CURLY'S CONSULTING / WHITE PAPER 02

Monetizing AI without guessing

Ammanuel "Tom" Selameab, CPP

Certified Pricing Professional

Curly's Consulting

February 2026

How to connect workflow value, buyer demand, delivery cost, and a commercially usable offer.

Abstract

An AI feature can be impressive, useful, and commercially weak. The gap appears when technical performance is mistaken for customer value, customer value is mistaken for budget, or budget is mistaken for willingness to buy a particular offer.

The monetization decision must connect four things: a workflow the buyer wants to improve; an outcome that can be demonstrated; an offer the buying organization can understand and purchase; and delivery economics that remain viable as usage changes. Model cost is one input. It does not determine how much value the customer receives or what charging structure the customer will accept.

For investors and operators, the practical objective is to learn which AI capabilities create incremental contribution, including their effects on the existing product. A paid add-on that cannibalizes core seats or imposes large support costs may produce less value than a capability included in a differentiated premium package.

The decision this paper supports

Choose an AI offer only after testing the workflow, economic buyer, price structure, and cost distribution together. Use an explicit value and contribution model to decide whether to bundle, charge separately, meter usage, or continue learning.

1 Define the workflow before naming the AI product

"An AI assistant" is a technology description. "Draft a first-pass response to a complex support case using approved company material" describes an activity with a user, a starting condition, and an observable result. The second description supports research into time saved, quality, supervision, and willingness to adopt.

Map the workflow from trigger to accepted output. Identify the person doing the work, the person accountable for the result, the person approving expenditure, and the systems involved. These may be different people. A user can love an assistant while the budget owner sees no reduction in cost or improvement in throughput.

Define the alternative the customer would actually use. It may be the current manual process, an incumbent software feature, an outsourced service, a general-purpose AI tool, or choosing not to do the work. Value is incremental to that alternative. Comparing the AI product with an artificially expensive manual process exaggerates the opportunity.

Distinguish task performance from production performance. Peng and colleagues' controlled experiment reported substantially faster completion of a specified JavaScript programming task with GitHub Copilot. The study concerns that task and experimental setting; it does not establish an equivalent percentage improvement in an enterprise's total software-development costs.[1] The monetization implication is to measure the customer's workflow rather than importing a headline productivity number.

2 Build a value model the buyer can challenge

Begin with observable quantities: workflow frequency, current handling time, fully loaded labor cost, required review, error rates, and adoption. Then identify whether the outcome releases cash, increases capacity, improves quality, or reduces risk. Those are different benefits with different evidence requirements.

Time saved is not automatically labor expense removed. If a team saves five minutes on each case but keeps the same staffing and does not complete additional valuable work, its cash expense may remain unchanged. The value can still be real, but it is capacity or employee experience rather than a booked labor saving.

Use the following structure for one workflow:

$$V_{\text{gross}} = N\,a\,s\,t\,w, \tag{1}$$

$$V_{\text{net}} = V_{\text{gross}} - C_{\text{review}} - C_{\text{error}} - C_{\text{change}}. \tag{2}$$

Here N is the number of eligible tasks, a adoption, s the share of attempted tasks with usable output, t hours avoided per successful task, and w the relevant hourly value. The cost terms capture review, expected error consequences, and workflow change. State whether w represents avoidable cost or the value of redeployable time.

Avoid adding benefits that describe the same improvement. A faster support process may reduce handling time and increase case capacity; counting both as independently realized benefits can double-count the saved time. Revenue benefits require their own conversion mechanism, such as additional qualified work completed and an evidenced probability of sale.

3 Research the purchase, not just the reaction

Interviews should investigate a recent instance of the workflow before presenting the concept. Ask what happened, what it cost, what was difficult, what alternatives were tried, and who authorized spending. Specific behavior helps distinguish a recurring problem from a polite response to a compelling demonstration.

Then test a concrete offer: eligible work, expected output, limits, review responsibilities, price, billing unit, and implementation requirements. Ask the respondent to choose among realistic alternatives, including the current process. A feature rating does not require a budget tradeoff. An offer choice does.

Recruit both users and economic buyers. Include accounts that rejected the core product, accounts with strong internal alternatives, and low-adoption customers where relevant. Research limited to enthusiastic design partners can identify valuable use cases while overstating broad demand. Record incentives or special terms given to participants.

Test the credibility of the value claim separately from the price. A buyer who does not believe the output will be reliable is not revealing a clean reservation price for a trusted product. The intervention may need better evidence, a narrower task, or a service commitment before price optimization becomes meaningful.

4 Choose the offer architecture deliberately

There are several defensible commercial structures. Their suitability depends on how value and cost vary, who buys, and whether the capability changes demand for the existing product.

Structure	When it merits testing	Main economic question
Included capability	Strengthens core differentiation or adoption	Does core retention or price realization pay for delivery?
Premium package	AI belongs to a distinct, valuable workflow	Will buyers upgrade rather than receive a costly free feature?
Separate add-on	A subset has a distinct use case and budget	Is incremental willingness to pay sufficient after cannibalization?
Usage-based charge	Usage is understandable and tracks value or cost	Will variable bills suppress useful adoption?
Base fee plus allowance	Buyers want predictability and usage varies	Do included units, overages, and commitments support contribution?
Outcome-based charge	Outcomes are observable and attributable	Who bears failures, disputes, and risks outside the product's control?

The label "AI" is not a reason to create an add-on. Conversely, including a feature does not mean it has no monetization role. It may support a premium tier or protect a broader workflow. Model these paths against the same baseline, including expected changes in core-seat counts and existing service revenue.

Outcome pricing requires operational definitions. For a support product, is a resolution a closed ticket, a customer-confirmed answer, or a case that stays closed for seven days? What happens after reopening or human intervention? If buyer and vendor cannot reproduce the same count, the billing unit is premature.

5 Model the distribution of delivery cost

Two customers with the same number of requests can generate different costs because their context length, output length, model routing, retrieval, tool calls, retries, and review requirements differ. Build cost telemetry at the workflow and account level rather than relying on one blended average.

Historical supplier announcements illustrate why cost is an engineering and usage variable. Anthropic's March 2025 announcement described prompt caching and token-efficiency changes that could alter cost and throughput.[3] It does not establish a universal saving for an application; actual economics depend

on request patterns and implementation. Supplier cost reductions can improve margin without proving a higher customer willingness to pay.

For an account with billable usage q, price function $B(q)$, variable delivery cost $D(q)$, and incremental account-support cost S, contribution is:

$$\text{Contribution}(q) = B(q) - D(q) - S.$$

Inspect the middle and tails of the distribution. A flat fee may look profitable at median usage while losing money on the most active accounts. Heavy users may also be the customers receiving the most value, making a clear usage ladder more appropriate than silently restricting the product.

Avoid unlimited commitments before measuring plausible high-usage conditions. Test long contexts, repeated failures, exceptional review needs, and deliberate high-volume use. Build alerts and a defined capacity or commercial response. The customer should know which consumption is included and how additional consumption is handled.

6 Worked example: value and vendor margin disagree

Illustrative assumptions. A customer handles 6,000 eligible tasks each month. A successful AI-assisted task avoids three minutes of work valued at $40 per hour. Adoption is 60 percent and 80 percent of attempted tasks produce usable output. Gross monthly time value is:

$$6{,}000 \times 0.60 \times 0.80 \times \frac{3}{60} \times \$40 = \$5{,}760.$$

Each of the 3,600 attempted tasks requires half a minute of review, costing $1,200 in time value. Allow $400 for expected error and change costs. Net estimated workflow value before the product fee is $4,160 monthly. This is a capacity-value estimate unless the buyer can show how the time converts into cash savings or additional output.

Consider an add-on priced at $1,200 per month. For the vendor, assume an illustrative $0.08 cost per attempted task, $160 in account-level workflow infrastructure, and $100 in incremental support. Monthly contribution is:

$$\$1{,}200 - (3{,}600 \times \$0.08) - \$160 - \$100 = \$652.$$

That is a 54.3 percent contribution margin before development and shared operating costs. These cost inputs are synthetic and are not vendor API prices.

At 18,000 attempts under the same flat fee and other assumptions, contribution becomes negative $500. Attractive customer value does not make the vendor's offer economically sound at every usage level. A usage allowance, a higher-volume tier, a lower-cost workflow, or a different scope may be necessary. Buyer research must test the revised offer because the commercial change can alter adoption.

7 Price the promise the product can reliably deliver

An AI product's value depends on accepted output, not merely generated output. Define quality gates and failure handling before choosing the billing event. Charging for failed attempts can produce a mathematically attractive revenue model that buyers find difficult to trust.

NIST's Generative AI Profile identifies risks including confabulation, privacy, and information security and provides a framework for managing them.[2] The commercial inference is that review, assurance, and operating controls can be part of the cost and purchase requirement. A demonstration that omits them may overstate both customer value and vendor margin.

Research which assurances matter to the target buyer and which are actually feasible. Do not sell an outcome guarantee that relies on unobserved customer behavior. Where the outcome cannot be guaranteed, a narrower commitment to workflow scope, response time, or acceptance criteria may be more credible and measurable.

8 Use a paid pilot to resolve a named uncertainty

A pilot should answer a specific question: Will buyers pay for this workflow? Does actual accepted output match the value estimate? Are variable costs sustainable? Will the add-on reduce core-seat revenue? Define the question and the observation needed before recruiting participants.

Record activation, attempted and successful tasks, human review, usage distribution, price realization, support effort, renewal intent, and actual purchase continuation. Where a credible comparison is feasible, measure workflow outcomes against a baseline or randomized access rather than relying only on user recollection.

Set expansion, revision, and stop criteria. For example, require a minimum acceptable contribution under a high-usage scenario, verified buyer understanding of the bill, and demonstrated workflow benefit at the tested price. These thresholds are company decisions, not universal industry benchmarks.

A successful monetization strategy makes the relationship among value, price, and cost observable. It gives the buyer a reason to purchase and the operator a reason to scale. The AI capability earns its commercial role through that relationship, rather than through the novelty of the technology.

9 Buy evidence about the workflow and the offer

An AI pricing decision requires evidence that a demonstration cannot provide: who owns the problem and budget, what the customer actually gains after review, which offer the buyer will choose, and whether contribution survives the resulting usage. Competitor prices and model-provider costs answer only fragments of that question.

The research brief should specify the workflow, economic buyer, alternatives, candidate package and charging unit, expected bill, and the operating data needed to estimate delivery cost. The financial model should then connect buyer choices with adoption, accepted output, core-product effects, and contribution. These are inputs to be measured, not assumptions that become reliable because the spreadsheet is detailed.

Curly's Consulting's **Pricing Transformation** combines size-of-prize modeling, buyer research and willingness-to-pay analysis, monetization design, and support executing the highest-value opportunities. For an AI offer, the relevant scope should be built around the workflow and unresolved commercial decisions, including the cost evidence the operating team must supply.

The next commercial step

Request a Pricing Transformation quote. Describe the AI workflow, target buyer, proposed offer, and evidence already available so the engagement can be scoped around the decision.

Sources and methodology

This paper combines the cited published work with Curly's Consulting's analytical framework. The commercial examples, thresholds, and calculations are illustrative; they are not client results or findings from a new empirical study. Published studies support the specific claims beside their citations. Their results should not be treated as universal benchmarks for a portfolio company.

References

[1] Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*. Research paper. https://arxiv.org/abs/2302.06590.

[2] Autio, C., et al. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. https://doi.org/10.6028/NIST.AI.600-1.

[3] Anthropic (2025). *Token-saving updates on the Anthropic API*. March 13. Historical product announcement, used to illustrate cost drivers rather than current vendor rates. https://www.anthropic.com/news/token-saving-updates.

About the author

Ammanuel "Tom" Selameab, CPP, is a Certified Pricing Professional and the founder of Curly's Consulting. He helps investors and operators make pricing and monetization decisions through primary buyer research, financial modeling, and commercial analysis.
tom@curlysconsulting.com | curlysconsulting.com

© 2026 Curly's Consulting. All rights reserved.