# THE ECONOMICS OF AUTONOMOUS WORK

From Human-Performed Digital Work
to Compounding Productive Capacity

SUCCESSOR Ω HARNESS

**ACTIVATE. EVOLVE. SCALE.**

Powered by TRANSFORMATIONAL INTELLIGENCE ∞ and the applicable $AGIALPHA utility.

Vincent Boucher | MONTREAL.AI

25 October 2026 Public research edition 1.0

**The employer's question**

I pay a person to reconcile these files. What would I receive from AGI ALPHA, how soon, at what total cost, with how much of my attention—and what would the next job inherit?

**Measured delivery.** A public release-inventory reconciliation; 36/36 fresh controlled jobs accepted under frozen machine criteria; median complete process turnaround 0.301 s; 108 additional operating attempts; eight operating controls; one successor report.

**Economic comparison.** Official labor-market evidence from Quebec, the United States and Europe, published human-work studies, and explicit cost/attention assumptions. No newly recruited human panel is required. No matched human speedup or fully observed monetary saving is claimed.

**TURN COMPLETED WORK INTO LASTING PRODUCTIVE CAPACITY.**

**Abstract**

A person is paid to deliver useful work through a computer. Replacing that transaction requires an acceptable deliverable, a credible total cost and a manageable burden of supervision. This paper evaluates an actual supported AGI ALPHA product path against traditional human-performed digital work using a market-data-grounded comparison. The executed job reconciles two tabular exports and delivers exact joined records, unmatched records, field differences, a source-linked action queue and a portable reusable tool. A public website source-inventory case identifies 16 changed records among 135 matched paths. A prospectively frozen confirmation accepts all 36 fresh controlled jobs across four contexts and three complexity levels; median process turnaround, including artifact generation, independent machine checking and packaging, is 0.301 s, with range 0.161–1.338 s. The retained executable is invoked on all 36 fresh jobs. A distinct reporting adapter composes their evidence with the first delivery into a source-linked successor report. Three repeated batches at each of one, two and four local processes accept 108 additional attempts, with median throughput 131.870, 212.352 and 331.667 deliverables/minute, respectively. These finite local results do not establish an unattended enterprise.

Official compensation data and historical human studies ground the commercial model without inventing worker observations. Where a matched no-AI human duration is unavailable, the model derives break-even time, maximum affordable attention and accepted-capacity conditions. Complete monetary cost remains unobserved because human review, setup, maintenance and investment allocation are not metered. At the documented U.S. office-support compensation rate, an explicitly modeled 30-minute traditional workflow versus five minutes of residual attention plus USD1 of allocated nonlabor cost yields approximately USD17.98 versus USD4.00 per accepted delivery, conditional on equal full acceptance. That is a scenario, not an observed 4.5-fold gain. The contribution is an inspectable bridge from real useful artifacts and executable reuse to adoption economics, with unfavorable scenarios, historical negative studies and unproved acceleration preserved.

## 1 Start with the work an employer needs

Consider a release owner with two exports of a website's asset inventory. Before approving a release, the owner needs to know which paths were added, removed or changed, where each difference came from and what remains to be decided. An employee might use a spreadsheet, a database join, a script or a native version-control diff. The business output is the checked exception package. The employer still decides whether each proposed change belongs in the release.

Our first job uses public source metadata from the actual AGI ALPHA repository: the 135 tracked files beneath `docs/assets/` at the v1.42 source and at the pinned current main. The delivered package finds 135 exact matched paths, no unmatched paths and 16 changed Git-blob identities. It provides each old and new value, its source row, a grouped review action, exact JSON and spreadsheet-safe CSV. The machine acceptance checker independently reconstructs the expected answer from the two input files. This useful artifact exists independently of any claim about future learning.

The same bounded workflow is relevant to product catalogs, stock exports and publication metadata. O*NET describes record verification, discrepancy reporting and reconciliation as paid duties of data-entry and bookkeeping workers; business-intelligence work includes report preparation and reusable reporting tools [38–40]. These occupational descriptions establish task relevance. They do not establish how long this particular job takes or that every responsibility of those roles is automatable.

We define *computer-mediated work* as work performed through digital interfaces, using a keyboard, mouse or equivalent input, with results delivered or actions completed through software. A salaried role can contain computer work, physical work, interpersonal obligations and accountable judgment. This study measures one explicit portion: exact reconciliation and descriptive reporting. It does not substitute a successful spreadsheet operation for an entire occupation.

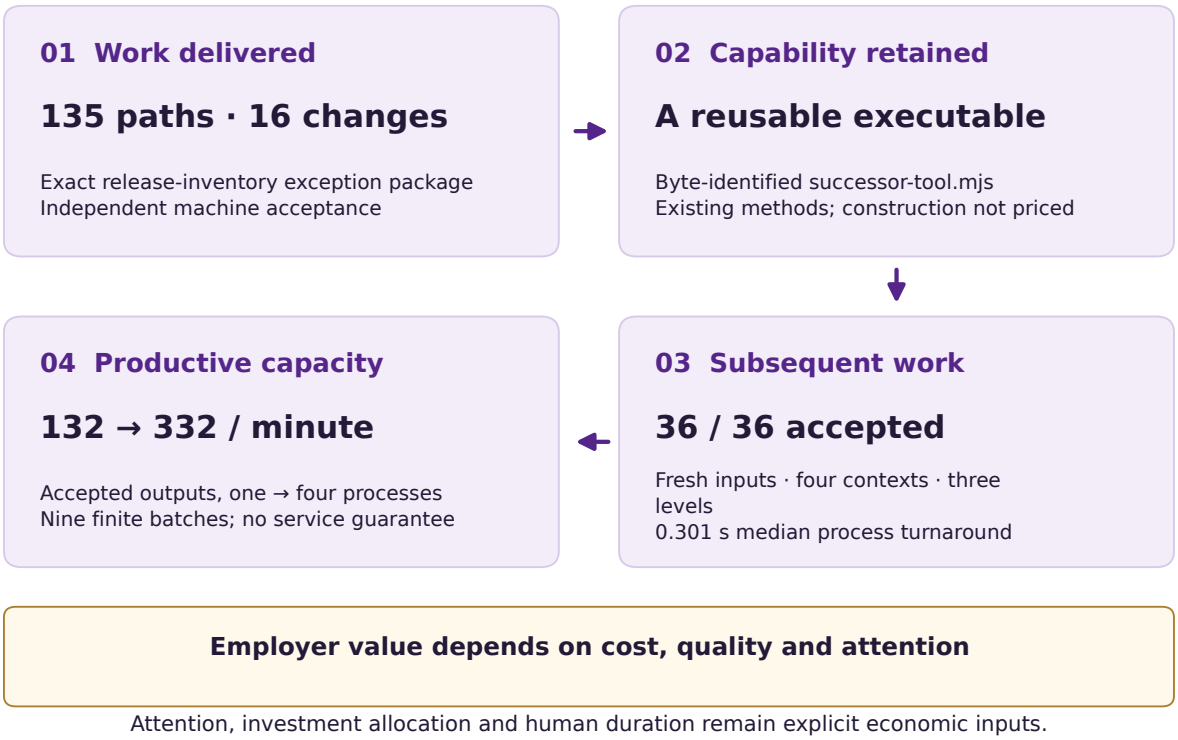



**Figure 1:** The executed sequence and its economic boundary. Solid relationships are observed artifact production, byte-identical tool reuse, fresh acceptance and finite process capacity. The connection to employer savings depends on disclosed attention, cost, reliability and adoption assumptions.

## 2 The product comparison and the evidence hierarchy

The primary comparator is a competent person using ordinary non-AI software and efficient legitimate workflows. A human may reuse formulas, templates, macros, SQL, scripts and professional knowledge. AGI ALPHA may use strong existing tools and capable AI when available and justified. The employer's output contract, information, permitted actions and deadline must be shared; internal methods need not be identical.

For the chosen exact job, the approved released route needs no task-time model inference. It uses maintained deterministic tools, internal checks, reusable source and study-level delivery packaging. This is an actual product capability, with the scope stated. It is not evidence that a frontier model autonomously understood an arbitrary business request or navigated a general desktop. Development and orchestration by the research assistant are separately recorded as unpriced investment work.

| Traditional human workflow | Executed AGI ALPHA workflow |
|---|---|
| Inspect both exports and choose exact keys | Read the frozen brief and configured source keys |
| Use spreadsheet functions, SQL, a script or a native diff | Execute the maintained exact reconciliation tools |
| Check duplicates, unmatched records and changed values | Reconstruct expected records with an independent oracle |
| Prepare the exception list and source-linked handoff | Export CSV/JSON/HTML, hash and package the deliverable |
| Resolve business questions and authorize the release | Leave unresolved truth and release authority with the owner |

**Table 1:** The same employer outcome with different internal methods. The left column is a task description, not a timed human observation.

**Empirical matched no-AI human comparison: NOT RUN. Market-data economic comparison: completed as a conditional model.**

| Evidence class | What this paper has | Permitted inference |
|---|---|---|
| Executed system | Actual inputs, artifacts, timing, CPU, acceptance, lineage and controls | Finite correctness and operating behavior on the declared tasks |
| External human evidence | Official occupations/compensation and published original human studies | Labor-price anchors and contextual workflow evidence within their original scope |
| Conditional economics | Explicit human time, attention, burden, maintenance, allocation and acceptance inputs | Break-even and capacity implications if those assumptions hold |
| Not observed | Matched no-AI human trial, employer stopwatch review, invoices, revenue and cash savings | No empirical human ratio, all-in price or profit claim |

**Table 2:** Evidence classes are retained in machine-readable fields. A missing observation remains null rather than becoming zero.

The phrase *AGI ALPHA FLYWHEEL advantage* denotes the complete system's advantage over the stated traditional human workflow. Here that advantage is conditional in the commercial model, because matched human effort and full monetary costs are not observed. Executable reuse is measured separately. Neither a favorable model nor reuse alone proves that recursive learning caused the whole product's advantage.

# 3 A finished job, then fresh useful missions

## 3.1 The delivered artifact contract

The original catalog CLI writes joined records, unmatched records for each side, conflict records, exact result JSON, a local comparison record, an HTML report, project history and a standalone successor tool [31]. The original browser guide adds the complete field-level difference list and groups related occurrences without dropping evidence. Our thin study wrapper adds a plain-language summary, a manifest, a delivery archive and an independent acceptance record. Later jobs invoke the exact emitted tool and receive the same employer-facing summary and exception artifacts.

Keys are exact strings: case, whitespace, leading zeroes and Unicode are not silently normalized. Duplicate keys retain every matching pair. Common non-key columns are compared; unshared columns are explicitly listed. A conflict means unequal source text, not proof that either source is wrong. Formula-like CSV cells receive a protective leading apostrophe, while exact JSON retains originals. The supported limits are 200,000 bytes and 1,500 rows per input, 32 columns, and 10,000 matching pairs. Exceeding a limit produces refusal rather than truncation.

### 3.2 Prospective boundary and development record

Before confirmation generation, 69 files covering the protocol, producer glue, independent oracle, source inventories, upstream code and analysis were hashed into a frozen boundary. The checkpoint was saved and the boundary recorded in the existing coordination issue before any confirmation result [36]. The SHA-256 begins `41b8b56a39993e29`; the companion contains the complete value and all member hashes.

Development retained a genuine generator failure on a tiny test fixture. Independent pre-freeze review also identified two accounting defects: an outer timeout could lose a failed attempt, and missing CPU measurements could be treated as zero. Both were corrected and checked before freezing. The shared employer deadline was clarified as one eight-hour business day; the system's separate 30-second execution cap is a resource control, not a 30-second limit imposed on the human comparator.

### 3.3 Portfolio, acceptance and sequence

Confirmation consists of 36 fresh jobs: four contexts, three complexity levels and three instances. Base sizes are 80, 360 and 1,200 rows per input; the high level adds 24 duplicates per side. Changed keys, column order, unmatched records, conflicting text, quoted commas, multiline values, Unicode and formula-like strings create meaningful variation. The randomized order mixes difficulty. These are controlled source-generated tasks, not 36 paying customer engagements or a representative sample of worldwide work.

Acceptance is conjunctive. Every pair, unmatched row, conflicting field and source value must be correct; coverage and traceability must be complete; downloadable artifacts and summary links must be valid. Material error tolerance is zero. An independently authored Python CSV/SQLite checker reconstructs the expected semantic records without importing producer modules. Independent role separation reduces implementation coupling, but is not a blinded human assessor or an outside organization. Technical usability is checked; human usability and willingness to pay are not measured.

The public showcase is explicitly development-selected, separate from the fresh 36-job confirmation. It produces the retained executable. All 36 fresh jobs invoke those exact bytes. A final successor task maps the 37 accepted outputs into a grouped workload report through the released public-data workbench. That task combines reconciliation evidence with a second known capability; its totals, counts and extrema are checked independently in Python.

# 4 Measured delivery: quality, speed and resource use

| Controlled complexity | Accepted | Median seconds | Range seconds | Median CPU s |
|---|---|---|---|---|
| Low | 12/12 | 0.197 | 0.161–0.390 | 0.212 |
| Medium | 12/12 | 0.279 | 0.248–0.325 | 0.342 |
| High | 12/12 | 0.978 | 0.868–1.338 | 1.113 |
| All controlled jobs | 36/36 | 0.301 | 0.161–1.338 | 0.358 |

**Table 3:** Complete local process turnaround includes startup, native execution, exports, guide evidence, independent acceptance, hashing and ZIP construction. CPU includes worker startup and completed children through packaging, excluding the final receipt-write tail.

All 36 confirmation jobs passed on their first recorded attempt. The descriptive Wilson 95% interval for the observed acceptance fraction is 90.36–100%; it is not a general reliability guarantee. The bootstrap interval for the median process time is 0.257–0.626 s. The wide interval reflects the deliberately mixed difficulty distribution, not uncertainty in a universal agent speed. All task records, including their specific inputs and outputs, are retained.

The natural public-inventory showcase took 0.276 s within the measured worker window and 0.413 CPU-seconds. Unlike the confirmation table, its standalone invocation did not record the outer Python process-turnaround interval, so those boundaries must not be equated. The successor report took 0.321 s in its worker window and 0.575 CPU-seconds including startup. It consumed 37 source-linked parent outputs and passed exact aggregate checks.

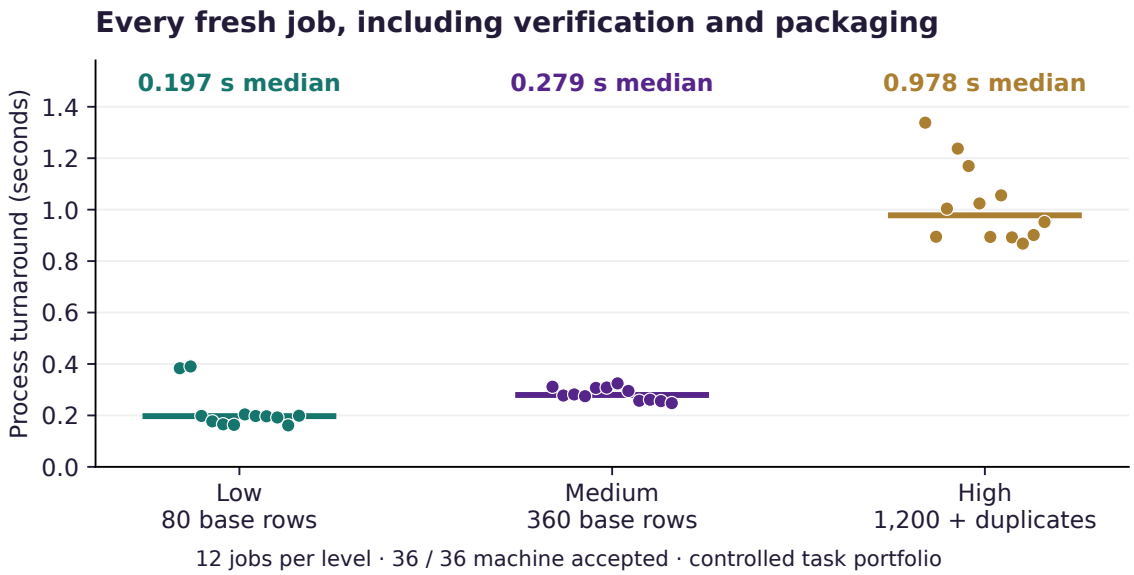



**Figure 2:** Every fresh confirmation job, grouped by fixed complexity. These are measured system delivery times, not a comparison with invented human durations. High-complexity jobs remain visible.

The 36-job campaign consumed 20.987 CPU-seconds in total, or 0.583 CPU-seconds per accepted job. An assumed price per CPU-hour can value that quantity, but the session supplies no host invoice. The median is useful for a typical-job scenario; portfolio cost uses the total divided by accepted outputs. Startup, verification, export and packaging are included in the defined execution measurement. Source acquisition, research authoring, future maintenance, employer review and installation work remain outside that window and must be allocated separately.

# 5 What the next mission inherits

The retained `successor-tool.mjs` has SHA-256 beginning `98fa37ff41e89a8f`. It is emitted by the actual maintained capability builder and subsequently executed 36 times on fresh confirmation inputs. Every invocation is bound to the tool hash in its receipt. The portable executable contains established exact methods; it still runs all four maintained internal strategies. It is not a newly discovered algorithm or an optimized single-method specialization.

This distinction matters commercially. A useful tool can spare a later operator from reconstructing a workflow even when its algorithm is conventional. However, this study does not observe the human reconstruction time, price the earlier engineering investment, or compare inherited and rebuilt processes under matched conditions. The fact established is useful executable reuse. Its investment payback remains a conditional calculation.

| Mission | Asked for / delivered | Retained or reused | Claim supported |
|---|---|---|---|
| Public inventory | Complete list of changed source assets and evidence | Emits portable tool and project record | Useful first artifact; no live-site correctness claim |
| Fresh 36 jobs | New catalog, stock, asset and metadata exception packages | Exact tool bytes plus original guide | Verified fresh-input reuse across supported schemas |
| Successor report | Field-difference workload grouped by context | Parent evidence plus public-data adapter | Known-capability composition and exact lineage |
| Wrong-key control | Apply old key to changed schema | Refuse, then explicit valid-key fallback | Scope correction; no autonomous semantic inference |

**Table 4:** Completed work leaves a usable artifact and a reusable means of producing it. Construction and verification work is retained, rather than assumed free.

We distinguish four claims. *Reuse* requires an inherited artifact to be invoked on fresh input. *Broader capability* requires a meaningfully new accepted contract. *Improved productivity* requires comparable complete resources and outputs. *Increasing improvement rate* requires multiple measured improvement cycles. The first is established here; the composed successor is bounded transfer of known capabilities. Complete productivity and acceleration are not experimentally established.

# 6 Human-work evidence from labor markets and published studies

## 6.1 Occupations and labor-price anchors

The human benchmark is grounded in multiple geographies, with the currency and statistical basis preserved. Quebec Job Bank rates use Statistics Canada's 2023–2024 Labour Force Survey and were updated on 19 November 2025. They are gross wages, not employer costs. Employer-load factors used in sensitivity analysis are assumptions, not measured Canadian benefit rates [15–18, 29].

BLS December 2025 private-industry office and administrative support compensation is USD35.96 per employee hour worked, comprising USD24.79 in wages and USD11.17 in benefits. No additional generic benefits multiplier is applied. The professional-and-related category is USD72.19; it is a broader skill sensitivity rather than a matched clerk rate [37].

Eurostat's 2025 employer-cost data cover enterprises with at least ten employees in its defined covered economy. They supply useful geographic sensitivity, not occupation-matched reconciliation

costs. The EU mean is EUR34.90; the Netherlands, Luxembourg, Bulgaria and Romania span EUR12.00–56.80 [20, 21].

| Benchmark | Reference period | Hourly rate | Statistical basis |
|---|---|---|---|
| Quebec data-entry clerk | 2023–24 | CAD24.00 | Occupational median wage |
| Quebec bookkeeper | 2023–24 | CAD28.02 | Occupational median wage |
| Quebec database analyst | 2023–24 | CAD43.96 | Occupational median wage |
| U.S. data-entry keyer | 2025 | USD19.88 | Occupational median wage |
| U.S. bookkeeping clerk | 2025 | USD24.36 | Occupational median wage |
| U.S. office/admin, private | Dec. 2025 | USD35.96 | Mean wages plus benefits |
| EU covered economy | 2025 | EUR34.90 | Mean employer labor cost |
| Netherlands, covered economy | 2025 | EUR47.90 | Geographic sensitivity |
| Bulgaria, covered economy | 2025 | EUR12.00 | Geographic sensitivity |

**Table 5:** Selected anchors; the companion retains all 12 rate series, sources, occupation codes, methods and scope. U.S. occupational wages are reproduced by official O*NET from BLS [39, 40]. CAD, USD and EUR are not treated as equivalent.

These rates value time; they do not supply time. Their source populations may already use AI in some work. The occupational wage tables do not establish individual AI usage. We therefore use them as labor-price parameters for the explicitly defined traditional no-AI scenario, not as evidence that all sampled employees worked without AI.

### 6.2 Published human performance: retain the original task

Historical studies establish that competent people perform related data work, and that time and quality can vary substantially. Their task contracts remain different from this study's exact reconciliation package.

| Primary study | Observation | Comparison boundary |
|---|---|---|
| WREX, CHI 2020 [14] | Twelve professional data scientists; manual Jupyter condition completed 12/36 short transformation attempts within five-minute limits | Professional ordinary programming, before widespread generative assistants; censored microtasks, not complete reconciliation deliveries |
| TASHEEH, EDBT 2024 [23] | Five expert cleaners; reported ten-file manual cleanup 67 ± 18 minutes and 83 ± 17% accuracy | Row-structure repair, not our contract; preferred tools allowed, explicit AI prohibition not established; dispersion is not a stated confidence interval |
| SpreadsheetBench 2024 [30] | Four Excel experts; 50 instructions with three cases; 62% all-case and 71.33% per-case success | No task-level duration; completion within one day is not divided into invented minutes per task; AI usage not explicitly established |
| OSWorld 2024 [42] | CS students; 111.94-second median and 72.36% success on heterogeneous desktop tasks | Unfamiliar students and mixed tasks are not competent professional reconcilers; no transferred speed denominator |

**Table 6:** External observations are evidence in their original setting, not rows of an AGI ALPHA human trial.

The WREX timing entries include five-minute censoring for failed attempts; converting them into time per successful employer job would conceal failures. TASHEEH's Figure 9 reports cumulative

workflow time; its ten-file observation is not divided into an assumed universal file rate. A later wrangling study is excluded as a no-AI baseline because participants in nominally conventional tool conditions also consulted AI [24]. Annotated action trajectories and expected time ranges likewise remain annotations, not stopwatch observations [1, 43].

No retrieved primary study supplies a fully comparable human duration and acceptance distribution for our exact portfolio. That does not prevent the economic analysis. It determines its form: sourced prices, explicit assumptions and break-even surfaces, rather than a fabricated experimental human/agent ratio.

# 7 The economics of an accepted deliverable

## 7.1 Complete cost, including attention and failure

Let $t_H$ be traditional human minutes per attempted job, including attributable preparation, correction, review and supervision. Let $h_A$ be the corresponding residual human minutes when using AGI ALPHA. Hourly employer valuations are $w_H$ and $w_A$. Let $q_H, q_A$ be accepted outputs divided by attempted jobs at the same accounting boundary as time and resource use, and $u$ measured system CPU-seconds per attempt. Retries must be included consistently in those totals. Then:

$$C_H = \frac{w_H t_H/60 + v_H}{q_H}, \tag{1}$$

$$C_A = \frac{w_A h_A/60 + p_{\mathrm{CPU}} u/3600 + k + A/N}{q_A}. \tag{2}$$

$v_H$ is attributable nonlabor human-workflow cost; $k$ covers per-attempt maintenance and operating overhead; $A/N$ allocates construction, installation and verification investment across the declared number of attempts. Model usage is included in $k$ when present. Our task-time model calls are zero, but authoring cost is not zero merely because the session does not supply a bill. The sum of all incurred costs, including failed and abandoned attempts, divided by accepted outputs is the empirical version of Equation 2.

The study reports marginal *resource quantities*, not an observed marginal price. Fully allocated monetary cost is also unknown. The companion's ledgers record the missing categories explicitly. A numerical all-in cost in the following tables is therefore conditional. An employer can replace each input with its own documented work, review and infrastructure costs.

## 7.2 Break-even time and the attention frontier

Solving $C_H = C_A$ gives:

$$t_H^{\star} = \frac{60}{w_H}\left[\frac{q_H}{q_A}\left(\frac{w_A h_A}{60} + \frac{p_{\mathrm{CPU}} u}{3600} + k + \frac{A}{N}\right) - v_H\right]. \tag{3}$$

For equal acceptance, equal labor rates and no separate human-workflow nonlabor charge, this simplifies to residual attention plus the labor-time equivalent of the system's other cost. The threshold is often more useful than an asserted 100-fold speedup: an employer can ask whether its actual current workflow exceeds it.

For a specified all-in cost per accepted delivery of five local-currency units, with $q_H = 1$ and $v_H = 0$, the traditional-work thresholds are 12.50 minutes at the Quebec data-entry wage, 10.71 at the Quebec bookkeeping wage, 8.34 at U.S. office/admin compensation, 6.26 for the Netherlands employer-cost scenario and 25.00 for Bulgaria. Wage-only rows omit employer loading unless

explicitly selected. These are independent CAD5/USD5/EUR5 scenarios, not an FX comparison or an observed five-dollar AGI ALPHA price.

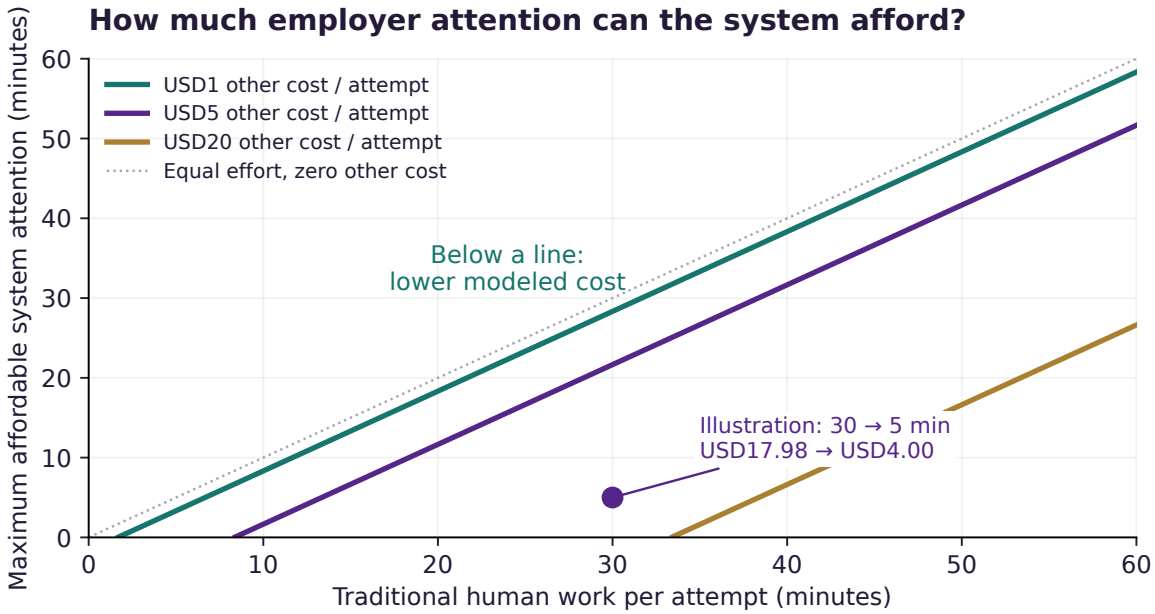



**Figure 3:** Conditional adoption frontier at USD35.96 per hour, equal acceptance and equal attention valuation. Measured median CPU is valued at an assumed USD1/CPU-hour. Lines vary allocated non-attention cost. Human duration and attention are inputs, not observations.

### 7.3 Magnitude, including weak and losing cases

Table 7 is a postfreeze presentation model combining the documented U.S. office/admin rate with measured median CPU and explicit workflow assumptions. Other cost includes maintenance and allocated investment. Both acceptance fractions equal one and $v_H = 0$ in this illustration; the companion varies the acceptance fractions. A lower accepted-delivery probability increases cost and reduces capacity.

| Scenario | $t_H$ min | $h_A$ min | Other USD | $C_H$ USD | $C_A$ USD | $C_H/C_A$ |
|---|---|---|---|---|---|---|
| Equal effort | 5 | 5 | 1 | 3.00 | 4.00 | 0.75 |
| 15-minute workflow | 15 | 5 | 1 | 8.99 | 4.00 | 2.25 |
| 30-minute workflow | 30 | 5 | 1 | 17.98 | 4.00 | 4.50 |
| Low attention | 60 | 1 | 1 | 35.96 | 1.60 | 22.48 |
| Heavy review | 30 | 15 | 5 | 17.98 | 13.99 | 1.29 |
| Short job, heavy review | 5 | 15 | 5 | 3.00 | 13.99 | 0.21 |

**Table 7:** Conditional cost per accepted delivery. No row is an observed matched human experiment. Actual system CPU is priced at an assumed USD1/CPU-hour; full human and allocation inputs remain assumptions.

A 30-to-5-minute attention substitution with USD1 of other allocated cost gives approximately a 4.5-fold cost ratio and sixfold attention-limited capacity. A 60-to-1-minute case can produce a much larger conditional advantage. An efficient five-minute incumbent can instead beat an attention-heavy system. The commercial region is consequently defined by the task's real human burden, the system's exception burden and the repeated-use allocation, not by branding or a compulsory positive result.

For a fixed attention budget $H$ minutes, budget $B$, operating time $T$ and measured accepted machine throughput $\lambda$, usable capacity is bounded by

$$Q_A \leq \min\left(\frac{B}{C_A}, \frac{Hq_A}{h_A}, \lambda T, D\right), \tag{4}$$

where $D$ is available qualifying demand. Accepted throughput $\lambda$ must correspond to the same operating regime as $q_A$; if $\rho$ is attempt throughput, $\lambda = \rho q_A$. An acceptance sensitivity is not a new measured throughput. The repeated local throughput measurement is not evidence that employer attention, data access or demand scale at the same rate. Saved worker minutes create capacity. Cash savings require an actual change in overtime, contractor expenditure, hiring or another expense; an unchanged salary bill is not reduced by arithmetic.

## 8 Scale is an operating observation

| Local processes | Accepted attempts | Median jobs/min | p95 queued latency (s) | Mean CPU/job (s) |
|---|---|---|---|---|
| 1 | 36/36 | 131.87 | 5.456 | 0.523 |
| 2 | 36/36 | 212.35 | 3.061 | 0.572 |
| 4 | 36/36 | 331.67 | 2.037 | 0.621 |

**Table 8:** Twelve mixed jobs per batch, three repeats per concurrency. Queued latency starts at batch submission; it is distinct from process turnaround. Every operating attempt includes machine acceptance and packaging.

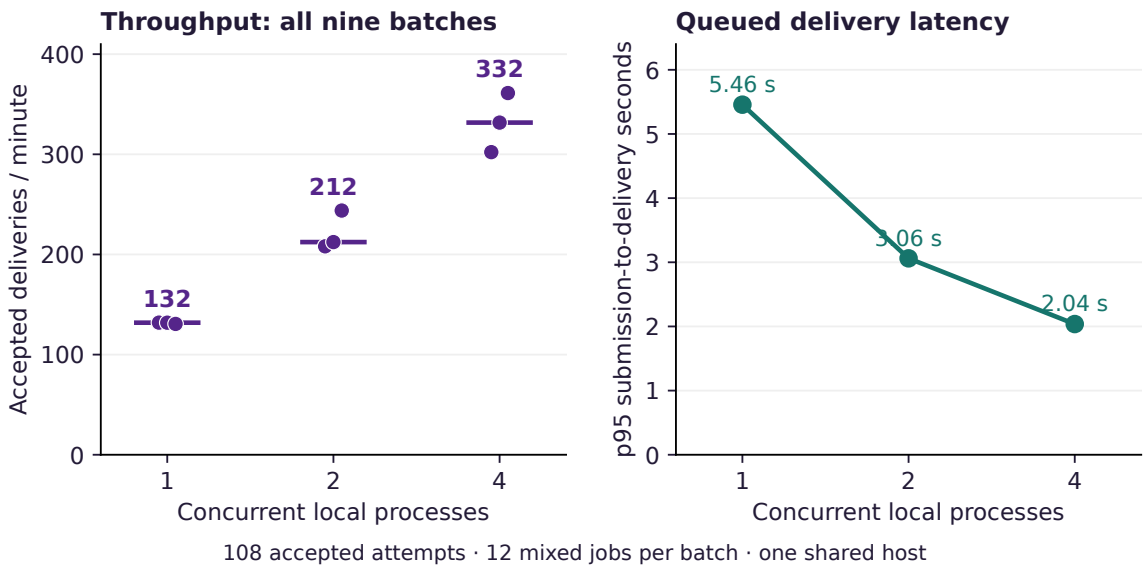



**Figure 4:** Finite accepted throughput and submission-to-delivery latency on one shared host. Markers retain every batch. More processes improve this batch's throughput with higher CPU per job. This is not measured global scaling or a production-service guarantee.

Median throughput rises from 131.87 to 331.67 accepted deliveries per minute, a measured 2.52-fold local increase from one to four processes. Mean CPU per attempted delivery rises from 0.523 to 0.621 seconds, about 18.7%. Initial queued backlog is 11, 10 and 8 jobs. No unplanned rescue intervention occurs within these scripted batches; this observation does not measure an employer's preparation or review time. Physical CPU allocation, infrastructure price, peak host memory and sustained multi-day behavior are not measured.

All eight separate controls pass: pre-aborted STOP; checkpoint interruption followed by historical restore and explicit new execution; input and time boundaries; source mismatch; corrupted saved claims; preserved local artifact restoration; wrong-key refusal with explicit fallback; and exact-string ambiguity preservation. The time-boundary control injects clock readings, rather than spending sixty seconds of real work. Recovery restarts from committed inputs; it does not continue from an internal program counter. Rollback restores local saved bytes; it is not production database or contract rollback.

The browser performs local computation and stops when its execution context closes. A separately started native command can continue only while its process remains alive. This finite study does not install a background service, provision an approved remote executor or establish a continuously operating enterprise. Monetary-spend enforcement, remote-worker rescue and transactional production rollback are not tested.

## 9 The $40 trillion thesis and the commercial opportunity

The relevant global quantity is not GDP, total salaries, the number of computer users or a provider's revenue. A defensible task-expenditure estimate would combine countries and occupations:

$$L_{\text{digital}} = \sum_{c,o} (\text{hours worked}_{c,o})(\text{loaded cost/hour worked}_{c,o})(\text{digital-task time share}_{c,o}). \quad (5)$$

Task allocations must avoid overlap, use compatible years and currencies, match worked-hour or paid-hour denominators consistently, and distinguish employee compensation from imputed self-employment income. ILOSTAT definitions separate earnings from employer labor costs [25, 26].

World Bank displays 2025 world GDP of USD118.35 trillion, while ILO reports a 52.6% global labor-income share [28, 41]. Their mechanical product, USD62.25 trillion, is an arithmetic scale check, not an official harmonized employer-cost estimate. Country coverage, aggregation and price bases were not reconciled here. ILO's recent labor-share estimates include modeling [27]. A USD40 trillion digital-work figure would be roughly 64% of that mechanical envelope, a substantial allocation requiring its own evidence.

Digital-device adoption does not resolve that allocation. Eurostat reports nearly 30% of EU workers using devices all or most of their work time in 2022; JRC's 2024–2025 survey uses a broader any-use concept exceeding 90% [19, 22]. Neither is a worldwide wage-weighted share of tasks a system can deliver. The proposed USD40 trillion annual opportunity is therefore an **illustrative scenario**, worldwide, motivated by 2025-scale economic quantities.

| Conditional quantity | Conservative | Central | Ambitious |
|---|---|---|---|
| Illustrative digital-task expenditure | USD12T | USD24T | USD40T |
| Assumed technical eligibility | 5% | 20% | 50% |
| Assumed adoption of eligible work | 5% | 20% | 40% |
| Adopted baseline expenditure | USD30B | USD960B | USD8T |
| Retained employer-cost fraction | 50% | 30% | 10% |
| Provider-fee fraction | 20% | 20% | 20% |
| Potential customer economic value | USD9B | USD480B | USD5.6T |
| All-provider service fees | USD6B | USD192B | USD1.6T |

**Table 9:** Transparent arithmetic scenarios, not forecasts, confidence bounds or extrapolations from the 36 jobs. The common USD60T total-labor calibration is hypothetical. A particular provider's revenue additionally requires its actual service-market share.

The identity is adopted baseline expenditure equals retained employer costs plus service fees plus potential customer value. Provider costs sit inside the service fee; they are not added twice. Provider revenue is not employer savings, and gross contribution before fixed costs is not profit. Integration, access rights, auditability, liability, procurement and task heterogeneity constrain adoption. The executed portfolio supplies no credible worldwide eligibility percentage.

The opportunity can still be commercially meaningful without an established trillion-dollar total. A narrowly defined recurring job with low exception burden, inspectable delivery and a favorable break-even threshold offers a testable purchase decision. Repeated use can amortize setup and reusable verification infrastructure; per-delivery checking remains a recurring cost. That is a concrete economic route toward accumulating productive capacity; its scale must be earned in actual deployments.

## 10 Transformational Intelligence as an execution discipline

The foundations motivate changing the formulation, representation and means of production while preserving the mission's purpose [2, 13]. In this study, their stages correspond to actual records:

| Stage | Executed operation | Retained evidence |
|---|---|---|
| Identify | Specify the employer deliverable and exact-string semantics | Frozen brief, source inventories and acceptance contract |
| Out-Learn | Inspect schema, source rights and supported bounds | Key choices, field coverage and refusal rules |
| Out-Think | Represent differences as source-linked occurrences and groups | Complete JSON, row references and grouped action queue |
| Out-Design | Package a usable file set and portable execution tool | CSV, HTML, manifest and executable bytes |
| Out-Strategise | Use maintained exact tools; keep ambiguity explicit | No forced truth resolution; declared fallback |
| Out-Execute | Run, independently check, package and retain failures | Attempt receipts, acceptance and control records |
| Out-Evolve | Invoke retained bytes on fresh jobs; compose reporting | Parent hashes, 36 invocations and successor lineage |

**Table 10:** The stages are visible in execution. They are not seven separate learned capabilities or a claim of autonomous self-improvement.

The Out-stage papers inform observation, executable representations, purpose preservation and fresh acceptance [5–10]. The Possibility Foundry motivates charging construction and preserving useful alternatives [11]. Mission-Sovereign and neural-symbolic SUCCESSOR work distinguish portable evidence from authority [4, 32]. Mission Gym and SUCCESSOR $\Omega \times$ AGI Jobs motivate explicit mission contracts, termination and evidence dependencies [3, 12].

The workbench compiles unfunded typed work drafts through existing AGI Jobs compiler source. It does not settle them. AGIJobManager's USDC route remains separate from Jobs v0 and Alpha's $AGIALPHA routes. Source inspection of the five Alpha contracts supports a mechanism map, not live activation: no wallet, funding, staking, minting, posting, settlement or deployed-contract change occurs. The current runtime/release owners and original Harness/Fleet holds remain intact.

## 11 Limits, preserved evidence and the next commercial test

This study executes a narrow useful path, not a universal autonomous employee. Its inputs are bounded public or synthetic tables. Its correctness contract is exact and executable. Real organizations also face missing authority, semantic ambiguity, inconsistent business rules, private data, access permissions, evolving systems and adverse incentives. No result here establishes performance on those untested conditions.

The independent checker is separately authored but within the same research effort. No blinded human assessment, user-study satisfaction or paying customer's acceptance is claimed. The chosen portfolio can establish finite correctness, not a calibrated population reliability. The local process host is shared and short-lived; throughput measurements are not a capacity contract. Current models are not benchmarked, and a future provider upgrade requires fresh evaluation.

Previous research remains unchanged. Strategy Succession's 517/864 qualified outcomes tied its equally funded learner and fell below its stronger fixed portfolio's 591; its primary retained capability was not admitted. The subsequent Software Capability Foundry recorded a whole-instance CPU ratio of 1.0574 against its direct specialist, failing its predeclared superiority gate [33, 34]. These software-only findings are not reinterpreted as human-work results. The present market-data comparison answers a different question and inherits none of their unproved claims.

The next commercial observation should fill the largest practical unknowns: actual task intake and review minutes, exception-resolution effort, installation and maintenance costs, accepted output under a buyer's own contract, and willingness to pay. Existing ordinary software remains part of the competent human alternative. An employer with a fast established script may find little incremental value; one repeatedly constructing and checking the same deliverable may find substantial value. The economic model makes that difference explicit without requiring new participant recruitment to complete this edition.

## 12 Reproduction and publication status

The companion includes original task-execution sources with their hashes, the frozen protocol, generator, independent oracle, complete task materials, raw outputs, per-attempt receipts, load/-control evidence, economic source registers, scenario scripts, editable figures, LaTeX, bibliography and a standalone browser demonstration. `README.md` gives exact reproduction commands. The browser labels new local computation, recorded study evidence and modeled economics separately. Its displayed research observations are not rerun merely by opening the page.

The publication is additive under `studies/agi-alpha-work-economics-v1/`. Coordination and the prospective boundary are public in issue #4850 [35, 36]. Preparation, independent role review, repository review, merge, website deployment and live verification are distinct states. A source-review handoff is not a release. The companion's publication status file records the latest verified state; this paper does not claim an approved hosted deployment.

**What an employer can conclude**

AGI ALPHA can execute the declared exact reconciliation job and leave a portable reusable tool plus inspectable evidence. The 36 fresh confirmation jobs took 0.161–1.338 seconds from process launch through acceptance and packaging; queued batch latency is reported separately. Remaining employer attention and complete monetary costs must be supplied honestly. Sourced labor prices translate those unknowns into explicit adoption thresholds. The commercial aim is useful work whose retained means reduce future burden, with each claimed gain justified at the same accepted-delivery boundary.

# A Cost ledger and measurement boundaries

| Component | Evidence in this edition | Treatment |
|---|---|---|
| Native execution and export | Per-stage wall and CPU; exact output files | Included in worker and outer confirmation turnaround |
| Independent verification | Python CSV/SQLite checker, per-job checks and CPU | Included; not a substitute for employer judgment |
| Hashing and archive packaging | Per-stage timing; SHA-256 output manifests | Included in delivery window |
| Task-time model inference | Zero calls/tokens on this deterministic route | Zero task-model charge; says nothing about research authoring cost |
| Compute price | No actual host invoice | Null actual charge; explicit per-CPU-hour assumptions only |
| Research and construction | Development attempts, source changes and assistant-assisted preparation | Historical/source construction and current authoring bill unmetered; allocate explicitly in scenarios |
| Installation and intake | Public input acquisition and source selection documented | Not a stopwatch employer observation; unknown allocation |
| Human supervision/review | No unplanned scripted-batch rescue; no participant stopwatch | Human minutes unknown; scenario parameter includes preparation, correction and review |
| Retries and failures | All eligible attempts and controls retained | Confirmation has no retries; development failures remain separate investment evidence |
| Maintenance | No longitudinal observation | Explicit $k$ or allocation parameter; not assumed free |
| Financial/chain activity | None | No payments, balances or revenues counted |

For multiple attempts, use the sum of all known incurred resources and report every missing component. An accepted-only median is descriptive; it must not erase failed attempts when estimating delivery cost. The benchmark scripts preserve unfavorable conditions. The study ledger does not convert assistant elapsed time into fabricated human labor.

# B Acceptance details and protected boundaries

The oracle uses SQLite binary string joins to enumerate source-row pairs, then independently constructs unmatched records and differing common fields. It compares semantic multisets, source hashes, all reported metrics, duplicate counts and grouped occurrence coverage. It checks each native strategy's pair result but does not use the producer's pair list as its reference. It also checks safe CSV contents, summary counts, exact originals and local output links.

The source freeze is checked again after execution. Confirmation outputs are not edited to make them pass. The final public package preserves unsuccessful development attempts, injected failure

evidence and the initial review findings. Injection tests are labeled as such. Simulated clock readings, local snapshot restoration and ordinary configuration fallback are not renamed production resilience.

No credentials, participant names, private employer data or private screen captures are part of the public dataset. The records are public repository metadata or synthetic data. The optional future participant protocol is retained separately and creates no current recruitment or payment commitment. Upstream `reviewMinutes:10` is a configured example allowance, not ten observed human minutes; actual human-review fields remain null.

## C Repository and contract contribution map

| Repository | Pinned main prefix | Contribution/status |
|---|---|---|
| AGI-Alpha-Agent-v0 | `f836f06e8351` | Actual exact catalog/guide and pure workbench execution |
| AGIJobManager | `5d5670df9dbb` | USDC commissioning route inspected; no transaction |
| AGIJobsv0 | `bc4abbd5c28f` | Existing compiler dependency and separate $AGIALPHA route; unfunded drafts |
| AGI-Fleet-Linux | `45759fb096fa` | Source/release context only; no admitted external worker |

| Alpha contract | Inspected purpose | Actual study action |
|---|---|---|
| AscensionAccess | ENS eligibility and scoped admission | Source inspection only |
| NovaSeed | Capsule/plan identity and lineage | No sealing or minting |
| AscensionRiskOracle | Scoped validator review and capacity | No attestation or stake |
| AscensionJobMarket | Commissioning, review, payouts/refunds | No posting or settlement |
| AscensionMark | Plan-limited funding and routing | No funding, trade or route |

**Table 12:** Source functions are not evidence of current chain execution. Historical deployment records remain historical. Sharing $AGIALPHA does not transfer approvals or authority between protocols.

## D Source method and scope notes

All market rates retain their original nominal currency. No currency conversion or purchasing-power substitution is used. Quebec low/high wage figures are distributional wage estimates, not confidence bounds. Statistics Canada's wage guide and Job Bank's method distinguish survey and regional estimation approaches; the selected Quebec provincial rows are labeled Labour Force Survey, 2023–2024. Eurostat country figures concern broad covered sectors and enterprise-size coverage; they are never presented as exact clerk rates.

Published human-work evidence retains its original denominator. SpreadsheetBench's expert subset is 50 instructions, not its full 912-instruction benchmark. WREX's five-minute failures remain censored failures. The TASHEEH batch and OSWorld's mixed-task median remain external observations. AI assistance is explicitly unestablished where the publication does not say. These sources ground task relevance and economic sensitivity rather than fabricate a matched trial.

Global market scenarios use a hypothetical common USD60T labor envelope. Technical eligibility, adoption, retained-cost fraction, provider price and market share are assumptions. The task experiment does not estimate any of those fractions. A provider's attainable revenue must be measured from actual offered services, price, accepted volume and realized market access.

## References

[1] Reyna Abhyankar, Qi Qi, and Yiying Zhang. OSWorld-Human: Benchmarking the efficiency of computer-use agents, 2026. URL https://arxiv.org/abs/2506.16042v2. Version 2, 18 May 2026; MLSys 2026.

[2] Vincent Boucher. AGI ALPHA: A scalable substrate for intelligence organizations. MONTREAL.AI research manuscript, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/f836f06e8351e64db1da770e16e304b5feb01f28/docs/manuscript/AGI_ALPHA_Unified_Publication_Final.md. Original Markdown inspected, 25 October 2026; architecture and bounded local evidence.

[3] Vincent Boucher. Gym, specialist asi and successor omega: The executable succession institution, August 2026. URL https://montrealai.github.io/Gym_Specialist_ASI_Successor_Omega_v2_1_0_Print.pdf. Version 2.1.0, 10 August 2026; PDF SHA-256 73bad8930e02e5c81e59add0a029c76b3f1b0f9b4f54cf17dfcf2b2768cfb7d8.

[4] Vincent Boucher. Mission-sovereign successors: Goalos singularity navigator omega + seize, August 2026. URL https://montrealai.github.io/goalos-singularity-navigator-omega-seize-agi-alpha-or-agi-club-owner-access-v1/research/GoalOS_Singularity_Navigator_Omega_SEIZE_Mission_Sovereign_Successors_v5_0_0_A4.pdf. Version 5.0.0, 9 August 2026; PDF SHA-256 9dd9caa6170bd1905e52926516d687c47313d58cd5e2a59e6038eabf0014639b.

[5] Vincent Boucher. Out-design, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Design.pdf. Executable release certificates and usable transition capacity; original LaTeX inspected.

[6] Vincent Boucher. Out-evolve: Building the means of discovery and production, October 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Evolve.pdf. 20 October revised research manuscript; PDF SHA-256 5979bcaf704934ba42e4722389cbb1a9ef0cbc889e1b286428615a82a31e99bc.

[7] Vincent Boucher. Out-execute: Constructing the conditions of completion, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Execute.pdf. Boundary contracts, fresh acceptance and recovery; original LaTeX inspected.

[8] Vincent Boucher. Out-learn, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Learn.pdf. Revised experiment-construction and calibration study; original LaTeX inspected.

[9] Vincent Boucher. Out-strategise, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Strategise.pdf. Joint observation and construction under uncertainty; original LaTeX inspected.

[10] Vincent Boucher. Out-think, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Out_Think.pdf. Purpose-preserving executable representations; original LaTeX inspected.

[11] Vincent Boucher. The possibility foundry, October 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/papers/possibility-foundry-v1.pdf. 23 October executable digital laboratory; original LaTeX inspected.

[12] Vincent Boucher. Successor omega x agi jobs: The constitutional compiler and proof-carrying work system for mission-sovereign intelligence institutions, August 2026. URL https://montrealai.github.io/successor-omega-x-agi-jobs-v2.pdf. Version 2.0.0, 12 August 2026; PDF SHA-256 c25b911f20ac623bdb26f520376f2b1f0e774eba982db1c103a5e73d4b1d1a0d.

[13] Vincent Boucher. Transformational intelligence: Learning to change the game, 2026. URL https://montrealai.github.io/AGI-Alpha-Agent-v0/Transformational_Intelligence.pdf. Original LaTeX inspected at Alpha commit f836f06e8351e64db1da770e16e304b5feb01f28.

[14] Ian Drosos, Titus Barik, Philip J. Guo, Robert DeLine, and Sumit Gulwani. Wrex: A unified programming-by-example interaction for synthesizing readable code for data scientists. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. doi: 10.1145/3313831.3376442. URL https://pg.ucsd.edu/publications/Wrex-synthesizing-readable-data-science-code_CHI-2020.pdf.

[15] Employment and Social Development Canada. Job bank: Wages, accounting technician in quebec, 2025. URL https://www.jobbank.gc.ca/marketreport/wages-occupation/24500/QC. NOC 12200; Statistics Canada Labour Force Survey 2023–2024. Wages updated 19 November 2025. Accessed 25 October 2026.

[16] Employment and Social Development Canada. Job bank: Wages, database analyst in quebec, 2025. URL https://www.jobbank.gc.ca/marketreport/wages-occupation/17873/QC. NOC 21223; Statistics Canada Labour Force Survey 2023–2024. Wages updated 19 November 2025. Accessed 25 October 2026.

[17] Employment and Social Development Canada. Job bank: Wages, data input clerk in quebec, 2025. URL https://www.jobbank.gc.ca/marketreport/wages-occupation/19213/QC. NOC 14111; Statistics Canada Labour Force Survey 2023–2024. Wages updated 19 November 2025. Accessed 25 October 2026.

[18] Employment and Social Development Canada. Job bank: Questions and answers on wages, 2026. URL https://www.jobbank.gc.ca/trend-analysis/search-wages/wage-methodology. Methodology; wages exclude non-wage benefits. Accessed 25 October 2026.

[19] Eurostat. Eu employment: use of digital devices, June 2023. URL https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20230627-1. EU Labour Force Survey 2022 ad hoc extraction. Published 27 June 2023. Accessed 25 October 2026.

[20] Eurostat. Eu hourly labour costs ranged from 12 to 57 euros in 2025, March 2026. URL https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20260331-2. Published 31 March 2026. Dataset lc_lci_lev; enterprises with 10 or more employees, covered economy. Accessed 25 October 2026.

[21] Eurostat. Hourly labour costs, 2026. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Hourly_labour_costs. Statistics Explained; data extracted 31 March 2026. Accessed 25 October 2026.

[22] Ignacio Gonzalez Vazquez, Enrique Fernandez Macias, Sally Wright, and Davide Villani. Digital monitoring, algorithmic management and the platformisation of work in europe. Technical Report JRC143072, EUR 40466, European Commission, Joint Research Centre, 2025. URL https://publications.jrc.ec.europa.eu/repository/handle/JRC143072. AIM-WORK survey 2024–2025. Accessed 25 October 2026.

[23] Mazhar Hameed, Gerardo Vitagliano, Fabian Panse, and Felix Naumann. TASHEEH: Repairing row-structure in raw CSV files. In *Proceedings of the International Conference on Extending Database Technology*, 2024. URL https://openproceedings.org/2024/conf/edbt/paper-108.pdf.

[24] Shiyi He, El Kindi Rezig, Paul Rosen, and Andrew M. McNutt. How wrangling tools shape wrangling: A technical dimensions analysis, 2026. URL https://arxiv.org/abs/2607.26198v1. Version 1, 28 July 2026.

[25] International Labour Organization. Ilostat: Wages and working time statistics, cond database. URL https://ilostat.ilo.org/methods/concepts-and-definitions/description-wages-and-working-time-statistics/. Definitions of earnings, labour costs, coverage and comparability. Accessed 25 October 2026.

[26] International Labour Organization. Sdg indicator 10.4.1: Labour share of gdp, metadata. Technical report, United Nations Statistics Division, 2026. URL https://unstats.un.org/sdgs/metadata/files/Metadata-10-04-01.pdf. Metadata updated 27 March 2026. Accessed 25 October 2026.

[27] International Labour Organization. Ilo modelled estimates: Methodological overview. Technical report, ILO Department of Statistics, January 2026. URL https://webapps.ilo.org/ilostat-files/Documents/TEM.pdf. Accessed 25 October 2026.

[28] International Labour Organization. The world of work and the 2030 agenda: a ten-year review, March 2026. URL https://ilostat.ilo.org/blog/the-world-of-work-and-the-2030-agenda-a-ten-year-review/. Published 19 March 2026. Accessed 25 October 2026.

[29] Eesha Kirubaharan and Andre Bernard. Earnings and wages: A guide to using indicators from the survey of employment, payrolls and hours and the labour force survey. Technical report, Statistics Canada, June 2024. URL https://www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2024002-eng.htm. Released 27 June 2024. Accessed 25 October 2026.

[30] Zeyao Ma, Bohan Zhang, Jing Zhang, Jifan Yu, Xiaokang Zhang, Xiaohan Zhang, Sijia Luo, Xi Wang, and Jie Tang. SpreadsheetBench: Towards challenging real world spreadsheet manipulation, 2024. URL https://arxiv.org/abs/2406.14991v2. Version 2, 17 October 2024.

[31] MONTREAL.AI. Successor alpha catalog reconciliation cli, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/f836f06e8351e64db1da770e16e304b5feb01f28/examples/successor-alpha/cli.mjs. Exact Git blob verified; task-execution code matches public release v1.43.0.

[32] MONTREAL.AI. Neural-symbolic successor omega: Customer-owned mission intelligence, August 2026. URL https://montrealai.github.io/neuralsymbolicsuccessoromegav1.pdf. Definitive corporate white paper, 25 August 2026; PDF SHA-256 36e7c26fe28961dbd8a0dc89160a0462e08acb53fc9b00e2508a8740b52847e9.

[33] MONTREAL.AI. The software capability foundry: additive research handoff, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/pull/4900. Source draft; negative superiority gate preserved.

[34] MONTREAL.AI. Agi alpha flywheel: Strategy succession, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/tree/f836f06e8351e64db1da770e16e304b5feb01f28/studies/agi-alpha-flywheel-strategy-succession-v1.

[35] MONTREAL.AI. Agi alpha human-work economics: additive coordination, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/issues/4850#issuecomment-6396844402.

[36] MONTREAL.AI. Prospective boundary: The economics of autonomous work, 2026. URL https://github.com/MontrealAI/AGI-Alpha-Agent-v0/issues/4850#issuecomment-6396944650.

[37] U.S. Bureau of Labor Statistics. Employer costs for employee compensation: December 2025, March 2026. URL https://www.bls.gov/news.release/archives/ecec_03202026.htm. Published 20 March 2026. Table 4: private industry occupations. Accessed 25 October 2026.

[38] U.S. Department of Labor, Employment and Training Administration. O*net online: Business intelligence analysts, 15-2051.01, 2026. URL https://www.onetonline.org/link/summary/15-2051.01. Tasks only: displayed wages refer to the broader Data Scientists occupation and are not used as a matched analyst rate. Accessed 25 October 2026.

[39] U.S. Department of Labor, Employment and Training Administration. O*net online: Bookkeeping, accounting, and auditing clerks, 43-3031.00, 2026. URL https://www.onetonline.org/link/summary/43-3031.00. Occupational tasks; wages reproduced from BLS 2025 data. Page updated 2026. Accessed 25 October 2026.

[40] U.S. Department of Labor, Employment and Training Administration. O*net online: Data entry keyers, 43-9021.00, 2026. URL https://www.onetonline.org/link/summary/43-9021.00. Occupational tasks; wages reproduced from BLS 2025 data. Page updated 2026. Accessed 25 October 2026.

[41] World Bank. World development indicators: World, gdp (current us dollars), 2026. URL https://data.worldbank.org/country/world. 2025 value displayed as USD 118.35 trillion. Accessed 25 October 2026.

[42] Tianbao Xie et al. OSWorld: Benchmarking multimodal agents for open-ended tasks in real computer environments, 2024. URL https://arxiv.org/abs/2404.07972v2. Version 2, 30 May 2024.

[43] Mengqi Yuan et al. OSWorld 2.0: Benchmarking computer use agents on long-horizon real-world tasks, 2026. URL https://arxiv.org/abs/2606.29537v2. Version 2, 13 July 2026.