From Competition to Collaboration: Designing Sustainable Mechanisms Between LLMs and Online Forums*

Niv Fono
Technion—Israel Institute of Technology
Haifa, Israel
niv.fono@campus.technion.ac.il

Yftah Ziser
University of Groningen
Groningen, Netherlands
y.ziser@rug.nl

Omer Ben-Porat
Technion—Israel Institute of Technology
Haifa, Israel
omerbp@technion.ac.il

ABSTRACT

Background: As users increasingly turn to Large Language Models (LLMs) over traditional Question-and-Answer (Q&A) forums, these platforms face a critical decline in engagement. This creates a paradox: while Generative AI (GenAI) systems draw users away from forums, they also depend on the very data those forums produce to improve their performance.
Objectives and Research Questions: We aim to address this problem by modeling the ecosystem as a strategic game where GenAI agents propose questions and forums select which to publish. We investigate whether their objectives actually align and aim to design non-monetary, incentive-aware mechanisms that support collaboration while respecting the autonomy of both parties.
Methods: We employ a game-theoretic framework evaluated via data-driven simulations. Using real-world Stack Exchange data and open-source LLMs (LLaMA, Pythia) as agents, we simulate the complex interplay between forum and GenAI interests.
Results: We find a fundamental conflict: GenAI learning goals rarely align with what drives forum engagement. Despite this, our mechanism bridges the gap, recovering 56–66% of forum utility and 46–52% of GenAI utility compared to an ideal full-information benchmark.
Conclusions: Sustainable collaboration between AI systems and human platforms is possible, even with asymmetric information and under constrains. We show that carefully designed mechanisms can preserve the knowledge-sharing ecosystem even without the need for financial incentives.

KEYWORDS

Generative AI, Strategic Behavior, Collaboration

ACM Reference Format:
Niv Fono, Yftah Ziser, and Omer Ben-Porat. 2026. From Competition to Collaboration: Designing Sustainable Mechanisms Between LLMs and Online Forums. In *Appears at the 8th Games, Agents, and Incentives Workshop (GAIW-26). Held as part of the Workshops at the 25th International Conference on Autonomous Agents and Multiagent Systems., Paphos, Cyprus, May 2026*, IFAAMAS, 10 pages.

*This work was also accepted to AAMAS 2026 as an extended abstract.

Appears at the 8th Games, Agents, and Incentives Workshop (GAIW-26). Held as part of the Workshops at the 25th International Conference on Autonomous Agents and Multiagent Systems., Armstrong, Curry, Hosseini, Mattei, Tsang, Wąs (Chairs), May 2026, Paphos, Cyprus. …$ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

1 INTRODUCTION

Online Q&A forums are witnessing a sharp decline in user participation, as people post fewer questions and answers than before. Stack Overflow, for example, experienced a 25% drop in posts within months of ChatGPT's release-a striking indicator of this trend [10]. Instead of contributing to community-driven platforms like Stack Overflow or Mathematics Stack Exchange, users increasingly turn to generative AI systems for instant answers [25]. While this shift offers individual convenience, it carries systemic risks: Q&A forums generate much of the high-quality, human-produced knowledge that large language models depend on for training [4, 21], evaluation [9], and benchmarking [26]. A sustained decline in forum activity may therefore weaken the very epistemic foundations of GenAI, creating a negative feedback loop where AI systems erode the human knowledge sources they rely on. Ultimately, this erosion could not only risk AI progress but also threaten the survival of these online communities, depriving users of open, collaborative spaces for trustworthy knowledge exchange.

Existing responses to this sustainability challenge remain narrow in scope and insufficiently aligned with the structural realities of the ecosystem. Proposals that focus on restricting data access by GenAI companies [3] or introducing simple compensation or incentive schemes [13] tend to treat the problem as one of resource extraction or fairness accounting. While well intentioned, such measures primarily operate at the level of surface symptoms rather than addressing the deeper strategic interdependence between the actors involved. These approaches tent to frame GenAI firms and online knowledge communities as adversarial stakeholders in competition over data ownership or user attention, rather than as co-dependent entities within a shared knowledge base. Recent works have begun to recognize the structural nature of this interdependence. For example, Taitler and Ben-Porat [34] demonstrate that uncoordinated adoption of GenAI technologies can lead to a Braess-like paradox in which both AI development and community vitality suffer in the long run due to declining user participation. However, even such analyses typically stop short of modeling online forums as fully strategic agents with their own utility functions, governance norms, and content curation policies. This kind of models treat forums implicitly as reactive rather than proactive. But platforms often pre-commit to moderation policies or content curation schemes as strategic tools [32].

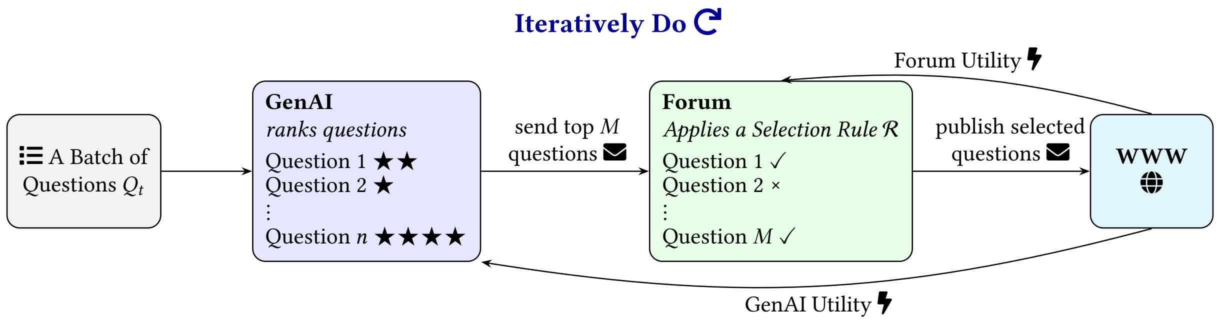


Figure 1: Iterative interaction between a GenAI provider and an online Q&A forum. In each round, the GenAI ranks user-generated questions by their expected model-learning value and submits the top M to the forum. The forum then applies its own selection rule $\mathcal{R}$, publishing only those questions that align with its community objectives. Feedback from the published posts informs both sides in subsequent rounds.

To address this challenge, we propose a form of collaboration that protects the interests of both parties without involving monetary exchange. Specifically, we consider an exchange of goods: GenAI systems could submit queries that they fail to resolve to Q&A forums, where human experts can provide high-quality answers. This exchange benefits both sides: GenAI companies gain access to expert-level supervision data to improve their models, and forums regain traffic and engagement through challenging, high-value questions. Recent research analyzing Stack Overflow data in the post-ChatGPT period [18] shows that while overall user contributions have declined, the complexity and difficulty of questions have increased, indicating that users are already turning to forums for queries that LLMs cannot handle effectively. Our proposed collaboration leverages this natural gap, channeling questions beyond the LLM's capabilities to human experts, thereby amplifying the value of the exchange. Because GenAI firms are unlikely to disclose their internal strategies and forums may reveal only partial information about their curation or moderation policies, such collaboration must be asymmetric by design. We therefore introduce a novel framework that enables both sides to preserve confidentiality while still engaging in a mutually beneficial interaction. In our model, illustrated in Figure 1, this interaction is formalized as a strategic collaboration in which, within an iterative game, the GenAI company proposes candidate questions for publication and forums decide whether to accept or reject them. Each actor optimizes its own objective: GenAI companies seek questions that maximize their model improvement potential, while forums prioritize those that enhance community utility. Importantly, our framework is **agnostic** to the specific definitions of these utilities, allowing for flexible and evolving notions of value. Implementing such a system, however, requires careful modeling of incentives, disclosure protocols, and trust assumptions, which we address in detail in Section 2.

To assess the consequences of these strategic interactions, we simulate a highly realistic game in which the two players are instantiated using real-world analogues. The **GenAI side** is represented by a set of large language models, including open-source systems such as LLaMA 3.1 and Pythia, while the **Q&A side** is modeled using data from several Stack Exchange communities—*Stack Overflow*, *Mathematics Stack Exchange*, *English Stack Exchange*, *Ask Ubuntu*, and *LaTeX Stack Exchange*. This setup allows us to capture both the linguistic and epistemic diversity of real interactions between generative models and human-driven knowledge ecosystems. Our simulations reveal that the epistemic value of forum data for GenAI systems is systematically misaligned with community engagement. Questions that are most beneficial for model learning often differ from those that attract high participation or satisfaction within the forums. This divergence highlights a central tension: while GenAI agents aim to maximize learning efficiency, forums strive to sustain community vitality and relevance. Even under full-information conditions—where all available data and objectives are transparently shared between the GenAI player and the forums, our **idealized simulations** show that disagreement may still emerge regarding the optimal collaboration strategy.

To further explore the feasibility of a sustainable cooperation, we simulate a range of counterfactual collaboration schemes, including greedy baselines and our proposed learned bargaining strategy. The results indicate that mutually beneficial outcomes are achievable: GenAI systems can obtain expert-level data that enhances their learning. At the same time, forums experience improved engagement metrics, all without requiring monetary exchange or complete transparency. Quantitatively, our learned mechanism achieves approximately 46–52% of the GenAI-side utility while preserving 56–66% of the forum-side utility, under full cooperation. Across all examined models, the learned strategy consistently improves results, yielding gains of up to 23% in GenAI utility relative to baseline allocations. Together, our findings demonstrate that a strategically-aware cooperation between GenAI systems and Q&A communities can produce substantial joint value, offering a viable path toward mutually reinforcing knowledge exchange.

Our contribution. Our work makes several key contributions toward understanding and modeling the interaction between generative AI systems and Q&A forums:

- **Realistic collaboration design.** We survey the key dimensions that shape GenAI–Forum collaboration and distill three guiding principles for realistic interaction: (i) prohibiting monetary transfers to preserve community trust and autonomy, (ii) addressing intrinsic incentive misalignment between model improvement potential and community engagement value, and (iii) formalizing the collaboration as a process in which GenAI proposes questions and the forum determines publication. These principles, discussed in Section 2, capture the social, economic, and strategic constraints that motivate our later framework.

- **Game-theoretic framework.** We develop a game-theoretic framework that models GenAI providers and Q&A forums as strategic, interdependent agents—each optimizing distinct yet interconnected utilities. This formulation reinterprets their interaction as a cooperative rather than adversarial game, encompassing both full-information and asymmetric-information settings and providing a foundation for analyzing stable, mutually beneficial equilibria.
- **Data-driven game simulation.** We empirically instantiate the proposed framework through large-scale simulations of a realistic GenAI–Forum game, using real data from multiple Stack Exchange communities and open-source LLMs. Our analyses reveal a systematic misalignment between GenAI's improvement potential (perplexity) and forum engagement value (view counts), confirming that their interaction constitutes a genuine strategic game rather than an optimization problem. Across extensive data-driven experiments, our learned acceptance-aware mechanism recovers over half of the ideal full-information utility for both agents—achieving 46–52% of GenAI's learning potential and 56–66% of forum engagement—without disclosing private information. These findings demonstrate that adaptive, game-theoretic interaction can produce substantial mutual gains even under asymmetric information.

Together, these contributions establish both a theoretical and empirical foundation for sustainable, non-monetary collaboration between GenAI systems and human knowledge ecosystems.

2 GENERAL COLLABORATION GUIDELINES

Before introducing our formal framework, we first outline several key principles that guide its design. These properties capture essential aspects of GenAI-Forum collaboration and provide the conceptual foundation on which the model is built. Presenting them upfront clarifies the reasoning behind our assumptions and modeling choices, and highlights the structural challenges that any sustainable collaboration mechanism must address. Each of the following subsections focuses on one such consideration and explains how it is reflected in the formal setting.

2.1 No Monetary Transfers

Markets are often viewed as efficient mechanisms for resolving conflicts of interest through monetary exchange. A natural question therefore arises: why not allow GenAI companies to pay online forums to publish their questions? While this arrangement may seem straightforward in theory, it overlooks deeper structural and strategic challenges inherent to such collaborations. In practice, introducing payments would distort community incentives, compromise forum autonomy, and create long-term instability in the collaboration. We highlight three fundamental reasons:

(1) **Erosion of community trust.** Forum users contribute voluntarily, motivated by intrinsic factors such as reputation, reciprocity, and community belonging rather than financial gain. If forums begin accepting payments from AI companies for publishing or answering GenAI-generated questions, users may perceive that their unpaid contributions are being monetized by others. This perception could undermine trust and fairness, prompting users to demand direct compensation or disengage entirely. Since community participation relies on gamified reputation systems rather than financial rewards, introducing money risks replacing intrinsic motivation with extrinsic incentives, ultimately reducing engagement [2, 19].

(2) **Loss of autonomy and value misalignment.** Financial dependence on GenAI companies would expose forums to similar legitimacy risks observed in nonprofit-corporate partnerships [20, 30]. When a community-driven platform becomes financially tied to a powerful corporate actor, its decisions may gradually align with the sponsor's interests rather than its own mission. This power asymmetry can challenge the forum's core values, leading to subtle forms of self-censorship or content bias. Over time, such influence erodes community trust and transforms the forum from an independent public knowledge space into a dependent, co-opted partner.

In the economic literature, settings with transferable utilities are typically easier to handle than non-transferable utilities (see, e.g., [31]). We do not argue that all GenAI-Forum collaboration should rely on non-transferable utilities, but rather that such should be explored.

2.2 Incentive Misalignment

At first glance, the objectives of GenAI and the Forum appear well-aligned. Both parties benefit from the publication of well-posed, interesting questions that attract community engagement and yield high-quality answers. Such questions contribute to the Forum's vitality and simultaneously provide GenAI with valuable external feedback that can inform model improvement. However, a closer examination suggests that the two sides evaluate their value along different dimensions. For the Forum, the primary objective is to sustain community participation and content quality—favoring questions that are clear, accessible, and likely to elicit thoughtful human responses. For GenAI, by contrast, value arises from informational gain: questions that expose uncertainty or challenge the model's current knowledge state are most useful, even if they are niche, ambiguous, or attract limited human engagement. As a result, a question highly valuable to GenAI may not resonate with the community, while the questions most engaging to users may provide little new information to the model. This divergence creates a subtle yet significant misalignment in incentives. While both parties aim to promote productive knowledge exchange, they differ in what kind of questions best achieve that goal. The discussion in this section is theoretical, outlining the expected directions of misalignment that we later substantiate empirically in Section 5.

2.3 Asymmetric Information and Roles

In practice, full transparency between GenAI and the Forum is neither feasible nor desirable. GenAI companies are naturally cautious about revealing too much of their internal processes or decision criteria. The questions they propose for community discussion often reflect areas of uncertainty or weakness in the model. Disclosing the entire set of such questions could inadvertently expose proprietary information, highlight model failures, or reveal sensitive aspects of the system's behavior. Therefore, GenAI must act strategically: selecting which questions to share in a way that balances the benefits of community feedback with the risks of disclosure.

At the same time, the Forum maintains full discretion over what is ultimately published. Its decisions are guided by internal policies that aim to protect the quality and sustainability of the community. These may include prioritizing questions that are likely to attract engagement, ensuring topical diversity, or upholding moderation and fairness standards. The Forum may also limit the frequency or volume of AI-submitted content to prevent user fatigue or preserve the platform's authentic character.

This structure creates an inherently asymmetric form of collaboration: GenAI controls what to propose, and the Forum controls what to publish. Such asymmetry reflects real-world conditions: where cooperation is possible and mutually beneficial, yet bounded by privacy, autonomy, and strategic caution on both sides. We formalize this asymmetric interaction in the next section.

3 PROBLEM FORMULATION

We now introduce the formal asymmetric-information framework. We consider a sequential interaction over T discrete rounds between a GenAI provider and a Q&A forum, which we refer to as Player G and Player F, respectively. Each round includes two stages: in the first stage, Player G proposes candidate questions for publication. In the second stage, Player F curates which of these are ultimately published. Formally, in every round $t \in [T]$,

Stage I. Player G constructs a candidate pool of questions, which we denote by

$$Q_t = \{q_{t,1}, q_{t,2}, \dots, q_{t,n_t}\} \subset \mathcal{Q},$$

where $\mathcal{Q}$ is the space of all possible questions. In practice, Q_t models the set of candidate questions identified by Player G as uncertain, error-prone, or worth external feedback. For instance, this pool may be derived from analyzing past user interactions and selecting those with low implicit satisfaction scores, ambiguous answers, or high model uncertainty—cases where the system would most benefit from human clarification. We abstract the way Q_t is constructed, and assume it is given exogenously.

As discussed in Subsection 2.3 Player G does not wish to share all of its candidate questions with Player F. Therefore, Player G selects a subset $A_t \subseteq Q_t$, where we assume $|A_t| = M \ll |Q_t|$ for some $M \in \mathbb{N}$, and submit A_t to Player F.

Stage II. Player F observes A_t, but does not publish A_t as is. Instead, it uses a *selection rule* $\mathcal{R}$, $\mathcal{R} : 2^{\mathcal{Q}} \to 2^{\mathcal{Q}}$, and selects a subset of published questions S_t such that $S_t = \mathcal{R}(A_t)$. The rule $\mathcal{R}$ represents Player F's internal decision mechanism, reflecting its own objectives and community standards. Here too, we assume Player F is constrained to pick $|S_t| \leq K$ for some $K \in \mathbb{N}$ for every round t.

Utilities. Published questions draw the attention of the user community in Player F, who interact with it and possibly generate feedback. Such interaction and feedback benefit Player F, who aims to optimize KPIs like user engagement measurements, such as views, upvotes, or answers. We keep the actual objective of Player F abstract, and let $u_F(q)$ measure the utility of Player F from each question $q \in \mathcal{Q}$.

Similarly, Player G also derives utility from publishing questions on Player F, as community interactions provide valuable external signals. For example, user-generated clarification comments, alternative answers, or follow-up discussions can serve as high-quality supervision signals that reveal model weaknesses or knowledge gaps. These signals can be incorporated into Player G's training or evaluation pipelines to improve its accuracy, calibration, and domain coverage. Here too, we abstract away the richness of the feedback and quantify the utility of a (published) question $q \in \mathcal{Q}$ by $u_G(q)$.

We assume the utilities are deterministic and are privately known to the respective players. The assumption of deterministic utilities simplifies the analysis by removing uncertainty from the utility space, allowing us to focus on the strategic interaction itself rather than stochastic variability. Furthermore, we extend the utility function to the total utilities of each side such that

$$U_G(\mathbf{A}, \mathcal{R}) = \sum_{t \in [T]} \sum_{q \in S_t} u_G(q), \qquad U_F(\mathbf{A}, \mathcal{R}) = \sum_{t \in [T]} \sum_{q \in S_t} u_F(q),$$

where $\mathbf{A} = (A_1, \dots A_T)$. The utilities are modeled as linear for both simplicity and practical justification. First, linear aggregation enables tractable analysis and transparent comparison between strategies, avoiding the computational complexity of modeling higher-order dependencies. Second, since the number of questions published through Player F represents only a small fraction of the overall data available to Player G, potential interactions between questions, such as redundancy or topical overlap, can be reasonably neglected.

Remark 1. The dependence of the utility functions U_G, U_F on $\mathcal{R}$ and A_t follows from the fact that $S_t = \mathcal{R}(A_t)$. Therefore, Player F can set $\mathcal{R}$ to ensure that the proposed question sets are perceived as fair. For instance, by rejecting the entire set of questions if it seems to disproportionately serve Player G's interests, much like a responder in a dictator game refuses an inequitable offer [14].

3.1 Full-Information Variant

To assess our framework, it is instructive to consider the full-information setting. In this idealized scenario, both Players G and F share complete information about their utilities and candidate questions Q_t. A natural cooperative objective is to maximize the Nash product [29], given by

$$\arg \max_{S \subseteq Q_t, |S|=K} U_G(S) \cdot U_F(S). \tag{1}$$

This formulation captures a Pareto-efficient balance between the interests of both sides, and serves as a theoretical upper bound on achievable joint utility.

However, directly optimizing the Nash product over discrete items is computationally intractable. In particular, we show that

THEOREM 3.1. *Maximizing the problem in Equation* (1) *is NP-hard.*

While maximizing a bilinear function of additive utilities is generally NP-hard (see, e.g., [12]), to keep the paper self-contained, we provide a formal proof for our specific setting in the appendix. Beyond computational hardness, solving Equation (1) also requires complete transparency: both u_G and u_F must be publicly known, and Player G must disclose its entire candidate pool Q_t. These requirements are unrealistic (recall Section 2).

Heuristic alternatives can partially relax these assumptions. For instance, round-robin selection (see Subsection 4.4 for a formal presentation) allows players to alternately choose items based on their marginal utilities, reducing the need for full utility disclosure. Yet, even this approach still requires Player G to reveal Q_t, and its outcome remains sensitive to turn order and strategic manipulation.

4 EXPERIMENTAL SETUP

In this section, we present our experimental methodology.

4.1 Players and Utilities

Player G. To capture GenAI's improvement potential from our framework, we have used several open-source LLMs:

- **Pythia 6.9B.** EleutherAI-pythia-6.9b [8], A fully open *white-box* model with a well-documented training corpus. Its transparency enables us to verify what data it was trained on.
- **LLaMA 3.1 8B.** [28]A *black-box* model for which the training data are not publicly disclosed, representing a common closed-source setting.
- **LLaMA 3.1 8B-Instruct.** [23] An instruction-tuned variant of LLaMA 3.1 8B, optimized for conversational and reasoning tasks, similarly lacking full disclosure of training sources.

This varied set of LLMs allows us to capture a broad range of behaviors and transparency levels, representing both research-oriented and production-like GenAI systems. By comparing open and closed models, we can evaluate whether our framework remains robust under differing information assumptions and access constraints.

We approximate Player G's utility using **perplexity**—a standard measure of model uncertainty—computed over the initial 64 tokens of each question (title plus question content). A higher perplexity value indicates greater model uncertainty [22], which has been shown to be a strong predictor of prompt failure [16] and a key signal for informative data selection in different learning settings [27]. Consequently, questions with higher perplexity provide stronger learning signals, and thus yield greater utility for Player G. Finally, we assume $M = 100$, i.e., $|A_t|$ is at most 100 questions in every round.

Player F. On the forum side, Player F is simulated using real data from Stack Exchange, which serves as a representative example of large-scale, community-driven Q&A platforms. We use the *view count* of each question as a proxy for engagement, as it directly captures the level of human attention and interest that a question attracts. To ensure comparability across interaction rounds and forum domains, view counts are normalized before aggregation. In our simulations, Player F is constrained by a moderation-like capacity limit, allowing at most $K = 50$ GenAI-based questions to be published per round (i.e., $|S_t| \leq 50$). This setting mirrors realistic posting and curation dynamics on Stack Exchange, where community attention and moderation resources are limited.

4.2 Model Selection and Data Integrity

To minimize potential data contamination, we focus on Stack Exchange questions posted in July 2024 to July 2025. This choice ensures that none of the evaluated models could have been trained on the same data. In particular, *LLaMA 3.1 8B-Instruct*, which is the most recent model in our set, was released at the end of 2024, and therefore could not have been exposed to Stack Exchange content from or after July 2024. By combining models with both transparent (*white-box*) and opaque (*black-box*) training data sources, we obtain a balanced perspective on how our framework performs across different levels of model transparency and accessibility.

To instantiate the simulated collaboration scenario, we conducted a longitudinal experiment using real-world Stack Exchange data. Each week—corresponding to one interaction round—we constructed a candidate pool of questions drawn from that period. We then applied the two stages according to players' respective strategies, as we detail shortly in Subsections 4.3 and 4.4. This process was repeated weekly over a one-year horizon, from July 23, 2024, to July 23, 2025.

To ensure diversity in both linguistic structure and topical expertise, we used data from five distinct Stack Exchange communities, sampled in proportion to the popularity of each domain: `math` (54,458 samples), `stackoverflow` (314,949 samples), `ubuntu` (12,622 samples), `english` (2,938 samples), and `latex` (11,441 samples). This cross-domain selection allows our analysis to capture a broad spectrum of question styles, technical subjects, and community behaviors. To obtain a realistic representation of the overall Stack Exchange ecosystem, we combined these datasets into a single corpus, aligning samples across the same date span to reflect genuine temporal activity patterns.

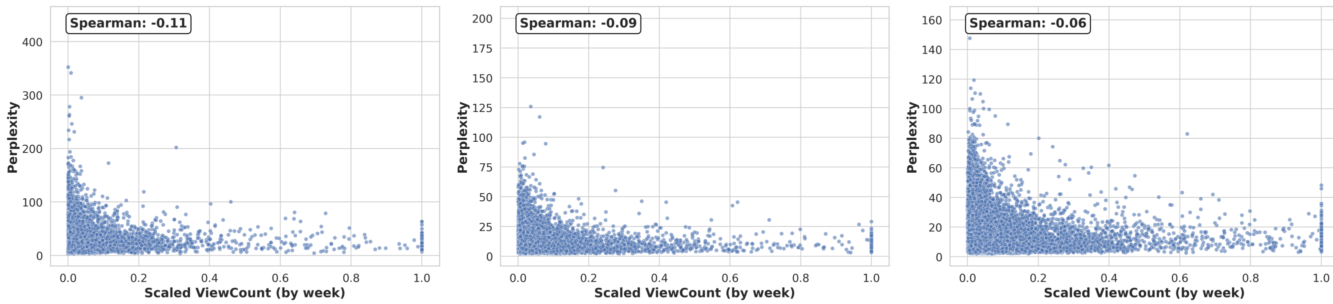
4.3 Strategies

In the asymmetric-information setting, Player G receives a candidate question pool Q_t at each round and extracts from it a subset A_t such that $|A_t| = M = 100$. Player F then applies its selection rule $\mathcal{R}$ to choose the subset of A_t such that the published group of questions $|S_t| \leq K = 50$. Next, we propose several approaches for picking A_t and $\mathcal{R}$. We begin with Player G:

- **G-Greedy** Under this strategy, Player G selects the 100 questions with the highest perplexity scores.
- **G-Utility Maximization** Here, Player G selects the 100 questions that maximize its *expected utility*,

$$\mathbb{E}[U(q)] = u_G(q) \cdot \widehat{\pi}(q),$$

where $\widehat{\pi}(q)$ denotes Player G's estimate of the probability that a given question will be accepted by Player F. To construct these estimates, Player G trains a lightweight classifier that is updated every quarter (i.e., after 13 interaction rounds). Due to limited supervision in early stages, we employ a TF–IDF text representation [33] combined with a Naive Bayes classifier to approximate $\widehat{\pi}(q)$. During the first quarter, when no acceptance history is available, we assume $\widehat{\pi}(q) = 1$ for every $q \in Q$.



(a) llama 3.1-8B-Instract ubuntu **(b) llama 3.1-8B math** **(c) Pythia 6.9B math**

Figure 2: Relationship between question perplexity and normalized ViewCount across five StackExchange domains. Each plot reports the Spearman correlation coefficient ρ. A general pattern of negative correlation emerges, highlighting systematic misalignment between forum engagement and LLM uncertainty.

- **Random** Picking $M = 100$ questions from Q_t uniformly at random.

For Player F, we construct $\mathcal{R}$ as follows. We learn a classifier C and pick $\theta \in \mathbb{R}$ such that

$$S_t = \mathcal{R}(A_t) = \mathrm{Top}_K \{q \in A_t : C(q) \geq \theta\},$$

where $\mathrm{Top}_K\{\cdot\}$ picks the elements with the highest C score in the set.

To pick C, we train a *BERT*-based classifier [11] using data from the Stack Exchange datasets, covering the period from April 23, 2024, to July 22, 2024. View counts are normalized on a weekly basis to account for temporal variation in engagement. Each question is labeled according to its normalized view count percentile: questions in the top 40% are labeled as positive samples ($y = 1$), questions in the bottom 40% as negative samples ($y = 0$), and the remaining 20% are excluded from the training set.

The classifier is trained for three epochs on an NVIDIA RTX A6000 GPU, using the cross-entropy loss function. The resulting model achieves an ROC-AUC of 0.74 on the validation set. Based on validation results, we set the decision threshold θ to $\theta = 0.8110$, yielding a precision of 0.7881 and a recall of 0.3940. The threshold θ is determined to satisfy

$$\text{Precision} = 2 \times \text{Recall}.$$

This calibration reflects Player F's preference for *high-quality acceptance* over quantity. In practical terms, Player F prioritizes precision, ensuring that published questions are consistently engaging—even at the expense of recall. Hence, it is acceptable for Player F to publish fewer questions per round, provided that those selected maintain a high likelihood of generating strong engagement.

4.4 Full Information

We move on to the full-information setting (recall Subsection 3.1). We construct an *idealized experiment* where players engage in perfect cooperation, jointly selecting the subset S_t. Since maximizing Equation (1) is NP-hard as argued before, we evaluate several computationally heuristics to efficiently measure performance:

- **Myopic Round Robin (MPP).** An iterative process that goes as follows. In even rounds, Player G selects a previously-unselected question with the highest u_G score. Similarly Player F selects a previously-unselected question with the highest u_F score in odd rounds. This process continues until the $K = 50$ questions are chosen.
- **Max Sum of Products (MaxSP).** Each question q receives a scalar score $v(q)$ such that $v(q) = u_G(q) \cdot u_F(q)$. We then select

$$S_t = \arg \max_{S \subset Q_t, |S| = K} \sum_{q \in S} v(q).$$

- **Greedy Nash Product (GreedyNP).** We start with $S_t^0 = \emptyset$, and increase it iteratively to obtain S_t^K. In every iteration n, for $n = 1$ till $n = K = 50$, we pick q^n from $Q_t \setminus S_t^{n-1}$ such that

$$q^n \in \arg \max_{q \in Q_t \setminus S_t^{n-1}} \left(U_G(S) + u_G(q)\right) \cdot \left(U_F(S) + u_F(q)\right).$$

That is, GreedyNP iteratively selects the item that yields the largest marginal increase in the Nash product ($U_G \cdot U_F$), updating cumulative utilities at each step.

- **Random** As a control mechanism, we include uniform sampling from Q_t.

5 FINDINGS

5.1 Incentive Misalignment

We observe a consistently weak negative relationship across all evaluated LLMs and Q&A domains between question perplexity and view count (mean Spearman correlation = −0.064, std = 0.039). The correlation is not only far from 1 but hovers around zero, often even negative, indicating that the objectives of GenAI models and Q&A platforms are largely unaligned. If the correlation were close to 1, GenAI systems would naturally select and forward to forums the most engaging and valuable questions for human experts to answer. Instead, this weak association suggests that their interaction forms a genuinely non-trivial game rather than a straightforward optimization problem. To better characterize the relationship between model uncertainty and community interest, we examine scatter plots for three representative GenAI–forum pairs (Figure 2). As expected, most questions exhibit low view counts, suggesting they are of limited value to the Q&A platforms. However, as we move along the view-count axis, no clear trend emerges, highly

LLM	Heuristic	Perplexity	Views
Pythia 6.9B	MPP	134,541.75	263.03
Pythia 6.9B	MaxSP	79,957.79	**295.45**
Pythia 6.9B	GreedyNP	**150,074.12**	251.59
Pythia 6.9B	Random	43,430.74	19.35
LLaMA 3.1 8B	MPP	121,931.06	265.12
LLaMA 3.1 8B	MaxSP	70,641.08	**299.66**
LLaMA 3.1 8B	GreedyNP	**135,279.39**	254.89
LLaMA 3.1 8B	Random	39,889.08	17.68
LLaMA 3.1 8B-Instruct	MPP	196,204.18	265.31
LLaMA 3.1 8B-Instruct	MaxSP	113,915.33	**294.46**
LLaMA 3.1 8B-Instruct	GreedyNP	**221,401.76**	251.65
LLaMA 3.1 8B-Instruct	Random	60,677.23	18.31

Table 1: Performance of the idealized full-information and cooperative scenario. "Perplexity" is the cumulative perplexity achieved by Player G, while "Views" is the cumulative normalized views achieved by Player F.

viewed questions do not consistently correspond to higher perplexity, reinforcing the weak alignment between the objectives of GenAI models and Q&A communities.

5.2 Sequential Interaction

5.2.1 Full information. Aggregating the weekly cooperative outcomes across all models yields consistent trends, as reported in Table 1. The results demonstrate a persistent utility divergence between the two players, even under perfect cooperation. Specifically, the Nash-style **GreedyNP** heuristic consistently maximizes Player G's cumulative perplexity, indicating superior GenAI's improvement potential. In contrast, the **MaxSP** heuristic produces the highest cumulative view counts, favoring Player F's engagement-oriented objectives. The **MPP** heuristic achieves an intermediate trade-off between the two utilities, while **Random** performs substantially worse across all metrics.

5.2.2 Asymmetric information. These findings highlight that even under conditions of complete transparency and alignment, inherent trade-offs between model improvement potential (Player G) and user engagement (Player F) remain unavoidable. Thus, full information does not eliminate incentive asymmetry, it merely clarifies the structure of the cooperative frontier against which more constrained mechanisms can be evaluated. Table 2 reports both the cumulative perplexity achieved and the normalized view count. Across all evaluated language models, **G-Utility** strategy consistently outperforms the **G-Greedy** strategy. For every model, incorporating acceptance probability estimates yields higher total perplexity than naively maximizing perplexity alone. For example, with Pythia-6.9b, **G-Greedy** achieves a cumulative perplexity of 75,715, whereas the **G-Utility** strategy increases this to 78,164, while simultaneously boosting community engagement from 163.6 to 196.1 normalized views. Similar gains are observed for both variants of LLaMA-3.1, where expected-utility selection yields perplexity improvements of +20% to +23% relative to greedy, alongside substantial increases in views. The random baseline is dominated on both metrics, confirming that success requires alignment with forum preferences rather than merely injecting variance. Overall, these findings demonstrate that strategic anticipation of forum behavior enables strictly better outcomes for both parties, even under asymmetric information and without enforcing quota fulfillment (since S_t could be a strict subset of A_t).

Table 3 compares the best strategy under partial utility, **G-Utility**, with the leading fully collaborative strategies (**GreedyNP** for Player G and **MaxSP** for Player F). Our results demonstrate that **G-Utility** successfully recovers a substantial fraction of the cooperative benefits *without* requiring direct coordination or information exchange between Player F and Player G. Across all evaluated LLMs, the learned acceptance-aware strategy achieves approximately 56–66% of the maximal attainable view counts (under **MaxSP**) and 46–52% of the maximal achievable perplexity (under **GreedyNP**).

6 RELATED WORK

Data Depletion and Feedback Loops. Users increasingly abandon their roles as knowledge contributors once AI tools provide immediate assistance [10], creating a "participation decline" [36]. This trend poses significant risks to the long-term viability of AI development [5]. Both open-source and commercial LLMs critically rely on rich, publicly available datasets [1], such as *The Pile* [15] and *CodeInsight* [7], which incorporate data from Q&A platforms like Stack Overflow. As these sources experience declining contributions, AI models face a "data starvation" risk: reduced human input may degrade model quality, discouraging further participation and creating a self-reinforcing cycle.

Participation Dynamics and Incentives. Existing approaches to this sustainability challenge have significant limitations. Proposals to restrict community data usage [3] or implement basic incentive mechanisms [13] address only surface-level symptoms and neglect the structural dynamics driving participation decline. Behavioral interventions, such as Taitler and Ben-Porat's "selective response" [35], attempt to steer user behavior but still treat AI providers and knowledge communities as competing stakeholders rather than collaborators. Overall, these approaches fail to provide cooperative solutions, overlooking the fundamental interdependence between forums and AI systems: forums supply essential training data, while AI systems influence user behavior. We address this challenge by framing it as a cooperative equilibrium problem, introducing a game-theoretic framework that models strategic interactions between AI providers and knowledge communities. By formalizing their interdependencies, we design selection mechanisms that promote sustainable contributions while preserving confidentiality. Our framework supports practical deployment and empirical evaluation, demonstrating measurable improvements in both model performance and community engagement, leveraging the fact that users increasingly ask questions that LLMs cannot fully answer [18].

7 LIMITATIONS

In our work, we operated under several assumptions that may not fully hold in competitive or changing real-world settings:

- **Two-player game:** Our work is conceptual with the goal of presenting a proof-of-concept collaboration framework between a single GenAI company and a single forum. However, practical

LLM	Heuristic	Views	Perplexity
Pythia 6.9B	G-Greedy	163.61	75715.23
Pythia 6.9B	G-Utility	**196.12**	**78164.39**
Pythia 6.9B	Random	144.39	18484.38
LLaMA 3.1 8B	G-Greedy	119.80	50948.34
LLaMA 3.1 8B	G-Utility	**166.35**	**62642.90**
LLaMA 3.1 8B	Random	134.86	15347.52
LLaMA 3.1 8B-Instruct	G-Greedy	139.87	92051.71
LLaMA 3.1 8B-Instruct	G-Utility	**176.90**	**110636.87**
LLaMA 3.1 8B-Instruct	Random	141.97	25483.55

Table 2: Performance of asymmetric information. The G-Utility strategy maximizes both view counts and perplexity."Perplexity" is the cumulative perplexity achieved by Player G, while "Views" is the cumulative normalized views achieved by Player F.

LLM	Views			Perplexity		
	G-Utility	GreedyNP	MaxSP	G-Utility	GreedyNP	MaxSP
Pythia 6.9B	196.12 (66%)	251.59 (78%)	295.45 (100%)	78,164 (52%)	150,074 (100%)	79,958 (53%)
LLaMA 3.1 8B	166.35 (56%)	254.89 (85%)	299.66 (100%)	62,643 (46%)	135,279 (100%)	70,641 (52%)
Llama-3-8B-Instruct	176.90 (60%)	251.65 (85%)	294.46 (100%)	110,637 (50%)	221,402 (100%)	113,915 (51%)

Table 3: Comparison across collaboration strategies. The asymmetric information strategy G-Utility achieves 56–66% of the view count maximizing strategy (MaxSP) and simultaneously 46–52% of the perplexity maximizing strategy (GreedyNP)."Perplexity" is the cumulative perplexity achieved by Player G, while "Views" is the cumulative normalized views achieved by Player F.

scenarios could involve multiple GenAI entities and multiple forum simultaneously. Such multi-agent settings could introduce strategic interdependence among GenAI players; for example, if an LLM company anticipates that a competitor will submit certain questions, it may adjust its own submissions strategically. Alternatively, if one forum refuses to publish questions from a certain LLM provider, the provider could decided to never collaborate with that forum again.

- **Single domain of utility proxies:** Our analysis is limited to one domain of utility proxies. We do not examine cases where the GenAI company submits domain-specific content (e.g., questions related to its own products or models) or where the forum's selection rule $\mathcal{R}$ evolves dynamically over time. Such extensions could meaningfully affect both the utility structure and the observed patterns of cooperation.
- **Temporal scope:** The empirical evaluation is based on data covering a one-year period. This temporal scope may not fully capture the long-term stability or potential cyclical behavior of cooperative strategies.
- **Linear utilities:** Recall that our model assumes utilities are linear. While this simplifies the setting, it might be unrealistic in some scenarios. Indeed, publishing two similar questions in a forum results in cannibalism of view count, as the sum of view counts could be higher than publishing only one of the questions. Non-linear utility functions (e.g., the submodular formulation analyzed by Amanatidis et al.) may better capture redundancy effects and are left for future work.

8 CONCLUSION AND DISCUSSION

In this work, we addressed a critical sustainability challenge at the intersection of generative AI development and online knowledge forums. While LLMs rely heavily on high-quality forum data for continual improvement, the rise of these models threatens the very forums that supply such data, creating a dynamic reminiscent of the "Tragedy of the Commons" [17].

We surveyed the key guidelines and then formalized the problem through a game-theoretic model of asymmetric information. We designed a practical framework where GenAI companies propose questions and forums selectively accept them, assuming non-transferable utilities. Our empirical analysis demonstrates that player utilities are systematically misaligned, suggesting that the interaction between GenAI systems and human-driven forums constitutes a genuine strategic game rather than an optimization problem. Nonetheless, our results show that a lightweight, acceptance-aware strategy for the GenAI company and a simple threshold classifier can recover a substantial portion of the ideal full-information variant—achieving up to 66% of forum utility and 52% of GenAI utility relative to full-information collaboration.

Together, our findings provide evidence that sustainable collaboration between GenAI companies and online forums is not only possible but highly beneficial to both parties.

REFERENCES

[1] Toufique Ahmed, Christian Bird, Premkumar Devanbu, and Saikat Chakraborty. 2024. Studying LLM Performance on Closed- and Open-source Data. arXiv:2402.15100 [cs.SE] https://arxiv.org/abs/2402.15100

[2] Jin Ai, Dogan Gursoy, Yue Liu, and Xingyang Lv. 2022. Effects of offering incentives for reviews on trust: Role of review quality and incentive source. *International Journal of Hospitality Management* 100 (2022), 103101. https://doi.org/10.1016/j.ijhm.2021.103101

[3] Ines Akaichi and Sabrina Kirrane. 2025. A comprehensive review of usage control frameworks. *Computer Science Review* 56 (2025), 100698. https://doi.org/10.1016/j.cosrev.2024.100698

[4] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, Daniele Mazzotta, Badreddine Noune, Baptiste Pannier, and Guilherme Penedo. 2023. The Falcon Series of Open Language Models. arXiv:2311.16867 [cs.CL] https://arxiv.org/abs/2311.16867

[5] Adriana Alvarado Garcia, Heloisa Candello, Karla Badillo-Urquiola, and Marisol Wong-Villacres. 2025. Emerging Data Practices: Data Work in the Era of Large Language Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 846, 21 pages. https://doi.org/10.1145/3706598.3714069

[6] Georgios Amanatidis, Georgios Birmpas, Philip Lazos, Stefano Leonardi, and Rebecca Reiffenhäuser. 2025. Algorithmically Fair Maximization of Multiple Submodular Objective Functions. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems* (Detroit, MI, USA) *(AAMAS '25)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 115–123.

[7] Nathanaël Beau and Benoît Crabbé. 2024. CodeInsight: A Curated Dataset of Practical Coding Solutions from Stack Overflow. arXiv:2409.16819 [cs.CL] https://arxiv.org/abs/2409.16819

[8] Stella Biderman, Hailey Schoelkopf, Quentin Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, Aviya Skowron, Lintang Sutawika, and Oskar van der Wal. 2023. Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling. arXiv:2304.01373 [cs.CL] https://arxiv.org/abs/2304.01373

[9] Damai Dai, Chengqi Deng, Chenggang Zhao, R. X. Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Y. Wu, Zhenda Xie, Y. K. Li, Panpan Huang, Fuli Luo, Chong Ruan, Zhifang Sui, and Wenfeng Liang. 2024. DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models. arXiv:2401.06066 [cs.CL] https://arxiv.org/abs/2401.06066

[10] R Maria del Rio-Chanona, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus* 3, 9 (09 2024), pgae400. https://doi.org/10.1093/pnasnexus/pgae400 arXiv:https://academic.oup.com/pnasnexus/article-pdf/3/9/pgae400/59316621/pgae400.pdf

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL] https://arxiv.org/abs/1810.04805

[12] Brandon Fain, Kamesh Munagala, and Nisarg Shah. 2018. Fair Allocation of Indivisible Public Goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation* (Ithaca, NY, USA) *(EC '18)*. Association for Computing Machinery, New York, NY, USA, 575–592. https://doi.org/10.1145/3219166.3219174

[13] Mike Felgenhauer. 2024. Competition for publication-based rewards. *Economics Letters* 244 (2024), 112017. https://doi.org/10.1016/j.econlet.2024.112017

[14] Robert Forsythe, Joel L Horowitz, Nathan E Savin, and Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic behavior* 6, 3 (1994), 347–369.

[15] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, and Connor Leahy. 2020. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv preprint arXiv:2101.00027* (2020).

[16] Hila Gonen, Srini Iyer, Terra Blevins, Noah A. Smith, and Luke Zettlemoyer. 2024. Demystifying Prompts in Language Models via Perplexity Estimation. arXiv:2212.04037 [cs.CL] https://arxiv.org/abs/2212.04037

[17] Garrett Hardin. 1968. The Tragedy of the Commons. *Science* 162, 3859 (1968), 1243–1248. https://doi.org/10.1126/science.162.3859.1243 arXiv:https://www.science.org/doi/pdf/10.1126/science.162.3859.1243

[18] Denis Helic and Tiago Santos. 2025. Stack Overflow Is Not Dead Yet: Crowd Answers Still Matter. *arXiv preprint arXiv:2509.05879* (2025).

[19] Heidi Herlin. 2015. Better Safe Than Sorry: Nonprofit Organizational Legitimacy and Cross-Sector Partnerships. *Business & Society* 54, 6 (2015), 822–858. https://doi.org/10.1177/0007650312472609 arXiv:https://doi.org/10.1177/0007650312472609

[20] Heidi Herlin. 2015. Better Safe Than Sorry: Nonprofit Organizational Legitimacy and Cross-Sector Partnerships. *Business & Society* 54, 6 (Dec. 2015), 822–858. https://doi.org/10.1177/0007650312472609

[21] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. 2024. Large Language Models for Software Engineering: A Systematic Literature Review. arXiv:2308.10620 [cs.SE] https://arxiv.org/abs/2308.10620

[22] Hsiu-Yuan Huang, Yutong Yang, Zhaoxi Zhang, Sanwoo Lee, and Yunfang Wu. 2024. A Survey of Uncertainty Estimation in LLMs: Theory Meets Practice. arXiv:2410.15326 [cs.CL] https://arxiv.org/abs/2410.15326

[23] Hugging Face. 2024. *Meta-Llama-3-8B-Instruct*. https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct Accessed: 2025-10-08.

[24] Konstantinos Koiliaris and Chao Xu. 2018. Subset sum made simple. *arXiv preprint arXiv:1807.08248* (2018).

[25] Xinyu Li and Keongtae Kim. 2024. Impacts of generative AI on user contributions: evidence from a coding Q&A platform. *Marketing Letters* 36 (09 2024), 577–591. https://doi.org/10.1007/s11002-024-09747-1

[26] Haoyu Lu, Wen Liu, Bo Zhang, Bingxuan Wang, Kai Dong, Bo Liu, Jingxiang Sun, Tongzheng Ren, Zhuoshu Li, Hao Yang, Yaofeng Sun, Chengqi Deng, Hanwei Xu, Zhenda Xie, and Chong Ruan. 2024. DeepSeek-VL: Towards Real-World Vision-Language Understanding. arXiv:2403.05525 [cs.AI] https://arxiv.org/abs/2403.05525

[27] Katerina Margatina, Timo Schick, Nikolaos Aletras, and Jane Dwivedi-Yu. 2023. Active Learning Principles for In-Context Learning with Large Language Models. arXiv:2305.14264 [cs.CL] https://arxiv.org/abs/2305.14264

[28] Meta AI. 2024. *Introducing Meta Llama 3.1: Our most capable models to date*. https://ai.meta.com/blog/meta-llama-3-1/ Accessed: 2025-10-08.

[29] John Nash. 1953. Two-Person Cooperative Games. *Econometrica* 21, 1 (1953), 128–140. http://www.jstor.org/stable/1906951

[30] Ruth Phillips. 2007. Tamed or Trained? The Co-option and Capture of 'Favoured' NGOs. *Third Sector Review* 13, 2 (July 2007), 27+. Accessed 28 Sept. 2025.

[31] Ariel D Procaccia and Moshe Tennenholtz. 2013. Approximate mechanism design without money. *ACM Transactions on Economics and Computation (TEAC)* 1, 4 (2013), 1–26.

[32] Georg Reischauer and Johanna Mair. 2018. How Organizations Strategically Govern Online Communities: Lessons from the Sharing Economy. *Academy of Management Discoveries* 4 (09 2018), 220–247. https://doi.org/10.5465/amd.2016.0164

[33] KAREN SPARCK JONES. 1972. A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL. *Journal of Documentation* 28, 1 (01 1972), 11–21. https://doi.org/10.1108/eb026526 arXiv:https://www.emerald.com/jd/article-pdf/28/1/11/1336479/eb026526.pdf

[34] Boaz Taitler and Omer Ben-Porat. 2024. Braess's Paradox of Generative AI. arXiv:2409.05506 [cs.GT] https://arxiv.org/abs/2409.05506

[35] Boaz Taitler and Omer Ben-Porat. 2025. Selective response strategies for genai. *arXiv preprint arXiv:2502.00729* (2025).

[36] Yucheng Zhang, Chenhao Sun, Jiajun Chen, Yuxin Wang, and Yongfeng Zhang. 2024. The consequences of generative AI for online knowledge communities. *Scientific Reports* 14, 16321 (2024). https://doi.org/10.1038/s41598-024-69804-5

A APPENDIX

B NP-HARDNESS OF MAXIMIZING EQUATION (1)

Problem Statement. Let $\Omega = \{1, \dots, n\}$ be a finite set, and let $f, g : \Omega \to \mathbb{R}_{\geq 0}$ be two non-negative functions. Given an integer k, consider the optimization problem

$$\max_{S \subseteq \Omega,\ |S| \leq k} F(S) \quad \text{where} \quad F(S) = \Big(\sum_{i \in S} f(i)\Big) \Big(\sum_{i \in S} g(i)\Big). \qquad (2)$$

Theorem B.1. *The problem in Equation* (2) *is NP-hard.*

Proof. We reduce from the *cardinality-constrained subset sum* problem (CCSS):

- *Instance*: positive integers $a_1, \dots, a_n$, a target T, and size bound k.
- *Question*: is there a subset $S \subseteq \{1, \dots, n\}$ of size $|S| = k$ such that $\sum_{i \in S} a_i = T$?

Proposition B.2. *The CCSS problem is NP-hard.*

Proof. We prove this proposition by showing that CCSS is at least as hard as the Subset Sum problem, which is known to be

NP-hard [24]. Given an instance $(a_1, \ldots, a_n; T)$ of Subset Sum, form the multiset $\{a_1, \ldots, a_n, 0, \ldots, 0\}$ with n zeros and set $K = n$. Then any size-K subset summing to T picks exactly the original items that sum to T (and fills the rest with zeros), so Subset Sum reduces in polynomial time to Exact-K Subset Sum. □

On this input, define an instance of our bilinear maximization by choosing

$$f(i) = a_i, \quad M = \frac{2T}{k}, \quad g(i) = M - a_i \quad \text{for each } i = 1, \ldots, n.$$

For any subset S with $|S| = k$, let

$$A_S = \sum_{i \in S} a_i.$$

Then

$$\sum_{i \in S} f(i) = A_S, \qquad \sum_{i \in S} g(i) = kM - A_S,$$

and thus

$$F(S) = A_S\,(kM - A_S) = -A_S^2 + kM\,A_S.$$

The quadratic function

$$Q(x) = -x^2 + kM\,x$$

is strictly concave in x and attains its unique maximum at

$$x^* = \frac{kM}{2} = \frac{k}{2} \cdot \frac{2T}{k} = T.$$

Hence, among all subsets of size k, $F(S)$ is *uniquely maximized* exactly when $A_S = T$. Therefore, solving the bilinear maximization and checking whether the optimum equals

$$T\,(kM - T) = T\,(2T - T) = T^2$$

answers the subset-sum question. Since the latter is NP-hard, our maximization problem is NP-hard as well. □

C STATISTIC SIGNIFICANCE

Table 4: Spearman Correlation Between Normalized View Count and Perplexity

LLM	Subject	Spearman	p-val
Pythia 6.9B	ubuntu	-0.11725	8.6078×10^{-37}
Pythia 6.9B	english	-0.05326	0.0381
Pythia 6.9B	latex	-0.04045	0.0002
Pythia 6.9B	stackoverflow	-0.01343	1.9556×10^{-11}
Pythia 6.9B	math	-0.07459	2.2311×10^{-54}
LLaMA 3.1 8B	ubuntu	-0.12704	5.7306×10^{-43}
LLaMA 3.1 8B	english	-0.08214	0.0012
LLaMA 3.1 8B	latex	-0.02481	0.0227
LLaMA 3.1 8B	stackoverflow	-0.04896	2.2188×10^{-132}
LLaMA 3.1 8B	math	-0.08519	1.5570×10^{-70}
LLaMA 3.1 8B-Instruct	ubuntu	-0.11300	2.6493×10^{-34}
LLaMA 3.1 8B-Instruct	english	-0.05642	0.0267
LLaMA 3.1 8B-Instruct	latex	-0.03429	0.0016
LLaMA 3.1 8B-Instruct	stackoverflow	-0.04468	1.5948×10^{-110}
LLaMA 3.1 8B-Instruct	math	-0.08446	2.3419×10^{-69}