

Data Description Sheet

“Human + AI in Accounting: Early Evidence from the Field”

By Jung Ho Choi and Chloe Xie

This data description sheet is provided to fulfill the requirements of the Journal of Accounting Research data policy for submission of a manuscript.

Item 1. A description of which author(s) handled the data and conducted the analyses.

Jungho handled the survey, field, and experiment data and conducted the analyses with the following research assistants' help: Khanh Vu, Chnagyeong Kim, and Ben Zwarg. Chloe handled the field data with Ben Zwarg's help. Jungho, Chloe, and Khanh involved in the survey question development with the Stanford GSB Behavior Lab's support. Jungho and Khanh accessed the survey data via Qualtrics. Jungho, Chloe, Khanh, Changyong, and Ben all are able to access the field data. Chloe and Jungho reviewed all the analyses.

Item 2. A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

Our data consist of three parts: the survey, field, and experiment data. The survey data were collected through Prolific and Connect, two popular survey platforms in November 2024 and March 2025. The survey methods and instruments are discussed in the paper and online appendix. The field data were obtained from August 2024 to November 2024 and from March 2025 to April 2025. The experiment data were collected from June 2025 to July 2025.

Item 3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).

The field data were obtained with a condition that the research team can only disclose the outcomes at the aggregate-level without revealing any individual client information at the Partner Firm, which is confidential.

Item 4. A complete description of the steps necessary to download, obtain or collect as well as process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject

eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

As we discuss above, we use proprietary field data. The online appendix contains the full set of survey and experiment instruments.

Item 5. After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.

Our replication package contains a comprehensive log file including how to manipulate all the data for analyses.

Item 6. The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the manuscript. A brief description that enables other researchers to understand and run the code should be provided. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, and which commands were used in the analysis, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of disclosing the proprietary portion of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.

Our replication package contains all the code we used from the moment that we received the proprietary data from our partner company or collected the data from surveys and experiments.

Item 7. A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.

Our replication package contains a comprehensive log file including how to manipulate all the data for analyses.

Item 8. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

Chloe and Jungho will maintain the data and programs for at least six years.