T E C H N I C A L P A P E R

The Agentic Day

A Human-Equivalent Unit for Measuring AI Agent Output

Matthew Glover

Version 1.0 — May 2026

This paper proposes the Agentic Day (AD), a unit of measurement for expressing AI agent output in terms of equivalent human productive effort. It is intended as an open framework that organisations, researchers, and practitioners can adopt, adapt, and build upon for their own measurement needs.

Abstract

The rapid deployment of autonomous AI agents across software engineering, professional services, and knowledge work has created an urgent measurement problem: there is no widely adopted, intuitive unit for expressing what an AI agent produces in terms that non-technical stakeholders already understand and price. Existing metrics are fragmented across vendor-specific billing primitives, academic capability benchmarks, and consulting-derived productivity ratios, none of which have achieved the universal legibility required for workforce planning, engagement estimation, or cross-organisational comparison.

This paper proposes the Agentic Day (AD), a unit of measurement that expresses AI agent output as the equivalent of one productive human working day. The AD is designed to serve the same bridging function that horsepower served in the age of steam: translating the capability of a new technology into terms anchored to the technology it augments or displaces. The paper defines the unit, establishes its reference specification, presents a suite of derived metrics, examines the historical precedents that inform its design, and provides implementation guidance for organisations adopting the framework.

1. The Measurement Problem

Every major technological transition has faced a translation challenge. When the capability of a new technology cannot be expressed in terms the existing workforce, leadership, and commercial structures already understand, adoption stalls, value is misstated, and planning degrades into guesswork.

The current landscape of AI measurement suffers from exactly this problem. The field has produced a proliferation of metrics, but no single unit that an executive, delivery lead, or client can use to answer the basic question: how much work did the AI do, and how does that compare to what a person would have done?

1.1 The Fragmented Landscape

Existing AI measurement approaches fall into four broadly overlapping families:

1. **Labour-accounting metrics** that translate AI output into Full-Time Equivalents or their derivatives. These include the Digital FTE (D-FTE) concept used in enterprise consulting, APQC's Digital Worker Equivalent (DWE) in benchmarking, and Salesforce's Agentic Work Unit (AWU) as a billing primitive. Each captures a slice of the picture but is either too coarse (FTE-level), too vendor-specific (AWU), or too narrowly benchmarked (DWE in HR functions).

2. **Capability benchmarks** expressed in human time. The most influential is METR's task-completion time horizon, which estimates the human-expert duration of a task and fits a success curve to model performance. This approach has the right instinct — expressing AI capability in human-time units — but is oriented toward research evaluation rather than operational planning or commercial communication.
3. **Economic-impact studies** that quantify automatable work hours and dollar value at the macro level. McKinsey Global Institute's work estimates that demonstrated AI technologies could technically automate 57% of US work hours. The Anthropic Economic Index maps real AI conversations to occupational task taxonomies, revealing a roughly 43/57 automation-to-augmentation split. These provide valuable macro context but do not yield a unit usable at the engagement or team level.
4. **Commercial ROI frameworks** such as Forrester's Total Economic Impact (TEI) methodology and vendor-specific KPIs (Microsoft's "minutes saved per day," Salesforce's deflection rate). These produce dollar-denominated ROI figures but are typically vendor-commissioned, composite-organisation models — useful for procurement justification but not for day-to-day planning or cross-vendor comparison.

None of these frameworks provides a single, intuitive, universally applicable unit of AI output expressed in terms the user already understands. That is the gap the Agentic Day is designed to fill.

1.2 Why a Bridging Unit Matters Now

The urgency of this problem is increasing. Gartner predicts that over 40% of agentic AI projects will be cancelled by end-2027, citing escalating costs, unclear business value, and the phenomenon of "agent washing" — vendors relabelling conventional automation as agentic AI. A 2025 systematic review of 84 agentic AI evaluation papers found that 83% relied on technical metrics alone, only 30% included human-centred or economic metrics, and just 15% combined both, concluding that current measurement practice systematically overstates real-world deployment value.

The absence of a shared, credible measurement unit contributes directly to these failures. Without a common language for AI output, organisations cannot reliably plan capacity, estimate engagements, compare approaches, or communicate value to non-technical stakeholders. The Agentic Day proposes that language.

2. Historical Precedent: The Pattern of Bridging Units

The Agentic Day follows a well-documented pattern. Every major technological transition has required a bridging unit that translates the new technology's capability into terms anchored to the technology it augments or displaces. The success or failure of these units offers direct design lessons.

2.1 Horsepower (1782)

James Watt faced a precise analogue of today's problem: selling steam engines to mill-owners who understood horses. His solution was to measure the output of brewery dray horses and define one horsepower as 33,000 foot-pounds per minute — a figure that, by most physiological analyses, overstated sustained equine output by roughly 50%. The overstatement was tolerable because the engines vastly outperformed the benchmark, and it served a marketing function: every horsepower rating made the engine look more compelling.

Horsepower succeeded because of four properties: it was anchored to the displaced unit of effort (a horse, not a piston); it was a single scalar (engines could be rated in one number); it came with a reproducible test method (the indicator diagram, later the Prony brake and dynamometer); and it had marketing utility that favoured the seller just enough to drive adoption while remaining defensible. It remains in widespread use 244 years later, and was eventually formalised into the SI watt (1 hp ≈ 745.7 W).

2.2 Candlepower (1860s–1948)

Candlepower measured luminous intensity relative to a standard spermaceti candle, later standardised against carbon-filament lamps (1909, 1921) and a platinum blackbody (1937), before the SI candela replaced it in 1948 and the lumen became the dominant practical unit. Candlepower survives today only in flashlight marketing. The lesson: bridging units that lack precise, instrument-based redefinition once the technology matures become marketing artefacts rather than engineering standards.

2.3 MIPS (1970s–1990s)

Million Instructions Per Second (MIPS) was useful in the late 1970s when the VAX-11/780 served as the de facto "1 MIPS machine." But as processor architectures diverged — RISC versus CISC, pipelining, superscalar execution — MIPS became meaningless. Critics dismissed it as "Meaningless Indicator of Processor Speed" because it measured an internal proxy (instruction count) rather than delivered work, and because different instructions performed different amounts of useful computation. The industry migrated to workload-based benchmarks (SPEC CPU, Geekbench) and to FLOPS for scientific computing.

MIPS is the cautionary tale most relevant to AI measurement. Token counts, API calls, and turn counts are the MIPS of the agentic era: internal machine activities that correlate loosely and

unreliably with useful output. Any credible AI measurement unit must measure delivered work, not internal throughput.

2.4 Design Principles Extracted

From these precedents, four principles emerge for designing a successful bridging unit:

Principle	Horsepower	MIPS (Failure)	Implication for AI
Anchor to displaced effort	Horses, not pistons	Machine internals	Human hours, not tokens
Single scalar	One number rating	One number (but misleading)	One number with clear anchor
Measure delivered work	ft-lbs/min	Instruction count	Human-equivalent output
Reproducible test method	Indicator diagram	Present but misleading	Defined calibration protocol

3. The Agentic Day: Formal Definition

3.1 Core Definition

> **Definition**
>
> One Agentic Day (1 AD) is the productive output equivalent of six (6) human-productive-hours of work, delivered by one or more AI agent runs operating within defined parameters, regardless of the calendar time consumed or the specific AI model employed.

3.2 Properties

Property	Value	Rationale
Human-equivalent anchor	6 productive hours	Empirical developer productivity (see §3.3)
Output type	Raw output	Quality tracked separately via derived metrics
Model dependency	Model-agnostic	Measures what was delivered, not how
Task-type scope	Universal (v1)	Refinement by task type planned for v2
Granularity	Per-agent unit	Parallelism scales linearly
Version stability	Fixed unit	1 AD always = 6 human-equivalent hours

3.3 The Human Anchor: Why Six Hours

The six-hour anchor is grounded in a consistent body of productivity research. Studies of software developers and knowledge workers repeatedly find that genuinely productive output — time spent writing code, producing analysis, creating documentation, or solving technical problems — occupies approximately five to six hours of a standard eight-hour working day. The remainder is consumed by organisational overhead: meetings, communication, context-switching, administrative tasks, and the natural cognitive recovery that punctuates sustained deep work.

By anchoring the AD to six hours rather than eight, the unit captures what a developer actually delivers, not what a timesheet records. This choice has three important properties:

- **Conservative by design.** The unit slightly understates the human benchmark. If anything, AI output measured in ADs is modestly understated. This places the unit on the credible side of the line — analogous to Watt's slight overstatement of equine output, which made his engines look even more compelling by comparison.
- **Immediately legible.** Every project manager, delivery lead, and finance partner already thinks in developer-days. The AD slots directly into existing estimation, planning, and billing frameworks without requiring a new conceptual vocabulary.

- **Honest about human capacity.** It avoids the common fiction that an eight-hour day equals eight hours of output, which would inflate the denominator and systematically understate AI's relative contribution.

3.4 Model Agnosticism

The AD measures what was delivered, not how it was produced. A task completed by one model family and a task completed by another are measured by the same yardstick: the human effort they are equivalent to. This mirrors how horsepower works — 200 hp is 200 hp whether it comes from a turbocharged four-cylinder or a naturally aspirated V8.

In practice, different models will achieve one AD at different cost points and different wall-clock durations. These differences surface in derived metrics (cost per AD, wall-clock efficiency) rather than in the base unit, which remains a pure measure of output equivalence.

3.5 Raw Output

The AD measures raw agent output — the work produced before human review, rework, or acceptance. This is analogous to measuring brake horsepower at the crankshaft rather than wheel horsepower after drivetrain losses.

This design keeps the base unit clean and comparable. The efficiency of the human-agent review loop is captured by the Effective Agentic Day (EAD), a derived metric defined in Section 5. Tracking quality separately from quantity allows organisations to monitor and improve both dimensions independently.

3.6 The Fixed-Unit Principle

The Agentic Day is a fixed unit: 1 AD always equals six human-equivalent productive hours. This is a deliberate and consequential design choice.

AI model capability is improving rapidly. Research from METR indicates that frontier model task-completion time horizons are roughly doubling every four to seven months. Under the fixed-unit principle, this improvement means that the same calendar time yields progressively more ADs as models improve. The benefits of capability improvement accrue naturally — either to the organisation (in the form of lower cost per AD) or to the client (in the form of more effective output per engagement), without requiring the unit itself to be redefined.

The alternative — a floating unit that scales with model capability — would erode comparability across time periods, make longitudinal analysis unreliable, and require constant recalibration. The fixed unit avoids these problems, just as the watt remains fixed even as engines become more efficient.

4. Reference Specification

4.1 Agentic Run Parameters

An Agentic Day is composed of one or more agentic runs operating within a continuous window. The reference specification establishes baseline operational parameters:

Parameter	Reference Value	Notes
Maximum turns per run	300	Upper bound; many tasks complete in fewer turns
Timeout per run	2,700 seconds (45 min)	Wall-clock timeout for a single run instance
Concurrency	Single-threaded per agent	One active run per agent at a time
Runs per AD	Variable	Multiple sequential runs within a ~6hr-equivalent window
Reference models	Frontier-class models (2025–26)	Unit is model-agnostic; calibrated against frontier class

These parameters define the operational envelope within which the equivalence has been calibrated. They are not the definition of the unit itself. The AD is defined by its output equivalence (six human-productive-hours), not by its input parameters. As models, tooling, and orchestration evolve, the run parameters may change while the unit remains fixed.

4.2 Calibration

The v1.0 equivalence has been calibrated through sustained agentic development work across code generation, automated testing, documentation, and analysis tasks, supplemented by collaborative research. The calibration reflects the observation that, within the reference parameters, the cumulative output of an agent operating across multiple sequential runs produces work equivalent in scope, complexity, and utility to what a skilled developer would produce in six productive hours.

This paper recommends that adopting organisations validate the calibration against their own domains, task distributions, and agent configurations before using the AD for planning or reporting. Section 7 provides guidance on conducting domain-specific calibration studies.

5. Derived Metrics

The AD is a base unit. The following derived metrics build on it to provide operational, quality, and economic visibility. Organisations adopting the AD framework may use any subset of these metrics as appropriate to their context.

5.1 Effective Agentic Day (EAD)

```
EAD = AD × Acceptance Rate
```

The Effective Agentic Day measures the portion of raw agent output that passes human review without significant rework. If a project delivers 40 ADs of raw output and the acceptance rate is 85%, the effective delivery is 34 EADs. This metric tracks the quality of the human-agent collaboration loop and is expected to improve over time as prompt engineering, agent tooling, and domain expertise mature.

The separation of AD (quantity) from EAD (quality-adjusted quantity) is a fundamental design feature. It prevents gaming — an agent cannot achieve a high AD count by producing fast, low-quality output without the EAD reflecting the true picture.

5.2 Cost per Agentic Day (CPAD)

```
CPAD = (API / compute cost + tooling cost) ÷ ADs delivered
```

Cost per AD enables direct comparison with the fully-loaded cost of a human working day. If a human developer-day costs £800 (or $1,000, or €900) and one AD costs £15 in compute, the unit economics are immediately visible. CPAD is the metric that makes the economic case for agentic delivery legible to finance teams and decision-makers without requiring them to understand API pricing models.

5.3 AD Velocity

```
AD Velocity = ADs delivered ÷ calendar period
```

AD Velocity is the throughput metric for agentic delivery capacity, analogous to velocity in agile methodologies. Unlike story points, however, AD Velocity has a fixed external anchor (human-equivalent hours), making it comparable across teams, engagements, organisations, and time periods. A team reporting 60 AD Velocity per week can be meaningfully compared to a team reporting 30, regardless of their internal processes.

5.4 Augmentation Ratio

```
Augmentation Ratio = (ADs + Human-Days) ÷ Human-Days
```

The Augmentation Ratio expresses how much additional capacity the agentic model provides relative to the human team. A team of two developers (10 human-days per week) augmented by agents delivering 30 ADs per week has an augmentation ratio of 4× — the team operates as if it were four times its size. This is the headline metric for communicating the force-multiplier effect of agentic delivery to leadership and stakeholders.

5.5 Wall-Clock Efficiency

```
Wall-Clock Efficiency = ADs delivered ÷ calendar hours consumed
```

Wall-Clock Efficiency captures one of AI's fundamental advantages: it does not sleep, attend meetings, or context-switch. A single agent running continuously achieves approximately 0.67 ADs per calendar hour (4 ADs ÷ 6 hours of equivalent output per 24 clock hours). Multiple parallel agents multiply this figure. Wall-Clock Efficiency makes the "always-on" advantage of agentic delivery visible and measurable.

5.6 Metric Summary

Metric	Abbreviation	Formula	Primary Audience
Agentic Day	AD	Base unit (6 human-hours)	All stakeholders
Effective Agentic Day	EAD	AD × Acceptance Rate	Delivery, QA
Cost per AD	CPAD	Total cost ÷ ADs	Finance, Leadership
AD Velocity	ADV	ADs ÷ calendar period	Delivery, PMO
Augmentation Ratio	AR	(AD + HD) ÷ HD	Leadership, Strategy
Wall-Clock Efficiency	WCE	ADs ÷ calendar hours	Operations, Capacity

6. Scaling and Parallelism

6.1 Per-Agent Linearity

The AD is a per-agent unit. Parallelism scales linearly: three agents operating concurrently, each delivering one AD, consume three ADs in the same wall-clock period. This linearity is a core design property that makes capacity planning straightforward.

> **The 24-Hour Equation**
>
> A single agent running continuously for 24 hours delivers 4 ADs (24 ÷ 6). Three parallel agents deliver 12 ADs in the same 24 hours — the equivalent of nearly two and a half five-day working weeks of developer output in a single calendar day.

6.2 Illustrative Scaling

Scenario	Agents	Calendar Time	ADs	Human Equivalent
Single agent, business hours	1	8 hours	1.3	~1 developer-day
Single agent, overnight batch	1	16 hours	2.7	~2.7 developer-days
Single agent, 24 hours	1	24 hours	4.0	~4 developer-days
3 agents, 24 hours	3	24 hours	12.0	~12 developer-days
5 agents, 5-day week	5	120 hours	100.0	~20 developer-weeks
10 agents, continuous month	10	720 hours	1,200	~240 developer-weeks

These figures represent raw output (ADs). Effective delivery (EADs) will be lower depending on acceptance rates. The gap between AD and EAD is itself a valuable diagnostic: a narrowing gap over time indicates that the human-agent workflow is maturing.

6.3 Orchestration as a Separate Concern

The AD measures agent output exclusively. It does not include the human time required to prompt, configure, orchestrate, review, or integrate that output. This separation is essential for measurement clarity. In practice, organisations using the AD framework should track human orchestration effort alongside ADs, producing blended delivery estimates such as:

“40 ADs of agentic delivery + 5 human-days of orchestration and review.”

The ratio of ADs to human orchestration days is itself an informative metric, reflecting the maturity and efficiency of the organisation’s agentic delivery practice.

7. Implementation Guidance

This section provides practical guidance for organisations adopting the Agentic Day framework.

7.1 Step 1: Establish a Baseline

Before measuring AI output, measure human output. Select a representative set of tasks from the target domain (e.g., code generation, test authoring, documentation, analysis) and record the time skilled practitioners take to complete them. Use productive hours, not elapsed time: exclude meetings, breaks, and context-switching from the measurement. This baseline serves two purposes: it validates whether the six-hour productive-day anchor is appropriate for the specific domain, and it provides the human-time reference against which agent output is compared.

7.2 Step 2: Calibrate Agent Output

Run the same or equivalent tasks through the agentic system under the reference parameters (or the organisation's own operational parameters). Compare the agent's output — in terms of scope, completeness, and utility — to the human baseline. The calibration question is: does the agent's output from one operational window (a set of sequential runs totalling approximately six equivalent hours of work) match what the human baseline established?

If the output is systematically higher or lower, the organisation has two options: adjust the operational parameters to bring output into alignment with 1 AD (recommended for v1 adoption), or document the domain-specific conversion factor and apply it as a multiplier. The base unit (1 AD = 6 human-productive-hours) should remain unchanged.

7.3 Step 3: Instrument and Track

Integrate AD measurement into existing delivery tooling. At minimum, track:

- ADs delivered per engagement, sprint, or reporting period
- Acceptance rate (proportion of raw output accepted without significant rework)
- Compute/API cost per AD
- Human orchestration days alongside ADs

These four data points are sufficient to derive EAD, CPAD, AD Velocity, and the Augmentation Ratio.

7.4 Step 4: Review and Refine

Review the calibration at least annually, or upon significant changes in model capability, tooling, or domain scope. Calibration reviews should adjust operational parameters and cost models. They

should not adjust the core anchor (1 AD = 6 human-productive-hours) unless a fundamental shift in the nature of work justifies a major-version revision.

8. Limitations and Caveats

The Agentic Day framework is a practical measurement tool, not a precise scientific instrument. The following limitations should be understood by adopters:

8.1 Calibration Precision

The v1.0 equivalence is based on empirical observation and collaborative research, not controlled experimental studies with statistical confidence intervals. The six-hour anchor is well-supported by general productivity research, but the mapping of agent output to that anchor involves professional judgement. Organisations should treat the AD as an order-of-magnitude planning tool in early adoption, refining precision through accumulated measurement data.

8.2 Task-Type Variation

The universal task-type equivalence (1 AD of code = 1 AD of documentation = 1 AD of testing) is a deliberate v1 simplification. In practice, AI agents may perform substantially better at some task types than others relative to human baselines. Future versions of this framework may introduce task-type weighting, but the v1 approach prioritises simplicity and adoptability over precision.

8.3 Quality Heterogeneity

Because the AD measures raw output, two projects with identical AD counts may have very different effective delivery (EAD) values if their acceptance rates differ. Reporting ADs without the companion EAD metric can overstate the practical value of agentic delivery. Adopters are strongly encouraged to track and report both.

8.4 The Comparison Problem

The AD enables comparison across time periods, teams, and organisations, but only if all parties use consistent calibration methods and report the same set of metrics. Cross-organisational comparison requires transparency about operational parameters, domain scope, and acceptance-rate measurement methodology. Without this transparency, AD figures from different sources are not directly comparable.

8.5 Evolving Capability

Because the AD is a fixed unit while AI capability is rapidly improving, the cost and wall-clock time required to produce one AD will decrease over time. This is a feature of the design (see §3.6), but it

means that historical AD costs and velocities should not be projected forward without adjustment for expected model improvements.

9. Relationship to Existing Frameworks

The Agentic Day does not seek to replace existing measurement approaches. It occupies a specific niche in the measurement landscape and is designed to be compatible with and complementary to established frameworks:

Framework	Source	Relationship
Digital FTE (D-FTE)	Enterprise/Consulting	D-FTE operates at the headcount level; AD operates at the task/day level. 1 D-FTE ≈ 1 AD per working day at full capacity.
Digital Worker Equivalent	APQC	DWE is a benchmarking ratio for process automation; AD is an output unit for agentic delivery. Complementary.
Agentic Work Unit	Salesforce	AWU is a platform-specific billing primitive (per action); AD aggregates to the task/day level and is platform-agnostic.
METR Time Horizon	METR	METR expresses model capability in human-time units. AD applies the same logic to operational output measurement. Methodologically aligned.
Forrester TEI	Forrester	TEI provides ROI methodology; AD provides the unit of effort input to that methodology. The two are composable.
Minutes Saved	Microsoft	A micro-level metric. AD aggregates and normalises minute-level savings into a planning and estimation unit.
Anthropic Economic Index	Anthropic	Measures AI usage patterns across occupations and tasks. AD applies a similar human-anchor logic at the delivery team level.

The AD is designed to be expressible in terms of any future industry standard that emerges, just as horsepower is expressible in watts. If the field converges on a different base unit, the AD should be translatable into it without loss of meaning.

10. Future Work

The following areas are identified for development in subsequent versions of this framework:

- **Task-type weighting.** Investigating whether domain-specific conversion rates (e.g., code generation at 1.2×, documentation at 0.8×) improve estimation accuracy, and whether the added complexity is justified by the precision gain.
- **Formal calibration studies.** Controlled experiments with time-motion analysis, ideally across multiple organisations and domains, to establish confidence intervals on the six-hour equivalence and to produce domain-specific calibration factors.
- **Acceptance rate benchmarks.** Establishing baseline acceptance rates by task type and domain to make the EAD metric more predictive for engagement estimation.
- **Cross-organisational comparison protocol.** Defining the minimum set of metadata (operational parameters, domain scope, acceptance methodology) required for AD figures from different organisations to be meaningfully compared.
- **Integration with industry standards.** Monitoring convergence with APQC's DWE, METR's time-horizon methodology, and any emerging ISO or industry-body standardisation efforts. If an international standard crystallises, the AD should be mappable to it.
- **Non-software domains.** Extending the calibration and validation of the AD beyond software engineering into adjacent knowledge-work domains: legal analysis, financial modelling, scientific research, and professional writing.

11. Conclusion

The Agentic Day is a simple idea with a specific purpose: to give organisations a shared, intuitive unit for talking about what AI agents produce, expressed in terms they already understand and price.

It is not the only possible unit, and it makes no claim to be a precision instrument. It is a bridging metric, designed to occupy the same role that horsepower played in the transition from animal to mechanical power: an imperfect but immediately useful translation layer that allows the new technology to be planned, communicated, and valued using the language of the old.

The unit is deliberately simple (one number, one anchor), deliberately conservative (six hours, not eight), deliberately fixed (immune to model-version inflation), and deliberately open (available for any organisation to adopt and adapt). Its value will ultimately be determined not by its theoretical elegance but by whether practitioners find it useful in the daily work of planning, delivering, and communicating the value of AI-augmented work.

The history of measurement suggests that the units which endure are those that are anchored to human experience, simple enough to explain in one sentence, and honest enough to survive scrutiny. The Agentic Day aspires to meet those criteria.

Appendix A: Quick Reference

The Agentic Day at a Glance

1 AD = 6 human-productive-hours of equivalent output 24 hours continuous operation = 4 ADs per agent 3 agents × 24 hours = 12 ADs ≈ 2.5 developer-weeks Model-agnostic | Raw output | Fixed unit | Per-agent | Open framework

Derived Metrics Quick Reference

Metric	Formula	What It Tells You
Effective AD (EAD)	AD × Acceptance Rate	How much agent output was actually usable
Cost per AD (CPAD)	Total cost ÷ ADs	Unit economics vs. human day-rate
AD Velocity	ADs ÷ calendar period	Team throughput in human-equivalent terms
Augmentation Ratio	(AD + HD) ÷ HD	Force-multiplier effect of agentic delivery
Wall-Clock Efficiency	ADs ÷ clock hours	How effectively time is converted to output

Appendix B: Glossary

Term	Definition
Agentic Day (AD)	The base unit: productive output equivalent of 6 human-productive-hours, delivered by AI agents.
Agentic Run	A single execution of an AI agent within defined parameters (max turns, timeout). Multiple runs compose an AD.
Acceptance Rate	The proportion of raw agent output accepted by human reviewers without significant rework.
Effective Agentic Day (EAD)	Quality-adjusted output: AD multiplied by acceptance rate.
Cost per AD (CPAD)	The total compute and tooling cost to produce one AD.
AD Velocity (ADV)	ADs delivered per calendar period (week, sprint, month).
Augmentation Ratio (AR)	The force-multiplier: total output (AD + human-days) divided by human-days alone.
Wall-Clock Efficiency (WCE)	ADs produced per calendar hour of agent operation.
Human-Productive-Hour	One hour of focused, productive output by a skilled practitioner, excluding meetings, admin, and overhead.
Orchestration	Human effort to prompt, configure, review, and integrate agent output. Measured separately from ADs.
Fixed-Unit Principle	The design decision that 1 AD always equals 6 human-productive-hours, regardless of model improvement.

Bridging Unit	A unit of measurement that translates a new technology's capability into terms anchored to the technology it displays.

E N D O F P A P E R