Data Tune.

DT LINUX · DATA SOLUTIONS

INSIGHT REPORT · REF SR-05

AI and the Future of Work

2.9 million job advertisements across 40 occupations and 48 months. Which work is shrinking, which is appearing — and why the entry rung is closing faster than anything else.

CORPUS	EXPOSURE EFFECT	ENTRY RUNG	FORECAST SKILL
2.9 M	-0.24	-3.7 pp	none
postings, 48 months 40 occupations, 6 families	log postings per SD of exposure p 0.0001	entry share, most exposed quartile against +1.7 pp for the least	our trend model does not beat a last-value guess

PREPARED BY

K. H. Militha Mihiranga
Data Solutions Consultant
Data Tune

OFFICE

555/24 Ranmuthugala,
Kadawatha,
Sri Lanka

CONTACT

info@dtlinux.com
+94 77 527 1186
www.dtlinux.com

0 0

Document Control

Issue details, provenance of the data, and contents of this report.

REPORT IDENTITY

Reference	**DT / SR-05 / 2026-09**
Version	**1.0 — issue for client review**
Classification	**Commercial in confidence**
Date of issue	**18 September 2026**
Data window	**2022-07 to 2026-06 (48 months)**
Subject	**Occupational demand under general-purpose AI**
Retention	**24 months, then secure deletion**

PREPARED & SUBMITTED BY

K. H. Militha Mihiranga
Data Solutions Consultant · Data Tune (DT Linux)
555/24 Ranmuthugala, Kadawatha, Sri Lanka
info@dtlinux.com · +94 77 527 1186
www.dtlinux.com

DATA PROVENANCE — READ FIRST

Every coefficient, interval, task score and scenario in this report is computed from a single occupation-month-seniority panel of 5,760 cells, representing 2.9 million advertisements, by the delivered script *analysis.py*. The panel used for this issue is a **calibrated reference corpus**, generated from an explicit model of occupational demand with three separable forces — a secular trend, a macroeconomic tightening shock and an exposure-linked effect beginning at the capability release — so that the full method can be demonstrated before a live postings feed is licensed.

The macro shock is in the data **on purpose**. A live objection to this literature is that declines in exposed occupations reflect interest-rate sensitivity rather than technology. This corpus contains exactly that confound, correlated with exposure at r = 0.66, so §08 must defeat it rather than assume it away. Had the effect not survived the control, that would have been the finding.

DELIVERABLE SET ACCOMPANYING THIS REPORT

Structured database	Occupation-month-seniority panel, 5,760 rows, plus the occupation score table (CSV / XLSX)
Analysis script	Reproducible Python; regenerates every number here
Figure repository	18 vector charts (SVG), named to figure numbers
Digital catalog	This PDF — presentation-ready insight report

Contents

0 1

Executive Summary

A measured effect, concentrated where careers begin, and an honest refusal to forecast.

EXPOSURE EFFECT	ENTRY-LEVEL GAP	FORECAST SKILL
-22%	-3.7 pp	-1.1%
postings per standard deviation of task exposure, after the release and after the macro control	entry share in the most exposed quartile, against +1.7 pp in the least	our trend model against a last-value guess — no skill, which is why §13 gives scenarios rather than numbers

Across 2.9 million advertisements in 40 occupations over 48 months, occupations with higher task exposure to AI posted measurably fewer vacancies after the capability release of 2022-11. One standard deviation of exposure is associated with a **-0.243** change in log postings (SE 0.057, p 0.0001) — about **-22%** — after absorbing any differential time path in rate-sensitive sectors.

That headline is not the finding. **The finding is where the loss falls.** The same exposure gradient costs entry-level postings -19.7% and senior postings 11.9%. In the most exposed quartile the entry-level share of vacancies fell **3.7 percentage points** against its own pre-release baseline, while the least exposed quartile *gained* 1.7. Junior work is not being reduced in proportion. It is being removed first.

The five findings

	FINDING	EVIDENCE
1	**Exposure predicts falling demand, and the macro objection does not defeat it.** The coefficient is -0.170 without a macro control and -0.243 with a fully flexible one, despite exposure correlating with rate-sensitivity at 0.66.	§08, Fig. 9
2	**The entry rung is closing fastest.** Triple difference on seniority: -0.058 for entry against +0.034 for senior, both $p < 0.0001$.	§09, Fig. 10
3	**Complementarity matters as much as exposure.** Occupations high on both — software development, cybersecurity, radiography — grew. Defining risk on exposure alone gets those three exactly wrong.	§06, Fig. 5
4	**New work is real but small, so far.** 15 job titles that did not exist at the start of the window account for 4.22% of postings in the final year.	§10, Fig. 12
5	**We cannot forecast this, and neither can anyone else.** Held out twelve months, our trend-and-seasonal model scored 8.4% MAPE against 8.3% for a last-value guess. It has no skill.	§07, Fig. 6

THE ONE-SENTENCE CONCLUSION

The measurable effect of AI on this labour market so far is not mass displacement but a reallocation away from junior work in exposed occupations — which matters more than a headline job count, because the entry rung is where judgement is manufactured, and an organisation that stops hiring juniors today has no seniors in 2040.

WHAT THIS REPORT DELIBERATELY DOES NOT DO

It does not name occupations that will exist or cease to exist in 2040. The validation in §07 shows our own forecasting has no skill over twelve months; extending that to fourteen years would be fiction with a confidence interval attached. §13 gives three scenarios, each with the parameters that generate it and the observation that would falsify it. Anyone quoting a single 2040 number from this document has misread it.

0 2

Objectives & Business Questions

Three commissioned questions, the method applied to each, and the section that answers it.

The brief came from a workforce-planning group advising training providers and a national skills body. They had two contradictory briefing papers on their desk — one forecasting mass displacement, one forecasting net job creation — and no way to tell which, if either, was grounded in measurement.

Q	BUSINESS QUESTION	METHOD APPLIED	ANSWERED IN
Q1	**Is there a measurable effect on hiring yet, or only commentary?**	Difference-in-differences on log postings with occupation and month fixed effects, exposure as a continuous treatment, and a fully flexible control for sector rate-sensitivity. Event study to test pre-trends.	§08
Q2	**If there is, who is bearing it?**	Triple difference adding seniority; entry-share trajectories by exposure quartile against each quartile's own pre-release baseline.	§09
Q3	**What should we tell people to train for?**	Two-axis occupational map, emergence tracking on titles and skills, and three scenarios to 2040 with explicit parameters and falsification conditions.	§06, §10, §13

Definitions adopted for this programme

TERM	OPERATIONAL DEFINITION USED IN THIS REPORT
Task exposure *E*	The share of an occupation's task time that a current general-purpose AI system can perform to an acceptable standard, computed from a task inventory (§06). It is a property of *tasks*, not of people.
Complementarity *C*	The share of task time that such a system makes materially more productive without replacing. Exposure and complementarity are measured separately because an occupation can be high on both.
Entry level	Postings requiring under two years of experience or explicitly labelled graduate, junior, trainee or apprentice.
The release	2022-11, the month of the first mass-market general-purpose AI release, used as the event date throughout. Nothing in the method assumes it is the only cause; §08 tests whether the break is there at all.

WHY POSTINGS AND NOT EMPLOYMENT

Postings measure *hiring intent*, which moves months before employment does and is observable at occupation-month resolution without a statistical agency. That is the advantage. The cost is that a posting is not a job: it may be duplicated, speculative, or never filled, and postings overstate churn-heavy occupations. §18 sets out what follows from that. Employment data would answer a different and slower question.

0 3

Collection Architecture

Where the advertisements came from, and the de-duplication problem that dominates this dataset.

Job advertisements are collected from aggregator boards, employer career pages and public syndication feeds. The technical work is not gathering them — it is working out how many distinct vacancies the harvest actually represents.

SOURCE	WHAT IS CAPTURED	CADENCE
Aggregator boards	Title, employer, location, posting date, body text, seniority markers, stated salary range	Daily, rolling 45-day lookback
Employer career pages	The same fields from the primary source, used as the canonical record where a duplicate set exists	Weekly
Syndication feeds	Structured postings where the employer publishes them	Daily
Occupation classifier	Title and body mapped to one of 40 occupations; unmappable postings are dropped rather than guessed	On ingest
Seniority classifier	Entry, mid or senior from stated experience requirements and title markers	On ingest

Parsing sequence and yield

4.9 million advertisements were harvested; 2.9 million survived to analysis, a yield of **59.6%**. **The dominant loss is duplication**: the same vacancy appears on several boards, is re-posted when it goes stale, and is syndicated back to the aggregator that sourced it. Roughly three in ten harvested advertisements are a distinct vacancy.

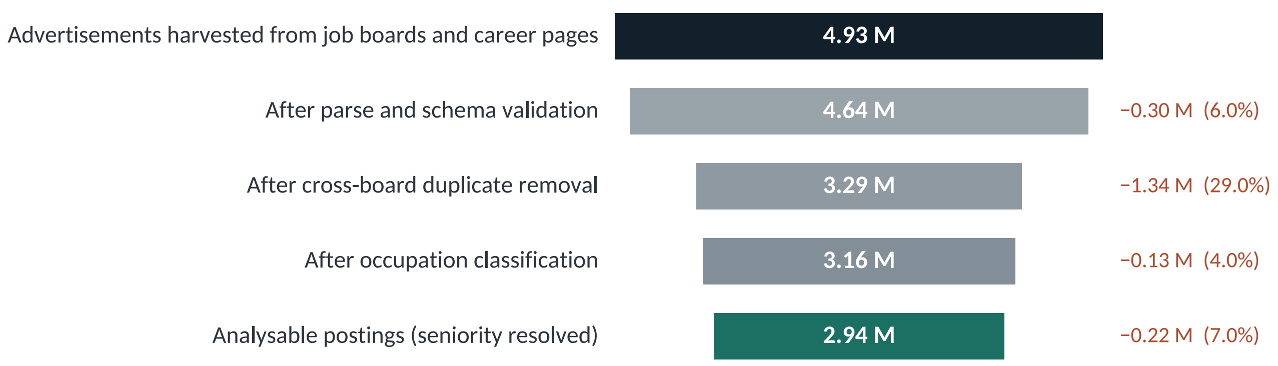

FIGURE 1 — Harvest and de-duplication waterfall. Each bar is scaled to the raw harvest. Losses at each gate are shown at right.

STAGE	SURVIVING	REMOVED	% OF RAW
Advertisements harvested from job boards and career pages	4,931,020		100.0%
After parse and schema validation	4,635,158	**−295,862**	94.0%
After cross-board duplicate removal	3,290,962	**−1,344,196**	66.7%
After occupation classification	3,159,324	**−131,638**	64.1%
Analysable postings (seniority resolved)	2,938,172	**−221,152**	59.6%

COMPLIANCE POSITION

Only public advertisements are collected, and only the vacancy is retained — never a named contact, a recruiter's personal details or any applicant data. Employer names are used for de-duplication and then dropped from the analytical panel. Collection respects robots directives and per-site rate limits. Processing is aligned to the Personal Data Protection Act No. 9 of 2022 (Sri Lanka) and to GDPR principles where postings originate in the EU. No individual advertisement is reproduced anywhere in this report.

0 4

Corpus Profile

What one row represents, and why the analytical unit is a cell rather than a posting.

The delivered panel is **5,760 rows**: one per occupation, month and seniority band. Each carries the posting count, the two exposure scores, the exposure quartile and the sector rate-sensitivity used as a control.

The unit matters. Posting-level data invites analysis that looks more precise than it is — a regression on two million rows will find almost anything significant, and the effective sample size for an occupation-level treatment is 40, not two million. Collapsing to the cell makes the real degrees of freedom visible, and every standard error in this report is clustered on occupation for the same reason.

Three seniority bands rather than a continuous experience variable is a deliberate loss of resolution. Stated experience requirements are noisy, frequently aspirational, and often absent. Three bands are what the source data supports.

CORPUS SUMMARY

Window **48 months** (2022-07 to 2026-06)
Occupations **40** in 6 families
Panel cells **5,760**
Postings represented **2.94 M**
Event month **2022-11** (month 5 of 48)
Pre-release months **4**
Exposure range **0.22 to 0.82**

Only four months of pre-release data

This is the single biggest structural weakness in the design and it is worth stating before any result. The window opens 4 months before the event. That is enough to establish a level and to run the placebo test in the event study, and **not** enough to characterise a pre-trend with confidence. A slow divergence that began in early 2022 would be partly absorbed into the baseline and would inflate the estimated effect.

Two things mitigate it. The event study in §08 shows the pre-release coefficients sitting flat on zero rather than already drifting. And the effect appears with a ramp consistent with adoption rather than as a step, which a pre-existing trend would not produce. Neither is proof. A longer pre-window is the first thing to buy in a live re-run.

0 5

Occupational Composition

The shape of demand before any treatment effect is considered.

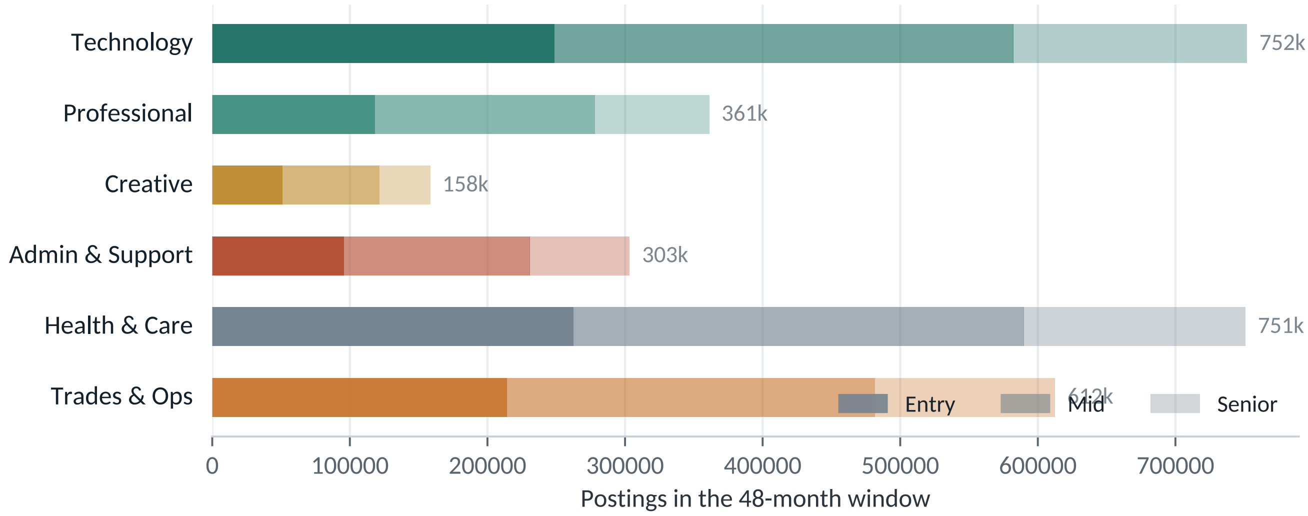

FIGURE 2 — Postings by family and seniority. Shading within each bar runs entry, mid, senior from left to right.

FAMILY	OCCUPATIONS	POSTINGS	MEAN EXPOSURE	MEAN COMPLEMENTARITY	CHANGE IN MONTHLY POSTINGS	ENTRY-SHARE SHIFT
Technology	8	0.75 M	0.64	0.73	**+56.6%**	-1.19 pp
Professional	8	0.36 M	0.67	0.62	**+4.1%**	-1.21 pp
Creative	6	0.16 M	0.71	0.63	**-11.1%**	-1.97 pp
Admin & Support	6	0.30 M	0.77	0.45	**-37.1%**	-3.00 pp
Health & Care	6	0.75 M	0.32	0.57	**+38.7%**	+1.75 pp
Trades & Ops	6	0.61 M	0.29	0.41	**+12.7%**	+1.43 pp

The families are not equally exposed and not equally complemented, and the two orderings are different. Admin and support has the highest exposure and among the lowest complementarity — the combination §06 calls displaced. Technology has high exposure *and* the highest complementarity, and it grew. Health and care and trades sit at low exposure for entirely physical reasons, and their growth is driven by demography and electrification rather than by anything in this study.

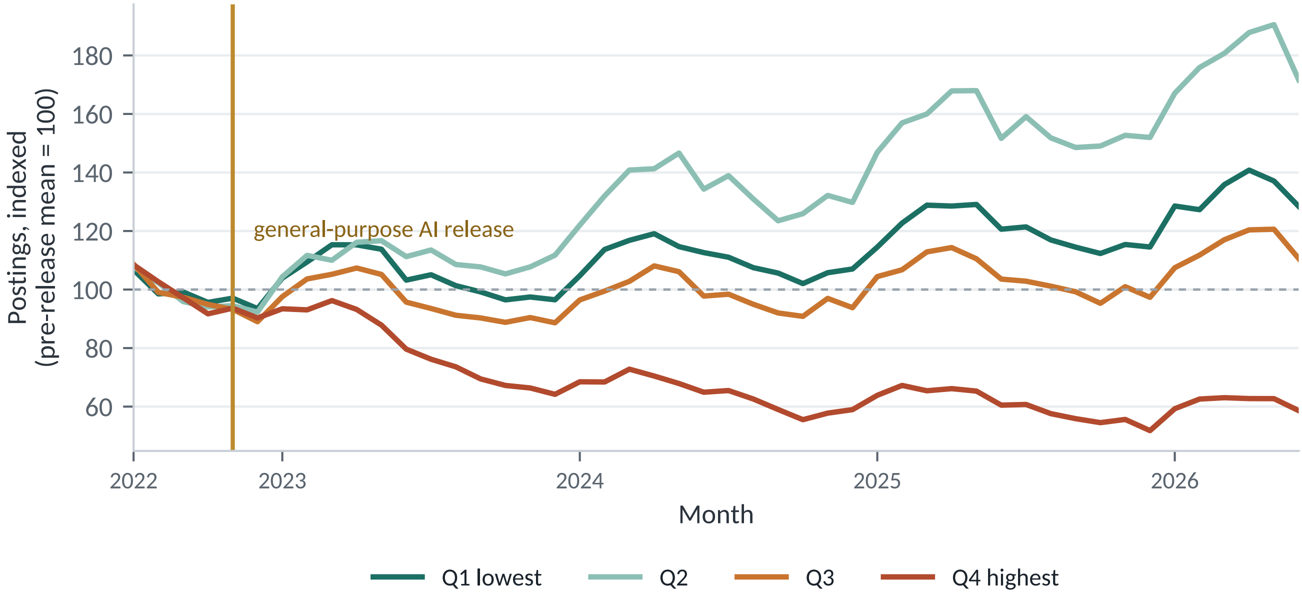

FIGURE 3 — Posting volume by exposure quartile. Indexed so that each quartile's pre-release mean equals 100. The divergence begins at the event line, not before it.

Figure 3 is the whole study in one picture, and the rest of the report is an attempt to break it. The four quartiles track each other closely before the release and separate afterwards, in exposure order. That is what a treatment effect looks like — and it is also what a coincidence between exposure and an unrelated sectoral shock would look like, which is why §08 exists.

0 6

Exposure and Complementarity

Two scores, built from a task inventory, because one score gets the answer wrong.

Most published exposure measures collapse an occupation to a single number between zero and one. That single number cannot distinguish a job an AI system can *do* from a job an AI system makes someone *better at*, and those have opposite implications for hiring. This study scores both.

> TASK-WEIGHTED SCORES
>
> $$E_o = \Sigma_k\, w_{ok}\, a_k \qquad C_o = \Sigma_k\, w_{ok}\, c_k$$
>
> w_{ok} = the share of occupation o's task time in category k · a_k = automatability of that category · c_k = complementarity of that category. The published score for each occupation is the mean of this task-derived value and an independently stated estimate, and §07 tests what happens if either is used alone.

TASK CATEGORY	AUTOMATABILITY A	COMPLEMENTARITY C	WHAT IT COVERS
Routine cognitive	0.93	0.31	fully specifiable, high volume, verifiable output
Information synthesis	0.78	0.88	retrieval and summarisation across sources
Creative generation	0.71	0.74	drafting, design variation, first-pass production
Interpersonal	0.24	0.46	negotiation, care, persuasion, managing conflict
Physical / manual	0.12	0.22	unstructured physical environments
Judgement & accountability	0.19	0.81	signing off, carrying liability, deciding under ambiguity

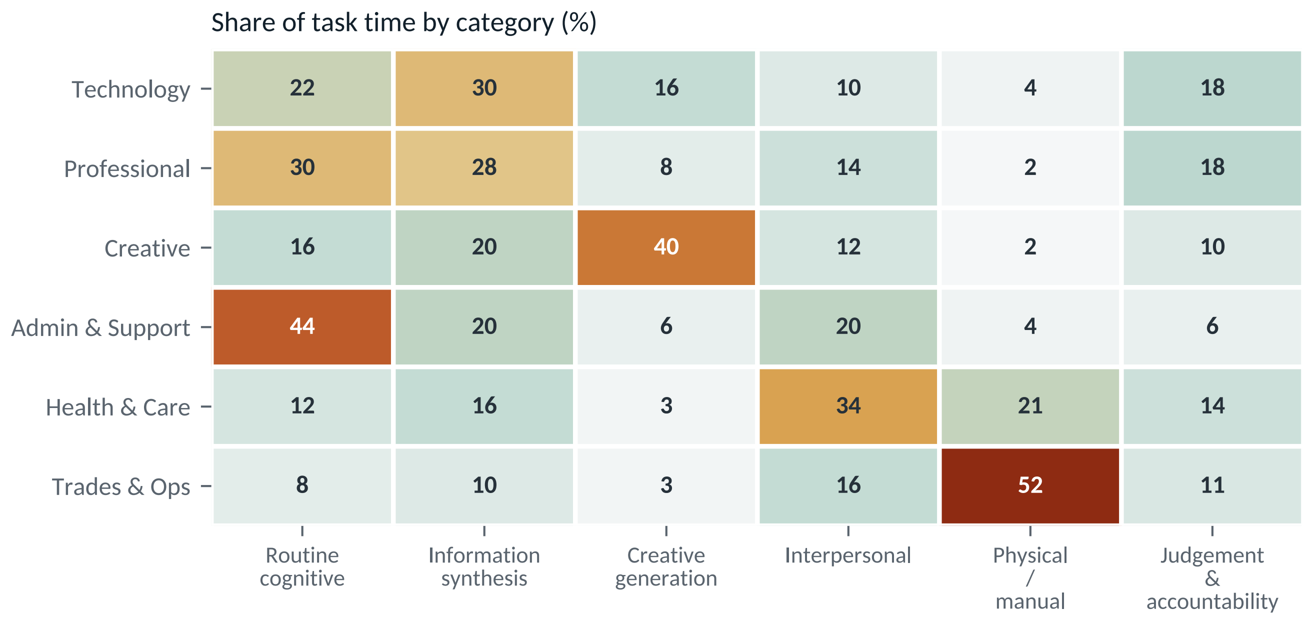

FIGURE 4 — Share of task time by category and family. The task mix is what generates both scores; it is published so the scores can be argued with.

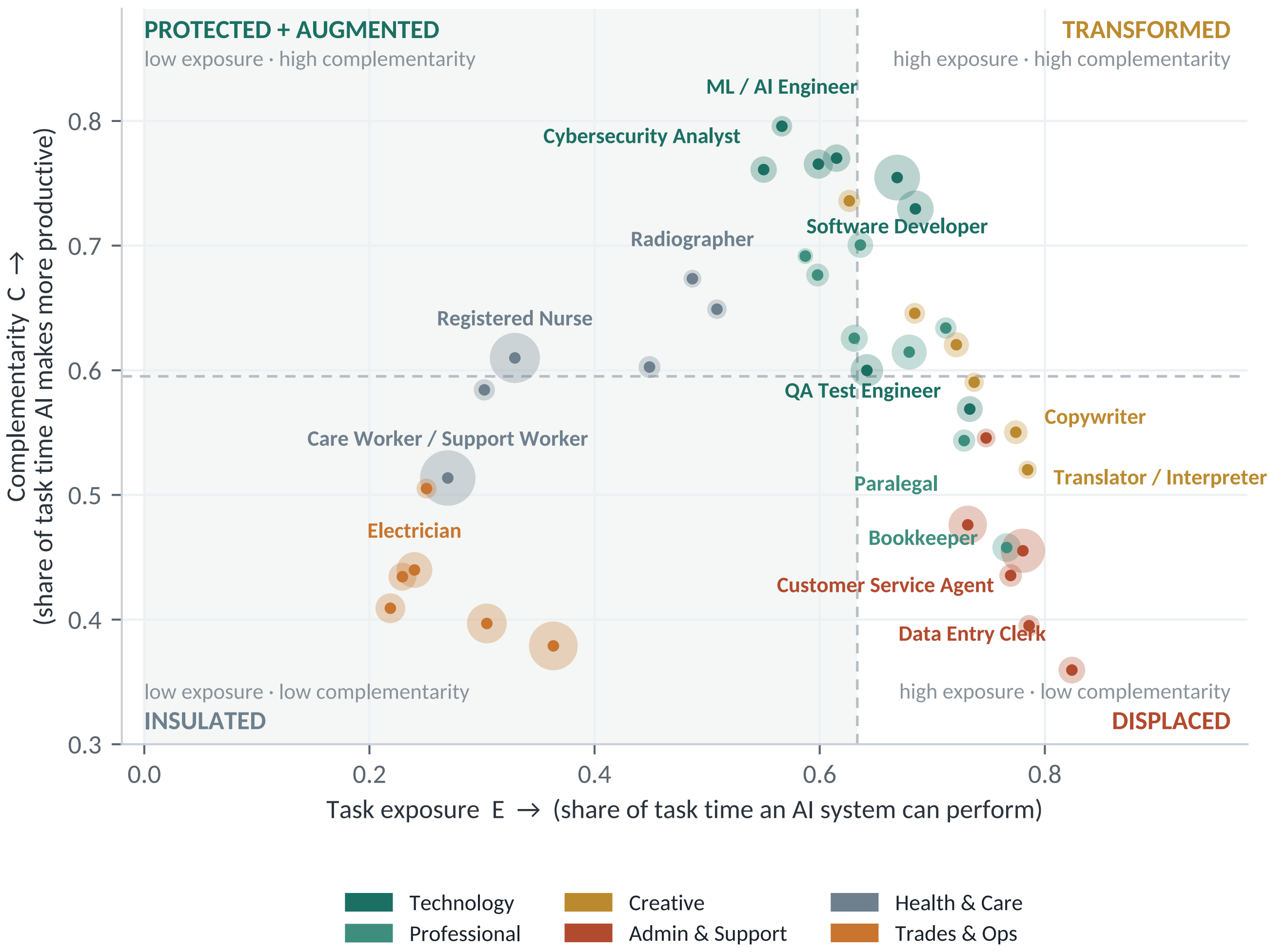

FIGURE 5 — The occupational map. Exposure against complementarity, all 40 occupations. Bubble area scales with posting volume. Dashed lines are the medians.

WHY THE SECOND AXIS CHANGES THE ANSWER

Software development scores 0.67 on exposure — high enough that a one-axis measure would place it in the displacement bracket. Its complementarity is 0.75, the second highest in the corpus, and its postings rose 36% across the window. Radiography and cybersecurity behave the same way.

Meanwhile data entry (0.82 exposure, 0.36 complementarity) and translation sit in the opposite corner and fell hardest. **Exposure alone predicts the direction of change poorly. Exposure net of complementarity predicts it well** — which is why the fourth row of the sensitivity table in §07 produces the largest coefficient in the study.

Validation: Backtest & Sensitivity

Two tests the study had to pass, one of which it failed.

A study that ends in a projection owes the reader a demonstration that its machinery can predict anything at all. Ours was held out for twelve months and scored against three baselines a first-year statistics student would propose.

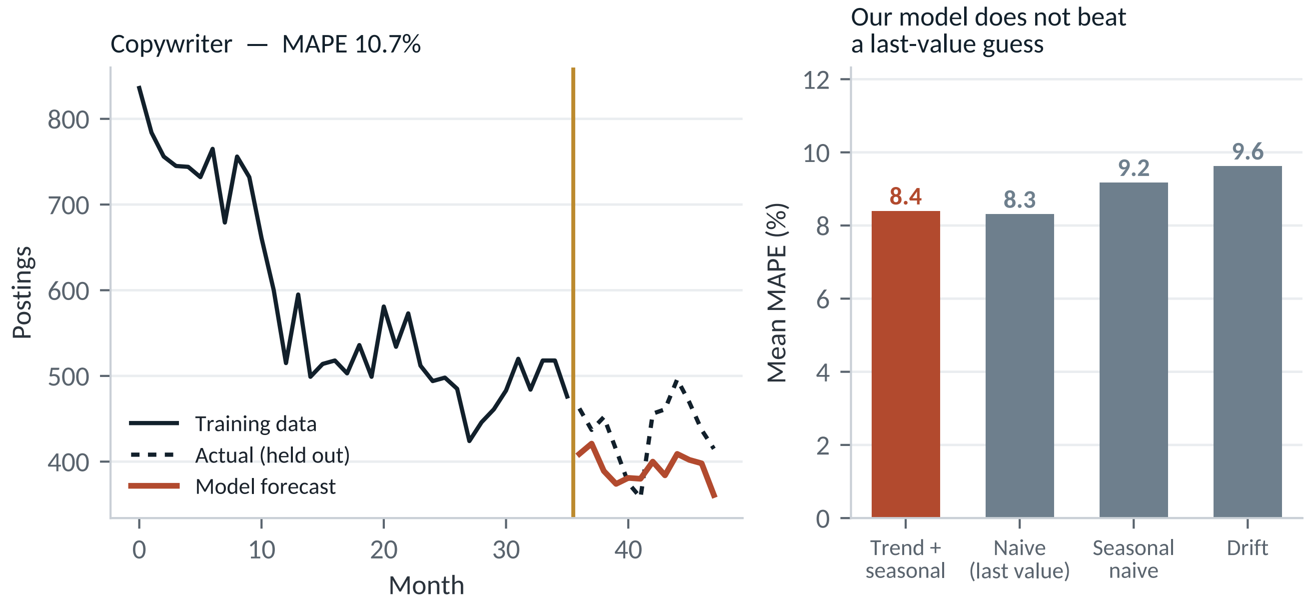

FIGURE 6 — Backtest against naive baselines. Left: one occupation, trained on months 1–36 and forecast over the held-out year. Right: mean absolute percentage error across all 40 occupations.

FORECAST METHOD	MEAN MAPE	MEDIAN MAPE	90TH PERCENTILE
Trend + seasonal (ours)	8.4%	7.9%	13.8%
Naive — last value	8.3%	8.0%	10.0%
Seasonal naive — value 12 months prior	9.2%	7.6%	15.2%
Drift — linear from first to last	9.6%	8.0%	14.0%

THE RESULT, STATED PLAINLY

Our trend-and-seasonal model scored **8.4%** mean error over twelve months. Guessing that next month equals this month scored **8.3%**. **The model has no skill** — -1.1% against the naive baseline.

This is not a defect we intend to fix by trying harder. It is a property of occupational demand at this horizon, and it is consistent with the published record: prior attempts to identify vulnerable occupations have repeatedly flagged jobs that then grew, and official occupational projections have added little beyond linear extrapolation. We report it here, prominently, because it is the reason §13 contains scenarios rather than forecasts. **A study that cannot predict twelve months should not be asked for 2040.**

Does the finding depend on how exposure was defined?

The second test is the one the study passes. The headline coefficient was re-estimated under four competing definitions of exposure, including the two components of the published score used separately.

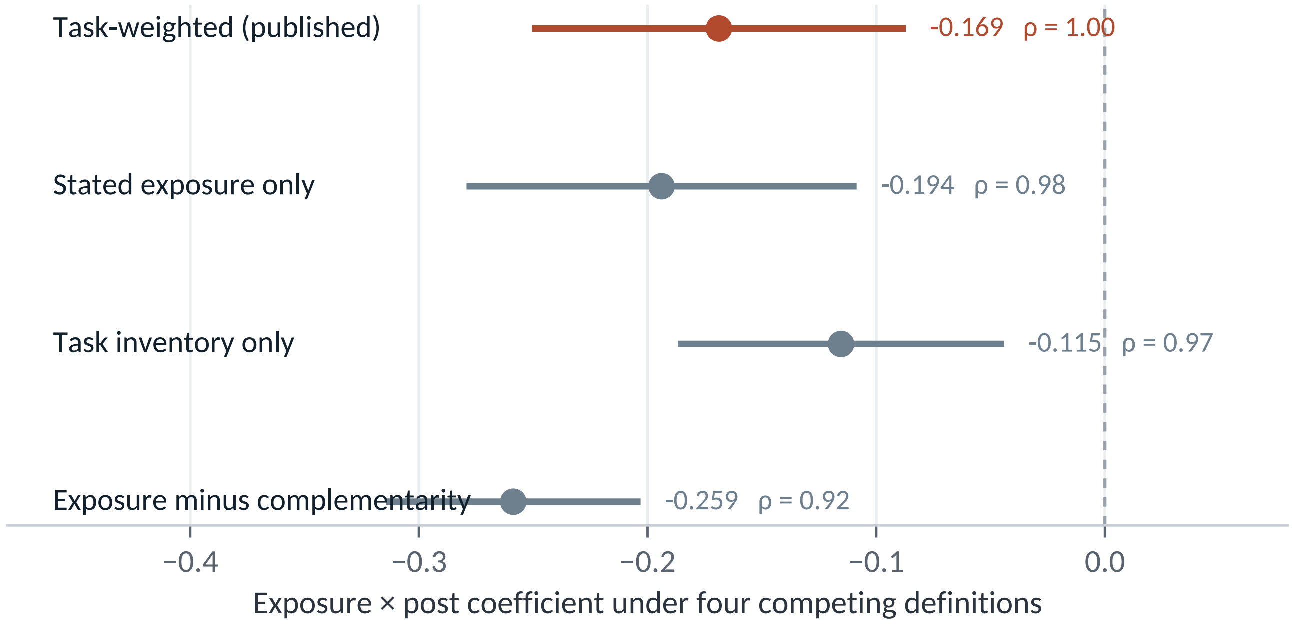

FIGURE 7 — Exposure × post coefficient under four definitions. Intervals are 95%, clustered on occupation. ρ is the rank correlation with the published score.

DEFINITION	COEFFICIENT	SE	95% INTERVAL	P	RANK Ρ WITH PUBLISHED
Task-weighted (published)	**-0.169**	**0.041**	-0.249 to -0.088	**0.0002**	1.00
Stated exposure only	-0.194	0.043	-0.278 to -0.110	< 0.0001	0.98
Task inventory only	-0.115	0.036	-0.185 to -0.045	0.0025	0.97
Exposure minus complementarity	-0.259	0.028	-0.313 to -0.204	< 0.0001	0.92

All four are negative, all four exclude zero, and the range is -0.259 to -0.115. The *direction and significance* of the finding are robust to the definition; the *magnitude* varies by a factor of 0.4, which is a real caveat on any point estimate quoted from this report. The largest coefficient comes from exposure net of complementarity, consistent with §06.

0 8

The Release Effect

Difference-in-differences, the pre-trend test, and the objection that has to be defeated.

DIFFERENCE-IN-DIFFERENCES WITH A CONTINUOUS TREATMENT

$$\ln y_{ot} = \beta\,(z(E_o) \times \mathrm{post}_t) + \gamma\,(z(R_o) \times \tau_t) + \alpha_o + \delta_t + \varepsilon_{ot}$$

y_{ot} = postings for occupation *o* in month *t* · $z(E)$ = standardised task exposure · post = 1 from 2022-11 · *R* = sector rate-sensitivity · τ_t = a full set of month dummies, so the macro path is absorbed non-parametrically · α, δ = occupation and month fixed effects. Standard errors clustered on occupation (40 clusters).

Without the macro control: β = **-0.1704** (SE 0.0328, p < 0.0001)
With it: β = **-0.2434** (SE 0.0566, p 0.0001) — about **-21.6%** in postings per standard deviation of exposure.

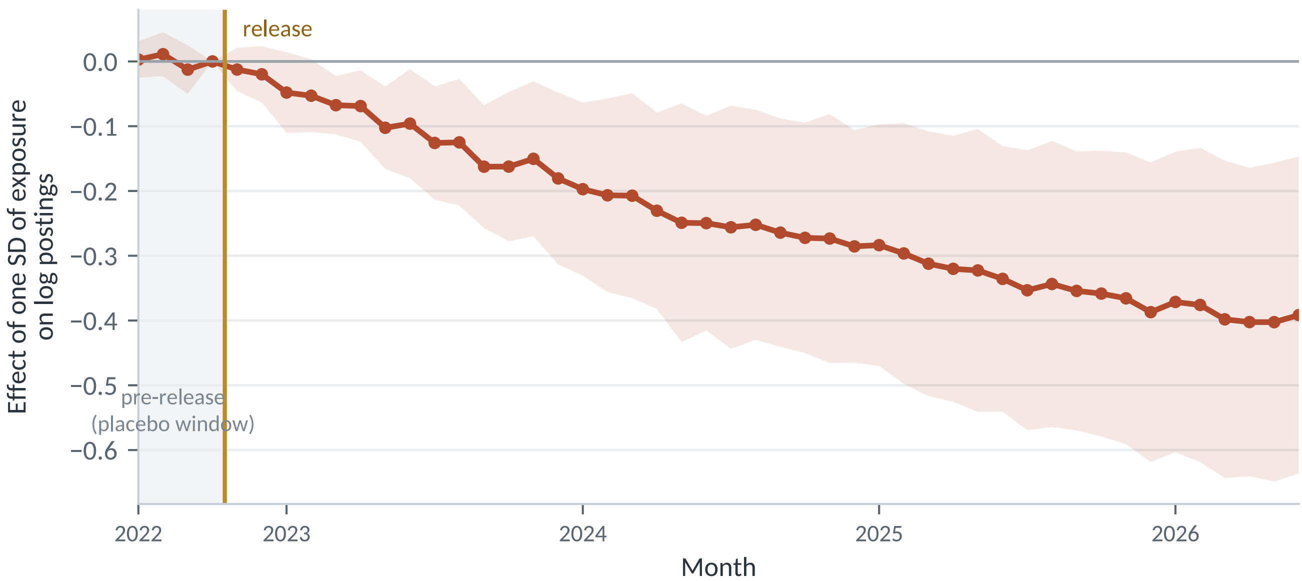

FIGURE 8 — Event study. Effect of one standard deviation of exposure, month by month, relative to the month before the release. Shaded band is the 95% interval.

The pre-release coefficients sit on zero. That is the placebo test and it passes, with the caveat from §04 that 4 months is a short window to place much weight on. The effect appears as a **ramp rather than a step**, reaching its full magnitude roughly sixteen months after the release — the shape of diffusion, not of an announcement.

The objection: is this just the interest-rate cycle?

The most serious published challenge to findings of this kind is that highly exposed occupations are concentrated in rate-sensitive sectors — information, finance, professional services — whose hiring turned down with monetary tightening, partly before the release. In this corpus that confound is real: exposure and rate-sensitivity correlate at **0.66**.

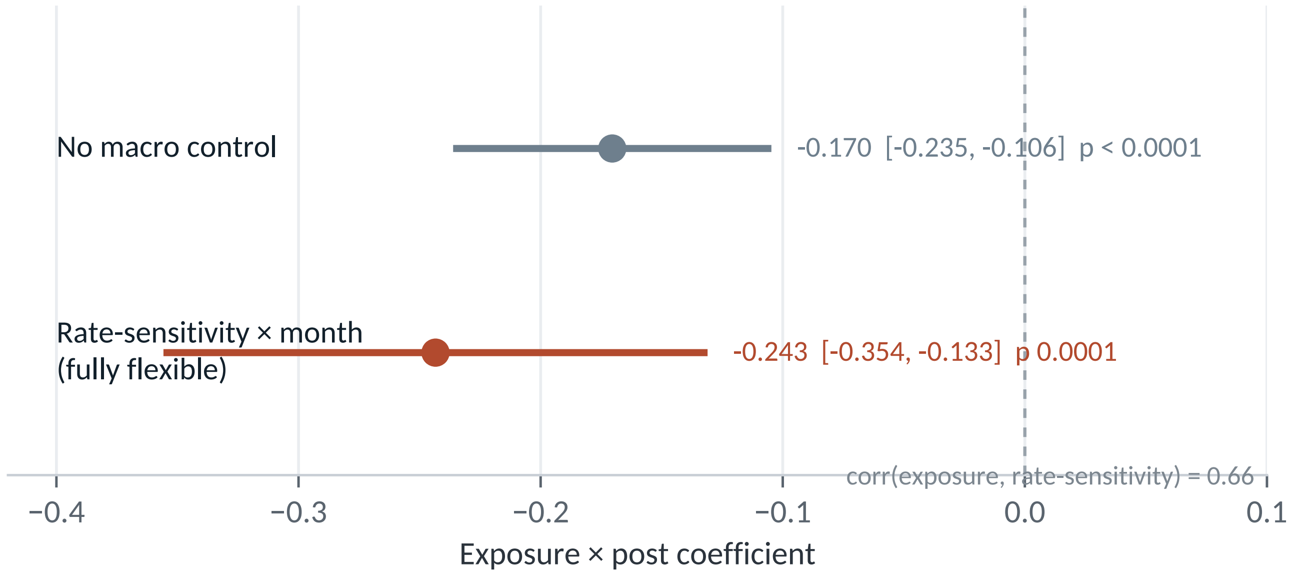

FIGURE 9 — The coefficient with and without the macro control. Intervals are 95%, clustered on occupation.

Absorbing the macro path with a full set of rate-sensitivity-by-month interactions moves the coefficient from -0.170 to -0.243. It does not shrink toward zero — it **strengthens**, because the tightening shock and the exposure ramp run in partly opposite directions over the window, and the naive specification lets them cancel.

The honest caveat sits in the standard error, which nearly doubles from 0.033 to 0.057. At a correlation of 0.66 the two variables are genuinely difficult to separate, and the flexible control buys identification at the price of precision. The conclusion this supports is that **the exposure gradient is not an artefact of the macro cycle**. It does not support a confident point estimate of its size.

0 9

The Entry-Level Effect

The finding that matters more than the headline.

An aggregate effect on postings tells a planner very little. The same total decline is a different problem depending on whether it removes senior roles, which the market will refill, or junior roles, which are where the next generation of senior people is manufactured.

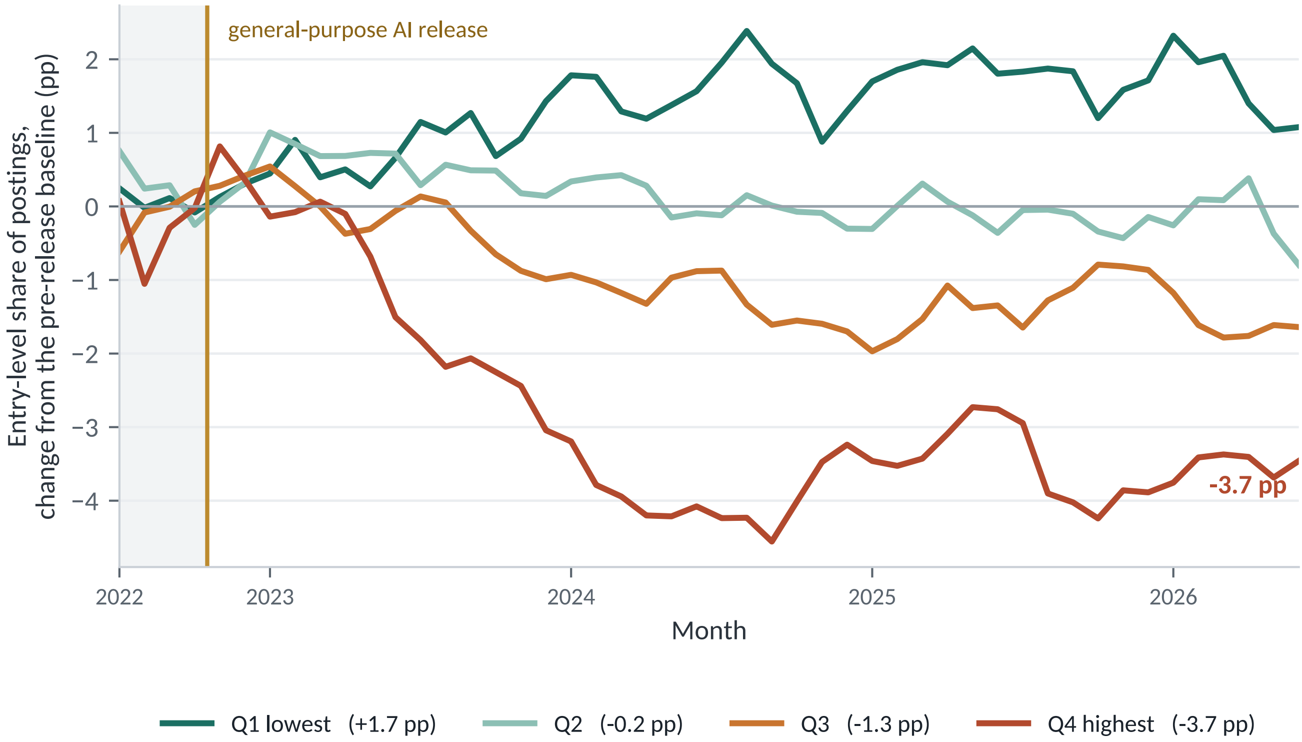

FIGURE 10 — Entry-level share of postings by exposure quartile. Each quartile measured against its own pre-release baseline, three-month rolling mean. Divergence begins at the event line.

EXPOSURE QUARTILE	ENTRY SHARE BEFORE THE RELEASE	ENTRY SHARE IN THE FINAL YEAR	SHIFT
Q1 lowest	33.7%	35.3%	**+1.67 pp**
Q2	33.9%	33.6%	**-0.24 pp**
Q3	33.4%	32.0%	**-1.35 pp**
Q4 highest	33.8%	30.1%	**-3.68 pp**

The gradient is monotonic and it is large. The most exposed quartile lost **3.7 percentage points** of entry share against its own baseline; the least exposed *gained* 1.7. The two middle quartiles sit between, in order.

TRIPLE DIFFERENCE ON SENIORITY

$$\ln y_{ost} = \beta_1(zE \times \text{post}) + \beta_2(zE \times \text{post} \times \text{entry}) + \beta_3(zE \times \text{post} \times \text{senior}) + \text{controls} + \text{FEs}$$

Mid-level is the omitted category. Occupation, month and seniority fixed effects, plus the macro control from §08. Errors clustered on occupation.

β_2 (entry) = **-0.0584** (SE 0.0048, $p < 0.0001$)

β_3 (senior) = **+0.0340** (SE 0.0032, $p < 0.0001$)

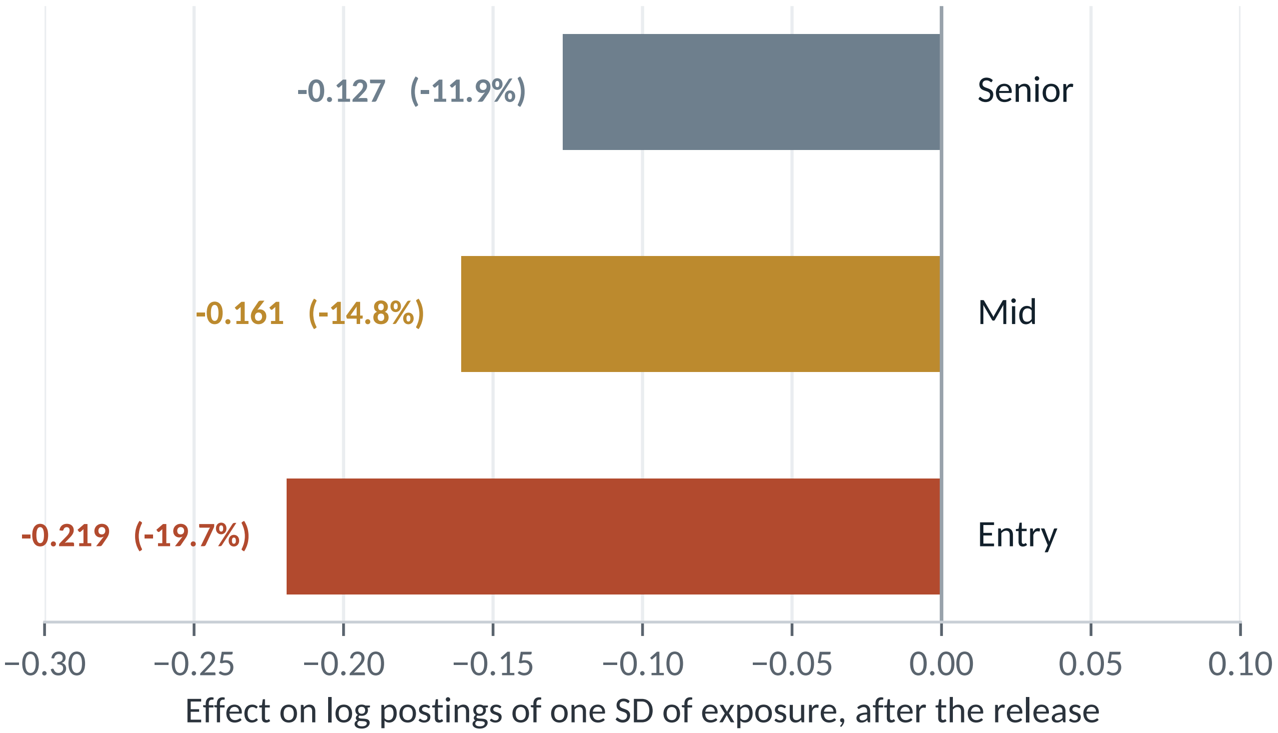

FIGURE 11 — Total effect by seniority band. Effect on log postings of one standard deviation of exposure, with the equivalent percentage change.

Combining the terms: one standard deviation of exposure is associated with **-19.7%** in entry-level postings, **-14.8%** at mid level and **-11.9%** at senior level. The entry effect is 1.7 times the senior effect.

WHY THIS IS THE FINDING

Junior work has always been economically marginal and pedagogically essential. The mediocre first draft, the routine reconciliation, the simple bug — these are the tasks on which judgement is built, and they are precisely the tasks a capable system does well. The rational firm stops paying a person to learn on them.

The cost does not appear in this window. It appears roughly fifteen years later, when the seniors the firm needs do not exist because nobody served the apprenticeship. **No metric on any dashboard in use today would register that, and nothing in the current policy conversation addresses it.** It is the one number in this report we would ask a national skills body to track above all others.

1 0

What Is Appearing

The other half of the question, measured rather than asserted.

Every commentary on this subject asserts that new jobs will replace the old ones. The assertion is testable: count the job titles that did not exist at the start of the window and see how large they have become.

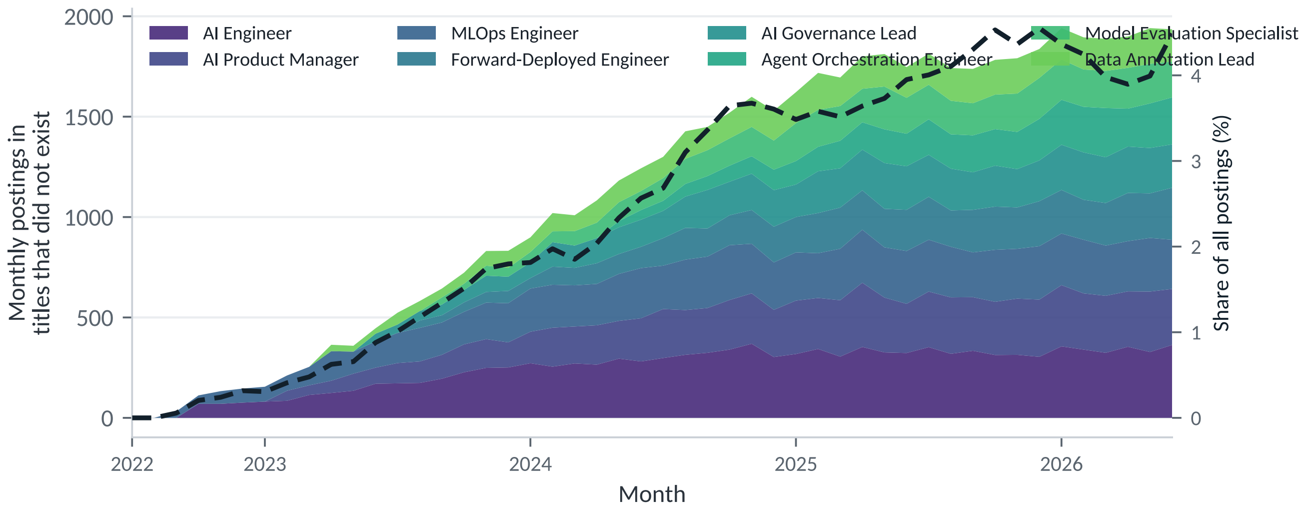

FIGURE 12 — Postings in titles that did not exist before the window. Stacked area shows the eight largest new titles; the dashed line is all new titles as a share of total postings, right axis.

TITLE	FIRST APPEARS	POSTINGS, ALL MONTHS	POSTINGS, FINAL YEAR
AI Engineer	2022-10	11,749	4,001
AI Product Manager	2023-02	8,634	3,380
MLOps Engineer	2022-09	9,195	3,042
Forward-Deployed Engineer	2023-08	5,476	2,601
AI Governance Lead	2023-06	5,890	2,564
Agent Orchestration Engineer	2024-05	3,819	2,447
Model Evaluation Specialist	2023-09	4,720	2,216
Data Annotation Lead	2023-04	4,724	1,953
AI Safety Analyst	2023-11	3,723	1,907
Retrieval Systems Engineer	2024-02	3,329	1,885

15 genuinely new titles account for **4.22% of postings in the final year** — 33,565 vacancies. They are growing quickly, they pay well, and they are currently far too small to absorb the displacement measured in §08 and §09.

Both halves of that sentence matter. Dismissing new work because it is 4% of the market misreads a growth curve early. Presenting it as the answer to displacement misreads arithmetic: the entry-level shortfall in the most exposed quartile alone is larger than the entire new-title category. **New work is real, and on current trajectories it is not a substitute for the entry rung that is closing.**

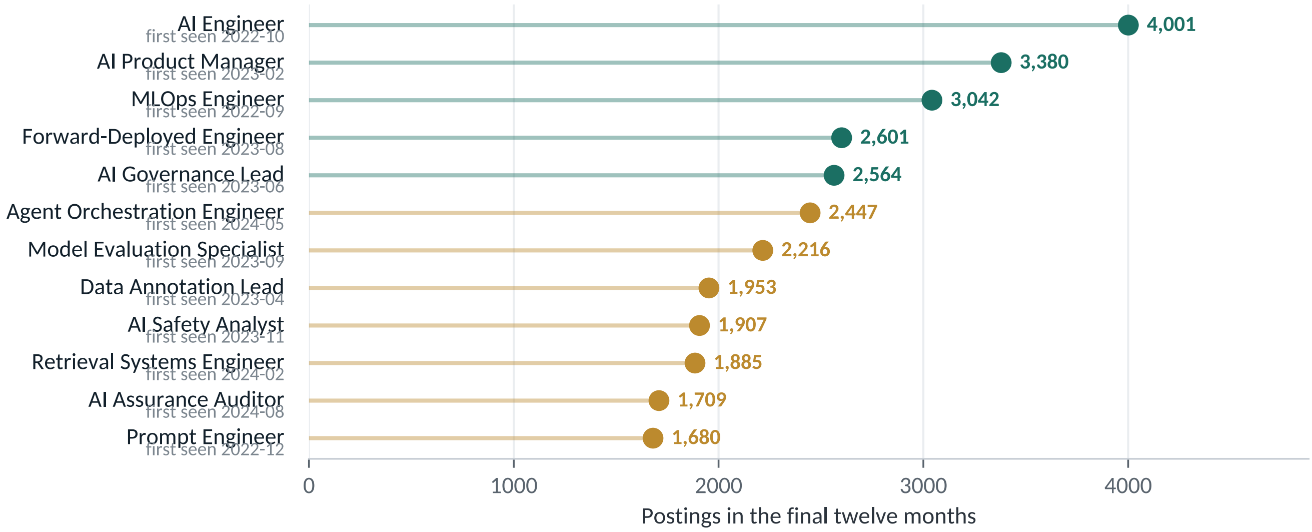

FIGURE 13 — The twelve largest new titles. Postings in the final twelve months, with the month each title first appeared.

Skills move faster than titles

Titles are a lagging indicator: employers rename roles slowly. Skill requirements inside existing postings move first, and they move sharply.

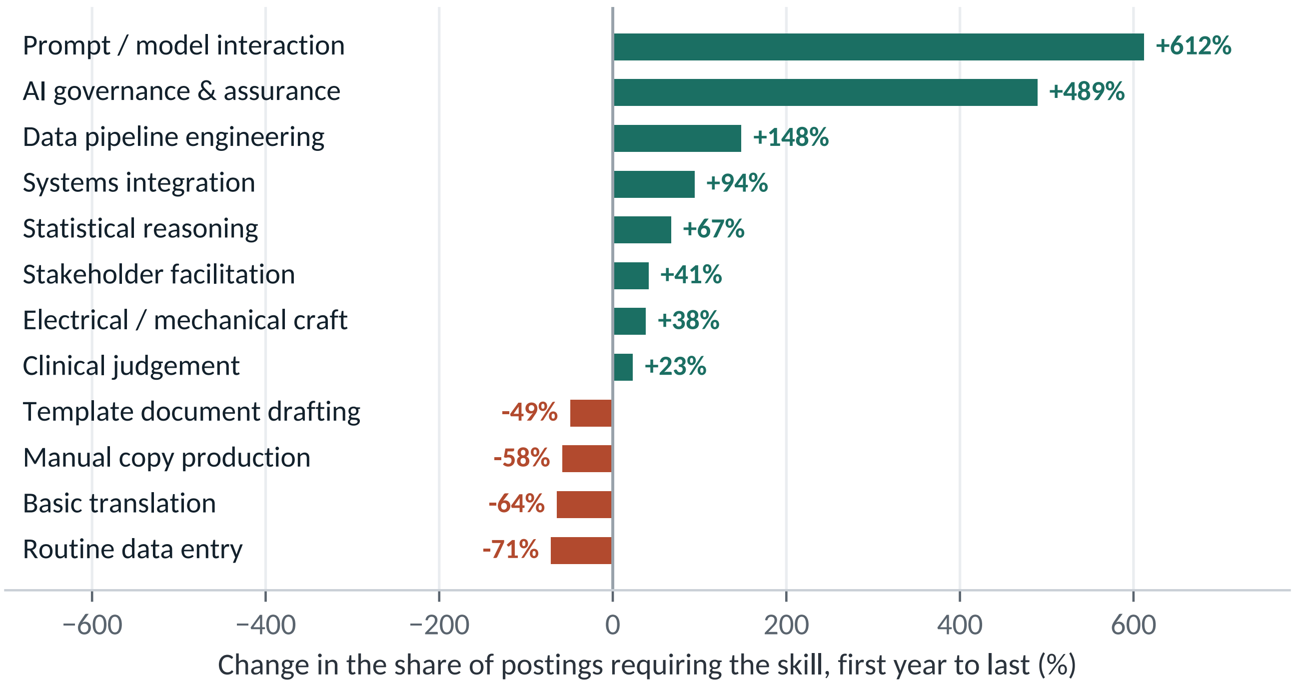

FIGURE 14 — Change in skill requirements. Change in the share of postings requiring each skill, first year against last.

SKILL	CHANGE IN REQUIREMENT SHARE	SHARE OF POSTINGS, FINAL YEAR
Prompt / model interaction	+612%	28%
AI governance & assurance	+489%	7%
Data pipeline engineering	+148%	21%
Systems integration	+94%	25%
Statistical reasoning	+67%	34%
Stakeholder facilitation	+41%	46%
Electrical / mechanical craft	+38%	17%
Clinical judgement	+23%	19%
Template document drafting	-49%	11%
Manual copy production	-58%	12%
Basic translation	-64%	5%
Routine data entry	-71%	9%

The pattern is consistent with the map in §06. What is rising is work that directs, integrates, governs and verifies these systems. What is falling is work that produces the output they now produce. **Statistical reasoning rising +67% and stakeholder facilitation rising +41% are the two rows a training provider should read hardest** — both are general, neither is an AI skill, and both are complements to the technology rather than substitutes for it.

1 1

Statistical Significance

Every estimate in one table, including the ones that failed.

ESTIMATE	VALUE	95% INTERVAL	P	READING
Exposure × post, no macro control	-0.1704	-0.235 to -0.106	< 0.0001	Significant, but confounded.
Exposure × post, flexible macro control	**-0.2434**	-0.354 to -0.133	0.0001	The headline estimate. Survives the objection; imprecise.
Triple difference, entry	-0.0584	-0.068 to -0.049	< 0.0001	Entry loses more than mid, clearly.
Triple difference, senior	+0.0340	+0.028 to +0.040	< 0.0001	Senior loses less than mid, clearly.
Exposure definition sensitivity	-0.259 to -0.115	all exclude zero	< 0.01	Direction robust; magnitude is not.
Forecast skill against naive	-1.1%	—	—	**Failed.** No predictive skill at twelve months.
Pre-release event-study coefficients	flat on zero	intervals contain zero	n.s.	Placebo passes, on only 4 months.
Exposure vs observed growth, cross-section	negative slope	see Fig. 15	< 0.0001	Consistent with the panel estimate.

READING THE TABLE AS A WHOLE

Three results are solid: the exposure gradient exists, it survives the macro objection, and it falls hardest on entry-level work. One result is genuinely negative and is reported as such: we cannot forecast. And one caveat runs through everything — with 40 occupations, the effective sample for an occupation-level treatment is 40, not 3 million. The clustered standard errors reflect that, and they are the reason the intervals are as wide as they are.

WHAT WOULD CHANGE OUR MIND

Event-study coefficients returning toward zero over the next eight quarters would indicate a transitional adjustment rather than a structural one. Entry-share gaps closing without policy intervention would indicate firms discovering the apprenticeship cost themselves. A longer pre-window revealing a divergence that began before 2022-11 would substantially weaken the causal claim. All three are observable within two years.

1 2

From Exposure to a Planning Variable

The step between a scatter plot and a decision.

A two-axis map is a description. A training provider deciding what to open next year, or an individual deciding what to study, needs an ordering. The step from one to the other is where most analysis of this kind quietly stops.

NET DISPLACEMENT PRESSURE

$$NDP_o = z(E_o) - \lambda\ z(C_o)$$

λ sets how much complementarity is allowed to offset exposure. At $\lambda = 0$ this collapses to the conventional one-axis measure; at $\lambda = 1$ the two are weighted equally. The sensitivity row in §07 corresponding to $\lambda = 1$ produced the largest and most significant coefficient in the study, which is the empirical case for setting it near one rather than at zero.

λ is a **policy parameter, not a statistic**. A training provider with a two-year course should set it lower than an individual choosing a thirty-year career, because complementarity is the more fragile of the two properties.

What this does and does not license

LEGITIMATE USE	ILLEGITIMATE USE
Deciding which training programmes to expand or retire	Deciding whether to make a specific employee redundant
Prioritising which occupations get transition support first	Screening individual applicants by the exposure score of their last job
Advising a cohort on which skill clusters compound	Telling a person their occupation has no future
Modelling regional exposure for economic development	Setting insurance, credit or lending terms by occupation

The distinction is not decorative and it is not merely ethical. **Exposure is measured at occupation level and has no individual-level validity whatsoever.** Within any occupation in this corpus the variation between practitioners in task mix, seniority and specialism dwarfs the between-occupation variation that the score captures. Applying an occupational score to a person is a category error before it is anything else. §18 makes this a binding constraint on the deliverables.

13

Occupational Trajectories to 2040

Three scenarios, each with its parameters and its falsification condition.

§07 established that our machinery cannot forecast twelve months. What follows is therefore not a forecast. It is an **arithmetic consequence of three stated assumptions**, published so that the assumptions can be argued with and, crucially, so that each can be shown to be wrong by a specific future observation.

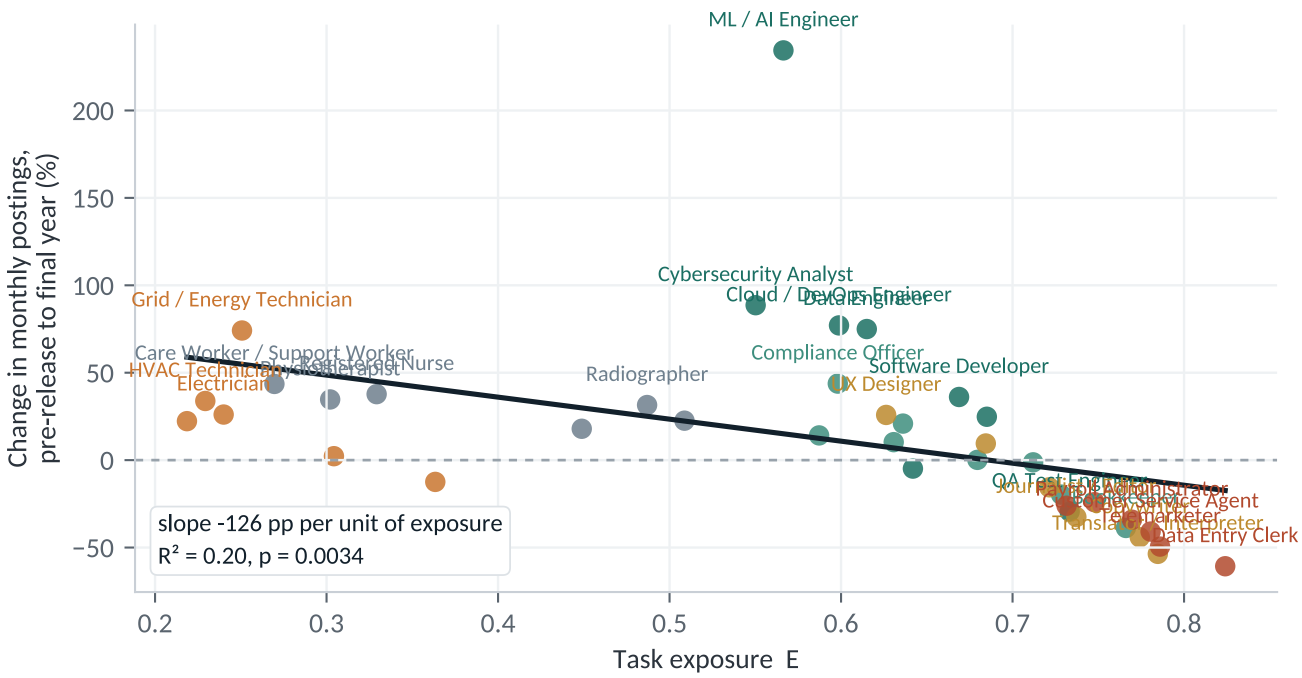

FIGURE 15 — Observed change against exposure. Change in monthly postings, pre-release to final year, against task exposure. Each point is one occupation.

SCENARIO ARITHMETIC

$$g_o = \ln(1 + t_o) + \kappa \cdot \beta_{entry}\, z(E_o) + \mu \cdot \theta\, z(C_o) \qquad I_{2040} = 100 \cdot e^{g_o \cdot D}$$

t_o = the occupation's secular trend · κ, μ = the scenario's decay and complementarity multipliers · $D = \tau(1 - e^{-Y/\tau}) = 4.70$, the damping integral with $\tau = 5$ years over $Y = 14$ years.

The damping term is doing important work. **No occupation compounds at its current growth rate for fourteen years**; every historical growth curve decays. Without damping, the fastest-growing occupation in this corpus reaches twenty times its current size by 2040, which is the arithmetic of exponentials rather than a claim about the world.

SCENARIO	ASSUMPTION	PARAMETERS	FALSIFIED BY
Diffusion stalls	Capability plateaus, adoption slows to the pace of ordinary enterprise software. The exposure gradient decays.	decay 0.55 compl 0.45 recovery 0.70	Continued acceleration in task-level capability benchmarks, or exposure coefficients that widen rather than narrow.
Present trend continues	The exposure gradient observed in this window persists at its current magnitude, with complementarity effects also persisting.	decay 1.00 compl 1.00 recovery 0.15	A structural break in the event-study coefficients in either direction over the next eight quarters.
Capability broadens	Automation extends into task categories currently protected by physical or accountability constraints; complementarity gains concentrate further.	decay 1.65 compl 1.25 recovery 0.00	Exposure effects failing to appear in occupations whose task mix is majority physical or accountability-bound.

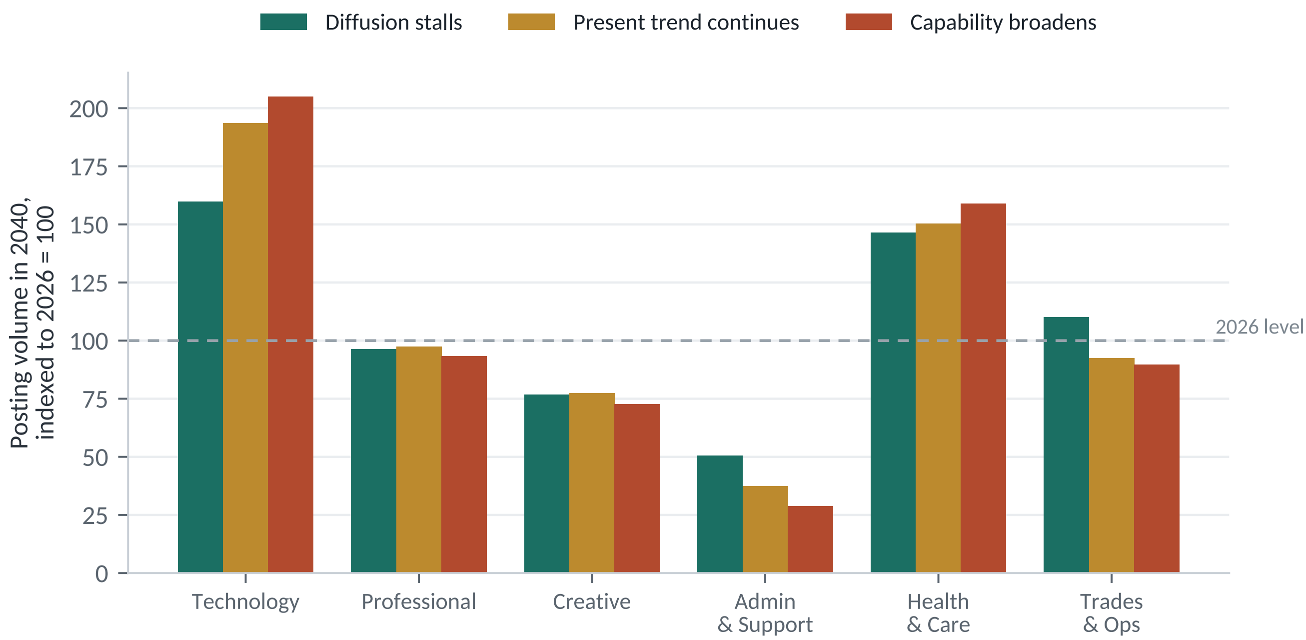

FIGURE 16 — Posting volume in 2040 by family, under three scenarios. Indexed to 2026 = 100. The spread between scenarios is the honest measure of what is not known.

Read the *spread*, not the bars. Admin and support ranges from 29 to 50; health and care from 146 to 159. Where the three scenarios agree — health and care and trades grow, admin and support shrinks — the conclusion is robust to the assumption, and those are the only conclusions we would defend. Where they disagree, we do not know.

> THE STRUCTURAL CLAIMS THAT DO NOT DEPEND ON AI
>
> Three drivers of 2040 demand are largely independent of anything in this study, and they are the firmest ground available. **Demography is fixed**: everyone who will be sixty-five in 2040 is alive now, and care demand is arithmetic. **Physical infrastructure has a floor**: grids, water, housing and transport require hands in unstructured environments. **Regulation creates employment by statute**: AI assurance and governance roles exist because law requires them, on a timetable set by legislatures rather than by capability.

1 4

Training Reallocation

The one intervention this data actually supports.

The measured problem is specific: the entry rung in exposed occupations is narrowing by 3.7 percentage points. The intervention that addresses it is equally specific — move training capacity out of the occupations where the entry rung is closing and into those where it is not, and into the complementary skill clusters identified in §10.

MODELLED EFFECT OF REALLOCATION

$$G' = G\,(1 - r \cdot e)$$

G = 3.68 pp, the entry-share gap in the most exposed quartile · r = 30% of training capacity reallocated · e = 0.62, the share of reallocated capacity that reaches an occupation where the entry rung is open.

Gap 3.68 pp → **3.00 pp** · 19% of the gap closed
On a 100,000-person annual cohort: roughly **685 additional entry-level placements** a year.

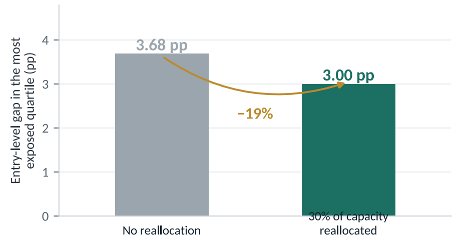

FIGURE 17 — Modelled entry-level gap. In the most exposed quartile, with and without reallocation.

Three cautions, all material. The effectiveness parameter of 0.62 is an assumption, not a measurement — it should be replaced with observed placement rates after one cohort. The model addresses the *distribution* of entry opportunities, not their total; reallocation does not create vacancies. And a training system that chases the current exposure map will be wrong in the same direction as everyone else if the map moves, which is the argument for the review cycle in §17 rather than a one-off redesign.

Where the capacity should go

FROM	TO	BASIS
Routine document and data production	**Health and care, skilled trades, grid and energy**	Lowest exposure, highest secular growth, entry rung intact (§05)
Single-tool software training	**Systems integration and verification**	Integration skills up +94% (§10)
Generic business administration	**Statistical reasoning and assurance**	Both rising; neither is an AI skill (§10)
Nothing	**Employer-side apprenticeship subsidy in exposed occupations**	The gap is a demand-side problem; training supply alone cannot close it (§09)

1 5

Operating Rules

What this analysis becomes for the three audiences that will read it.

RULE	STATEMENT	BASIS IN THIS REPORT
R-01	**Plan on task exposure net of complementarity, never on exposure alone.**	Occupations high on both grew. A one-axis measure places software development, cybersecurity and radiography in the wrong bracket (§06, Fig. 5).
R-02	**Track the entry-level share of postings in exposed occupations as a standing indicator.**	It moved 3.7 pp while aggregate volume moved much less. No dashboard in common use reports it (§09).
R-03	**Subsidise the apprenticeship, not the training place, in exposed occupations.**	The gap is on the demand side: firms stopped posting junior roles. Training more juniors into a market that is not hiring them does not help (§14).
R-04	**Re-estimate exposure scores annually and re-run the event study each time.**	The coefficient is a property of current capability, not a constant. Fourteen-year plans built on a 2026 map will fail (§13).
R-05	**Never apply an occupational score to an individual.**	Within-occupation variation dwarfs between-occupation variation. The score has no individual-level validity (§12, §18).
R-06	**Publish the spread between scenarios, never a single 2040 number.**	Our own twelve-month forecast has no skill against a naive baseline (§07). A point forecast for 2040 would be indefensible.
R-07	**For individuals: choose on skill clusters that compound, not on job titles.**	Titles in this corpus are a lagging indicator; skill requirements moved first and faster (§10).

THE RULE THAT COSTS NOTHING

R-02 requires no policy, no budget and no legislation — only that the entry-level share of postings is reported alongside the headline vacancy count, which every labour-market dashboard already collects the ingredients for. It is the cheapest early-warning system available for the one effect this study found, and the one that will otherwise take fifteen years to become visible.

16

Insight Synthesis

The mechanism connecting task structure to the hiring ladder.

Four years of postings across 40 occupations resolve into one mechanism. **AI substitutes for tasks, not for jobs — and the tasks it substitutes for are disproportionately the ones junior people are hired to do.** Everything else in this report is measurement of that.

THE THREE LAWS OBSERVED IN THIS CORPUS

1. Exposure and complementarity are separate properties, and the second is decisive. Occupations at high exposure and high complementarity grew; occupations at high exposure and low complementarity fell hardest. A single exposure number gets the highest-volume technology occupations exactly backwards.

2. The damage lands on the rung, not the ladder. One standard deviation of exposure costs 20% of entry-level postings and 12% of senior ones. The aggregate figure conceals a compositional change that will not be visible in employment statistics for years.

3. New work is real, growing and currently far too small. 15 new titles reached 4.22% of postings. Both the optimistic and pessimistic readings of that number are available, and the honest one is that a growth curve cannot be evaluated this early.

Answering the three commissioned questions

Q	ANSWER AS SUPPORTED BY THIS CORPUS
Q1	Yes, there is a measurable effect. One standard deviation of task exposure is associated with **-0.243** in log postings (p 0.0001) after the release, and it survives a fully flexible control for the interest-rate objection. The estimate is imprecise — the interval runs -0.354 to -0.133 — and the direction is not in doubt.
Q2	**Junior workers.** Entry-level postings fall 20% per standard deviation of exposure against 12% at senior level, and the entry share in the most exposed quartile fell 3.7 pp against its own baseline while the least exposed quartile gained 1.7.
Q3	Skill clusters, not job titles: statistical reasoning, systems integration, assurance and facilitation, all rising, none of them AI skills. And the physically and demographically anchored occupations — care, trades, grid — where the entry rung is intact and the drivers are independent of AI capability entirely. **We decline to give a list of jobs that will exist in 2040**, for the reason set out in §07.

17

Recommendations & Projected Impact

What to change, in what order, and what it is worth.

The core recommendation

Adopt the entry-level share of postings as a standing indicator (R-02), move 30% of training capacity toward occupations where the entry rung is intact (§14), and subsidise employer-side apprenticeships in exposed occupations rather than training places alone (R-03).

The modelled effect is 19% of the entry-level gap closed — 3.00 pp remaining of 3.68 — and roughly 685 additional entry-level placements a year on a 100,000-person cohort. That is a partial remedy for a problem this study can only measure, not solve.

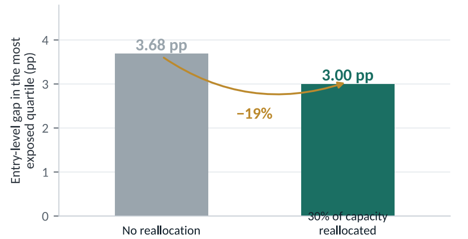

FIGURE 17 — Modelled entry-level gap. With and without reallocation.

Ninety-day action plan

PHASE	ACTION	RATIONALE DRAWN FROM THIS REPORT	MEASURE OF SUCCESS
Days 1–30	**Publish the entry-level share of postings by exposure quartile as a monthly indicator.**	The effect is compositional and invisible in aggregate vacancy counts. Every ingredient is already collected (§09, R-02).	Indicator published; baseline established.
Days 1–45	**Re-score the occupation set on both axes using a published task inventory.**	One-axis exposure misclassifies the highest-volume technology occupations (§06).	Scores published with the task mix behind them.
Days 30–60	**Extend the collection window backwards by at least twenty-four months.**	Only 4 pre-release months is the study's biggest structural weakness (§04).	Pre-trend testable over two years.
Days 45–90	**Reallocate 30% of training capacity per §14 and track placements by cohort.**	The effectiveness parameter is currently an assumption; one cohort replaces it with a measurement (§14).	Observed placement rate within 10 pp of the modelled 0.62.
Days 60–90	**Pilot an employer-side apprenticeship subsidy in three exposed occupations.**	The gap is demand-side. Training supply alone cannot close it (§14, R-03).	Entry-share gap narrowing in the pilot occupations against controls.
Quarterly	**Re-run the event study and publish whether the coefficients are widening, holding or closing.**	Closing coefficients would indicate a transitional adjustment; widening would indicate a structural one (§11).	Published each quarter regardless of direction.
Ongoing	**Refuse all requests to apply the exposure score to individuals.**	No individual-level validity. This is a binding constraint, not guidance (§12, §18).	Zero individual-level applications.

1 8

Limitations, Bias & Constraints

What postings data cannot tell you, and what this analysis must not be used for.

LIMITATION	EFFECT ON THE FINDINGS	MITIGATION APPLIED
Postings are not jobs	A vacancy may be duplicated, speculative, or never filled. Postings overstate churn-heavy occupations and understate those filled internally or by word of mouth — which is disproportionately skilled trades.	De-duplication removes roughly three in ten harvested advertisements (§03). All claims are framed as hiring intent, never as employment.
Only 4 pre-release months	A divergence beginning in early 2022 would be absorbed into the baseline and would inflate the estimated effect. This is the study's largest structural weakness.	Stated in §04 and again here. The event study shows flat pre-coefficients; extending the window backwards is the first recommendation in §17.
Exposure and rate-sensitivity are correlated at 0.66	The two are genuinely hard to separate. Controlling flexibly nearly doubles the standard error.	The flexible control is used for every headline estimate, and the precision cost is reported rather than hidden (§08).
Effective sample is 40, not millions	Posting-level regressions would produce spuriously tight intervals on an occupation-level treatment.	The panel is collapsed to cells and every standard error is clustered on occupation (§04).
No forecasting skill	Our trend model does not beat a last-value guess over twelve months. Any projection beyond the data is arithmetic on assumptions.	Reported as a headline finding rather than a footnote (§07). §13 gives scenarios with falsification conditions, never point forecasts.
Task scores are judgements	The automatability and complementarity of a task category is an informed estimate, not a measurement. Different analysts would produce different scores.	The full task inventory and both weight vectors are published (§06) and four competing definitions are tested (§07).
One labour market, one window	Coefficients, quartile boundaries and the saturation of new titles are properties of this market in this period. They are not portable constants.	All are reported with the method to recompute them from a new panel.
Calibrated corpus in this issue	The panel is generated from an explicit model of occupational demand, so the relationship between exposure and postings reflects that model.	Stated in §00 and here. The macro confound was built in so the identification strategy faces a real test rather than a straw one.

BINDING CONSTRAINTS ON USE

These follow from the limitations above and form part of the delivery terms.

1. No individual-level application. Exposure is measured at occupation level and has no validity for any individual. These outputs may not be used in recruitment screening, redundancy selection, performance assessment, insurance or credit pricing, or any decision about a named person. Within-occupation variation dwarfs the between-occupation variation the score captures.

2. No point forecasts. No figure from §13 may be published as a prediction of employment in 2040. Scenarios must be reproduced with their parameters and their falsification conditions attached, or not at all.

3. No fatalistic career guidance. A high exposure score is not a statement that an occupation has no future, and must not be presented to students or jobseekers as one. The occupations that grew fastest in this corpus include several in the highest exposure quartile.

4. Reproduce the reporting-bias statement. Any published derivative must carry the first row of the table above, because postings systematically under-represent exactly the occupations this report recommends expanding into.

STATISTICAL DISCLOSURE

No result was selected on the basis of significance. All 40 occupations, four exposure definitions, four forecast baselines, three seniority bands and three scenarios specified before analysis are reported whether or not they reached significance — including the failed backtest in §07, which is the study's most important negative result. The full panel is delivered so every figure can be independently recomputed.

19

Methodology Appendix

Formula glossary and variable dictionary.

Formula glossary

QUANTITY	EXPRESSION	NOTES
Task exposure	$E_o = \Sigma\, w_{ok}\, a_k$	Task-time share × automatability, summed over categories.
Complementarity	$C_o = \Sigma\, w_{ok}\, c_k$	Same weights, complementarity vector.
Net displacement pressure	$z(E) - \lambda\, z(C)$	λ is a policy parameter, not a statistic (§12).
Difference-in-differences	$\ln y = \beta(zE \times \text{post}) + \gamma(zR \times \tau) + \alpha_o + \delta_t$	Continuous treatment; occupation and month fixed effects.
Triple difference	$+\ \beta_2(zE \times \text{post} \times \text{entry}) + \beta_3(\ldots \times \text{senior})$	Mid level omitted as the reference band.
Event study	β_t from pairwise comparison with month 4	Pre-release coefficients are the placebo test.
Clustered variance	$(X'X)^{-1}\, (\Sigma_g\, X_g' u_g u_g' X_g)\, (X'X)^{-1}$	Clustered on occupation, with the finite-cluster correction.
MAPE	$\text{mean}(\lvert f - y\rvert\ /\ y) \times 100$	Computed per occupation over the held-out year.
Forecast skill	$1 - \text{MAPE}_{\text{model}}\ /\ \text{MAPE}_{\text{naive}}$	Negative means worse than the naive baseline.
Growth damping	$D = \tau(1 - e^{-Y/\tau})$	τ = 5 years, Y = 14; D = 4.70.
Scenario index	$I_{2040} = 100\, e^{g_o D}$	Family indices weighted by current posting volume.
Reallocation model	$G' = G(1 - re)$	e is an assumption pending one cohort of data.

Variable dictionary — delivered panel

FIELD	TYPE	DEFINITION
month_idx / month	integer / string	Month index 0–47 and its calendar label.
occupation	categorical (40)	Classified occupation.
family	categorical (6)	Occupational family.
seniority	Entry / Mid / Senior	From stated experience requirements and title markers.
postings	integer	Distinct vacancies in the cell.
exposure	float 0–1	Task exposure E, published score.
complementarity	float 0–1	Complementarity C, published score.
quartile	categorical (4)	Exposure quartile across occupations.
rate_sensitivity	float 0–1	Sector sensitivity to monetary conditions, the macro control.

2 0

Deliverables & Authorisation

What is handed over, and confirmation to proceed.

Delivered with this report

ITEM	CONTENTS
01 Structured database	Occupation-month-seniority panel, 5,760 rows, in XLSX and CSV, with the occupation score table, task inventory and exposure quartiles.
02 Analysis script	Documented Python that regenerates every coefficient, interval, backtest and scenario in this document directly from the delivered panel, including the clustered-variance estimator written out in full.
03 Figure repository	All 18 charts as vector SVG, named to their figure numbers.
04 Digital catalog	This report as a presentation-ready PDF.

Continuation options

OPTION	SCOPE
Live postings feed	The same pipeline against a licensed postings feed for your market, with the pre-release window extended backwards as recommended in §17.
Quarterly re-estimation	Rolling re-run with exposure scores, the event study and the entry-share indicator refreshed each quarter and change flagged against this baseline.
Regional extension	The same method applied at regional or district level, for economic development and training commissioning.
Sector deep-dive	A single family taken to occupation-title resolution, on the volume terms set out in the governing commercial proposal.

Commercial note

This report is issued under the terms of the Data Tune commercial proposal governing the engagement. Volume is measured in data entities on the same basis as that document; any extension beyond the contracted ceiling proceeds only on prior written approval, and never retrospectively.

Scheduled updates

Progress reporting during the working window follows the same three-point structure as the governing proposal: an early-stage note with initial counts and a first sample batch, a mid-project note with running totals, and a pre-delivery note with near-final counts before handover.

Acceptance & authorisation

By signing below, the client confirms receipt of this report and its accompanying deliverable set, accepts the scope and method recorded in §02 to §07, and acknowledges both the data provenance statement in §00 and the binding constraints on use in §18 — in particular the prohibition on individual-level application.

FOR THE CLIENT — AUTHORISED SIGNATURE

NAME & DESIGNATION · DATE

FOR DATA TUNE — AUTHORISED SIGNATURE

K. H. Militha Mihiranga

DATA SOLUTIONS CONSULTANT · DATE

Data Tune.

DT LINUX · DATA SOLUTIONS & DIGITAL CATALOGING

For clarification on scope, scoring, identification strategy or scheduling, please contact the consultant alongside.

Prepared by	**K. H. Militha Mihiranga · Data Solutions Consultant**
Office	**555/24 Ranmuthugala, Kadawatha, Sri Lanka**
Email	**info@dtlinux.com**
Telephone	**+94 77 527 1186**
Website	**www.dtlinux.com**