A Budget-Based Mechanism for Regulating Student Use of LLMs

Matthew Murphy*
Rakesh Vohra†

May 11, 2026

1 Introduction

The unrestricted use of LLMs may undermine three of the central objectives associated with education:

1. **Human capital formation.** Students acquire skills and knowledge through repeated cognitive effort.
2. **Productivity enhancement.** Students produce essays, code, proofs, and analyses that may be improved by AI assistance.
3. **Signaling.** Grades, assignments, and credentials communicate information about student ability to employers, graduate schools, and society.

The challenge is that LLMs may increase productivity while simultaneously reducing incentives for independent skill acquisition and degrading the informational content of educational signals. We propose a mechanism for regulating AI use to counter this that does not rely on outright prohibition. Instead, students are assigned a finite *AI Assistance Budget* over the course of their degree program. The mechanism introduces scarcity into AI usage decisions and encourages students to allocate AI assistance strategically across assignments and courses.

2 A Tax on AI Use

We propose that educational institutions impose a price on AI assistance. The objective is not to prohibit AI but to discourage excessive or routine dependence that undermines:

1. skill acquisition,
2. meaningful differentiation across students,
3. and the informational value of educational credentials.

We outline how such a budget-based pricing mechanism might work.

Each student is endowed with a finite number of AI credits for the duration of their degree program. For example:

*Department of Economics, University of Pennsylvania. Email: mjhm@sas.upenn.edu.
†Department of Economics & Department of Electrical and Systems Engineering, University of Pennsylvania. Email: rvohra@seas.upenn.edu.

- 1,000 credits over four years (or some appropriate number per course),
- non-transferable across students,
- potentially expiring or depreciating over time.

The scarcity of credits forces students to make trade-offs regarding when AI assistance is most valuable. AI usage could be classified into tiers.

- **Tier 0 (0 credits):** No AI assistance.
- **Tier 1 (1–3 credits):** Light assistance such as brainstorming, outlining, or debugging hints.
- **Tier 2 (5–10 credits):** Moderate assistance such as partial drafts, code scaffolding, or literature summaries.
- **Tier 3 (15–25 credits):** Heavy assistance such as full draft generation, derivation of solutions, or substantial rewriting.

The institution provides AI access through a sandboxed deployment—a dedicated interface running an approved model—which can log all interactions by design. This does not prevent the 'shadow use' of AI from other sources. To counter this, students must declare the level of AI usage associated with each submission. Each assignment submission might include:

1. **AI Usage Declaration** A statement describing the level and type of AI assistance used.
2. **Process Artifacts** Examples include:
    - selected prompt histories,
    - draft versions,
    - handwritten derivations,
    - version-control logs,
    - notes describing revisions.
3. **Reflection** Students briefly explain:
    - what the AI system contributed,
    - what errors or limitations it exhibited,
    - and what modifications were made.

Shadow use means this supporting documentation will not be logged automatically and must be manufactured. Because shadow AI use cannot be prevented, the system relies on conventional examinations and random oral audits. Audited students may be asked to:

- explain arguments,
- reproduce derivations,
- modify code or proofs,
- answer conceptual questions.

Under this structure, the mechanism does not need to detect shadow use perfectly. It needs only to make institutional use the lower-risk option for students who are uncertain about audit outcomes. Students most likely to shadow use heavily are also those least able to demonstrate comprehension under oral examination, so audit risk increases the expected cost of concealing heavy AI dependence. Section 4 develops the role of oral audits in responding to the specific challenge of shadow AI use.

To repeat, the objective is not perfect surveillance but rather:

- reducing routine overreliance,
- encouraging selective and strategic use,
- preserving incentives for independent work,
- and maintaining informational content in educational signals.

3 Determining the Tax

Several approaches are possible.

1. **Input-Based Pricing:** Credits may depend on:
   - prompt counts,
   - token usage,
   - interaction time.

   This approach is simple but may poorly measure actual cognitive outsourcing.

2. **Output-Based Pricing:** Pricing may depend on the amount of AI-generated content retained in the final submission. This better captures substitution but remains imperfect. In particular, accurate estimation of token use from final output and supporting materials faces significant limitations. Detection tools achieve reasonable precision on unedited AI-generated text, but accuracy degrades sharply once students edit the output—which is precisely the scenario that corresponds to Tier 2 and Tier 3 use. More fundamentally, token consumption is a function of interaction strategy rather than output length: the same submission could reflect either a tightly prompted single query or an extended iterative exchange, with no reliable way to distinguish the two from the final product alone. Output-based pricing, therefore measures a noisy proxy for the underlying variable of interest.

3. **Capability-Substitution Pricing** A more ambitious approach attempts to estimate:

   how much human cognitive effort was displaced by AI assistance.

   Under this approach:
   - grammar correction receives a low price,
   - conceptual synthesis receives a higher price,
   - derivation of proofs or core arguments receives the highest price.

4. **Dynamic Pricing:** Prices may vary by:

- course level,
- assignment type,
- proximity to examinations,
- or pedagogical objectives.

For example, AI assistance may be priced more heavily in introductory courses where foundational skills are developed.

4 The Shadow Use Problem

Shadow use, access to external AI systems outside institutional channels, is the central vulnerability of the proposed mechanism. A student who declares Tier 0, spends zero credits, and achieves Tier 3 output via an external tool imposes no credit cost on themselves while potentially producing high quality work with no effort. If widespread, shadow use would effectively tax students seeking to improve human capital, producing a deeply perverse outcome.

The mechanism cannot eliminate shadow use. The relevant question is whether it can raise the expected cost of shadow use sufficiently to alter the distribution of behavior. We argue that it can, through a combination of four instruments.

Oral Audits with Endogenous Targeting

Shadow use and comprehension are negatively correlated: a student who chooses to outsource thinking will underperform in a conventional or oral examination of their own submissions. Audits, therefore, serve as the primary enforcement tool, but their effectiveness depends critically on how audit targets are selected.

The naive approach—auditing students who declare high AI use—is incorrect. Students declaring Tier 0 on complex assignments are the ones most likely to have engaged in shadow use. More generally, the audit probability should be *increasing* in the implausibility of the declared tier relative to submission quality. Concretely:

- Submissions of high apparent sophistication carrying Tier 0 or Tier 1 declarations should face elevated audit probability.
- Audit questions should probe exactly the reasoning the submission claims the student performed independently, making the performance gap between genuine and outsourced work as large as possible.
- Audit failure might trigger retroactive credit penalties on prior submissions, raising the stakes beyond any single assignment.

Audits need not detect every case of shadow use. They need only make the expected cost of shadow use high enough relative to the benefit. This is standard: the requisite condition is sufficient deterrence, not perfect detection. Note this assumes the institution, through its faculty, has the will and credibility to hold students who fail an audit to account.

Stylometric Baselines

Each student generates authenticated writing over the course of a degree—examinations, in-class writing, oral transcripts. This body of work constitutes a personal stylometric baseline against which submitted work can be compared. Large deviations from a student's baseline—in sentence length distributions, lexical diversity, syntactic complexity, or function word patterns—flag submissions for audit without requiring any determination of token counts or AI tool identity.

The baseline approach has two advantages over population-level AI detection tools. First, it is self-referential: comparisons are made against the individual student's own authenticated output rather than against a general population norm, reducing the risk of differential impact on non-native speakers that has been documented with existing detection tools. Second, baseline precision improves over time as more authenticated samples accumulate, making the approach more powerful in later years of the degree program.

Stylometric flagging should function as an audit-targeting instrument rather than a primary enforcement mechanism. A deviation from baseline raises audit probability; it does not itself constitute a violation.

Limits of Deterrence

No combination of these instruments eliminates shadow use for a determined student with careful editing habits and awareness of their own stylometric profile. The mechanism just needs to ensure that sincere students are not severely disadvantaged relative to dishonest ones, and that the expected return to heavy shadow use is low enough that most students find genuine engagement the path of least resistance.

5 Conclusion

LLMs are likely to become permanent features of educational environments. Consequently, educational institutions require mechanisms that go beyond simple prohibition. We propose a budget-based framework in which students receive finite AI assistance budgets over the course of their degree programs. The mechanism is motivated by the concern that unrestricted AI use degrades the informational value of educational credentials.

The proposal accepts that hidden AI use cannot be completely eliminated. The shadow use problem—access to AI systems outside institutional channels—is the central vulnerability of the mechanism, and addressing it requires a coordinated response across four instruments: endogenous audit targeting, stylometric baseline monitoring, institutional AI provisioning, and assignment redesign. No single instrument is sufficient; in combination, they raise the expected cost of shadow use without requiring perfect detection.

The proposal therefore seeks to:

- preserve incentives for independent skill acquisition,
- maintain informative educational signals,
- and encourage strategic rather than routine AI dependence.