Standardization, AI, and the processing costs of sustainability disclosures in the field

Victor Wagner*
LMU Munich
victor.wagner@lmu.de

October 2025

[preliminary, please do not circulate]

Abstract This paper studies how standardization and Artificial Intelligence (AI) affect the processing costs of sustainability disclosures. Using field data from a large data collection project, I document how processing costs vary across disclosure formats and with experience. I complement this evidence with a field experiment that randomizes access to more versus less standardized reports and to a generative AI assistant. Processing costs, measured by time and accuracy in assessing firms' performance, are significantly reduced by standardization but not by access to the AI. AI also reduces the beneficial effects of experience and standardization.

Keywords: ESG disclosures, regulation, artificial intelligence, field study, field experiment

---

* This paper is part of my dissertation at LMU Munich. For helpful comments and suggestions, I would like to thank my dissertation chair, Thorsten Sellhorn, as well as Max Müller, Matthias Breuer, Patricia Breuer, Peter Fiechter, Joachim Gassen, Ann-Kristin Großkopf, Christian Leuz, Lisa Lüttke (discussant), Inga Meringdal, Gaizka Ormazabal, Lazo Papadopoulos, Ivo Schedlinsky, Katharina Schmidt (discussant), Harm Schütt, and Caren Sureth-Sloane, as well as seminar and conference participants at Amsterdam Business School, Chicago Booth, Rotterdam, WHU - Otto Beisheim School of Management, the 1st Accounting Research Camp on Transparency in Corporations and Markets, and the TRR 266 PhD brown bag. Julian Flaschke and Sebastian Peschel provided excellent research assistance. The paper's AEA RCT ID is AEARCTR-0016263. Approval for monitoring the data collection has been granted by the ethics commission of LMU Munich's faculty of medicine, No. 23-0435. Approval for conducting the experiment has been granted by the ethics commission of LMU Munich School of Management, No. ETH-SOM-007. I acknowledge financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG): project ID 403041268 – TRR 266.

1. Introduction

The rate of firms disclosing sustainability information has massively increased in recent years. This is because society is demanding greater accountability and transparency amid growing environmental and social crises. The rapid increase, however, has left many concerned about information overload, i.e., high costs associated with processing these disclosures due to low comparability and high discretion (e.g., Blackrock 2020; Client Earth 2021; WWF 2020). Responding to these concerns, the European Union (EU) has developed the Corporate Sustainability Reporting Directive (Directive EU 2022/2464, CSRD), a far-reaching disclosure mandate standardizing reports to "make sustainability information easily accessible for users" (recital 25). Processing these reports, which involves sifting through lengthy documents and identifying relevant information, is itself undergoing a fundamental transformation. This is driven by the advent of generative Artificial Intelligence (AI; e.g., Acemoglu, 2025; Autor, 2024), which potentially renders the benefits of standardization on processing costs irrelevant. In this paper, I document the extent and drivers of these costs, and then examine the effects of standardization through the CSRD and access to AI.

I focus on sustainability reporting not only because of its increasing relevance as a form of 'targeted transparency' employed to address societal challenges (Weil, Graham, and Fung 2013), but also because it provides a suitable setting to study the effects of standardization. Although firms have been publishing sustainability reports since at least the 2000s, these disclosures have frequently been labeled as greenwashing (Delmas and Burbano 2011) and criticized as mere impression management (Cho, Roberts, and Patten 2010). Moreover, absent mandatory regulation, firms had considerable discretion about what and how to present information. Studies find that firms bias their reports by employing visual information (e.g., Cho, Michelon, and Patten 2012) or rhetorical arguments (e.g., Laine 2009), ultimately reducing their usefulness (Milne and Gray 2013). Thus, standardization might be especially powerful in this setting.

Research on standardization in the financial reporting domain, such as through the introduction of International Financial Reporting Standards (IFRS), finds mixed effects on reporting outcomes such as transparency and comparability (e.g., Brüggemann, Hitz, and Sellhorn 2013; De George, Li, and Shivakamur 2016). While many earlier studies have documented positive benefits from adoption, later studies argue that concurrent changes in the institutional environment and the relevance of reporting incentives also play a significant role. As IFRS are principles-based standards, these factors play an important role in this setting (e.g., Ball, Robin, and Wu 2003; Carmona and Trombetta 2008). Against this backdrop, the sustainability reporting domain provides an interesting contrast.

The CSRD was explicitly motivated by the EU's dissatisfaction with the earlier Non-Financial Reporting Directive (NFRD, Directive EU 2014/95; e.g., Fiechter, Hitz, and Lehmann 2022), which did not deliver the hoped-for benefits with regard to comparability (e.g., EC 2019). Thus, in 2020, the EU decided to introduce a disclosure mandate and entrust its European Financial Reporting Advisory Group (EFRAG) to develop a set of reporting standards. The twelve European Sustainability Reporting Standards (ESRS) detail what environmental, social, and governance disclosures affected companies must provide. Moreover, the standards are much more prescriptive regarding the requirement of disclosures in a specific format, possibly decreasing the influence of incentives. However, given its recent introduction, the standards lack a proper enforcement and assurance infrastructure.

I leverage insights from a data collection project on the Sustainability Reporting Navigator (SRN) to better understand the determinants of processing costs and to inform my experimental design.[1] To inform the regulatory debate on the "CSRD-preparedness" of European firms, the SRN hosts a data entry platform where students – as part of their final assignment at one of the SRN's

[1] The SRN is an open-science platform that aims to increase the transparency and accessibility of sustainability reporting.

affiliated universities – collect firms' data points from the two most widely used standards, ESRS E1 on climate change (e.g., a firm's gross Scope 1 GHG emissions) and ESRS S1 on their own workforce (e.g., a firm's share of minority workers in its workforce). Starting from March 2023, 143 undergraduate and graduate students from four public German universities collected over 750,000 data points on 851 unique firms' reporting practices (4,453 firm-years). The datapoints are directly derived from the standards themselves, i.e., represent those elements an analyst would have to collect to create sustainability benchmarks and analyses.

Monitoring students' data collection allows me to construct a granular measure of information acquisition costs of individual data points, as well as on an aggregated level of firms' emissions and own workforce reporting, respectively.[2] Information processing is commonly divided into three steps that are then linked to specific costs: awareness, acquisition, and integration (Blankespoor 2019; Blankespoor, deHaan, and Marinovic 2020). I proxy for acquisition costs by the time it takes a student to browse the firm's report, identify the data point, and enter the corresponding value in the data collection platform. This captures the cognitive costs of reading and extracting specific information from a firm's report (Bloomfield 2002; Maines and McDaniel 2000).

I find that collecting the data points for one firm-year takes about 30 to 40 minutes on average. Data points that are uncommon, as well as qualitative data points (i.e., descriptions of methods used or explanations), take longer to collect. Experience gained through the process, i.e., completing data collection for additional firms, as well as experience gained outside the process, e.g., through previous internships or personal investing activities, significantly reduces the acquisition time of the disclosures. Data points highly dispersed across the report, as well as those located in more complex texts, are

[2] Ethics board approval for the monitoring of the data collection activities has been granted by the IRB of LMU Munich's Faculty of Medicine under No. 23-0435.

harder to process, which complements findings from prior literature (e.g., Li 2008). Presenting information in tabular format, however, aids processing.

Having documented that some of the dimensions, which are explicitly addressed by the CSRD, are associated with lower processing costs, I now turn towards my field experiment. The field experiment allows me to causally study how the 'composite' effect of the CSRD, as well as access to the AI, impacts processing.[3,4] My field experiment is integrated into a course on GHG emissions reporting under the CSRD, hosted on the SRN Academy, catering towards sustainability practitioners and accounting students. To successfully complete the course and gain a certificate, participants have to extract information and analyze the sustainability reports. Compared to the mere data acquisition task from the first part of the paper, I add an evaluation component to better resemble the task of an ESG analyst who actually evaluates and compares the firms' performance. I randomly allocate participants to groups receiving reports from before or after the CSRD, as well as having access to a custom AI tool. To study how experience influences processing and to reduce the risk of participants simply re-doing the task with new credentials, I prepare three industries.

My primary measure is an aggregate score based on the time it takes participants to finish the task and the accuracy of their answers to three questions. Specifically, participants evaluate which of an industry's three firms has the lowest absolute emissions levels, the lowest emissions intensity, and the most ambitious climate goal. The questions are based on a careful review of ESG analyst reports and are designed to capture a comprehensive view of the firms' GHG emissions performance. I select industries for which climate is a material ESG matter, thereby ensuring that the firms disclosed on their

[3] The paper's AEA RCT ID is AEARCTR-0016263 with ethics board approval being granted by the ethics commission of LMU Munich School of Management, No. ETH-SOM-007.

[4] I use the term 'field experiment' to indicate that the experiment departs from a standard lab setting because participants engage in a real-world task (analyzing GHG emission performance) with real-world data (actual firms' reports). This follows Floyd and List (2016) who emphasize that field experiments are defined not only by subject pools but also by the field context of the task, stakes, and information environment.

emissions performance already before the CSRD.[5] Finally, and similar to the first part of the paper, I elicit their prior experience in a pre-experimental questionnaire.

For my first treatment arm, receiving pre-CSRD versus post-CSRD reports, I find no significant differences in the time participants take to conduct their analysis and answer the questions between standardized and unstandardized reports. Participants who analyze standardized CSRD reports, however, are 15% more accurate in answering the three questions (p = 0.008). Given that the firms' environmental sections published after the CSRD are significantly longer than their pre-CSRD counterparts,[6] the results show that relying on a standardized way of presenting information makes up for increases in reporting volume, bolstering accessibility and understandability. Moreover, I find that the benefits of standardization are concentrated in the group with above-median prior experience, suggesting that the benefits of standardization are highest for users with prior experience.

But to what extent are the benefits of standardization still relevant? Since the first discussions about mandating sustainability reporting standards in the EU took place in 2020, technological progress has accelerated. Given the widespread use of (generative) artificial intelligence, one question is whether relying on AI can substitute for increased standardization. Many expected AI to significantly alter the nature of jobs and replace humans in manual 'knowledge tasks' (e.g., O'Donoghue and Roberts 2023). Given its potential to quickly sift through vast amounts of unstructured data, AI can reduce the frictions that led humans to experience processing costs. Recent large-scale randomized control trials (RCTs), for example, document improved efficiency for call-center agents from 10 to 15% (Brynjolfsson, Li, and Reynold 2025) and for software developers of up to 26% (Cui et al. 2025).

More and more commentators, however, cast doubt on the qualities of large language models (e.g., The Economist 2025). As most enterprise generative AI pilots fail to deliver financial returns,

[5] The three industries are passenger aviation, cement production, and steel manufacturing.
[6] For the nine reports (three firms in three industries) of the experimental task, I find an average (median) increase of page count by a factor of 3.69 (3.70) with a range of 1.67 to 10. In a related analysis for the first 100 firms adopting the CSRD, we show that the average sustainability report increases by a factor of 1.47 (Wagner, Müller, and Hagemeier 2025).

enterprise users continue to “prefer ChatGPT for simple tasks, but abandon it for mission-critical work” (Challapally et al. 2025, p. 10). This manifests in a decreasing AI adoption rate among U.S. firms (U.S. Census Bureau 2025), decreasing levels of trust over the last two years (Gillespie et al. 2025), and employees’ dissatisfaction when receiving “workslop,” i.e., documents or presentations created with AI involving low effort but which are passable looking (Niederhoffer et al. 2025). Moreover, it remains unclear which of the more knowledge-intensive tasks AI can perform sufficiently well. In their RCT with strategy consultants using generative AI, Dell’Acqua et al. (2023) argue for a “jagged technological frontier” as they find that performance improvements manifest only for a set of tasks which are “surprising and not immediately obvious to individuals or even to producers of LLMs themselves” (p. 4). It is an open question, therefore, how generative AI affects the processing of complex corporate disclosures.

To answer this question, I present results from the second treatment arm of my field experiment, which involves access to a custom AI tool that I specifically designed to aid the extraction and interpretation of sustainability information. The AI tool works similarly to ChatGPT, but includes a retrieval augmented generation (RAG) architecture to alleviate concerns of “hallucination” (see, e.g, Forster et al. 2025). This means that the chatbot injects the pages of the report that are most similar to the users’ query so that the large language model (LLM) can rely on the firms’ actual disclosures, not just its training data. The architecture thus combines the generative-interactive capabilities of LLMs with domain-specific knowledge, increasing accuracy and speed compared to “off-the-shelf” models.

I do not find that access to the AI tool reduces processing costs. To the contrary, participants using AI actually show lower aggregate scores. This holds true irrespective of whether the participants see standardized or unstandardized reports. The negative effect affects both dimensions, time spent and accuracy of the answers. This, however, is not necessarily due to the low accuracy of the AI chatbot. My analysis of the prompts and responses reveals that the bot responds, when prompted to give the

Scope 1 emissions (i.e., the subject of the first question and the only question which can be answered definitely by an AI), in 80% of the cases with the exact value.[7] These findings align with research highlighting that users have issues trusting AI output, especially in high-stakes settings (Dell'Acqua et al. 2023). In addition, using AI has detrimental effects on learning. I show that the internal experience gained from absolving more rounds only occurs in the group not using AI. Also, the benefits of standardization are concentrated in the groups without access to the AI chatbot. Using AI drives engagement, however. I find that AI-equipped participants don't show survey fatigue or signs of attrition over the three rounds compared to the group manually browsing the reports.

My paper offers several contributions. First, I contribute to research on the standardization of disclosures through mandatory regulation. Whereas studies examining the standardization of financial reporting through IFRS (for reviews, see, e.g., Brüggemann, Hitz, and Sellhorn 2013; De George, Li, and Shivakumar 2016) generally do find low or no effects of increased comparability or reporting quality, my study shows direct benefits of standardization for sustainability reporting. This is plausible since sustainability reporting started from lower levels of standardization and exhibits unique characteristics, e.g., features a diverse range of topics with different metrics, little to no monetization, and speaks to a broader audience than financial reporting (Christensen, Hail, and Leuz 2021).

Second, I add to a growing literature that examines the use of (generative) AI in professional contexts. A recent meta-analysis (Vaccaro, Almaatouq, and Malone 2024) reveals that human-AI combinations, on average, do not perform better than either group alone, calling for more research on how AI might lead to 'skill atrophy' (Macnamara et al. 2024). I add to this emerging literature by documenting that in a high-stakes setting, participants do not necessarily have an advantage with AI. In addition, I show that using AI is detrimental to learning.

[7] In those cases, where the first response did not contain the correct value, because the bot responded with the intensity or could not find relevant context, when queried a second time, the participants received the exact value in 60% of the cases.

My findings also have practical relevance to standard-setters in light of the currently debated changes to the CSRD, the "Omnibus initiative." Here, the European Commission proposes to significantly cut the number of data points and make more disclosures voluntary. My findings show that standardizing sustainability reporting through the CSRD has already helped to reduce processing costs. Especially since the CSRD standardized reporting along key dimensions that, in the voluntary setting, have increased processing costs, it is questionable if allowing for more reporting discretion is beneficial. This is in line with calls by many investors, NGOs, and other groups representing users of this reporting who argue that, as the disclosures are increasingly used for investment decisions (e.g., Eurosif et al. 2025), they require standardized information. From a welfare perspective, however, these benefits have to be weighed against the costs of the increased disclosures through the CSRD.

2. Institutional background and conceptual underpinnings

2.1 The Corporate Sustainability Reporting Directive

The Corporate Sustainability Reporting Directive (CSRD, Directive EU 2022/2464) is the largest-ever introduction of mandatory sustainability reporting. As the successor to the Non-Financial Reporting Directive (NFRD, Directive 2014/95, see, e.g., Dinh, Husmann, and Melloni 2023; Fiechter, Hitz, and Lehmann 2022), the CSRD went into force in January 2023, initially affecting over 50,000 large EU as well as international firms.[8] The scope of the CSRD is based on several 'waves,' with the first firms having to disclose according to the new standards for their fiscal years 2024. Over undue bureaucratic burdens, in February 2025, the EU proposed several amendments to the CSRD, drastically reducing the scope of firms affected as well as the extent to which they have to report, by cutting the number of data

[8] This includes EU firms that fulfil two of the three criteria: net turnover greater than €50 million, assets greater than €25 million, or number of employees greater than 250; listed small and medium enterprises; or firms with a non-EU parent where the combined group turnover is greater than €150 million.

points in the European Sustainability Reporting Standards (ESRS) which lie at the heart of the CSRD and detail the disclosure requirements firms have to adhere to.

The standards are comprised of two overarching standards on the general requirements (ESRS 1) and general disclosures (ESRS 2), as well as ten topical standards on the three dimensions of ESG. They contain detailed requirements on, e.g., the firm's plan to align with limiting global warming to "well below 2°C" (ESRS E1-1), it's current (E1-6) greenhouse gas emissions and planned reductions (E1-4) or the characteristics of its own employees (ESRS S1-6), their diversity (S1-9), or adequate wages (S1-10). To determine which topics a firm has to report on, ESRS 2 requires firms to conduct a 'materiality assessment,' which increases the relevance of many topics. Importantly, the ESRS follow a 'double' materiality principle, meaning firms have to consider how their activities impact nature and society in addition to the (financial) risks and opportunities that the firm faces from ESG issues.

2.2 The effects of standardization on reporting content and information processing

The CSRD is triggering a similar reshaping of corporate reporting practices as the introduction of the IFRS, twenty years ago. Starting with fiscal year 2005, Regulation 1606/2002 ("IAS Regulation") required EU firms to prepare their consolidated financial statements following International Financial Reporting Standards (IFRS). With the goal of "harmonising the financial information presented by companies [...] to ensure a high degree of transparency and comparability" (Art. 1), the European Commission strived to increase the efficiency of European capital markets. Regarding these economic consequences, studies reviewing the effects of the IFRS introduction consistently find positive capital market and macroeconomic effects (see, for reviews, e.g., Brüggemann, Hitz, and Sellhorn 2013 or De George, Li, and Shivakumar 2016).

Findings on the economic consequences are surprising insofar as findings on comparability and reporting quality are rather mixed (e.g., Carmona and Trombetta 2008). Given their nature as principles-based standards, IFRS leave a lot of discretion to the adopting firms, making it possible to merely "adopt a label" (Daske et al. 2013). The major driver for any reporting changes, then, is not necessarily the standards themselves, but underlying economic and political factors that together shape managers' incentives (Ball, Robin, and Wu 2003). Thus, IFRS effects are concentrated among firms that voluntarily adopt IFRS and not those that were forced to do so (Christensen et al. 2015).

There are reasons to believe that the introduction of the European Sustainability Reporting Standards (ESRS) has different reporting effects than the introduction of IFRS. First, financial reporting has existed as a cultural practice for thousands of years, which led to the establishment of numerous, albeit different, national financial reporting systems (Soll 2014). While sustainability reporting has existed since at least the 1990s, the reporting itself has faced never-ending criticism from academics and practitioners alike over concerns that it is a self-serving and mostly rhetorical practice (e.g., Cho, Michelon, and Patten 2012; Milne and Gray 2013). This is despite the fact that the Global Reporting Initiative (GRI) has become the "de facto law" (Larrinaga and Bebbington 2022, p. 133), having paved the way for the institutionalization of sustainability reporting.

Second, and this can be seen as a response to these criticisms, the ESRS are a wide-reaching example of standardization. Ambitiously, the standards prescribe disclosures on a wide array of topics – many of which, e.g., biodiversity or indigenous communities, have not been very prominent in mainstream reporting, yet – and aim to harmonize multiple aspects of the presentation format (i.e., the terminology, structure, and visual representation). This means that, before the introduction of the CSRD, a firm could choose not to or only partially (e.g., using their self-defined metrics) on key sustainability dimensions. Now, assuming that the topic is considered as 'material,' a firm is required to follow a set of specific metrics (e.g., absolute Scope 1 emissions in metric tonnes, percentage of women in the workforce) in a specific format (e.g., a tabular format for all emissions categories), in a specific location (e.g., in a dedicated section in the management report), and using a specific methodology and terminology (e.g., using one formula to calculate the gender pay gap or the same denominator for intensity-based metrics such as energy intensity).

While it is yet too early for a comprehensive assessment of how the CSRD has changed reporting practices, early evidence shows that reports prepared under the CSRD are more standardized, containing more tables and numbers, are more negative in tone and use more complex language, moving away from "glossy brochures" to more resemble financial reporting (Wagner, Müller, and Hagemeier 2025). I thus expect that the introduction of the CSRD will lead to a reduction in processing costs for users. Daske, Seregni, and Uckert (2023), who examine the effects of accounting terminology standardization, argue that processing costs decrease when users can rely on a common terminology, i.e., where one concept is represented by only one term. The ESRS, however, not only standardize terminology but also fundamentally how information is structured, leading to potentially high benefits in how users can process these disclosures.

I thus expect that more standardized reporting decreases information processing costs. Specifically, acquisition costs, i.e., the cognitive costs of reading and extracting specific information from a firm's report, should decrease (Bloomfield 2002; Maines and McDaniel 2000). This is in line with prior research that shows that data workers who collect and populate financial databases face lower processing costs for firms with more standardized terminology (Daske, Seregni, and Uckert 2023). Consistent with this, Drake, Roulstone, and Thornock (2014) find that sophisticated investors are more likely to acquire information from EDGAR for firms with complex, hard-to-value reporting environments. This implies that a lack of standardization raises acquisition costs, while harmonized and comparable reporting standards directly mitigate these frictions.

For integration costs, i.e., the costs of evaluating and assessing these disclosures (Maines and McDaniel 2000), extant literature similarly documents the benefits of standardization. For example, financial analysts face lower processing costs when analyzing financial statements of firms with more standardized financial statement line items (De Franco, Kothari, and Verdi 2011) or more similar accounting methods (Bradshaw, Miller, and Serafeim 2009). Whereas in traditional financial reporting the notion of integration typically denotes the refinement of information for incorporation into investment models or decision processes, its meaning in the sustainability context is more complex. This complexity arises from the broader and more heterogeneous audience of sustainability reporting, with users exhibiting diverse information needs (Christensen, Hail, and Leuz 2021). Even within the investor community, there is no unified approach to sustainable investing (Starks 2024). Some scholars even argue that privileging an investor-oriented perspective risks obscuring what sustainability reporting ought to address in the first place (Abhayawansa 2022), underscoring the ongoing debates on the interpretive authority of sustainability reporting.

The changes of the CSRD to the way information is presented might actually impair information processing, however. While the increased use of quantitative disclosures can, in principle, reduce frictions for users by clarifying economic implications (Liberti and Petersen 2018), the simultaneous introduction of more detailed and prescriptive standards risks increasing the complexity of reports. Prior evidence shows that when reporting regimes expand disclosure requirements, managers often respond with more complex disclosures, which can overwhelm users rather than facilitate understanding (Chychyla, Leone, and Minutti-Meza 2019; Dyer, Lang, and Stice-Lawrence 2017). This concern aligns with a broader stream of research documenting that complex and less readable language decreases the accessibility of financial reports: Li (2008) shows that lower readability is associated with less informative stock prices, Miller (2010) finds that obfuscation impedes the incorporation of information into market prices, and Guay, Samuels, and Taylor (2016) provide evidence that linguistic complexity distorts investors' perceptions of firm fundamentals.

Taken together, prior literature suggests that while standardization under the CSRD is intended to reduce information processing costs, the accompanying rise in textual and structural complexity could counteract these benefits.

2.3 The effects of information technology on processing

One of the more extensively studied information technology advancements is the introduction of the eXtensible Business Reporting Language (XBRL), which has promised to reduce information processing costs since the early 2000s (Wagenhofer 2003). XBRL is a markup language that firms use to tag their financial statements, i.e., to "identify each financial statement element and link it to descriptive information" (Blankespoor, Miller, and White 2014, p. 1473). The tagging is designed to facilitate the processing of financial information due to two key factors. First, if firms use the same tags for the same financial statement items, comparisons across firms are much easier. Second, XBRL makes the formerly static financial statements interactive: due to its nested structure, the XBRL taxonomy relates different tags to each other and thus adds information on the relation between tags.

Hodge, Kennedy, and Maines (2004) conduct an experiment on how users with and without access to an XBRL-powered search engine process information that is recognized in the income statement versus disclosed in a footnote. They show that users with access to XBRL are more likely to acquire footnote information and subsequently integrate it into their investment decisions. The authors also find that nearly half of their participants did not use the technology at all, cautioning about the actual help to unsophisticated users. Studies utilizing archival methods find mixed evidence. While some studies, for example, find higher stock synchronicity (Dong et al. 2016) or lower return volatility (Kim, Lim, and No 2012) after XBRL, others find evidence for greater information asymmetry (Blankespoor, Miller, and White 2014), making a definite assessment of the technology's benefits difficult.

Artificial Intelligence (AI), similarly, is a technology that can ease the processing of information. One form of AI that has reached widespread adoption across professions are chat chatbots based on large language models (LLMs) utilizing a Generative Pretrained Transformer (GPT) architecture. Their interactive chatbot character and the fact that the models are based on vast amounts of training data make them easy to use and helped the most prominent bot, ChatGPT from OpenAI, to reach one million users in only five days (Statista 2023). Their seemingly 'superhuman' understanding (Levy 2024) created a plethora of research examining the capabilities of generative AI and LLMs in accounting and finance. In their scoping review, Dong et al. (2024) identify 38 papers that outline potential applications in the accounting field. This body of research, for example, demonstrates that LLMs can be used standalone to measure business complexity (Bernard et al., 2023), proxy business risk (Kim, Muhn, and Nikolaev, 2023), or to audit sustainability reports (Föhr et al., 2024). These studies showcase that due to their advanced language understanding, LLMs are well-suited for information processing tasks, or as Bernard et al. (2023) put it, "that the LLM proxies for a sophisticated user experienced at reading financial filings." (p. 3).

For information acquisition specifically, LLMs can be employed in Retrieval Augmented Generation (RAG) pipelines. This architecture combines the strengths of pre-trained language models, which are trained on vast amounts of text, with 'knowledge' stored in external databases. In their pioneering paper, Lewis et al. (2020) demonstrate that their Wikipedia-enriched RAG architecture answered Jeopardy questions more factually in 42.7% of cases than the benchmark language model without specific access to Wikipedia. This is because with RAG, the model, before answering the question, retrieves relevant knowledge directly from a specified source, e.g., Wikipedia, and thus does not solely have to rely on its full 'memory.' In the sustainability domain, this architecture has been employed to extract information from sustainability reports at scale (e.g., Bronzini et al. 2024; Forster et al. 2025). The increased accuracy from the retrieval component, combined with the features of the LLM, suggests that using generative AI reduces acquisition costs.

With regard to integration costs, research from the human-computer interaction literature has examined how generative AI-equipped participants perform in various laboratory tasks, such as essay writing (e.g., Doshi and Hauser 2024; Kosmyna et al. 2025) or math and physics questions (e.g., Chang et al. 2025; Riedl and Weidmann 2025). Similarly, field-based studies have investigated how the performance of call center agents (e.g., Brynjolfsson, Li, and Raymond, 2025), strategy consultants (e.g., Dell'Acqua et al., 2023), or software developers (Cui et al., 2025) is affected when having access to generative AI tools, generally finding significant positive benefits. While access to generative AI thus generally enhances productivity for the workers, Dell'Acqua et al. (2023), in their experiment with strategy consultants, find evidence of a "jagged technological frontier." This means that, while AI is suitable for a subset of tasks, increasing output quality and worker productivity, for other "surprising and not immediately obvious [tasks]" (p. 4), there were muted or no benefits.

Despite these advances, others issue more cautionary notes about using LLMs, because of technical limitations in numerical reasoning and the inherent look-ahead bias in many research applications (Levy 2024), as well as the often inaccessible datasets the models are trained on, which can lead to biases (Papadopoulos 2025) and misinformation (de Villiers, Dimes, and Molinari 2024). Trust in generative AI, therefore, has eroded lately. A recent study on generative AI usage in big corporations claims that 95% of generative AI pilots are failing to deliver financial returns since most tools lack contextual understanding and fail to learn from previous exchanges, making integration into actual workflows difficult (Challapally et al. 2025). Thus, for "mission-critical work," users reported not trusting AI output and resorted to manual approaches. This aligns with a meta-analysis of human-AI cooperations that documents significant variation of effect sizes and that more than half of the studies fail to find that human-AI outperforms either group alone (Vaccaro et al. 2024).

In sum, while AI has the potential to transform knowledge work and already shows productivity improvements in many more repetitive and context-unaware tasks, benefits in more complex and/or high-stakes settings appear muted.

3. Descriptive evidence on information acquisition costs

3.1 Field setting, data sources, and measures

To examine information processing before the Corporate Sustainability Reporting Directive (CSRD), I leverage a multi-year data collection project, which we administered as part of the Sustainability Reporting Navigator (SRN). As part of their final Bachelor's or Master's theses, students were assigned a sample of firms for which they had first to find and upload the annual and/or sustainability report and then collect relevant information. After having received an introduction to the background and specific task, students collected data points, i.e., granular information items (e.g., Gross Scope 1 emissions in the fiscal year in metric tonnes of CO_2 equivalents or the number of female board members) associated with ESRS E1 climate change and ESRS S1 own workforce metrics. Following extensive verification procedures, we subsequently released this data (e.g., Sellhorn et al. 2023) to inform the public debate on the development of the ESRS with information about firms' current reporting levels.

Over a time span of nearly two years, from March 2023 to January 2025, 143 undergraduate and graduate students from four public German universities collected over 750,000 individual data points on 851 unique firms' emissions and workforce reporting practices for their fiscal years 2020 to 2023. This amounts to 2,698 hours, or 33 workdays of data entry. Panel A, Table 1 provides information on (a sample of) 73 students' characteristics as well as their responses to a pre-collection questionnaire. Specifically, I collected data on their investing experience and familiarity with reading and analyzing primary corporate reports, their environmental attentiveness, and their educational background. (For an overview of the questions, see Appendix A.) This allows me to investigate how students' prior experience affects their processing. I compute *EXPERIENCE*, as the first principal component of the answers to all questions (Allee, Do, and Raymundo 2022).

I derive my first set of measures from my close control of the data entry platform. I measure information acquisition based on the timestamps of when a student collected a data point. This allows me to construct my main measure, $TIME_{sit}$ as the total time it takes student *s* to collect all data points for firm *i* in year *t.* I remove data points with excessively long collection times of over 120 minutes that might have arisen, e.g., due to the student taking a break. The number of actual existing data points is $EXIST_{sit}$. To proxy for a student's experience in reading and analyzing reports, I construct $SESSION_i$ as student *i*'s incrementing counter for each successfully completed firm-year. Panel B, Table 1 provides descriptive evidence on the measures related to the overall data entry process. The main measure, the raw time spent in minutes, *TIME*, can be split into the two topics, "emissions" and "own workforce". Whereas collecting all the quantitative emissions data points for ESRS E1 takes 32.7 minutes, collecting all the quantitative own workforce data points for ESRS S1 takes 48.6 minutes, on average (untabulated).[9] The unit of observation is the student-firm-year session, of which there are 4,453.

My second set of measures is based on the annual and sustainability reports, which the students used as a source for the data points. Since students also had to record the page number of where they found a data point, I first check if the pages already contain a reference to 'ESRS' or 'CSRD' (*ESRS*).[10] Relying on the actual pages where students found data points helps to isolate the effects where the reporting is already (partially) aligned with the CSRD compared to more general references in the report that, e.g., the firm anticipates reporting under ESRS for the next years. I then construct the

[9] This total time difference is mainly driven by the number of data points required by each standard. Students had to collect 124 (158) data points for ESRS E1 (S1), which leads to an average collection time per data point of 26.4 (30.7) seconds.
[10] Whereas the draft ESRS were released in November 2022 and adopted in July 2023, the first interim draft of ESRS E1 was already released in March 2022. This can be seen in the fact that for reports analyzed in 2020, 2021, 2022, and 2023, 0%, 1%, 6%, and 17%, respectively of the relevant pages in the document contained a reference to 'ESRS' or the 'CSRD'.

DISPERSION of existing data points at the document-level based on the Herfindahl-index of the page numbers. I can then examine how the specific textual characteristics of the page where the data point is located affect processing. For these pages, I compute standard textual variables (Lang and Stice-Lawrence 2015), i.e., *LENGTH* as the average number of characters per page, *DIFFICULT* as the average Fog-Index per page (Li 2008), *TONE* as the net tone, i.e., share of positive minus negative words to total words, and *COMPLEX*, the share of complex words to total words, and *UNCERTAIN,* the share of uncertain words to total words (Loughran and McDonald 2011) as well as the number of *TABLES* and the share of pages that is covered by *IMAGES*. Panel C, Table 1 provides descriptives and Table 2 shows Pearson correlation coefficients for the variables.

Finally, I rely on the answers to a set of questions that students had to answer before they finished the data collection for a firm. Based on the fundamental qualitative characteristics of information (ESRS 1, Appendix B; IFRS Conceptual Framework, Chapter 2), I derived a set of six questions on relevance, completeness, neutrality, findability, presentation format, and understandability that students were required to answer for the firm they just analyzed. *REP_QUALITY* is the first principal component of the answers to the six questions. Appendix A contains the questions, and Panel D, Table 1 shows summary statistics of the variables.

3.2 Experience, report characteristics, and information acquisition

In the following, I first discuss how "internal" experience, i.e., experience gained by collecting data on more firm-years as well as "external" experience, i.e., experience gained from prior internships or accounting courses, is associated with information acquisition. Second, I present results from analyzing report characteristics, i.e., how the way the information is presented is associated with information acquisition.

The main main driver reducing information acquisition costs is internal experience. Figure 1 plots the average *TIME* (on the y-axis) against *SESSION*, a count variable for how many "sessions," or full firm-years, a student has already completed. The figure shows that experience strongly reduces information acquisition costs, most notably in the first five to ten sessions. Table 3 supports these conclusions. It presents results from regressing (the natural logarithm of) *TIME* on *SESSION*, controlling for the number of data points actually disclosed (*EXIST*) using different combinations of fixed effects in columns (1) to (3).

Experience does not always have to be generated in the data collection process itself, however. Based on the pre-collection survey the students had to fill out to register for the data collection, I construct *EXPERIENCE* as the first factor from a principal component analysis for a range of questions that proxy for a student's pre-task engagement with annual and sustainability reports, their investing experience, or their educational and practical background. As column (4) of Table 3 shows, there is a weakly significant but negative effect of experience on acquisition time. In columns (5) and (6), I convert the information on the rank of the session into two indicator variables, splitting *SESSION* into *SESS_EARLY (SESS_LATE)* for the first (last) third of sessions a student has completed. The coefficients on the interaction terms between these indicator variables and *EXPERIENCE* demonstrate that prior, *indirect* experience mainly alleviates processing costs in the early sessions, while only playing a muted role when the student already has gained first *direct* experience with the task.

The coefficients on *ESRS*, the indicator variable capturing if the pages refer to 'ESRS' or 'CSRD,' are statistically insignificant for all specifications but turn negative when holding constant firm-year and analyst effects. It is probable that insignificance is driven by the sparsity of the indicator, as only 5.5% of all reports refer to ESRS.

Second, I apply standard measures from textual analysis to characterize the pages the students had to read in order to get the required data points. The results in Table 4 show that reports where data points are reported in a more dispersed way, i.e., students had to consider multiple pages to get all the necessary data points for a topic, are associated with higher acquisition costs. Confirming results from prior literature, the results also show that reports that are more difficult to read, i.e., those with a higher Fog-index or those that contain more uncertain language, require users to incur higher processing costs. Finally, the results speak to the fact that tables can aid users' information acquisition since those reports that contain more tables are associated with lower processing costs. I do not find any evidence that tone or images seem to be associated with processing.

4. Evidence from the field experiment

4.1 Setting, design, and measures

I now turn to the second part of the paper, the field experiment. The participants in my field experiment stem from two sources. My main source of participants is practitioners interested in obtaining a certificate as part of an introductory course on GHG emissions reporting. The course is offered by the Sustainability Reporting Navigator (SRN), an academic, open-science platform on sustainability reporting. To obtain the certificate, participants go through the 3-hour-long course and take part in the final assessment. I also recruit students from several universities that have taken part in a sustainability reporting course in recent years, many of whom have finished their studies. In a similar fashion, they can obtain the certificate by completing the experiment. Students are rewarded with a voucher of 25 Euros for participation. Panel A of Table 5 shows summary statistics for the 172 participants who are predominantly male, mostly between 18 and 24, and mostly have an undergraduate degree.[11]

I develop my own custom experimental platform that allows me to closely monitor how participants process the reports.[12] After entering the platform, participants answer a set of background

[11] Given that I launched the platform in July 2025, the majority (78%) of participants are (former) students.
[12] I present screenshots of the platform in Appendix B.

questions (which are listed in Appendix A, the answers to which are summarized in Panel A of Table 5). I primarily use the six questions to create two factor variables, *PRIOREXP* proxying for the participants' prior experience in investing and analyzing reports, as well as *ESGATTITUDE* proxying for their attitude towards ESG investing and purchasing.

They then receive instructions on the task, that is, to answer three questions on firms' GHG emissions performance by reading their sustainability reports. They are then randomly assigned one of three industries (airlines, cement, and steel) and proceed to a dashboard. Before entering the dashboard, I elicit their prior perception of the three firms' GHG emissions performance by letting participants rank the firms. The dashboard contains a list of the three real-world firms from the industry on the left side[13] and the three assessment questions that they have to answer on the right.

The three questions are single-choice, i.e., participants can pick one of the three firms or "This question can't be answered," and can be divided into two categories. Questions one and two are on the absolute value of emissions and the emissions intensity, respectively. These questions are easily understandable for the participants, i.e., do not require a deep understanding of GHG emissions reporting or assume familiarity with sustainability reporting. Furthermore, they constitute the key performance indicators in the area of GHG emissions, akin to an absolute number of profit or a profit margin. Question three, however, requires contextual understanding, as it asks which firm has the most ambitious climate target. This question can't be answered by comparing a single metric, as the participants have to factor in an array of disclosures, the availability of a net-zero goal, the relative decrease of emissions targeted, and the use of carbon offsets, for example. *CORRECT*, my first main variable, is the number of correct answers they give.

[13] The firms are Air France-KLM, Lufthansa, and IAG (airlines); Cementir, Heidelberg Materials, and Vicat (cement); and Salzgitter, SSAB, and Voestalpine (steel)

The list of firms shows one button per firm that takes the participant to a sub-page which contains the sustainability report of the focal firm, as well as a sidebar where the participant can first define an indicator (e.g., absolute value of Scope 1 emissions in tonnes of CO_2) and then enter the value they have extracted from the firm's report. This way, the participants can store numerical information and notes for each firm and keep them in the dashboard to have them available to answer the questions.

This is helpful for my design in three ways: First, my design requires me to prohibit access to the reports while the participants answer the questions to prevent participants from simply looking up the answers. By storing information that way, they can keep the information on hand during the assessment. Second, this enables me to track the total time a participant spends analyzing a firm's report, one of my main measures for acquisition costs. Third, it allows me to granularly track the participants' information acquisition behavior as well as the accuracy of the values they extract.[14]

Finally, after they have finalized their extraction, or after a maximum of 30 minutes, they answer the three questions. After clicking the submit button, I again ask them to rank the companies' GHG emissions performance. As my second main variable, I track the time from when they entered the industry to when they started the assessment as *FINISHED*. They are then redirected to the initial screen with an overview of how many questions they answered correctly. Upon successfully answering all three questions, participants get to claim their certificate. In case they have not answered all questions correctly, they can choose to continue the assessment by conducting the analysis for another industry at any time. The number of rounds is tracked in *ROUND*.

Based on this setup, I aggregate both *CORRECT* and *FINISHED* into *SCORE*, which is the sum of the two standardized variables, but where shorter times correspond to a higher value (i.e., instead of

[14] I derive *N_ENTRIES, N_NOTES, N_QUERIES* as well as *TIME_ENTRIES, TIME_NOTES, TIME_QUERIES,* and *TIME_READING* as variables capturing the information acquisition process.

subtracting the variable's mean from each entry before dividing it by the standard deviation, I subtract the entry from the mean). Table 5 presents summary statistics for all variables.

4.2 Findings

Figure 2 and Table 6 present the results for the experiment. Figure 2 plots the mean of *SCORE*, the variable combining the time spent (*FINISHED*) and accuracy (*CORRECT*), for each of the four experimental groups. The figure highlights that the scores for participants analyzing standardized, i.e., post-CSRD reports are higher, regardless of the use of *AI*. In addition, for both standardized and unstandardized reports, the usage of AI decreases the average score. Panel A in Table 6 supports this by showing means and standard errors for *SCORE* and its components. The composite processing efficiency score ranges from 1.865 for the group analyzing pre-CSRD reports with AI to 2.451 for the group analyzing CSRD reports with AI. *FINISHED* ranges from 19.318 minutes to 21.321 minutes, and the number of correctly answered questions ranges from 1.9493 to 2.4162. The number of participants per group ranges from 26 to 48.

Panel B in Table 6 shows the results of a two-way ANOVA. Results reveal a significant effect of the *CSRD* treatment (F=7.090, p=0.0082), whereas neither the *AI* (F=0.039, p=0.8437), or the interaction of *AI* and *CSRD* (F=0.353, p=0.5528) is statistically different from zero. Panel C presents results from group comparisons. Consistent with the previous result, there is a positive and significant difference in means between the standardized reports and non-standardized reports groups, showing positive processing effects of standardization. For AI and non-AI user groups, standardization shows to have a similar effect size, specifically a *SCORE* difference of 0.360 (t=1.675, p=0.0949) and 0.292 (t=1.494, p=0.1361), respectively. Using AI, in both the standardized and unstandardized settings, reduces the score by 0.175 (t=-0.825, p=0.4100) and 0.107 (t=-0.536, p=0.5920), respectively.

Table 7 presents the same tests in an OLS framework on a participant-industry level, regressing *SCORE* on the two indicator variables for the groups (*AI* and *CSRD*) as well as their interaction (*AI* × *CSRD*). Column (2) adds industry fixed effects, thereby controlling for the fact that analyzing firms in an industry can be more difficult given more difficult to understand economic fundamentals. Finally, column (3) controls for *ROUND,* the number of rounds the participant has already completed, proxying for internal experience, i.e., experience gained throughout the process. Similar to the previous analyses, results show a small, negative (but statistically insignificant) coefficient on *AI* and the interaction, indicating that using AI has a negative effect on processing. Standardization has a positive and statistically significant effect on the score. With a statistically significant coefficient almost the same size as the main standardization effect, *ROUND* reveals that engaging in another experimental round increases *SCORE,* showing that gaining experience significantly increases the processing efficiency.

Columns (4) and (5) of Table 7 split the composite score into its components. First, it becomes apparent, that the treatment effects affect both components similarly, i.e., using AI increases the time it takes to finish the assessment as well as decreases the accuracy. The effect of standardization on the time spent is negative, albeit not statistically significant. For accuracy, there is a positive and significant standardization effect.

Turning to the results on *ROUND*, I find that internal experience gained through absolving more rounds leads to higher score, lower time to finish, as well as higher accuracy of the answers. Documenting the strong learning effect as a major determinant of processing mirrors the findings from the first stage of the paper. Thus, in Table 8, I further investigate how this interacts with standardization and the use of AI. Specifically, I repeat the analyses from Table 7, but split the sample by *AI*-treatment groups. The table shows results from OLS regressions of *SCORE* on *ROUND* and *CSRD* for the AI-treated group in columns (1) and (3) and for the manual assessment group in columns (2) and (4). I find that users without access to AI learn in the experimental rounds, whereas those with access do not.

Furthermore, the positive effect of standardization is concentrated in the non-AI group, suggesting that benefits from experience and standardization only occur when users manually solve the task.

Finally, I turn toward my analyses on how external experience and attitude towards ESG affect information processing. Columns (1) and (2) of Table 9 present results from OLS regressions on samples split by the median of *PRIOREXP*, the first factor from a principal component on a list of pre-experimental questions eliciting the participants' prior experience of investing in stocks, reading a firm's reports, and their familiarity with using AI. I find a positive and highly significant coefficient on *CSRD* for participants with above-median prior experience compared to a small and insignificant coefficient for the less experienced group, suggesting that standardization primarily works for those users who are already familiar with the subject. Experience gained in the experiment, however, seems to play a stronger role only for those less experienced participants since the coefficient on *ROUND*, the number of industries analyzed, is higher in this group.

In a similar vein, columns (3) and (4) split the sample according to the participants' ESG attitude, as elicited in a pre-experimental survey on their preferences for ESG investing and purchasing. Coefficients are in line with my main results, i.e., positive effects for the CSRD, but negative effects for using AI, albeit statistically insignificant. This suggests that participants who are ideologically aligned with the substantive content, in this case, ESG matters, do not perform differently from those who are not, potentially due to the experimental setting and context.

5. Conclusion

In this study, I characterize the information processing environment for sustainability reporting. Relying on a field study tracking the information acquisition behavior from ESG disclosures *before* the EU's Corporate Sustainability Reporting Directive (CSRD), I first provide descriptive evidence on the time it takes to collect information related to emissions and own workforce topics. Documenting an average acquisition time for relevant data points of 30 to 40 minutes, I then show how personal experience 'inside' and 'outside' the project helps to reduce the acquisition costs. Finally, I test how the dimensions that are getting standardized through the CSRD, i.e., disclosures being made in a centralized location and following a common structure, or using a tabular presentation format, are correlated to processing.

Having established these associations, in the second part of the paper, I examine the information processing of disclosure *after* the CSRD. In a field experiment hosted on the Sustainability Reporting Navigator (SRN)'s Academy, where participants analyze firms' emissions reporting and answer questions about their performance to get a course certificate, I randomize whether participants analyze firms' reports before or after the regulation. Given the increasing importance of using generative Artificial Intelligence (AI) to automate this kind of data collection and interpretation work, I additionally randomize access to a custom AI tool that I developed.

I find positive and significant effects of standardization, but not for using AI. Whereas users analyzing reports prepared post-CSRD complete the task in a comparable time, their answers to the three questions are more accurate, on average. Coupled with the fact that the sections discussing the firm's greenhouse gas (GHG) emissions are significantly longer post-CSRD, these results speak to the fact that the more standardized presentation format due to the CSRD improved participants' understanding of the underlying economic phenomena. However, using AI did not significantly increase the accuracy of the participants' answers, but it increased the time they take to finalize the assessment. Despite the fact that they were satisfied with the AI's performance and analyses of the AI's answers to users' questions on the amount of Scope 1 emissions being overwhelmingly accurate, the fact that participants used the AI extensively led them to take longer, on average, negatively affecting their efficiency.

My study extends academic work on the standardization of reporting as well as recent work examining how using generative AI changes productivity and work in applied settings, as well as provides practically relevant insights to practitioners, standard setters, and regulators. First, whereas standardization has been studied extensively in other settings, predominantly for financial reporting (Leuz and Wysocki 2016), sustainability provides a good setting to study how regulatory-induced standardization changes reporting practices and the processing of information. This is because sustainability reporting, despite its widespread adoption and relevance on policymakers' agenda, has lacked a mandatory regime so far. In addition, as its unique feature, e.g., the inclusion of broader stakeholder views, little to no monetization, and a plethora of metrics (Christensen, Hail, and Leuz 2021), makes the setting a fertile ground to study these effects. My findings support the claim that mandatory standardization decreases processing costs, providing first-stage results for further investigations, e.g., into the real effects of these mandates.

Second, the rapid rise of generative AI has spurred research on the new technology, e.g., through chatbots, affects the productivity of knowledge workers. While large-scale randomized control trials (RCTs) have shown that AI boosts productivity for call-center agents (e.g., Brynjolfsson, Li, and Reynold 2025) or software developers, a study on strategy consultants' use of generative AI (Dell'Acqua et al. 2023) has shown mixed results for a subset of tasks. Similarly, a meta-study, albeit on AI more generally, revealed that more than half of the results for the human-AI groups fail to outperform either group alone. My findings echo more cautious notes about the potential of generative AI to 'replace' humans for knowledge-intensive and high-stakes tasks.

Naturally, the study is subject to limitations that give rise to future research. First, while my descriptive evidence covers a large sample of European firms, it is limited to two ESRS standards (E1 and S1). Future work should extend these analyses to other environmental and social standards, particularly those where pre-CSRD disclosure was sparse. Second, the field experiment is restricted to three industries with relatively high levels of pre-existing disclosure, which likely understates the benefits of standardization and limits generalizibility. Third, my AI treatment relies on a single retrieval augmented generation (RAG)-based chatbot design. Future research could explore alternative architectures, degrees of interactivity (Croom 2025), and integration with structured data sources to assess when AI may complement human analysis more effectively. Finally, while the experimental setting allows for causal inference, using actual firms' pre- and post-CSRD reports inevitably adds complexity and noise. Longitudinal field studies could shed further light on how standardization and AI shape disclosure processing in practice over time.

In conclusion, this paper documents the extent to which sustainability disclosures are costly to process, that regulatory standardization under the CSRD improves their usability, and that generative AI does not yet offer a reliable substitute for these regulatory improvements. The study, therefore, has practical implications for standard setters, especially in light of the "Omnibus initiative" that aims to scale back the CSRD. My results demonstrate that the CSRD improves the usability of sustainability disclosures even in industries where climate reporting was already relatively mature. This suggests that rolling back requirements to increase reporting discretion risks undermining the very accessibility gains the CSRD has achieved. For practitioners, the findings caution against overreliance on generative AI for complex analytical tasks. While AI can reduce fatigue and improve engagement, it does not yet deliver consistent efficiency gains in the processing of structured corporate disclosures. Regulators and firms alike should therefore view AI as a complement to, rather than a substitute for, robust reporting standards. Ultimately, while technology may evolve, the case for standardized, reliable, and comparable sustainability reporting remains strong. Its scope, as always, will be determined not only by evidence but also by politics.

Appendix A. Variable definitions

Variable	Definition
Student analyst characteristics	
FEMALE	Indicator variable if the student is female.
ACCCOURSE	Number of accounting courses the student has attended.
SUSTCOURSE	Indicator variable if the student has attended a sustainability reporting course.
ACTIVIST	Indicator variable if the student has participated in the climate movement (e.g., Fridays for Future, Extinction Rebellion, Letzte Generation).
INTERNSHIPS	Number of internships in investment banking, consulting or auditing.
FAMILIARITY	Answer to the question "How familiar are you with sustainability reporting (leaving aside your thesis)?" on a 7-point Likert scale.
READSREPORTS	Answer to the question "How often are you working with companies' (financial or sustainability) reports?" on a 7-point Likert scale.
INVESTEXP	Answer to the question "How familiar are you with investing in stocks?" on a 7-point Likert scale.
EXPERIENCE	The first principal component of the answers to the seven, above presented questions on prior experience.
Data entry session characteristics	
TIME	Total time it takes to collect all data points from a firm's annual or sustainability report for one topic in a given year.
EXIST	Number of data points that are reported in a firm's annual or sustainability report for a topic in a given year.
SESSION	Incrementing variable for the number of sessions, i.e., full collections of a firm's topic in a given year, of a student.
TOPIC_EMISSION	Indicator variable that equals 1 if the collected data points are from ESRS E1 (climate change), and 0, if they are from ESRS S1 (own workforce).
ESRS	Indicator variable that equals 1 if the pages where the students found the data points contain a reference to the CSRD or ESRS, 0 if otherwise.
Sustainability report attributes	
DISPERSION	The Herfindahl-index of the page numbers where the students found the data points for a firm-year-topic.
LENGTH	The average number of characters of the pages where the students found the data points for a firm-year-topic.
DIFFICULT	The average Fog index of the pages where the students found the data points for a firm-year-topic.

TONE	The average net tone, i.e., share of positive minus negative words to total words based on Loughran and McDonald (2012) of the pages where the students found the data points for a firm-year-topic.
COMPLEX	The average share of complex words words to total words based on Loughran and McDonald (2012) of the pages where the students found the data points for a firm-year-topic.
UNCERTAIN	The average share of uncertain words words to total words based on Loughran and McDonald (2012) of the pages where the students found the data points for a firm-year-topic.
TABLES	The total number of tables of the pages where the students found the data points for a firm-year-topic.
IMAGES	The average area of the pages where the students found the data points for a firm-year-topic that is covered by an image.

Assessment of reporting quality

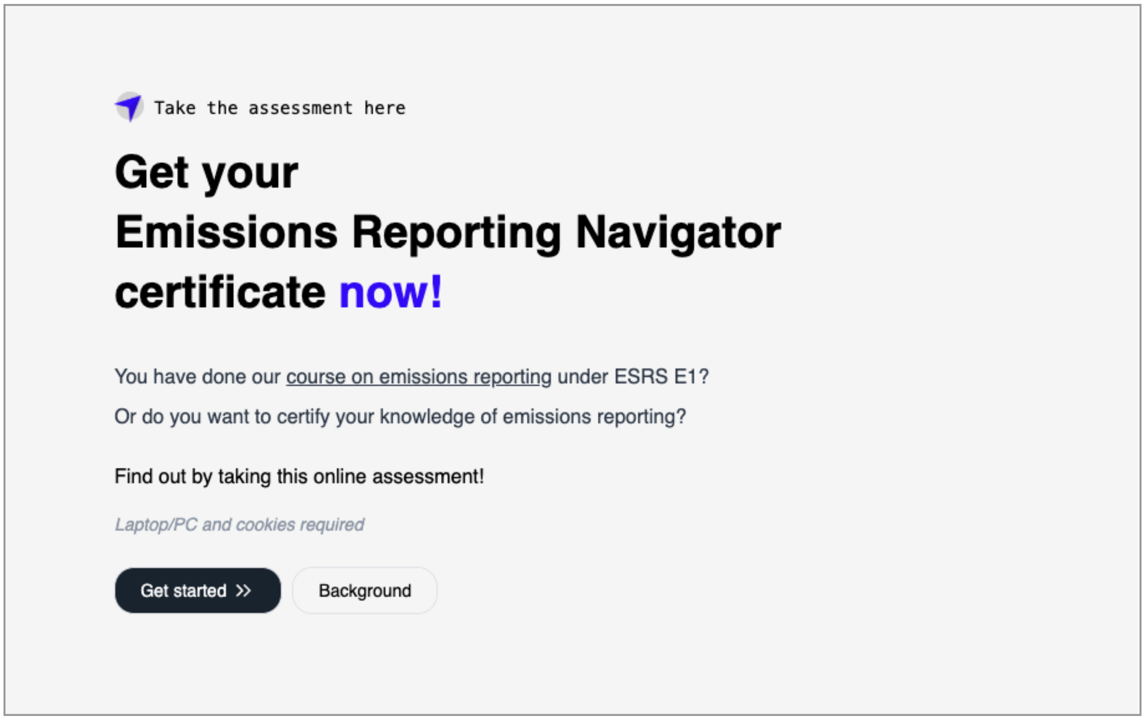
RELEVANCE	Answer to the question "In order to judge the firm's sustainability performance, the information from the reports seems relevant for my assessment." on a 7-point Likert scale.
COMPLETE	Answer to the question "I have the impression that the firm presents its sustainability performance in a complete, i.e., non-selective manner." on a 7-point Likert scale.
NEUTRAL	Answer to the question "The information in the reports seems to be presented in a neutral and unbiased way." on a 7-point Likert scale.
FINDABLE	Answer to the question "Overall, I was able to find the data points easily in the documents." on a 7-point Likert scale.
PRESENTED	Answer to the question "The way the information in the reports is presented makes it easy for me to compare the sustainability performance to other firms in my sample." on a 7-point Likert scale.
UNDERSTAND	Answer to the question "Overall, I found the information in the reports to be understandable, i.e., it was presented clearly and concisely." on a 7-point Likert scale.
REPQUAL	The first principal component of the answers to the six, above presented questions on reporting quality.

Experimental participant characteristics

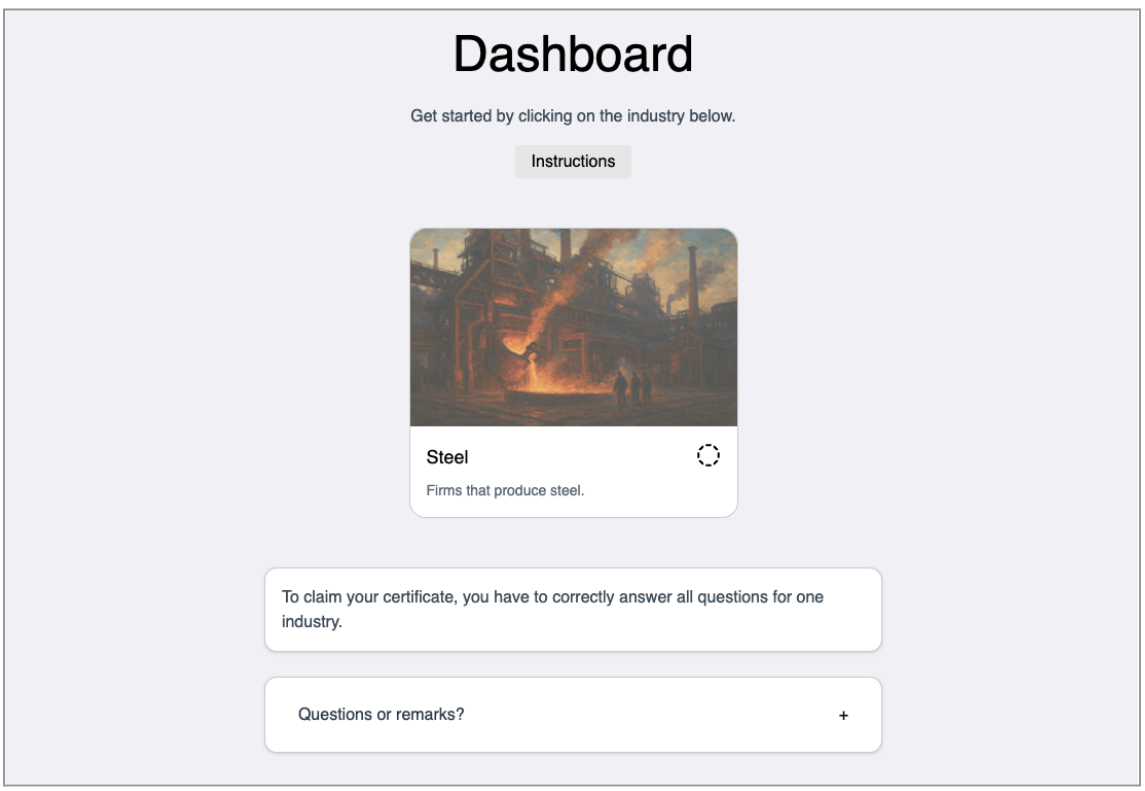
DEGREE	Answer to the question "What is the highest level of education you obtained?" with the responses "no formal education; A-levels/high school; vocational training; undergraduate degree; postgraduate Master's degree; doctorate" transformed to a numerical factor.
AGE	Answer to the question "How old are you?" with the responses "18-24; 25-34; 35-44; 45-54; 55-65; 65+" transformed to a numerical factor.

INVESTING	Answer to the question "How often are you investing in the stock market?" on a 7-point Likert scale.
PRIMSRC	Answer to the question "How often are you considering primary sources (e.g., annual reports) for your investment decisions?" on a 7-point Likert scale.
AIUSE	Answer to the question "How often do you use generative AI tools?" on a 7-point Likert scale.
PRIOREXP	The first principal component of the answers to the four, above presented questions on prior experience.
ESGINVEST	Answer to the question "How important are environmental, social, and governance (ESG) considerations for your investment decisions?" on a 7-point Likert scale.
ESGPURCHASE	Answer to the question "How important are ESG considerations for your purchasing decisions (e.g., for clothing, traveling, transportation)?" on a 7-point Likert scale.
ESGATTITUDE	The first principal component of the answers to the two, above presented questions on attitude towards environmental, social, and governance considerations.
Information acquisition variables	
FINISHED	Time (in minutes) a participant takes to finish the assessment of the three firms in an industry and proceed to the three final questions.
CORRECT	Number of correct answers to the three final questions.
SCORE	Sum of the standardized *FINISHED* and *CORRECT* variables, where *FINISHED* is constructed to attribute shorter times to a higher value (i.e., instead of subtracting the variable's mean from each entry before dividing it to the standard deviation, I subtract the the entry from the mean).
ROUND	Number of the (maximum of three) rounds the participant has participated in.

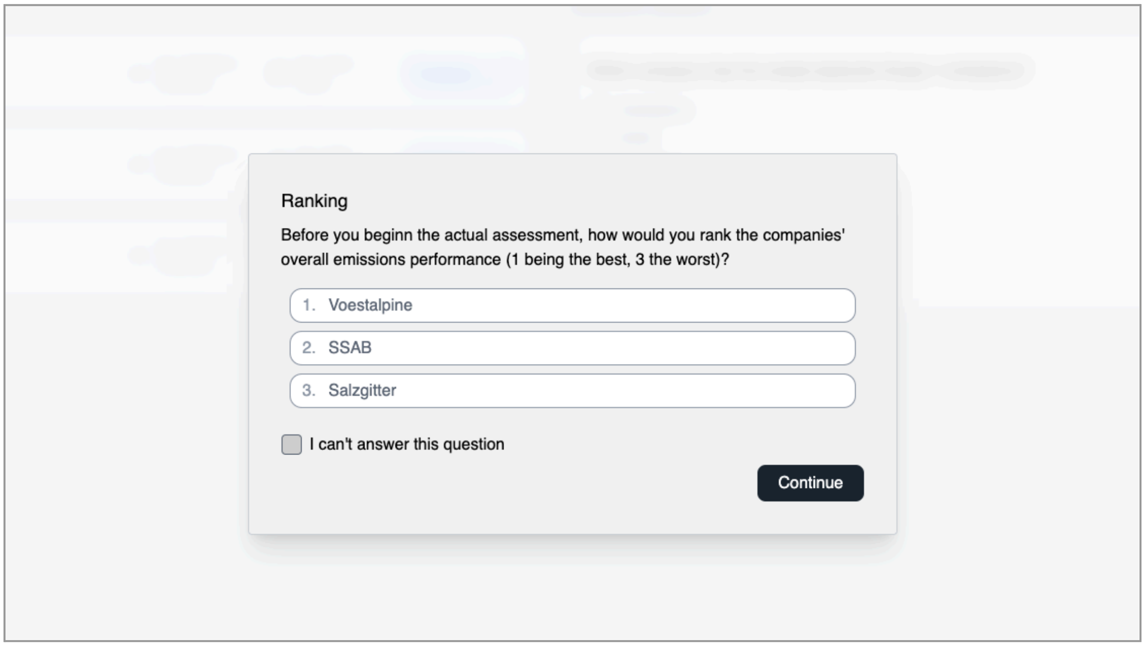
Appendix B. Experimental screens



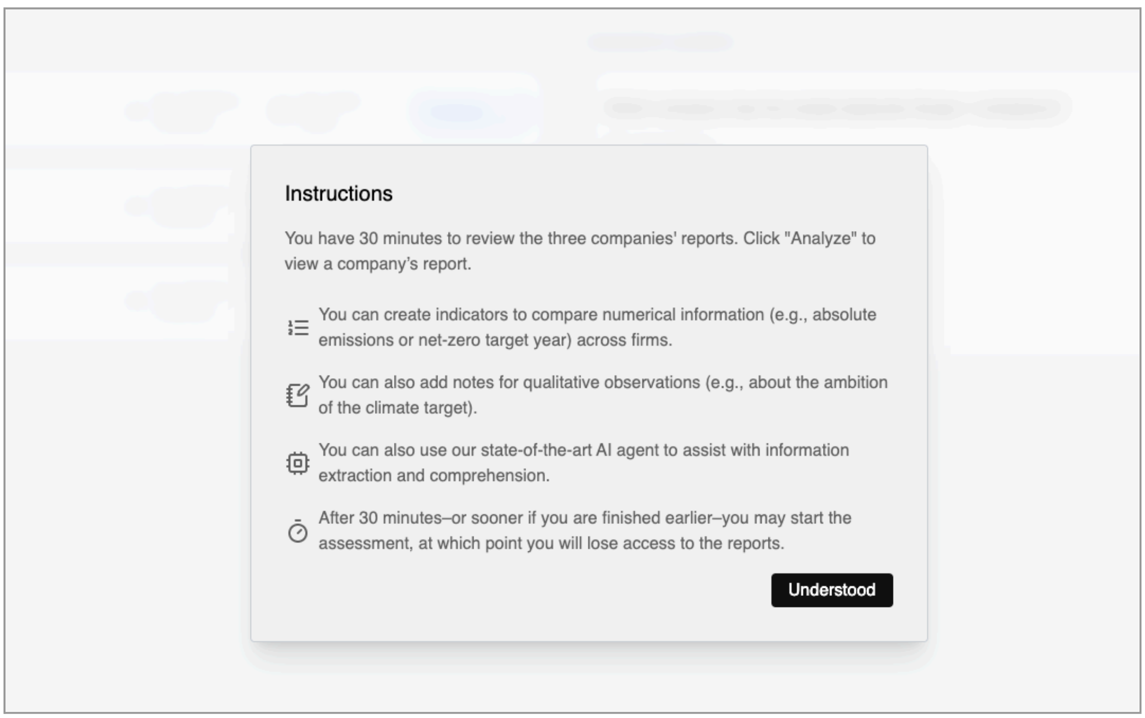
Landing page



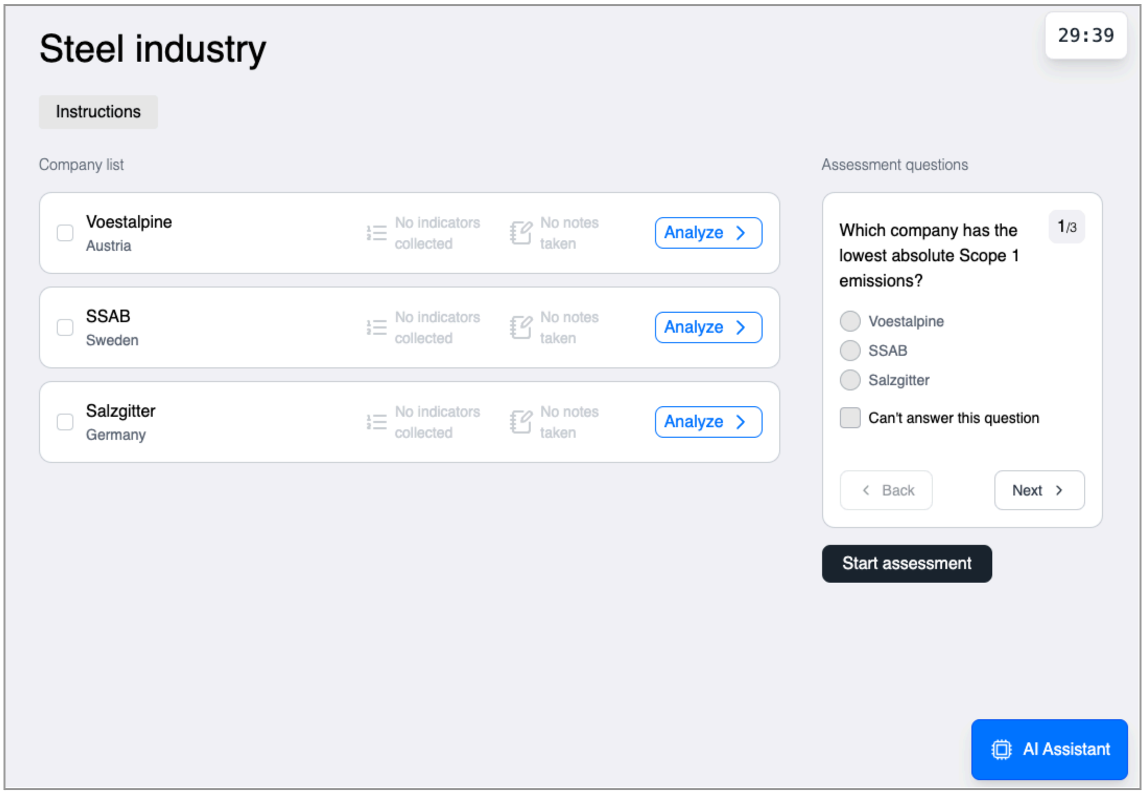
Dashboard overview after having answered the personal questions



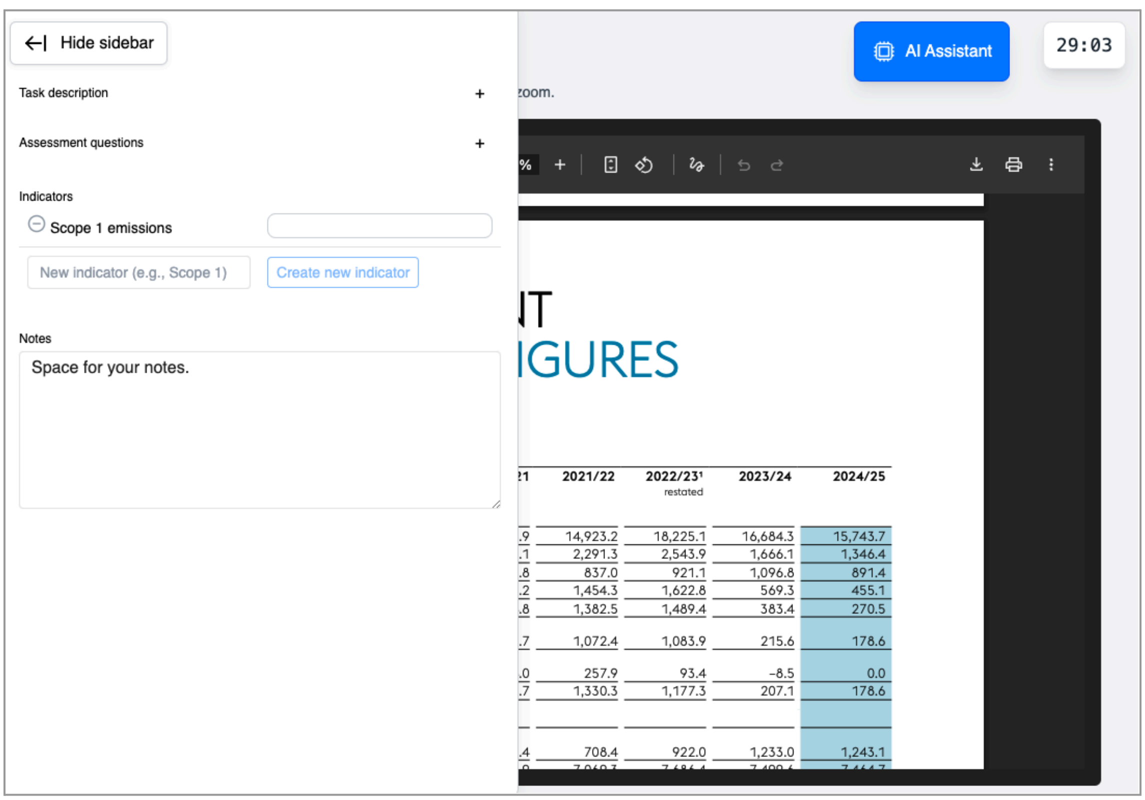
Participants' ranking of the firms in an industry to elicit a *prior* of their perception



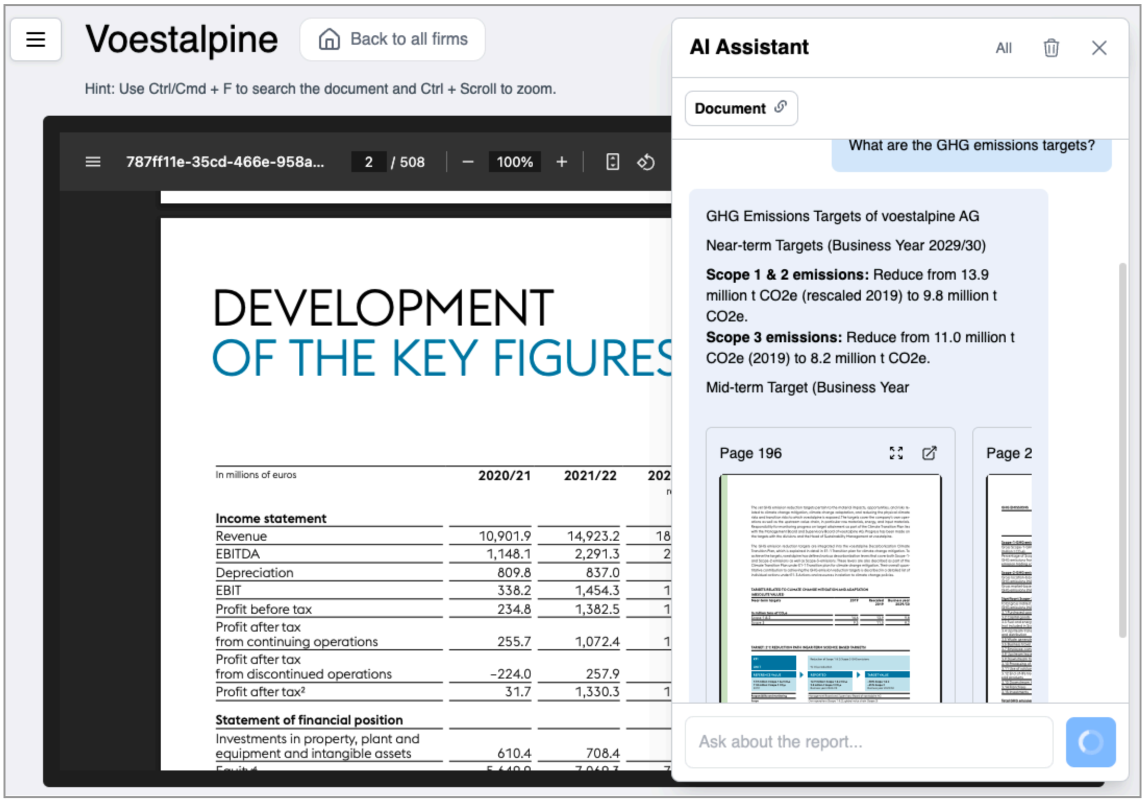
Instructions before starting an industry



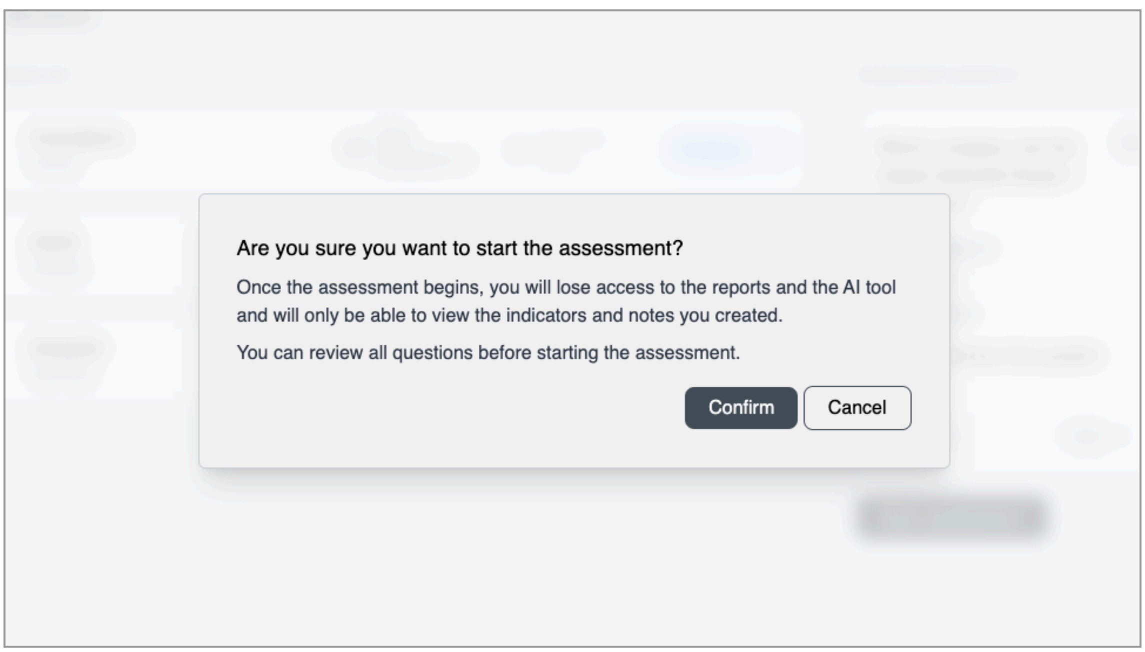
Industry overview screen with the company list (left) and the final questions (right), as well as the timer (top right) and the AI assistant (bottom right)



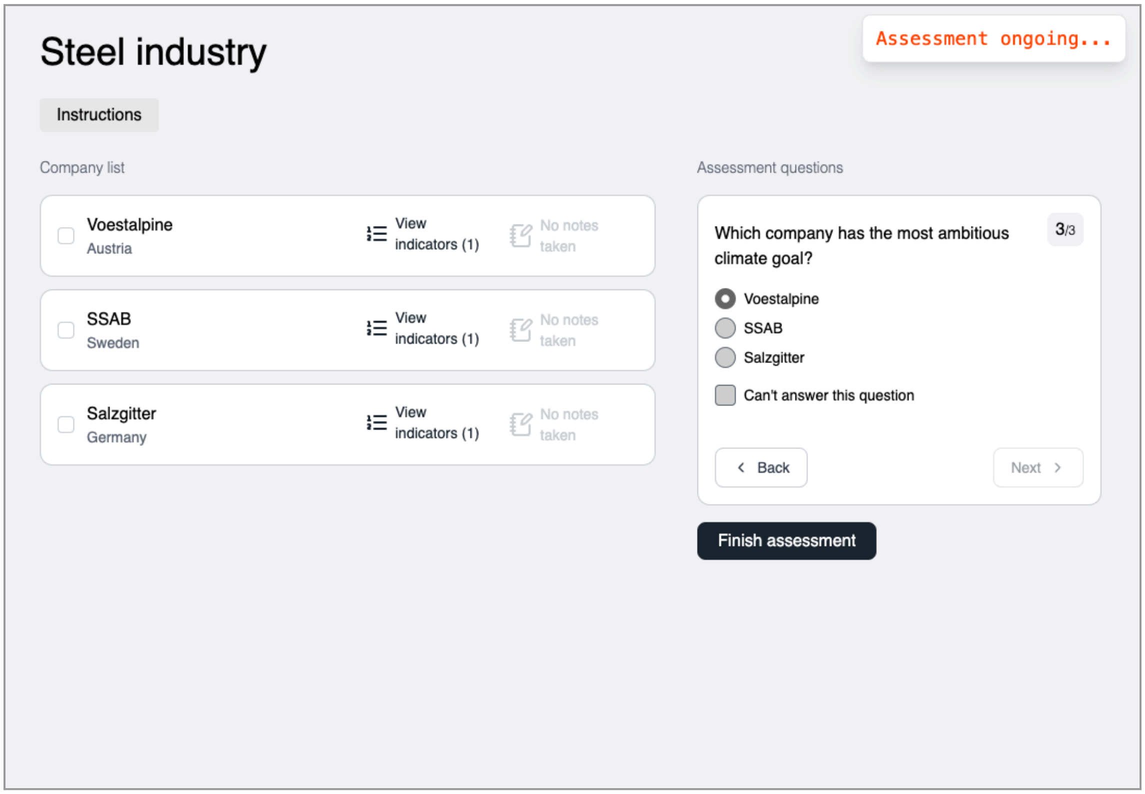
View of one firm's PDF report and the opened sidebar



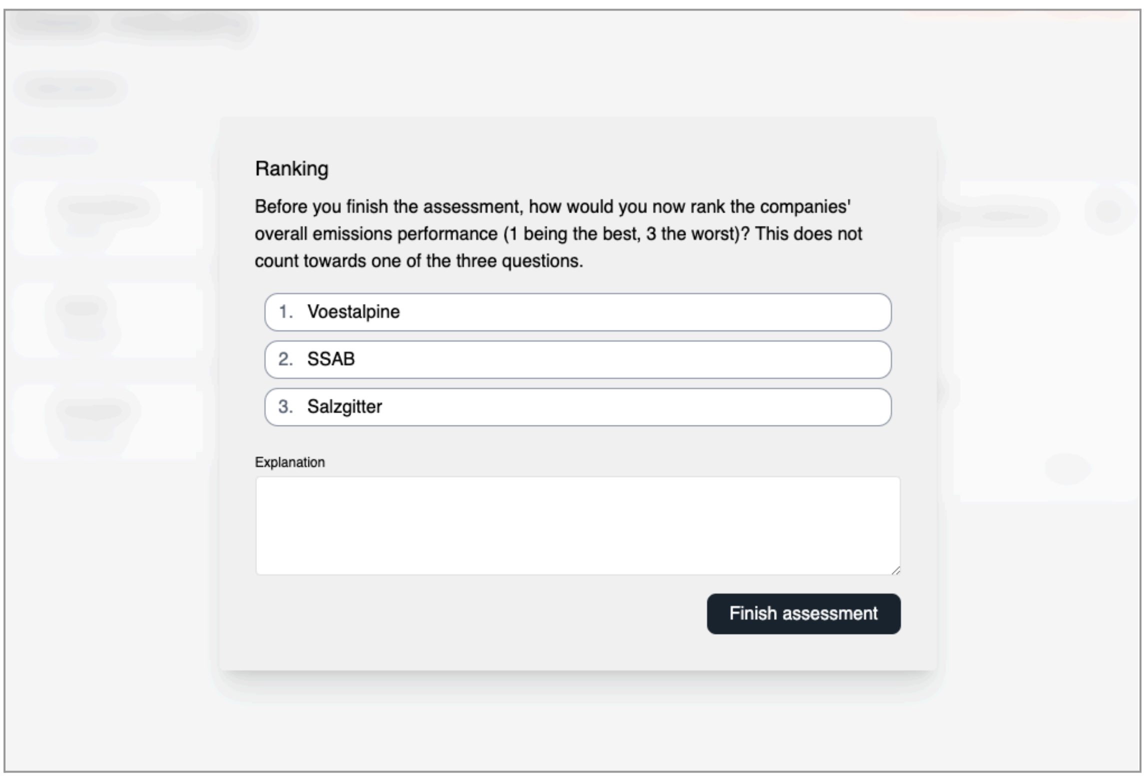
View of one firm's PDF report and the opened AI chatbot



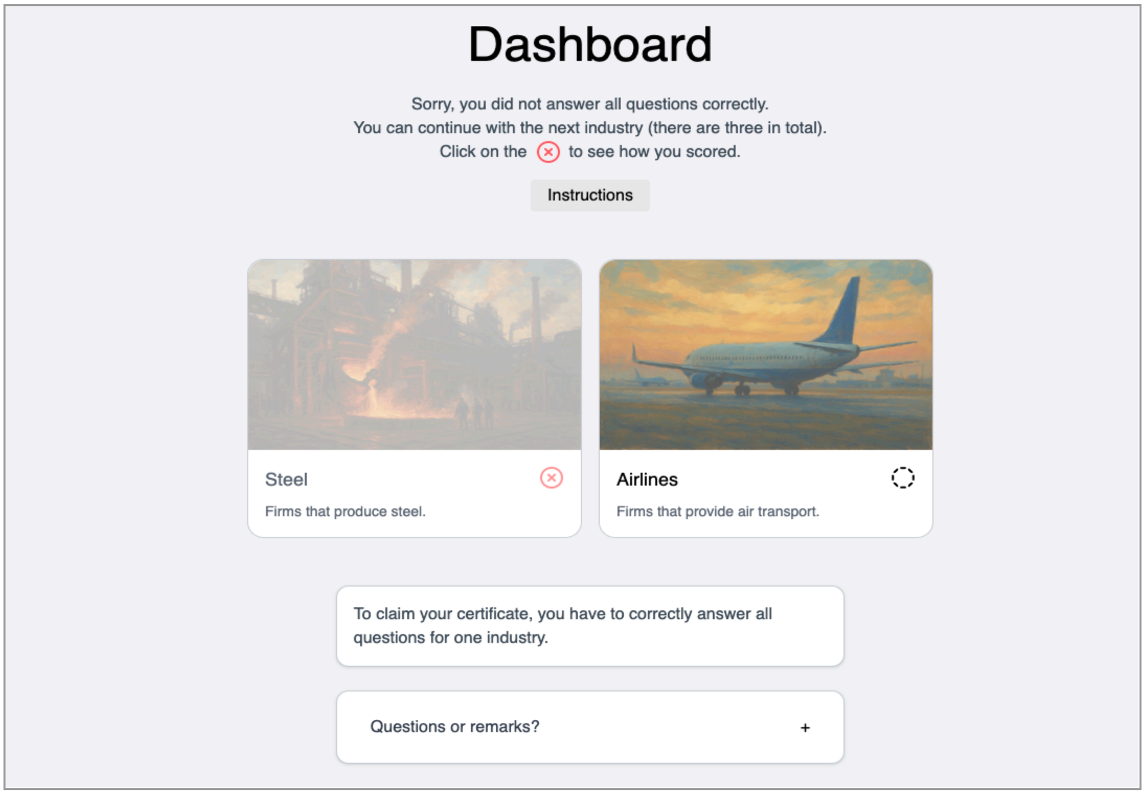
Pop-up before starting the assessment



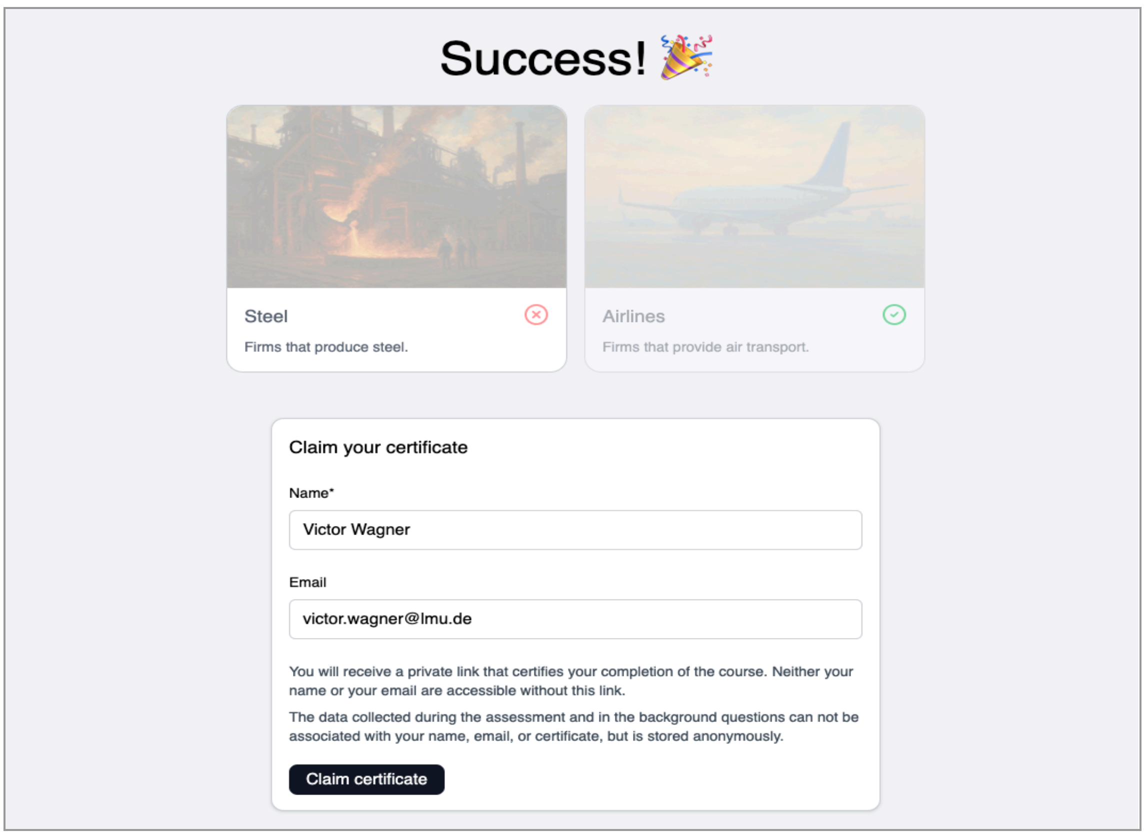
Ongoing assessment (no access to firms' reports, only indicators and notes)



Final ranking in an industry



Dashboard after having incorrectly answered the questions with the prompt to start again



Dashboard after having correctly answered the questions for the second industry

References

Abhayawansa, S. (2022). Swimming against the tide: Back to single materiality for sustainability reporting. *Sustainability Accounting, Management and Policy Journal*, *13*(6), 1361–1385.

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy*, *40*(121), 13–58.

Allee, K. D., Do, C., & Raymundo, F. G. (2022). Principal component analysis and factor analysis in accounting research. *Journal of Financial Reporting*, *7*(2), 1–39.

Autor, D. (2024). *Applying ai to rebuild middle class jobs*. Working Paper, available at: https://www.nber.org/papers/w32140

Ball, R., Robin, A., & Wu, J. S. (2003). Incentives versus standards: Properties of accounting income in four East Asian countries. *Journal of Accounting and Economics*, *36*(1–3), 235–270.

Blackrock. (2020). *Towards a common language for sustainable investing*. https://corpgov.law.harvard.edu/2020/01/22/towards-a-common-language-for-sustainable-investing/

Blankespoor, E. (2019). The impact of information processing costs on firm disclosure choice: Evidence from the xbrl mandate. *Journal of Accounting Research*, *57*(4), 919–967.

Blankespoor, E., deHaan, E., & Li, Q. (2024). *Generative AI in financial reporting*. Working Paper, available at: https://doi.org/10.2139/ssrn.4986017

Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, *70*(2–3).

Blankespoor, E., Miller, B. P., & White, H. D. (2014). Initial evidence on the market impact of the XBRL mandate. *Review of Accounting Studies*, *19*(4), 1468–1503.

Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons*, *16*(3), 233–243.

Bradshaw, M. T., Miller, G., & Serafeim, G. (2010). *Accounting Method Heterogeneity and Analysts' Forecasts*, Working Paper.

Brüggemann, U., Hitz, J.-M., & Sellhorn, T. (2013). Intended and unintended consequences of mandatory ifrs adoption: A review of extant evidence and suggestions for future research. *European Accounting Review*, *22*(1), 1–37.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, *140*(2), 889–942.

Carmona, S., & Trombetta, M. (2008). On the global acceptance of IAS/IFRS accounting standards: The logic and implications of the principles-based system. *Journal of Accounting and Public Policy*, *27*(6), 455–461.

Challapally, A., Pease, C., Raskar, R., & Pradyumna, C. (2025). *The GenAI Divide: State of AI in Business 2025*. Availablet at: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

Chang, S., Anderson, A., & Hofman, J. M. (2025). *Chatbench: From static benchmarks to human-ai evaluation*. Working Paper, available at: https://arxiv.org/abs/2504.07114.

Cho, C. H., Michelon, G., & Patten, D. M. (2012). Impression management in sustainability reports: An empirical investigation of the use of graphs. *Accounting and the Public Interest*, *12*(1), 16–37.

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society*, *35*(4), 431–443.

Choi, J. H. & Xie C. (2025). *Human + AI in accounting: Early evidence from the field.* Working Paper, available at:

Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, *26*(3), 1176–1248.

Christensen, H. B., Lee, E., Walker, M., & Zeng, C. (2015). Incentives or standards: What determines accounting quality changes around ifrs adoption? *European Accounting Review*, *24*(1), 31–61.

Chychyla, R., Leone, A. J., & Minutti-Meza, M. (2019). Complexity of financial reporting standards and accounting expertise. *Journal of Accounting and Economics*, *67*(1), 226–253.

ClientEarth. (2021). *Accountability Emergency: A review of UK-listed companies' climate change-related reporting (2019-20)*. Available at: https://www.clientearth.org/latest/documents/accountability-emergency-a-review-of-uk-listed-companies-climate-change-related-reporting-2019-20/

Croom, J. (2025). Interactivity and illusions of ability: How using generative ai affects investor judgments. *Journal of Accounting Research*, 1475-679X.70017.

Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2024). *The effects of generative AI on high skilled work: Evidence from three field experiments with software developers*. Working Paper, available at: https://doi.org/10.2139/ssrn.4945566

Daske, H., Seregni, C., & Uckert, M. (2023). *Unstandardized Accounting Terminology*. Unpublished Working Paper.

De Franco, G., Kothari, S. P., & Verdi, R. S. (2011). The Benefits of Financial Statement Comparability: The benefits of financial statement comparability. *Journal of Accounting Research*, *49*(4), 895–931.

De George, E. T., Li, X., & Shivakumar, L. (2016). A review of the IFRS adoption literature. *Review of Accounting Studies*, *21*(3), 898–1004.

De Villiers, C., Dimes, R., & Molinari, M. (2024). How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research. *Sustainability Accounting, Management and Policy Journal*, *15*(1), 96–118.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality.* Working Paper, available at: https://ssrn.com/abstract=4573321.

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, *54*(1), 64–87.

Dinh, T., Husmann, A., & Melloni, G. (2023). Corporate sustainability reporting in europe: A scoping review. *Accounting in Europe*, *20*(1), 1–29.

Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2024). A scoping review of ChatGPT research in accounting and finance. *International Journal of Accounting Information Systems*, *55*, 100715.

Dong, Y., Li, O. Z., Lin, Y., & Ni, C. (2016). Does information-processing cost affect firm-specific information acquisition? Evidence from xbrl adoption. *Journal of Financial and Quantitative Analysis*, *51*(2), 435–462.

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2015). The Determinants and Consequences of Information Acquisition via EDGAR. *Contemporary Accounting Research*, *32*(3), 1128–1161.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245.

European Commission (EC). (2019). *Guidelines on non-financial reporting: Supplement on reporting climate-related information*. https://eur-lex.europa.eu/legal-content/EN/TXT?uri=CELEX:52019XC0620(01)

Eurosif, IIGCC, PRI, Corporate Leaders Group, E3G, & GRI. (2025). *Weakening the EU's sustainability rules risks damaging competitiveness and growth warn companies and investors*. Available at: https://www.eurosif.org/wp-content/uploads/2025/07/2025.07.01-Investor-and-Business_Joint-Statement_PR.pdf

Fiechter, P., Hitz, J., & Lehmann, N. (2022). Real effects of a widespread CSR reporting mandate: Evidence from the european union's csr directive. *Journal of Accounting Research*, *60*(4), 1499–1549.

Floyd, E., & List, J. A. (2016). Using field experiments in accounting and finance. *Journal of Accounting Research*, *54*(2), 437–475.

Forster, K., Keil, L., Wagner, V., Müller, M. A., Sellhorn, T., & Feuerriegel, S. (2025). *Assessing corporate sustainability with large language models: Evidence from europe*. Working Paper, available at: https://ssrn.com/abstract=5361703

Föhr, T. L., Schreyer, M., Juppe, T. A., & Marten, K.-U. (2023). *Assuring sustainable futures: Auditing sustainability reports using ai foundation models.* Working Paper, available at: https://ssrn.com/abstract=4502549

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025*. Available at: https://kpmg.com/xx/en/our-insights/ai-and-technology/trust-attitudes-and-use-of-ai.html

Guay, W., Samuels, D., & Taylor, D. (2016). Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, *62*(2–3), 234–269.

Hodge, F. D., Kennedy, J. J., & Maines, L. A. (2004). Does search-facilitating technology improve the transparency of financial reporting? *The Accounting Review*, *79*(3), 687–703.

Kim, A. G., Muhn, M., & Nikolaev, V. V. (2023). *From transcripts to insights: Uncovering corporate risks using generative ai.* Working Paper, available at: https://ssrn.com/abstract=4593660.

Kim, J. W., Lim, J.-H., & No, W. G. (2012). The effect of first wave mandatory xbrl reporting across the financial information environment. *Journal of Information Systems*, *26*(1), 127–153.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an ai assistant for essay writing task*. Working Paper, available at: https://arxiv.org/abs/2506.08872.

Laine, M. (2009). Ensuring legitimacy through rhetorical changes?: A longitudinal interpretation of the environmental disclosures of a leading Finnish chemical company. *Accounting, Auditing & Accountability Journal*, *22*(7), 1029–1054.

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, *60*(2–3), 110–135.

Larrinaga, C., & Bebbington, J. (2021). The pre-history of sustainability reporting: A constructivist reading. *Accounting, Auditing & Accountability Journal*, *34*(9), 162–181.

Levy, B. (2025). *Caution ahead: Numerical reasoning and look-ahead bias in ai models*. Working Paper, available at: https://ssrn.com/abstract=5082861.

Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, *54*(2), 525–622.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for knowledge-intensive nlp tasks*. Working Paper, available at: https://arxiv.org/abs/2005.11401.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247.

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *The Review of Corporate Finance Studies*, *8*(1), 1–41.

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10‐ks. *The Journal of Finance*, *66*(1), 35–65.

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1), 46.

Maines, L. A., & McDaniel, L. S. (2000). Effects of comprehensive-income characteristics on nonprofessional investors' judgments: The role of financial-statement presentation format. *The Accounting Review*, *75*(2), 179–207.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, *85*(6), 2107–2143.

Milne, M. J., & Gray, R. (2013). W(H)ither ecology? The triple bottom line, the global reporting initiative, and corporate sustainability reporting. *Journal of Business Ethics*, *118*(1), 13–29.

Niederhoffer, K., Rosen Kellerman, G., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025). AI-generated "workslop" is destroying productivity. *Harvard Business Review*.

O'Donoghue, O., & Roberts, D. (2024). *New work, new world*. Available at: https://www.cognizant.com/en_us/insights/documents/new-work-new-world-with-generative-ai-wf2064768.pdf

Papadopoulos, L. (2025). *Rage Against The Machine? (De)biased AI and Financial Assessments*. Unpublished Working Paper.

Riedl, C., & Weidmann, B. (2025). *Quantifying human-AI synergy*. Working Paper, available at: https://doi.org/10.31234/osf.io/vbkmt_v1.

Sellhorn, T., Wagner, V., Hombach, K., Meringdal, I. E., Donau, C., & Müller, M. A. (2023). Gap-Assessment zur Klimaberichterstattung am deutschen Kapitalmarkt. *Der Betrieb*, *9*, 465–472.

Soll, J. (2014). *The reckoning: Financial accountability and the rise and fall of nations*. Basic books.

Starks, L. T. (2023). Presidential address: Sustainable finance and esg issues – *value* versus *values*. *The Journal of Finance*, *78*(4), 1837–1872.

Statista. (2023). *Threads Shoots Past One Million User Mark at Lightning Speed*. Available at: https://www.statista.com/chart/29174/time-to-one-million-users/

The Economist. (2025). *Faith in God-like large language models is waning*. Available at: https://www.economist.com/business/2025/09/08/faith-in-god-like-large-language-models-is-waning

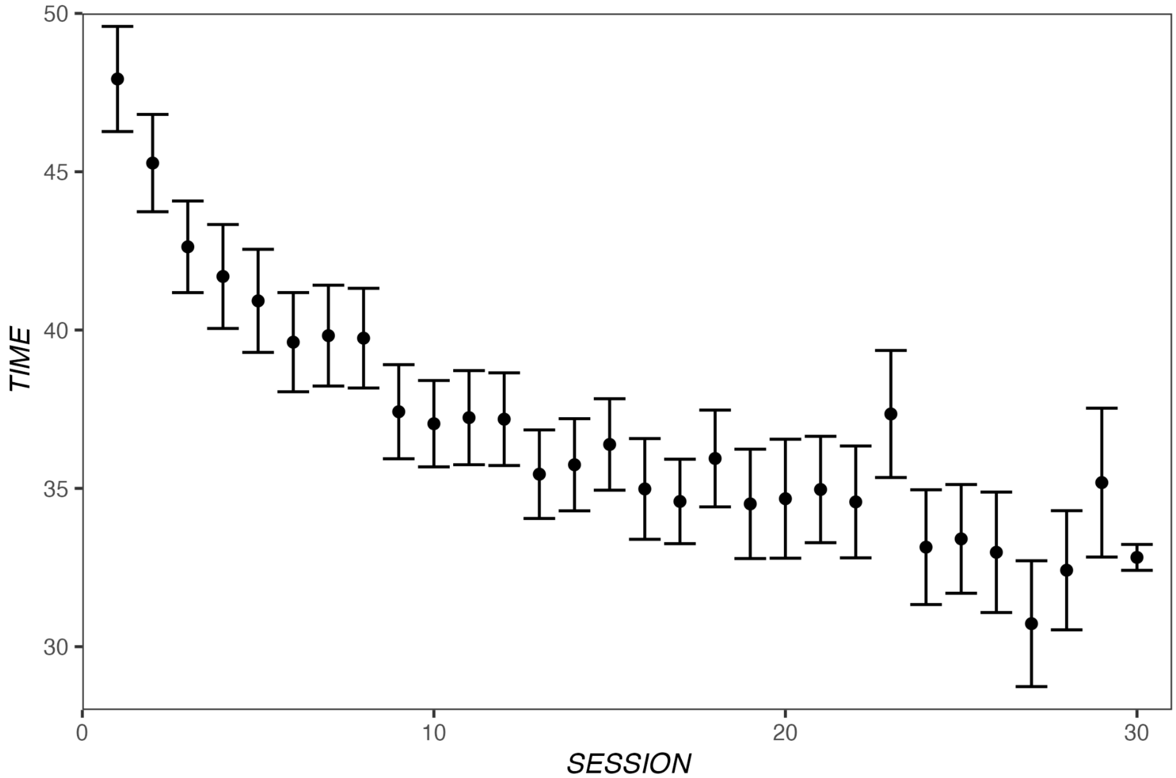
Wagenhofer, A. (2003). Economic consequences of internet financial reporting. *Schmalenbach Business Review*, *55*(4), 262–279.

Wagner, V., Müller, M. A., & Hagemeier, L. (2025). *How does the CSRD change reporting?* Available at: https://www.srnav.com/insights/how-does-the-csrd-change-reporting.

Weil, D., Graham, M., & Fung, A. (2013). Targeting transparency. *Science*, *340*(6139), 1410–1411.

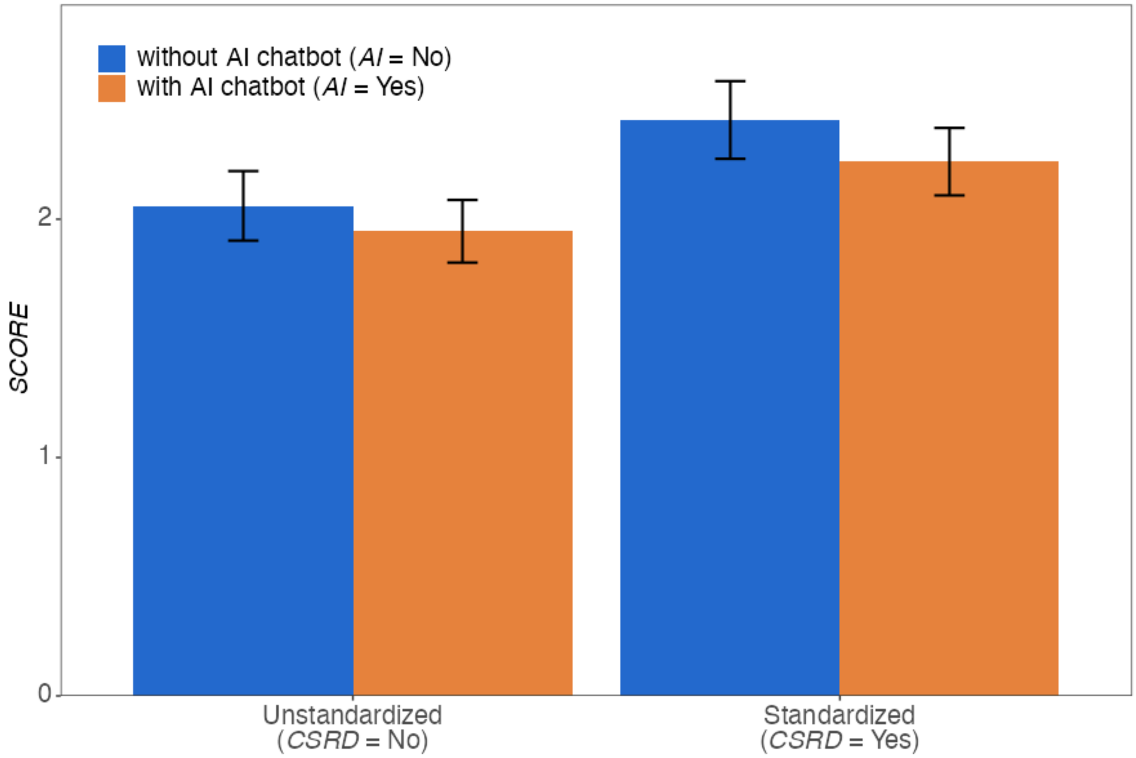
WWF. (2020). *Minimum regulatory requirements for climate alignment disclosure and metrics for financial institutions*. Available at: https://wwfeu.awsassets.panda.org/downloads/requirements_for_climate_disclosure___wwf_briefing__nov_2020_.pdf

Figure 1. Information acquisition and experience



Notes: This figure plots the average time a student analyst takes to enter data points for a firm-year (*TIME*) on the y-axis per *SESSION*, i.e., the number of firm-years they have already entered data for. For visualization purposes, I aggregate sessions over 30 in session 30. All variables are defined in Appendix A.

Figure 2. Processing efficiency by experimental group



Notes: This figure compares the means of *SCORE*, a composite of *FINISHED,* the time (in minutes) the participants took to finish the task and *CORRECT*, the number of correct answers they gave (based on the three questions). Errorbars represent the standard error. All variables are defined in Appendix A.

Table 1. Summary statistics for the field observations

	Mean	SD	Min	P25	P50	P75	Max	N
Panel A: Student analyst characteristics								
FEMALE	0.343	0.478	0.000	0.000	0.000	1.000	1.000	73
ACC_COURSE	2.956	1.476	0.000	2.000	3.000	4.000	5.000	73
SUST_COURSE	0.111	0.316	0.000	0.000	0.000	0.000	1.000	73
ACTIVIST	0.181	0.387	0.000	0.000	0.000	0.000	1.000	73
INTERNSHIPS	0.973	1.096	0.000	0.000	1.000	2.000	5.000	73
FAMILIARITY	2.904	1.511	1.000	2.000	3.000	4.000	7.000	73
READS_REPORTS	3.438	1.810	1.000	2.000	4.000	5.000	7.000	73
INVEST_EXP	3.959	1.611	1.000	3.000	4.000	5.000	7.000	73
SUST_RELEVANT	4.000	1.375	1.000	3.000	4.000	5.000	6.000	73
GREENWASHING	4.425	1.200	1.000	4.000	4.000	5.000	7.000	73
MATERIALITY	2.849	1.552	1.000	2.000	3.000	4.000	7.000	73
Panel B: Data entry session characteristics								
TIME	36.358	16.874	3.280	23.671	33.268	46.618	121.873	4,453
EXIST	19.425	14.366	0.000	8.000	17.000	28.000	97.000	4,453
SESSION	23.026	19.320	1.000	9.000	17.000	32.000	97.000	4,453
TOPIC_EMISSION	0.770	0.420	0.000	1.000	1.000	1.000	1.000	4,453
ESRS	0.055	0.229	0.000	0.000	0.000	0.000	1.000	3,182
Panel C: Sustainability report attributes								
DISPERSION	0.572	0.357	0.000	0.248	0.699	0.889	1.000	4,235
LENGTH	3.517	1.227	0.001	2.686	3.419	4.184	12.817	3,182
DIFFICULT	16.571	5.451	0.000	13.861	15.346	17.604	93.353	3,182
TONE	0.017	0.012	-0.041	0.010	0.017	0.024	0.074	3,182
COMPLEX	0.003	0.002	0.000	0.001	0.003	0.004	0.036	3,182
UNCERTAIN	0.007	0.005	0.000	0.004	0.006	0.008	0.131	3,182
TABLES	6.965	6.794	0.000	2.288	5.394	9.549	71.200	3,190
IMAGES	3.951	51.671	0.000	0.000	1.000	3.200	2812.000	3,190

Notes: This table provides descriptive statistics on a sample of the participants in Panel A, the data entry sessions in Panel B, and the attributes of the reports analyzed in Panel C. The number of participant characteristics in Panel A is different from the total number of participants in the data collection (N=143), because I have not been collecting personal characteristics from the beginning. All variables are defined in Appendix A.

Table 2. Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) *TIME*	1	.	.	.	.	.	.	.	.	.	.	.	.
(2) *EXIST*	.46	1	.	.	.	.	.	.	.	.	.	.	.
(3) *SESSION*	-.17	.01	1	.	.	.	.	.	.	.	.	.	.
(4) *TOPIC_EMISSION*	-.40	-.01	-.05	1	.	.	.	.	.	.	.	.	.
(5) *ESRS*	.06	.10	-.02	.02	1	.	.	.	.	.	.	.	.
(6) *DISPERSION*	.33	.12	.06	-.43	.06	1	.	.	.	.	.	.	.
(7) *LENGTH*	-.02	-.03	-.01	.00	-.01	-.08	1	.	.	.	.	.	.
(8) *DIFFICULT*	.08	.06	.03	-.01	.04	.05	-.16	1	.	.	.	.	.
(9) *TONE*	-.02	.11	-.04	.22	.03	-.07	-.06	.03	1	.	.	.	.
(10) *COMPLEX*	.09	-.01	.01	-.13	.01	.09	.08	.05	-.05	1	.	.	.
(11) *UNCERTAIN*	.01	-.01	.00	.00	.03	-.01	.21	.01	.04	.11	1	.	.
(12) *TABLES*	-.01	.06	-.02	.03	.00	-.06	-.01	.06	.18	-.05	-.11	1	.
(13) *IMAGES*	.00	.00	.00	.01	-.01	.02	-.01	.00	.00	.00	.00	.04	1

Notes: This table presents pairwise Pearson correlation coefficients. All variables are defined in Appendix A.

Table 3. Experience and information acquisition

	Log(*TIME*)					
	(1)	(2)	(3)	(4)	(5)	(6)
SESSION	-0.004***	-0.005***	-0.006***	-0.005***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
EXIST	0.017***	0.015***	0.013***	0.012***	0.012***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ESRS	0.046	0.032	-0.028	0.010	-0.005	0.013
	(0.031)	(0.033)	(0.024)	(0.043)	(0.042)	(0.043)
EXPERIENCE (EXP.)				-0.029	-0.017	-0.035
				(0.019)	(0.018)	(0.021)
SESS_EARLY					0.145***	
					(0.037)	
SESS_EARLY × EXP.					-0.040***	
					(0.015)	
SESS_LATE						-0.096**
						(0.045)
SESS_LATE × EXP.						0.021
						(0.016)
Topic FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm and year FE	-	Yes	Yes	Yes	Yes	Yes
Analyst FE	-	-	Yes	-	-	-
Adj. R^2	0.348	0.372	0.797	0.644	0.661	0.650
SE Cluster	Participant	Participant	Participant	Participant	Participant	Participant
N	3,182	3,182	3,182	1,356	1,356	1,356

Notes: This table presents results from OLS regressions of Log(*TIME*), the natural logarithm of the time a student analyst takes to enter data for one firm-year-topic, on session characteristics and experience. *SESSION* is a variable that increments for every firm-year-topic completed, *EXIST* is a count of actually existing data points, *ESRS* is an indicator variable if the report refers to 'ESRS' or 'CSRD'. *SESS_EARLY* (*SESS_LATE*) refer to sessions in the first (last) tercile of sessions. Standard errors are clustered by participant and included in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix A.

Table 4. Reporting characteristics and information acquisition

	Log(*TIME*)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DISPERSION	0.221***								0.231***
	(0.032)								(0.037)
LENGTH		0.000							0.000
		(0.000)							(0.000)
DIFFICULT			0.008***						0.007***
			(0.002)						(0.002)
TONE				-1.274					-0.827
				(1.000)					(0.985)
COMPLEX					6.132				4.459
					(4.633)				(5.004)
UNCERTAIN						3.449**			2.129
						(1.631)			(1.562)
TABLES							-0.004**		-0.002
							(0.002)		(0.002)
IMAGES								0.000	0.000
								(0.000)	(0.000)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Topic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.420	0.372	0.376	0.372	0.372	0.373	0.374	0.373	0.397
SE Cluster	Particip.	Particip.	Particip.	Particip.	Particip.	Particip.	Particip.	Particip.	Particip.
N	4,235	3,182	3,182	3,182	3,182	3,182	3,190	3,190	3,182

Notes: This table presents results from OLS regressions of Log(*TIME*), the natural logarithm of the time a student analyst takes to enter data for one firm-year-topic on various measures of textual reporting quality. Standard errors are clustered by participant and included in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix A.

Table 5. Descriptive statistics for the experiment

	Mean	SD	Min	P25	P50	P75	Max	N
Panel A: Participant characteristics								
FEMALE	0.337	0.474	0.000	0.000	0.000	1.000	1.000	172
AGE	0.436	0.735	0.000	0.000	0.000	1.000	3.000	172
DEGREE	2.291	1.269	1.000	1.000	3.000	3.000	5.000	172
INVESTING	2.465	1.634	0.000	1.000	3.000	4.000	6.000	172
PRIMSRC	2.192	1.765	0.000	1.000	2.000	3.000	6.000	172
AIUSE	4.791	1.377	0.000	4.000	5.000	6.000	6.000	172
PRIOREXP	-0.060	1.263	-2.422	-0.908	-0.045	0.811	3.350	172
ESGINVEST	3.052	1.652	0.000	2.000	3.000	4.000	6.000	172
ESGPURCHASE	3.453	1.542	0.000	3.000	4.000	4.000	6.000	172
ESGATTITUDE	0.008	1.390	-3.107	-0.674	-0.174	1.166	2.621	172
Panel B: Information acquisition variables								
FINISHED	20.500	7.727	5.183	14.502	21.555	27.056	38.068	297
CORRECT	1.966	0.865	0.000	1.000	2.000	3.000	3.000	297
SCORE	2.149	1.249	0.000	1.242	2.093	2.982	5.202	297
ROUND	1.579	0.713	1.000	1.000	1.000	2.000	3.000	297

Notes: This table provides descriptive statistics on the experimental participants (Panel A) and the variables used to characterize their information processing (Panel B). Panel B is on a participant-round level. All variables are defined in Appendix A.

Table 6. Results per experimental group

Panel A: Raw results per group

Experimental groups		*FINISHED*	*CORRECT*	*SCORE*	N
CSRD	*AI*				
No	No	20.235 (0.9012)	1.8649 (0.1006)	2.0560 (0.1467)	74
No	Yes	20.822 (0.8492)	1.8193 (0.0941)	1.9493 (0.1311)	83
Yes	No	19.318 (0.9093)	2.0656 (0.1212)	2.4162 (0.1632)	61
Yes	Yes	21.321 (0.9139)	2.1392 (0.0878)	2.2414 (0.1425)	79

Panel B: Two-way Analysis of Variance (ANOVA) results

Source of variation	SS	*df*	MS	*F*-stat	*p*-value
AI	0.03	1	0.029	0.039	0.8437
CSRD	5.23	1	5.230	7.090	0.0082
AI × *CSRD*	0.26	1	0.260	0.353	0.5528
Residuals	257.99	345	0.748		

Panel C: Between-group comparisons

Comparison	Difference in means	*df*	*t*-stat	*p*-value
AI use (post-CSRD)	-0.175 (0.212)	293	-0.825	0.4100
AI use (pre-CSRD)	-0.107 (0.199)	293	-0.536	0.5920
Standardization (w/ AI)	0.360 (0.215)	293	1.675	0.0949
Standardization (w/o AI)	0.292 (0.195)	293	1.494	0.1361

Notes: This table presents the results from comparing the experimental groups. Panel A shows means and standard errors of the main experimental variables for the four groups on a participant-industry level. *FINISHED*, the time (in minutes) the participants took to finish the task, *CORRECT*, the number of correct answers they gave (based on the three questions), and *SCORE*, an aggregate score based on the other two variables. Panel B shows the results from a two-way ANOVA, and Panel C from between-group comparisons. All variables are defined in Appendix A.

Table 7. Standardization, AI, and processing

	SCORE	*SCORE*	*SCORE*	*FINISHED*	*CORRECT*
	(1)	(2)	(3)	(4)	(5)
Intercept	2.056***	1.828***			
	(0.147)	(0.161)			
AI	-0.107	-0.074	-0.104	0.506	-0.042
	(0.192)	(0.200)	(0.201)	(1.457)	(0.068)
CSRD	0.360*	0.413**	0.403**	-1.071	0.134*
	(0.204)	(0.206)	(0.204)	(1.412)	(0.078)
AI × CSRD	-0.068	-0.122	-0.106	1.568	-0.006
	(0.287)	(0.299)	(0.298)	(2.076)	(0.106)
ROUND		0.345***	0.336***	-1.024*	0.084**
		(0.095)	(0.092)	(0.609)	(0.039)
Industry FE	-	-	Yes	Yes	Yes
Adj. R^2	0.009	0.045	0.062	-0.003	0.031
SE Cluster	Participant	Participant	Participant	Participant	Participant
N	297	297	297	297	297

Notes: This table presents results from OLS regressions of *SCORE*, an aggregate performance measure in columns (1) to (3), as well as of its components, *FINISHED,* the time and *CORRECT*, the performance in the experiment on two indicator variables, *AI* and *CSRD,* delineating the two treatment arms. *ROUND* is the number of the experimental rounds (maximum of three). Standard errors are clustered by participant and included in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix A.

Table 8. Experience, standardization, and processing

	SCORE			
	AI = 1	*AI* = 0	*AI* = 1	*AI* = 0
	(1)	(2)	(3)	(4)
ROUND	0.122	0.546***	0.123	0.579***
	(0.123)	(0.139)	(0.125)	(0.134)
CSRD			0.294	0.438**
			(0.210)	(0.210)
Industry FE	Yes	Yes	Yes	Yes
Adj. R^2	0.000	0.121	0.009	0.144
SE Cluster	Participant	Participant	Participant	Participant
N	162	135	162	135

Notes: This table presents results from OLS regressions of *SCORE*, an aggregate performance measure based on the time and performance in the experiment on *CSRD,* an indicator for the standardization treatment and *ROUND*, the number of the experimental rounds (maximum of three). Columns (1) and (3) show results for the *AI*-treated group, columns (2) and (4) for the non-*AI* group. Standard errors are clustered by participant and included in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix A.

Table 9. Experience, ESG attitudes, and information processing

	SCORE			
	PRIOREXP		*ESGATTITUDE*	
	= LOW	= HIGH	= LOW	HIGH
	(1)	(2)	(3)	(4)
AI	-0.284	0.054	-0.108	-0.165
	(0.288)	(0.268)	(0.295)	(0.259)
CSRD	0.055	0.616***	0.395	0.239
	(0.286)	(0.249)	(0.302)	(0.291)
AI × *CSRD*	0.240	-0.360	-0.326	0.299
	(0.407)	(0.408)	(0.385)	(0.456)
ROUND	0.420***	0.267*	0.288**	0.397***
	(0.114)	(0.152)	(0.129)	(0.134)
Adj. R^2	0.056	0.040	0.040	0.061
SE Cluster	Participant	Participant	Participant	Participant
N	153	144	167	130

Notes: This table presents results from OLS regressions of *SCORE*, an aggregate performance measure based on the time and performance in the experiment, on two indicator variables, *AI* and *CSRD,* delineating the two treatment arms. *ROUND* is the number of the experimental rounds (maximum of three). Columns (1) and (2) split the sample by the median of *PRIOREXP*, the first factor from a principal component analysis on questions regarding the participant's prior experience. Columns (3) and (4) split the sample on *ESGATTITUDE*. Standard errors are clustered by participant and included in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix A.