Journal of Organizational Behavior Management

ISSN: 0160-8061 (Print) 1540-8604 (Online) Journal homepage: www.tandfonline.com/journals/worg20

Behavioral economics of artificial intelligence (AI) at work: Modeling demand for large language models in organizational contexts

Brent A. Kaplan, Derek D. Reed, Matthew M. Laske, Madison E. Graham, Florence D. DiGennaro Reed, Abigail Blackman & Steven R. Hursh

To cite this article: Brent A. Kaplan, Derek D. Reed, Matthew M. Laske, Madison E. Graham, Florence D. DiGennaro Reed, Abigail Blackman & Steven R. Hursh (16 Aug 2026): Behavioral economics of artificial intelligence (AI) at work: Modeling demand for large language models in organizational contexts, Journal of Organizational Behavior Management, DOI: 10.1080/01608061.2026.2717085

To link to this article: https://doi.org/10.1080/01608061.2026.2717085

Published online: 16 Aug 2026.

JOURNAL OF ORGANIZATIONAL BEHAVIOR MANAGEMENT
https://doi.org/10.1080/01608061.2026.2717085

Routledge
Taylor & Francis Group

Behavioral economics of artificial intelligence (AI) at work: Modeling demand for large language models in organizational contexts

Brent A. Kaplan [a], Derek D. Reed [b], Matthew M. Laske [b], Madison E. Graham [c], Florence D. DiGennaro Reed [d], Abigail Blackman [e], and Steven R. Hursh [b]

[a]codedbx, Lexington, KY, USA; [b]Institutes for Behavior Resources, Inc., Baltimore, MD, USA; [c]Department of Applied Behavioral Science, University of Kansas, Lawrence, KS, USA; [d]Behavior Analyst Certification Board, Littleton, CO, USA; [e]Council of Autism Service Providers, Lexington, SC, USA

ABSTRACT

Generative artificial intelligence and large language models (LLMs) are becoming routine tools for knowledge work, yet workers differ in which systems they choose and how strongly they rely on them. We applied operant demand methods to quantify the behavioral value of LLM assistance among employed adults who currently use LLMs. Participants ranked familiar LLMs and completed hypothetical purchase tasks for their most preferred model, least preferred model, and a multi-model platform. Consistent with behavioral economic principles, consumption decreased systematically as cost increased, producing well-ordered demand curves. Top-preferred LLMs showed relatively inelastic demand and higher essential value; least-preferred LLMs showed steeper declines in consumption; the multi-model option reflected intermediate valuation. Demand metrics may offer organizations a practical way to inform LLM access and resource-allocation decisions, though findings are limited to current LLM users under a hypothetical direct-cost contingency and may not extend to non-monetary constraints such as time and effort.

KEYWORDS

Artificial intelligence; behavioral economics; demand; large language models; organizational behavior management

Large language models (LLMs) such as OpenAI's ChatGPT™, Anthropic's Claude™, and Google's Gemini™have become increasingly common tools for knowledge work. Unlike earlier rule-based chatbots, these models are trained on large text corpora and can produce contextually appropriate prose, condense complex information into clear summaries, and generate syntactically correct code (Zhao et al., 2023). As of August 2025, a nationally representative survey found that 54.6% of U.S. adults aged 18–64 had used generative AI and 37.4% of employed workers had used it at work (Bick et al., 2025, 2026). Experimental research shows that LLMs can deliver meaningful productivity gains; a randomized trial with mid-career professionals found that ChatGPT cut task completion time by about 37% (a decrease of about

CONTACT Brent A. Kaplan brentkaplan@codedbx.com codedbx, Lexington, KY 40503, USA

10 minutes) and improved the quality of written work by about 20% (Noy & Zhang, 2023), while a field study of customer support agents found that AI assistance increased the number of issues resolved per hour by 15% and helped less experienced workers learn on the job (Brynjolfsson et al., 2025), though such gains may vary substantially across task types and worker populations.

Many workers now use LLMs to support their work, most often to draft communications, search and summarize information, document instructions, translate, and write code (Brachman et al., 2025; Chatterji et al., 2025). Much of this use is also undisclosed. In experimental task scenarios, between 62% and 92% of participants reported they would passively conceal their use of LLMs, and between 46% and 70% reported they would actively conceal it (Zhang et al., 2025). This kind of employee-initiated, unsanctioned use (a pattern sometimes called "shadow AI") indicates that a worker's decision to use an LLM often operates somewhat independently of formal organizational adoption.

From an organizational behavior management (OBM) perspective, LLM use is a workplace behavior worth examining in its own right. For routine or semi-structured tasks, LLMs can reduce the time and effort a task requires, which may free workers for higher-value strategic or interpersonal activities. These efficiencies are counterbalanced by substantial barriers, including concerns over accuracy and hallucinated outputs, data governance and intellectual property risks, environmental costs of large-scale model training and inference (Ren et al., 2024), and fears of job displacement (Lin & Parker, 2025).

Worker attitudes toward LLMs also vary widely. Some workers are enthusiastic adopters, whereas others are skeptical or resistant, viewing these tools as imposed rather than genuinely useful. A nationally representative Pew survey found that more workers express worry than excitement about the growing role of AI in the workplace (Lin & Parker, 2025), and leadership tends to overestimate employee enthusiasm; in one survey, 76% of executives believed their employees were enthusiastic about adoption, compared with just 31% of individual contributors (Lovich et al., 2025). This heterogeneity in attitudes, and in willingness to engage, is part of why quantifying demand among workers who already use LLMs is informative but partial: it characterizes the behavior of current users, not the reasons others decline.

Whether LLM assistance contributes to an organization's work depends on more than technical feasibility or who pays for access. We submit that workers' election to use LLMs (particularly when that use is discretionary) is the ultimate metric of potential impact. Put another way, regardless of institutional support, the worker is the agent engaging the LLM. The worker must expend their own resources to prompt the LLM and verify the veracity of its outputs. Understanding the contribution of LLM support to an organization's productivity and output thereby depends on the worker's allocation of time, resources, and effort to engage the LLM, particularly when constraints are

imposed on its use. In practice, the constraints on LLM use take many forms: direct financial cost (personal subscriptions or departmental chargebacks), time and effort required to craft prompts and verify outputs, organizational policies that limit or gate access, and compliance or security burdens. While the present study operationalizes constraint as a direct monetary cost to the worker, the behavioral economic framework generalizes to any response cost that modulates engagement.

Behavioral economics is a subdiscipline of behavior analysis that aims to understand how constraints on a reinforcer affect operant behavior (Reed et al., 2013) – for example, how LLM costs might alter workers' use of the service. Generally speaking, behavioral economics is the intersection of microeconomic principles and behavioral psychology. Behavior analysts leverage behavioral economic principles whenever they work to identify the optimal reinforcer magnitude or cost to maintain target levels of behavior, or to optimize contingency–response relations. Thus, it is no surprise that behavioral economics has been proposed as a novel approach to studying behavior in organizations (DiGennaro Reed et al., 2015; Hirst & DiGennaro Reed, 2016; Wine et al., 2012), particularly in performance management and worker output.

Operant demand is a central concept in behavioral economics and is defined as an organism's defense of a reinforcer under increasing constraint (Reed et al., 2025). As discussed above, demand theory can be translated to operationalize LLM demand as a worker's willingness to expend resources to access the LLM in the face of increasing costs. While operant demand typically studies demand for reinforcers such as food or drugs by manipulating the unit price of operants such as key or lever presses, the concept has been translated to myriad everyday human interactions (Reed et al., 2020), often using hypothetical tasks where participants report how much of a reinforcer they would buy or consume across a range of increasing prices – such as beer purchases to simulate alcohol demand (Kaplan et al., 2018) or willingness to get tested for diseases at various testing costs (Strickland et al., 2022). This hypothetical demand approach has also been translated to worker behavior in OBM contexts. For example, Henley and colleagues arranged a hypothetical work task to model "worker" willingness to complete job tasks at varying incentive amounts (Henley, DiGennaro Reed, Kaplan, et al., 2016). A follow-up study used crowdsourcing methods to have Amazon Mechanical Turk workers complete menial tasks at increasing effort requirements to model workforce demand in a behavioral economic framework (Henley, DiGennaro Reed, Reed, et al., 2016).

Behavioral economists use demand curve analyses to quantify and model operant demand data (Kaplan, 2025b; see also Reed, Kaplan, et al., 2025 for a comprehensive treatment). Demand curves can be quantified using both empirical (i.e., directly observed data from the task) and derived (i.e.,

statistically modeled using nonlinear curve-fitting techniques) metrics. The most common demand indices are *intensity* (the amount of the reinforcer consumed at minimal cost), *elasticity* (the degree to which consumption decreases as price increases; more elastic demand indicates sharper reductions in use as costs rise, while inelastic demand indicates sustained use despite cost increases), *Pmax* (the maximum price tolerated), *Omax* (the maximum behavioral output across all prices), and *breakpoint* (the last price associated with any reinforcer consumption). Importantly, demand metrics – both empirical and derived – have strong psychometric properties, even in hypothetical tasks (Miller et al., 2023).

Quantifying how much workers value LLM assistance, and how that value holds up as access becomes more costly, is a first step toward understanding the role these tools play in workers' output. Therefore, the purpose of this study was to apply behavioral economic demand methods to quantify how employed adults who currently use LLMs value access to LLM assistance for organizationally relevant tasks. Using a series of hypothetical purchase tasks and a nonlinear mixed-effects model, we estimated demand and expenditure for participants' most-preferred, least-preferred, and multi-model LLM access across a range of prices. These analyses were intended to generate quantitative indices of LLM valuation that can inform how organizations structure and support LLM access.

Method

Participants

We recruited all participants from Prolific, an online research platform that maintains a pre-screened participant pool and provides demographic and behavioral filters for targeted recruitment. Participants were eligible if they: (a) were at least 18 years of age and resided in the United States, (b) expressed experience with AI chatbots (specifically: ChatGPT, Claude, Google Bard, Google Gemini, Grok, Microsoft Bing AI), (c) reported any weekly AI use, (d) were not employed as computer programmers, (e) used a computer or mobile device for 50% or more of their work, (f) used technology at work more than once a day, (g) indicated at least part-time employment, (h) had a lifetime approval rate of 95–100% in Prolific, and (i) had at least 250 previous Prolific submissions. Quality-control filters (h and i) were applied to reduce inattentive or careless responding, consistent with recommended practices for online behavioral research (Peer et al., 2022). Recruitment dates occurred between August 22, 2025, and August 29, 2025. The median session duration was 16 min. Participants were compensated $3.00 USD for completing the study, rendering a realized wage of approximately $11.50/h.

Interested and eligible individuals completed the survey through Qualtrics. Upon entering the survey, they were asked to consent to participate. All procedures were approved by the University of Kansas Institutional Review Board. The first block of questions asked about participants' experiences and behaviors related to large language models (LLMs).

We recruited a total of 126 participants, most of whom worked full time (84%). Table 1 presents the full demographic breakdown by age, sex/gender, race, education, household income, and employment.

Apparatus

The survey was administered via the Qualtrics Experience Management platform (https://www.qualtrics.com). Participants completed the study on their own devices (desktop or mobile). Custom JavaScript embedded within Qualtrics dynamically displayed productivity and cost calculations during the HPTs.

LLM questions

For all LLM-related questions, we told participants to assume the LLM provider or model was the most recent release or their most preferred release for that model. Participants initially selected the LLMs they had experience with or were researching. The full list included OpenAI GPT, Anthropic Claude, Deepseek, Google Gemini, Meta Llama, Alibaba Qwen, Microsoft Co-Pilot, xAI Grok, and Moonshot Ai Kimi.

Preference assessment. After indicating their familiarity with LLMs, participants completed a preference assessment for their top and bottom preferred LLMs. The available LLMs were only those with which participants had experience. Participants were asked to rank their preferred LLMs from 1 (top preferred) to 10 (bottom preferred).

LLM usage. We then asked participants to indicate how many prompts they write on average each day for each LLM provider/model. Options included 0, 1–5, 6–10, 11–20, 21–50, and 51+ prompts. Then we asked an analogous question about hours spent working with each LLM provider/model during an average workday. Options included 0, 1–2, 3–4, 5–6, 7–8, and 9+ hours per day.

Behavioral economic tasks

Participants completed a series of different behavioral economic tasks to gauge their decision making for LLMs under various pricing scenarios. For the purposes of this paper, we will focus on three hypothetical purchase tasks (HPTs). Participants completed one HPT for each of their top and bottom preferred LLMs, as well as one for a *multi-model* LLM (i.e., a platform offering

Table 1. Demographic summary.

Characteristic	N	$N = 126$[1]
AgeBin	126	
18–24		6.0 (4.8%)
25–34		32.0 (25.4%)
35–44		33.0 (26.2%)
45–54		32.0 (25.4%)
55+		23.0 (18.3%)
How do you describe yourself?	126	
Male		54.0 (42.9%)
Female		68.0 (54.0%)
Non-binary / third gender		4.0 (3.2%)
Prefer to self-describe		0.0 (0.0%)
Prefer not to say		0.0 (0.0%)
Are you of Spanish, Hispanic, or Latino origin?	126	13.0 (10.3%)
Race	126	
American Indian/Native American or Alaska Native		1.0 (0.8%)
Asian		2.0 (1.6%)
Black or African American		13.0 (10.3%)
Multiracial		10.0 (7.9%)
Other		3.0 (2.4%)
White		97.0 (77.0%)
Education	126	
High school or less		6.0 (4.8%)
Some college/Associate's		40.0 (31.7%)
Bachelor's degree		43.0 (34.1%)
Graduate/Professional		37.0 (29.4%)
Income	126	
<$50,000		26.0 (20.6%)
$50,000–$99,999		48.0 (38.1%)
$100,000–$149,999		32.0 (25.4%)
$150,000 or more		19.0 (15.1%)
Prefer not to say		1.0 (0.8%)
What best describes your employment status over the last three months?	126	
Working full-time		106.0 (84.1%)
Working part-time		20.0 (15.9%)
Unemployed and looking for work		0.0 (0.0%)
A homemaker or stay-at-home parent		0.0 (0.0%)
Student		0.0 (0.0%)
Retired		0.0 (0.0%)
Other		0.0 (0.0%)
Do you ever use a multi-model platform (e.g., OpenRouter, Requesty, Vertex AI, LiteLLM)?	126	
No, and I don't know what a multi-model platform is.		70.0 (55.6%)
No, but I know what a multi-model platform is.		45.0 (35.7%)
Yes		11.0 (8.7%)
Approximately how much do you spend per month on LLM capabilities, in total (across all kinds)? These can include subscriptions, tokens, API calls, etc.	106	
$0		50.0 (47.2%)
$1.00 to $9.99		13.0 (12.3%)
$10.00 to $19.99		20.0 (18.9%)
$20.00 to $49.99		13.0 (12.3%)
$50.00 to $99.99		9.0 (8.5%)
$99.99 to $199.99		1.0 (0.9%)
$200.00 or more		0.0 (0.0%)
Unknown		20

Note. Values are count (percentage) unless otherwise specified.

access to multiple underlying language models within one interface; top and bottom preferred LLM order was randomized with the multi-model LLM always presented last). For each of these HPTs, we first provided participants with a vignette:

> Imagine that your company offers an opportunity for you to individually purchase and use access to [**TOP/BOTTOM/MULTI-MODEL LLM**] to enhance your productivity by streamlining tasks such as writing reports, summarizing meetings, drafting emails, and generating insights from data. [FOR MULTI-MODEL ONLY: In this multi-model platform, you can specify which LLM provider and model to use for each prompt. The options available are always the latest model versions and include all of the following:
>
> - OpenAI GPT
> - Anthropic Claude
> - Deepseek
> - Google Gemini
> - Meta Llama
> - Alibaba Qwen
> - Microsoft Co-Pilot
> - xAI Grok
> - Moonshot Ai Kimi
>
> This [**TOP/BOTTOM/MULTI-MODEL LLM**] access provides batches of **structured prompts per hour**, allowing you to work more efficiently and potentially freeing up time for other tasks. **Each batch of prompts** is equivalent to approximately **30 min of additional productivity per hour**. The system operates seamlessly within your workflow and adapts to your specific role, whether you are a worker focusing on task execution or a manager overseeing projects and making strategic decisions.
>
> You can purchase a **workweek's supply** of [**TOP/BOTTOM/MULTI-MODEL LLM**] assisted productivity batches, assuming a **standard 40-h workweek**. For example, purchasing 40 batches would equate to using **one batch per hour across the entire week**. If you purchase 80 batches, you would be using two batches per hour across the week, effectively maximizing AI assistance for most of your tasks.
>
> Assume the following information:
>
> - You must pay for [**TOP/BOTTOM/MULTI-MODEL LLM**] access using your own finances (your company is NOT paying for it).
> - You have the same income/savings that you do now.
> - You do not have access to any other prompt-based LLM AI assistance than what is available for purchase here.
> - It is a typical workweek, with your usual projects, tasks, work requirement, and timelines.
> - The batches you are about to purchase are for your use only. In other words, you can't share your batches with anyone else. You also can't save the batches to roll over across weeks. Therefore, everything you buy is for your own use during the single 40-h week.
>
> Please answer these questions honestly, as if you were actually in this situation.
>
> *Over the next group of pages, indicate how many total batches for the workweek you would purchase at different price points.*

At the end of the vignette, participants were required to correctly answer a multiple-choice verification question asking: "This task asks you to imagine paying for batches of LLM access for _____" with options: "1 day," "1 week," "1 month," and "1 year."

After successfully answering the question, participants completed the first of nine price points in the HPT. The question read: "How many **[TOP/BOTTOM/MULTI-MODEL LLM]** productivity batches would you request for an entire workweek at the following price? **$XX per batch**" where XX increased consecutively across pages with the following prices: $0 (free), $0.05, $0.10, $0.25, $0.50, $1.00, $2.50, $5.00, and $10.00. These price points were selected to span from free access to clearly suppressive costs, producing complete demand curves that allow estimation of all key parameters (intensity, elasticity, breakpoint). The range was designed to capture the full behavioral response function rather than to precisely mirror any single commercial pricing structure. Underneath the question, participants responded on a slider indicating how many productivity batches they would purchase (ranging from 0 to 80). Custom JavaScript was used such that when the participant moved the slider, text above the slider would indicate "Productivity Gained: XX.XX Hours" and "Total Spent: $XX.XX." For example, at $1.00 per productivity batch, if the participant selected 40 batches, the text would show "Productivity Gained: 20.00 Hours" and "Total Spent: $40.00."

LLM attitudes and demographics

The final block of questions asked participants about their opinions on LLMs in general (not specific to any specific LLM provider or model) and about general demographic information, including ethnicity, race, education, income, age, sex/gender, and employment. In this block we also collected participants' self-rated expertise with and attitudes toward LLMs (including perceived impact on their work) and a brief free-text description of an advanced task each participant had completed with an LLM. Analysis of these measures is beyond the scope of the present paper, which focuses on the demand experiment.

Analyses

Our primary analysis focused on responses on the three HPTs using nonlinear mixed-effects models (Kaplan, 2025b; Kaplan et al., 2021). Mixed-effects models have several advantages over fixed effects only models (which are typically used for behavioral economic demand curve analyses). Notably, mixed-effects models allow us to estimate group-level parameters while also accounting for subject-level variability by including random effects.

A thorough discussion of mixed effects models can be found in (Kaplan et al., 2021).

To model the demand data, we used a variation of the Zero Bounded Exponential model (Gilroy et al., 2021) that we call a *ZBEn Log-Like 4 (LL4) variant*. This transformation preserves the interpretability of standard demand parameters (Q_0 and α), while smoothly handling zero-consumption values that arise when participants purchase no batches at higher prices. The full mathematical specification is provided in Appendix A.

We used R (R Core Team, 2025) Version 4.5.2 for MacOS and the following packages for the manuscript and analysis: *beezdemand* (Kaplan, 2025a; Kaplan et al., 2019), *qualtRics* (Ginn et al., 2025), *tidyverse* (Wickham et al., 2019), and *flextable* (Gohel & Skintzos, 2025). We also used the fantastic apaquarto extension (Schneider, n.d..) for creating this document in accordance with American Psychological Association's 7th Edition style requirements.

We use some foundational LLM models (e.g., ChatGPT 5, Opus 4.8) to assist in generating some R code, primarily related to data wrangling, visualization, and analysis, and to assist with drafting portions of this manuscript. We reviewed all LLM-generated output and always revised it when it did not meet our requirements. All final decisions, data handling, and interpretations were made by the authors. All data and code to reproduce this manuscript (produced in Quarto; Allaire et al., 2024), including a supplementary robustness analysis using a simplified exponential demand equation (Rzeszutek et al., 2025), are freely and openly available at: https://github.com/brentkaplan/be-ai-obm.

Results

General LLM preferences and usage

Table 2 shows the top and bottom LLM choices across the sample (complete preference rankings for all LLMs are provided in Appendix B). Participants frequently reported using more than one LLM, so user proportions by model reflect overlapping groups rather than mutually exclusive categories. As shown in Table 3, OpenAI GPT was the most commonly used system, with 93.7% of participants ($n = 118$) indicating any use, followed by Google Gemini (84.1%, $n = 106$) and Microsoft Co-Pilot (46.8%, $n = 59$). Among OpenAI GPT users who provided prompt estimates, the most common pattern was 1–5 prompts per day (40 (33.9%)), followed by 6–10 prompts per day (28 (23.7%)), with smaller proportions reporting no daily prompts (10 (8.5%)) or heavier use (≥21 prompts per day; 11 (9.3%), 8 (6.8%)). Other LLMs showed broadly similar prompt distributions, although overall adoption and the proportion of heavier users varied by model.

Table 2. Top and Bottom LLM choices.

LLM	Top (n)	Top (%)	Bottom (n)	Bottom (%)
OpenAI (Chat GPT)	84	66.7%	12	9.5%
Google Gemini	22	17.5%	47	37.3%
Microsoft Co-Pilot	11	8.7%	29	23.0%
xAI Grok	5	4.0%	16	12.7%
Anthropic Claude	2	1.6%	9	7.1%
Deepseek	2	1.6%	4	3.2%
Meta Llama	0	0.0%	8	6.3%
Moonshot Ai Kimi	0	0.0%	1	0.8%

Self-reported hours of use by model showed a parallel pattern. For OpenAI GPT users who answered the hours question, most reported 1–2 h of use per day (63 (53.4%)), with additional users indicating 3–4 h (23 (19.5%)) and relatively few reporting either no daily use (18 (15.3%)) or very high use (≥7 h per day; 2 (1.7%), 4 (3.4%)). Users of Google Gemini and Microsoft Co-Pilot showed similar low-to-moderate daily hours (62 (58.5%), 24 (40.7%)), again with only a small subset of respondents reporting extended daily engagement with any single model. Together, these patterns suggest that, for most respondents, LLMs functioned as tools used intermittently or for a few hours per day rather than as all-day resources.

Behavioral economic demand

Mixed model contrasts

We successfully fit the nonlinear mixed-effects model to the *LL*4 transformed productivity batch data. We incorporated both fixed and random effects for $log_{10}(\alpha)$ and $log_{10}(Q_0)$ across 126 individual participants and included the categorical predictor LLM (Top LLM, Bottom LLM, and Multi-Model LLM) as a fixed factor. Both $log_{10}(Q_0)$ and $log_{10}(\alpha)$ were allowed to vary by LLM condition, while their subject-specific deviations were modeled as independent random intercepts. Using maximum likelihood estimation, the model achieved good convergence (AIC = 909.8, BIC = 965). The relatively low residual variance (0.24) and moderate random-effect variances ($Q_0 = 0.38$, $\alpha = 0.65$) suggest stable within-subject fits and meaningful between-subject variability.

In terms of estimates of Q_0 and α, the estimated number of batches purchased at free price (Q_0) were 50.24, 58.57, and 45.82 batches, for the Top LLM, Multi-Model LLM, and Bottom LLM conditions, respectively (Table 4). Our contrasts indicated statistically significant differences in Q_0 between the Top LLM and Multi-Model LLM conditions ($p < .001$), the Top LLM and Bottom LLM conditions ($p = .00864$), and between the Multi-Model and Bottom LLM conditions ($p < .001$) (Table 5).

In terms of α, the estimated fixed effect parameter estimates were 0.01175, 0.01111, and 0.01554, for the Top LLM, Multi-Model LLM, and

Table 3. LLM daily prompts and hours by model.

			Prompts per Day						Hours per Day					
LLM	Users	Users (%)	0	1–5	6–10	11–20	21–50	51+	0	1–2	3–4	5–6	7–8	9+
OpenAI GPT	118	93.7%	10 (8.5%)	40 (33.9%)	28 (23.7%)	21 (17.8%)	11 (9.3%)	8 (6.8%)	18 (15.3%)	63 (53.4%)	23 (19.5%)	8 (6.8%)	2 (1.7%)	4 (3.4%)
Google Gemini	106	84.1%	18 (17.0%)	49 (46.2%)	22 (20.8%)	9 (8.5%)	5 (4.7%)	3 (2.8%)	29 (27.4%)	62 (58.5%)	9 (8.5%)	3 (2.8%)	0 (0.0%)	3 (2.8%)
Microsoft Co-Pilot	59	46.8%	19 (32.2%)	24 (40.7%)	7 (11.9%)	1 (1.7%)	3 (5.1%)	5 (8.5%)	24 (40.7%)	24 (40.7%)	6 (10.2%)	0 (0.0%)	3 (5.1%)	2 (3.4%)
xAI Grok	35	27.8%	6 (17.1%)	17 (48.6%)	4 (11.4%)	4 (11.4%)	0 (0.0%)	4 (11.4%)	16 (45.7%)	12 (34.3%)	5 (14.3%)	0 (0.0%)	0 (0.0%)	2 (5.7%)
Anthropic Claude	23	18.3%	7 (30.4%)	8 (34.8%)	6 (26.1%)	1 (4.3%)	0 (0.0%)	1 (4.3%)	11 (47.8%)	9 (39.1%)	1 (4.3%)	0 (0.0%)	1 (4.3%)	1 (4.3%)
Deepseek	16	12.7%	3 (18.8%)	7 (43.8%)	0 (0.0%)	4 (25.0%)	1 (6.2%)	1 (6.2%)	6 (37.5%)	8 (50.0%)	0 (0.0%)	0 (0.0%)	1 (6.2%)	1 (6.2%)
Meta Llama	13	10.3%	4 (30.8%)	4 (30.8%)	2 (15.4%)	1 (7.7%)	1 (7.7%)	1 (7.7%)	7 (53.8%)	3 (23.1%)	0 (0.0%)	1 (7.7%)	1 (7.7%)	1 (7.7%)
Vertex AI	4	3.2%	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (50.0%)	0 (0.0%)	2 (50.0%)	0 (0.0%)	0 (0.0%)	2 (50.0%)	0 (0.0%)	0 (0.0%)	2 (50.0%)
Other	3	2.4%	1 (33.3%)	1 (33.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (33.3%)	1 (33.3%)	1 (33.3%)	1 (33.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Moonshot Ai Kimi	2	1.6%	0 (0.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	0 (0.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)
Perplexity AI	2	1.6%	0 (0.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	0 (0.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)
Alibaba Qwen	1	0.8%	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)
OpenRouter	1	0.8%	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Poe	1	0.8%	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Note. Values are count (percentage) unless otherwise specified.

Bottom LLM conditions, respectively. These α values correspond to Essential Values of 0.851, 0.9, and 0.644, respectively. Contrasts indicated no statistically significant differences in α between the Top LLM and Multi-Model LLM conditions ($p = .0881$). However, we did detect statistically significant differences in α between the Top LLM and Bottom LLM conditions ($p < .001$) and between the Multi-Model and Bottom LLM conditions ($p < .001$).

Demand curves derived from the mixed model are shown in Figure 1. Descriptive statistics for batches purchased at each price by condition are provided in Appendix C. At the population level, the number of batches purchased declined systematically as price increased across all three conditions, with clear separation between Top, Multi-Model, and Bottom LLMs at low prices. At higher prices, the Top LLM showed relatively more inelastic

Table 4. Estimated marginal means for demand parameters.

LLM	Q_0 ($\log_{10}$)	Q_0 (Natural)	α ($\log_{10}$)	α (Natural)	EV
Top LLM	1.7 (1.63, 1.77)	50.24 (42.76, 59.02)	−1.93 (−2.05, −1.81)	0.0117 (0.009, 0.0154)	0.851 (0.651, 1.113)
Multi-Model LLM	1.77 (1.7, 1.84)	58.57 (49.85, 68.81)	−1.95 (−2.07, −1.84)	0.0111 (0.0085, 0.0145)	0.9 (0.689, 1.176)
Bottom LLM	1.66 (1.59, 1.73)	45.82 (38.99, 53.85)	−1.81 (−1.93, −1.69)	0.0155 (0.0119, 0.0203)	0.644 (0.492, 0.841)

Note. Q_0 = consumption at free/zero price; α = elasticity parameter; EV = essential value. Values are presented as estimate (95% CI lower bound, upper bound). $\log_{10}$ = log base-10 scale; Natural = back-transformed natural units.

Table 5. Pairwise contrasts for demand parameters.

Contrast	Parameter	Scale	Estimate	SE	df	95% CI	t	p
Top LLM - (Multi-Model LLM)	Q_0	$\log_{10}$	−0.0666	0.0151	3271	[−0.1027, −0.0305]	−4.4179	<0.0001
Top LLM – Bottom LLM		$\log_{10}$	0.04	0.0152	3271	[0.0035, 0.0764]	2.6277	0.0086
(Multi-Model LLM) – Bottom LLM		$\log_{10}$	0.1066	0.0152	3271	[0.0701, 0.143]	7.0035	<0.0001
Top LLM - (Multi-Model LLM)		Ratio	0.8578	—	—	[0.7894, 0.9322]	—	<0.0001
Top LLM – Bottom LLM		Ratio	1.0964	—	—	[1.0082, 1.1924]	—	0.0086
(Multi-Model LLM) – Bottom LLM		Ratio	1.2782	—	—	[1.1753, 1.3901]	—	<0.0001
Top LLM - (Multi-Model LLM)	α	$\log_{10}$	0.0242	0.0142	3271	[−0.0098, 0.0582]	1.7061	0.0881
Top LLM – Bottom LLM		$\log_{10}$	−0.1213	0.0145	3271	[−0.156, −0.0866]	−8.3784	<0.0001
(Multi-Model LLM) – Bottom LLM		$\log_{10}$	−0.1455	0.0143	3271	[−0.1798, −0.1113]	−10.1843	<0.0001
Top LLM - (Multi-Model LLM)		Ratio	1.0574	—	—	[0.9777, 1.1435]	—	0.0881
Top LLM – Bottom LLM		Ratio	0.7563	—	—	[0.6982, 0.8192]	—	<0.0001
(Multi-Model LLM) – Bottom LLM		Ratio	0.7153	—	—	[0.661, 0.7739]	—	<0.0001

Note. Q_0 = consumption at free/zero price; α = elasticity parameter. SE = standard error; df = degrees of freedom; CI = confidence interval; t = t-ratio; p = p-value adjusted using false discovery rate (FDR) correction. $\log_{10}$ = log base-10 scale; Ratio = back-transformed ratio scale.

demand overlapping with purchases associated with the Multi-Model LLM. Individual fits (light lines/points) mirrored this overall pattern, indicating that the model captured both the shared structure of demand and meaningful between-person variability in reinforcement value.

As a complementary view of valuation, Figure 2 presents model-predicted expenditure curves. Expenditure rose with price at lower costs, peaked at intermediate prices, and then declined as purchasing dropped off, producing the expected inverted-U demand profile. Notably, the Top LLM condition sustained marginally higher expenditure at higher prices than the Bottom LLM condition, with the Multi-Model option again occupying a middle position.

Discussion

The present study modeled workers' demand for LLMs using HPTs across three conditions: their *most preferred*, *least preferred*, and a *multi-model* platform providing access to several LLMs. Consistent with established behavioral economic principles, consumption decreased systematically as cost increased, and the resulting demand curves were well-ordered and interpretable. Demand for participants' top-preferred models was relatively inelastic, indicating a strong willingness to purchase, whereas demand for least-preferred models dropped rapidly with increasing cost. Demand for the multi-model

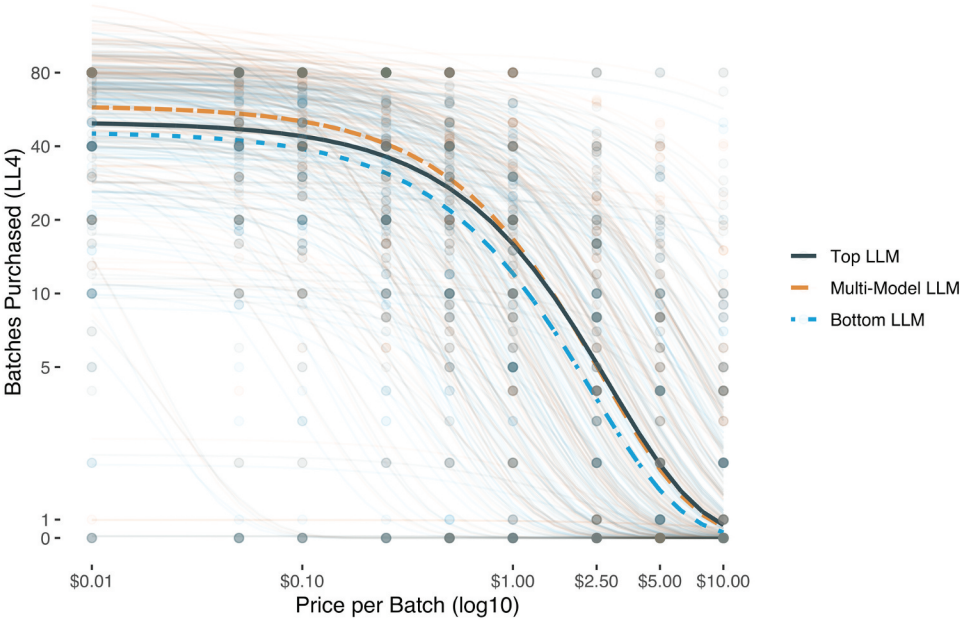

Figure 1. Mixed model predictions (fixed and random effects).

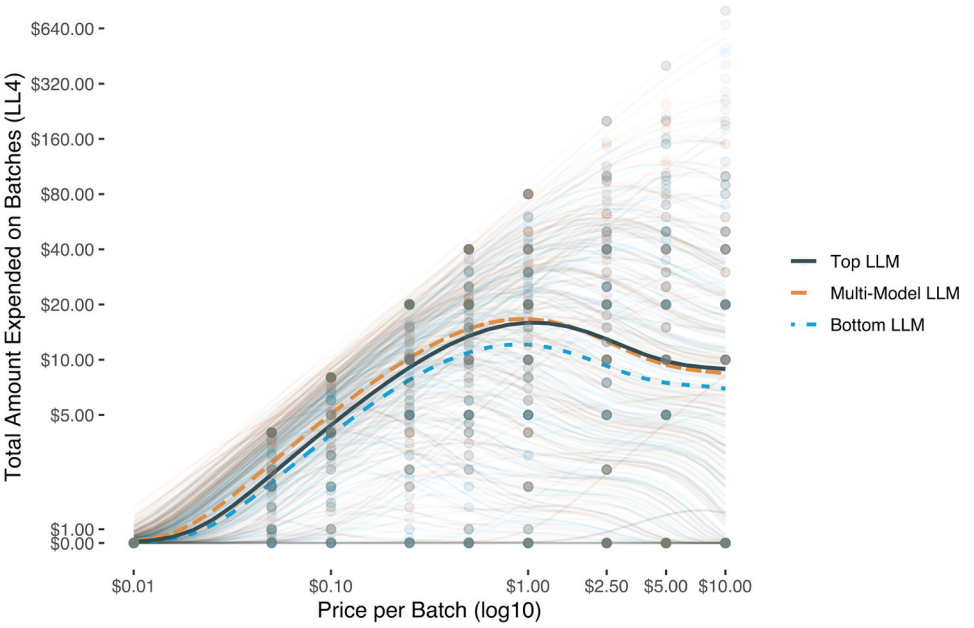

Figure 2. Mixed model predictions of expenditure (fixed and random effects).

option fell between these extremes, a pattern consistent with (though not a direct test of) substitution among LLMs.

Before discussing implications, we note that the present study examines demand among employed adults who currently use LLMs, operationalized under a direct monetary cost contingency. Our findings should therefore be interpreted as an initial application of behavioral economic purchase task methodology to LLM valuation – one in which monetary cost serves as the constraint (consistent with established purchase task methodology). Future extensions of this work might examine additional response costs relevant to workplace adoption, such as effort, time delays, or organizational policy requirements.

These participants' allocation of resources toward LLMs is consistent with operant responding governed by reinforcement history and effort requirements, at least under the simulated cost contingency studied here. In other words, LLM access appears to function as a reinforcer whose demand can be described with the same behavioral economic methods used for more traditional commodities. This is in keeping with Cox and Jennings' (2024) observation that behavior-analytic performance systems increasingly incorporate AI tools; although their analysis focused on service-delivery settings, the demand framework we apply here may extend to other organizational contexts.

It is important to acknowledge that these data should not be read as an unqualified endorsement of LLM adoption. Emerging evidence documents meaningful risks associated with LLM use in workplace contexts, including

environmental costs of training and operating large models (Ren et al., 2024), potential erosion of critical thinking skills when workers rely heavily on generated outputs (Lee, 2025), and a commonly noted concern that the effort required to verify and correct outputs may exceed time savings for some tasks. Additionally, a nationally representative Pew survey found that more workers express worry (52%) than excitement (29%) about AI's growing role in the workplace, and about a third expect AI to lead to fewer job opportunities in their field (Lin & Parker, 2025). These concerns are not captured by a demand framework that models only the price contingency, and future research should incorporate them when modeling how workers take up and sustain LLM use.

Support for the behavioral economic approach

Our quantitative results underscore the value of behavioral economics for characterizing demand for LLMs. Elasticity values showed that top-preferred LLMs maintained high demand even as costs rose, indicating low price sensitivity (a pattern analogous to that of essential commodities in traditional demand analyses). Least-preferred LLMs showed more elastic demand, consistent with more discretionary use or with participants treating these platforms as imperfect substitutes for their preferred ones. This functional differentiation has applied implications. Under the direct-cost contingency modeled here, demand for highly valued LLMs was relatively insensitive to price, whereas demand for less valued systems fell off more quickly. Whether these differences extend to non-monetary costs (e.g., the time and effort to use a tool) is an open question; if they do, less valued systems may need additional support such as training, incentives, or reduced friction to sustain use. Understanding which tools sustain relatively inelastic demand, and which are more price-sensitive, may help inform how organizations allocate LLM resources, though the present data do not identify what features drive these valuation differences.

Implications for OBM

These findings may also have implications for the ethical deployment of LLMs in workplace settings, though we note that our study examined a single contingency (direct cost) and did not measure ethical attitudes directly. Jennings and Cox (2024) note that ethical deployment of AI in behavioral services requires weighing benefits against risks, commensurate with the behavioral economic tradeoff between reinforcer magnitude and cost. Modeling such tradeoffs quantitatively enables data-based decision-making that aligns with beneficence (maximizing productivity and learning) and nonmaleficence (avoiding over reliance, bias, or harm). Moreover, despite the clear reinforcing value of LLMs, our participants' survey responses also

indicate concerns over accuracy, data security, and employer surveillance. These findings are consistent with existing data showing widespread concealment of LLM use in the workplace (Zhang et al., 2025), as well as the broader concerns over accuracy, data security, and employer surveillance documented in the literature. Such covert behavior is consistent with the possibility that some organizations may have arranged contingencies that discourage disclosure, though alternative explanations (e.g., social norms, uncertainty about policies) cannot be ruled out from these data. Interpreting the use of LLMs within a behavioral economic framework offers novel insight into the potential contingencies influencing demand.

The broader OBM literature offers practical strategies to remediate these patterns. Leaders can implement incentives and reinforcement systems to (a) ensure transparent, ethical LLM use, (b) integrate LLM safety behaviors (e.g., checking for hallucinations, securing data) into performance management systems, and (c) provide feedback and recognition for safe practices. As Jennings and Cox (2024) proposed, organizations could treat AI ethics as an *operant domain*—a set of observable, trainable, and measurable behaviors. Clear data-protection protocols, transparency about how LLM outputs influence decisions, and IT safeguards aligned with HIPAA and HITECH can support ethical engagement while reducing risk.

Our HPT results show that even modest hypothetical costs reduced the number of productivity batches participants chose to purchase. When cost was low or zero, most participants requested the maximum number of batches; as price rose, requested batches declined according to established demand functions, with individual variability captured by the nonlinear mixed-effects model. If these hypothetical patterns hold under real contingencies, organizational policies that attach cost, delay, or administrative friction to LLM access could suppress use that workers would otherwise find productive. Although we measured demand at a zero price, our design does not address whether sustained free access changes how deliberately workers use these tools.

Organizations can apply these findings behaviorally by calibrating reinforcement systems around LLM use. For example, providing free access for approved, documented purposes while requiring additional effort to use unapproved tools can differentially shape appropriate behavior. We suggest that organizations consider developing a policy for effective LLM use. This policy may include what an LLM is; what LLMs can be used (e.g., what software the organization is providing access to); what are acceptable and prohibited uses of LLMs (e.g., cannot enter confidential information, protected health information, intellectual property); how to verify the information provided (i.e., employees remain accountable for their work); security, compliance and transparency expectations (e.g., software used on company computers, disclosure of usage in documentation); and enforcement of the policy (i.e., the consequences for violation). Once a policy is developed, it could be

discussed with employees, and behavioral skills training (BST) could be used to teach safe prompting, critical review skills, and ethical usage, among other competencies. Following initial training, organizations could establish feedback mechanisms – proportionate to their context and culture – to support ongoing compliance and ethical practice. The form and intensity of such monitoring would vary across organizations and should be calibrated to avoid intrusiveness while maintaining accountability. In this sense, OBM professionals can become essential in translating complex technological change into functional, observable behavior systems.

Participants' willingness to pay for LLM access across all conditions reflects strong demand for efficiency and support in task completion. However, the same data highlight individual differences in value: not all LLM platforms were equally motivating. The heterogeneity we observed in valuation is consistent with the idea that context (organizational norms, expectations, and feedback) may shape the reinforcing value of specific LLM platforms, though we did not measure these contextual variables directly. Organizational cultures that promote experimentation, transparency, and data security will likely enhance appropriate LLM engagement, whereas those that punish disclosure will sustain secrecy and variable performance. Toward this end, periodic assessments of worker engagement and satisfaction with LLMs – whether through behavioral economic tasks, surveys, or other methods – could help organizations track shifts in demand and identify emerging concerns. Tools such as shinybeez (Kaplan & Reed, 2025), a Shiny application for conducting demand and discounting analyses without requiring programming expertise, could lower the barrier for OBM practitioners to carry out these assessments internally. By pairing these quantitative metrics with open communication and secure data infrastructure, leaders can maintain an ethical culture where LLMs serve both employee and organizational well-being and protect client PHI, when applicable.

Limitations and future directions

Several limitations warrant consideration. First, and most critically, participants were asked to imagine paying for LLM access using their own finances. In many organizational settings, employers subsidize or fully cover LLM costs. The present data, therefore, estimate demand under a direct monetary cost to the performer, which is only one of several constraints employees face. Findings should not be interpreted as a complete model of adoption in workplaces where access is employer funded. Second, selectively recruiting participants from Prolific with existing LLM knowledge over-represents technologically literate individuals comfortable with LLMs; the weekly LLM use criterion further restricts the sample to current adopters. Consequently, generalizability to broader or less tech-savvy workforces is limited. Third, the

study used HPTs, which, although validated across many domains, rely on self-reported preferences and may not fully capture real monetary contingencies (Gelino et al., 2025). Fourth, commercial LLM pricing and bundling structures evolve rapidly; the specific price points used in this study may not correspond to current or future pricing models. Additionally, many employees access LLMs through bundled platforms (e.g., enterprise suites with built-in AI features) rather than choosing individual models, and workers may use different models for different purposes – patterns not captured by the single-model HPT design. Fifth, although the LL4 modeling approach improved the handling of zero consumption and non-loglinear trends, alternative models with more robust validation (see Koffarnus et al., 2022 for an overview) may yield different elasticity estimates. Finally, our data were cross-sectional; demand for LLMs is likely to evolve over time, driven by experience, model updates, and organizational norms, necessitating longitudinal follow-up.

Building on these limitations, we suggest several targets for future research. First, additional behavioral economic experiments should impose real response costs (e.g., subscription fees, time delays) or incentivized purchasing tasks to verify whether hypothetical demand replicates under tangible contingencies. Second, future work should also examine how organizational variables – such as disclosure policies, supervision contingencies, or peer norms – influence LLM use by treating them as non-monetary "prices" in purchasing task formats. Third, future studies should recruit participants who do not regularly use LLMs, as demand among non-users may reveal important adoption barriers, attitudinal profiles, and demand variability that current-user samples cannot capture. Finally, measuring how demand shifts as workers gain LLM experience or as organizations implement LLM training is necessary to understand how experiential histories shape and sustain engagement.

Conclusion

The present findings suggest that behavioral economic demand modeling can quantify how workers who already use LLMs value that access. Participants' willingness to allocate resources to LLM assistance was orderly, predictable, and cost-sensitive under a direct monetary contingency, consistent with LLM access functioning as a measurable reinforcer within a demand framework. Applying these methods within OBM could help organizations reason about how to price, provision, and support LLM access, though the reach of such applications will depend on how well hypothetical demand under a single cost contingency translates to the many constraints workers face in practice. Pairing behavioral economic measurement with the ethical considerations raised in the behavior-analytic literature (Cox & Jennings, 2024; Jennings & Cox, 2024) offers one path toward keeping LLMs in a supporting role relative to human performance.

Author contributions

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: ***Brent A. Kaplan***: conceptualization, data curation, methodology, formal analysis, software, validation, visualization, writing – original draft, and writing – review & editing. ***Derek D. Reed***: conceptualization, funding acquisition, methodology, visualization, writing – original draft, and writing – review & editing. ***Matthew M. Laske***: writing – review & editing. ***Madison E. Graham***: data curation and writing – review & editing. ***Florence D. DiGennaro Reed***: writing – review & editing. ***Abigail Blackman***: writing – review & editing. ***Steven R. Hursh***: writing – review & editing

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Brent A. Kaplan http://orcid.org/0000-0002-3758-6776
Derek D. Reed http://orcid.org/0000-0002-5854-3425
Matthew M. Laske http://orcid.org/0000-0003-0195-5186
Madison E. Graham http://orcid.org/0000-0002-2353-8415
Florence D. DiGennaro Reed http://orcid.org/0000-0001-9143-8275
Steven R. Hursh http://orcid.org/0000-0002-5391-2842

References

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2024). Quarto. Zenodo. https://doi.org/10.5281/zenodo.5960048

Bick, A., Blandin, A., & Deming, D. J. (2025, November 13). *The state of generative AI adoption in 2025*. Federal Reserve Bank of St. Louis. https://www.stlouisfed.org/on-the-economy/2025/nov/state-generative-ai-adoption-2025

Bick, A., Blandin, A., & Deming, D. J. (2026). The rapid adoption of generative AI. *Management Science*. https://doi.org/10.1287/mnsc.2025.02523

Brachman, M., El-Ashry, A., Dugan, C., & Geyer, W. (2025). Current and future use of large language models for knowledge work. *Proceedings of the ACM on Human-Computer Interaction*, *9*(7), 1–24. https://doi.org/10.1145/3757403

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). *How people use ChatGPT* (Working Paper 34255). National Bureau of Economic Research. https://www.nber.org/papers/w34255

Cox, D. J., & Jennings, A. M. (2024). The promises and possibilities of artificial intelligence in the delivery of behavior analytic services. *Behavior Analysis in Practice*, *17*(1), 123–136. https://doi.org/10.1007/s40617-023-00864-3

DiGennaro Reed, F. D., Henley, A. J., Hirst, J. M., Doucette, J. L., & Jenkins, S. R. (2015). Basic research considerations for performance management of staff. In F. D. DiGennaro Reed & D. D. Reed (Eds.), *Autism service delivery: Bridging the gap between science and practice* (pp. 437–463). Springer. https://doi.org/10.1007/978-1-4939-2656-5_16

Gelino, B. W., Reed, D. D., Naudé, G. P., Harbaugh, J. C., Schlitzer, R. D., Mercincavage, M., Strasser, A. A., Johnson, M. W., & Strickland, J. C. (2025). Hypothetical purchase task metrics predict responding on an incentivized purchase task for full and reduced nicotine cigarettes. *Nicotine & Tobacco Research*, *27*(12), 2281–2288. https://doi.org/10.1093/ntr/ntaf132

Gilroy, S. P., Kaplan, B. A., Schwartz, L. P., Reed, D. D., & Hursh, S. R. (2021). A zero-bounded model of operant demand. *Journal of the Experimental Analysis of Behavior*, *115*(3), 729–746. https://doi.org/10.1002/jeab.679

Ginn, J., O'Brien, J., & Silge, J. (2025). qualtRics: Download 'Qualtrics' survey data, 3.2.2. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.qualtRics

Gohel, D., & Skintzos, P. (2025). *Flextable: Functions for tabular reporting* 0.9.10. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.flextable

Henley, A. J., DiGennaro Reed, F. D., Kaplan, B. A., & Reed, D. D. (2016). Quantifying efficacy of workplace reinforcers: An application of behavioral economic demand to evaluate hypothetical work performance. *Translational Issues in Psychological Science*, *2*(2), 174–183. https://doi.org/10.1037/tps0000068

Henley, A. J., DiGennaro Reed, F. D., Reed, D. D., & Kaplan, B. A. (2016). A crowdsourced nickel-and-dime approach to analog OBM research: A behavioral economic framework for understanding workforce attrition. *Journal of the Experimental Analysis of Behavior*, *106*(2), 134–144. https://doi.org/10.1002/jeab.220

Hirst, J. M., & DiGennaro Reed, F. D. (2016). Cross-commodity discounting of monetary outcomes and access to leisure activities. *Psychological Record*, *66*(4), 515–526. https://doi.org/10.1007/s40732-016-0201-4

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review*, *115*(1), 186–198. https://doi.org/10.1037/0033-295X.115.1.186

Jennings, A. M., & Cox, D. J. (2024). Starting the conversation around the ethical use of artificial intelligence in applied behavior analysis. *Behavior Analysis in Practice*, *17*(1), 107–122. https://doi.org/10.1007/s40617-023-00868-z

Kaplan, B. A. (2025a). beezdemand: Behavioral Economic Easy Demand 0.1.3 (GitHub). https://github.com/brentkaplan/beezdemand

Kaplan, B. A. (2025b). Quantitative models of operant demand. In D. D. Reed, B. A. Kaplan, & S. P. Gilroy (Eds.), *Handbook of operant behavioral economics: Demand, discounting, methods, and applications* (pp. 91–130). Elsevier.

Kaplan, B. A., Foster, R. N. S., Reed, D. D., Amlung, M., Murphy, J. G., & MacKillop, J. (2018). Understanding alcohol motivation using the alcohol purchase task: A methodological systematic review. *Drug & Alcohol Dependence*, *191*, 117–140. https://doi.org/10.1016/j.drugalcdep.2018.06.029

Kaplan, B. A., Franck, C. T., McKee, K., Gilroy, S. P., & Koffarnus, M. N. (2021). Applying mixed-effects modeling to behavioral economic demand: An introduction. *Perspectives on Behavior Science*, *44*(2), 333–358. https://doi.org/10.1007/s40614-021-00293-w

Kaplan, B. A., Gilroy, S. P., Reed, D. D., Koffarnus, M. N., & Hursh, S. R. (2019). The R package beezdemand: Behavioral economic easy demand. *Perspectives on Behavior Science*, *42*(1), 163–180. https://doi.org/10.1007/s40614-018-00187-7

Kaplan, B. A., & Reed, D. D. (2025). Shinybeez: A shiny app for behavioral economic easy demand and discounting. *Journal of the Experimental Analysis of Behavior*, *123*(2), 355–376. https://doi.org/10.1002/jeab.70000

Koffarnus, M. N., Kaplan, B. A., Franck, C. T., Rzeszutek, M. J., & Traxler, H. K. (2022). Behavioral economic demand modeling chronology, complexities, and considerations: Much ado about zeros. *Behavioural Processes*, *199*, 104646. https://doi.org/10.1016/j.beproc.2022.104646

Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Lin, L., & Parker, K. (2025). *Workers' views of AI use in the workplace*. Pew Research Center. https://www.pewresearch.org/social-trends/2025/02/25/workers-views-of-ai-use-in-the-workplace/

Lovich, D., Meier, S., & Taylor, C. (2025). *Leaders assume employees are excited about AI. They're wrong*. Harvard Business Review. https://hbr.org/2025/11/leaders-assume-employees-are-excited-about-ai-theyre-wrong

Miller, B. P., Reed, D. D., & Amlung, M. (2023). Reliability and validity of behavioral-economic measures: A review and synthesis of discounting and demand. *Journal of the Experimental Analysis of Behavior*, *120*(2), 263–280. https://doi.org/10.1002/JEAB.860

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reed, D. D., Graham, M. E., & Acuff, S. F. (2025). Introduction to operant demand. In D. D. Reed, B. A. Kaplan, & S. P. Gilroy (Eds.), *Handbook of operant behavioral economics: Demand, discounting, methods, and applications* (pp. 33–49). Elsevier.

Reed, D. D., Kaplan, B. A., & Gilroy, S. P. (Eds.). (2025). *Handbook of operant behavioral economics: Demand, discounting, methods, and applications*. Elsevier.

Reed, D. D., Naudé, G. P., Gelino, B. W., & Amlung, M. (2020). Behavioral economic considerations of novel addictions and nonaddictive behavior: Research and analytic methods. In S. Sussman (Ed.), *The Cambridge handbook of substance and behavioral addictions* (pp. 73–86). Cambridge University Press. https://doi.org/10.1017/9781108632591.011

Reed, D. D., Niileksela, C. R., & Kaplan, B. A. (2013). Behavioral economics: A tutorial for behavior analysis in practice. *Behavior Analysis in Practice*, *6*(1), 34–54. https://doi.org/10.1007/BF03391790

Ren, S., Tomlinson, B., Black, R. W., & Torrance, A. W. (2024). Reconciling the contrasting narratives on the environmental impact of large language models. *Scientific Reports*, *14*(1), 26310. https://doi.org/10.1038/s41598-024-76682-6

Rzeszutek, M. J., Regnier, S. D., Franck, C. T., & Koffarnus, M. N. (2025). Overviewing the exponential model of demand and introducing a simplification that solves issues of span, scale, and zeros. *Experimental and Clinical Psychopharmacology*, *33*(2), 209–223. https://doi.org/10.1037/pha0000754

Schneider, W. J. (n.d.). Apaquarto. GitHub. https://github.com/wjschne/apaquarto

Strickland, J. C., Reed, D. D., Hursh, S. R., Schwartz, L. P., Foster, R. N. S., Gelino, B. W., LeComte, R. S., Oda, F. S., Salzer, A. R., Schneider, T. D., Dayton, L., Latkin, C., & Johnson, M. W. (2022). Behavioral economic methods to inform infectious disease response: Prevention, testing, and vaccination in the COVID-19 pandemic. *PLOS ONE*, *17*(1), e0258828–e0258828. https://doi.org/10.1371/journal.pone.0258828

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Woo, K. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wine, B., Gilroy, S., & Hantula, D. A. (2012). Temporal (in)stability of employee preferences for rewards. *Journal of Organizational Behavior Management*, *32*(1), 58–64. https://doi.org/10.1080/01608061.2012.646854

Zhang, Z., Shen, C., Yao, B., Wang, D., & Li, T. (2025). Secret use of large language model (LLM). *Proceedings of the ACM on Human-Computer Interaction*, *9*(2), 1–26. https://doi.org/10.1145/3711061

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., & Wen, J.-R. (2023). A survey of large language models. *Frontiers of Computer Science*, *20*, 2012627. https://doi.org/10.48550/arXiv.2303.18223

Appendices

Appendix A

ZBEn Log-Like 4 (LL4) Variant – Mathematical Specification

In their formulation, Gilroy et al. (2021) used an Inverse Hyperbolic Sine (IHS) transformation for the dependent variable in lieu of standard logarithmic transformation in models such as the original exponential model (Hursh & Silberberg, 2008). Instead of using the IHS transformation, which is limited in that its scaling is *identical to log_{10} for values greater than 10* but transitions to a *hyperbolic curve for values less than 10*, we used a variation shown in Equation 1.

$$LL4(x;\lambda,b) = \frac{\log_b\left(x^{\lambda}+1\right)}{\lambda} \tag{1}$$

where λ is the *exponential on consumption* (in this case, 4) and b is the logarithmic *base* (in this case, 10). In more simple terms, this transformation is:

$$LL4(x) = \frac{\log_{10}(x^4+1)}{4} \tag{2}$$

This log-like transformation is identical to the $\log_{10}$ transformation for values ≥ 2 and smoothly transitions to zero as a (4th degree) polynomial curve for values < 2. **In order to keep the interpretability of Q_0 and α consistent with the $\log_{10}$ transformation, we keep the right-hand side of the equation the same**. Thus, the demand curve equation with normalized α used in this analysis (*ZBEn Log-Like 4 (LL4) variant*) is shown in Equation 3:

$$LL4(y) = \log_{10}(Q_0) \cdot e^{-\frac{\alpha}{\log_{10}(Q_0)} \cdot Q_0 \cdot x} \tag{3}$$

Appendix B

LLM Preference Rankings

LLM	Number of Rankings	Mean Ranking	Median Ranking
OpenAI GPT	118	1.43	1
Google Gemini	106	2.10	2
Microsoft Co-Pilot	59	2.49	2
xAI Grok	35	2.74	3
Anthropic Claude	23	2.83	3
Deepseek	16	3.06	3
Meta Llama	13	4.54	4
Alibaba Qwen	1	8.00	8
Moonshot Ai Kimi	2	8.00	8

Appendix C

Descriptive Statistics for Productivity Batches Purchased by LLM Ranking

Condition	Price	Mean	Median	SD	Prop. zeros	NAs	Min	Max
Top LLM	$0.00	59.02	80.0	27.79	0.03	0	0	80
	$0.05	57.75	80.0	27.13	0.04	0	0	80
	$0.10	53.44	60.0	27.99	0.04	0	0	80
	$0.25	43.82	40.0	29.41	0.06	0	0	80
	$0.50	31.02	23.5	26.13	0.11	0	0	80
	$1.00	21.06	11.5	23.31	0.13	0	0	80
	$2.50	9.98	4.0	14.45	0.24	0	0	80
	$5.00	5.76	1.5	11.07	0.41	0	0	80
	$10.00	3.70	0.0	10.53	0.57	0	0	80
Multi-Model LLM	$0.00	67.67	80.0	23.93	0.02	0	0	80
	$0.05	63.40	80.0	25.83	0.04	0	0	80
	$0.10	58.50	80.0	27.22	0.05	0	0	80
	$0.25	47.36	42.5	29.36	0.06	0	0	80
	$0.50	33.55	30.5	26.84	0.12	0	0	80
	$1.00	21.63	15.0	22.83	0.19	0	0	80
	$2.50	11.38	4.0	17.24	0.35	0	0	80
	$5.00	6.54	1.0	12.65	0.46	0	0	80

(*Continued*)

Condition	Price	Mean	Median	SD	Prop. zeros	NAs	Min	Max
	$10.00	3.45	0.0	9.54	0.56	0	0	80
Bottom LLM	$0.00	57.61	80.0	28.95	0.04	0	0	80
	$0.05	56.21	80.0	29.60	0.06	0	0	80
	$0.10	50.64	59.0	29.42	0.06	0	0	80
	$0.25	39.89	40.0	29.37	0.09	0	0	80
	$0.50	28.28	20.0	25.98	0.13	0	0	80
	$1.00	16.90	10.0	18.85	0.21	0	0	80
	$2.50	8.71	3.5	14.86	0.38	0	0	80
	$5.00	5.46	1.0	11.93	0.44	0	0	80
	$10.00	3.13	0.0	9.81	0.61	0	0	80