***Rutgers Access:***

Textual data on companies’ filings: [Seekinf](https://www.seekedgar.com/UserGuide.pdf)

Most used data vendor: [WRDS](https://wrds-www.wharton.upenn.edu/pages/support/getting-started/getting-started-guide/), including the following most used data in accounting research:

* Accounting, Banking, Economics, ESG, Finance, Healthcare, Insurance, Marketing, and Statistics
* Audit data: [Audit Analytics](https://wrds-www.wharton.upenn.edu/pages/get-data/audit-analytics/)
* Financial report data: [Compustat](https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/)
* Security price, return, and volume data: [CRSP](https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/)
* Analyst forecast data: [I/B/E/S](https://wrds-www.wharton.upenn.edu/pages/get-data/ibes-thomson-reuters/)
* Real-time news analytics: [RavenPack](https://wrds-www.wharton.upenn.edu/pages/get-data/ravenpack-news-analytics-40/)
* And other [subscribed](https://wrds-www.wharton.upenn.edu/pages/get-data/) data.

***Personal/Coauthor Data:***

LinkedIn Profiles searched by U.S. audit partners’ name and office ([AuditorSearch](https://pcaobus.org/resources/auditorsearch) - Public access)

Biographical data of corporate executives and board members: [BoardEx](https://wrds-www.wharton.upenn.edu/pages/get-data/boardex/)

Processed data on [fraud](https://www.sec.gov/divisions/enforce/friactions), financial misconducts, earnings management, financial reporting quality, audit quality

Environmental, Social, and Governance (ESG) ratings: [Refinitiv](https://wrds-www.wharton.upenn.edu/pages/get-data/thomson-reuters/refinitiv-esg/), [MSCI-KLD](https://wrds-www.wharton.upenn.edu/pages/get-data/msci-formerly-kld-and-gmi/), Russell 3000 [Bloomberg ESG Scores](https://www.bloomberg.com/professional/product/esg-data/)

Corporate misconducts of all types till 2022: [Violation Tracker](https://violationtracker.goodjobsfirst.org/)

***Public Data that could be used to test our LLM model:***

Transactional dataset simulator: [Paysim](https://github.com/EdgarLopezPhD/PaySim)

Synthetic dataset from [Ernst Young Academic Resource Center (EYARC)](https://eyus.sharepoint.com/sites/EYARC)

* 37,869 observations with 15 variables and 18 [General Ledger](https://eyus.sharepoint.com/sites/EYARC/EYARC/Forms/AllItems.aspx?csf=1&web=1&e=1hQdtT&cid=ae255f4c%2D1830%2D4766%2D8b86%2D2e7dc497e406&FolderCTID=0x012000B2E36F818B78F64AAB7991B5401A29D3&id=%2Fsites%2FEYARC%2FEYARC%2FAnalytics%20mindset%2FCase%20studies%2FPeach%20State%20University%20Hotel%20%28EY%20Helix%20General%20Ledger%20Analyzer%20for%20Students%29%2FCase%202%20and%203%5FEY%20Helix%20General%20Ledger%20Analyzer%2C%20user%20guide%20and%20data%20set&viewid=644b16f9%2De37e%2D44c3%2Da8e9%2D92563239a999) classes