AI and the Research Team: supplemental appendix*

Johan Fourie†

A Proofs

Throughout, $r = p_A/q$, $s = \sigma(a) = (1-\phi)G(a)$, and $\Delta = q - p_A$.

A.1 Lemma 1 (Span of control)

Assume $X > 0$. Any optimum uses all attention: if $\sum_i t_i < 1$, adding attention to a project with $E_i > 0$ raises output. Using all attention, $\sum_i y_i = C^{1-\beta} \sum_i t_i (E_i/t_i)^\beta$ over the projects with $t_i > 0$; projects with $E_i = 0$ optimally receive none. Strict concavity of $x \mapsto x^\beta$ and Jensen's inequality bound the sum by $C^{1-\beta} X^\beta$, with equality exactly when $E_i/t_i = X$ for every project with positive attention, that is, when $t_i = E_i/X$. At $X = 0$ output is zero under every allocation and the optimal allocation is not unique, which is why the lemma requires $X > 0$. □

A.2 Theorem 1 (At most one peak)

The limits follow from $\sigma \to 1-\phi$ and $c \to c_\infty$. Since $X^* = C(\beta/c)^{1/(1-\beta)}$ and $m^* = X^*(1-s)$ with $c = q - \Delta s$,

$$\frac{d \ln m^*}{ds} = \frac{\Delta}{(1-\beta)(q-\Delta s)} - \frac{1}{1-s} = \frac{\beta q - p_A - \beta \Delta s}{(1-\beta)(q-\Delta s)(1-s)}. \tag{1}$$

The denominator is positive on the finite-capability domain $s \in [0, 1-\phi)$: $q - \Delta s \geq q - \Delta = p_A > 0$ and $1 - s > 0$. (At $\phi = 0$ the endpoint $s = 1$ is approached only as $a \to \infty$.) The numerator is strictly decreasing in s, so the derivative changes sign at most once, from positive to negative. Because s is strictly increasing in a, m^* is quasi-concave in a, and adding the leader does not change the sign of the derivative of $N^* = 1 + m^*$. □

*This is the supplementary material for Fourie, Johan. 2026. "AI and the Research Team." Working Paper, Department of Economics, Stellenbosch University. This paper was created with the help of Anthropic's Claude Code (Opus 4.8 and Fable 5) and OpenAI's Codex (GPT-5.5 and GPT-5.6), and was checked for errors with refine.ink. Cite the paper, not this supplement.

†Department of Economics, Stellenbosch University. Email: johanf@sun.ac.za.

A.3 Corollary 1 (Peak coverage and the field threshold)

The numerator of (1) at $s = 0$ is $\beta q - p_A$, positive if and only if $r < \beta$. Its zero is

$$s^* = \frac{\beta q - p_A}{\beta(q - p_A)} = \frac{\beta - r}{\beta(1 - r)}.$$

An interior peak requires the numerator to be positive at $s = 0$ and negative at the endpoint $s = 1 - \phi$, that is, $s^* < 1 - \phi$. Rearranging,

$$\phi < 1 - s^* = 1 - \frac{\beta - r}{\beta(1 - r)} = \frac{\beta(1 - r) - \beta + r}{\beta(1 - r)} = \frac{r(1 - \beta)}{\beta(1 - r)} = \phi^*,$$

which also establishes the identity $s^* + \phi^* = 1$. If $r \geq \beta$, the numerator is nonpositive at $s = 0$ and negative for every $s > 0$, so m^* never rises. If $r < \beta$ but $\phi \geq \phi^*$, the numerator remains positive on $[0, 1 - \phi)$, so m^* rises to its long-run limit. □

A.4 The adoption-ceiling condition

Assume $r < \beta$. Let effective coverage for a fully codifiable team be a continuous, nondecreasing path $\widetilde{G}(t) \in [0, 1]$ with $\widetilde{G}(0) < s^*$ and limit L. If $L < s^*$, member employment is nondecreasing and never turns downward. If $L = s^*$, member employment approaches or reaches its maximum but does not subsequently decline. If $L > s^*$, continuity implies that the nonempty set $\{t : \widetilde{G}(t) = s^*\}$ is attained at a finite date and is the peak set. Strict monotonicity makes the peak unique; under weak monotonicity it may be a plateau. Both the initial condition and monotonicity are needed: a path that starts above s^* declines throughout, and a nonmonotone path can move employment in either direction. The paper's statement that a peak "typically requires near-complete long-run adoption" is a summary of the prior distribution of s^* (median 0.91; $s^* > 0.90$ in 58 percent of draws, with the 90 percent interval reaching down to 0.75), not a universal threshold.

A.5 The early-stage sign

Let $m_0 = C(\beta/q)^{1/(1-\beta)}$ denote member employment at $a = 0$. If $r < \beta$ and G is right-differentiable at zero, with $G'(0)$ the right derivative,

$$\left.\frac{d \ln N^*}{da}\right|_{a=0} = \frac{m_0}{1 + m_0} \frac{\beta q - p_A}{(1 - \beta)q} (1 - \phi) G'(0), \tag{2}$$

which is weakly decreasing in ϕ holding the other primitives fixed, and strictly decreasing when $G'(0) > 0$. The cross-field growth ordering is exact at $a = 0$ and extends to a neighborhood of zero by continuity. At higher capability it need not hold: changing ϕ also changes the automated share and the distance to the peak, so a lower-ϕ field can grow more slowly than a higher-ϕ field while both are still rising. At low capability, author counts should therefore grow faster in work with less physical exposure; this is the sign tested in the paper's researcher-level regression.

B Functional Form and Time-Varying Prices

Theorem 1 is a comparative static in capability at fixed task prices. Two extensions bound its scope. First, a declining relative AI cost r_t raises s^* and lowers $\phi^* = 1 - s^*$: cheaper AI moves positive-ϕ fields toward the monotone-growth regime. As r falls toward zero, every fixed $\phi > 0$ eventually loses its interior peak, while the fully codifiable type retains one for every positive r, with the peak converging to the boundary $s = 1$. Along a path where capability and prices move together, member employment need not be quasi-concave: the single-crossing argument holds prices fixed, and a sufficiently sharp price decline can generate additional turning points, exactly as with an endogenous member price. Second, additive task costs are not load-bearing for the long-run employment floor. If physical and codifiable execution combine with a finite elasticity of substitution η and the AI price is positive, physical member employment remains positive for every finite η; with free AI it vanishes only when η exceeds $1/(1-\beta)$. The single-peak shape itself does use the Cobb–Douglas execution demand and the linear cost aggregation.

C Monte Carlo Detail

The Monte Carlo (`code/montecarlo_peak.R`; 20,000 draws, all retained) propagates the priors in Table 1 through the peak condition. The peak occurs when coverage reaches s^*, so the peak horizon is $h_{\text{peak}} = h_{50}\, s^*/(1 - s^*)$ under $G(a) = a/(a + h_{50})$, and the peak date solves $\log_2 h_{\text{peak}} = \iota + \lambda\, t$ along the fitted METR trend. Trend uncertainty enters through a pairs bootstrap of the post-2023 observations: each draw refits the log-linear trend on a resample of the (date, horizon) pairs, giving intercept ι and slope λ jointly; draws with $\lambda \leq 0$ are excluded (none occurred in 20,000 draws). The implied doubling time has median 121 days with 90 percent interval $[104, 141]$. Formally, effective coverage is $\widetilde{G}_t = d_t\, G[a(t - \ell)]$ with adoption $d_t \in [0, 1]$ and lag $\ell \geq 0$; the Monte Carlo sets $d_t = 1$ and represents adoption delay entirely through ℓ, so the scenario-B peak date is the scenario-A date plus ℓ. Scenario B draws that lag uniformly: the lower bound of half a year is two quarters of team-formation and preprint lag, and the upper bound of two years reflects the survey evidence that regular coding-agent use among quantitative social scientists reached only 20 percent in early 2026, with the surge dated to late December 2025 (Lyttelton, Massenkoff and Wilmers, 2026). Scenario B therefore conditions on eventual full adoption: diffusion enters only as a delay. Scenario C instead draws a long-run adoption ceiling $d_\infty \sim U(0.8, 1.0)$, independently of the other parameters, and models effective coverage as a constant multiplicative ceiling from the outset, $\widetilde{G}_t = d_\infty G[a(t - \ell)]$. Under the priors $r < \beta$ always holds, so a peak occurs if and only if $s^* < d_\infty$; conditional on occurring, codifiable coverage at the peak is s^*/d_∞, dated on the same bootstrap trend and lag. (With adoption still rising at the crossing, this dating would be early; the constant-ceiling model is the stated benchmark.)

Results. Scenario A (capability clock): median peak 2027.2, 90 percent interval [2026.4, 2028.1]; the probability of a peak by end-2028 is greater than 0.99. Scenario B (diffusion-adjusted): median 2028.5, 90 percent interval [2027.4, 2029.6]; the probability of a peak by end-2028 is 0.77 and by end-

Table 1: Priors and scenario definitions.

Object	Prior	Notes
β	$U(0.4, 0.8)$	execution elasticity
$r = p_A/q$	$U(0.05, 0.20)$	AI-to-member task cost
h_{50}	log-uniform on [2, 24] hours	median codifiable task
METR trend (ι, λ)	pairs bootstrap, post-2023 fit	joint draw
Lag (scenarios B, C)	$U(0.5, 2.0)$ years	adoption and output
Ceiling d_∞ (scenario C)	$U(0.8, 1.0)$	long-run effective adoption

2030 is 1.00. Scenario C (adoption ceiling): the peak occurs in 50.4 percent of draws; conditional on occurrence, the median is 2028.6 with 90 percent interval [2027.5, 2030.2]. Because the peak condition is $s^* < d_\infty$, the probability of a peak at a fixed ceiling follows from the s^* prior alone: 0.11 at $d_\infty = 0.80$, 0.23 at 0.85, 0.42 at 0.90, 0.79 at 0.95, and 1 at complete adoption. Bridge-free objects: s^* has median 0.91 with 90 percent interval [0.75, 0.97], and $\phi^* = 1 - s^*$ has median 0.087 with 90 percent interval [0.028, 0.249]. In 58 percent of draws s^* exceeds 0.90, so near-complete effective coverage is typically, though not always, required for a peak. All dates are conditional on the assumed bridge from software-task horizons to research tasks; the bridge-free objects are not.

D Test Protocol

The dated test has two stages, separating the coverage condition from the team response.

Stage 1: coverage. Establish that effective coverage of codifiable research tasks has crossed s^*, using capability evaluations on long-horizon software and research tasks together with adoption surveys of research workflows. Under the maintained bridge, crossing corresponds to the capability-clock dates above; direct evaluation and adoption evidence takes precedence over the bridge if the two disagree. Crossing is established when measured effective coverage exceeds 0.75, the fifth percentile of the s^* prior, so that the absence of a peak is informative for at least 95 percent of parameter draws; a crossing of the median 0.91 strengthens the rejection.

Stage 2: team response. Re-pull the frozen cohort of 2,279 investigators at a single retrieval date in the first half of 2030, using records through the fourth quarter of 2029 and reporting the retrieval vintage. Reestimate equation (3), and compute the seasonally adjusted, exposure-weighted mean author count in the two zero-score fields, algebra and number theory and theoretical computer science. The peak criterion is an interior maximum followed by four consecutive quarters strictly below it, where a peak is detected only if the maximum exceeds the mean of the four following quarters by more than twice the clustered standard error of the difference. Symmetric noise therefore neither manufactures a peak nor converts an imprecisely estimated flat path into a decisive rejection; the rejection statement applies to precisely estimated flat or rising paths. If Stage 1 confirms crossing and Stage 2 finds no peak by this criterion, the joint hypothesis stated in the paper is rejected. If Stage 1 finds no crossing, the dated prediction is not yet in play, and

Table 2: Estimates behind Section 3 of the paper.

	Estimate	Std. error	t
Panel A: field level (standard errors clustered by 30 fields)			
Cross-field $\widehat{\phi}_f \times$ post-2022, baseline	0.069	0.033	2.09
with field-specific trends	−0.033	0.024	−1.41
with trends, excluding 2020–2022	−0.039	0.027	−1.43
median author-count variant	−0.035	0.124	−0.28
rank-proxy variant	−0.028	0.030	−0.93
occupation-proxy variant	−0.016	0.019	−0.82
Within-field computational × post-2022	0.041	0.029	1.44
excluding 2015	0.023	0.028	0.82
Panel B: researcher level (standard errors clustered by investigator)			
Exposure × capability, through December 2025	0.19	0.26	0.73
author counts capped at 25	0.21	0.24	0.88
through June 2026, server-location definition	−0.03	0.14	−0.21
through June 2026, preprint-typed definition	0.07	0.15	0.47
Scale margin: asinh papers per PI-year	0.03	0.33	0.08
Scale margin: asinh authorships per PI-year	−0.69	0.64	−1.07

Panel A standard errors are derived from the reported t statistics and Panel B t statistics from the reported standard errors. Joint tests of the annual and quarterly pre-period researcher-level interactions give $p = 0.94$ and $p = 0.60$. The June 2026 rows use provisional records. The scale-margin rows use the annual investigator panel through 2025 (16,471 PI-years) with investigator and field-by-year fixed effects, clustered by investigator.

the exercise repeats when crossing occurs.

E Additional Evidence

Table 2 collects the field-level and researcher-level estimates summarized in Section 3 of the paper.

E.1 Researcher-level specification

The paper's researcher-level regression is

$$\ln N_{ikt} = \rho\, \widehat{\phi}_i\, g_t + \mathrm{PI}_i + (\mathrm{field} \times \mathrm{month})_{f(i)t} + \mathrm{server}_k + \varepsilon_{ikt}, \tag{3}$$

where N_{ikt} is the author count of investigator i's preprint on server k in month t, $\widehat{\phi}_i$ is the investigator's physical-exposure score from up to twenty pre-2020 abstracts, g_t is the METR horizon mapped into coverage under the benchmark bridge, and standard errors are clustered by investigator. Field-by-month effects absorb field-level variation, so ρ is identified from within-field differences in predetermined exposure. Table 2 reports the estimates. Annual and quarterly pre-period interactions are jointly indistinguishable from zero ($p = 0.94$ and $p = 0.60$); these tests do not establish parallel trends. The June 2026 extension is estimated under two preprint definitions, and the 2026

records are provisional: retrieval vintages differ materially for recent months. The scale-margin rows estimate the same interaction on the annual panel with asinh annual papers and asinh annual credited authorships per investigator as outcomes, years with no preprint entering as zeros from each investigator's first panel year. Unlike the author-count ordering, the model's scale-growth ordering in ϕ holds at every capability level, not only near $a = 0$. Figure 1 reports the annual event study.

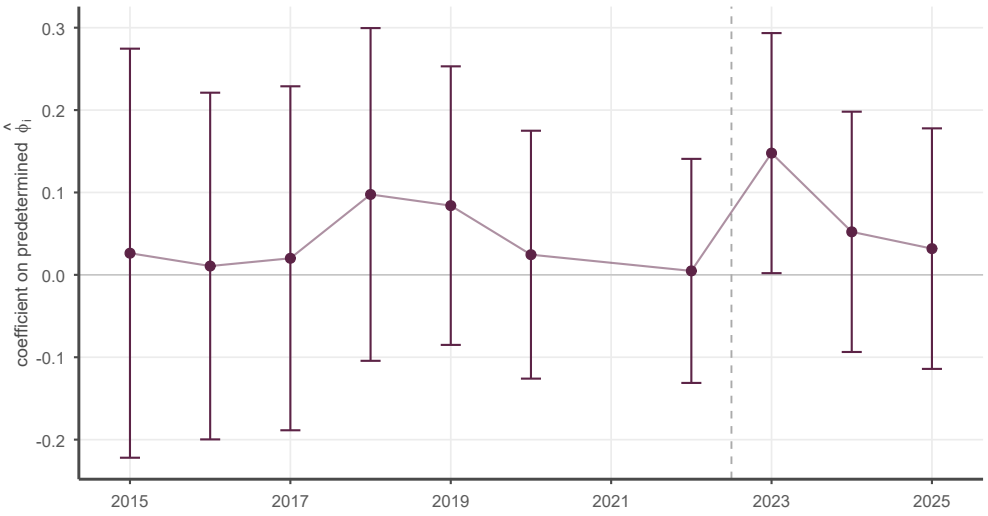

Figure 1: Researcher-level event study. Points are annual coefficients on predetermined physical exposure relative to 2021. Intervals use standard errors clustered by investigator. The dashed line marks the start of 2023.

E.2 Field-level comparisons

The cross-field regression uses OpenAlex author counts for each field and year, 2005–2024; $\bar{N}_{ft}$ is the field-year mean author count after capping each paper's count at 25:

$$\ln \bar{N}_{ft} = \alpha_f + \delta_t + \gamma\, \widehat{\phi}_f\, \mathbf{1}\{t \geq 2023\} + \lambda_f t + \varepsilon_{ft}, \tag{4}$$

with standard errors clustered by the thirty fields. Without field trends the interaction is positive and statistically significant, opposite to the model's early-stage sign; the pre-2015 event-study coefficients show physical fields converging toward larger teams for two decades, so the raw interaction loads on a long pre-trend. With field-specific linear trends and excluding 2020–2022 the interaction turns negative and imprecise, as do the median-based, rank-proxy, and occupation-proxy variants (Table 2).

The within-field comparison classifies 21,243 sampled papers from 2015, 2019, and 2021–2024 by execution mode and estimates, for computational versus physical papers,

$$\ln N_{ift} = \alpha_{ft} + \mu_{fe} + \theta\, \mathbf{1}\{e = \text{comp}\}\, \mathbf{1}\{t \geq 2023\} + \varepsilon_{ift}, \tag{5}$$

with field-by-year and field-by-execution-mode effects, where e indexes execution mode (k is reserved for the preprint server in equation (3)). The estimate is positive and imprecise, and excluding 2015 roughly halves it (Table 2). Event coefficients relative to 2021 are small in 2022–2024, each interval covering zero; the 2015 coefficient is negative and distinct from zero, so the pre-period does not support a general parallel-trends claim. Figure 2 reports both event studies. Paper types can change endogenously within fields, and publication lags separate team formation from publication year, so these are descriptive tests of the predicted ordering.

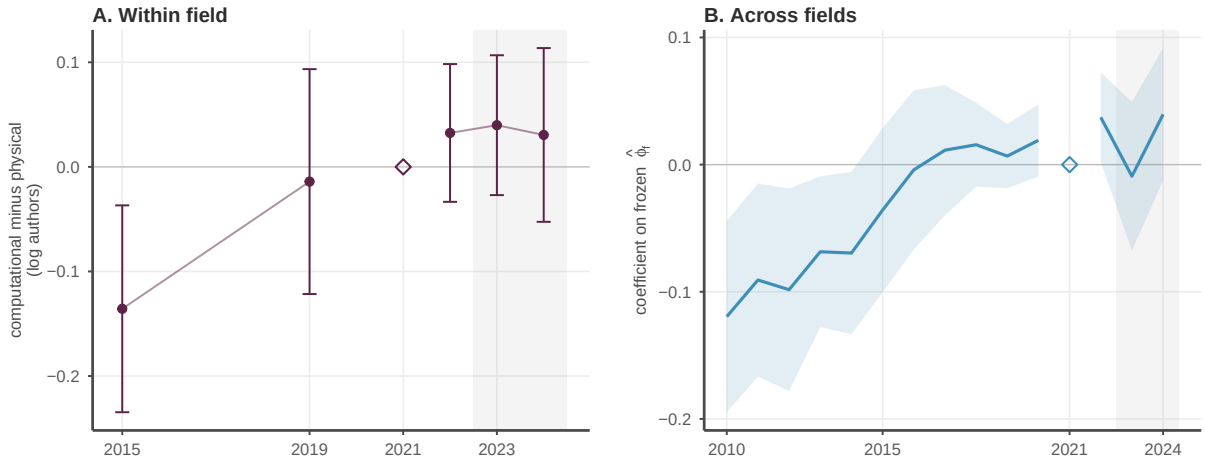

Figure 2: Field-level author-count comparisons. Panel A: within-field difference in log authors between computational and physical papers, by year, relative to 2021. Panel B: coefficient on the frozen field proxy by year, relative to 2021. Intervals use standard errors clustered by field. Shading marks 2023–2024. The 2021 reference year is plotted at zero (open diamond) without an interval; lines do not interpolate through it.

References

Lyttelton, Thomas, Maxim Massenkoff, and Nathan Wilmers. 2026. "Coding Agents in the Social Sciences." *Anthropic Economic Research.*