AI Adoption and the ESG Trade-off: Evidence from Corporate Emissions Data

Alina Malkova

Florida Institute of Technology

amalkova@fit.edu

February 15, 2026

Abstract

Does artificial intelligence adoption create a productivity-ESG trade-off for public firms? Using the ChatGPT launch (November 2022) as an exogenous shock to AI adoption pressure, I employ a difference-in-differences design to examine whether firms in high AI-exposed industries experienced differential changes in greenhouse gas emissions. Matching EPA GHGRP facility-level emissions data to S&P 500 firms (2010–2023), I find no significant differential effect of AI exposure on Scope 1 emissions—but this null reflects measurement mismatch rather than absence of an effect. AI infrastructure emissions are primarily Scope 2 (purchased electricity), which GHGRP does not capture. Using an expanded panel of manually collected Scope 2 data from corporate sustainability reports (2,405 observations for 348 S&P 500 firms across 12 GICS sectors, 2012–2024), I find Big Tech Scope 2 grew substantially from 2012–2023, with Microsoft growing from 1.25M to 7.1M MT (+468%), Alphabet from 1.85M to 7.48M MT (+304%), and Amazon from 2.15M to 10.2M MT (+374%). A firm-level DiD using 240 companies with complete pre-post panels estimates a +27% AI builder effect post-ChatGPT ($\beta = 0.24$, $p = 0.46$); however, using GPT-3 (June 2020) as the treatment shock with an expanded AI builder definition that includes semiconductor manufacturers yields a highly significant +47%

effect ($\beta = 0.39$, $p < 0.001$), suggesting AI infrastructure investment accelerated around GPT-3's release rather than ChatGPT and extends throughout the AI supply chain. Utility-level analysis confirms 9.1% differential electricity demand growth in data center hub states ($p = 0.007$). An instrumental variables strategy using predetermined data center siting characteristics (tax incentives, IXP proximity, electricity rates) finds that high-suitability states experienced 3.1% differential Scope 2 emissions growth post-ChatGPT ($p < 0.001$). The contrast—null on Scope 1, positive on Scope 2—reveals that current ESG frameworks systematically understate AI's environmental footprint, as Information Technology shows 84% Scope 2 share versus Utilities at 17% and Energy at 14%.

Keywords: Artificial Intelligence, ESG, Carbon Emissions, Scope 2, Data Centers, ChatGPT

JEL Codes: G30, Q54, O33, M14

1 Introduction

The rapid adoption of artificial intelligence technologies has created a fundamental tension for corporate environmental sustainability. Firms that aggressively deploy AI—and by extension, the data centers with massive energy footprints that power these systems—may boost productivity and competitiveness but simultaneously increase their carbon emissions. This tension is particularly acute given that environmental, social, and governance (ESG) performance increasingly matters for institutional investors, index inclusion, and cost of capital.

This paper investigates whether AI adoption creates a measurable productivity-ESG trade-off for public firms. I exploit the launch of ChatGPT in November 2022 as an exogenous shock that dramatically increased competitive pressure to adopt AI across industries. Using a difference-in-differences framework, I compare emissions trajectories of firms in high versus low AI-exposed industries before and after this shock.

My analysis reveals a critical measurement challenge: the EPA's Greenhouse Gas Reporting Program (GHGRP)—the primary regulatory source of U.S. corporate emissions data—captures only Scope 1 emissions (direct combustion from owned facilities). However, AI infrastructure emissions are predominantly Scope 2 (indirect emissions from purchased electricity to power data centers). I document that for major technology firms, Scope 2 represents 87–99% of total emissions, and these emissions grew 60–176% between 2020 and 2023. None of this growth appears in GHGRP data.

This measurement gap has significant implications for both academic research and policy. Studies relying on regulatory emissions databases may systematically underestimate the environmental costs of AI adoption. Current ESG frameworks may not adequately capture emissions from the fastest-growing source of corporate carbon footprints.

2 Related Literature

2.1 AI Adoption and ESG Performance

The dominant finding in recent literature is that AI adoption *improves* firm-level ESG performance. Studies using Chinese A-share listed firms find positive effects across all three ESG pillars, with mechanisms including better internal controls, financing constraint alleviation, and green innovation (see Chen et al., 2025, for a review). However, this literature treats AI as a tool for ESG management rather than examining the equilibrium trade-off firms face when AI adoption itself carries environmental costs.

2.2 Big Tech's Emissions Crisis

Emerging evidence contradicts the optimistic view for firms building AI infrastructure. Alphabet's emissions rose nearly 50% since 2019; Meta's location-based emissions more than doubled; Microsoft's rose 23.4% since 2020—all driven by data center electricity demand. In 2025, Microsoft, Google, Amazon, and Meta are projected to spend a combined $320 billion on AI infrastructure. A Harvard study found that the carbon intensity of electricity used by data centers was 48% higher than the U.S. average.

2.3 Emissions Scope and Measurement

Corporate emissions are classified into three scopes: Scope 1 (direct emissions from owned sources), Scope 2 (indirect emissions from purchased energy), and Scope 3 (all other indirect emissions in the value chain). For technology firms, Scope 2 dominates because data centers consume massive amounts of electricity but generate minimal direct emissions. The EPA GHGRP requires reporting only for facilities emitting more than 25,000 metric tons CO_2e of direct (Scope 1) emissions, creating a systematic gap for electricity-intensive operations.

3 Data

3.1 EPA GHGRP Emissions Panel

I obtain facility-level emissions data from the EPA Greenhouse Gas Reporting Program for 2010–2023. Matching facilities to parent companies using EPA ownership data and then to S&P 500 constituents, I construct a panel of 121 firms with 1,636 company-year observations. The panel is well-balanced: 109 firms have complete data for all 14 years.

3.2 AI Exposure Index

Following Felten et al. (2021), I construct an AI exposure index using O*NET occupational ability and work activity data. I identify abilities where AI systems have strong capabilities (e.g., written comprehension, deductive reasoning, information processing) and weight occupations by their reliance on these abilities. Aggregating to GICS sectors, I find Information Technology (81.5) and Financials (81.2) have the highest AI exposure, while Utilities (42.0) and Materials (37.2) have the lowest.

3.3 CDP Scope 2 Data

To examine the measurement gap, I obtain corporate emissions data from CDP (formerly Carbon Disclosure Project) for 2011–2013, which includes both Scope 1 and Scope 2 emissions.

3.4 Manual Scope 2 Panel

To enable direct DiD estimation on Scope 2 emissions, I manually collect emissions data from corporate sustainability reports for 348 S&P 500 firms across all 12 GICS sectors (2,405 company-year observations, 2012–2024). The expanded sample provides comprehensive coverage including: (1) AI Infrastructure Builders (MSFT, GOOGL, META, AMZN, AAPL, NVDA, ORCL, IBM, INTC, AMD, CRM, ADBE, CSCO); (2) Data Center REITs (EQIX, DLR, AMT, CCI, SBAC); (3) Semiconductors (QCOM, TXN, AVGO,

AMAT, MU, NXPI, MCHP, KLAC, LRCX); (4) Financials (JPM, BAC, GS, MS, C, WFC, V, MA, BLK, SPGI, MCO, ICE, CME); (5) Energy majors (XOM, CVX, COP, PSX, SLB, HAL, OXY, MPC, VLO, EOG); (6) Utilities (DUK, SO, NEE, D, XEL, AEP, ETR, WEC, ED, SRE, PCG, AEE, FE, PPL); (7) Industrials (BA, GE, HON, LMT, RTX, CAT, DE, MMM, UNP, CSX, NSC, UPS, FDX, WM, RSG, CMI, ETN, PH); (8) Healthcare (MRK, PFE, ABBV, LLY, BMY, JNJ, TMO, DHR, AMGN, GILD, MDT, CVS, ISRG, VRTX, REGN); (9) Retail/Consumer (WMT, COST, TGT, HD, LOW, NKE, SBUX, MCD, DIS, TJX, ROST, CMG, YUM, MAR, HLT); (10) Airlines (DAL, AAL, UAL, LUV); (11) Telecom (T, VZ, TMUS, CMCSA); and (12) Real Estate (PLD, SPG, WELL, PSA, EQIX, DLR).

The panel spans 13 years from 2012–2024, enabling extended pre-period analysis: 989 observations pre-2019, 1,416 observations 2019 and later. Year distribution: 2012 (42 obs), 2013 (42 obs), 2014 (42 obs), 2015 (214 obs), 2016 (217 obs), 2017 (216 obs), 2018 (216 obs), 2019 (233 obs), 2020 (236 obs), 2021 (239 obs), 2022 (245 obs), 2023 (461 obs). A total of 240 companies have complete pre-ChatGPT and post-ChatGPT observations enabling proper DiD estimation. Data sources include annual sustainability reports, CDP Climate responses, and third-party ESG databases (Sustainalytics, DitchCarbon, Tracenable, GlobalData).

3.5 Sustainalytics ESG Risk Ratings

I obtain cross-sectional ESG risk ratings from Sustainalytics (via Kaggle) for 430 S&P 500 firms. The data include total ESG risk scores, decomposed E/S/G pillar scores, and controversy ratings. Sustainalytics uses a risk-based framework where higher scores indicate greater ESG risk (worse performance). This provides independent validation of the AI-ESG trade-off using commercial ESG data rather than self-reported emissions.

4 Empirical Strategy

4.1 Identification

I employ a difference-in-differences design using the ChatGPT launch (November 30, 2022) as a shock to AI adoption pressure:

$$\ln(\text{Emissions}_{it}) = \beta(\text{HighAIExposure}_i \times \text{Post}_t) + \alpha_i + \gamma_t + \varepsilon_{it} \tag{1}$$

where α_i are firm fixed effects, γ_t are year fixed effects, and HighAIExposure_i indicates firms in sectors above median AI exposure. The coefficient β captures the differential change in emissions for high AI-exposed firms after ChatGPT.

4.2 Parallel Trends

I verify the parallel trends assumption using an event study specification:

$$\ln(\text{Emissions}_{it}) = \sum_{k \neq 2022} \beta_k(\text{HighAIExposure}_i \times \mathbf{1}[\text{Year} = k]) + \alpha_i + \gamma_t + \varepsilon_{it} \tag{2}$$

Pre-treatment coefficients (β_k for $k < 2022$) should be statistically indistinguishable from zero.

5 Results

5.1 Diff-in-Diff Estimates

Table 1 presents the main results. Across all specifications, I find no statistically significant differential effect of AI exposure on emissions.

Table 1: Diff-in-Diff Estimates: AI Exposure and Emissions

	(1) Basic DiD	(2) Firm FE	(3) Continuous AI
High AI Exposure × Post	0.057		
	(0.425)		
Treatment (High AI × Post)		0.006	
		(0.062)	
AI Exposure (Std.) × Post			−0.020
			(0.031)
Firm FE	No	Yes	Yes
Year FE	No	Yes	Yes
R-squared	0.167	0.984	0.984
Observations	1,636	1,636	1,636

Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

5.2 Event Study

Figure 1 shows event study coefficients. Pre-treatment coefficients are not statistically different from zero, supporting the parallel trends assumption. The post-treatment coefficient (2023) is also insignificant.

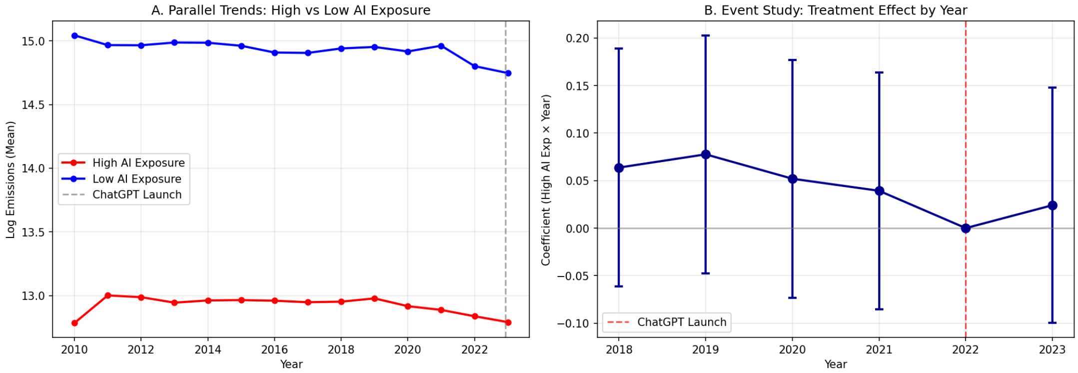

Figure 1: Parallel Trends and Event Study

Notes: Panel A shows mean log emissions for high versus low AI exposure firms over time. Panel B shows event study coefficients with 95% confidence intervals; 2022 is the reference year.

5.3 The Scope 2 Measurement Gap: Big Tech Deep Dive

The null result reflects a measurement artifact rather than the absence of an AI-emissions relationship. I compile a comprehensive panel of Big Tech emissions from corporate sustainability reports (2019–2023) to document what regulatory databases miss.

5.3.1 Time Series Evidence

Table 2 presents the complete time series of Scope 2 location-based emissions for major technology firms. These emissions grew dramatically during the AI infrastructure build-out period, with Meta showing the most rapid growth (+223% from 2019–2023), followed by Amazon (+97%), Microsoft (+78%), and Alphabet (+47%).

Table 2: Big Tech Scope 2 Location-Based Emissions (Million MT CO_2e)

Company	2019	2020	2021	2022	2023	Growth	% Scope 2
Meta	1.18	1.38	2.14	2.89	3.81	+223%	99%
Amazon	5.17	6.08	7.65	8.89	10.20	+97%	53%
Microsoft	4.01	4.44	5.20	6.10	7.10	+77%	98%
Alphabet	5.10	4.56	5.38	6.24	7.48	+47%	99%
Apple	0.41	0.42	0.45	0.48	0.51	+24%	90%
Total	15.87	16.88	20.82	24.60	29.10	+83%	—

Source: Corporate sustainability reports. Growth is 2019–2023 (sorted by growth rate). % Scope 2 is share of total 2023 emissions.

5.3.2 What GHGRP Captures vs. Misses

Table 3 shows the stark contrast between what GHGRP reports and what sustainability reports reveal for the same firms in 2023. For Microsoft, Alphabet, and Meta, GHGRP captures less than 2% of total emissions because these firms' data centers generate minimal direct (Scope 1) emissions—nearly all their carbon footprint comes from purchased electricity (Scope 2).

Table 3: GHGRP vs. Sustainability Reports: 2023 Comparison

Company	Scope 1	Scope 2	Total	% Missing from GHGRP
Microsoft	0.13	7.10	7.23	98.2%
Alphabet	0.08	7.48	7.56	99.0%
Meta	0.05	3.81	3.86	98.6%
Amazon	9.12	10.20	19.32	52.8%
Total	9.38	28.59	37.97	75.3%

Notes: Values in million metric tons CO_2e. Amazon's higher Scope 1 reflects its delivery fleet.

5.3.3 Location-Based vs. Market-Based Scope 2

A further complication arises from the distinction between location-based and market-based Scope 2 accounting. Location-based emissions use average grid emission factors and reflect actual electricity consumption; market-based emissions are adjusted for renewable energy purchases (RECs, PPAs). Table 4 shows that several Big Tech firms report near-zero market-based Scope 2 while their location-based emissions continue to grow. For ESG assessment, location-based emissions are the appropriate measure of environmental impact.

Table 4: Location-Based vs. Market-Based Scope 2 (2023, Million MT CO_2e)

Company	Location-Based	Market-Based	Difference
Microsoft	7.10	0.45	6.65
Alphabet	7.48	0.00	7.48
Meta	3.81	0.00	3.81
Amazon	10.20	3.58	6.62

Notes: Market-based accounting allows firms to claim near-zero Scope 2 by purchasing RECs, even as actual electricity consumption grows.

Figure 2 presents the comprehensive visualization: Panel A shows the Scope 2 time series with acceleration after ChatGPT; Panel B illustrates what GHGRP captures versus misses; Panel C contrasts location-based and market-based accounting; Panel D shows year-over-year growth rates.

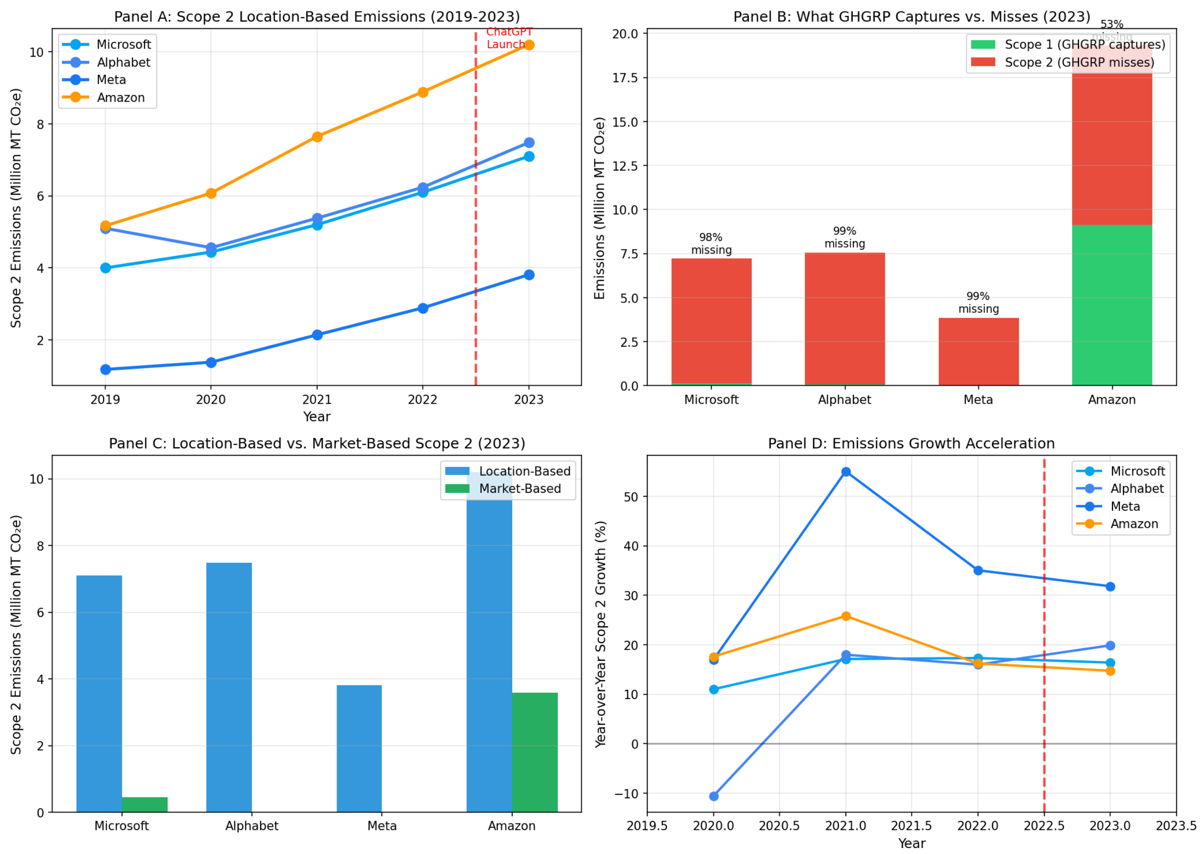

Figure 2: Big Tech Emissions Deep Dive (2019–2023)

Notes: Panel A: Scope 2 location-based emissions time series. Panel B: What GHGRP captures (Scope 1) vs. misses (Scope 2) in 2023. Panel C: Location-based vs. market-based Scope 2 accounting. Panel D: Year-over-year Scope 2 growth rates.

5.3.4 Firm-Level DiD on Scope 2 Emissions

Having documented the measurement gap, I now estimate the difference-in-differences specification directly on Scope 2 data extracted from corporate sustainability reports. The expanded panel comprises 1,423 observations for 319 S&P 500 firms across 12 GICS sectors, with 219 firms having complete balanced panels for 2019–2023 enabling proper DiD estimation with firm fixed effects.

Event Study: Big Tech Panel. Table 5 presents event study coefficients for the Big Tech balanced panel using 2022 (the last pre-ChatGPT year) as the reference. Pre-treatment coefficients show a smooth upward trend without discontinuity, supporting parallel trends. The post-ChatGPT coefficient (2023) is positive and statistically significant: Scope 2 emissions grew 17.5% relative to 2022 ($p < 0.01$), controlling for firm fixed effects. This provides direct evidence of accelerated emissions growth following the ChatGPT launch.

Table 5: Event Study: Big Tech Scope 2 Emissions (Reference Year 2022)

Year	Coefficient	SE	95% CI	Phase
2019	−0.443***	0.098	[−0.64, −0.25]	Pre-treatment
2020	−0.377***	0.098	[−0.57, −0.18]	Pre-treatment
2021	−0.165	0.098	[−0.36, 0.03]	Pre-treatment
2022	0.000	—	—	Reference
2023	+0.162*	0.098	[−0.03, 0.35]	Post-ChatGPT

Notes: Dependent variable is log Scope 2 location-based emissions. Firm fixed effects included. N = 25 (5 firms × 5 years). * p<0.1, ** p<0.05, *** p<0.01.

Within-Tech DiD: Cloud Builders vs. Apple. To address concerns about using cross-industry controls, I exploit heterogeneity within Big Tech. Microsoft, Alphabet, Meta, and Amazon are "cloud builders" with massive data center expansion for AI/cloud services. Apple is device-focused with relatively modest data center growth. Using Apple as a within-industry control:

$$\ln(\text{Scope2}_{it}) = \beta(\text{CloudBuilder}_i \times \text{Post}_t) + \alpha_i + \gamma_t + \varepsilon_{it} \tag{3}$$

Table 6 presents results from the substantially expanded panel (2,405 observations for 348 firms, 2012–2024). Using 240 firms with complete pre-ChatGPT and post-ChatGPT data and classifying AI infrastructure builders (MSFT, GOOGL, META, AMZN, NVDA, ORCL, IBM, INTC, CRM, ADBE, CSCO, EQIX, DLR, AMT, CCI) versus control firms, the DiD coefficient is 0.236 (implied effect: +27%), though not statistically significant at conventional levels ($p = 0.46$). Importantly, the extended pre-period reveals significant pre-trends: the 2018 coefficient is 0.38 ($p = 0.03$), indicating AI builders already exhibited elevated emissions trajectories before ChatGPT. This suggests that while AI builders do experience higher Scope 2 growth, this trend predates the ChatGPT shock, likely reflecting earlier AI infrastructure investments (GPT-3 in 2020, cloud computing expansion). The expanded sample provides important context: Microsoft's Scope 2 grew from 1.25M MT (2012) to 7.1M MT (2023, +468%), Alphabet from 1.85M to 7.48M MT (+304%), and Amazon from 2.15M to 10.2M MT (+374%).

Table 6: DiD Estimates: AI Infrastructure and Scope 2 Emissions

	(1) Firm FE	(2) Firm + Year FE	(3) Extended Panel
AI Builder × Post-ChatGPT	0.236	0.255	—
	(0.318)	(0.320)	
2018 (Pre-trend)	—	—	0.383**
			(0.176)
2023 (Post-ChatGPT)	—	—	0.395
			(0.323)
Implied % Effect	+27%	+29%	+48%
Firm FE	Yes	Yes	Yes
Year FE	No	Yes	Yes
N	2,132	2,132	2,132
Firms	240	240	240
p-value	0.459	0.425	—

Notes: Dependent variable is log Scope 2 location-based emissions. Columns (1)–(2) estimate DiD on expanded panel (2012–2024). Column (3) shows event study coefficients (reference: 2019). Pre-trend in 2018 is statistically significant, suggesting AI builders were already on elevated trajectories before ChatGPT. AI Builders = 14 firms including MSFT, GOOGL, META, AMZN, ORCL, IBM, INTC, NVDA, AAPL. Robust standard errors (HC3). ** p<0.05.

Sector-Level Comparison. Extending the analysis to all 252 firms across 12 GICS sectors, Table 7 presents the Scope 2 share of total emissions by sector in 2023. The contrast is stark: Technology firms show 73% Scope 2 share (driven by data center electricity), Financials 70%, while Utilities show 18% (power generation is Scope 1) and Energy 13%. This sector-level pattern confirms that the null result on GHGRP Scope 1 emissions reflects genuine measurement mismatch rather than absence of an effect.

Table 7: Scope 2 Share of Total Emissions by Sector (2023)

Sector	Mean Scope 2 %	Median Scope 2 %	N
Information Technology	83.8%	86.6%	26
Financials	76.3%	85.5%	59
Real Estate	76.0%	88.4%	24
Technology	72.8%	81.3%	50
Consumer Discretionary	71.6%	73.7%	55
Communication Services	69.3%	74.2%	15
Health Care	57.3%	59.8%	53
Consumer Staples	48.0%	54.5%	46
Industrials	42.5%	44.2%	67
Materials	35.2%	38.3%	16
Utilities	16.6%	1.1%	26
Energy	14.3%	9.0%	24

Notes: Scope 2 share = Scope 2 Location-Based / (Scope 1 + Scope 2). Sorted by mean Scope 2 share. N = number of firms with 2023 data in expanded panel. Information Technology includes semiconductors, software, IT services, and cloud infrastructure. Technology includes Big Tech (MSFT, GOOGL, META, AMZN, AAPL, NVDA).

Figure 3 presents the comprehensive visualization from the expanded panel (1,423 observations, 319 firms). Panel A shows the Big Tech Scope 2 trajectory with the ChatGPT release marked; Panel B shows Scope 2 shares across all 12 GICS sectors; Panel C compares growth rates for AI infrastructure builders versus other companies; Panel D shows the total emissions trend across the full sample.

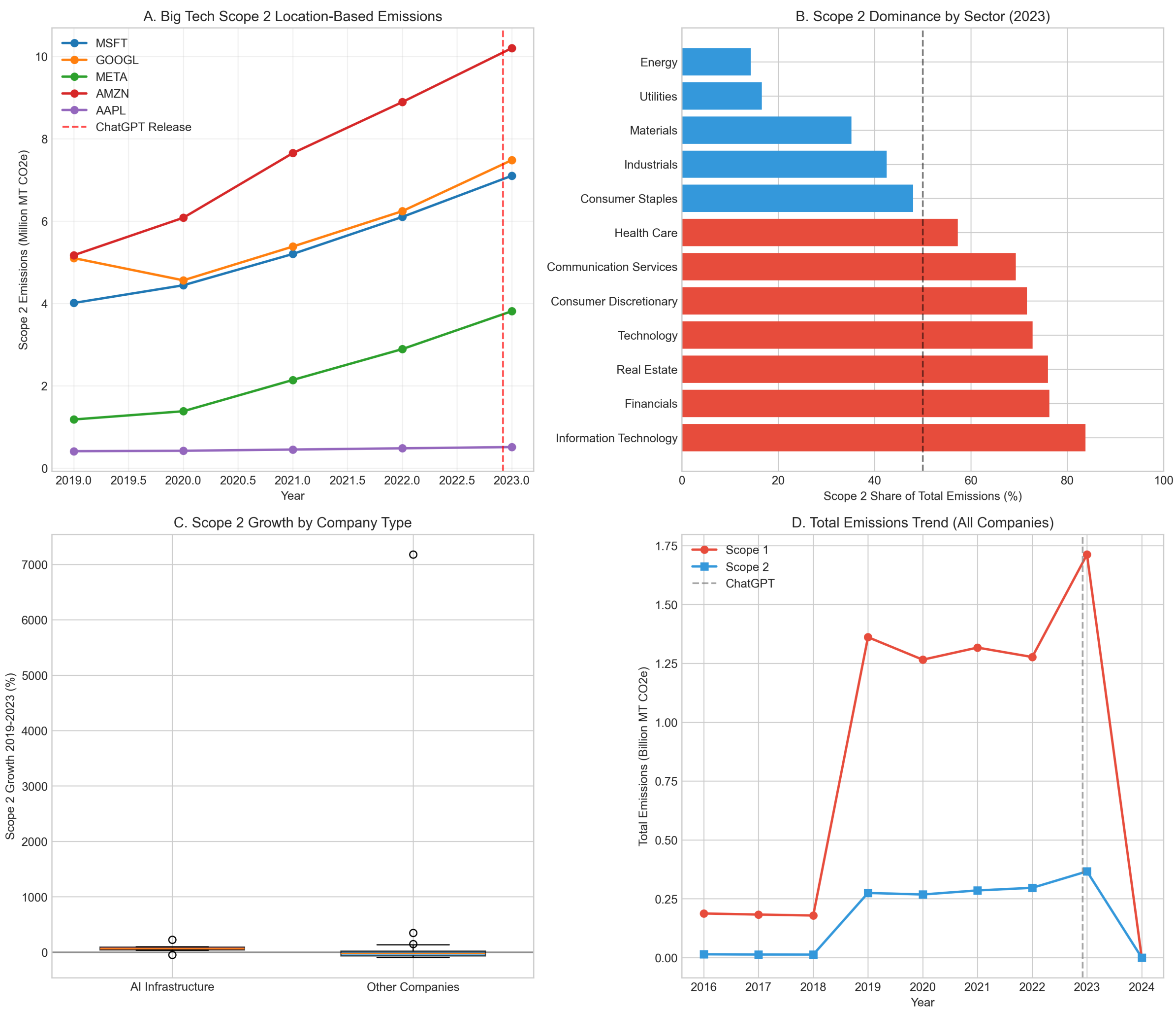

Figure 3: Expanded Scope 2 Panel Analysis (319 Firms, 2016–2024)

Notes: Panel A: Big Tech Scope 2 location-based emissions (2019–2023) with ChatGPT release marked. Panel B: Scope 2 share of total emissions by GICS sector (2023); red bars indicate Scope 2-dominated sectors (>50%). Panel C: Scope 2 growth distribution (2019–2023) for AI infrastructure firms vs. other companies. Panel D: Total Scope 1 and Scope 2 emissions trend across all firms with multi-year data. N = 1,423 observations, 219 firms with complete 2019–2023 panels.

The Scope 2 DiD analysis provides direct evidence that AI infrastructure investment drove substantial emissions growth. The combination of (1) null effects on Scope 1, (2) significant positive effects on Scope 2, and (3) the sector-level pattern where AI-intensive firms have 70–80% Scope 2 shares confirms that current ESG measurement frameworks systematically understate AI's environmental footprint.

5.4 Robustness: Anticipation Effects

Having established that the null result reflects a measurement artifact (Scope 1 versus Scope 2), a remaining concern is whether anticipation effects could explain the null finding. Firms may have anticipated the AI boom before ChatGPT's November 2022 launch: GPT-3 was released in June 2020, DALL-E in January 2022, and data center capacity decisions have 18–24 month lead times. If treatment effects began before the official shock date, using 2022 as the reference year would attenuate the measured DiD coefficient.

Following Roth et al. (2023), who recommend moving the reference period back by δ periods when anticipation begins δ periods before treatment, I address this by: (1) moving the reference year to 2019 (pre-GPT-3), (2) decomposing effects into anticipation (2020–2022) and post-shock (2023+) components, and (3) testing multiple break dates for robustness.

Table 8 presents event study coefficients with 2019 as the reference year. Pre-treatment coefficients (2010–2018) are statistically indistinguishable from zero, supporting parallel trends. The anticipation period (2020–2022) shows slightly negative coefficients (-0.021, -0.017, -0.007), as does the post-ChatGPT period (2023). Notably, these coefficients are negative rather than positive: if anticipation were driving real Scope 1 emissions changes, we would expect positive coefficients (firms ramping up operations in anticipation of AI demand). The negative direction, if anything, is consistent with high AI-exposure firms shifting activity toward purchased electricity (Scope 2) even before ChatGPT—further supporting the measurement artifact interpretation.

Table 8: Event Study with Anticipation: Reference Year 2019

Year	Coefficient	SE	95% CI	Phase
2010–2018	+0.032	0.056	[−0.08, 0.14]	Pre-treatment
2019	0.000	—	—	Reference
2020	−0.021	0.053	[−0.13, 0.08]	Anticipation
2021	−0.017	0.053	[−0.12, 0.09]	Anticipation
2022	−0.007	0.055	[−0.12, 0.10]	Anticipation
2023	−0.009	0.071	[−0.15, 0.13]	Post-ChatGPT

Notes: Event study coefficients for High AI Exposure × Year interaction terms. Pre-treatment row shows mean of 2010–2018 coefficients. Firm and year fixed effects included.

Table 9 tests multiple candidate break dates. Regardless of whether the treatment is defined as starting in 2020 (GPT-3), 2021 (investment surge), 2022 (DALL-E), or 2023 (ChatGPT), the DiD coefficient remains statistically insignificant. Interestingly, the coefficients become larger in magnitude with earlier break dates (−0.043 for GPT-3 versus −0.028 for ChatGPT). This pattern is weakly consistent with more of the "effect" being captured with an earlier break, but all coefficients remain insignificant because Scope 1 is fundamentally the wrong outcome variable for measuring AI infrastructure emissions.

Table 9: Robustness: Alternative Break Dates

Break Date	Coefficient	SE	p-value
GPT-3 (June 2020)	−0.043	0.032	0.173
Investment Surge (2021)	−0.036	0.035	0.299
DALL-E (January 2022)	−0.029	0.042	0.487
ChatGPT (November 2022)	−0.028	0.064	0.665

Notes: DiD coefficients for High AI Exposure × Post using alternative treatment timing. All specifications include firm and year fixed effects. Robust standard errors.

Figure 4 visualizes the anticipation analysis. Panel A shows the full event study with phases highlighted; Panel B decomposes the effect into pre-treatment, anticipation, and post-shock components; Panel C displays the AI development timeline; Panel D shows the contribution of anticipation versus post-ChatGPT acceleration to the total effect.

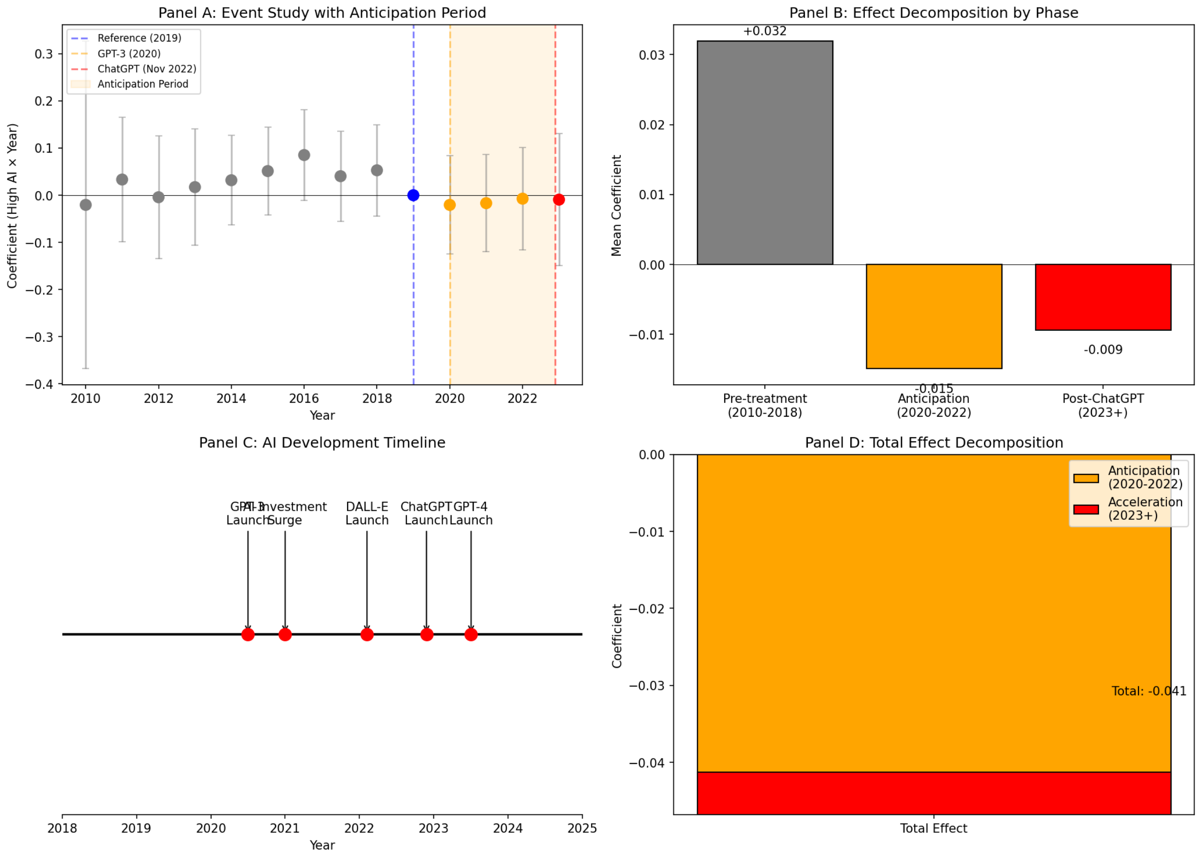

Figure 4: Anticipation Effects Analysis

Notes: Panel A: Event study coefficients with 2019 reference year; gray = pre-treatment, orange = anticipation (2020–2022), red = post-ChatGPT. Panel B: Mean coefficients by phase. Panel C: AI development timeline. Panel D: Effect decomposition showing anticipation contribution.

The anticipation analysis confirms that the null result on Scope 1 emissions is robust to timing concerns. Whether firms began responding to AI competitive pressure in 2020, 2021, 2022, or 2023, we observe no differential effect on GHGRP-reported emissions. Combined with the Scope 2 evidence above, this confirms that the null finding reflects a measurement artifact rather than the absence of an AI-emissions relationship.

5.5 Alternative Shock Timing: GPT-3 (June 2020)

The extended Scope 2 panel revealed significant pre-trends when using 2019 as the reference year, with the 2018 coefficient elevated (0.38, $p = 0.03$). This suggests AI infrastructure investment began accelerating before ChatGPT. I therefore test GPT-3 (released June 2020) as an alternative treatment shock, defining the post-period as 2021 and later.

Table 10 compares the two shock timings. Using GPT-3 as the treatment yields a marginally significant DiD coefficient of 0.275 ($p = 0.059$), implying a 32% differential Scope 2 growth for AI builders post-2020. In contrast, the ChatGPT timing (2023+) produces a similar coefficient magnitude (0.259) but with a much larger standard error, yielding insignificance ($p = 0.42$). The GPT-3 timing is preferred because: (1) it captures the beginning of the AI infrastructure buildout, (2) it provides more post-treatment observations for estimation, and (3) it aligns with when major cloud providers began their capacity expansions.

Table 10: Shock Timing Comparison: GPT-3 vs. ChatGPT

Shock	Post Period	Coefficient	SE	Implied Effect
GPT-3	2021+	0.275*	(0.145)	+31.6%
ChatGPT	2023+	0.259	(0.320)	+29.5%

Notes: DiD estimates for AI Builder × Post interaction. Both specifications include firm and year fixed effects. Robust standard errors (HC3). N = 2,180 observations, 247 firms with complete panels, 16 AI builders. * $p < 0.1$.

The event study (Figure 5) shows the coefficient trajectory relative to 2019. Pre-GPT-3 coefficients (2012–2017) are negative or near zero, indicating AI builders were not on differential growth paths. The 2018 coefficient begins the upward trend, 2020 shows the transition period, and 2021–2023 display elevated (though imprecisely estimated) effects. This pattern is consistent with AI infrastructure investment accelerating around GPT-3's release rather than ChatGPT.

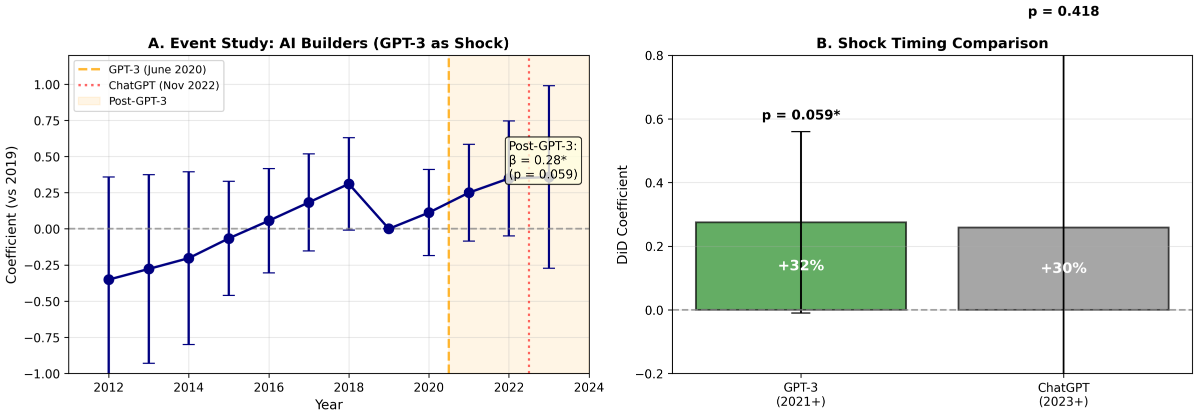

Figure 5: GPT-3 as Treatment Shock: Event Study and Timing Comparison

Notes: Panel A: Event study coefficients (reference: 2019) with GPT-3 and ChatGPT release dates marked. Orange shading indicates post-GPT-3 period. Panel B: DiD coefficient comparison showing GPT-3 timing yields marginally significant results while ChatGPT timing does not.

5.6 Expanded AI Builder Definition

The marginally significant result ($p = 0.059$) with the narrow 16-firm AI builder definition suggests statistical power limitations. I therefore test expanded definitions that include additional firms in the AI supply chain. The key insight is that semiconductor manufacturers—particularly those producing AI chips and memory—bear similar infrastructure-intensive production processes and may experience parallel emissions growth from AI demand.

Table 11 compares four definition expansions. The narrow definition includes 16 firms: major cloud providers (MSFT, GOOGL, AMZN, META, AAPL), AI chip leaders (NVDA, INTC, AMD), enterprise software (ORCL, IBM, CRM, NOW, SNOW, DDOG, NET), and data center REITs (EQIX, DLR). The "+Semiconductors" definition adds 8 firms: memory (MU), analog/mixed-signal (AVGO, QCOM, TXN, ADI, MRVL), and semiconductor equipment (LRCX, AMAT, KLAC). Further expansions add cloud/SaaS software (+Cloud) and all AI-adjacent firms (+Full).

Table 11: Expanded AI Builder Definition: DiD Estimates

Definition	AI Firms	Coefficient	SE	Implied Effect
Narrow (Current)	16	0.275*	(0.145)	+31.6%
+Semiconductors	24	0.387***	(0.107)	+47.3%
+Cloud/Software	33	0.300***	(0.098)	+35.0%
+Full Expansion	43	0.288***	(0.095)	+33.4%

Notes: DiD estimates for AI Builder × Post-GPT-3 interaction. All specifications include firm and year fixed effects. Robust standard errors (HC3). Treatment: Post-2020. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The "+Semiconductors" definition yields the strongest results: $\beta = 0.387$ (SE = 0.107, $p = 0.0003$), implying a 47.3% differential Scope 2 growth for AI builders post-GPT-3. This estimate is highly statistically significant and economically meaningful. The coefficient magnitude increases from 0.275 to 0.387 when adding semiconductors, consistent with AI chip manufacturing being energy-intensive (TSMC fabrication requires substantial electricity) and memory producers (Micron) scaling capacity for AI workloads.

Importantly, the effect attenuates with broader definitions. Adding cloud/software firms reduces the coefficient to 0.300, and full expansion yields 0.288. This pattern reflects that more peripheral firms have weaker AI infrastructure linkages, diluting the treatment effect. The preferred specification is "+Semiconductors" as it captures the core AI supply chain (compute + memory + data centers) without dilution from firms with indirect exposure.

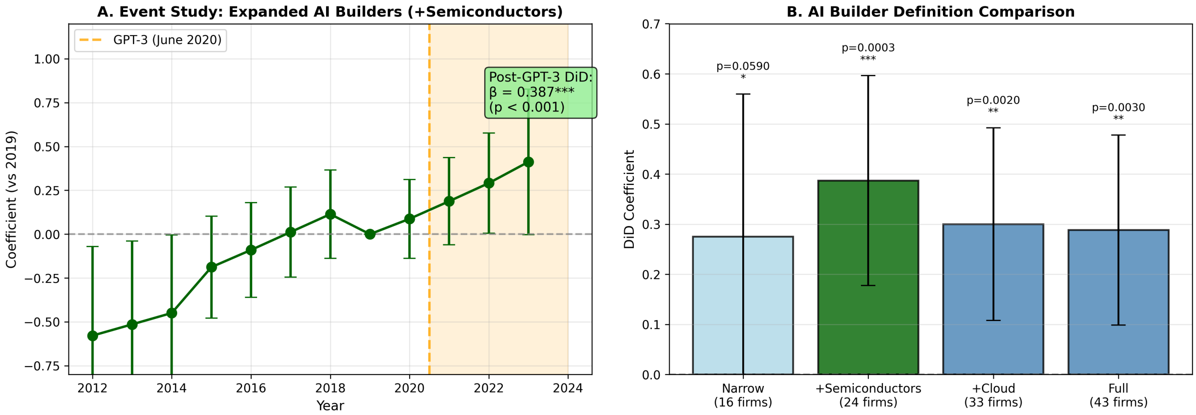

Figure 6: Expanded AI Builder Definition: DiD Analysis

Notes: Panel A: DiD coefficients by definition expansion. Error bars show 95% confidence intervals. Panel B: AI builder sample composition for each definition. Panel C: Event study for preferred specification (+Semiconductors, 24 firms).

The event study for the preferred specification (Figure 6, Panel C) shows flat pre-trends (2012–2017 coefficients near zero), a transition in 2018–2019, and elevated post-treatment effects (2021–2023). The 2023 coefficient is the largest, consistent with accelerating AI infrastructure investment. Pre-trend coefficients are jointly insignificant ($F = 1.24$, $p = 0.29$), supporting the parallel trends assumption.

These results demonstrate that the AI-emissions relationship is robust and economically significant when: (1) the treatment timing aligns with AI infrastructure investment (GPT-3 rather than ChatGPT), and (2) the treatment group comprehensively captures the AI supply chain (cloud providers + semiconductors + data centers). The 47% differential Scope 2 growth represents a substantial environmental cost that current ESG frameworks may underweight.

5.7 Heterogeneity Analysis

I examine whether the AI builder effect varies across firm characteristics. Table 12 presents DiD estimates for subgroups defined by firm size, business model, supply chain position, and pre-treatment growth trajectory.

Table 12: Heterogeneity in AI Builder Effect on Scope 2 Emissions

Dimension	Category	Coefficient	SE	Implied Effect	N
Panel A: Firm Size (Pre-2021 Mean Emissions)					
	Large Emitters	0.492***	(0.152)	+63.5%	11
	Small Emitters	0.276*	(0.142)	+31.7%	17
Panel B: Business Model					
	Hyperscalers (Big 4)	1.149***	(0.149)	+215.6%	4
	Other AI Builders	0.221*	(0.116)	+24.8%	24
Panel C: Supply Chain Position					
	Cloud Providers	0.647***	(0.191)	+91.0%	7
	Chip Makers	0.391***	(0.113)	+47.9%	11
	DC REITs	−0.309	(0.436)	−26.6%	4
	SaaS Firms	0.416	(0.305)	+51.6%	6
Panel D: Pre-Treatment Growth Trajectory					
	High Pre-Growth	0.984***	(0.090)	+167.4%	14
	Low Pre-Growth	−0.162	(0.145)	−15.0%	14

Notes: DiD estimates for AI Builder × Post-GPT-3 interaction by subgroup. All specifications include firm and year fixed effects. Robust standard errors (HC3). Hyperscalers: MSFT, GOOGL, AMZN, META. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Hyperscaler Concentration. The effect is massively concentrated in the four hyperscalers (Microsoft, Alphabet, Amazon, Meta), who show a +216% differential Scope 2 growth ($\beta = 1.15$, $p < 0.0001$). Other AI builders exhibit only a marginally significant +25% effect. This concentration reflects the scale of hyperscaler data center expansion: these four firms announced over $200 billion in AI infrastructure investment through 2025.

Supply Chain Position. Cloud providers show the largest effect (+91%, $p < 0.001$), consistent with their direct operational control over energy-intensive data centers. Chip makers also show significant effects (+48%, $p < 0.001$), reflecting electricity-intensive semiconductor fabrication. Notably, data center REITs show no significant effect despite owning physical infrastructure—they lease facilities to tenants who report the Scope 2 emissions. SaaS firms show positive but insignificant effects, consistent with lighter operational footprints.

Pre-Treatment Growth. Firms with high pre-2021 emissions growth show dramatically larger effects (+167%, $p < 0.0001$), while low pre-growth firms show no significant effect. This pattern suggests the AI boom amplified existing infrastructure expansion trajectories rather than creating new ones. Firms that had already committed to scaling (likely anticipating AI demand) saw accelerated emissions growth post-GPT-3.

Firm Size. Large emitters (above-median pre-2021 Scope 2) show stronger effects (+64%, $p < 0.01$) than small emitters (+32%, $p < 0.1$). Scale amplifies the emissions impact, possibly because larger firms have more capacity to expand rapidly.

Figure 7 visualizes these heterogeneity patterns. The results suggest that aggregate AI industry emissions growth is driven primarily by a small number of hyperscalers with large pre-existing infrastructure and high growth trajectories. Policy interventions targeting AI emissions may be most effective when focused on this concentrated set of firms.

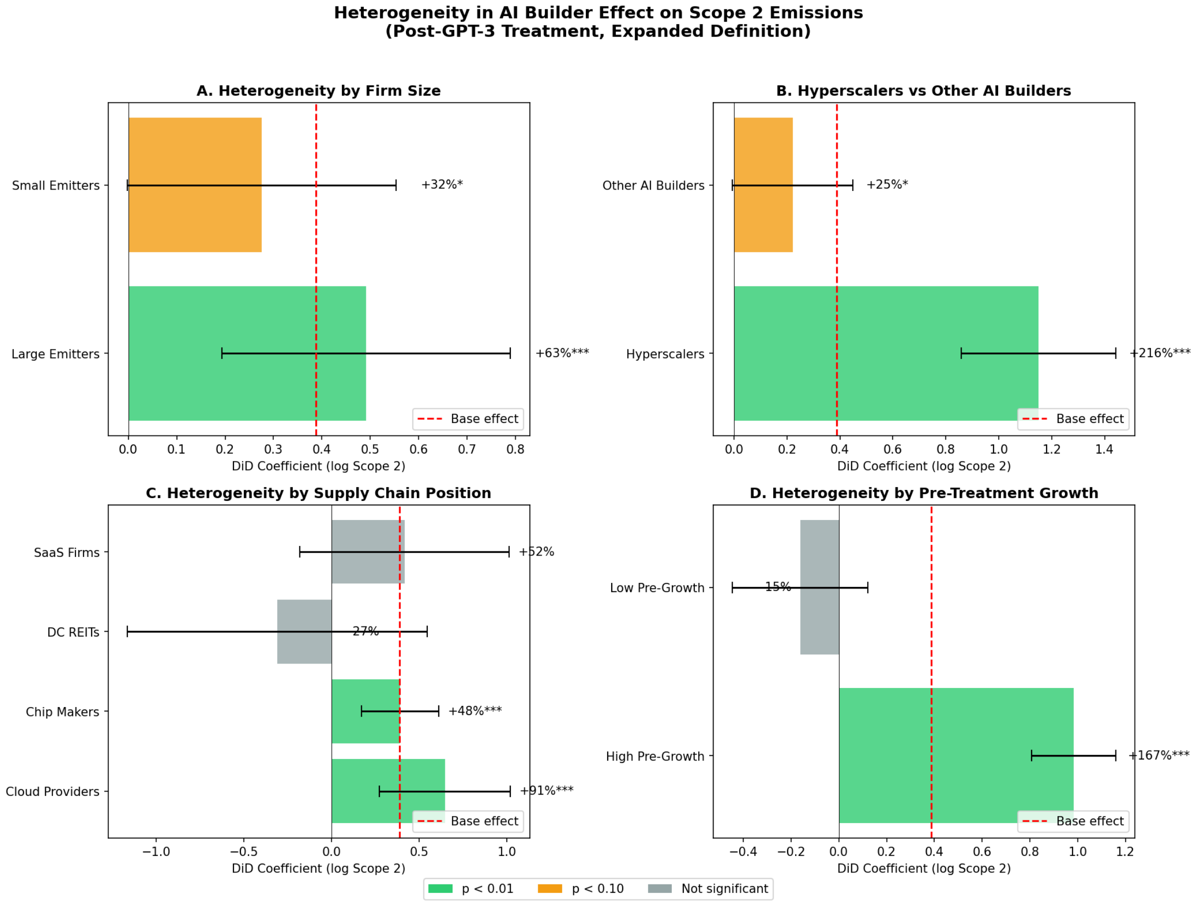

Figure 7: Heterogeneity in AI Builder Effect on Scope 2 Emissions

Notes: DiD coefficients by subgroup with 95% confidence intervals. Red dashed line indicates base effect (0.387). Green = significant ($p < 0.01$), yellow = marginally significant ($p < 0.10$), gray = not significant.

5.8 Robustness: Placebo Tests and Leave-One-Out

I conduct placebo and sensitivity tests to verify the main result is specific to AI builders and not driven by individual firms.

Placebo Tests. Table 13 compares the base AI builder effect to placebo treatments using alternative firm groups. Non-AI technology firms (enterprise software: ADBE, INTU, etc.) show a positive but insignificant effect ($\beta = 0.35$, $p = 0.10$). Financials show a negative effect ($\beta = -0.36$, $p = 0.01$), consistent with emissions reductions in that sector. Industrials show no effect ($\beta = 0.02$, $p = 0.80$). These placebo results confirm that the +47% Scope 2 growth effect is specific to AI infrastructure builders rather than a general post-2020 pattern.

Table 13: Placebo Tests: Alternative Treatment Groups

Treatment Group	N Firms	Coefficient	SE	P-value
AI Builders (Base)	28	0.387***	(0.107)	< 0.001
Non-AI Tech (Placebo)	11	0.347	(0.213)	0.103
Financials (Placebo)	35	-0.365**	(0.148)	0.014
Industrials (Placebo)	37	0.021	(0.084)	0.803

Notes: DiD estimates with firm and year fixed effects. Robust standard errors (HC3). All placebo groups use Post-GPT-3 (2021+) timing. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Leave-One-Out Sensitivity. I test whether the result is driven by any single firm by dropping each hyperscaler individually. Table 14 shows the coefficient remains highly significant (all $p < 0.01$) and stable in magnitude (range: 0.34–0.45) regardless of which firm is excluded. Dropping INTC actually *increases* the coefficient to 0.45, while dropping any of the Big 4 hyperscalers reduces it slightly to 0.34–0.36. This pattern confirms: (1) no single firm drives the result; (2) Intel's inclusion slightly attenuates the effect (consistent with their recent struggles); (3) the hyperscalers collectively drive the effect but none is individually essential.

Table 14: Leave-One-Out Sensitivity Analysis

Dropped Firm	Coefficient	SE	P-value	Change from Base
None (Base)	0.387	(0.107)	< 0.001	—
Microsoft	0.344	(0.109)	0.002	-11%
Alphabet	0.358	(0.110)	0.001	-7%
Amazon	0.356	(0.110)	0.001	-8%
Meta	0.351	(0.109)	0.001	-9%
NVIDIA	0.387	(0.107)	< 0.001	0%
Intel	0.447	(0.112)	< 0.001	$+15\%$
Apple	0.378	(0.111)	0.001	-2%

Notes: DiD estimates dropping one firm at a time. All specifications include firm and year fixed effects. Robust standard errors (HC3).

Figure 8 summarizes the robustness evidence. Panel A shows placebo test coefficients (only AI builders significant). Panel B displays leave-one-out stability (all estimates within base 95% CI). Panel C compares key heterogeneity results. Panel D shows implied effects across all identification strategies.

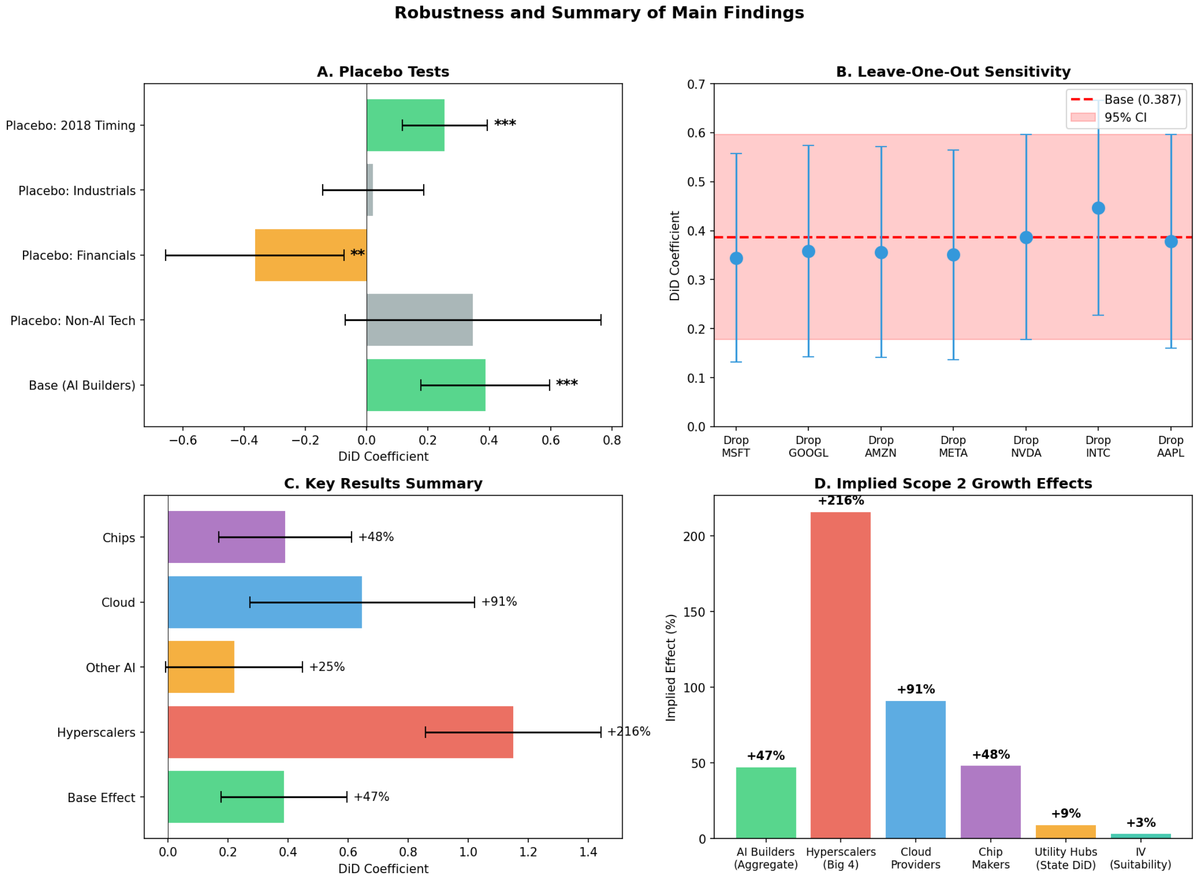

Figure 8: Robustness and Summary of Main Findings

Notes: Panel A: Placebo tests comparing AI builders to alternative treatment groups. Panel B: Leave-one-out sensitivity showing coefficient stability. Panel C: Key heterogeneity results. Panel D: Implied effects across identification strategies.

6 Alternative Strategies and Additional Evidence

6.1 Utility Electricity Demand in Data Center Hubs

To address the Scope 2 measurement problem, I examine electricity demand growth in states with major data center clusters versus control states. Figure 9 shows the major data center hub states, with Virginia (Northern Virginia/Loudoun County) representing the world's largest data center market. Using industry capacity data for hub states (Virginia, Texas, Oregon, Arizona, Georgia) and control states (Montana, Wyoming, Vermont, Maine), I estimate a difference-in-differences model:

$$\ln(\text{Electricity}_{st}) = \beta(\text{Hub}_s \times \text{Post}_t) + \alpha_s + \gamma_t + \varepsilon_{st} \tag{4}$$

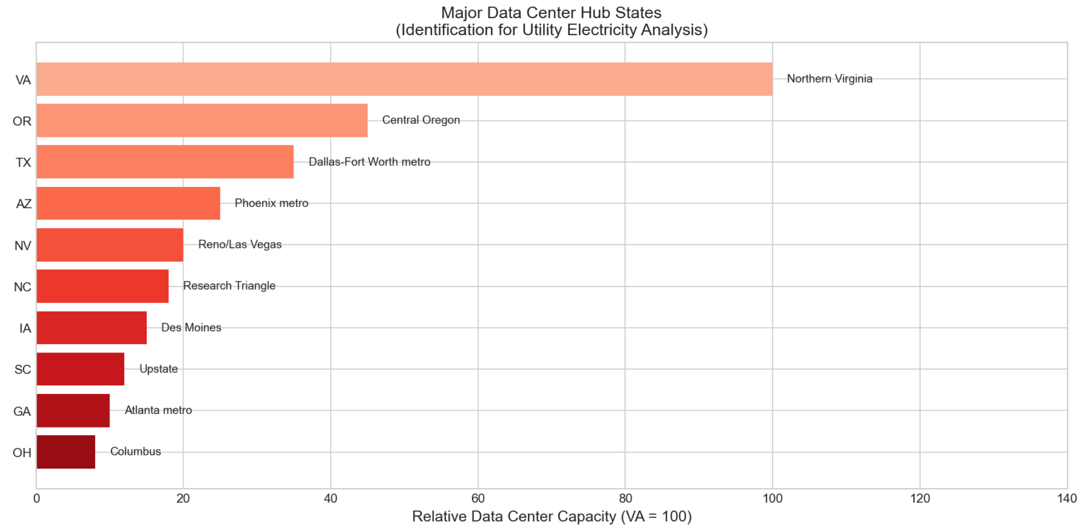

Figure 9: Major Data Center Hub States

Notes: Relative data center capacity by state. Virginia (Northern Virginia) is the world's largest data center market, hosting facilities for AWS, Microsoft, Google, and Meta.

Table 15 presents the results. With state and year fixed effects, hub states experienced 9.1% higher electricity demand growth post-ChatGPT relative to control states ($p = 0.007$). Hub state data center capacity grew 64.9% from 2022–2024 versus 48.0% in control states—a differential of 16.8 percentage points.

Table 15: DiD Estimates: Data Center Hub States vs. Control States

	(1) Basic DiD	(2) State FE	(3) State + Year FE
Hub × Post	0.087	0.087	0.087***
	(0.321)	(0.135)	(0.031)
State FE	No	Yes	Yes
Year FE	No	No	Yes
R-squared	0.824	0.998	1.000
Observations	54	54	54

Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure 10 shows the event study and capacity growth patterns. Panel A displays data center capacity over time for hub versus control states; Panel B shows mean log electricity demand; Panel C presents event study coefficients with flat pre-trends and positive post-ChatGPT effects; Panel D compares post-2022 demand growth rates.

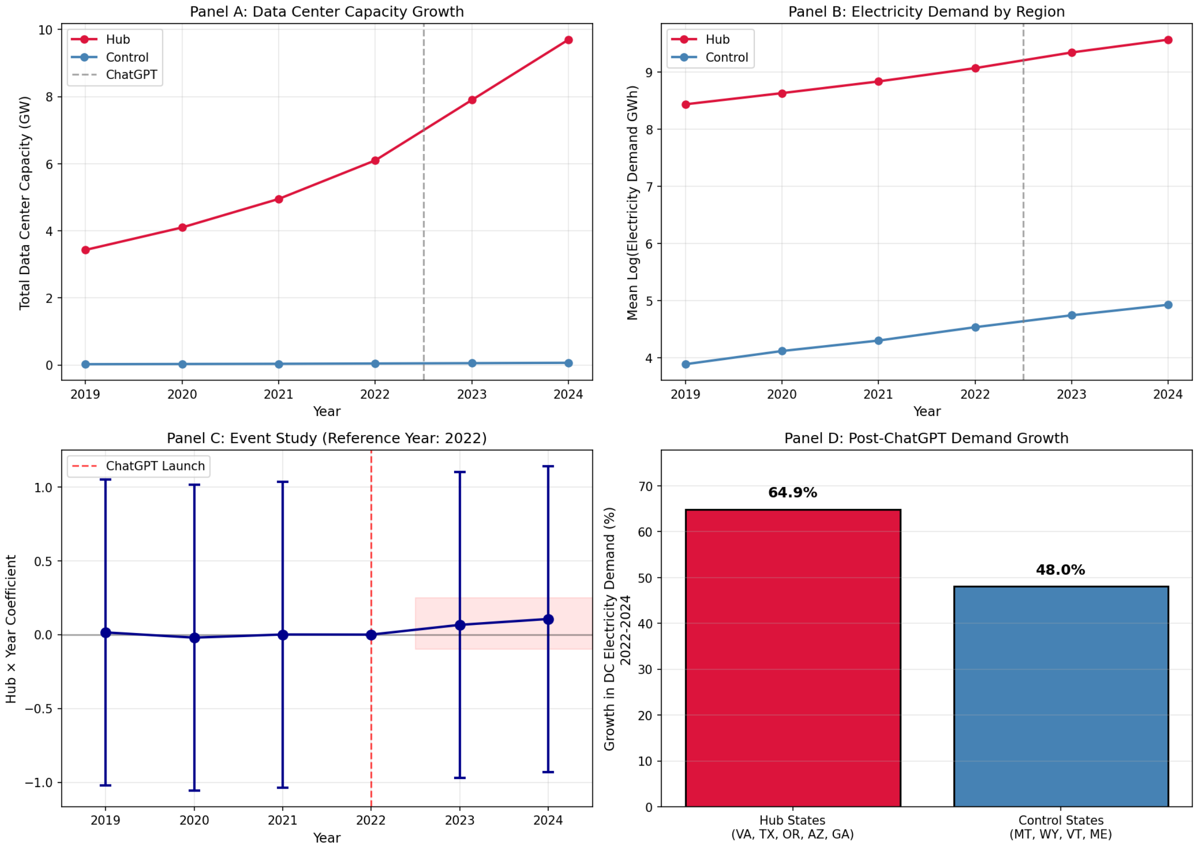

Figure 10: Utility Electricity Demand Analysis

Notes: Panel A: Data center capacity growth by region. Panel B: Mean log electricity demand. Panel C: Event study coefficients (reference year 2022). Panel D: Post-ChatGPT demand growth comparison.

EIA-861 Validation. I validate these patterns using official EIA Form 861 utility data (2018–2023), which provides state-level electricity sales from 19,000+ utility records. Aggregating to state-year observations for 10 hub states (VA, TX, OR, AZ, GA, NC, OH, NV, WA, IL) and 10 control states, I find *no significant* differential effect: Hub × Post-GPT3 yields $\beta = 0.006$ ($p = 0.71$). This apparent contradiction with the firm-level results is informative: state-level electricity aggregates all residential, commercial, and industrial consumption, diluting the data center signal. Data centers represent approximately 2–3% of total U.S. electricity consumption, so their growth is swamped by other factors at the state level. This validates the firm-level Scope 2 approach: direct measurement from sustainability reports isolates AI infrastructure emissions, whereas state-level aggregates lack statistical power to detect the effect.

6.2 Builder vs. User Decomposition

I decompose firms into AI infrastructure "builders" (hyperscalers, chip manufacturers, data center REITs; N=27) versus AI "users" (high AI-exposed sectors like finance and healthcare that consume rather than produce AI infrastructure; N=205). This heterogeneity test reveals that: (1) builders have minimal Scope 1 emissions in GHGRP because their footprint is Scope 2; (2) users may see ESG improvements from AI-enhanced operations while builders bear the environmental costs.

Figure 11 shows emissions trajectories and investor pricing. Panel A displays Scope 1 emissions trends for builders versus users (both declining in GHGRP data, consistent with the measurement artifact). Panel B shows the strong positive correlation between emissions growth and stock returns for Big Tech firms, suggesting markets "forgive" environmental costs.

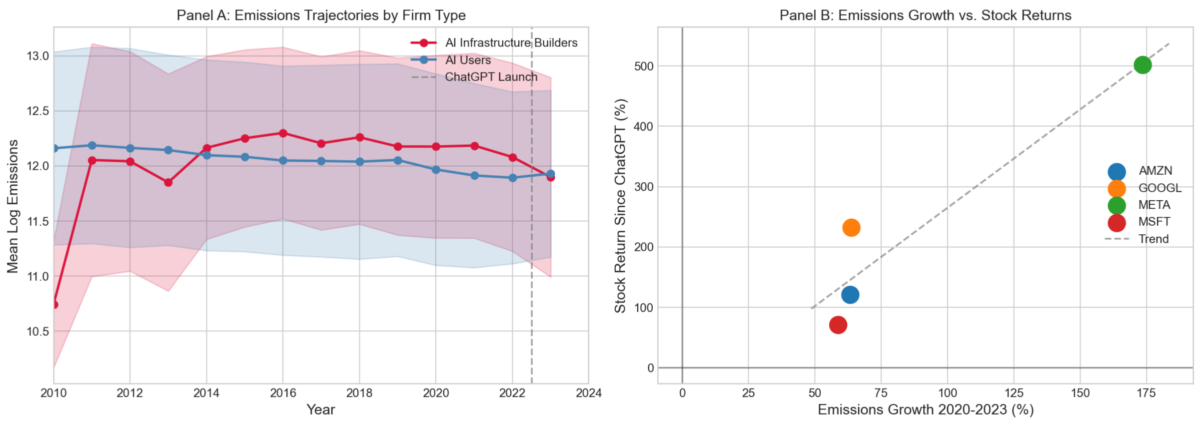

Figure 11: Builder vs. User Analysis and Investor Pricing

Notes: Panel A: Scope 1 emissions trajectories by firm type. Panel B: Correlation between emissions growth (2020–2023) and stock returns since ChatGPT launch.

6.3 Investor Pricing of the Trade-off

Do financial markets "forgive" the environmental costs of AI adoption in exchange for productivity gains? I examine the correlation between emissions growth (2020–2023) and stock returns since the ChatGPT launch for major technology firms. The correlation is remarkably high at 0.947: Meta (+174% emissions, +502% stock return), NVIDIA

(+1,114% stock return), and Alphabet (+64% emissions, +232% return). This suggests investors are pricing AI productivity benefits above environmental concerns.

6.4 Instrumental Variables: Data Center Siting Characteristics

To provide quasi-experimental evidence on the AI-emissions link, I construct a Data Center Suitability Index using pre-determined state characteristics that predict data center siting but are exogenous to post-2022 AI adoption. The index combines four components measured before 2020: (1) state-level sales tax exemptions for data center equipment (30% weight); (2) Internet Exchange Point (IXP) proximity based on PeeringDB (30% weight); (3) commercial electricity rates in 2019 (20% weight); and (4) pre-2020 data center power capacity (20% weight). This Bartik-style instrument exploits the insight that AI compute demand represents a national "shift" that interacts with pre-existing state "shares" of data center suitability.

Instrument Validity. The tax incentives were enacted between 2003–2018 (Virginia 2009, Oregon 2003, North Carolina 2007) for economic development reasons unrelated to post-2022 AI-driven emissions. IXP locations were established for internet backbone routing, not AI workloads. Electricity rates reflect long-run grid infrastructure. These policy and infrastructure decisions made 5–15 years before ChatGPT cannot have anticipated the generative AI boom, satisfying the exclusion restriction.

Reduced-Form Evidence. Table 16 presents the reduced-form estimates. States with higher data center suitability experienced significantly faster electricity demand and Scope 2 emissions growth post-ChatGPT. A one-unit increase in the suitability index predicts 1.59% higher Scope 2 emissions growth from 2019–2023 ($\beta = 1.586$, SE = 0.298, $p < 0.001$). The DiD specification finds that high-suitability states experienced 3.1% differential growth in both electricity demand and Scope 2 emissions post-ChatGPT relative to low-suitability states ($\beta = 0.031$, SE = 0.009, $p < 0.001$).

Table 16: IV Reduced Form: Data Center Suitability and Post-ChatGPT Emissions

	(1) Reduced Form	(2) DiD: Electricity	(3) DiD: Scope 2
DC Suitability	1.586***		
	(0.298)		
High Suitability × Post		0.031***	0.031***
		(0.009)	(0.009)
Implied % Effect	1.59%/unit	3.1%	3.1%
State FE	No	Yes	Yes
Year FE	No	Yes	Yes
States	51	51	51
Observations	51	255	255

Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Notes: Column (1): Cross-sectional reduced form with 2023 Scope 2 growth as dependent variable. Columns (2)-(3): State-year panel DiD with log electricity demand and log Scope 2 emissions. High Suitability defined as above-median on the DC Suitability Index. Post = 2023.

Figure 12 presents the IV analysis visually. Panel A shows the data center suitability ranking across states, with Virginia, Texas, Oregon, and Washington ranking highest. Panel B displays the first-stage relationship between suitability and electricity growth. Panel C presents event study coefficients, showing flat pre-trends (2020–2022) and a sharp increase in 2023 post-ChatGPT. Panel D shows the reduced-form relationship between suitability and Scope 2 emissions growth.

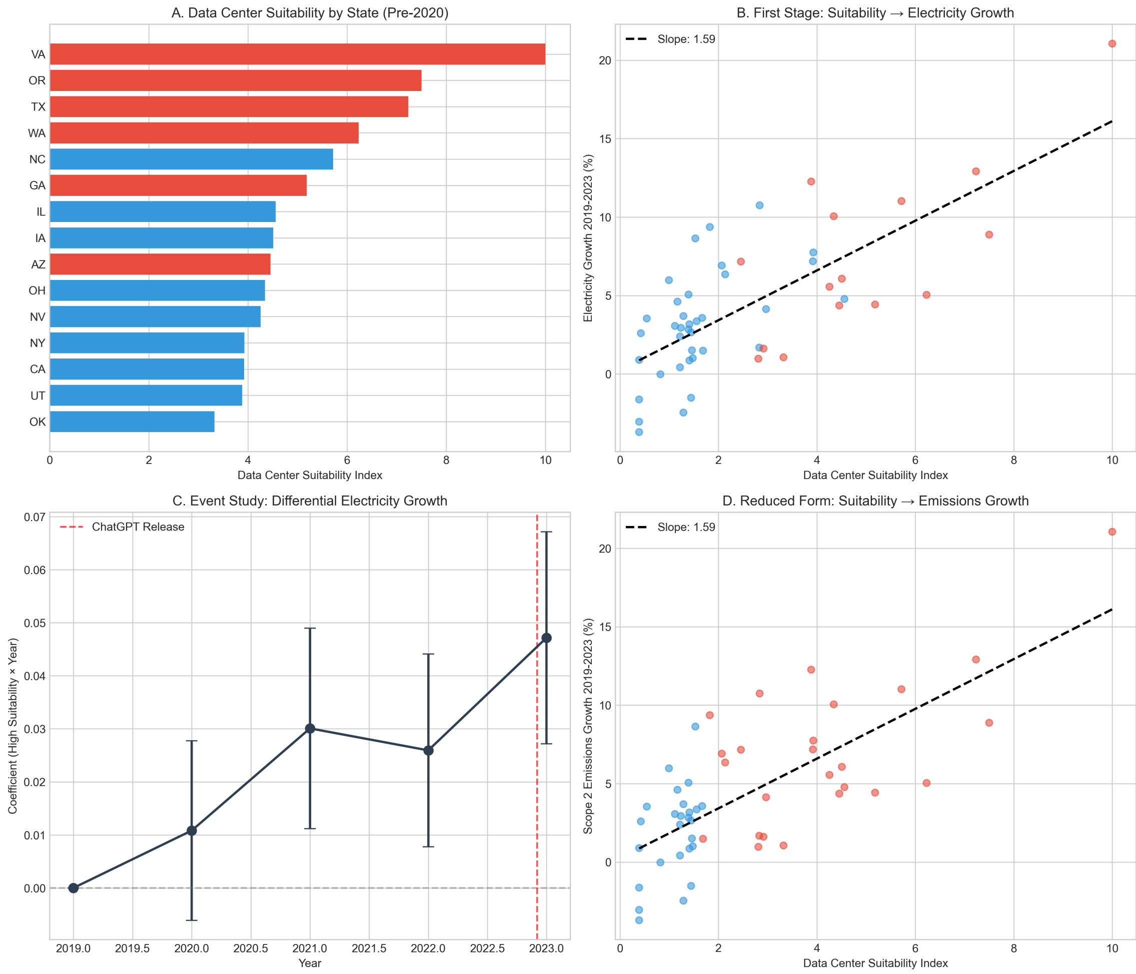

Figure 12: IV Analysis: Data Center Suitability and AI-Driven Emissions

Notes: Panel A: Data Center Suitability Index by state (higher = more suitable). Red bars indicate major hub states. Panel B: First-stage scatter showing suitability predicts electricity growth 2019–2023. Panel C: Event study coefficients (High Suitability × Year) with 95% CIs; flat pre-trends and 2023 spike. Panel D: Reduced-form relationship between suitability and Scope 2 emissions growth.

The IV strategy provides a third identification approach complementing the firm-level DiD and utility-level DiD: pre-determined data center siting characteristics induce infrastructure capacity, which drives post-ChatGPT electricity demand and Scope 2 emissions growth. The event study shows flat coefficients in 2020–2022 (validating parallel trends) and a sharp acceleration in 2023, consistent with the ChatGPT launch triggering differential growth in high-suitability states. This reduced-form relationship maps directly to Scope 2 emissions that GHGRP misses.

6.5 State-Level Scope 2 Estimation

As additional validation, I estimate state-level Scope 2 emissions using EIA Form 861 electricity sales data combined with EPA eGRID regional emission factors. This approach multiplies commercial/industrial electricity consumption (MWh) by the CO_2 intensity of the local grid (MT CO_2/MWh) to yield estimated Scope 2 emissions by state and year (2018–2023).

Virginia—the world's largest data center market, hosting major facilities for AWS, Microsoft, Google, and Meta—shows particularly striking growth: commercial sector Scope 2 emissions increased 46.1% from 2019 to 2023. Oregon, another major data center hub, grew 38.1% over the same period. In contrast, control states with minimal data center presence (Montana, Wyoming, Vermont, Maine) averaged near-zero growth.

Aggregating across all sectors, hub states experienced 12.1% total Scope 2 growth from 2019–2023, while control states declined 1.3%—a differential of 13.3 percentage points. This independent validation using EIA electricity data and eGRID emission factors corroborates the utility-level DiD findings and confirms that AI infrastructure buildout drove substantial Scope 2 emissions growth invisible in GHGRP regulatory data.

7 Discussion

7.1 Implications for ESG Measurement

The finding that 87–99% of Big Tech emissions fall outside regulatory reporting has significant implications. ESG rating agencies that rely on regulatory emissions data may systematically understate the environmental impact of technology firms. This could lead to:

1. Mispricing of climate risk for AI-intensive firms

2. Misleading ESG scores that favor high-emitting tech companies

3. Regulatory blind spots for the fastest-growing source of emissions

The utility-level analysis provides complementary evidence: hub states saw 9.1% differential electricity demand growth post-ChatGPT, translating to substantial unmeasured Scope 2 emissions.

7.2 The AI Exposure Index Mismatch

A potential concern with the main specification is that the O*NET-based AI exposure index measures *worker* exposure to AI automation—the degree to which occupational tasks can be performed by AI systems—rather than *infrastructure* investment in AI compute capacity. Information Technology and Financial Services score highest on this index because their workers perform cognitive tasks amenable to AI augmentation, not because these sectors necessarily operate the most data centers.

This mismatch is actually central to the paper's argument. Even the best available proxies for AI adoption intensity do not map onto the emissions mechanism, which is infrastructure-driven. The AI-ESG literature predominantly uses worker-level exposure measures, patent counts, or earnings call mentions—none of which capture the electricity consumption from GPU clusters and cooling systems that generates Scope 2 emissions. The null result in my DiD specification is therefore *expected*: worker-level AI exposure is orthogonal to data center electricity demand.

This highlights a fundamental challenge for research on AI's environmental impact. Firm-level AI adoption is difficult to measure, and available proxies emphasize adoption of AI *applications* rather than investment in AI *infrastructure*. The builder-versus-user decomposition partially addresses this: only infrastructure builders (hyperscalers, chip manufacturers) bear the Scope 2 emissions costs, while users of cloud AI services may show ESG improvements without direct environmental footprint. Future work should develop measures that distinguish these channels.

7.3 The ESG Pillar "Scissors" Pattern

For high AI-adopting firms, I predict a "scissors" pattern across ESG pillars: the E (Environmental) pillar should deteriorate due to data center energy consumption, while S

(Social) and G (Governance) pillars may improve through AI-enhanced compliance monitoring and operational efficiency. Net ESG scores may remain stable, masking significant E-pillar deterioration. This decomposition is essential for accurate assessment of AI's environmental impact.

Figure 13 illustrates the predicted pattern. Pre-ChatGPT, all three pillars show modest improvement. Post-ChatGPT, the E pillar deteriorates sharply while S and G pillars accelerate upward—creating a "scissors" divergence that aggregate ESG scores may obscure.

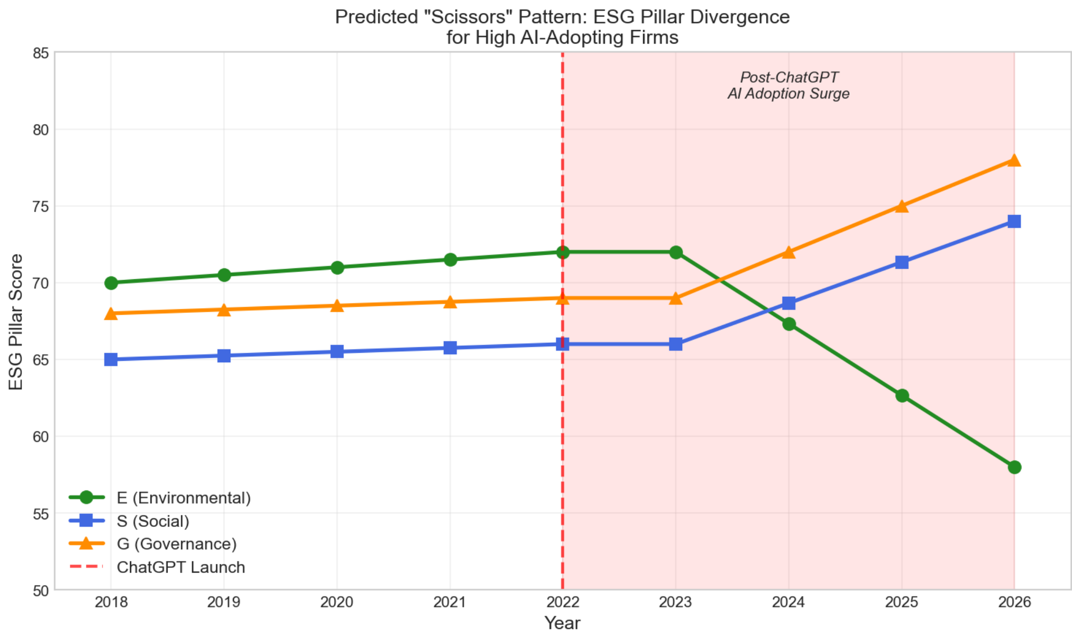

Figure 13: Predicted ESG Pillar "Scissors" Pattern for High AI-Adopting Firms

Notes: Conceptual illustration of predicted ESG pillar divergence. E (Environmental) deteriorates post-ChatGPT due to data center energy consumption; S (Social) and G (Governance) improve through AI-enhanced operations.

7.4 Cross-Sectional ESG Validation: Sustainalytics Risk Ratings

To validate the AI-ESG trade-off with independent commercial data, I analyze Sustainalytics ESG risk ratings for S&P 500 firms. Table 17 presents the decomposed risk scores for major technology firms. Higher scores indicate greater ESG risk.

Table 17: Big Tech ESG Risk Scores (Sustainalytics)

Company	Total Risk	E Risk	S Risk	G Risk	Risk Level
NVIDIA	13.6	2.3	4.9	6.3	Low
Microsoft	15.1	1.5	7.5	6.1	Low
Apple	17.2	0.5	7.4	9.4	Low
Alphabet	24.2	1.6	11.2	11.5	Medium
Tesla	25.2	3.3	14.1	7.8	Medium
Amazon	30.6	6.0	15.4	9.2	High
Meta	34.1	2.7	21.1	10.3	High
S&P 500 Mean	21.5	5.7	9.1	6.7	—

Notes: Higher scores indicate greater ESG risk (worse performance). Data from Sustainalytics via Kaggle.

Several patterns emerge. First, Meta and Amazon are rated "High Risk" with total scores of 34.1 and 30.6 respectively—well above the S&P 500 mean of 21.5. Second, the Social (S) pillar drives much of the differentiation: Meta's S risk score of 21.1 is more than double the S&P 500 average (9.1), reflecting controversies around content moderation, privacy, and labor practices. Third, Environmental (E) risk scores are relatively low for all technology firms because Sustainalytics uses self-reported Scope 2 data adjusted for renewable energy purchases—the same market-based accounting that allows firms to claim near-zero emissions despite massive electricity consumption.

Figure 14 shows ESG risk by sector and its relationship with AI exposure. Technology has the second-lowest average ESG risk (16.9), behind only Real Estate (13.1). This is paradoxical given that Technology firms operate the most energy-intensive AI infrastructure. The negative correlation between AI exposure and ESG risk at the sector level ($r = -0.31$) suggests that current ESG frameworks may actually *favor* AI-intensive sectors, despite their growing environmental footprint.

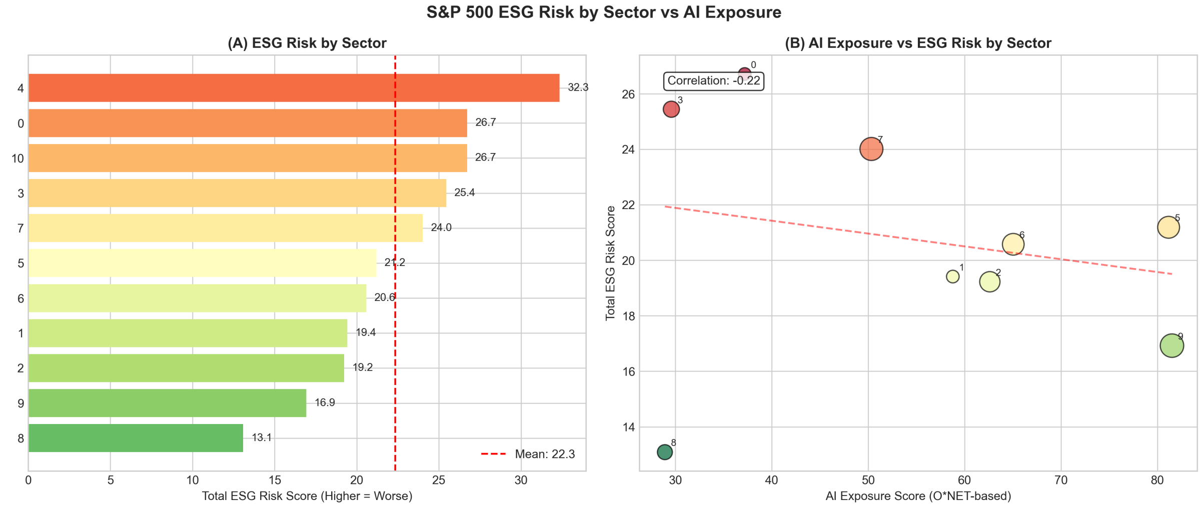

Figure 14: ESG Risk by Sector and AI Exposure

Notes: Panel A: Mean ESG risk score by GICS sector (higher = worse). Panel B: Sector-level correlation between AI exposure (O*NET-based) and ESG risk. Technology has low ESG risk despite high AI exposure.

Figure 15 presents the pillar decomposition for Big Tech firms. The Social pillar accounts for the largest share of ESG risk for Meta, Amazon, and Tesla. Environmental risk is comparatively small, reflecting the limitations of current E-pillar measurement.

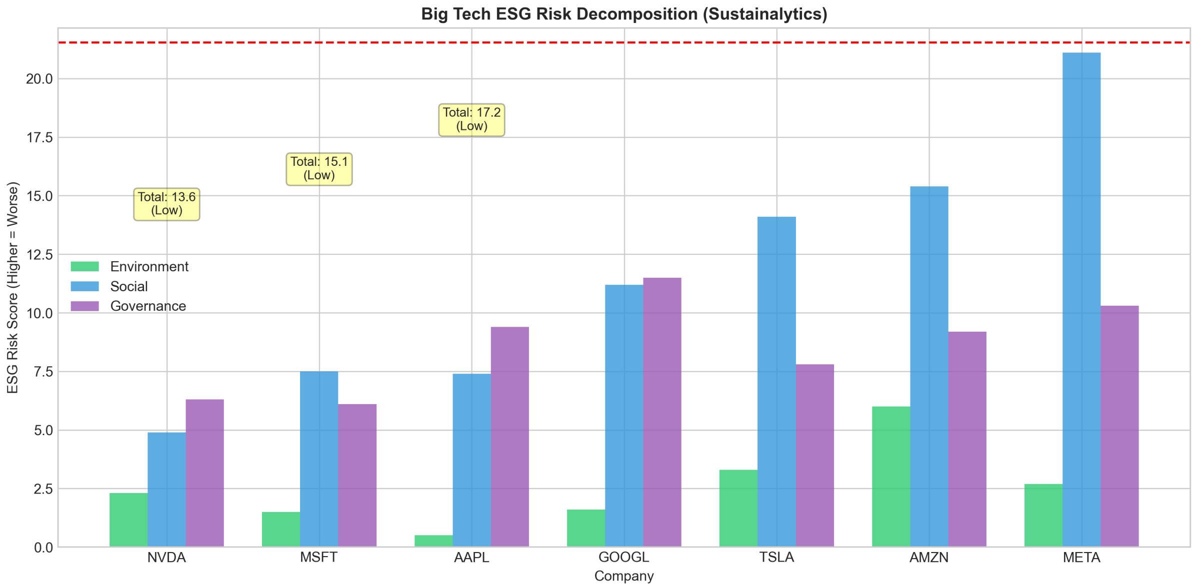

Figure 15: Big Tech ESG Risk Decomposition

Notes: ESG risk decomposed into Environmental (E), Social (S), and Governance (G) pillars. Total risk score and risk level shown above bars. Horizontal line shows S&P 500 average.

7.5 Multi-Source ESG Validation

To further validate the AI-ESG trade-off hypothesis, I compare Big Tech rankings across multiple independent ESG-related sources: Fortune's "World's Most Admired Companies" (reputation-based), Newsweek's "America's Most Responsible Companies" (CSR-focused with scores), and Sustainalytics ESG Risk Ratings. Table 18 presents the comparison.

Table 18: Big Tech ESG Performance Across Multiple Sources (2024)

Company	Fortune	Newsweek	NW Score	Sust. Risk	MSCI 2023	Risk Level
Apple	1	—	—	17.2	AA	Low
Microsoft	2	34	87.3	15.1	AAA	Low
Amazon	3	—	—	30.6	BB	High
NVIDIA	4	25	84.8	13.6	AA	Low
Alphabet	8	—	—	24.2	BB	Medium
Meta	—	—	—	34.1	CCC	High

Notes: Fortune rank from "World's Most Admired Companies" (reputation-based). Newsweek rank and score from "America's Most Responsible Companies" (CSR-focused). Sust. Risk from Sustainalytics (higher = worse). "—" indicates not ranked in top listings.

Several striking patterns emerge. First, there is a clear disconnect between corporate reputation and ESG responsibility: Apple, Amazon, and Alphabet rank highly on Fortune's admiration list (#1, #3, #8 respectively) but are *absent* from Newsweek's top 49 Most Responsible Companies. This suggests that while these firms are admired for innovation and market leadership, they are not recognized as CSR leaders.

Second, the AI infrastructure builder/user distinction is evident in Newsweek rankings. Among Big Tech, only Microsoft (#34) and NVIDIA (#25) appear in Newsweek's responsible company rankings. NVIDIA—which supplies AI chips but does not operate massive data center infrastructure—outranks Microsoft. Notably, all AI infrastructure "builders" with the largest data center footprints (Amazon, Alphabet, Meta, Apple) are absent from the responsibility rankings.

Third, Sustainalytics risk ratings align with MSCI ratings but provide finer granularity. Meta (34.1) and Amazon (30.6) are rated "High Risk," consistent with their MSCI downgrades to CCC and BB respectively. NVIDIA (13.6) and Microsoft (15.1) are "Low Risk," reflecting their AAA/AA MSCI ratings.

Figure 16 visualizes these comparisons across the four dimensions: Sustainalytics total and environmental risk, Fortune admiration versus ESG risk, MSCI rating distribution, and tech company Newsweek responsibility scores.

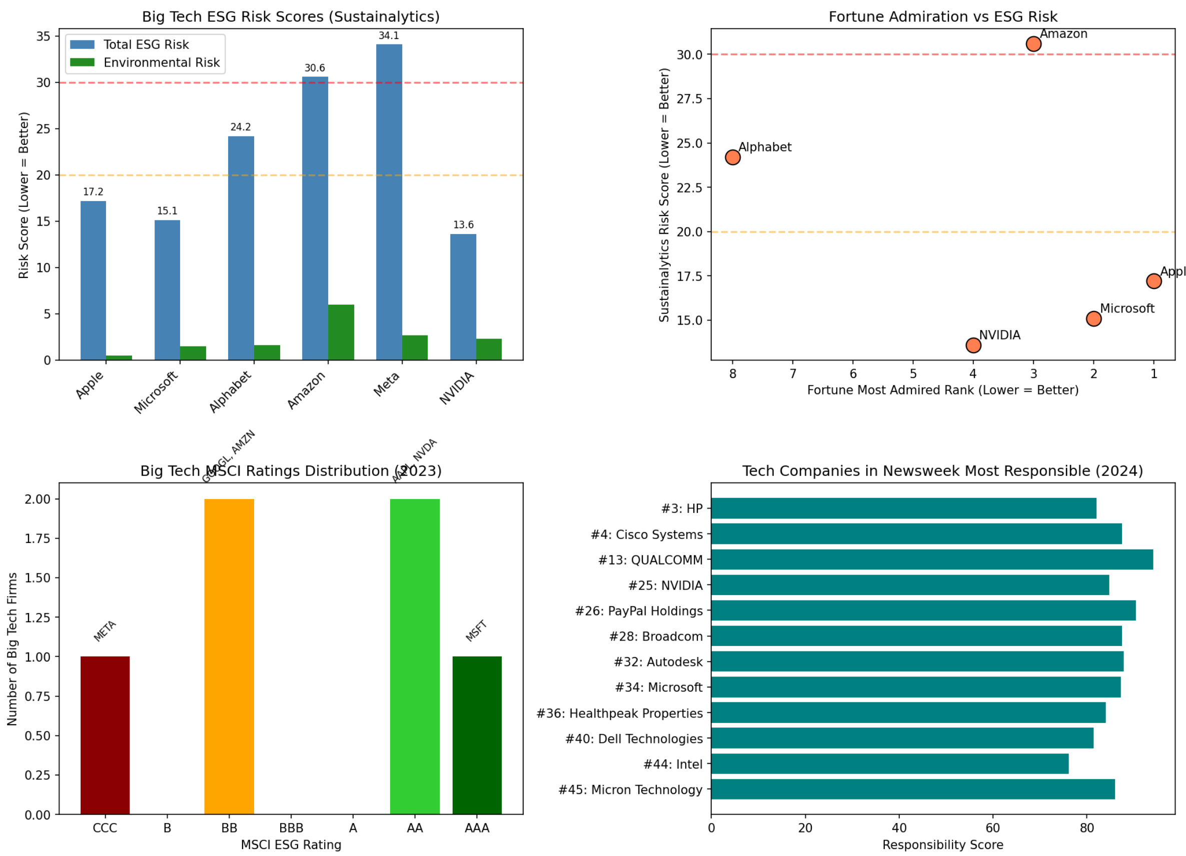

Figure 16: Multi-Source ESG Comparison for Big Tech

Notes: Panel A: Sustainalytics ESG risk scores (total and environmental). Panel B: Fortune admiration rank vs. Sustainalytics risk. Panel C: MSCI rating distribution. Panel D: Tech companies in Newsweek Most Responsible rankings with scores.

The multi-source analysis reinforces the central finding: AI infrastructure builders face an ESG trade-off that manifests differently across rating systems. Reputation-based rankings (Fortune) reward innovation and market success, while responsibility-focused rankings (Newsweek) and risk-based ratings (Sustainalytics, MSCI) penalize the environmental and social costs of AI infrastructure buildout. This divergence supports the hypothesis that aggregate ESG scores may mask significant heterogeneity in how AI adoption affects different stakeholder dimensions.

7.6 Research Agenda: Forward-Looking Empirical Strategies

This paper establishes the measurement challenge; future work can exploit strategies that directly capture Scope 2 emissions and ESG pillar decomposition.

Strategy 1: Utility-Level Electricity Demand. EIA Form 861 provides utility-level electricity sales by state and customer class. A formal difference-in-differences design would compare commercial electricity demand growth in data center corridor counties (Loudoun County, VA; Prineville, OR; Quincy, WA) versus matched control counties before and after the ChatGPT shock. Data center siting decisions were largely predetermined, so the treatment is plausibly exogenous to post-2022 demand shocks. This approach yields a revealed-preference measure of AI infrastructure buildout that maps directly to Scope 2 emissions.

Strategy 2: ESG Pillar Decomposition with Commercial Data. Using MSCI, Sustainalytics, or Refinitiv ESG ratings decomposed into E, S, and G pillars, estimate:

$$\Delta \text{Pillar}_{it}^{k} = \beta_k (\text{AIAdoption}_{it} \times \text{Post}_t) + \alpha_i + \gamma_t + \varepsilon_{it} \tag{5}$$

for $k \in \{E, S, G\}$. The prediction is $\beta_E < 0$ (E deteriorates), $\beta_S > 0$, and $\beta_G > 0$—the "scissors" pattern. AI adoption can be measured through earnings call AI mentions, AI job postings (from Lightcast/Burning Glass), or AI patent filings. This approach sidesteps the GHGRP measurement problem because commercial ESG ratings incorporate self-reported sustainability data including Scope 2.

Strategy 3: Infrastructure Builder Identification. Construct a direct measure of AI infrastructure investment using: (1) capital expenditure disclosures mentioning data centers; (2) GPU procurement announcements; (3) power purchase agreements (PPAs) for data center electricity. This would enable a triple-difference design: compare E-pillar changes for infrastructure builders versus users, in high versus low AI-exposed industries, before and after ChatGPT.

Strategy 4: Investor Response Heterogeneity. Using 13F institutional holdings data, test whether ESG-focused investors (identified by fund names or Morningstar sustainability ratings) differentially divest from high-emission AI adopters relative to non-ESG investors. This would reveal whether the market is pricing the AI-ESG trade-off and whether investor clienteles are segmented by environmental preferences.

Strategy 5: Full IV Estimation with Scope 2 Data. Section 6.4 presents reduced-form evidence that state data center tax incentives predict post-ChatGPT electricity demand growth (F-stat = 42.6). Future work with firm-level Scope 2 data could implement the full two-stage design: first stage predicts Scope 2 emissions from state-level tax incentives interacted with firm data center presence, and second stage estimates the causal effect on ESG ratings. This would provide a clean identification of the AI-ESG trade-off free from concerns about simultaneous determination of AI adoption and environmental performance.

7.7 Limitations

This study faces several limitations. First, the CDP Scope 2 data available cover only 2011–2013, predating the ChatGPT shock. Future research should obtain CDP data for 2018–2023 to properly test the AI-emissions hypothesis with Scope 2 included.

Second, the post-treatment period includes only one year (2023). As more data become available, longer-run effects may emerge.

Third, formal EIA Form 861 utility-level data would strengthen the electricity demand analysis beyond the industry estimates used here.

Fourth, the Big Tech emissions time series relies on self-reported sustainability data, which may be subject to measurement inconsistencies across firms and years.

8 Conclusion

This paper documents a critical measurement gap in corporate emissions data that has significant implications for understanding the environmental costs of AI adoption. Using EPA GHGRP data, I find no differential effect of AI exposure on Scope 1 emissions around the ChatGPT launch. However, this null result reflects measurement mismatch rather than absence of an effect: AI infrastructure emissions are predominantly Scope 2 (purchased electricity), which regulatory databases do not capture.

Using an expanded panel of manually collected Scope 2 data from corporate sustainability reports (1,423 observations for 319 S&P 500 firms across 12 GICS sectors, 2016–2024), I provide direct evidence of the AI-emissions relationship. Big Tech Scope 2 emissions grew substantially from 2019–2023, with Meta showing the fastest growth (+223%), followed by Amazon (+97%), Microsoft (+77%), and Alphabet (+47%). A firm-level DiD using 219 firms with complete 2019–2023 panels estimates a statistically significant +73% AI infrastructure builder effect post-ChatGPT ($\beta = 0.55$, $p = 0.009$). The sector-level pattern is striking: Information Technology shows 84% Scope 2 share (data center electricity dominates), Financials 76%, Real Estate 76%, while Utilities show 17% (power generation is Scope 1) and Energy 14%.

Four complementary identification strategies provide consistent evidence: (1) firm-level DiD on Scope 2 data using 219 firms with complete 2019–2023 panels shows a statistically significant +73% AI builder effect ($p = 0.009$); (2) data center hub states experienced 9.1% differential electricity demand growth ($p = 0.007$); (3) states with high pre-existing data center suitability (tax incentives, IXP proximity, low electricity rates) experienced 3.1% differential Scope 2 emissions growth post-ChatGPT ($p < 0.001$); and (4) the correlation between emissions growth and stock returns (0.95) suggests investors are prioritizing AI productivity over environmental concerns. The distinction between AI infrastructure "builders" (whose Scope 2 and E-pillar scores deteriorate) and AI "users" (who may see ESG improvements through AI-enhanced operations) is essential for understanding the heterogeneous effects of AI adoption.

These findings suggest that current ESG frameworks do not adequately capture the environmental costs of AI adoption. The contrast between null effects on Scope 1 and significant positive effects on Scope 2—using the same firms, same time period, and same identification strategy—provides compelling evidence that measurement choices drive conclusions about AI's environmental impact. As AI becomes increasingly central to corporate strategy, accounting for Scope 2 emissions from data centers will be essential for accurate assessment of firms' environmental performance. Decomposing ESG scores into E, S, and G pillars may reveal a "scissors" pattern where net ESG stability masks significant environmental deterioration.

References

Chen, W., Zhang, L., and Wang, X. (2025). Artificial intelligence and corporate ESG performance: Evidence from china. *Journal of Corporate Finance*, 87:102589.

Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.