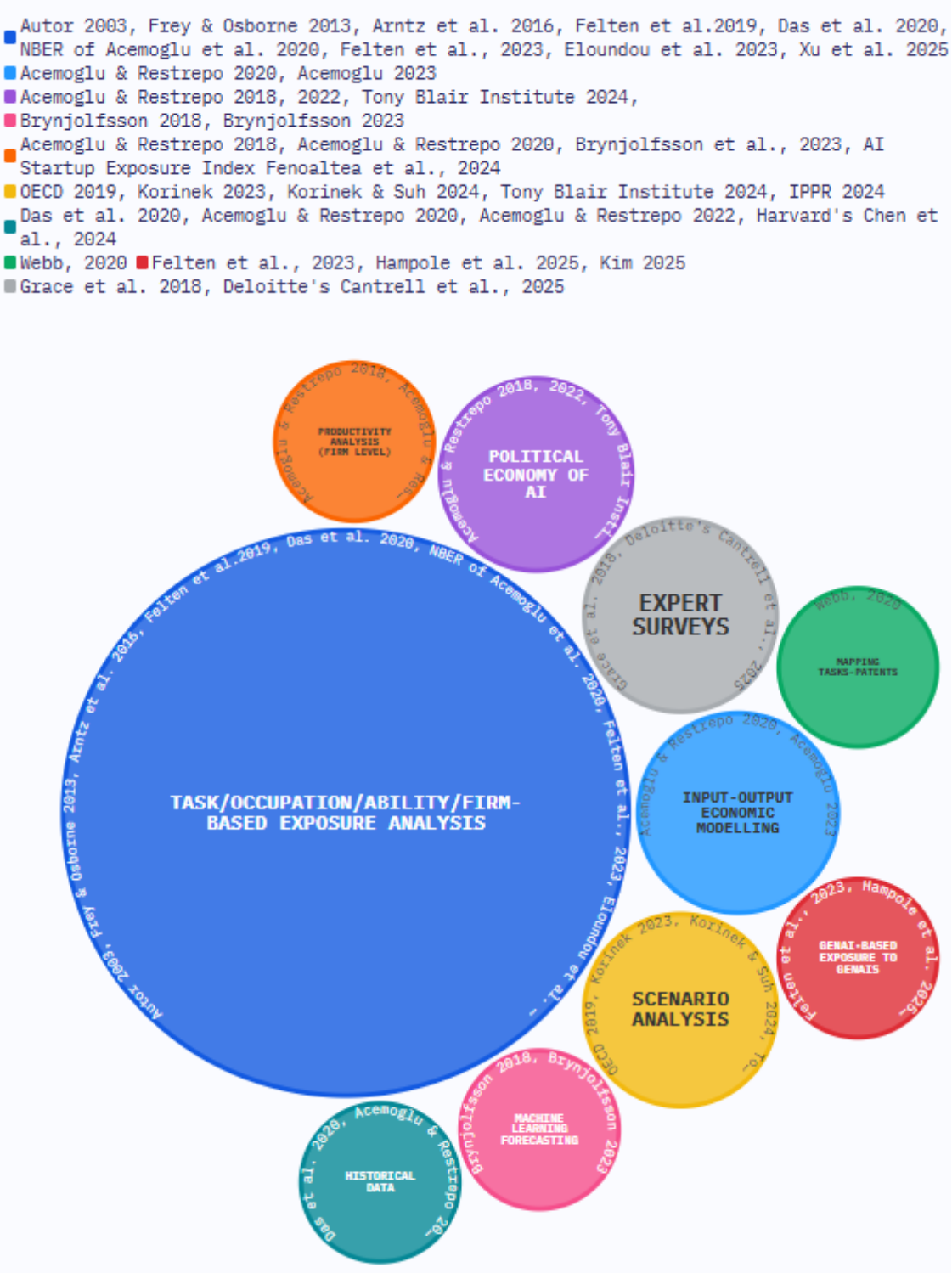
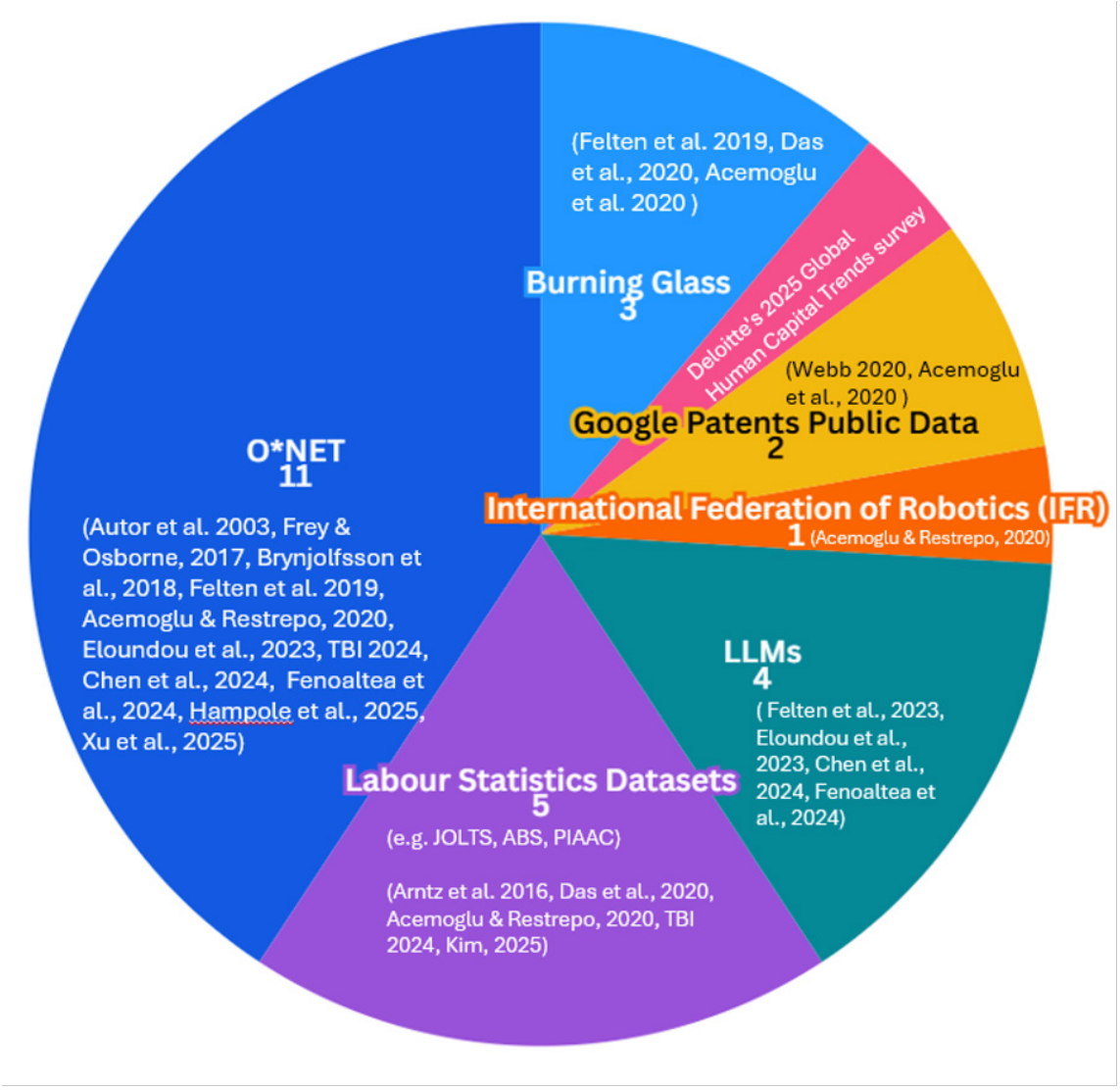


Figure 2: Datasets used

https://public.flourish.studio/visualisation/22803347/



References

Acemoglu, D., Anderson, G. W., Beede, D. N., Buffington, C., Childress, E. E., Dinlersoz, E., Foster, L. S., Goldschlag, N., Haltiwanger, J. C., Kroff, Z., Restrepo, P., & Zolas, N. (2022). Automation and the workforce: A firm-level view from the 2019 Annual Business Survey (NBER Working Paper No. 30659). National Bureau of Economic Research. https://doi.org/10.3386/w30659

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2020). AI and jobs: Evidence from online vacancies (NBER Working Paper No. 28257). National Bureau of Economic Research. https://doi.org/10.3386/w28257

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation and work (NBER Working Paper No. 24196). National Bureau of Economic Research. https://doi.org/10.3386/w24196

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

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in U.S. wage inequality. Econometrica, 90(5), 1973–2016. https://doi.org/10.3982/ECTA19815

Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. OECD Social, Employment and Migration Working Papers, No. 189. OECD Publishing. https://doi.org/10.1787/5jlz9h56dvq7-en

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

Autor, D. & Thompson, N. (2025). Expertise, Journal of the European Economic Association, Volume 23, Issue 4, August 2025, Pages 1203–1271, https://doi.org/10.1093/jeea/jvaf023

Bick, A. and Blandin, A. and Deming, D. and Schumacher, T. (2026). What Work Does Generative AI Do?. NBER Working Paper No. w35677, Available at SSRN: https://ssrn.com/abstract=7377853

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? AEA Papers and Proceedings, 108, 43–47. https://doi.org/10.1257/pandp.20181019

Brynjolfsson, E. Li, D. Raymond, L. (2025). Generative AI at Work, The Quarterly Journal of Economics, Volume 140, Issue 2, Pages 889–942, https://doi.org/10.1093/qje/qjae044

Cantrell, S., Flynn, J., & Scoble-Williams, N. (2025, March 24). AI is revolutionizing work: You need a human value proposition for the age of AI. Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2025/why-you-need-employee-value-proposition-for-age-of-ai.html

Chen, W. X., Srinivasan, S., & Zakerinia, S. (2024). Displacement or complementarity? The labor market impact of generative AI (Harvard Business School Working Paper No. 25-039). Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/25-039_05fbec84-1f23-459b-8410-e3cd7ab6c88a.pdf

Das, S., Steffen, S., Clarke, W., Reddy, P., Brynjolfsson, E., & Fleming, M. (2020). Learning occupational task-shares dynamics for the future of work. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (pp. 36–42). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375826

Dessart, F., Fernandez Macias, E. and Gomez, E. (2025). Anticipating the impact of AI on occupations: a JRC methodology. (JRC142580). European Commission. https://publications.jrc.ec.europa.eu/repository/handle/JRC142580

Eloundou, T. Manning, S., Mishkin, P. & Rock, D. (2024) GPTs are GPTs: Labor market impact potential of LLMs. Science. 384,1306-1308. https://doi.org/10.1126/science.adj0998

Felten, E. W., Raj, M., & Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. AEA Papers and Proceedings, 108, 54–57. https://doi.org/10.1257/pandp.20181021

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. Strategic Management Journal, 42(12), 2195–2217. https://doi.org/10.1002/smj.3286

Felten, E. W., Raj, M., & Seamans, R. (2023). How will language modelers like ChatGPT affect occupations and industries? arXiv preprint arXiv:2303.01157. https://doi.org/10.48550/arXiv.2303.01157

Fenoaltea, E. M., Mazzilli, D., Patelli, A., Sbardella, A., Tacchella, A., Zaccaria, A., Trombetti, M., & Pietronero, L. (2024). Follow the money: A startup-based measure of AI exposure across occupations, industries and regions (Version 2). arXiv preprint arXiv:2312.04637. https://arxiv.org/abs/2312.04637

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., Troszyński, M. Generative AI and Jobs: A Refined Global Index of Occupational Exposure. ILO Working Paper 140. Geneva: International Labour Office, 2025. ILO. https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729–754. https://doi.org/10.1613/jair.1.11222

Hampole, M., Papanikolaou, D., Schmidt, L. D. W., & Seegmiller, B. (2025). Artificial intelligence and the labor market (NBER Working Paper No. 33509). National Bureau of Economic Research. https://www.nber.org/papers/w33509

Henseke, G., Davies, R., Felstead, A., Gallie, D., Green, F., & Zhou, Y. (2025). Nine Raters, One Index: Carrying LLM Disagreement into Labour-Market Estimates. arXiv preprint arXiv:2507.22748. https://doi.org/10.48550/arXiv.2507.22748

Horton, J., Filipas, A., & Manning, B. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? (NBER Working Paper No. 31122). National Bureau of Economic Research. https://doi.org/10.3386/w31122

Hosseini Massoum, S.M. & Lichtinger, G. (2025). Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data. http://dx.doi.org/10.2139/ssrn.5425555

Humlum, A. & Vestergaard, E. Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI. NBER Working Paper 33777 (2025), https://doi.org/10.3386/w33777

Jung, C., & Srinivasa Desikan, B. (2024). Transformed by AI: How generative artificial intelligence could affect work in the UK – and how to manage it. Institute for Public Policy Research. https://www.ippr.org/articles/transformed-by-ai

Kharazian, A., Simon, L., & Stevens, R. (2026). A new look at AI's impact on jobs: Firm-level AI spending and workforce adjustment (Ramp Economics Lab Working Paper). Ramp. https://ramp.com/data/ai-jobs-impact/paper

Kim, K. (2025). Forecasting labor demand: Predicting JOLT job openings using deep learning model. arXiv. arXiv:2503.19048 https://arxiv.org/abs/2503.19048

Klein Teeselink, B. & Carey, D. (2026). AI, Automation, and Expertise. Available at SSRN: https://ssrn.com/abstract=6134506 or http://dx.doi.org/10.2139/ssrn.6134506

Korinek, A. (2023). Scenario planning for an A(G)I future. Finance & Development, 60(4), 30–33. https://www.imf.org/en/Publications/fandd/issues/2023/12/Scenario-Planning-for-an-AGI-future-Anton-korinek

Korinek, A., & Suh, D. (2024). Scenarios for the transition to AGI (NBER Working Paper No. 32255). National Bureau of Economic Research. https://doi.org/10.3386/w32255

Machovec, C., Rieley, M. J., & Rolen, E. (2025). Incorporating AI impacts in BLS employment projections: Occupational case studies. Monthly Labor Review. https://doi.org/10.21916/mlr.2025.1

Massenkoff, M. & McCrory, P. (2026). Labor market impacts of AI: A new measure and early evidence. Anthropic. https://www.anthropic.com/research/labor-market-impacts

OECD. (2019). OECD employment outlook 2019: The future of work. OECD Publishing. https://doi.org/10.1787/9ee00155-en

Patwardhan, T., Dias, R., Proehl, E., Kim, G., Wang, M., Watkins, O., ... & Tworek, J. (2026). Gdpval: Evaluating ai model performance on real-world economically valuable tasks. In International Conference on Learning Representations (Vol. 2026, pp. 24005-24040). https://arxiv.org/abs/2510.04374

Pissarides, C., Thomas, A., & The Review Team. (2025). The final report of the Pissarides Review. Institute for the Future of Work. https://www.ifow.org/publications/the-final-report-of-the-pissarides-review

Restrepo, P. (2023). Automation: Theory, evidence, and outlook (NBER Working Paper No. 31910). National Bureau of Economic Research. https://doi.org/10.3386/w31910

Richmond, A. M. (2026). The AI Jobs Transition Framework: Mapping AI's Near-Term Impact on Jobs. OpenAI Economic Research. https://cdn.openai.com/pdf/the-ai-jobs-transition-framework_report.pdf

Tomlinson, K., Jaffe, S., Wang, W., Counts, S., & Suri, S. (2025). Working with AI: measuring the applicability of generative AI to occupations. arXiv preprint arXiv:2507.07935.

Tony Blair Institute for Global Change. (2024, July 9). The potential impact of AI on the public-sector workforce. https://institute.global/insights/economic-prosperity/the-potential-impact-of-ai-on-the-public-sector-workforce

Tony Blair Institute for Global Change. (2024, November 8). The impact of AI on the labour market. https://institute.global/insights/economic-prosperity/the-impact-of-ai-on-the-labour-market

Tony Blair Institute for Global Change. (2024, November). The impact of AI on the labour market: Technical annex. https://assets.ctfassets.net/75ila1cntaeh/1z14gfxVQ2AVW4kl8M0tRL/67242d23e2fd49b3ce91b6e77f3003e6/The_Impact_of_AI_on_the_Labour_Market_-_Technical_Annex_-_TBI_-_Nov_2024.pdf

Wan, A., Hatgis-Kessell, S., Aguirre, T., Liang, P., & Bommasani, R. (2026). Economic Evaluations of Language Models. arXiv preprint arXiv:2607.19375. https://arxiv.org/abs/2607.19375

Webb, M. (2020). The impact of artificial intelligence on the labor market (Working Paper). Stanford University. https://www.michaelwebb.co/webb_ai.pdf

Xu, D., Yang, H., Rizoiu, M.-A., & Xu, G. (2025). From occupations to tasks: A new perspective on automatability prediction using BERT. arXiv. https://arxiv.org/abs/2502.09021

Institute for the Future of Work

IFOW is an independent research and development institute dedicated to transforming working lives for the better, co-founded by former employment barrister Anna Thomas MBE, Nobel prize-winning economist Sir Christopher Pissarides, and technologist Naomi Climer CBE.

Our core team at Somerset House works with a growing network of strategic partners striving for systems change.

Our vision is a future in which everyone flourishes in work they shape.

Our mission is to understand together how to transform working lives for good.

Our theory of change is that creating and sustaining good work is the best way to achieve this goal and ensure that innovation and social good advance together.

About our Toolkits

Our Toolkits present practical applications arising from IFOW's research team, research fellows, practioners and partners, offering policymakers, firms and academics tools to operationalise the latest thinking in the Future of Work space.

If you would like to further information, please contact team@ifow.org

Institute for the Future of Work
Somerset House
Strand
London
WC2R 1LA
ifow.org / @ifow.org

Citation

Papagiannaki, E., *Recent Methodologies on AI and Labour - a Desk Review.* London: Institute for the Future of Work.
DOI: 10.5281/zenodo.18801200

Permission to share

This document is published under the Creative Commons Attribution Non Commercial No Derivatives 4.0 England and Wales Licence. This allows anyone to download, reuse, reprint, distribute, and/or copy IFOW publications without written permission subject to the conditions set out in the Creative Commons Licence.