BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Can GenAI fill banks' emissions data gaps?

by Cristina Angelico and Enrico Bernardini

March 2026

Number 1003

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Can GenAI fill banks' emissions data gaps?

by Cristina Angelico and Enrico Bernardini

Number 1003 – March 2026

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of Banca d'Italia and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within Banca d'Italia, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6643 (online)

Designed by the Printing and Publishing Division of Banca d'Italia

CAN GENAI FILL BANKS' EMISSIONS DATA GAPS?

by Cristina Angelico* and Enrico Bernardini*

Abstract

This paper presents new evidence highlighting significant limitations in emissions data for major listed banks in the euro area, sourced from four leading providers. Notable issues are data gaps, inconsistencies across providers and high volatility over time in scope 3 emissions, which are often inexplicably lower than scope 2 emissions and do not correlate with the banks' exposure to high-emitting sectors. The paper then examines whether Generative Artificial Intelligence (GenAI), which relies on broader and diverse information sets, can help bridge the existing data gaps by comparing the outcomes from three GenAI tools. We find that GenAI-based emissions data, either estimated or retrieved, are correlated with data from traditional sources and may therefore help to partially fill current gaps and identify anomalies in the available data. Nevertheless, they suffer from similar issues in terms of quality and consistency to those documented for the data supplied by professional providers. Additional concerns regard replicability and transparency, underscoring the need for initial caution in their use. Despite the current limitations, looking forward, GenAI may become a valuable complementary data source as models that are fine-tuned for this task are developed. Future improvements will also depend on the availability of more reliable underlying information, which in turn requires parallel progress in defining simple, clear and actionable measurement standards as well as regulatory developments to promote climate disclosure by non-financial corporations within banks' portfolios.

JEL Classification: C8, G21, Q54, G32.

Keywords: scope 3 emissions, GenAI, banks' carbon emissions, climate-related risk.

DOI: 10.32057/0.QEF.2026.1003

---

* Banca d'Italia, Secretariat to the Governing Board, Climate change and Sustainability Hub.

1 Introduction[1]

Even amid the recent geopolitical fragmentation and the uncertainty surrounding climate policies globally, credible climate-related disclosure remains critical to allow a comprehensive understanding of potential climate-related risks and opportunities for companies of any size (Angelini (2023)), market participants, and policy makers, thus facilitating informed and strategic decision making (ECB (2025)). High-quality climate-related information is essential for adequately identifying, assessing, and managing climate-related financial risks, as well as facilitating market pricing, guiding investors and credit institutions in their risk management activities and investment decisions, enabling prudential supervision activities, and tracking progress toward the green transition.

Greenhouse gas (GHG) emissions data are key to these aims. Despite their relevance and the work of the GHG Protocol to standardize their measurement (GHG-Protocol (2013)), there are still notable challenges regarding the availability, consistency and quality of emissions data due to ambiguities in calculation methodologies and opaque representation, especially for the scope 3 emissions, i.e., the indirect emissions related to upstream and downstream supply chain (Papadopoulos (2022); Bingler et al. (2022); Nguyen et al. (2023); Talbot & Boiral (2018); Busch et al. (2022)). At the same time, the ongoing shift in climate policy winds in several jurisdictions and related legislative initiatives (i.e., the Omnibus package in the European Union), risk to hinder the sustainability-disclosure trajectory going well beyond the desirable goal of simplifying the regulatory framework (Angelini (2025)) and soften the disclosure burden, thus further limiting the availability of high-quality climate-related data.

In this framework, this work seeks to shed light on whether and how Generative Artificial Intelligence (GenAI) tools may be a cost-efficient complementary source of emissions data and, in particular, contribute to improving the availability and consistency of emissions data of the European financial system, where banks play a central role in financial intermediation.

We focus on banks' scope 3 emissions (i.e. those emissions related to their credit and investment portfolios) for two reasons. First, they are particularly relevant for investors, researchers, and policymakers to evaluate banks' climate risk exposure and the ability to harness the green transition opportunities. For these purposes, despite their limitations, it is necessary to employ scope 3 emissions since banks' scope 1 (i.e. direct) and scope 2 (i.e. indirect, stemming from the company's energy consumption) emissions refer only to internal operations and do not provide insights into financed activities.[2] Second, they represent an interesting and particularly complex example, as banks face additional challenges when computing their scope 3 emissions compared to non-financial corporations. Indeed, beyond the emissions related to their company's value chain (i.e., purchased goods and services, or business travel, etc.), they also need to assess emissions concerning their credit and investment portfolios, ideally by gathering data related to their non-financial counterparts (i.e., borrowers and security issuers) or otherwise exploiting sectoral emissions data as a proxy.

The ability of GenAI tools to access, extract and elaborate information from several sources, including texts, makes it worthwhile to explore whether they can help fill the data gap in banks' emissions by supplying comparable metrics, also benefiting from the continuous interaction with the user, which may help in data quality assessment. Indeed, with their powerful capabilities in extracting and elaborating vast amounts of unstructured and structured data, possibly broader and more diverse than the typical data sets used by professionals, GenAI tools may be able to overcome the significant efforts that analysts, researchers, and policy makers need to pay to recover climate-related information - often embedded in text, tables or figures, and fragmented in multiple reports, websites and other sources - and verify the quality of the data sourced by the private providers. At

[1]The views expressed herein are those of the authors and do not necessarily represent the views of the Bank of Italy. We thank Paolo Angelini, Ivan Faiella, Patrizio Pagano, Luigi Bellomarini, Carolina Camassa, Andrea Coletta and Aldo Glielmo for their useful comments. All remaining errors are our own.

[2]Scope 1 and 2 emissions of non-financial corporations, instead, are well informative on their climate performance as they reflect energy consumption related to their business.

the same time, despite their potential, the scant availability of reliable information, likely used as input by such tools, may limit their ability to fill the current data gap.

Our analysis builds on new evidence that scope 3 emissions of listed banks in the Euro area display several unexpected and undesirable features: they are often missing also for large and medium listed banks, they differ across providers, are highly volatile over time, usually inexplicably lower than scope 2, and unrelated to the banks' exposure to high-emitting sectors. Such undesirable features and the low reliability of available banks' emissions data hinder their use for research and policy purposes at the current stage. Against this framework, we test whether widely available GenAI tools can be a novel and valuable source of information, and whether they have the potential to revolutionise research also in this area, beyond other domains, as discussed by Korinek (2023), and provide evidence on the first-in-its-kind application to fill the documented data gap and improve data quality.

Using three alternative GenAI tools (Claude, ChatGPT, and Gemini), we first generate emissions data for our sample of listed Euro area banks for 2022 by asking the tools to retrieve or estimate the data and testing alternative models and features of each tool. We then analyse and compare the outcomes generated. Notably, given the vast number of available models and the rapid pace of new releases, our goal is not to test or rank the latest models, nor to develop an advanced fine-tuned model for our task, but rather to provide an overview of the performance of a selection of models - easily accessible via subscription to users such as climate-risk analyst and researchers - with varying capabilities and characteristics in terms of data retrieval, web search, or mathematical reasoning.

We show that the Large Language Models (LLMs) underlying the three different GenAI tools provide different answers to the same questions, supplying data with varying characteristics, and only in some cases correlated with one another. Some of these LLMs provide information that correlates with that of professional providers, classified as self-reported by the banks or internally estimated. They therefore might help partially fill gaps and identify anomalies in the available data. At the same time, GenAI-based data exhibit data quality issues similar to those documented for the data supplied by the professional providers (both self-reported by the banks or estimated), suggesting that the latter are likely present in the set of information on which the LLMs were calibrated or that professional providers and LLMs employ similar sources. However, the interaction enabled by LLMs makes it possible to correct certain inconsistencies in the data more quickly than the providers' services. Further notes of caution concern the replicability and transparency of processes, given the heterogeneity of outputs produced across similar requests — by the same or different models —, the output sensitivity to the computational effort required by the task, errors in the classification of data (retrieved or estimated), and the continuous, rapid evolution of LLMs. Despite these limitations, which call for caution, the results are encouraging, considering that the GenAI-based data were generated by widely available tools using prompts that did not provide any specific information on the banks' credit and sustainability policies or characteristics.

We conclude that, for now, particular attention is needed when using banks' scope 3 data sourced from professional providers as well as GenAI tools. However, although caution is required at this stage, GenAI may become a valuable complement in the future to enhance the robustness and availability of data as more high-quality climate-related information becomes available, which LLMs and users can leverage, and models specifically fine-tuned for this aim will be developed [3].

Improving and expanding standardised emissions disclosure, both for financial institutions and for non-financial firms included in their portfolios, may allow to enhance clarity and consistency, thereby reducing challenges for users and, at the same time, enabling cost-efficient and innovative techniques. Despite the current headwinds, regulatory initiatives in the European Union (i.e. the

[3] Several studies provide evidence that fine-tuned models exhibit superior adaptability and performance, compared with general-purpose LLMs, particularly in terms of computational efficiency and result accuracy. Fine-tuning may involve training LLMs on task-specific data sets or applying specialized learning techniques, such as Reinforcement Learning from Human Feedback (RLHF) and retrieval-augmented generation (RAG). See for instance Wu et al. (2023), Yang et al. (2025), Rasul et al. (2024).

Omnibus package) and other countries should be consistent in this direction.

The remainder of the paper is organised as follows. Section 2 describes the data. Section 3 shows the motivating evidence related to the low data coverage and emissions data issues. Section 4 reports and discusses the results of the test of GenAI tools. Finally, Section 5 contains concluding remarks.

2 Data

The paper primarily focuses on scope 3 emissions, which, according to the GHG Protocol, for banks and other financial intermediaries, include those related to loans and investment portfolios, i.e., financed emissions (downstream type, classified as category 15 emissions). These represent the bulk of total banks' emissions , e.g. 86% for our bank sample in 2022 (Figure A1 in Appendix 1).

The bank sample comprises 129 medium- and large-sized listed banks active in the Euro area, for which we have relatively good coverage of emissions data, which is not available for unlisted corporations.

The analysis utilises three sets of data: emissions data sourced from major data providers, loan data from Anacredit, and GenAI-based emissions data. Emissions data refer to the period from 2018 to 2023 and are gathered by four well-known private data providers: LSEG Datastream, ISS, MSCI ESG Research, and Bloomberg. Data coverage varies between emission types and sources, as shown in Section 3 and in Appendix 1. Loan data, at the sector and bank- level, are sourced from Anacredit for the period 2018-2023 and focus on high-carbon sectors, which are key for the transition to a low-carbon economy according to the International Energy Agency (IEA (2023)); these sectors are manufacturing, energy, transport, and mining.[4] Finally, we get GenAI-based banks' scope 3 emissions for our sample of banks for 2022 using three alternative tools: Claude developed by Antropich, ChatGPT by OpenAI and Gemini by Google. More details on the prompts and the data supplied by these alternative tools are provided in Section 4.

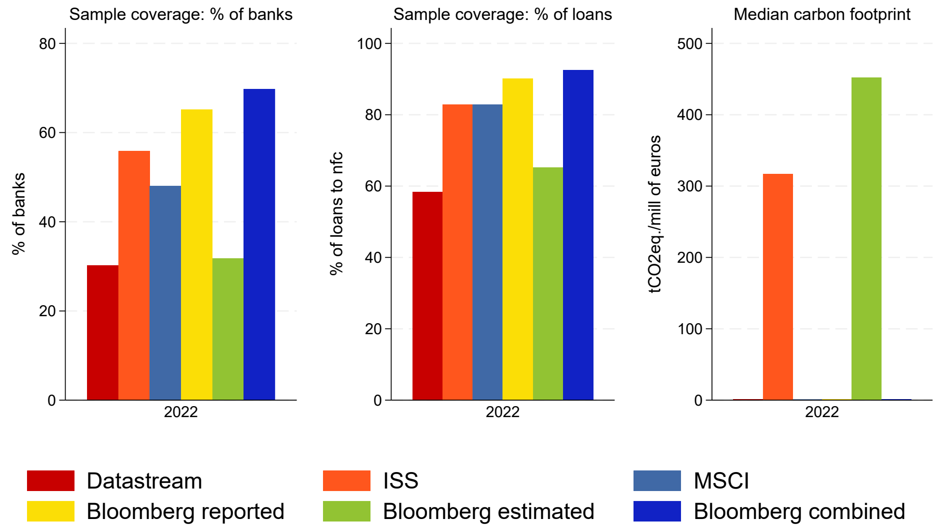
3 Motivation: missing data and data anomalies

Our exercise on the GenAI-based data is mainly motivated by the following undesirable features and anomalies in the scope 3 emissions data supplied by professional providers for our sample of banks. First, the data coverage for scope 3 emissions for the bank sample varies across sources and is often limited, ranging between 32 and 69 % of the sample (Figure 1).[5]

[4] AnaCredit is the Euro area credit register which collects at the monthly frequency the amount and other details of any loan granted to corporations by Euro area credit institutions that exceeds the threshold of 25,000 euros.

[5] Despite their relevance, downstream scope 3 emissions (i.e. those that are only related to the portfolios) is limited (Figure A2 in Appendix 1).

Figure 1: Banks' scope 3 emissions: data coverage and carbon footprint for the 2022 by source



Note: For each bank, carbon (or emission) footprint is computed as the ratio between the bank's scope 3 emissions and its total loans to non-financial corporations (nfc) as of the end of 2022.

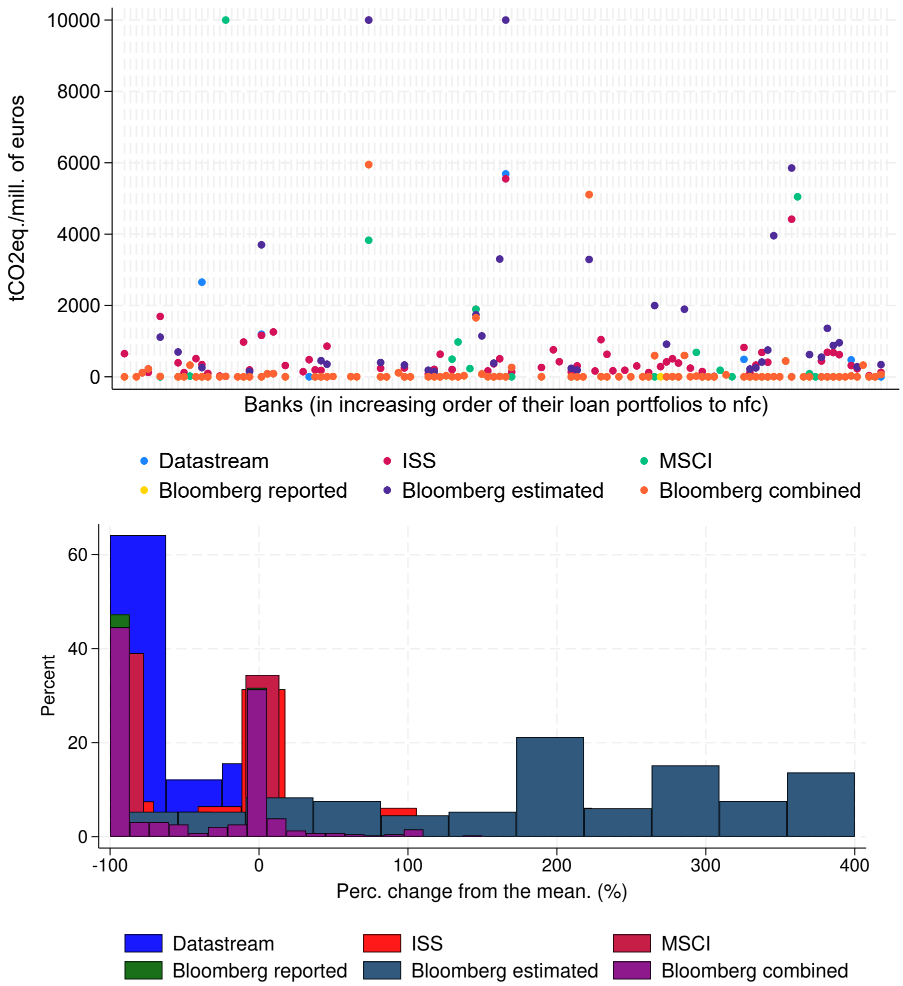
This fact is surprising, given that the sample includes only medium- and large-listed Euro area banks, most of which, as of the date of reference (end of 2022), were already subject to disclosure requirements, while others were expected to comply with the forthcoming Corporate Sustainability Reporting Directive (CSRD). Looking forward, the current regulatory uncertainty regarding the scope and timing of the CSRD's application will further hinder carbon disclosure by firms and banks. Nevertheless, data are broadly available for the largest banks, which cover a substantial share of the loans to non-financial corporations granted by our sample and also supply more credit to high-emitting sectors (Figure A3 in Appendix 1).[6]

Second, bank-level emissions display a wide heterogeneity between data sources, as already evidenced by the literature for non-financial firms (Figure 2, upper panel). Some heterogeneity is expected given the different methodologies employed by various providers to compute the emissions data, which usually rely on either self-reported information from credit institutions, when obtainable from their reports, or other sources such as the Climate Disclosure Project (CDP), or, otherwise, estimate data using internal models. Nevertheless, it is striking to observe such remarkable divergences in banks' scope 3 emissions, with some providers often exceeding 100% of the average, especially for estimated data (Figure 2, lower panel).[7]

[6]Such largest banks for which data are available do not have credit portfolios more concentrated towards high-emitting sectors. Data availability, indeed, is not related to the portfolio composition.

[7]Notwithstanding the heterogeneity mentioned above, pairwise correlations of scope 3 emissions from different sources between 2018 and 2022 are broadly positive and significant (Table A1 in Appendix 1).

Figure 2: Heterogeneity of scope 3 emissions of individual banks across data sources



Note: Data on banks' emissions by different providers as of 2022. In the upper plot, each dot refers to a bank in the sample. The figure was set with an upward bound for the emissions data for graphic purposes. On the x-axis, banks are sorted according to the size of their credit portfolio to non-financial corporations, so that the first bank on the left grants the smallest amount of loans. Several data points appear to be zero due to the chart's scale; the underlying values are smaller relative to the others but positive. In the lower plot, the figure displays, for each data source, the histograms of the percentage change from the average scope 3 emissions computed as average across all the data sources for each bank and time. Data are bounded to be between -100% and +400%.

The heterogeneity in the data and the sample coverage across alternative sources is also reflected in the aggregate figures: according to the median scope 3 footprint, a loan of 1 million euros finances between 0.9 and 452 tons of CO2eq (Figure 1).[8]

Third, carbon emissions footprints - computed as the ratio between the bank's scope 3 emissions and its total loans to non-financial corporations - are not related with banks' sectoral portfolio composition: the carbon footprints are surprisingly not higher for the banks more exposed to high-

[8]Large differences across sources survive if we restrict the sample and compute the median also on a subset of common observations.

emitting sectors such as mining, energy and transport (Table 1).[9]

Table 1: Banks' scope 3 emissions footprint and exposure to high-emitting sectors

	(1)	(2)	(3)	(4)	(5)	(6)
	Scope 3 emissions footprint					
Mining	-158.6	-22846.7	-1900.9	-2.948	-903117.7	-0.861
	(0.165)	(0.295)	(0.416)	(0.959)	(0.315)	(0.987)
Manufacturing	-18.68*	-3214.6	-1321.0	-40.76*	-39226.9	-34.68*
	(0.053)	(0.249)	(0.273)	(0.077)	(0.400)	(0.083)
Energy	-41.51	-2354.8	-1435.5	-34.34	-348193.6	-30.35
	(0.307)	(0.374)	(0.319)	(0.155)	(0.323)	(0.163)
Transport	18.36	8433.6	-1147.0	79.61	298677.6	80.22
	(0.261)	(0.142)	(0.204)	(0.276)	(0.332)	(0.273)
R^2	0.132	0.040	0.027	0.119	0.089	0.106
Obs.	173	293	302	364	132	384

Note: Results from a set of regressions where the dependent variables are the bank's scope 3 emissions (or carbon) footprint computed as the ratio between its scope 3 emissions and its total loans to non-financial corporations. The independent variables are the share of the bank's loans to non-financial corporations granted to high-emitting sectors (mining, manufacturing, energy and transport). All the regressions include the constant and year fixed effects. Each column refers to a data source: (1) is Datastream, (2) ISS, (3) MSCI, (4) Bloomberg reported, (5) Bloomberg estimated, (6) Bloomberg combined. Errors are clustered at the bank level. P-values in parentheses. * = $p < 0.10$, ** = $p < 0.05$, *** = $p < 0.01$.

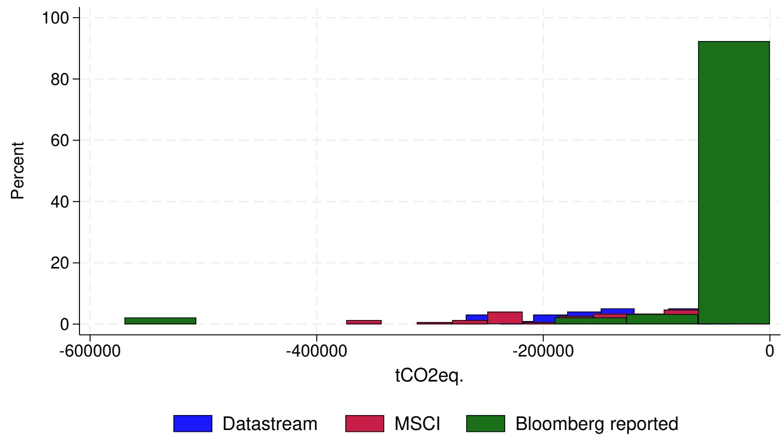
Furthermore, analysing bank-level data by sources, we find several cases where scope 2 emissions are larger than scope 3. This evidence sounds conceptually inexplicable as scope 3 emissions include the whole financial intermediation activity rather than just banks' own operations (as for scope 1 and 2).[10] The data differences for these anomalies are significant in magnitude, with scope 3 emissions being lower than scope 2 emissions by 60% on average (with a range between 1% and 100%, Figure 3). The relatively high frequency of this pattern suggests that it might be due to misreporting or underestimation of scope 3 emissions, rather than other reasons, such as the presence in our sample of some banks operating high-emitting data centres.

Finally, scope 3 emissions data are highly volatile over time, for all data sources. Banks' scope 3 absolute emissions and footprints vary substantially over the years, often increasing or decreasing by more than 100% (Figure 4). Such substantial changes occur over time for each bank and source, with fluctuations that often alternate between jumps and drops.

[9] Similarly, the differences in banks' exposure to the high-emitting sectors do not explain the heterogeneity of scope 3 footprints between sources (Table A2 in Appendix 1).

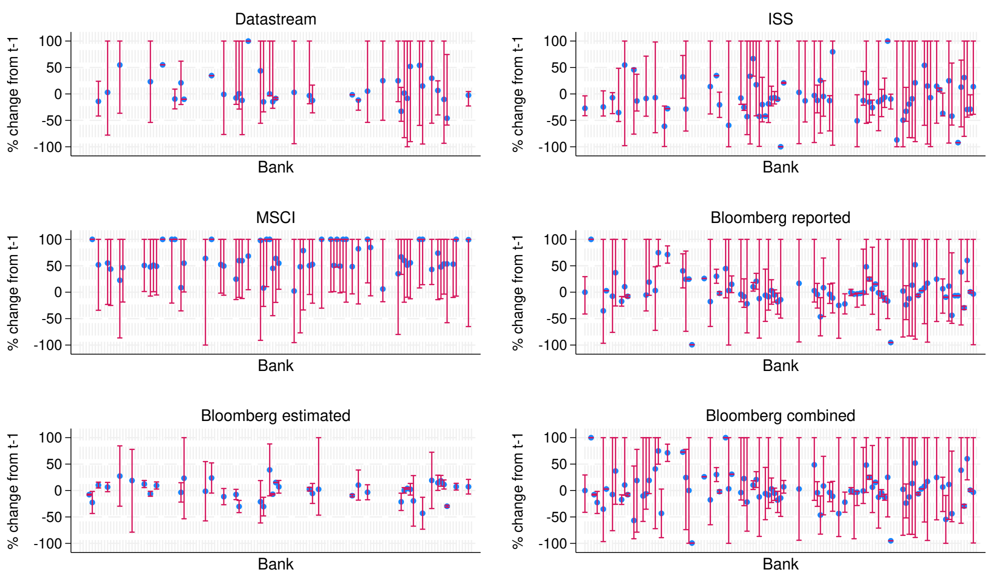
[10] This feature occurs for about 50% of bank sub-sample covered by Datastream and MSCI and 27% covered by Bloomberg reported data. It never occurs with the data estimated by Bloomberg nor with the remaining sources.

Figure 3: Data anomalies: negative differences between scope 2 and 3 emissions



Note: The figures plots the histograms of the differences between scope 2 and 3 emissions - computed for each bank and data source - when the differences are negative, by data source in 2022.

Figure 4: Data anomalies: bank-level annual percentage change in scope 3 emissions footprint



Note: The figures display the annual percentage change from the previous year in scope 3 emissions footprint reported for each bank by data sources. Each figure refers to a data source, while each dot/line refers to a bank. The blue dot is the median value over the spanned period, while the red bars display the minimum and the maximum. Outliers are not reported, and data are bounded to be between -100 and 100%. On the x-axis, banks are sorted according to their loans to non-financial corporations, so the first bank has the smallest total loans, while the last has the largest amount. The bank's scope 3 emissions (or carbon) footprint is computed as the ratio between its scope 3 emissions and its total loans to non-financial corporations.

The variability might be explained by revisions in the methodology employed or in the emission perimeter considered by reporting banks or providers when estimating the emissions. In particular, a methodological change could explain why, in a few years, revisions by a given provider tend to move in the same direction for all banks. If we set lower and upper bounds for the data at -100% and 100% respectively, the average year-on-year percentage change of bank-level scope 3 emissions footprint ranges between 0.4 to 50% depending on the data source (Figure A4 in Appendix 1).[11] However, a significant heterogeneity is observed within each source and over time.

4 Assessing GenAI-based emissions data

Given the documented undesirable features of the currently available scope 3 emissions data for the Euro area banking system and the skyrocketing use of GenAI tools for an increasing range of applications, we test whether GenAI could be a novel and valuable alternative source of emissions data for the banking sector to fill the existing data gap and check data anomalies. Previous works used Machine Learning techniques to estimate corporate emissions (i.e., Assael et al. (2022); Nguyen et al. (2021)) or LLMs to evaluate banks' environmental disclosure (Ángel Iván Moreno & Caminero (2023); Giannetti et al. (2023)) and financial institutions' climate-related indicators (BIS (2024)) or to recover scope 3 emissions for non-financial corporations based on transaction data (Jain et al. (2023)).[12] Here, instead, we test whether it is possible to retrieve banks' self-reported scope 3 emissions or estimate them using three alternative LLMs: Claude, ChatGPT, and Gemini. The type of data (if retrieved or estimated), coverage, and explanations differ across GenAI models and runs. Notably, given the increasing number of available LLMs and their rapid development, our goal is not to test or rank the latest models, but rather to provide an overview of the performance of a select group of tools that are easily accessible to climate data users, such as risk analysts, policymakers, and researchers. The tools we employ vary in their capabilities and characteristics, particularly in terms of data retrieval, web search, and mathematical reasoning. The data was collected between February and July 2025.

First, we asked Claude, an LLM developed by Anthropic, to retrieve scope 3 emissions data for the list of banks within our sample for 2022.[13] To evaluate the robustness of the results, we employed two different models developed by Anthropic.[14] In both cases, the *Web search* option was not enabled, as for Claude, the system does not automatically search online if not explicitly authorised.[15] To our request, Claude 3.5 (without the *Web search* option) provided us with a table displaying the relevant data (hereafter, *Claude 3.5 - estimated*) by clarifying that the supplied data was not extracted from the banks' sustainability reports. Still, it was estimated to be representative of typical emissions for these institutions and should be verified against the banks' most recent official reports (see the Appendix for more details). Claude 3.7 (without the *Web search* option), instead, replied by providing a table that displayed: 1) retrieved data based on publicly available sustainability reports when available (within its training set), and 2) estimated data, mainly for the smaller banks. Hereafter, we will refer to these data as *Claude 3.7 - retrieved & estimated*. Since the *Web search* option was not enabled, Claude did not retrieve the information classified as selfreported by the banks on the web, but only from its training sample; this, therefore, suggests that the latter includes explicitly banks' sustainability reports and/or climate-related information.

[11]We replace the values greater than 100 with 100 and those lower that -100 with -100 to bound the percentage changes.

[12]See Jain et al. (2023) for more information on other research exploring the potential of LLMs in the domain of climate and sustainability. Bhuiyan (2024) instead provides some case studies on how AI can help companies measure and mitigate their own emissions. Relative to our work, BIS (2024) provides an example of how LLMs can extract climate-related information (i.e., scope 1 emissions) from a closed set of sustainability reports, while we assess more broadly the availability of widely available LLMs to retrieve or estimate scope 3 emissions data.

[13]Prompt: *Can you please recover the scope 1, 2 and 3 emissions of the following banks in 2022? Can you please save the data in an Excel file (that I can export) where there is a row for each bank and 3 columns, one for scope 1, one for scope 2 and one for scope 3?*

[14]Claude 3.5 is defined as the quickest model among those by Anthropic and adequate for most of the tasks, and Claude 3.7, instead, is defined as the most intelligent model and more suitable for mathematical tasks.

[15]Without explicitly enabling the *Web search* option, we did not allow Claude to search on the web for the information requested. For this reason, it only exploited information from its training set.

We then requested the same question to ChatGPT, an LLM supplied by OpenAI, which provided us with publicly available data reported by banks and retrieved from their reports, CDP disclosure or other sources (hereafter, *ChatGPT - retrieved*). Further, when explicitly asked,[16] it also provided estimated data (hereafter, *ChatGPT - retrieved & estimated*) for our bank sample.[17] In these cases, the *Web search* option was enabled as the system automatically recognised the need to search online. To obtain the requested information, we needed to iterate the request multiple times, as a large number of missing values were provided when we supplied the full list of banks. This fact occurred because retrieving the data for the entire sample from the web was evaluated as a highly intensive task by ChatGPT.[18] We thus run the request for small subsets of our samples separately, by using the same prompt each time. Additionally, we requested frequent revisions when we encountered unexpected values or missing data.[19] Although the prompt remained consistent across different queries, the outcomes varied depending on the length of the list provided, which suggests potential issues with data replicability and consistency.[20] To address these issues and verify whether they were due to the limit imposed on the number of messages users can freely send using ChatGPT, we employed ChatGPT Plus 4.o to recover the scope 3 emissions for our sample of banks in 2022. Nevertheless, since we encountered very similar replicability challenges related to the length of the list of banks supplied, we asked ChatGPT Plus 4.o to estimate the emissions for the entire list (hereafter, *ChatGPT Plus - estimated*).

Third, we launched the same prompt on Gemini 2.5 Pro and Gemini Deep Search (with 2.5 Pro), two LLMs developed by Google. The replies included publicly available data retrieved from over 400 websites (hereafter, *Gemini 2.5 Pro - retrieved* and *Gemini Deep Search - retrieved* respectively). We then requested an estimate of the emission data for the other LLMs. To this request, Gemini 2.5 Pro replied by providing the estimated data for the missing observations (hereafter, *Gemini 2.5 Pro - retrieved & estimated)*.[21] For more information on the prompts and responses, please refer to Appendix 2.

From the analysis of the bank-level data supplied by the alternative GenAI tools, two main findings stand out. First, the GenAI-based data presents issues and challenges similar to those of other sources analysed in the previous sections of the paper, suggesting that they rely on similar underlying data or that traditional providers may serve as the foundation upon which GenAI tools are trained. The sample coverage varies across runs and models. While the retrieved scope 3 emissions are available only for a limited number of banks, mainly large institutions that account for between 40 and 80% of the loans to non-financial corporations within our sample (Figure 5), estimated emissions are available for the entire (or most of the) sample.[22]

As documented for the traditional data providers, bank-level data vary substantially across models and runs (Figure 6). Such differences are also found among data classified as self-reported by the bank (i.e. retrieved), suggesting that the models may differ in the set of documents assessed or in the way the texts are processed. Given the heterogeneity in the coverage and bank-level data, the median carbon footprints differ across alternative runs and models (Figure 5).[23]

[16] Prompt: *Can you please estimate the scope 1, 2 and 3 emissions of the following banks in 2022?*

[17] The estimated data are mainly based on size, regional footprints and bank characteristics. Although we requested estimates for all the banks in the sample, in a few cases, ChatGPT exploited the self-reported data retrieved by the banks' reports when available.

[18] Please note that we encountered similar issues when trying to obtain the retrieved data from Claude 3.7 once we enabled the *Web search* option.

[19] In cases of unexpected values or missing data, we did not alter the prompt; we re-launched it.

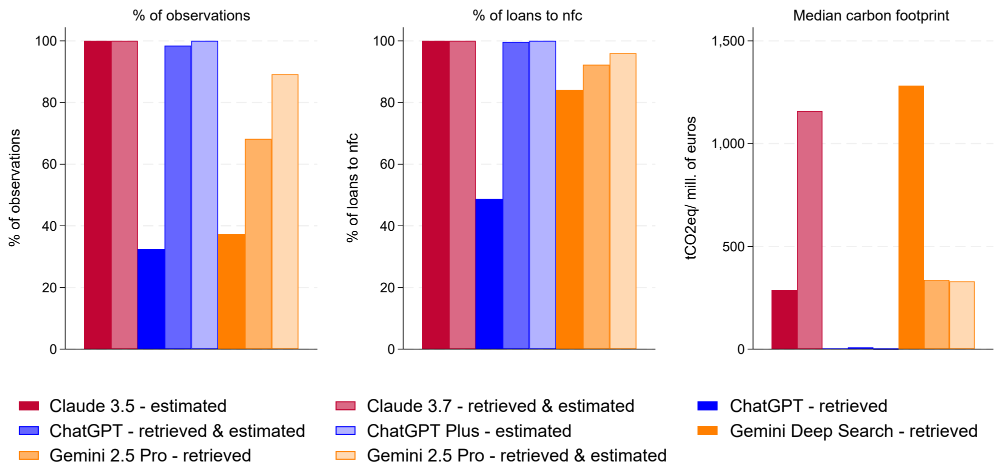
[20] Examining the banks in our samples individually or in different subsets may yield data of different, and possibly higher, quality.

[21] Although explicitly asked to estimate all the data, Gemini exploited the retrieved data and estimated only the missing observations. Besides, even if expressly requested multiple times, it was not possible to get any estimated data with Gemini Deep Search.

[22] Very similar figures are available for scope 1 and 2.

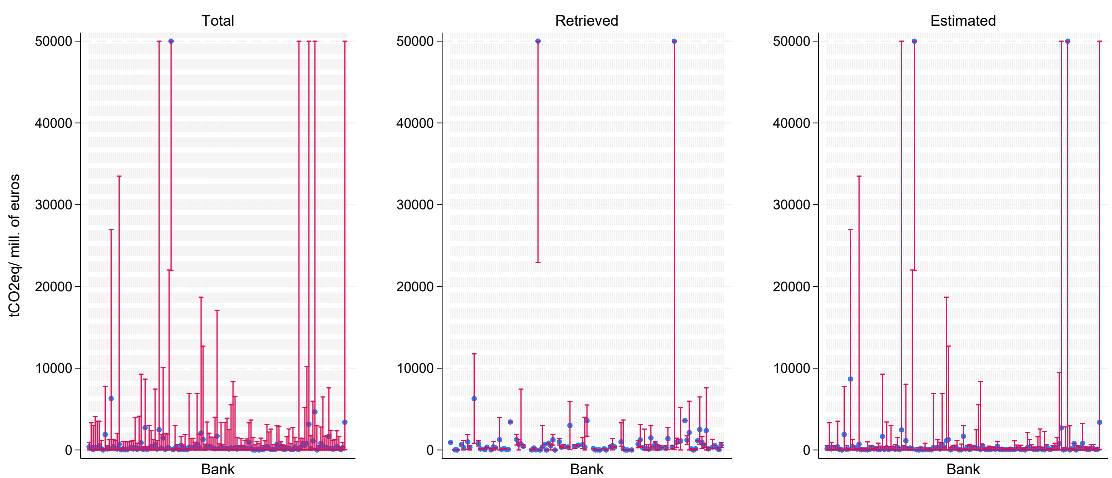
[23] The median values are substantially lower than the average footprints, highlighting the presence of outliers, and reach much larger values than those observed with the traditional data provides.

Figure 5: GenAI-based scope 3 emissions: data coverage and carbon footprint for the 2022



Note: For each bank, carbon footprint is computed as the ratio between the bank's GenAI-emissions and the total loans to non-financial corporations (nfc) as of the end of 2022.

Figure 6: GenAI-based scope 3 emissions footprint by bank for the 2022



Note: The figures display box-plots of the median GenAI-based scope 3 emissions footprint for each bank (blue dot) and their minimum and maximum values (red bars). Each dot refer to one bank, sorted on the x-axis in ascending order by credit portfolio size, with the first one referring to the bank with the smallest credit portfolio in the sample. For graphical representation purposes, outliers are bounded so that both the maximum and the median values are limited to be lower or equal than 50000 tCO2eq/mill. of euros. Emission (or carbon) footprint is computed as the ratio between the bank's GenAI-emissions and its total loans to non-financial corporations (nfc) as of the end of 2022.

The vast heterogeneity in the bank-level data is confirmed by the often null correlations between the emissions supplied by different models and runs (Table 2). A virtually null correlation, indeed, links GenAI-based emissions retrieved by ChatGPT and Gemini (i.e. rows f-g and column c) and a slightly negative correlation, although not significant, ties the estimated emissions, i.e. supplied by Claude 3.5 and ChatGPT (row f and column a). Data supplied by Gemini in the different runs are always highly correlated.

Table 2: Correlation between GenAI-based scope 3 emissions data and other data sources

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	Average
				Across GenAI-based emissions					**0.177**
(b)	0.905***								
(c)	0.071	0.074							
(d)	0.161*	0.140	-0.001						
(e)	-0.043	-0.059	-0.068	0.052					
(f)	0.883***	0.873***	0.021	0.183	-0.102				
(g)	0.898***	0.857***	-0.043	-0.212**	-0.075	0.8602***	0.8602***		
(h)	-0.8305***	-0737***	-0.083	-0.103	-0.042	-0.761***	0.761***	0.8744***	
				With traditional data providers					**0.166**
(1)	-0.097	-0.101	-0.2872	-0.109	0.264*	-0.152	-0.152	0.041	
(2)	0.667***	0.72***	-0.021	0.426***	-0.094	0.685***	0.685***	0.7522***	
(3)	-0.045	-0.037	-0.136	-0.0213	0.006	-0.075	-0.075	-0.0468	
(4)	0.143	0.313***	-0.099	0.036	0.0295	0.249*	0.249*	0.197	
(5)	0.464***	0.539***	0.287	0.1742	-0.0967	0.525***	0.5254**	0.474***	
(6)	0.145	0.315***	-0.056	0.0397	0.0308	0.251*	0.251*	0.200*	

Note: Correlations using bank-level emission data as of 2022 per each data provider vis-à-vis GenAI-based emissions, under the different runs. In the table (a) refers to Claude 3.5 - estimated, (b) Claude 3.7 - retrieved & estimated (c) ChatGPT - retrieved, (d) ChatGPT - retrieved & estimated, (e) ChatGPT Plus - estimated, (f) Gemini Deep Search - retrieved, (g) Gemini 2.5 Pro - retrieved, (h) Gemini 2.5 Pro - retrieved & estimated (1) Datastream, (2) ISS, (3) MSCI, (4) Bloomberg reported, (5) Bloomberg estimated and (6) Bloomberg combined.

Notably, instead, data provided by ChatGPT under different runs (and models) are also not correlated with each other, possibly because the information provided by ChatGPT was often revised under different runs (with similar prompts) and varied depending on the length of the list of banks supplied and, therefore, the computational effort requested. This result highlights the issues of reliability and replicability of GenAI tools' outcomes, which can substantially differ depending on the tool selected, the formulation of the prompt, and the timing of the query, as well as the frequent updates of the underlying models and their computational limits.

On the other hand, surprisingly, data retrieved by Gemini Deep Search from the banks' reports or websites are highly correlated with the data estimated by Claude (row f and column a), suggesting possible misclassifications or inaccurate descriptions of the process used to obtain the data.

Similar to the findings for traditional data sources, banks' carbon footprints based on GenAI are not correlated with their exposure to high-emitting sectors (Table A3 in Appendix 3).

Moreover, GenAI-based data also suffer from the above-mentioned anomalies in the ratio between the scope 2 and 3 emissions, although to a lesser extent than the providers' data. For instance, within the data estimated by Claude 3.7, the scope 2 emissions were larger than the scope 3 emissions for about 20% of the sample. However, once a question was raised about these data anomalies, the tool recognised the mistake and corrected the data.[24] A few anomalies were also found in the data (retrieved or estimated) by ChatGPT and, in some cases, corrected afterwards. Some issues were also found in the data estimated by Gemini 2.5 Pro, but never in its retrieved data.[25] In sum, despite the initial similar anomalies, GenAI tools seem to be able to self-correct, thanks to the quick interaction with the users, which is granted by the tools.

The second relevant finding is that GenAI-based emissions are significantly and positively correlated with data of some providers, although the correlation varies according to the tool and the type of data considered (either retrieved or estimated, Table 2). Scope 3 emissions estimated by Claude are highly and significantly correlated with those of providers that rely mainly or entirely on estimated values (i.e., Bloomberg and ISS), suggesting possible common drivers in the estimation models and information sets used by some GenAI tools and data providers.[26] This, instead, is not the case for the data estimated by ChatGPT, pointing once again to the fact that not all LLMs are the same and relevant differences in their outcomes can be in place.[27]

For the reported emissions, data retrieved by Gemini and Claude from the banks' reports and websites are positively correlated with those self-reported by the banks according to Bloomberg (row 5 and columns a and f).[28] However, this is not true for the data retrieved and classified as self-reported by the banks according to ChatGPT. This missing correlation possibly reflects the frequent revisions made by ChatGPT to the data. Differences (i.e. between Claude and ChatGPT) may also arise due to the activation (or lack thereof) of the *Web search option*.

These results highlight, on the one hand, the ability of LLMs to retrieve information, elaborate and combine it. On the other hand, they also reveal differences in the outcomes of alternative models and highlight relevant difficulties in recovering detailed information from a large number of long and complex documents, which lead to the risk of misclassification and potential mistakes. Despite these documented limitations and areas for improvement, it is encouraging that some of the tested LLMs

[24] Claude clarified that the scope 3 emissions, if correctly assessed, should be substantially larger than the scope 2 as they relate to the financed emissions. The errors were primarily due to underestimating the scope 3 emissions for medium and smaller banks, for which only upward scope 3 emissions were available or considered.

[25] Also, several explanations on why scope 2 emissions might be higher than scope 3 emissions were provided, including incomplete scope 3 reporting and extremely energy-intensive activities (i.e. data centres or cryptocurrency mining).

[26] See rows 2 and 5, columns a and b.

[27] See, for instance, row 5 and columns d and e.

[28] According to Bloomberg, reported data are available for a larger number of banks than those provided by Gemini and Claude. Furthermore, note that the data retrieved by Gemini and Claude are also correlated with that supplied by ISS, which combines estimated and reported data (rows 5 and columns a and f), and, in particular, the data retrieved by Gemini is surprisingly correlated with that estimated by Bloomberg once again pointing to possible misclassification.

provide meaningful information, either retrieved or estimated, that correlates with the data provided by traditional providers, considering that the request did not include any information on the bank's credit and sustainability policies or characteristics.

In summary, the analysis suggests that at the current stage, GenAI-based emissions data can partially contribute to filling the data gaps and scrutinising data anomalies. In particular, the quick interaction with the users is a unique feature of GenAI, which allows for rapid self-correcting. However, at the current stage, GenAI tools cannot fully overcome the documented limitations of traditional data providers, and concerns remain regarding data quality. This result reflects the fact that LLMs are trained on the currently available, scant, and often low-quality data and information. Additional caution is also warranted regarding the replicability of the outcomes, given the heterogeneity among data provided through subsequent runs and similar prompts (i.e., for ChatGPT). Finally, transparency issues pertain to the classification of data (i.e., whether it is retrieved or estimated) and the description of the process used to provide the data.

5 Conclusions

Despite carbon emissions data are key for measuring risks and seizing the opportunities arising from the green transition, enabling effective risk management and timely action, and also to ensure transparency and accountability, allowing supervisory activities (ECB (2025)) and tracking progress toward green transition, there are still several challenges that jeopardize the reliability, thus hindering their use by investors, researchers and institutions.

In this paper, we document several data issues that affect banks' scope 3 emissions, regarding the limited coverage, the heterogeneity among sources, the high volatility of the data for individual banks and providers over time, the anomalies in the scope 2 and 3 ratio and the lack of correlation between banks' scope 3 emissions footprints and their exposure to high-emitting sectors. These findings provide insight into how to frame user-based checks on data quality, particularly for investors and authorities. Ultimately, they underscore the need for more transparency and stability in the estimation methodologies employed by credit institutions and data providers and, as a foundation, for more straightforward and actionable measurement rules embedded in international standards.

The paper then focuses on an innovative application of GenAI tools for retrieving self-reported or estimated carbon emissions by exploiting the ability of GenAI to harness multiple data types and diverse sources.The results suggest that widely available LLMs provide information positively correlated with data of professional providers, which can help to some extent users in filling the data gap; nevertheless, at present, significant concerns related to replicability and data transparency arise, calling for caution and further investigation. In the future, the greater availability of high-quality emissions data, combined with more detailed user requests, additional hard data (i.e., satellite or sensor data), and language models fine-tuned for this specific task (i.e., able to understand technical terminology), may allow us to overcome some of the present limitations.

In summary, these findings corroborate the importance of improving the standardised disclosure of carbon emissions and estimation models, both for financial institutions and the non-financial corporations included in their portfolios. Regulatory developments are at a crossroads, as the Omnibus I package in the European Union could loosen the CSRD-led disclosure requirements, while other countries are implementing ISSB standards. Despite the current headwinds, regulatory initiatives should be consistent to foster climate disclosure, which will also be crucial in enabling cost-efficient, innovative techniques such as generative artificial intelligence.

References

Angelini, P. (2023), 'Le PMI e la transizione climatica e ambientale'. Speech at the Confindustria Conference 'Finance and ESG Disclosure', Rome, September 26, 2023.

URL: *https://www.bancaditalia.it/pubblicazioni/interventi-direttorio/int-dir-2023/Angelini-26.09.2023.pdf*

Angelini, P. (2025), ‘La sostenibilità nell’industria finanziaria: vecchi modelli per nuovi scenari?’. Speech at the conference on ’Sustainability in the Financial Industry: old models for new scenarios?’, Milan, April 2, 2025.
URL: *https://www.bancaditalia.it/pubblicazioni/interventi-direttorio/int-dir-2025/Angelini-02.04.2025.pdf*

Assael, J., Heurtebize, T., Carlier, L. & Soupé, F. (2022), ‘Greenhouse gases emissions: estimating corporate non-reported emissions using interpretable machine learning’.
URL: *https://arxiv.org/abs/2212.10844*

Bhuiyan, M. (2024), Carbon footprint measurement and mitigation using AI, Technical report, Available at SSRN: https://ssrn.com/abstract=4746446 or http://dx.doi.org/10.2139/ssrn.4746446.

Bingler, J. A., Colesanti Senni, C. & Monnin, P. (2022), ‘Understand what you measure: Where climate transition risk metrics converge and why they diverge’, *Finance Research Letters* **50**, 103265.
URL: *https://www.sciencedirect.com/science/article/pii/S1544612322004561*

BIS (2024), Project gaia: Enabling climate risk analysis using generative AI, Technical report.

Busch, T., Johnson, M. & Pioch, T. (2022), ‘Corporate carbon performance data: Quo vadis?’, *Journal of Industrial Ecology* **26**(1), 350–363.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jiec.13008*

ECB (2025), ‘Opinion of the European Central Bank of 8 may 2025 on proposals for amendments to corporate sustainability reporting and due diligence requirements (con/2025/10)’.
URL: *https://www.ecb.europa.eu/pub/pdf/legal/ecb.leg*$_{c}$*on*$_{2}$025$_{1}$0*.en.pdf*

GHG-Protocol (2013), Technical guidance for calculating scope 3 emissions, Technical report.

Giannetti, M., Jasova, M., Loumioti, M. & Mendicino, C. (2023), ’glossy green’ banks: The disconnect between environmental disclosures and lending activities, Technical Report 919/2023, Swedish House of Finance Research Paper.
URL: *https://ssrn.com/abstract=4424081*

IEA (2023), Net zero roadmap: A global pathway to keep the 1.5 °c goal in reach, Technical report.

Jain, A., Padmanaban, M., Hazra, J., Godbole, S. & Weldemariam, K. (2023), ‘Supply chain emission estimation using large language models’.
URL: *https://arxiv.org/abs/2308.01741*

Korinek, A. (2023), ‘Generative ai for economic research: Use cases and implications for economists’, *Journal of Economic Literature* **61**(4), 1281–1317.
URL: *https://www.aeaweb.org/articles?id=10.1257/jel.20231736*

Nguyen, Q., Diaz-Rainey, I., Kitto, A., McNeil, B. I., Pittman, N. A. & Zhang, R. (2023), ‘Scope 3 emissions: Data quality and machine learning prediction accuracy’, *PLOS Climate* **2**(11), 1–31.
URL: *https://doi.org/10.1371/journal.pclm.0000208*

Nguyen, Q., Diaz-Rainey, I. & Kuruppuarachchi, D. (2021), ‘Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach’, *Energy Economics* **95**(C), S0140988321000347.
URL: *https://EconPapers.repec.org/RePEc:eee:eneeco:v:95:y:2021:i:c:s0140988321000347*

Papadopoulos, G. (2022), Discrepancies in corporate ghg emissions data and their impact on firm performance assessment, Technical Report 2022/12, JRC Working Papers in Economics and Finance, European Commission, Ispra.

Rasul, K., Ashok, A., Williams, A. R., Ghonia, H., Bhagwatkar, R., Khorasani, A., Bayazi, M. J. D., Adamopoulos, G., Riachi, R., Hassen, N., Biloš, M., Garg, S., Schneider, A., Chapados, N., Drouin, A., Zantedeschi, V., Nevmyvaka, Y. & Rish, I. (2024), 'Lag-llama: Towards foundation models for probabilistic time series forecasting'.
URL: *https://arxiv.org/abs/2310.08278*

Talbot, D. & Boiral, O. (2018), 'Ghg reporting and impression management: An assessment of sustainability reports from the energy sector', *Journal of Business Ethics* **147**.

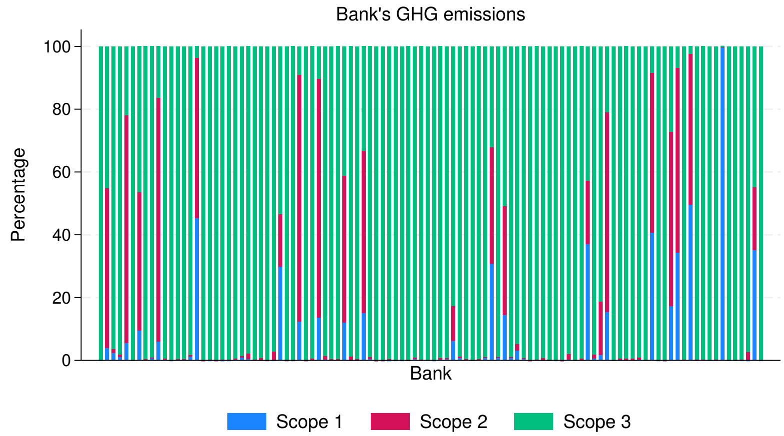
Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D. & Mann, G. (2023), 'Bloomberggpt: A large language model for finance'.
URL: *https://arxiv.org/abs/2303.17564*

Yang, H., Liu, X.-Y. & Wang, C. D. (2025), 'Fingpt: Open-source financial large language models'.
URL: *https://arxiv.org/abs/2306.06031*

Ángel Iván Moreno & Caminero, T. (2023), Assessing the data challenges of climate-related disclosures in european banks. A text mining study, Working Papers 2326, Banco de España.
URL: *https://ideas.repec.org/p/bde/wpaper/2326.html*

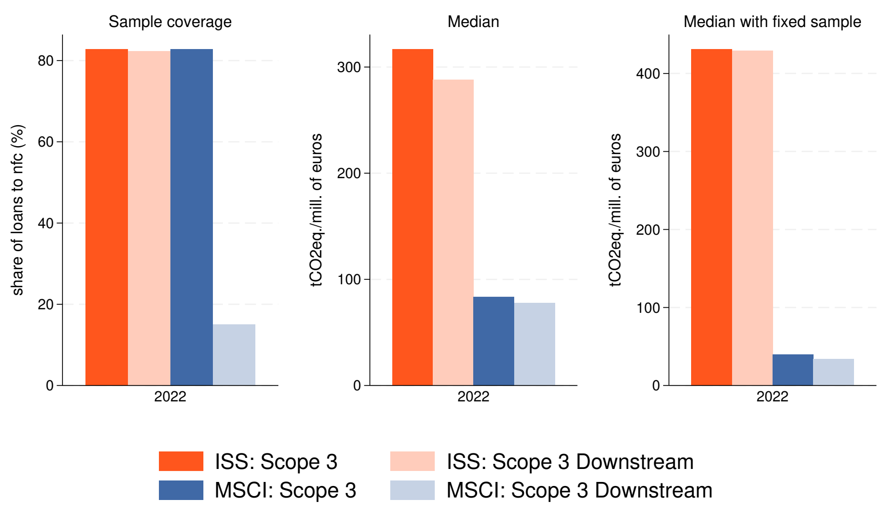
Appendix 1: Scope 3 emissions among providers: coverage, correlation and anomalies

Figure A1: Share of scope 1, 2 and 3 on overall emissions



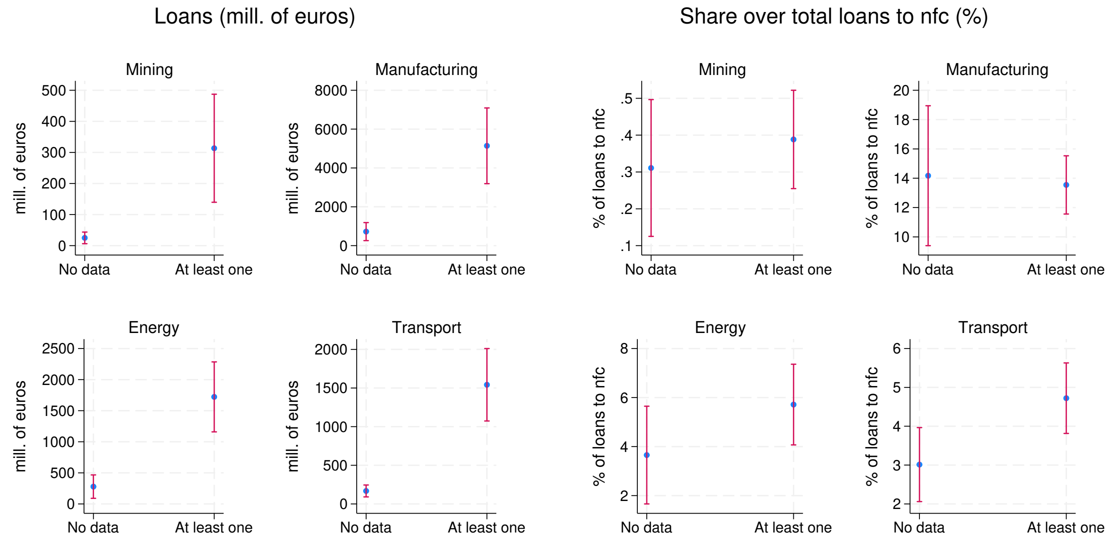
Note: For each bank the figure plots the share of scope 1, 2 and 3 emissions over the total GHG bank's emissions. For each bank the emissions type data is the average over the different providers (Bloomberg, MSCI, ISS and LSEG-Datastream) as of the end of 2022. Each bar refers to a bank in the sample.

Figure A2: Sample coverage and footprint for scope 3 and downstream emissions



Note: For each bank, emission footprint is computed as the ratio between the bank's emissions and the total loans to non-financial corporations (nfc) as of the end of 2022. In the central panel we compare the overall median scope 3 emissions and the median scope 3 downstream emissions for a given source (the median is computed on a sub-sample for which we have both emission types for a given source). In the right-hand side panel, the median is computed on the sub-sample of banks for which we have the downstream data from both providers.

Figure A3: Scope 3 emissions availability and exposure to high-emitting sectors



Note: The figure displays the average (blue dot) amount of loans granted by the banks within our sample to high-emitting sectors (in millions of loans) and share of loans granted to such sectors over the total loans to non-financial corporations (nfc) by a given bank in 2022. Confidence intervals are in red. Two groups are considered: 1) the banks for which we do not any scope 3 emissions data (No data); 2) those for which we have data on scope 3 emissions from at least one provider (At least one).

Table A1: Pairwise correlation of scope 3 emissions across different data sources

	(1)	(2)	(3)	(4)	(5)	Average
		Scope 3 emissions				**0.569**
(2)	0.298***					
(3)	0.903***	0.306***				
(4)	0.906***	0.466***	0.842***			
(5)	0.521***	0.757***	-0.073	0.523***		
(6)	0.886***	0.462***	0.237***	0.974***	0.519***	
		Downstream scope 3 emissions				**0.994**
(3)		0.994***				

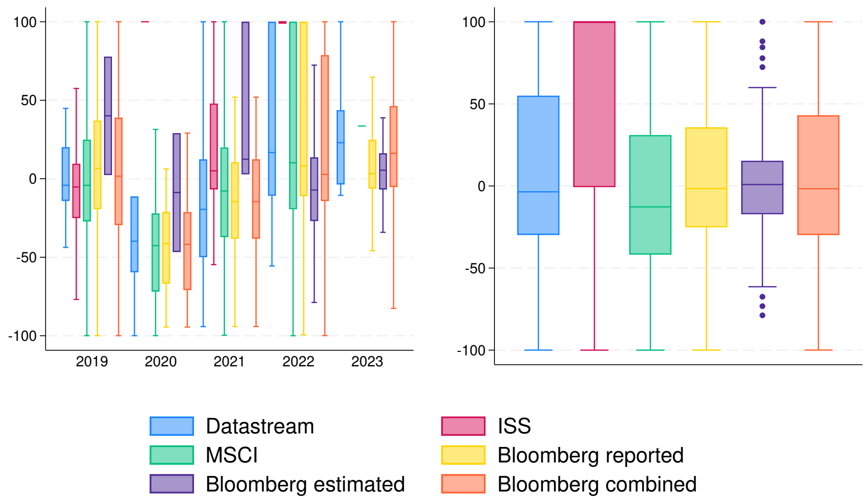
Note: Pairwise correlations using bank-level emission data from 2018 to 2022 per each data provider. In the table (1) refers to Datastream, (2) to ISS, (3) to MSCI, (4) to Bloomberg reported, (5) to Bloomberg estimated and (6) to Bloomberg combined.

Table A2: Differences in banks' scopes 3 emission footprints across sources and exposure to high-emitting sectors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Mining	-91.91*	-118.8	-123.2	-1066.1	-127.0	-42720.3	-27892.9	-1044417.6	-27464.8	395.7***	-1486664.6	262.0	-1135220.8	-0.198
	(0.052)	(0.118)	(0.281)	(0.336)	(0.267)	(0.318)	(0.337)	(0.321)	(0.331)	(0.003)	(0.313)	(0.389)	(0.304)	(0.731)
Manufacturing	-1.796	-1.507	-8.788	-106.7	-9.596	-5722.5	-5902.5	-15997.1	-5594.8	-42.87*	-5101.4	-181.1	-52215.1	0.374
	(0.786)	(0.837)	(0.290)	(0.274)	(0.251)	(0.265)	(0.249)	(0.713)	(0.253)	(0.090)	(0.923)	(0.189)	(0.408)	(0.272)
Energy	8.239	-17.87	-26.35	-141.0	-19.96	-24070.0	-6121.4	-350797.6	-5935.3	-53.88	-548802.9	-213.1	-580540.5	0.965
	(0.770)	(0.472)	(0.521)	(0.679)	(0.643)	(0.232)	(0.349)	(0.334)	(0.342)	(0.274)	(0.330)	(0.207)	(0.310)	(0.454)
Transport	-5.295	8.375	11.11	-87.92	9.165	12069.3*	9943.4	255619.5	9775.2	215.7**	308590.8	148.5	351007.3	-0.653
	(0.759)	(0.517)	(0.450)	(0.151)	(0.556)	(0.072)	(0.128)	(0.352)	(0.131)	(0.025)	(0.347)	(0.192)	(0.320)	(0.362)
Observations	156	149	170	70	170	181	203	76	209	221	74	232	113	351
R^2	0.105	0.066	0.067	0.106	0.078	0.090	0.057	0.088	0.055	0.444	0.111	0.099	0.125	0.020

Note: Results from a set of multivariate regressions where the dependent variable is the difference in the bank's scope 3 emissions footprints between two different sources. The bank's scope 3 emissions (or carbon) footprint is computed as the ratio between its scope 3 emissions and its total loans to non-financial corporations. The independent variables refer to the share of bank's loans to non-financial corporations belonging to high-emitting sectors (mining, manufacturing, energy and transport). Each column refer to a different pair of data sources. All the regressions include the constant and a year fixed effect. Errors are clustered at the bank level. P-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A4: Data anomalies: average percentage change in scope 3 emissions footprint



Note: The left-hand side figure displays box-plots of the average bank-level annual percentage change in scope 3 emissions footprint over time by data source. The right-hand side figure displays box-plots of the average bank-level annual percentage change in scope 3 emissions computed over the entire sample period by data source. Outliers are not reported, data are limited to be between -100 and 100%. The bank's scope 3 emissions (or carbon) footprint is computed as the ratio between its scope 3 emissions and its total loans to non-financial corporations.

Appendix 2: Analysis of GenAI-based emission data

This Appendix discusses in greater detail the prompts and the relevant replies provided by Claude, ChatGPT and Gemini.

A2.1 Claude

We first ask Claude 3.5, without enabling the *Web search* option, to recover the scope 1, 2 and 3 emissions for the banks in our sample in 2022 using the following prompt: *Can you please recover the scope 1, 2 and 3 emissions of the following banks in 2022? Can you please save the data in an excel file (that I can export) where there is a raw for each bank and 3 columns, one for scope 1, one for scope 2 and one for scope 3?* Claude 3.5 provided us with a table with relevant data by clarifying that the supplied data were not extracted from the banks' sustainability reports, but it was estimated to be representative of typical emissions for these institutions and should be verified against the banks' most recent official reports.[29] The values did not change once we specified that, as investors, we sought to assess the carbon footprint of the bank list. However, the tool further stressed that the data consist of rough estimates that should be considered order-of-magnitude approximations at best and are not suitable for investment decisions without verification against official sources.[30] Once Claude 3.5 was asked to assess the degree of confidence for the reported estimates in a scale from 1 (lowest confidence) to 10 (highest confidence), it assigned a higher confidence score to the

[29] It first included only a subset of the largest banks in the sample and then, when prompted, added the remaining ones.

[30] It recommended that investors take a more systematic approach to obtain accurate emissions data. The recommendation included replacing estimated values with official and verified data from sustainability reports or other sources (i.e. CDP), assessing the precision of scope 3 emissions (particularly category 15, i.e., financed emissions), as these typically represent more than 95% of the total carbon footprint of banks and are most material for investors, and documenting the methodologies used by each bank.

data related to large global banks (i.e., between 6 and 7) since these banks have comprehensive public disclosures and more standardized reporting, making estimates more reliable; slightly lower scores for mid-sized European banks (i.e. 5 or 6) as they generally have good disclosure although differences in methodology and completeness; low values for smaller regional banks (between 3 and 4) with limited public data, more reporting inconsistencies and less standardized methodologies; minimum scores (2 or 3) for specialized institutions with very diverse business models that make comparison and estimation difficult, as well as for counties with less mature sustainability reporting frameworks. The tool further clarified that the confidence is particularly low for scope 3 emissions, which constitute the vast majority of a bank's carbon footprint but are reported using widely varied methodologies. Finally, when inquired, Claude 3.5 specified that it is unable to accurately perform the task of providing actual verified emissions data.[31]

When launching the prompt with Claude 3.7, instead, we specified to assume that we were investors or supervisory authorities trying to assess the carbon footprint of the banks in the list by using estimated or reported data of emissions.[32] Without enabling the *Web search* option, Claude 3.7 replied to our request by providing a table that displayed: 1) reported data based on publicly available sustainability reports and TCFD disclosures as of 2022, when available, and 2) estimated data, mostly for the smaller banks. Claude 3.7 did not retrieve the information on the web, but only from its training sample, therefore, suggesting that the latter includes bank's sustainability reports and/or climate-related information. For the scope 1 and 2 emissions, the estimates were based on the bank size (e.g., number of employees, total assets), location (e.g., country's energy mix, and data available for peer financial institutions). Scope 3 emissions, instead, were estimated using the credit and investment portfolios when available, using sectoral emission factors, and standard methodologies (i.e., the PCAF - Partnership for Carbon Accounting Financials), reported data by banks with similar size and business model or sectoral data for the banking sector. For scope 1 and 2 emissions, more than 57% of the data were classified as reported, while this was the case only for 17% of the scope 3 emissions. The reliability of the estimated scope 3 emissions was classified most of the times as medium (60% of the banks), while in almost all of the other cases the reliability score was low, especially for smaller banks with less public information for which the estimates were mostly based on sectoral averages.[33] Once again, the tool specified that these were not official data and need to be verified against official sources, as described above for Claude 3.5.[34]

A2.2 ChatGPT

When we asked ChatGPT, with the *Web search* option activated, to recover the scope 1, 2, and 3 emissions for all the banks in our sample, using the same prompt as for Claude. In this case, the web search option was enabled as the system automatically recognised the need to search online. Its first reply retrieved the emission data only for one bank. The answer stated that, unfortunately, detailed emissions data for the other banks were not readily accessible in public sources and recommended consulting the banks' sustainability reports or official websites. We requested to revise the answer to the same question multiple times and obtained data for a few more banks. Since most of the observations were missing, despite the numerous requests, we proceeded by asking for a few banks each time. Therefore, the list of banks was split into several sub-samples, and the same prompt was launched for each sub-sample. We further requested revisions and double checks in case of missing

[31] The following reasons were mentioned: 1) access limitations: it does not have direct access to subscription databases like CDP, Bloomberg, or MSCI ESG; 2) update of data: training data has limited information on 2022 and 2023 sustainability reports for many of these banks, which are necessary for investment decisions; 3) verification requirements: any data used for investment decisions should be carefully verified against official sources.

[32] *Imagine I am an investor or a supervisory authority trying to assess the carbon footprint of the following banks. I need to obtain estimates or reported data on the institutions' emissions.*

[33] The reliability score was medium for most of the banks whose available information on the portfolios was limited, but data for similar institutions allow to apply more detailed models.

[34] Scope 1 and 2 emissions in the 2nd and 3rd runs are the same, as in the last run only the scope 3 emissions were corrected.

and unexpected values, relaunching the same question and asking to verify the answer. According to ChatGPT, self-reported data were unavailable for several banks along this procedure. Although the prompt remained consistent across different queries, the outcomes varied depending on the length of the list provided, which suggests potential issues with data replicability and consistency. Examining the banks in our samples in different subsets or individually may yield responses of different, and possibly higher, quality.

Given the large number of missing data points, we asked ChatGPT to estimate the emissions data for our sample of banks.[35] Despite the invitation to display estimated data for all the bank sample, ChatGPT combined estimated and reported data when available. The reply highlighted that the estimated data would be approximate and based on available industry benchmarks, bank size (as measured by assets, employees, and revenues), region, and emissions data from comparable institutions. They should be used only for screening or indicative analysis, not for reporting or regulatory purposes.

We also tested whether ChatGPT Plus 4.o, a more advanced model by OpenAI available by subscription with a less restrictive limit on the number of messages the user can send, could overcome some of the limits of ChatGPT's freely available model. Nevertheless, we encountered similar challenges and limitations related to its ability to retrieve banks' self-reported data, as well as issues with data quality and replicability. The retrieved information varied across similar requests, and the estimated data supplied by ChatGPT Plus 4.o was not correlated with that provided by ChatGPT.

We encountered similar challenges in retrieving data from the web, using two additional models, namely 4.o mini and o3. Indeed, in these cases, we obtained similar responses, highlighting that we were asking for a large (or huge) dataset, which could only be partially delivered by proceeding step by step, starting with the largest banks. The same occurred when we explicitly asked to search on the web with the *Web search* tool. For this reason, we did not pursue the data collection and the empirical analysis for these alternative models.

A2.3 Gemini Advance

Similarly, Gemini Deep search with 2.5 Pro replied to our initial prompt, retrieving data from the web only for a limited number of banks and recommending checking the individual banks' 2022 annual or sustainability reports. Then, after a few iterations, the data were updated for the other banks by consulting a large number of websites to recover the annual reports and sustainability or ESG reports published by each bank in the provided list for the year 2022 or public replies to the CDP questionnaire. A detailed description of the data, methodologies, and sources was provided. Similar replies were provided by Gemini 2.5 Pro. The latter also provided estimated data once asked based on the size of the bank. In all these queries the *Web search* option was activated.

Appendix 3: Additional information on GenAI-based emission data

[35] *Can you please estimate the scope 1, 2 and 3 emissions of the following banks in 2022?*

Table A3: GenAI-based scope 3 footprint and exposure to high-emitting sectors

	(a)	(b)	(c)	(d)	(5)	(f)	(g)	(h)
	Scope 3 emissions footprint							
Mining	-18058.5	-150279.5	-18978.4	-873.0	-109.6	-54901.3	107.693	-110810.6
	(0.315)	(0.312)	(0.104)	(0.313)	(0.304)	(0.331)	(0.641)	(0.313)
Manufacturing	-3399.9	-28093.7	117.2	-168.7	-19.94	-12820.2	-42.135	-17964.5
	(0.306)	(0.306)	(0.805)	(0.306)	(0.312)	(0.314)	(0.305)	(0.331)
Energy	-2869.5	-23683.6	400.8	-139.4	-16.80	-12520.1	-28.698	-15696.1
	(0.311)	(0.311)	(0.317)	(0.312)	(0.317)	(0.324)	(0.639)	(0.327)
Transport	-3966.6	-33081.9	13726.0*	-191.9	-23.23	-4086.8	-17.108	-21515.2
	(0.327)	(0.323)	(0.085)	(0.327)	(0.332)	(0.403)	(0.689)	(0.318)
Obs.	129	129	42	127	129	48	88	115
R^2	0.024	0.024	0.481	0.024	0.024	0.092	0.0295	0.0282

Note: Results from a set of regressions where the dependent variables bank-level GenAI-based scope 3 emissions footprint computed as the ratio between scope 3 emissions and the banks' total loans to non-financial corporations. The independent variables are the share of bank loans to non-financial corporations belonging to high-emitting sectors (mining, manufacturing, energy and transport). All the regressions include the constant. Each column refers to a model (a) refers to Claude 3.5 - estimated, (b) Claude 3.7 - retrieved & estimated (c) ChatGPT - retrieved, (d) ChatGPT - retrieved & estimated, (e) ChatGPT Plus - estimated, (f) Gemini Deep Search - retrieved, (g) Gemini 2.5 Pro - retrieved, (h) Gemini 2.5 Pro - retrieved & estimated. P-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$