

Data Description Sheet

“Beyond Automation: AI and the Human Value of Sell-Side Analysts”

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

Il Sun Yoo (Singapore Management University) handled the data and conducted the analyses reported in the manuscript.

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 can vouch for the stated source of the raw data.*

All data used in the study are from public sources. The sources of our raw data are reported in the manuscript. The following is a summary:

Data:

- Technology job posting data from Kalyani et al. (2025): <https://www.techdiffusion.net/>, accessed in May 2024.
- Analyst forecast data: I/B/E/S, accessed in May 2024.
- Additional analyst forecast data for the 2024 AskResearchGPT tests: I/B/E/S, accessed in November 2025.
- Financial data: Compustat Annual Data Files and CRSP Files, accessed in February 2024.
- Management guidance data: I/B/E/S, accessed in January 2024.
- Earnings conference call transcripts, obtained from Kimball Chapman in June 2025.
- Broker-hosted conference data, hand collected from the Bloomberg Terminal in June 2025.
- Filing date data from Loughran-McDonald (LM) 10X Summary data: <https://sraf.nd.edu/sec-edgar-data/>, accessed November 2025.
- Public float data from Ewens et al. (2024): https://github.com/michaelewens/public_float_regulation, accessed in June 2025

The analyses are done using the above data sources and SAS/STATA programs. These programs and datasets are available upon request and will be uploaded to JAR website if published.

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, and any restrictions imposed by the organization on the authors). In particular, the authors should indicate if an organization or data provider imposes restrictions on the publication of the results, has not given the authors full control of the relevant data, requires that the results must be reviewed or approved prior to public release of the paper or publication.*

We do not utilize proprietary data in this study.

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.

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 can be found in Sections 3, 4, and 5 of the manuscript.

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.

All data manipulations were conducted via computer programs, and the SAS and Stata codes used for these manipulations are included in the online supplements.

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.

All final dataset construction based on the raw data was performed using SAS. The final analyses were conducted in Stata.

For dataset construction,

- Timeliness_sample.sas: used for sample construction for timeliness tests and iXBRL tests (Tables 2 and 3)

- Analystfirmyear_sample.sas: used for sample construction for accuracy, boldness, and call participation tests (Tables 4, 5, 6, and 8)
- Coverage_sample.sas: used for sample construction for new coverage tests (Table 7)
- Askresearchgpt_timeliness_sample.sas and Askresearchgpt_sample.sas: used for sample construction for AskResearchGPT analyses (Table 9)

For analyses,

- First step analyses.do: used to conduct analyses for timeliness tests and iXBRL tests (Tables 2 and 3)
- Second step analyses.do: used to conduct analyses for accuracy, boldness, new coverage, and call participation tests (Tables 4, 5, 6, 7, and 8)
- Askresearchgpt analyses.do: used to conduct analyses for AskResearchGPT analyses
- Appendix B and Figure 2.do: used to conduct analyses in Appendix B and Figure 2

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.

We provide a comprehensive log file that shows the execution of the entire code in the replication package.

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.

The data and programs will be maintained by Il Sun Yoo for at least six years, consistent with National Science Foundation guidelines.