

Measuring How Americans Use Artificial Intelligence: Recommendations for the AMERICAN TIME USE SURVEY (ATUS) AI Questions

LIYA PALAGASHVILI

Senior Research Fellow, Director of the Labor Policy Project, Mercatus Center at George Mason University

REVANA SHARFUDDIN

Research Fellow, Mercatus Center at George Mason University

YULIYA YATSYSHINA

Economic Researcher, Mercatus Center at George Mason University

Proposed Information Collection; ATUS Artificial Intelligence (AI) Questions

Agency: US Department of Labor, Bureau of Labor Statistics

Comment Period Opens: July 10, 2026

Comment Period Closes: September 8, 2026

Comment Submitted: September 8, 2026

OMB No. 1220-NEW

Introduction

The Bureau of Labor Statistics (BLS) proposes to add a two-minute artificial intelligence module to the American Time Use Survey beginning in January 2027. The module would produce nationally representative data linking AI use to a 24-hour activity diary.¹ We are grateful for the opportunity to submit comments on this proposed collection. We are members of the Labor Policy Project at the Mercatus Center at George Mason University, which is dedicated to advancing knowledge about the effects of institutions and public policy on society. Our work examines how technology and policy shape labor markets. Accordingly, our comment is designed to help BLS develop measures that improve understanding of how AI affects work and daily activities while limiting respondent burden.

We support the collection. Existing evidence measures firm adoption, self-reported use at work and home using broad categories, or activity on individual AI platforms. These sources do not provide a nationally representative link between AI use and a continuous record of work, education, household production, care, and leisure. That link is the proposed module's central practical value. This leaves a

¹ Bureau of Labor Statistics, "Proposed Information Collection; ATUS Artificial Intelligence (AI) Questions," 91 Fed. Reg. 42775-76 (July 10, 2026), <https://www.govinfo.gov/content/pkg/FR-2026-07-10/html/2026-13928.htm>.

gap in understanding how AI use relates to workers' activities and the allocation of time across work and nonwork domains.

Prior research has raised the possibility that irregular or informal income-generating activity is not always reported as work in CPS-style questions. BLS has used ATUS data to examine this possibility, finding that some respondents classified as unemployed or not in the labor force reported income-generating activities, while also finding that such activity was relatively uncommon on a given day.² The proposed AI module could extend this measurement work by identifying whether AI-assisted activity occurs within ATUS's other income-generating activity categories and whether patterns differ by labor-force status; it would not by itself establish the size or growth of informal or creator-economy activity.

Recent congressional testimony by members of our team likewise emphasizes the value of distinguishing occupational exposure from reported AI use and measuring changes in tasks, duties, hours, training, and work arrangements.³ These considerations reinforce the need for worker- and activity-based measures alongside firm-side adoption data.

Our recommendations address four priorities: defining AI use and the headline estimates; linking use to diary episodes without overstating within-job task detail; preserving work-arrangement information and public-use linkage; and validating burden, nonresponse, and comparability. We prioritize behavior over general attitudes, while recognizing that ATUS itself relies on self-reported recall. Because the notice does not reproduce the draft questionnaire, these recommendations address the construct, the measurement approach, and the documentation and dissemination plan rather than final question wording. Where a recommendation is already reflected in the draft instrument, we intend it as support for that choice. We encourage BLS to make the proposed instrument and available testing results accessible so these recommendations can be evaluated against its actual design.⁴

Summary of Recommendations

The recommendations below distinguish the core measures we consider essential to the module's value, conditional additions that should depend on realistic interview testing, and implementation requirements for documentation, precision, and continuity. BLS should define the headline statistics and evaluate average burden across all response paths before finalizing the instrument.

Core measures (we recommend these be included in the fielded module):

1. A short, tested operational definition of respondent-recognized, deliberate AI use, including stand-alone tools and AI features integrated into other software, supported by tested inclusion and exclusion examples.
2. A diary-day use screener, followed for users by a summary pass over the completed diary that identifies the AI-assisted episode or episodes and allows more than one selection. Any distinction

² Mary Dorinda Allard and Anne E. Polivka, "Measuring Labor Market Activity Today: Are the Words 'Work' and 'Job' Too Limiting for Surveys?" *Monthly Labor Review*, US Bureau of Labor Statistics, November 14, 2018, <https://www.bls.gov/opub/mlr/2018/article/measuring-labor-market-activity-today.htm>.

³ Revana Sharfuddin, "Building an AI-Ready America: Adopting AI at Work," testimony before the US House Committee on Education and Workforce, Subcommittee on Health, Employment, Labor, and Pensions, February 3, 2026, <https://www.mercatus.org/research/federal-testimonies/building-ai-ready-america-adopting-ai-work>; Liya Palagashvili, "AI and the American Workforce: Evidence, Measurement Gaps, and a Path Forward," testimony before the US Senate Committee on Health, Education, Labor, and Pensions, Subcommittee on Employment and Workplace Safety, July 29, 2026, <https://www.help.senate.gov/download/palagashvili-testimony.pdf>.

⁴ Bureau of Labor Statistics, "ATUS Artificial Intelligence (AI) Questions," 91 Fed. Reg. at 42775–76.

between assistance with the recorded activity and unrelated concurrent AI use should be included only if testing shows that it can be made reliably within the burden target.

3. Response codes that separately identify reported use, reported nonuse, uncertainty, and missing responses, without requiring respondents to classify use as “direct” versus “embedded.”
4. Frequency of use over a clearly defined past-seven-day reference period, asked of all eligible respondents, including diary-day nonusers, using a numerical count or nonoverlapping numerical categories.
5. For work-related and other income-generating use, identification of the relevant job or activity using existing information first, with a targeted follow-up only where needed to distinguish work arrangements; skip patterns should ensure the module reaches respondents who report only other income-generating activity on the diary day, not only those who report conventional main-job or other-job work.

Conditional measures (recommended only if cognitive and operational testing show acceptable burden):

1. A short task-purpose item for AI-assisted work episodes, supplementing the existing activity code if within-job task analysis is an objective.
2. Employer policy or tool/training access, with a clear job reference and appropriate coverage of users and nonusers.
3. Separately tested measures of deliberate versus automatic assistance, or the extent of respondent review of AI output, only if they add value beyond the core.
4. A self-reported time-effect item with symmetric response options, treated as the lowest-priority addition.

Implementation and documentation requirements:

1. BLS should release linkable person-level and episode-level AI measures, with clear estimands, denominators, weights, variance-estimation guidance, and full documentation of wording, universes, skip patterns, editing, and imputation. Documentation should identify which CPS labor-force-status and job-desire variables are available for analysis, their reference dates, their applicable universes, and appropriate sample-size caveats for not-in-the-labor-force subgroups.
2. Before fielding, BLS should publish evidence from cognitive and operational testing, including completion times across response paths. After the first year, BLS should report module and item nonresponse, multiple-episode reporting, and achieved precision for key subgroup estimates. Comparisons involving other income-generating activities and narrow labor-force-status subgroups should be reported only when sample sizes support them, with pooling or broader categories used where necessary.
3. BLS should establish a stable core for 2027 and 2028—including the operational definition, diary-day use measure, episode attribution, broader-period frequency measure, and essential work/activity linkage—and document any changes in wording, examples, coding, editing, imputation, or weighting. Any revision should be accompanied by a comparability or bridge analysis where feasible, and BLS should consider retaining a short recurring core after 2028.

Measures we do not recommend within a two-minute module: General attitudes toward AI, optimism or concern, and expectations about future job displacement. Other surveys collect these. ATUS's comparative advantage is the measurement of activity and time.

I. Preserve ATUS's Distinctive Research Value

The module should fill a gap rather than duplicate existing collections, while retaining useful overlap for comparisons. The Census Bureau's Business Trends and Outlook Survey and its AI supplements collect firm-side information on AI use, business functions, organizational change, training, and reported or expected employment effects.⁵ These data show adoption on the employer side, but a firm may report AI use even when few workers use it. Workers may also adopt AI without a formal firm initiative. Firm adoption and worker use are related but distinct outcomes.

Nationally representative household surveys already measure generative AI use at work and outside work, frequency, products used, assisted tasks, reported time savings, and workplace support or barriers.⁶ Platform data can classify AI conversations by work status and task, but they cover users of a particular service and do not provide a full-day record of time use.⁷

ATUS can add a diary-linked measure that those sources lack. Its advantage is structured recall and linkage to a full-day activity record, not direct observation or freedom from reporting error. It can show whether respondents report AI use for work or nonwork activities, and how use is distributed across education, household production, caregiving, job search, and leisure. It can also distinguish reported use from estimates of occupational exposure. Exposure measures describe whether a technology could affect a task; they do not show whether a person used it.⁸

ATUS's reach beyond employed workers is an underappreciated asset for this module. The CPS uses a reference-week labor-force framework and is designed to classify a person as employed if they report any work for pay or profit. Nevertheless, intermittent activities, activities respondents do not identify as "work," or income-generating activity outside a conventional job may not be fully captured or separately identified. BLS has used ATUS data to examine this possibility, finding that some respondents classified as unemployed or not in the labor force reported income-generating activities, while also finding that such activity was relatively uncommon on a given day and had not markedly increased over the years studied.⁹

Because ATUS interviews respondents regardless of labor-force status and records a full-day diary, the AI module could extend this measurement work by identifying whether AI-assisted activity occurs within income-generating categories and whether patterns differ by labor-force status. This should be treated as a descriptive measurement opportunity, not evidence that official statistics omit a large or growing creator economy.

⁵ Kathryn Bonney et al., "Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey," US Census Bureau Center for Economic Studies Working Paper 24-16R, 2024, <https://www2.census.gov/ces/wp/2024/CES-WP-24-16R.pdf>; Adam Grundy, Cory Breau, and Dhanapati Khatiwoda, "Large Firms With at Least 20 Employees Biggest AI Users," America Counts, US Census Bureau, May 26, 2026, <https://www.census.gov/library/stories/2026/05/ai-use-businesses.html>.

⁶ Alexander Bick, Adam Blandin, and David J. Deming, "The Rapid Adoption of Generative AI," Federal Reserve Bank of St. Louis Working Paper 2024-027F, revised October 2025, <https://doi.org/10.20955/wp.2024.027>.

⁷ Aaron Chatterji et al., "How People Use ChatGPT," NBER Working Paper 34255, September 2025, <https://www.nber.org/papers/w34255>.

⁸ Palagashvili, "AI and the American Workforce."

⁹ Allard and Polivka, "Measuring Labor Market Activity Today."

Before finalizing individual questions, BLS should specify the principal statistics: the share reporting AI use on the diary day; the share reporting use in the past seven days; the share reporting work-related AI use among respondents who worked on the diary day; and the duration of episodes marked as AI-assisted. These quantities have different denominators and interpretations. In particular, the last is neither minutes spent interacting with AI nor time saved. BLS should distinguish estimates for all eligible people, employed people, and people who actually worked on the diary day, and document the population and reference period for each.

The main economic value of this design is descriptive: it would link reported AI use to the timing and type of activities, work arrangements, and labor-force status. These data could support comparisons across activities and population groups (including work, education and training, job search, household production, care, and leisure), but they would not by themselves measure output, productivity, time saved, task composition within broad work episodes, or individual before-and-after changes in time use. The strongest justification for placing the module in ATUS is precisely this ability to link reported AI use to time and activity while retaining the survey's demographic and labor-market information. Detailed task composition and on-the-job training require additional information, as discussed below.

II. Link AI Use to Specific Diary Activities

The activity episode, rather than the occupation or the AI product, should anchor the diary linkage. An occupation-level question would obscure substantial variation within occupations. The same worker may use AI for writing, information search, coding, scheduling, or training and not use it for meetings, physical tasks, or care. Product-based questions would also age quickly and would not reveal what the respondent did. However, a diary episode is not necessarily a discrete task. The existing work codes distinguish main-job and other-job work but do not separately identify writing, coding, meetings, or training within those jobs. An AI flag on a broad work episode therefore cannot, by itself, identify which task AI assisted.¹⁰

BLS should attach an AI-use indicator to each relevant diary episode. For an AI-assisted episode, the public-use file should retain the underlying ATUS activity code, start and stop time, work or nonwork classification, and any existing location and co-presence variables that can be disclosed. BLS should not replace the ATUS activity code with a separate list of AI tasks. The value comes from adding AI use to the existing diary structure. If within-job task analysis is a priority, BLS should test a short task-purpose follow-up alongside the existing codes. Such an item could identify the kinds of work AI assisted without requiring a full reconstruction of the workday. It would identify reported purposes, not the time devoted to each task. Without it, research claims should be limited to broad activities and work episodes.

A candidate low-burden sequence could first give the tested definition and ask about AI use during the ATUS diary day, from 4 a.m. to 4 a.m.¹¹ If use is reported, the interviewer could review the completed diary and identify the relevant episode or episodes. The question should allow uncertainty and should not require respondents to classify the technology before reporting recognizable uses. Only

¹⁰ Bureau of Labor Statistics, "American Time Use Survey Activity Lexicon," 2025, <https://www.bls.gov/tus/lexicons/lexiconnoex2025.pdf>.

¹¹ Bureau of Labor Statistics, "American Time Use Survey User's Guide: Understanding ATUS 2003 to 2025," June 2026, <https://www.bls.gov/tus/atususersguide.pdf>.

diary-episode follow-ups should be skipped for diary-day nonusers; the past-seven-day frequency question should still reach them. This sequence would anchor recall in activities the respondent has already reported, but its burden must be tested.

BLS should permit more than one episode to be selected. Restricting the response to a main use would understate the number and breadth of AI-assisted episodes, even if person-level use prevalence were unchanged. The instrument should also make clear that AI use may be brief within a longer episode. The duration of an episode should not automatically be interpreted as the number of minutes spent interacting with AI.

This last point matters for economic interpretation. An AI tool may affect the duration of a task even if the interaction lasts only a few minutes. Conversely, an hour-long work episode marked as AI-assisted does not mean AI was used for the full hour. BLS should document the indicator as “AI used during the episode,” not “time spent using AI.”

The follow-up should distinguish assistance with the recorded activity from unrelated AI use occurring at the same time. For example, using AI for personal planning during a work meeting should not automatically be classified as AI-assisted work. Because ATUS generally records a primary activity rather than a complete inventory of simultaneous activities, BLS should test a limited way to identify the purpose of such use without changing the underlying diary code.¹² Nor should unattended processing by an AI system be counted as additional human activity time.

III. Define AI So Respondents Report the Same Concept

The term artificial intelligence is too broad to support a reliable prevalence measure without an operational definition. It can refer to generative tools that a person actively prompts, predictive systems, recommendation engines, or features embedded in ordinary software. Respondents may recognize some uses and remain unaware of others.

The Census Bureau notes that businesses may fail to report incidental or embedded AI because they do not know that a third-party system uses it. Research on individual adoption raises the same concern: survey responses generally capture use the respondent recognizes, not passive or hidden AI use. Palagashvili likewise emphasizes that definitions vary across surveys and that workers may not know AI is embedded in the software they use.¹³

BLS should define the headline measure as respondent-recognized, deliberate use of an AI tool or feature, whether the tool is stand-alone or integrated into other software. Product integration, deliberate versus automatic operation, and certainty that AI was involved are different dimensions. The core measure should capture deliberate use that the respondent recognizes; embedded or uncertain use can be retained as separate categories only if testing shows that respondents can report it reliably within the burden constraint. These categories should not be combined into a single prevalence estimate without reporting their components.

The instrument should use a short operational definition supported by tested examples. BLS should clarify whether it intends to cover generative AI alone or other AI applications as well. If generative use is the target, examples can describe deliberately asking an AI system to draft, summarize, translate, or create content. If other applications are included, BLS should specify their inclusion criteria and test the broader definition. Descriptions such as “transforms text” or “provides

¹² Bureau of Labor Statistics, “ATUS User’s Guide.”

¹³ Bonney et al., “Tracking Firm Use of AI in Real Time”; Bick, Blandin, and Deming, “The Rapid Adoption of Generative AI”; Palagashvili, “AI and the American Workforce.”

predictions or recommendations” are too broad on their own: ordinary editing tools, formulas, and fixed-rule systems may perform those functions. We do not propose final wording before reviewing the instrument and testing evidence.

The core wording should describe functions instead of brands. Product examples can help recall, but they will become outdated and may cue respondents toward consumer chatbots while excluding specialized workplace tools. BLS should place current examples in interviewer instructions or help text with standardized rules for when examples are read. Any change in examples should be tested and documented because it can change reported prevalence even when the core question is unchanged.

BLS should also separate AI use from ordinary software use. Search engines, automated spelling correction, calculators, conventional data analysis, and fixed-rule automation should not be included merely because respondents view them as advanced technology. Tested inclusion and exclusion examples should aim to reduce false positives and improve comparisons across demographic groups with different levels of technical familiarity. Exclusions should concern the function used, not an entire product: a conventional search action and deliberate use of an AI answer feature within the same product may fall on different sides of the operational definition.

IV. Measure Frequency and Workplace Context

A yes-or-no measure cannot distinguish experimentation from regular integration into daily activity. The module should distinguish use on the diary day from use during a broader reference period, such as the past seven days, and should record frequency within that period. The diary-day measure supports activity analysis. The broader measure describes a different quantity, not a correction to diary-day nonuse. A regular weekday user may correctly report no use on a sampled Sunday. All eligible respondents, including diary-day nonusers, should receive the broader-period question.

The broader frequency measure should use a numerical count, such as the number of days used from zero through seven, or tested, nonoverlapping numerical categories. BLS should avoid terms such as “regularly,” “occasionally,” “several days,” or “almost every day,” which respondents may interpret differently. The reference period should be defined explicitly, including its endpoints and relationship to the diary day. BLS should publish both diary-day and reference-period estimates rather than treating one as a substitute for the other. Days of use measure frequency, not hours of interaction or the intensity of assistance.

Work-related AI use should be classifiable by work arrangement. ATUS already distinguishes main-job work, other-job work, and other income-generating activities, and its employment questions update some information from the earlier CPS interview. BLS should first document which existing variables identify the job or business associated with an AI-assisted episode and the timing of those characteristics. Main-job class of worker should not automatically be assigned to secondary work. Where a gap remains, BLS should test a targeted follow-up, preferably once for the relevant job or activity rather than repeatedly for each episode.¹⁴ Self-employment, independent contracting, and informal income-generating activity are related but not interchangeable; any requested distinction needs its own clear definition. The aim is to describe AI use across work arrangements, not to determine legal worker status or infer individual transitions from a single diary.

¹⁴ Bureau of Labor Statistics, “ATUS User’s Guide”; Bureau of Labor Statistics, “ATUS Activity Lexicon”; Bureau of Labor Statistics, “American Time Use Survey Questionnaire, 2011–25,” June 2026, <https://www.bls.gov/tus/questionnaires/tuquestionnaire.pdf>.

The module should explicitly extend AI-use questions to ATUS's existing category of other income-generating activities, not only to episodes coded as main-job or other-job work. ATUS already distinguishes this category in its activity lexicon and diary coding, and BLS has used these data to examine possible discrepancies between diary-reported income-generating activity and CPS-style labor-force status.¹⁵ The lexicon includes income-generating hobbies, crafts, performances, services, and other income-generating activities; BLS should document which activity codes and verbatim responses are included. An AI-use indicator attached to these episodes would allow researchers to examine whether AI-assisted activity occurs outside conventional employment and whether patterns differ by labor-force status. BLS should confirm that the diary-linkage sequence and skip patterns reach respondents who report only other income-generating activity, rather than routing them out because they do not report conventional work.

If burden testing permits an employer-policy item, it should refer to a specified job and clearly defined AI use. Categories should allow permission subject to limits, prohibition, no communicated policy, and uncertainty; encouragement should not be treated as the opposite of restrictions. To study barriers to adoption, eligible employees who report no AI use must also receive the relevant policy or access question. If the item is asked only of users, the resulting analysis should be limited to users.¹⁶

If testing shows room within the average burden target, BLS could test a job-specific item on tool access or training. These are distinct constructs and should not be combined into an ambiguous response. A worker who does not use AI because the employer prohibits it is different from a worker who has access but sees no useful application, lacks training, or prefers not to use it. These distinctions help separate technology availability from individual adoption and can inform research on differences across firms and workers. Such information is desirable, not required for the diary-linked core.

A compact task-role item could be considered only after the higher-priority task-purpose and context items. Producing output that the respondent reviews is itself a form of assistance, so these should not be presented as mutually exclusive categories without further testing. A question about the extent of respondent review or revision may be more interpretable, but it would remain a self-report and would not measure automation or job displacement. OECD worker research examines task automation, new tasks, training, and worker consultation; these broader constructs need not all be included in this short module.¹⁷

V. Support Economic Analysis Without Overstating What the Data Show

The module can materially improve economic research, but the questionnaire and documentation should distinguish outcomes the data measure from outcomes researchers may infer. ATUS measures time allocation, not output, quality, wages produced during an episode, or the counterfactual time the same activity would have required without AI. An AI-assisted episode that is shorter may reflect time savings, task selection, worker skill, job design, or other differences.

A self-reported time-savings item is a low-priority addition that could add useful information if BLS treats it as a respondent assessment. The item should ask about a specific AI-assisted activity, define a clear comparison, and allow “less time,” “about the same time,” “more time,” and “cannot estimate.” Asking only how much time was saved presumes a benefit and would bias the result. BLS

¹⁵ Allard and Polivka, “Measuring Labor Market Activity Today”; Bureau of Labor Statistics, “ATUS Activity Lexicon.”

¹⁶ Bick, Blandin, and Deming, “The Rapid Adoption of Generative AI”; Marguerita Lane, Morgan Williams, and Stijn Broecke, “The Impact of AI on the Workplace: Main Findings from the OECD AI Surveys of Employers and Workers,” OECD Social, Employment and Migration Working Papers, no. 288 (Paris: OECD Publishing, 2023), <https://doi.org/10.1787/ea0a0fe1-en>.

¹⁷ Lane, Williams, and Broecke, “The Impact of AI on the Workplace.”

should publish such responses as reported time effects, not measured productivity gains. The comparison should include prompting, checking, and revising output, and allow that the activity might not have been undertaken without AI.

The activity linkage can nevertheless support richer economic research than a general-use question. Researchers could compare reported AI use across broad activities and occupational groups and examine associations with time devoted to work, job search, household production, or leisure. Within-job task comparisons would require the additional task-purpose information discussed in section II; on-the-job training is not separately identified by the ordinary work codes.¹⁸ ATUS interviews each respondent once about one diary day, so repeated annual collections can reveal population patterns, not individual before-and-after changes in time use following AI adoption.¹⁹ Stable measures over multiple years may support causal work only when combined with an appropriate research design and external variation.

The module should preserve links to available ATUS and Current Population Survey variables needed for this work, subject to existing disclosure rules. BLS should document which characteristics are available for analysis—including occupation, industry, employment status, class of worker, usual hours, education, age, sex, race and ethnicity, and household composition—and which subgroup comparisons are likely to be publishable given sample sizes. BLS should identify which characteristics are updated at the ATUS interview, which come from the earlier CPS, and how job changes or missing information affect linkage.

Distributional analysis is especially important. Aggregate associations can conceal differences between workers whose employers provide approved tools and training and workers who rely on personal accounts or face restrictions. They can also conceal differences between employees and the self-employed, workers with caregiving responsibilities, and workers in occupations with similar exposure but different opportunities to adopt. Context variables can improve interpretation when their universes and sample sizes support the comparison. This potential should not be treated as a promise of reliable estimates for every subgroup.

One distributional comparison is particularly important for labor market research: AI use by labor-force status, including the not-in-the-labor-force subgroup. Because ATUS respondents can be linked to CPS labor-force information, although the relevant CPS interview generally predates the ATUS interview, BLS should document which labor-force-status and job-desire variables are available in the ATUS public-use file or require a separate CPS linkage, their reference dates, and their applicable universes.²⁰ This linkage could support descriptive analysis of AI use in other income-generating activities among people classified as not in the labor force, subject to timing, classification, and sample-size limitations. If the AI module reaches non-employed respondents and captures AI use in other income-generating activities as recommended above, researchers could examine whether reported AI use differs across not-in-the-labor-force categories. BLS should document how the module variables can be linked to the available CPS information and should avoid implying that the resulting data identify a clean measure of informal work or a person's reason for not working. Without this documentation, the linkage may be technically available but practically difficult for researchers to use.

Recent descriptive evidence of changes in solo self-employment and nonemployer-type business formation in AI-exposed sectors suggests that AI measurement should not be limited to payroll

¹⁸ Bureau of Labor Statistics, “ATUS Activity Lexicon.”

¹⁹ Bureau of Labor Statistics, “ATUS User's Guide.”

²⁰ Bureau of Labor Statistics, “ATUS User's Guide.”

employment.²¹ The proposed ATUS measures could describe AI use across current work arrangements, but would not establish whether AI caused movement into independent work.

VI. Protect Comparability, Precision, and Public-Use Value

The notice anticipates 7,672 respondents per year and states that the data will support analysis across demographic and occupational groups. BLS should publish expected unweighted sample sizes and precision for the tabulations it intends to produce. The total is not a sample of employed AI users: employment, working on the diary day, AI use, and specific activities each further restrict relevant comparisons. Multiple episodes from a respondent do not provide additional independently sampled people. Detailed occupational or demographic estimates may require pooling both years or using broader categories. BLS should report achieved sample sizes and design-based uncertainty for released estimates, and assess the precision of proposed subgroup comparisons before committing to detailed tabulations. Comparisons involving other income-generating activities and narrow labor-force-status subgroups may be especially imprecise and may require pooling years or using broader categories.

BLS should identify a stable core before fielding: the operational definition, diary-day use indicator, episode linkage and attribution, broader-period frequency, and essential work/activity linkage. Those measures, including standardized example protocols, should remain unchanged during 2027 and 2028 unless evidence identifies a serious measurement failure. If a core revision becomes necessary, BLS should document it, assess whether a bridge or overlap study is feasible, and clearly identify any resulting break in the series. A stable wording alone does not guarantee stable interpretation as AI products change.

The public-use microdata should include linkable AI-use measures at the diary-episode level and person-level summaries. BLS should release the questionnaire, interviewer instructions, codebook, universe and skip patterns, editing and imputation rules, appropriate analysis weights, applicable replicate weights, and variance-estimation guidance for person- and episode-level estimates. Without these materials, researchers may produce estimates that are not comparable or that incorrectly treat episodes as independent observations. The documentation should state how to retain the person-level sampling structure when analyzing multiple episodes, how to handle weekday/weekend weighting and pooled years, and whether module-specific weighting adjustments are warranted.²² Reported nonuse, uncertainty, refusal, other missingness, and structural skips should remain distinguishable.

BLS should assess nonresponse to ATUS, failure to complete the AI module among ATUS respondents, and nonresponse to particular AI items separately. Available CPS characteristics can help compare observable characteristics of respondents and nonrespondents and evaluate possible weighting adjustments. Abraham, Maitland, and Bianchi found broadly similar aggregate time-use estimates under alternative weighting approaches in an early ATUS study, but that result does not establish that new AI estimates are unaffected by nonresponse.²³ Low response rates alone do not demonstrate bias, and weighting cannot automatically remove differences associated with unobserved

²¹ Liya Palagashvili, “Artificial Intelligence and the Rise of Independent Work: Early Evidence on Solo Business Formation and Self-Employment” (Mercatus Policy Research, Mercatus Center at George Mason University, July 2026), <https://www.mercatus.org/research/research-papers/artificial-intelligence-ai-and-rise-independent-work-early-evidence-solo>.

²² Bureau of Labor Statistics, “ATUS User’s Guide.”

²³ Katharine G. Abraham, Aaron Maitland, and Suzanne M. Bianchi, “Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter?,” *Public Opinion Quarterly* 70, no. 5 (2006): 676–703, <https://doi.org/10.1093/poq/nfl037>.

AI use. BLS should evaluate whether existing weights remain adequate for the module and publish the evidence and limitations.

BLS should also consider retaining a short core after 2028. Two years can establish a baseline but cannot show whether adoption plateaus, spreads to new activities, or changes time allocation over a longer period. A stable recurring module, even if fielded less frequently, would be more valuable than a one-time set of detailed questions that cannot be compared over time.

VII. Validate Question Performance and Burden

BLS should conduct and publish cognitive testing and operational burden testing across occupations, age groups, education levels, employment arrangements, and levels of digital familiarity. Testing should include respondents classified as unemployed or not in the labor force who report other income-generating activity, because this is a key universe for the proposed linkage. Testing should examine whether respondents interpret the definition consistently, recognize the intended AI uses in both stand-alone tools and integrated features, distinguish work from nonwork use, and correctly distinguish assistance with a diary activity from unrelated concurrent use. It should also examine false positives from ordinary software and false negatives from unfamiliar product names or workplace features. Testing should also assess the frequency item among diary-day nonusers and whether new probes affect the completeness or coding of the core diary.

The January 2027 field date makes prioritization important, but we cannot establish the feasibility of a particular testing schedule or clearance route from the notice alone. BLS should use cognitive interviewing to assess interpretation and realistic operational interviews to measure burden across complete response paths. Those response paths should include respondents who report no conventional job but do report other income-generating activity. If feasible, a randomized prefield test could examine the effects of alternative examples while holding the intended construct constant. If examples are delivered through optional interviewer help, BLS should distinguish assignment to a version from actual exposure and record which examples were read. Any production experiment should have a documented allocation, analysis, and comparability plan; it should not be treated as a costless change to an otherwise stable core. A first-year evaluation should report completion times, module and item nonresponse, multiple-episode reporting, and uncertain responses. Any second-year revision should follow the documented continuity rule in section VI.

Because workers may fear that reporting unauthorized use could reach an employer, BLS should use a brief, accurate confidentiality assurance consistent with its existing protections, explaining that individual responses are used for statistical purposes and are not disclosed to employers. Cognitive interviews should test whether this assurance improves reporting and whether job-specific policy categories and uncertainty options are understood consistently.

The notice estimates an average of two minutes per response, not a maximum for every respondent. Its arithmetic is consistent: 7,672 responses multiplied by two minutes is approximately 256 hours.²⁴ The empirical question is whether that average represents the full interview paths. BLS should time the definition, screeners, diary review, clarifications, repeated episode follow-ups, and any optional items in realistic interviews. Average burden should reflect the proportions following each path, with the distribution reported separately for nonusers, occasional users, and frequent users. Broader-period questions still impose burden on diary-day nonusers. BLS should also assess whether rising use or more AI-assisted episodes could increase average burden in 2028.

²⁴ Bureau of Labor Statistics, “ATUS Artificial Intelligence (AI) Questions,” 91 Fed. Reg. at 42776.

If burden exceeds the target, we would first remove the self-reported time-effect item, then additional technical or delegation classifications, and then optional employer-policy, access, or training questions. A supplementary work-task-purpose item should follow only if the core fits and within-job task identification remains a priority. BLS should protect the operational definition, diary-day use and episode attribution, broader-period frequency, and essential linkage to existing work information. Any new work-arrangement follow-up should be limited to verified gaps. If even this core does not fit, BLS should publish the tested burden and tradeoffs rather than imply that all requested measures can be delivered within the original estimate.

Conclusion

The proposed collection has clear practical utility through its linkage of AI use to a full-day activity record. By clearly distinguishing measured time allocation and self-reported AI use from causal productivity effects, BLS can create a stable, tested core module that supports economic research on patterns of AI use, work arrangements, and daily activities over time. The priority is a modest set of interpretable measures whose research claims match the information collected.