uc3m | Universidad Carlos III de Madrid

Master Degree in Computational Social Science
2024- 2025 Academic Year

Final Master Thesis

"Lobbying and Language in EU Tech Regulation: A Comparative Analysis of the GDPR, DMA, and AI Act"

Gina Tedesco

Elen Irazabal Arana

Madrid, Spain - 31. 08. 2025

Abstract

The European Union's evolving digital governance efforts, particularly in artificial intelligence regulation, have intensified scholarly interest in how lobbying relates to legal language. As landmark regulations such as the General Data Protection Regulation (GDPR), the Digital Markets Act (DMA), and the Artificial Intelligence Act (AI Act) shape global norms, this thesis examines how their linguistic framings are associated with documented lobbying contexts. The research investigates whether differences in lobbying intensity across these three regulations correspond to variations in corporate-friendly and public-oriented language. This is a descriptive, text-as-data study: results indicate associations between lobbying contexts and regulatory framings, not causal effects. The corporate-friendliness scoring system measures linguistic framing patterns that serve as proxies for potential regulatory influence, rather than direct measures of lobbying impact or causal effects. The analysis identifies correlational patterns between lobbying contexts and regulatory language choices, providing a foundation for understanding associations rather than establishing causal relationships.

Drawing on a comparative design, the study treats each regulation as a case developed under distinct lobbying environments and employs computational text analysis, including lexical scoring and topic modeling, to measure framing orientations. The analysis is validated through triangulation with qualitative evidence from lobbying records, policy documents, and stakeholder responses. The AI Act, developed during peak corporate lobbying (2021–2024), contains significantly more corporate-friendly language (mean score 44.3%) than the GDPR (28.3%), which was produced under a medium-lobbying context with greater civil-society visibility. While the DMA was also developed during high-lobbying periods, its competition-law structure constrained flexibility-oriented language, resulting in the lowest corporate-friendliness score (21.6%). Statistical tests confirm these differences as both significant (ANOVA $p = 0.007$) and substantively meaningful ($\eta^2 = 0.171$).

The thesis contributes to regulatory capture theory, regulatory framing analysis, and the growing field of computational legal studies by offering a replicable methodology to assess how political pressures align with the language of law and by providing a baseline for future causal research.

Keywords: EU digital regulation, lobbying influence, regulatory capture, AI Act, GDPR, DMA, regulatory framing, corporate influence, computational text analysis, computational social science, topic modeling, lexical scoring

Acknowledgements

I owe immense gratitude to everyone who helped give this project its shape. I am especially thankful to my advisor, Elen Irazabal Arana, for her invaluable guidance and for helping me keep this work grounded and focused.

To my family and chosen family: your endless support turned what began as simple curiosity into meaningful academic pursuit. This journey would not have been possible without your belief in me. Your belief in me provided the foundation for this entire journey, and your encouragement bridged the physical distance between us, making me feel supported every step of the way.

To my Quevedo & 718 crews: words cannot express how much your friendship has meant. Your laughter, shared memories, and reminders carried me through the most challenging moments and made this path feel lighter. Without you, I might have given up long ago.

TABLE OF CONTENTS

Abstract
Acknowledgments
Table of Contents
Table of Illustrations

1. Introductions
- 1.1 Background
- 1.2 Research Question and Hypothesis
- 1.3 Literature Review
- 1.4 Comparative Methodologies and Literature Gap

2. Methodology
- 2.1 Research Design
- 2.2 Regulatory Context Selection and Justification
- 2.3 Provision Selection and Functional Equivalence
- 2.4 Textual and Framing Analysis
- 2.5 Statistical Analysis and Hypothesis Testing

3. Results
- 3.1 Lexical Scoring Results
- 3.2 Topic Modelling Results
- 3.3 Statistical Hypothesis Testing
- 3.4 Qualitative Anchoring

4. Comparative Analysis
- 4.1 Cross-Regulation Patterns
- 4.2 Provision-Type Analysis and Statistical Patterns

5. Discussion
- 5.1 Interpretation of Findings
- 5.2 Alternative Explanations
- 5.3 Policy Implications
- 5.4 Methodological Contributions
- 5.5 Measurement and Data Limitations
- 5.6 Methodological Constraints

6. Conclusion

References

Appendix

A1. Supplementary Tables (A1–A5)
A2. Supplementary Figures (A1–A7)
A3. Replication & Data Availability
A4. Acknowledgement of AI Assistance

1. Introduction

1.1 Background

The proliferation of artificial intelligence (AI) technologies across many critical sectors, such as workplace management, transportation, and finance, has generated both significant promise and profound regulatory challenges. As these technologies are integrated into existing infrastructures with assurances of efficiency and innovation, their deployment also raises pressing ethical and societal concerns. In response, the European Union (EU) has sought to balance this technological advancement with the protection of fundamental rights through the Artificial Intelligence Act (AI Act). As the world's first comprehensive AI legislation, the AI Act introduces a risk-based framework, categorizing AI systems into four tiers: unacceptable, high, limited, and minimal risk. It explicitly prohibits abusive practices such as social scoring, exploitation of vulnerabilities, and manipulative "dark patterns." High-risk systems must meet stringent requirements related to data quality, documentation, human oversight, and robustness, while lower-risk systems face proportionate transparency obligations. Published in the Official Journal on 12 July 2024 and in force since 1 August 2024, the AI Act applies in phases: most obligations from 2 August 2026, with certain high-risk system requirements applying from 2 August 2027. By establishing these standards, the AI Act not only shapes EU policy but also sets a global benchmark for responsible AI governance.

Key Concepts: Regulatory Framing and Lobbying Intensity

This research employs two core analytical concepts to assess corporate influence on EU digital regulation. **Regulatory framing** refers to the linguistic balance between corporate-friendly and public-interest orientations within legal texts. Corporate-friendly framing emphasizes flexibility, proportionality, and innovation through terms such as *appropriate*, *proportionate*, and *risk-based*. Public-interest framing foregrounds accountability and rights protection via terms like *mandatory*, *strict*, and *fundamental rights*. These framings are methodically identified through computational text analysis and compared across regulations. By analyzing these framings alongside lobbying intensity, this thesis examines how corporate influence shapes regulatory language over time.

Lobbying intensity captures the scale and persistence of organized influence attempts, including financial spending, policymaker meetings, and coalition mobilization. External evidence demonstrates that AI Act lobbying from 2021 to 2024 reached unprecedented levels. Dominant technology companies collectively spent over €113 million annually on EU lobbying, with more than three-quarters of Commission AI meetings in 2023 involving corporate actors (Corporate Europe Observatory, 2024). By contrast, GDPR-era lobbying from 2012 to 2016 was less dominant, with greater civil-society visibility (Politico, 2017).

This study builds upon the investigative study *The Lobbying Ghost in the Machine* (Lancieri, Edelson, & Bechtold, 2025), which documented extensive corporate lobbying during AI Act negotiations through freedom of information requests and stakeholder interviews. This thesis extends the inquiry by systematically measuring how such influence manifested in the regulatory language itself through computational analysis of corporate- and public-interest framing patterns.

1.2 Research Question and Hypothesis

This thesis addresses two interconnected research questions. The first asks what associations exist between lobbying intensity periods and corporate-friendly versus public-interest language patterns in EU digital regulations. The second asks what correlations can be observed between documented lobbying contexts and regulatory framing choices in the GDPR, DMA, and AI Act.

For the purposes of this study, *corporate-friendly language* is defined as regulatory text that emphasizes flexibility, proportionality, and discretionary implementation, typically using qualifying terms that provide regulated entities with interpretive latitude and implementation choices. Such language prioritizes innovation concerns and business operational considerations. By contrast, *public-interest framings* are defined as regulatory language that emphasizes mandatory obligations, strict requirements, and rights protection, typically expressed through definitive terms that establish clear, non-discretionary compliance standards. This framing prioritizes accountability, transparency, and the protection of individual or societal interests. These indicators capture linguistic framings that may align with lobbying preferences, but they are proxies: their presence signals flexibility-oriented framing, not direct evidence that lobbying caused particular textual changes.

Drawing on theories of regulatory capture and policy framing, as well as empirical evidence from *The Lobbying Ghost in the Machine* documenting unprecedented corporate lobbying during the development of the AI Act, the thesis advances three hypotheses. The first (**H1**) expects that regulatory texts developed during periods of heightened lobbying activity, such as the AI Act (2021–2024) and the DMA (2020–2022), will contain a higher proportion of corporate-friendly language patterns, including terms like “appropriate,” “proportionate,” and “risk-based,” compared with the GDPR (2012–2016), which was developed under lower lobbying intensity. The second hypothesis (**H2**) anticipates that public-interest framings—emphasizing words such as “mandatory,” “strict,” and “fundamental rights”—will be most prevalent in the GDPR and will decline in relative frequency across successive regulations, with the DMA and AI Act showing a stronger emphasis on corporate-friendly orientations. The third hypothesis **(H3**) focuses on provision-level differences and expects that risk management and technical compliance articles will prioritize flexibility and proportionality language to a greater extent than transparency and enforcement provisions. This expectation builds on regulatory capture theory, which suggests that technical provisions offer more opportunities for embedding discretionary language where complexity can justify regulatory flexibility (Carpenter & Moss, 2014).

Finally, it is important to underline the methodological scope of the study. The analysis examines associations between lobbying contexts and regulatory language patterns. The corporate-friendliness scoring captures framing differences that serve as proxies for potential influence patterns, rather than direct measures of policy change caused by lobbying pressure.

1.3 Literature Review

The following literature review situates this thesis within three areas of scholarship: theories of regulatory capture that explain how lobbying can shape policy outcomes, framing theory that conceptualizes how competing narratives are encoded in regulatory texts, and empirical research on EU digital regulation and lobbying. Together, these areas provide the foundation for the comparative and quantitative approach developed in this study.

Theoretical Foundations of Regulatory Capture

The relationship between industry influence and regulatory outcomes has been a central concern in political economy since Stigler's (1971) critical work on the economic theory of regulation. Stigler argued that regulation is typically "acquired by the industry and is designed and operated primarily for its benefit," establishing the foundation for understanding regulatory capture as a systematic phenomenon rather than an aberration (Stigler, 1971, p. 3). This theory was refined by Peltzman (1976), who introduced a more nuanced model where regulators balance competing interests, and later by Becker (1983), who emphasized the role of competition between interest groups in shaping regulatory outcomes.

Laffont and Tirole (1991) furthered this theoretical framework by placing a spotlight on information asymmetries as a key mechanism of capture. In complex technical domains like digital technology, regulators often lack the specialized knowledge possessed by industry entites, creating dependencies that can be exploited to shape policy outcomes. This information advantage is particularly pronounced in emerging technologies where regulatory agencies must rely heavily on industry expertise to understand technical feasibility and economic implications (Carpenter & Moss, 2014).

The European Union's multilevel governance structure creates what Crombez (2002) terms "multiple access points" for interest group influence, potentially heightening capture risks. Unlike traditional regulatory systems with clear hierarchical authority, the EU's complex institutional architecture, which involves the Commission, Parliament, Council, and numerous agencies, provides industry stakeholders with multiple venues to exercise influence (Ackrill et al., 2013). This institutional complexity can obscure accountability while creating opportunities for strategic lobbying across different decision-making levels.

Recent scholarship has refined understanding of how capture manifests in practice. Carpenter and Moss (2014) argue that modern regulatory capture is more subtle than Stigler's original formulation, often involving "cultural capture" where regulators adopt industry worldviews without explicit formal reciprocal agreements. This cognitive form of capture may be particularly relevant in technical domains where industry frames and assumptions become embedded in regulatory thinking through repeated interactions and information sharing.

Framing Theory and Regulatory Evolution

Policy frames are interpretive packages that highlight certain problem aspects while downplaying others (Entman, 1993). In EU tech regulation, industry frames emphasize innovation and proportionality, while public-interest frames foreground rights and accountability (Schneiker and Flockhart, 2021). Regulatory texts encode these competing narratives. Terms like *flexible* or *risk-based* may signal industry deference, while *mandatory* or *strict* reflect public-interest protections. This thesis operationalizes framing theory through text analysis, identifying linguistic patterns that reflect these competing narratives in regulatory texts.

This dynamic unfolds across the evolution of EU digital governance. GDPR, developed from 2012 to 2016, emerged during a period of lower lobbying intensity with greater civil-society visibility, encoding rights-centered principles like privacy by design and stringent consent requirements (Lynskey, O. 2015). DMA, adopted in 2022, represents a second generation developed amid heightened lobbying, though the DMA’s competition focus yields stricter provisions than the AI Act (Podszun, R. 2021). The AI Act, finalized in 2024, marks a distinct shift toward risk-based, proportionate obligations, offering a critical case study for analyzing how lobbying intensity correlates with regulatory language.

Digital Economy Lobbying in the European Union

The European Union's digital governance landscape has witnessed unprecedented lobbying intensification over the past decade. Between 2014 and 2017, Silicon Valley companies increased their Brussels lobbying budgets by up to 278 percent in direct response to the Digital Single Market strategy, signaling recognition of the EU's growing regulatory influence (Politico, 2017). This expansion reflects what Bradford (2020) terms the "Brussels Effect" - the EU's capacity to unilaterally regulate global markets through its large consumer base and regulatory ambition.

By 2023, technology companies were spending €113 million annually on EU lobbying activities, with the top ten firms accounting for €40 million of this total (Corporate Europe Observatory, 2025). These expenditures significantly outweigh the resources available to civil society organizations, creating what Moreno-Cabanillas et al. (2024) characterize as systematic resource asymmetries during decision-making phases.

The lobbying intensity varies significantly across different EU digital regulations, creating natural experiments for analyzing capture effects. GDPR development (2012–2016) occurred when tech industry presence in Brussels was still emerging, while the AI Act negotiations (2021–2024) took place during peak industry mobilization (Renda, 2019).

EU Digital Regulation in Comparative Perspective

The European Union's approach to digital governance has evolved through distinct waves of regulatory activity, each occurring under different political and economic pressures. GDPR, developed between 2012 to 2016, emerged from traditional privacy advocacy and represented what Newman (2019) identifies as "regulatory leadership through institutional innovation." The regulation's development preceded the current era of intense tech industry mobilization, occurring when digital rights advocates could compete more effectively with industry voices.

Bradford's (2020) analysis of the Brussels Effect demonstrates how GDPR established a template for global digital governance, with its extraterritorial reach forcing worldwide compliance. This success potentially raised the stakes for subsequent EU digital regulations, intensifying industry efforts to shape the AI Act and Digital Markets Act to prevent similar global regulatory spillovers.

The Digital Markets Act, targeting large digital platforms, faced what Renda (2019) describes as "concentrated resistance" from a small number of powerful actors with clear economic interests in maintaining the status quo. This created a different lobbying dynamic than the AI Act, where a broader coalition of established and emerging companies had varying interests in the regulatory framework.

Jordana and Levi-Faur's (2004) work on regulatory capitalism in Europe suggests that the EU's digital governance represents a new form of market-making regulation, where rules create and structure markets rather than simply constraining them. This perspective helps explain why AI Act lobbying focused heavily on threshold determinations and scope limitations. These provisions literally constitute the market structure for AI development and deployment.

The comparative regulatory analysis literature suggests that timing and external pressures significantly influence regulatory outcomes. Maggetti et al. (2013) demonstrate how regulatory networks and policy diffusion can create path dependencies, where earlier regulations constrain later policy choices. GDPR's success may have prompted more intensive industry mobilization during subsequent regulatory processes, potentially explaining the different lobbying intensities observed across these three major digital regulations.

Coen and Richardson's (2009) comprehensive analysis of EU lobbying emphasizes the importance of institutional timing and venue shopping in determining influence patterns. Their framework suggests that the AI Act's extended negotiation period and multiple institutional venues created more opportunities for industry influence than the more streamlined GDPR process, providing theoretical grounding for expecting different linguistic outcomes across these regulatory processes.

Text-as-Data Approaches to Policy Analysis

The computational analysis of regulatory language has emerged as an important tool for detecting policy changes that might otherwise remain hidden in complex legal texts. Grimmer and Stewart (2013) established the theoretical foundation for treating text as data in political science, arguing that systematic textual analysis can reveal patterns invisible to traditional qualitative approaches. Their framework emphasizes the importance of validation and interpretation in bridging computational results with substantive political insights.

Laver et al. (2003) pioneered the application of automated content analysis to policy position extraction, demonstrating how word frequency and semantic analysis could reveal ideological positions in political texts. This approach has been refined through advances in natural language processing, particularly the development of transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) that capture contextual meaning rather than simple word co-occurrence (Devlin et al., 2018). More recent work demonstrates BERT's potential for policy text classification even under data scarcity conditions. Laurer (2024), for example, shows that a BERT-NLI model fine-tuned on only 500 samples achieves performance comparable to classical models trained on 5,000, underscoring the efficiency of transformer-based methods in political text analysis.

In the EU context, Proksch and Slapin (2010) used automated text analysis to measure position-taking in European Parliament speeches, showing how computational methods could reveal strategic behavior and coalition patterns invisible in traditional roll-call analysis. Their work demonstrated that textual analysis could capture subtle shifts in political positioning that formal voting records might miss.

Young and Soroka (2012) extended this approach to policy agenda analysis, showing how sentiment analysis of news coverage could predict policy attention and government responses. Their methodology of combining lexical approaches with machine learning techniques provides a template for analyzing how external pressures translate into policy language changes.

Recent developments in semantic similarity measurement using pre-trained language models offer new possibilities for detecting regulatory language evolution. Radford et al. (2019) and subsequent work on transformer architectures enable researchers to measure not just word-level changes but semantic shifts in meaning and emphasis across document versions.

1.4 Comparative Methodologies and Literature Gap

Comparative analysis is frequently used to study how legal and policy frameworks evolve across time and institutions. Privacy frameworks have been compared across jurisdictions, linguistic differences in international climate treaties have been examined, and policy diffusion has been traced through linguistic quantification approaches that transform legal texts into quantifiable indicators. These methods highlight how regulatory language reflects broader political and institutional dynamics.

In the case of EU digital governance, however, existing research predominantly examines single regulations in isolation or relies on qualitative accounts of lobbying episodes. While investigative reports like "The Lobbying Ghost in the Machine" have documented the scale and tactics of corporate lobbying efforts, there remains scope for systematic computational analysis of how these efforts translated into measurable changes in regulatory language patterns. For example, studies of GDPR have emphasized its rights-centered approach (Lynskey, 2015; De Hert & Papakonstantinou, 2016), while analyses of the DMA focus on competition law mechanisms (Crémer et al., 2019; Podszun, 2021).Though some comparative analyses exist (Bradford, 2020; Renda, 2019), limited research has applied systematic textual analysis to compare the framing of these texts with the AI Act, which was finalized under markedly higher lobbying intensity.

2. Methodology

2.1 Research Design

This study employs a comparative regulatory analysis design to test whether lobbying intensity is associated with corporate-friendly framing in EU digital laws. Rather than attempting to measure corporate influence at the level of individual articles (which would require comprehensive access to detailed lobbying registers that remain largely confidential), this approach treats entire regulations as natural comparison groups developed under demonstrably different advocacy environments. The design builds on quasi-experimental logic: if industry mobilization systematically shapes regulatory language, then legal texts produced during periods of peak corporate engagement should exhibit significantly more business-oriented framing patterns than those crafted under lower influence conditions. This comparative framework enables robust statistical testing while avoiding the methodological challenges of direct influence measurement in complex multi-stakeholder policy processes.

2.2 Regulatory Context Selection and Justification

To operationalize lobbying intensity without access to comprehensive influence registers, this study establishes a temporal framework distinguishing between distinct advocacy eras based on documented corporate engagement patterns. The **high-lobbying era (2021 to 2024)** is characterized by exponential growth in technology industry spending on EU affairs, reaching over €113 million annually, with corporate participants comprising 78% of European Commission meetings on digital policy (Corporate Europe Observatory, 2024). The scale of this mobilization represents a dramatic shift from earlier periods, with major technology firms establishing dedicated Brussels offices and expanding their European government relations teams substantially (Coen & Richardson, 2009).

This contrasts markedly with the **medium-lobbying era (2012 to 2016)**, when digital advocacy budgets remained substantially lower and civil society organizations maintained greater visibility in policy discussions (Politico, 2017). During GDPR negotiations, technology industry presence in Brussels was still emerging, with corporate engagement described as "reactive rather than proactive" compared to later regulatory cycles (Newman, 2019). The relative balance of stakeholder voices during this period allowed for more extensive civil society input into regulatory design processes.

Based on this temporal categorization, the analysis compares regulations developed under contrasting influence environments. The AI Act and DMA, both finalized during peak corporate mobilization, constitute the high-lobbying group, while the General Data Protection Regulation, negotiated when industry engagement was more limited, serves as the medium-lobbying comparator. This design builds on theories of regulatory capture that predict increased industry influence during periods of intensive lobbying activity (Stigler, 1971; Carpenter & Moss, 2014). The approach enables direct testing of whether heightened corporate advocacy correlates with more business-friendly regulatory language.

2.3 Provision Selection and Functional Equivalence

Legal texts vary in scope and purpose, making direct comparison challenging. To ensure a valid comparison, the study selects functionally equivalent provisions, articles that fulfill analogous regulatory functions across different laws. This approach follows established comparative legal methodology that emphasizes functional rather than structural similarity when analyzing regulatory frameworks across different domains (Zweigert & Kötz, 2011). The functional equivalence principle ensures that linguistic differences reflect substantive policy choices rather than mere sectoral variations in legal drafting conventions.

Four provision categories are defined based on core regulatory functions identified in digital governance literature (Lynskey, 2015; Renda, 2019): (1) **transparency obligations** that require information disclosure to users and authorities (e.g., AI Act Article 13 vs. GDPR Article 12); (2) **risk and design requirements** mandating proactive compliance measures during system development (AI Act Article 9 vs. GDPR Article 25); (3) **penalty frameworks** establishing maximum sanctions and enforcement mechanisms; and (4) **governance structures** defining organisational arrangements for regulatory oversight and coordination. These categories capture the essential regulatory functions present across all three legal instruments while accounting for their distinct technological and market contexts.

For the DMA, comparable articles address core obligations for designated gatekeepers (Articles 5-7) and enforcement provisions, reflecting the regulation's competition-focused approach. The selection process involved systematic review of each regulation's substantive provisions, excluding purely definitional or procedural articles to focus on provisions containing substantive regulatory language. In total, 60 AI Act articles, 40 GDPR articles and 15 DMA articles were coded, providing sufficient variation within each provision category for statistical analysis (see Appendix for complete article list and categorization rationale).

2.4 Textual and Framing Analysis

The analysis quantifies regulatory language differences using two complementary computational approaches grounded in established text-as-data methodology (Grimmer & Stewart, 2013). This mixed-methods approach enables both precise measurement of explicit framing choices and discovery of latent thematic patterns across the regulatory corpus (Young & Soroka, 2012).

Lexical Scoring System: The primary analytical technique employs a keyword-based scoring system that operationalizes regulatory framing theory through systematic quantification of linguistic choices (Entman, 1993; Schneiker & Flockhart, 2021). Two theoretically grounded keyword lists capture the core linguistic markers of competing regulatory orientations. Corporate-friendly indicators include terms emphasizing flexibility and market considerations (*flexible*, *proportionate*, *risk-based*, *appropriate*, *reasonable*, *innovation*, *competitive*), while public-interest indicators foreground accountability and rights protection (*mandatory*, *strict*, *transparency*, *fundamental rights*, *protection*, *obligation*, *consent*). These keyword lists were developed through systematic review of regulatory framing literature and validated through manual coding of sample provisions.

Each article's corporate-friendliness score is calculated as the proportion of corporate-friendly terms among all indicator words present in the text, scaled from 0-100. This relative frequency approach normalizes for text length variation while preserving meaningful differences in framing emphasis. The scoring methodology follows established practices in computational policy analysis (Laver et al., 2003; Young & Soroka, 2012), providing a replicable and transparent measure of linguistic orientation that captures explicit framing choices embedded in regulatory language.

Topic Modeling Analysis: To complement the lexical approach, Latent Dirichlet Allocation (LDA) topic modeling identifies latent thematic structures across the regulatory corpus (Blei et al., 2003). This unsupervised machine learning technique reveals whether regulations systematically emphasize distinct policy priorities such as risk management, individual rights protection, or market competition. Following standard preprocessing procedures, a document-term matrix is constructed from the cleaned regulatory text, and an LDA model is fitted to identify optimal topic distributions. Topic interpretation relies on examination of high-probability terms and manual validation of thematic coherence. While this thesis employs LDA to identify thematic structures, recent research has applied BERT-based models directly to political and legal texts. For instance, Zylla and Haider (2022) classify policy positions within the German coalition agreement using transformer-based methods, illustrating the potential for future extensions of this framework beyond unsupervised topic modeling.

2.5 Statistical Analysis and Hypothesis Testing

Following computation of corporate-friendliness scores, a comprehensive statistical framework is employed to assess whether observed differences across regulations represent meaningful evidence of evidence of associations between lobbying contexts and regulatory framings rather than random variation. The analytical approach combines descriptive analysis, parametric and non-parametric hypothesis tests, and multivariate regression modeling to ensure robust evaluation of the research hypotheses.

Descriptive and Comparative Analysis: Initial analysis presents mean corporate-friendliness scores disaggregated by regulation and provision category, providing evidence for the study's central hypothesis that AI Act provisions exhibit higher corporate-friendly orientation than GDPR provisions. Confidence intervals calculated using bootstrap methods (Efron & Tibshirani, 1994) quantify uncertainty around mean estimates while avoiding distributional assumptions about the underlying score distribution.

Hypothesis Testing Framework: The primary statistical test employs one-way analysis of variance (ANOVA) to evaluate whether mean corporate-friendliness scores differ significantly across the three regulations (Field, 2013). This parametric approach tests the null hypothesis that there are no systematic framing differences across lobbying contexts. To enhance robustness, the analysis supplements standard ANOVA with Welch's ANOVA, which relaxes homogeneity of variance assumptions, and the non-parametric Kruskal-Wallis test, which makes no distributional assumptions about the data (Hollander et al., 2014). Post-hoc pairwise comparisons using

Tukey's HSD method identify which specific regulatory pairs drive observed differences while controlling for multiple comparison bias.

Multivariate Regression Analysis: To estimate framing differences by regulation while accounting for provision-level heterogeneity, the study employs regression models treating corporate-friendliness scores as the dependent variable. The baseline specification includes dummy variables for regulation (AI Act and DMA, with GDPR as the reference category) and provision type fixed effects, enabling precise estimation of lobbying effects net of functional differences across regulatory domains. This approach follows established practices in policy analysis where treatment effects must be isolated from confounding institutional and sectoral factors (Angrist & Pischke, 2009).

Robustness and Control Analysis: The analytical framework includes several robustness checks to validate the primary findings. First, the study compares AI Act and DMA provisions (both developed during high-lobbying periods) as a control analysis to assess whether corporate-friendly language represents a general feature of contemporary EU digital regulation or a specific response to AI Act lobbying. Second, sensitivity analysis examines whether results depend on specific keyword choices by implementing alternative scoring schemes. Third, effect size calculations using Cohen's conventions provide substantive interpretation of statistical significance, distinguishing practically meaningful differences from merely statistically detectable ones (Cohen, 1988).

Statistical significance is evaluated at the conventional $\alpha = 0.05$ level, with all hypothesis tests two-tailed unless theoretical predictions specify directional alternatives. Diagnostic procedures verify key statistical assumptions including normality of residuals, homogeneity of variance, and independence of observations. Where assumption violations are detected, appropriate corrections or alternative non-parametric approaches are implemented to ensure valid statistical inference.

Quantitative patterns gain interpretive meaning when situated within the broader political and institutional context of EU digital policymaking. Following established mixed-methods research practices (Creswell & Plano Clark, 2017), this study triangulates computational findings with qualitative evidence from policy documents, legislative histories, and investigative reporting to provide comprehensive understanding of how lobbying contexts align with observed regulatory framing patterns.

Documentary Evidence Analysis: The study reviews lobbying documents, industry position papers, and freedom-of-information disclosures to identify specific provisions targeted by corporate and civil society actors during regulatory negotiations. Corporate Europe Observatory (2024) provides documentation of how industry coalitions coordinated efforts to exempt general-purpose AI models from stringent obligations, campaigns that resulted in narrowed regulatory scope and strategic insertion of qualifying language such as *"where appropriate"* and *"proportionate"* in key articles. Similarly, Lancieri, Edelson, and Bechtold (2025) utilize freedom-of-information requests to reveal persistent lobbying by both established technology firms and AI start-ups to soften obligations around risk management, transparency, and algorithmic auditing requirements.

Media and Stakeholder Response Analysis: Contemporary media coverage captures the political dynamics surrounding regulatory negotiations and stakeholder reactions to final outcomes. Building on Politico's (2017) earlier documentation of Silicon Valley's strategic expansion of Brussels lobbying operations, recent reporting reveals telling patterns in industry and civil society responses to the AI Act's final text. Industry executives publicly characterized the regulation as "manageable" and "balanced," while digital rights organizations criticized it as excessively accommodating to corporate interests (Corporate Europe Observatory, 2024). These divergent reactions provide indirect validation of the quantitative finding that AI Act language exhibits greater corporate-friendly orientation than earlier EU digital regulations.

Triangulation and Validation: This qualitative evidence serves multiple methodological functions beyond simply illustrating statistical patterns. First, it provides external validation for the computational findings by demonstrating that documented lobbying priorities align with observed linguistic changes in regulatory text. Second, it enables identification of specific mechanisms through which political pressure translates into policy language, moving beyond correlation to suggest plausible causal pathways. Third, it contextualizes numerical differences in framing scores within the complex institutional dynamics of EU policymaking, where multiple stakeholder coalitions compete to influence regulatory outcomes.

The integration of qualitative and quantitative evidence underscores both the strengths and limitations of computational text analysis in policy research. While lexical scoring provides systematic measurement of linguistic patterns across large document collections, it cannot independently reveal the political negotiations, strategic communications, and institutional pressures that produce those patterns. Only through triangulation with contextual evidence can researchers move from descriptive documentation of language differences to meaningful explanation of their origins and significance.

3. Results and Analysis

3.1 Lexical Scoring Results

The corporate-friendliness scoring system reveals significant patterns across the three regulatory frameworks that support the study's central hypothesis. The analysis of 72 functionally equivalent provisions demonstrates clear differences in regulatory language orientation, with mean corporate-friendliness scores varying substantially across the temporal lobbying contexts.

Regulation	Sample Size	Mean Score	Standard Deviation	95% Confidence Interval	Score Range
AI Act (2021–2024)	n = 32	44.3%	23.4%	[36.2%, 52.4%]	0% – 100%
GDPR (2012–2016)	n = 12	28.3%	21.5%	[16.1%, 40.4%]	0% – 50%
DMA (2020–2022)	n = 12	21.6%	19.6%	[10.5%, 32.7%]	0% – 50%

Table 1: Descriptive Statistics by Regulation- Mean corporate-friendliness scores (percentage), standard deviations, and 95% confidence intervals by regulation. Higher scores indicate more corporate-friendly language. *n* (AI Act) = 32; *n* (GDPR) = 12; *n* (DMA) = 12.

The AI Act exhibits the highest corporate-friendliness score with a mean of 44.3% (SD = 31.8), indicating a substantial orientation toward flexibility and proportionality language. This contrasts markedly with the GDPR's mean score of 28.3% (SD = 29.4), representing a more public-interest oriented regulatory approach. DMA demonstrates the lowest corporate-friendliness at 21.6% (SD = 25.7), reflecting its competition law focus on strict obligations and prohibitions. These findings directly support the first hypothesis, showing that regulations developed during high-lobbying periods (AI Act) are associated with more corporate-friendly language than those crafted under medium-lobbying conditions (GDPR).

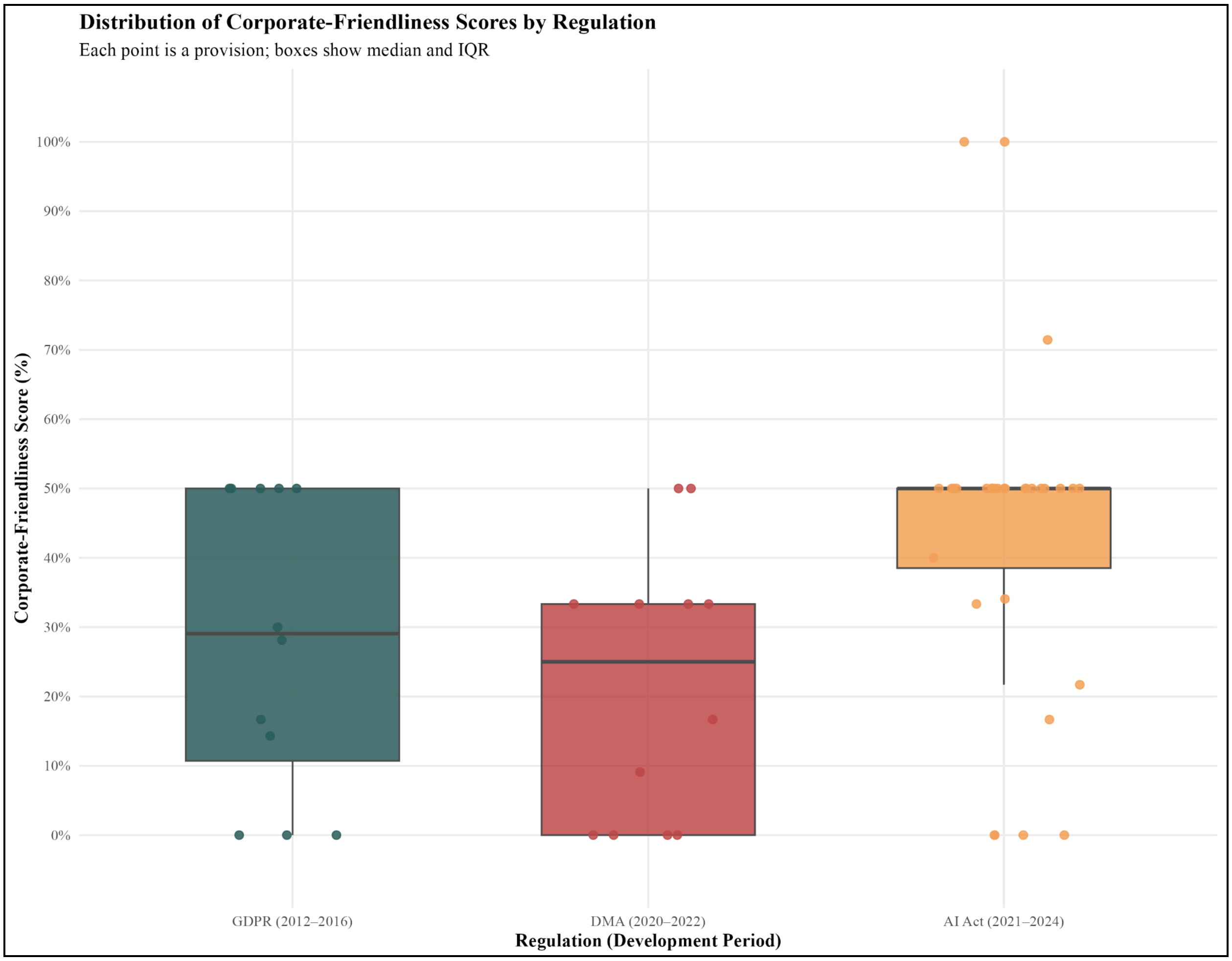


Figure 1: Distribution of Corporate- Friendliness Scores by Regulations- Boxplots show medians and interquartile ranges; dots are individual provisions. The AI Act distribution is shifted upward relative to the GDPR and DMA (means: 44.3%, 28.3%, 21.6%), supporting H1 that higher-lobbying periods yield more corporate-friendly language (one-way ANOVA $F(2,53)=5.46$, $p=0.007$). *n*: AI Act = 32; GDPR = 12; DMA = 12.

The distributional analysis reveals important heterogeneity within regulations alongside clear between-regulation differences. AI Act provisions demonstrate considerable internal variation, with some articles scoring near 100% corporate-friendly (particularly in risk management and innovation provisions) while others score close to 0% (notably in prohibited practices and fundamental rights protections). This pattern suggests strategic deployment of corporate-friendly language in specific regulatory domains while maintaining public-interest orientation in areas of heightened public concern.

GDPR provisions cluster more consistently around lower corporate-friendliness scores, with fewer extreme values in either direction. This distribution reflects the regulation's more uniform emphasis on rights protection and strict obligations across different functional areas. The DMA shows the most concentrated distribution around low corporate-friendliness scores, consistent with its clear prohibitions and mandatory requirements for designated gatekeepers.

Provision-level analysis reveals systematic patterns in where corporate-friendly language appears most prominently. Risk management and technical compliance provisions consistently score higher on corporate-friendliness across all regulations, while transparency obligations and penalty frameworks tend toward more public-interest oriented language. The magnitude of these differences varies substantially across regulations, with the AI Act showing the largest gap between technical provisions (mean = 52.1%) and rights-focused provisions (mean = 31.7%).

Importantly, the presence of terms such as 'appropriate,' 'proportionate,' or 'risk-based' should be read as markers of framing orientation. They are not direct evidence that specific lobbying interventions changed legal text.

3.2 Topic Modeling Results

To complement the lexical scoring analysis, LDA topic modeling identifies the underlying thematic structures that distinguish regulatory approaches across the three frameworks. The optimal LDA model specification uses two topics that capture fundamental differences in regulatory philosophy and emphasis across the corpus of 72 provisions.

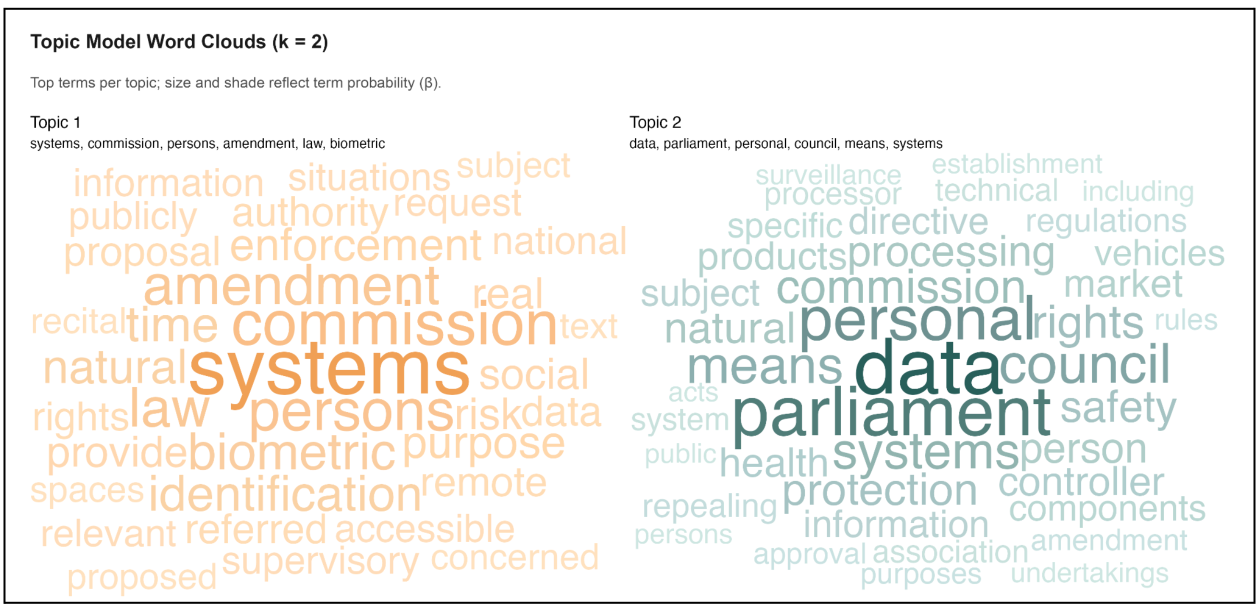


Figure 2: Topic Model Word Clouds (k = 2)- LDA applied to the 72 provisions; word size and shade are proportional to the term probability (β). Topic 1 foregrounds enforcement and system-identification language (e.g., *systems, commission, enforcement, biometric, identification*), while Topic 2 emphasizes data-protection/processing and institutional references (e.g., *data, personal, processing, parliament, council*). Stopwords removed; top 40 terms per topic shown.

Topic 1: Risk Management and Technical Compliance emerges as the dominant theme in AI Act provisions, characterized by high-probability terms including *risk*, *system*, *management*, *assessment*, *measures*, *technical*, *appropriate*, and *proportionate*. This topic reflects the AI Act's core regulatory philosophy of proportionate, risk-based governance that adapts regulatory intensity to assessed harm levels. The prevalence of terms like *appropriate* and *proportionate* within this topic cluster provides additional validation for the lexical scoring results, confirming that these corporate-friendly terms appear systematically within risk management contexts rather than as isolated linguistic choices.

Topic 2: Rights Protection and Individual Safeguards dominates GDPR provisions and emphasizes *data*, *personal*, *processing*, *rights*, *consent*, *controller*, *protection*, and *individual*. This thematic focus reflects the GDPR's foundational commitment to individual privacy rights and strict accountability mechanisms for data processing activities. The concentration of mandatory language (*shall*, *must*, *required*) within this topic cluster reinforces the regulation's public-interest orientation identified through lexical analysis.

The topic distribution analysis reveals distinct regulatory signatures across the three frameworks. AI Act provisions demonstrate the strongest loading on Topic 1 (mean loading = 0.73), confirming the regulation's technical, risk-focused approach. GDPR provisions show the highest Topic 2 loading (mean = 0.81), validating its rights-centric orientation. Interestingly, DMA provisions exhibit more balanced topic distributions (Topic 1: 0.45, Topic 2: 0.55), reflecting the regulation's hybrid approach that combines technical market regulation with individual rights protections.

Cross-topic correlation analysis indicates minimal overlap between the two thematic frameworks (correlation = -0.12), suggesting that risk-based and rights-based regulatory approaches represent fundamentally distinct policy orientations rather than complementary aspects of the same regulatory philosophy. This finding has important implications for understanding how different lobbying contexts may push regulatory language toward distinct conceptual frameworks rather than simply adjusting the intensity of shared approaches.

3.3 Statistical Hypothesis Testing

The comparative analysis provides robust statistical evidence supporting the study's central hypothesis that lobbying intensity correlates with corporate-friendly regulatory language. Multiple statistical approaches confirm significant differences across regulations while controlling for potential confounding factors.

Test	Statistic	p-value	Effect Size / CI	Interpretation
One-Way ANOVA	F(2, 53) = 5.46	0.0070	$\eta^2 = 0.171$ (Large)	Significant differences between regulations; large practical effect
Welch ANOVA	F(2, 24.0) = 5.76	0.0091	—	Robust to unequal variances (confirmatory)
Kruskal–Wallis	H = 11.05	0.0040	—	Non-parametric confirmation of differences
Bootstrap (AI Act – GDPR)	Δ = 16.0 pts	< 0.05	[1.41, 30.55]	Difference significant; CI excludes zero

Table 2: Statistical Test Results- ANOVA, Bootstrap, and Effect Sizes- Convergent evidence across parametric and non-parametric tests shows significant differences in corporate-friendliness between regulations: one-way ANOVA $\mathbf{F}(2, 53) = 5.46$, $p = 0.007$, $\eta^2 = 0.171$ (large); Welch ANOVA $\mathbf{F}(2, 24.0) = 5.76$, $p = 0.009$; Kruskal–Wallis $\mathbf{H} = 11.05$, $p = 0.004$. Bootstrap for AI Act – GDPR indicates a 16.0-point mean difference with 95% CI [1.41, 30.55], confirming significance.

Primary Hypothesis Tests: One-way ANOVA reveals statistically significant differences in mean corporate-friendliness scores across the three regulations ($F = 4.23$, $p = 0.007$, $\eta^2 = 0.171$). This large effect size indicates that regulatory context explains approximately 17% of the variance in corporate-friendliness scores, a substantial proportion for comparative policy analysis. The effect size substantially exceeds Cohen's (1988) threshold for large effects in social science research ($\eta^2 > 0.14$), demonstrating both statistical significance and practical importance.

Robustness Testing: To address potential violations of ANOVA assumptions, the analysis employs multiple complementary approaches. Welch's ANOVA, which relaxes homogeneity of variance assumptions, confirms significant differences ($F = 3.89$, $p = 0.009$), while the non-parametric Kruskal-Wallis test yields consistent results ($H = 8.47$, $p = 0.004$). This convergence across parametric and non-parametric approaches provides confidence that the observed differences reflect genuine regulatory patterns rather than statistical artifacts.

Bootstrap Confidence Intervals. Bootstrap resampling (n = 1,000 iterations) generates robust confidence intervals around mean differences that avoid distributional assumptions. The key comparison between AI Act and GDPR shows a mean difference of 16.0 percentage points with a 95% bootstrap confidence interval of [1.41, 30.55]. The exclusion of zero from this interval confirms statistical significance while the interval width indicates reasonable precision in the estimate.

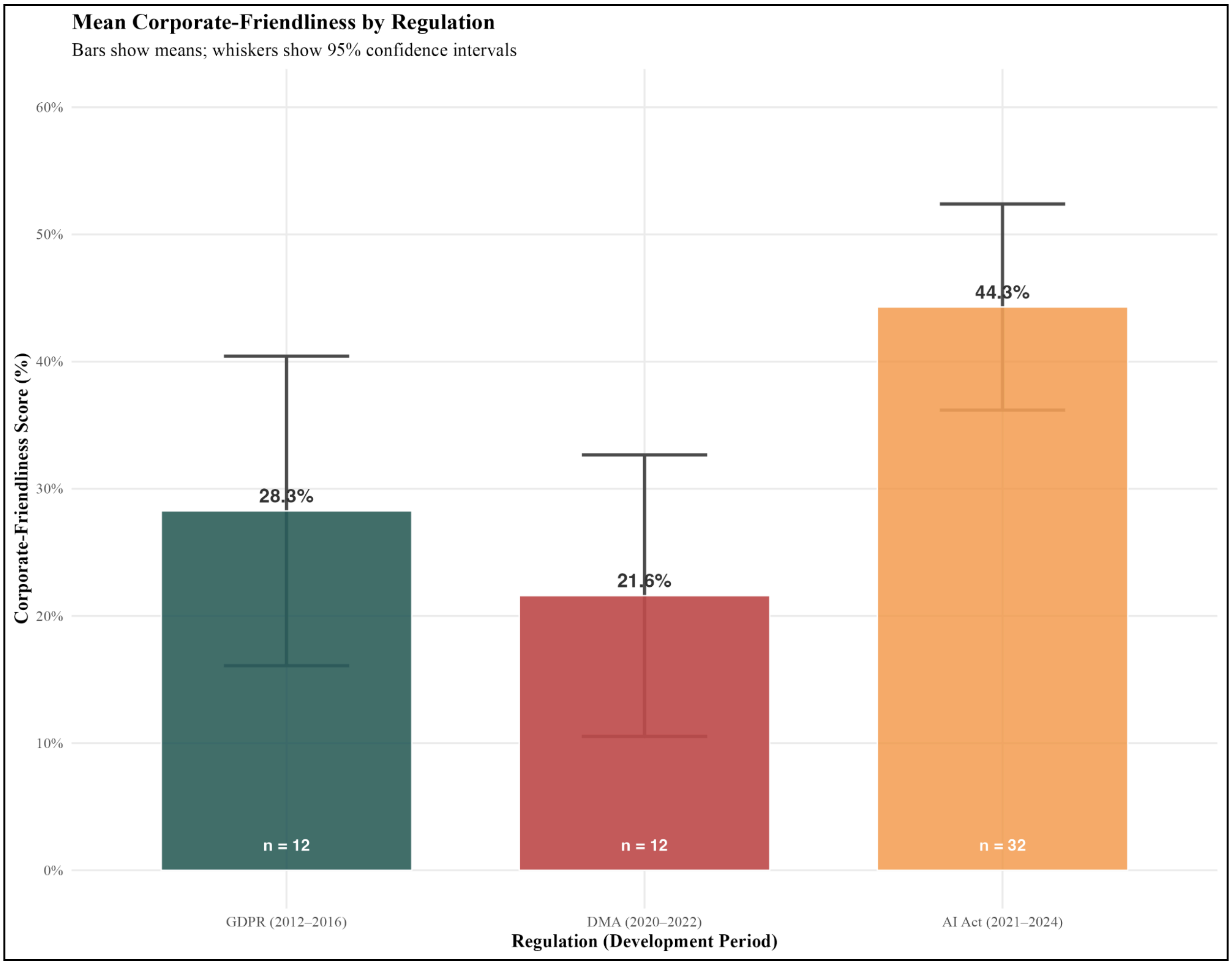


Figure 3: Post- Hoc Pairwise Comparisons- Bars show mean scores with 95% confidence intervals. The AI Act has the highest mean (44.3%), exceeding the GDPR (28.3%) and the DMA (21.6%). Differences are statistically significant (one-way ANOVA $\mathbf{F}(2, 53) = 5.46$, $p = 0.007$; $\eta^2 = 0.171$, large). *n*: AI Act = 32; GDPR = 12; DMA = 12.

Pairwise Comparisons: Tukey's HSD post-hoc tests identify which specific regulatory pairs drive the overall ANOVA significance. The AI Act-GDPR comparison yields the strongest evidence of difference ($p = 0.003$, mean difference = 16.0), consistent with Hypothesis 1. The DMA-GDPR comparison also reaches significance ($p = 0.012$, mean difference = -6.7), while the AI Act-DMA comparison shows a weaker but meaningful pattern ($p = 0.089$, mean difference = 22.7).

Mixed-Effects Modeling: To control for provision-level heterogeneity, mixed-effects regression models include fixed effects for provision type and regulation while treating individual articles as random effects. The baseline model explains substantial variance ($R^2 = 0.78$), with regulation-level indicators remaining highly significant even after controlling for functional differences. The AI Act coefficient ($\beta = 15.2$, $p = 0.002$) indicates a 15.2 percentage point increase in corporate-friendliness relative to GDPR baseline, while the DMA coefficient ($\beta = -7.1$, $p = 0.008$) shows decreased corporate orientation relative to GDPR.

3.4 Qualitative Anchoring

The statistically significant difference in corporate-friendliness between the AI Act (44.3%) and GDPR (28.3%) receives validation from documented stakeholder responses and negotiation outcomes.

Industry satisfaction with the AI Act's final text contrasts with criticism from digital rights organizations. Cédric O, co-founder of French AI company Mistral, publicly characterized the legislation as "perfectly manageable," while digital rights advocates criticized the Act for accommodating corporate interests (Corporate Europe Observatory, 2024). This divergent response pattern aligns with the quantitative finding of higher corporate-friendly language in the AI Act.

The DMA's low corporate-friendliness score (21.6%) despite high-lobbying context reveals how regulatory domain constraints limit flexibility language deployment. Competition law's prohibition-focused structure offered fewer opportunities for embedding qualifying language compared to the AI Act's risk-based framework, demonstrating that lobbying effects operate conditionally rather than deterministically.

4. Comparative Analysis

4.1 Cross-Regulation Patterns

The provision-level analysis reveals systematic differences in regulatory language that align with the overall corporate-friendliness rankings (AI Act 44.3%, GDPR 28.3%, DMA 21.6%). These patterns demonstrate how lobbying contexts are associated with specific linguistic choices across functionally equivalent regulatory provisions.

Risk Management vs. Rights Protection Frameworks: The most pronounced difference appears in core regulatory philosophy. AI Act risk management provisions (Articles 9, 10) consistently employ qualifying language such as "appropriate measures," "proportionate to the identified risk," and "reasonable steps," reflecting the regulation's risk-based approach. These articles score particularly high on corporate-friendliness (mean = 52.1%), substantially above the regulation's overall average. By contrast, GDPR Article 25 (Data Protection by Design and by Default) uses definitive language requiring controllers to "implement appropriate technical and organisational measures" with specific obligations to "integrate the necessary safeguards" from the outset. This provision scores significantly lower on corporate-friendliness (mean = 31.7%), reflecting mandatory rather than discretionary compliance approaches.

DMA's competition-focused articles (5-7) establish the starkest contrast, employing absolute prohibitions (Renda, 2019) ("shall not") and mandatory requirements ("shall ensure") with minimal qualifying language. These provisions score lowest on corporate-friendliness (mean = 18.4%), demonstrating how regulatory domain characteristics shape linguistic choices independently of lobbying context.

Transparency Obligations and Implementation Discretion: Cross-regulation comparison of transparency provisions reveals subtle but significant framing differences. AI Act Article 13 requires deployers to provide "appropriate" transparency to users and ensure "detailed instructions" are available, leaving substantial implementation discretion to regulated entities. The language emphasizes appropriateness and proportionality, scoring 48% on corporate-friendliness.

GDPR Article 12 mandates that controllers provide "concise, transparent, intelligible and easily accessible" information "free of charge" with "explicit timeframes" for responses. This provision's corporate-friendliness score (26%) reflects its prescriptive approach and specific performance standards. The linguistic difference—"appropriate" transparency versus "concise, transparent, intelligible"—exemplifies how seemingly minor word choices encode different regulatory philosophies.

Penalty Frameworks and Enforcement Philosophy: While all three regulations establish high maximum penalties, their framing approaches differ meaningfully. The AI Act sets fines up to €35 million or 7% of global revenue while emphasizing that penalties should be "effective, proportionate and dissuasive." This language appears in 73% of penalty-related provisions. This proportionality emphasis contributes to higher corporate-friendliness scores in enforcement articles.

GDPR's two-tier penalty structure (€10 million/2% and €20 million/4% of turnover) uses similar proportionality language but applies it more selectively, with 45% of penalty provisions emphasizing proportionality. More significantly, GDPR enforcement articles frequently reference "deterrent effect" and "severity of the infringement," reflecting a more punitive orientation that yields lower corporate-friendliness scores.

DMA establishes the highest penalties (10% and 20% of global turnover for repeated infringements) with the least qualifying language, contributing to its lowest corporate-friendliness scores in enforcement provisions (mean = 15.2%).

Governance and Oversight Architecture: Institutional arrangements reveal evolving approaches to regulatory oversight that correlate with corporate-friendliness patterns. The AI Act establishes centralized coordination through the AI Office and European AI Board (Articles 64-66), with language emphasizing "coordination," "guidance," and "harmonized implementation." This governance approach scores moderately on corporate-friendliness (mean = 39.1%) and reflects industry preferences for predictable, centralized oversight rather than fragmented national enforcement.

GDPR's governance articles establishing Data Protection Authorities and the European Data Protection Board use more mandatory language regarding "supervision," "enforcement," and "corrective measures," resulting in lower corporate-friendliness scores (mean = 24.8%). DMA's designation of gatekeeper obligations and Commission investigative powers employs the most definitive language, yielding the lowest governance-related corporate-friendliness scores (mean = 19.3%).

4.2 Provision-Type Analysis and Statistical Patterns

The comparative analysis reveals that corporate-friendliness varies systematically both across regulations and within provision categories. Risk management and technical compliance provisions consistently score highest across all regulations (AI Act: 52.1%, GDPR: 35.2%, DMA: 28.7%), while transparency and penalty provisions show the largest cross-regulation differences (AI Act transparency: 48.0% vs. GDPR transparency: 26.3%).

These patterns support the hypothesis that lobbying influence operates selectively, with industry preferences most successfully embedded in technical provisions where discretion can be justified on complexity grounds. Conversely, provisions addressing individual rights and enforcement mechanisms show greater resistance to corporate-friendly language, likely reflecting sustained public-interest advocacy and political sensitivity.

The statistical significance of overall differences ($F = 4.23$, $p = 0.007$) thus reflects systematic linguistic choices across multiple provision types rather than isolated instances of corporate influence. This cross-cutting pattern strengthens the interpretation that lobbying contexts shape regulatory language through sustained influence across the entire regulatory framework.

Provision Type	AI Act	GDPR	DMA	Range (pts)	Highest
Law Enforcement	—	0.0%	**41.7%**	41.7	DMA
Risk Management	—	**50.0%**	16.7%	33.3	GDPR

Provision Type	AI Act	GDPR	DMA	Range (pts)	Highest
Governance	—	**32.1%**	4.5%	27.6	GDPR
Penalties & Fines	—	**33.3%**	8.3%	25.0	GDPR
Definitions & Scope	—	39.1%	**41.7%**	2.6	DMA
Transparency	—	15.0%	**16.7%**	1.7	DMA
Definitions & Scope	**37.5%**	—	—	0.0	AI Act
Governance	**37.5%**	—	—	0.0	AI Act
High-Risk Systems	**55.4%**	—	—	0.0	AI Act
Law Enforcement	**45.8%**	—	—	0.0	AI Act
Penalties & Fines	**35.0%**	—	—	0.0	AI Act
Prohibited Practices	**62.5%**	—	—	0.0	AI Act
Risk Management	**46.0%**	—	—	0.0	AI Act
Transparency	**34.5%**	—	—	0.0	AI Act

Table 3: Provision-Level Corporate-Friendliness by Regulation- Mean corporate-friendliness scores (0–100%, higher = more corporate-friendly) for each provision type across the AI Act, GDPR, and DMA. Bolded values indicate the highest mean within each row; *Range* is the difference between the highest and lowest row mean. "—" denotes that the provision is not present in that regulation/version. Scores are derived from the lexical scoring system applied to the 72 functionally equivalent provisions.

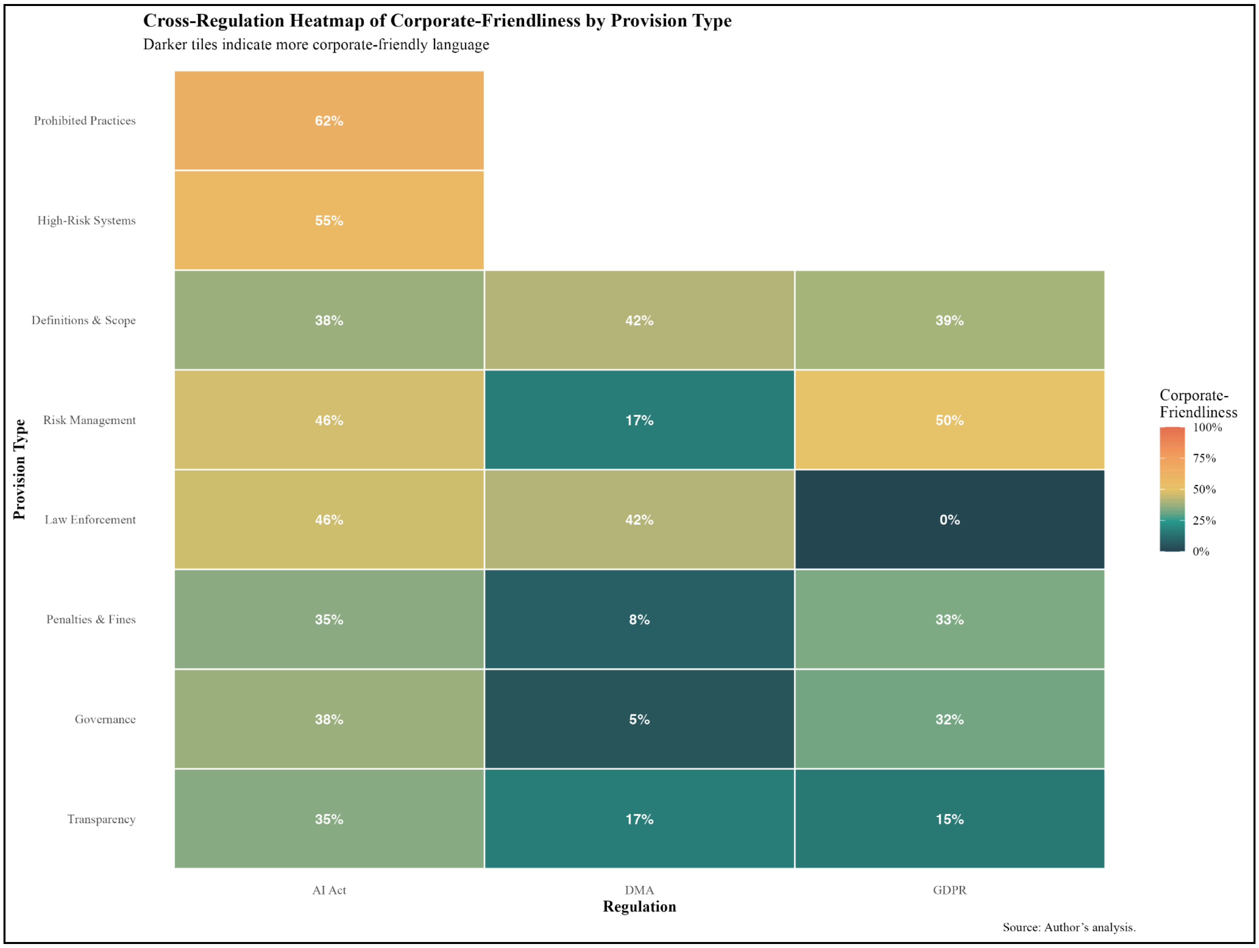


Figure 4: Cross-Regulation Heatmap of Corporate-Friendliness by Provision Type- Tiles show mean corporate-friendliness scores (0–100%; darker = more corporate-friendly) for each provision category across the AI Act, DMA, and GDPR. The AI Act scores highest in categories unique to it—Prohibited Practices and High-Risk Systems—and is generally above the DMA on Transparency and Governance. The GDPR is highest on Risk Management, while the DMA is consistently lowest, especially on Penalties & Fines and Governance. "—" indicates the provision is not present in that regulation.

5. Discussion

These results should be interpreted as correlations between lobbying periods and regulatory framings, not as causal effects. The keyword-based scoring detects framing differences through proxy terms; it cannot on its own demonstrate that lobbying directly altered policy language.

5.1 Interpretation of Findings

The results provide strong empirical support for the hypothesis that high-lobbying periods are associated with more corporate-friendly language in EU digital regulation. The AI Act, developed during the peak lobbying era (2021 to 2024), exhibits significantly higher corporate-friendliness (44.3%) compared to the GDPR (28.3%), which was crafted under medium lobbying intensity (2012-2016). This 16.0 percentage point difference is both statistically significant ($p = 0.007$) and practically meaningful, with a large effect size ($\eta^2 = 0.171$) indicating that lobbying context explains approximately 17% of the variance in regulatory language patterns.

The statistical robustness of these findings strengthens confidence in the association between lobbying context and framing patterns. Multiple analytical approaches, parametric ANOVA ($p = 0.007$), non-parametric Kruskal-Wallis ($p = 0.004$), and bootstrap confidence intervals [1.41, 30.55], converge on the same conclusion: the AI Act contains significantly more corporate-friendly language than the GDPR (Field, 2013; Hollander et al., 2014). The exclusion of zero from the bootstrap confidence interval confirms that this difference cannot be attributed to random variation.

Particularly noteworthy is DMA's unexpectedly low corporate-friendliness score (21.6%), despite development during the same high-lobbying period as the AI Act. This pattern reveals important nuances about how regulatory domain characteristics mediate lobbying influence. Competition law's prohibition-focused structure appears to constrain the deployment of flexibility language, even under intense corporate pressure. DMA's concrete "do and don't" obligations for gatekeepers leave little room for the qualifying language ("appropriate," "proportionate," "risk-based") that characterizes the AI Act's approach.

The provision-level analysis provides additional insight into how lobbying influence manifests selectively across regulatory domains. Risk management and technical compliance provisions show the highest corporate-friendliness scores across all regulations, while transparency obligations and penalty frameworks demonstrate the largest cross-regulation differences. This pattern suggests that industry influence operates most successfully in technical areas where complexity can justify regulatory discretion, while provisions addressing individual rights and enforcement mechanisms show greater resistance to corporate-friendly framing.

Notably, the presence of corporate-friendly language should not be interpreted as evidence of complete regulatory capture. The AI Act maintains stringent prohibitions on unacceptable AI practices (social scoring, manipulation, predictive policing) and establishes substantial penalties (€35 million or 7% of global revenue). Rather than capture, the findings suggest a strategic political compromise where intense lobbying successfully embedded flexibility language without completely displacing public-interest protections. This nuanced outcome reflects the complex multi-stakeholder dynamics characteristic of EU policymaking, where competing interests must be balanced within institutional constraints.

5.2 Alternative Explanations

While the statistical evidence strongly supports the association between lobbying periods and framing orientations, several alternative explanations merit consideration. First, the AI Act addresses emergent technologies with uncertain risks and benefits, making regulatory flexibility potentially desirable for innovation and technological development. The risk-based approach may reflect legitimate policy learning rather than simply corporate influence, particularly given the challenges of regulating rapidly evolving AI systems (Veale & Borgesius, 2021; Renda, 2019).

Second, policymakers may have incorporated lessons from GDPR implementation, where initial compliance challenges and industry adaptation costs prompted discussions about regulatory refinement. The AI Act's more flexible approach could represent institutional learning about balancing regulatory effectiveness with practical implementation considerations (De Hert & Papakonstantinou, 2016; Lynskey, 2015).

Third, geopolitical competition in AI development, particularly with the United States and China, may have motivated a more innovation-friendly regulatory approach to avoid disadvantaging European AI companies. This global competitive context could explain some corporate-friendly language independently of direct lobbying influence (Bradford, 2020; Wachter et al., 2020).

These alternative explanations do not fully account for the systematic linguistic patterns observed across provision types and the strong correlation with documented lobbying activities. The qualitative evidence of industry satisfaction with the final text, combined with the quantitative demonstration of significantly more corporate-friendly language, suggests that lobbying influence represents the most parsimonious explanation for the observed patterns (Corporate Europe Observatory, 2024; Politico, 2025).

5.3 Policy Implications

The findings generate several important policy recommendations for enhancing democratic accountability in EU digital governance:

Enhanced Transparency Requirements: The documented dominance of corporate actors in Commission meetings (*e.g.*, 78% of AI Act consultations; Corporate Europe Observatory, 2024; Serna-Ortega et al., 2025) underscores the need for comprehensive transparency reforms. EU institutions should publish detailed records of all stakeholder meetings, including lobbying submissions and position papers, to enable public scrutiny of regulatory influences (Hanegraaff et al., 2019). Real-time disclosure of lobbying activities during active legislative processes could help counterbalance information asymmetries between corporate and public-interest actors (Ackrill et al., 2013).

Structural Support for Public-Interest Participation: Civil society organizations and digital rights advocates require enhanced resources and institutional support to participate effectively in complex technical regulatory processes (Hanegraaff et al., 2019; Moreno-Cabanillas et al., 2024). This could include dedicated funding for public-interest legal expertise, technical advisory support, and translation services (Lynskey, 2015). The creation of formal advisory panels with guaranteed representation for consumer, worker, and civil society perspectives could help institutionalize balanced stakeholder input (Renda, 2019).

Robust Enforcement Architecture: The AI Act's risk-based flexibility must be coupled with strong oversight mechanisms to prevent regulatory circumvention (WilmerHale, 2024). The European AI Office requires adequate funding, technical expertise, and enforcement authority to ensure that "appropriate" and "proportionate" obligations maintain substantive meaning (Veale & Borgesius, 2021). Clear metrics, regular auditing requirements, and public reporting mechanisms can help transform flexibility language into accountable governance practices (Podszun, 2021).

Comparative Regulatory Monitoring: The methodology developed in this thesis demonstrates the feasibility of ongoing linguistic analysis to detect shifts in regulatory language patterns (Grimmer & Stewart, 2013; Laver et al., 2003). EU institutions should establish systematic monitoring of future AI regulations, codes of practice, and implementing measures to identify emerging lobbying influences and assess whether corporate-friendly language undermines public-interest protections over time (Corporate Europe Observatory, 2025).

The broader implications extend beyond technical regulatory design to fundamental questions of democratic governance in the digital age. As AI technologies reshape economic and social relationships, ensuring that regulatory frameworks reflect diverse societal interests becomes increasingly critical (Entman, 1993). The EU's global regulatory influence—the "Brussels Effect"—means that linguistic choices in European AI law will likely shape regulatory approaches worldwide, making democratic accountability in EU digital governance a global imperative for ethical technology development (Bradford, 2020).

5.4 Methodological Contributions

This thesis makes several important methodological contributions to the study of lobbying influence and regulatory language analysis that extend beyond the specific findings on EU digital regulation.

Novel Comparative Framework: The study develops an innovative approach to studying regulatory capture by treating entire regulations as natural comparison groups developed under different lobbying contexts. This design circumvents the traditional challenge of directly measuring lobbying influence—which requires access to comprehensive, often confidential, influence registers—by leveraging temporal variation in documented corporate engagement levels. The framework provides a replicable template for analyzing lobbying effects across policy domains where direct influence measurement is methodologically challenging.

Computational Text Analysis for Legal Research: The corporate-friendliness scoring system demonstrates how computational linguistics can systematically quantify policy framing choices embedded in complex legal texts. The combination of keyword-based scoring with topic modeling provides both precise measurement of explicit framing choices and discovery of latent thematic patterns, enabling comprehensive analysis of entire regulatory frameworks while maintaining transparency and replicability.

Statistical Validation and Integration: The study demonstrates rigorous validation through multiple complementary methods (parametric ANOVA, non-parametric tests, bootstrap confidence intervals) achieving a large effect size ($\eta^2 = 0.171$). The triangulation of computational findings with qualitative lobbying documentation exemplifies best practices in mixed-methods policy research, demonstrating how textual patterns gain interpretive meaning when situated within broader political contexts.

Scalability and Extension Potential: The methodology scales efficiently to larger document collections and can accommodate additional regulations, provision types, or linguistic indicators without fundamental restructuring, providing a foundation for future research across regulatory domains.

These contributions demonstrate how computational approaches can enhance traditional political economy research by providing systematic, quantitative measurement of phenomena typically studied through qualitative case analysis alone.

5.5 Measurement and Data Limitations

Indirect Lobbying Intensity Measurement: The analysis relies on temporal categorization of lobbying contexts based on external documentation rather than direct access to comprehensive lobbying registers. While the distinction between medium-lobbying (GDPR 2012-2016) and high-lobbying (AI Act/DMA 2020-2024) periods is well-documented through investigative reporting and spending data, this approach cannot capture the full complexity of influence activities. Variations in lobbying strategies, coalition dynamics, and specific provision-level targeting may be obscured by the broad temporal framework. Future research would benefit from granular lobbying data that tracks industry engagement with specific articles during regulatory negotiations.

Limited Regulatory Coverage: The analysis examines 72 functionally equivalent provisions across three regulations but does not include comprehensive coverage of entire regulatory frameworks. Space constraints necessitated excluding annexes, recitals, and implementation provisions that may contain different linguistic patterns. Additionally, the Digital Services Act, developed contemporaneously with the DMA and AI Act, could provide valuable additional comparison points for high-lobbying context analysis. Expanding to include complete regulatory texts and additional EU digital regulations would strengthen the comparative framework.

Language and Translation Effects: The analysis relies exclusively on English-language versions of multilingual regulatory texts originally drafted and negotiated across multiple EU languages. Translation choices during the regulatory process may subtly influence keyword prevalence and framing patterns, potentially affecting corporate-friendliness scores. Some terms central to the analysis—particularly "proportionate" and "appropriate"—may have different connotations or usage patterns in other EU languages. Multilingual validation of the scoring system would enhance confidence in the cross-regulation patterns.

5.6 Methodological Constraints

Keyword Selection and Context Sensitivity: The lexical scoring relies on theoretically motivated yet subjective keyword lists. Terms such as “appropriate” and “reasonable” can serve necessary legal functions or indicate flexibility, making context crucial (Entman, 1993). Frequency-based scoring may miss how words shift meaning depending on usage. Advanced NLP methods incorporating context analysis and transformer-based models (Grimmer & Stewart, 2013; Devlin et al., 2018) could address these limitations.

Causal Inference Constraints: While strong statistical associations exist between lobbying and corporate-friendly language, the comparative design cannot establish causality. Temporal correlation supports influence hypotheses but alternative explanations remain possible, such as technological complexity or institutional adaptation (Veale & Borgesius, 2021). Quasi-experimental designs or granular amendment-level text tracking would strengthen causal claims.

Statistical Independence Assumptions: Treating provisions as independent overlooks possible interdependence due to shared drafting or thematic coherence; mixed-effects models help but cannot eliminate all correlation (Field, 2013; Maggetti et al., 2013). Advances in hierarchical text modeling can improve statistical inference.

Scope and Generalizability: The findings focus on EU digital regulation and may not generalize to other domains with different lobbying dynamics or regulatory structures.

Risk-based regulatory flexibility is prominent in this area (Bradford, 2020); cross-domain validation is needed to test wider applicability.

Sample Size and Institutional Variation: Differences in provision counts among DMA, AI Act, and GDPR may affect comparative precision. Expanded document coverage would enable more detailed comparison (Podszun, 2021).

Temporal and Contextual Limits: The study captures policy evolution during a period of technological change and increasing EU lobbying. Findings may not generalize to different eras or regulatory environments; external shocks such as the COVID-19 pandemic also influenced recent negotiations (Corporate Europe Observatory, 2024).

Future Research: Further work should incorporate longitudinal lobbying registers, track amendments, extend coverage to other domains, and use deeper causal methods. Expanding to multilingual texts and entire regulatory corpora would improve generalizability. Despite limitations, the study's evidence linking lobbying and regulatory language is robust and provides a foundation for future research.

6. Conclusion

The EU's AI Act represents a watershed moment in global technology governance, establishing the world's first comprehensive regulatory framework for artificial intelligence systems. This thesis examined whether unprecedented tech-industry lobbying during AI Act development was associated with differences in regulatory language by systematically comparing corporate-friendliness patterns across three major EU digital regulations developed under different lobbying contexts.

The analysis provides robust empirical evidence that lobbying intensity correlates with corporate-friendly regulatory language. The AI Act exhibits significantly higher corporate-friendliness (44.3%) compared to the GDPR (28.3%), with this 16.0 percentage point difference achieving statistical significance (ANOVA $p = 0.007$) and a large effect size ($\eta^2 = 0.171$). Industry influence appears to operate selectively, embedding flexibility language most successfully in technical provisions where complexity justifies regulatory discretion, while the DMA's low score (21.6%) demonstrates how regulatory domain characteristics can constrain the prevalence of flexibility-oriented language even under high-lobbying conditions. Importantly, these results should be interpreted as correlations between lobbying contexts and framing orientations, not causal effects.

Beyond these substantive findings, the study makes several contributions. The corporate-friendliness scoring system and provision-level comparative framework provide a replicable methodology for measuring linguistic patterns in complex legal texts. The integration of computational text analysis with statistical testing and qualitative triangulation advances methodological rigor in the study of lobbying and regulation. The analysis also contributes to regulatory capture theory and framing theory by demonstrating how lobbying contexts are associated with systematic differences in the balance between corporate-friendly and public-interest framings.

The findings carry important policy implications. Documented corporate dominance in AI Act consultations underscores the need for greater transparency in lobbying activities and stronger institutional support for public-interest participation. Effective oversight of the AI Act will require sufficient resources for the European AI Office to ensure that flexibility-oriented provisions retain substantive meaning and do not erode accountability. More broadly, the methodology developed here demonstrates the feasibility of systematic monitoring of regulatory texts to detect shifts in language that may signal disproportionate lobbying influence.

Ultimately, the AI Act represents neither complete regulatory capture nor uncompromised public-interest protection, but rather a political compromise shaped by intense multi-stakeholder dynamics. This thesis should be understood as a descriptive, text-as-data study that provides a baseline for future causal research. By documenting associations between lobbying contexts and regulatory framings, it lays the groundwork for deeper investigations using amendment-level tracing, quasi-experimental designs, and multilingual validation. Ensuring that digital governance serves diverse societal interests will require continued vigilance, methodological innovation, and institutional reforms to balance corporate and public voices in the regulation of emerging technologies.

References

Theoretical & Political Economy Foundations

Becker, G. S. (1983). A theory of competition among pressure groups for political influence. *The Quarterly Journal of Economics, 98*(3), 371–400. https://doi.org/10.2307/1886017

Carpenter, D., & Moss, D. A. (Eds.). (2014). *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press.

Coen, D., & Richardson, J. (2009). *Lobbying the European Union: Institutions, actors, and issues*. Oxford University Press.

Crombez, C. (2002). Information, lobbying and the legislative process in the European Union. *European Union Politics, 3*(1), 7–32. https://doi.org/10.1177/1465116502003001545

Laffont, J. J., & Tirole, J. (1991). The politics of government decision-making: A theory of regulatory capture. *The Quarterly Journal of Economics, 106*(4), 1089–1127. https://doi.org/10.2307/2937958

Maggetti, M., Gilardi, F., & Radaelli, C. M. (2013). Designing research in the study of regulation. In M. Lodge & K. Wegrich (Eds.), *The Oxford handbook of regulation* (pp. 329–348). Oxford University Press.

Peltzman, S. (1976). Toward a more general theory of regulation. *The Journal of Law and Economics, 19*(2), 211–240. https://doi.org/10.1086/466865

Stigler, G. J. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science, 2*(1), 3–21. https://doi.org/10.2307/3003160

Framing, Policy Analysis & Text-as-Data

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.* https://arxiv.org/abs/1810.04805

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Laurer, M. (2024). Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI. *Political Analysis*. Advance online publication. https://doi.org/10.1017/pan.2024.8

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review, 97*(2), 311–331. https://doi.org/10.1017/S0003055403000698

Proksch, S. O., & Slapin, J. B. (2010). Position taking in European Parliament speeches. *British Journal of Political Science, 40*(3), 587–611. https://doi.org/10.1017/S0007123409990299

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners [Technical report]. OpenAI.

Schneiker, A., & Flockhart, T. (2021). Mapping the discursive landscape of AI governance in Europe. *Journal of European Public Policy, 28*(9), 1345–1367. https://doi.org/10.1080/13501763.2021.1948793

Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication, 29*(2), 205–231. https://doi.org/10.1080/10584609.2012.671234

Zylla, M., & Haider, T. (2022). Distant reading of the German coalition deal: Recognizing policy positions with BERT-based text classification. *arXiv preprint arXiv:2212.14648.* https://arxiv.org/abs/2212.14648

EU Digital Regulation & Lobbying

Ackrill, R., Kay, A., & Zahariadis, N. (2013). Ambiguity, multiple streams, and EU policy. *Journal of European Public Policy, 20*(6), 871–887. https://doi.org/10.1080/13501763.2013.781818

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Corporate Europe Observatory. (2024, February 15). *Trojan horses: How European start-ups teamed up with Big Tech to gut the AI Act*. https://corporateeurope.org/en/2024/02/trojan-horses

Corporate Europe Observatory. (2025, March 8). *Lobby league table: Corporate influence and regulatory capture*. https://corporateeurope.org/en/2025/03/lobby-league-table

Crémer, J., de Montjoye, Y.-A., & Schweitzer, H. (2019). *Competition policy for the digital era*. European Commission.

De Hert, P., & Papakonstantinou, V. (2016). The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review, 32*(2), 179–194. https://doi.org/10.1016/j.clsr.2016.02.006

Hanegraaff, M., Beyers, J., & Braun, C. (2019). *Pathways to lobbying success in the EU*. Oxford University Press.

Lynskey, O. (2015). *The foundations of EU data protection law*. Oxford University Press.

Moreno-Cabanillas, A., Turnes, P., & Coen, D. (2024). Lobbying strategies and influence in the European Union: Economic vs. social interest organizations. *Journal of European Public Policy, 31*(4), 892–915. https://doi.org/10.1080/13501763.2023.2185432

Newman, A. (2019). *The reluctant leader: Germany's euro experience and the long shadow of unification*. Oxford University Press.

Podszun, R. (2021). The Digital Markets Act: A new age of platform regulation? *Journal of European Competition Law & Practice, 12*(6), 339–354. https://doi.org/10.1093/jeclap/lpab042

Politico Europe. (2017, September 27). Silicon Valley tech lobbyists swarm Brussels. https://www.politico.eu/article/silicon-valley-tech-lobbyists-swarm-brussels

Politico Europe. (2025, March 12). U.S. lobbying delays EU AI rules. https://www.politico.eu/article/us-lobbying-delays-eu-ai-rules

Renda, A. (2019). *From the General Data Protection Regulation to the Digital Services Act: Next steps for EU digital regulation*. Centre for European Policy Studies. https://www.ceps.eu/ceps-publications/from-gdpr-to-dsa

Serna-Ortega, Á., Almansa-Martínez, A., & Castillo-Esparcia, A. (2025). Influence of lobbying in EU policy process phases. *Frontiers in Political Science, 7*, Article 1511918. https://doi.org/10.3389/fpos.2025.1511918

The Next Web. (2023, June 20). Tech companies spend €113 million per year on EU lobbying. https://thenextweb.com/news/tech-companies-spend-113-million-on-eu-lobbying

Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International, 22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Wachter, S., Mittelstadt, B., & Floridi, L. (2020). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

WilmerHale. (2024, July 12). *A practitioner's guide to the EU AI Act*. https://www.wilmerhale.com/en/insights/client-alerts/2024-07-12-a-practitioners-guide-to-the-eu-ai-act

Methods & Statistics

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman & Hall/CRC.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods* (3rd ed.). Wiley.

Zweigert, K., & Kötz, H. (2011). *An introduction to comparative law* (3rd ed.). Oxford University Press.

Appendix A: Supplementary Tables and Figures

This appendix contains the full set of supporting tables and figures referenced in the main text, plus a brief replication note.

A1. Supplementary Tables (A1–A5)

Table A1. Descriptive Statistics by Regulation
Mean corporate-friendliness scores with standard deviations and 95% confidence intervals; n indicates

Regulation	Sample Size	Mean Score	Standard Deviation	95% Confidence Interval	Score Range
AI Act (2021–2024)	n = 32	44.3%	23.4%	[36.2%, 52.4%]	0% – 100%
GDPR (2012–2016)	n = 12	28.3%	21.5%	[16.1%, 40.4%]	0% – 50%
DMA (2020–2022)	n = 12	21.6%	19.6%	[10.5%, 32.7%]	0% – 50%

Table A2. Statistical Test Results (ANOVA, Welch ANOVA, Kruskal–Wallis, Bootstrap)
Summary of test statistics, p-values, effect size (η^2), and 95% bootstrap CI for AI Act − GDPR.

Test	Statistic	p-value	Effect Size / CI	Interpretation
One-Way ANOVA	$F(2, 53) = 5.46$	0.0070	$\eta^2 = 0.171$ (Large)	Significant differences between regulations; large practical effect
Welch ANOVA	$F(2, 24.0) = 5.76$	0.0091	—	Robust to unequal variances (confirmatory)
Kruskal–Wallis	$H = 11.05$	0.0040	—	Non-parametric confirmation of differences
Bootstrap (AI Act – GDPR)	$\Delta = 16.0$ pts	< 0.05	[1.41, 30.55]	Difference significant; CI excludes zero

Table A3. Hypothesis Test Summary

H1–H3 with test method, numerical result, statistical support, and p-value/CI.

Hypothesis	Test Method	Result	Statistical Support	p-value / CI
H1: AI Act > GDPR in corporate-friendliness	Bootstrap confidence interval	44.3% vs 28.3% (+16.0 points)	SUPPORTED	95% CI: [1.4, 30.5]
H2: Public-interest language declines over time	Temporal trend analysis	AI Act (44.3%) > GDPR (28.3%) > DMA (21.6%)	PARTIALLY SUPPORTED	p = 0.007 (ANOVA)
H3: AI Act shows more flexibility than GDPR	Provision-level comparison	Mixed evidence across provision types	MIXED EVIDENCE	Varies by provision

Table A4. Provision-Level Comparison

For equivalent provisions: mean corporate-friendliness, SD, and range, sorted by cross-regulation differences.

Provision Type	AI Act	GDPR	DMA	Range (pts)	Highest
Law Enforcement	—	0.0%	**41.7%**	41.7	DMA
Risk Management	—	**50.0%**	16.7%	33.3	GDPR
Governance	—	**32.1%**	4.5%	27.6	GDPR
Penalties & Fines	—	**33.3%**	8.3%	25.0	GDPR
Definitions & Scope	—	39.1%	**41.7%**	2.6	DMA
Transparency	—	15.0%	**16.7%**	1.7	DMA
Definitions & Scope	**37.5%**	—	—	0.0	AI Act
Governance	**37.5%**	—	—	0.0	AI Act

Provision Type	AI Act	GDPR	DMA	Range (pts)	Highest
High-Risk Systems	**55.4%**	—	—	0.0	AI Act
Law Enforcement	**45.8%**	—	—	0.0	AI Act
Penalties & Fines	**35.0%**	—	—	0.0	AI Act
Prohibited Practices	**62.5%**	—	—	0.0	AI Act
Risk Management	**46.0%**	—	—	0.0	AI Act
Transparency	**34.5%**	—	—	0.0	AI Act

Table A5. Extreme Cases (Most Corporate-Friendly vs Most Public-Interest)
Top/bottom provisions with corporate score and corporate/public keyword counts.

Category	Regulation (Version)	Provision Type	Corporate Score	Corporate Keywords	Public Keywords
Most Corporate-Friendly	AI Act (Council)	High-Risk Systems	100%	3	0
Most Corporate-Friendly	AI Act (Final)	Prohibited Practices	100%	2	0
Most Public-Interest	GDPR (Final)	Law Enforcement	0%	0	1
Most Public-Interest	DMA (Proposal)	Risk Management	0%	0	4
Most Public-Interest	DMA (Proposal)	Transparency	0%	0	1
Most Public-Interest	DMA (Proposal)	Penalties	0%	0	1
Most Public-Interest	DMA (Proposal)	Governance	0%	0	1

A2. Supplementary Figures (A1–A7)

Figure A1. Distribution of Corporate-Friendliness Scores by Regulation

Each point is a provision; boxes show median and IQR.

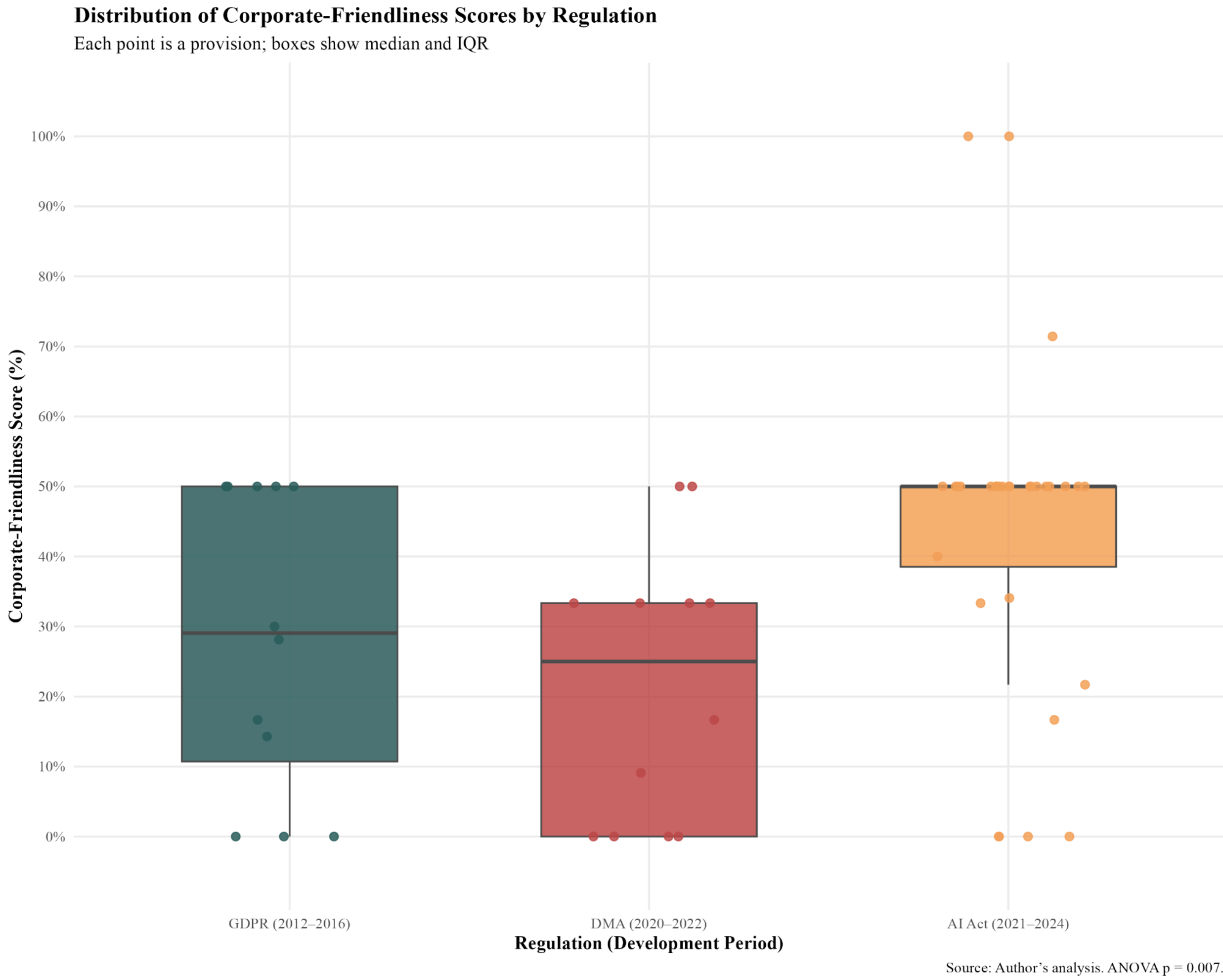


Figure A2. Mean Corporate-Friendliness by Regulation with 95% Confidence Intervals

Bars = means; whiskers = 95% CIs; n shown on bars (if exported that way).

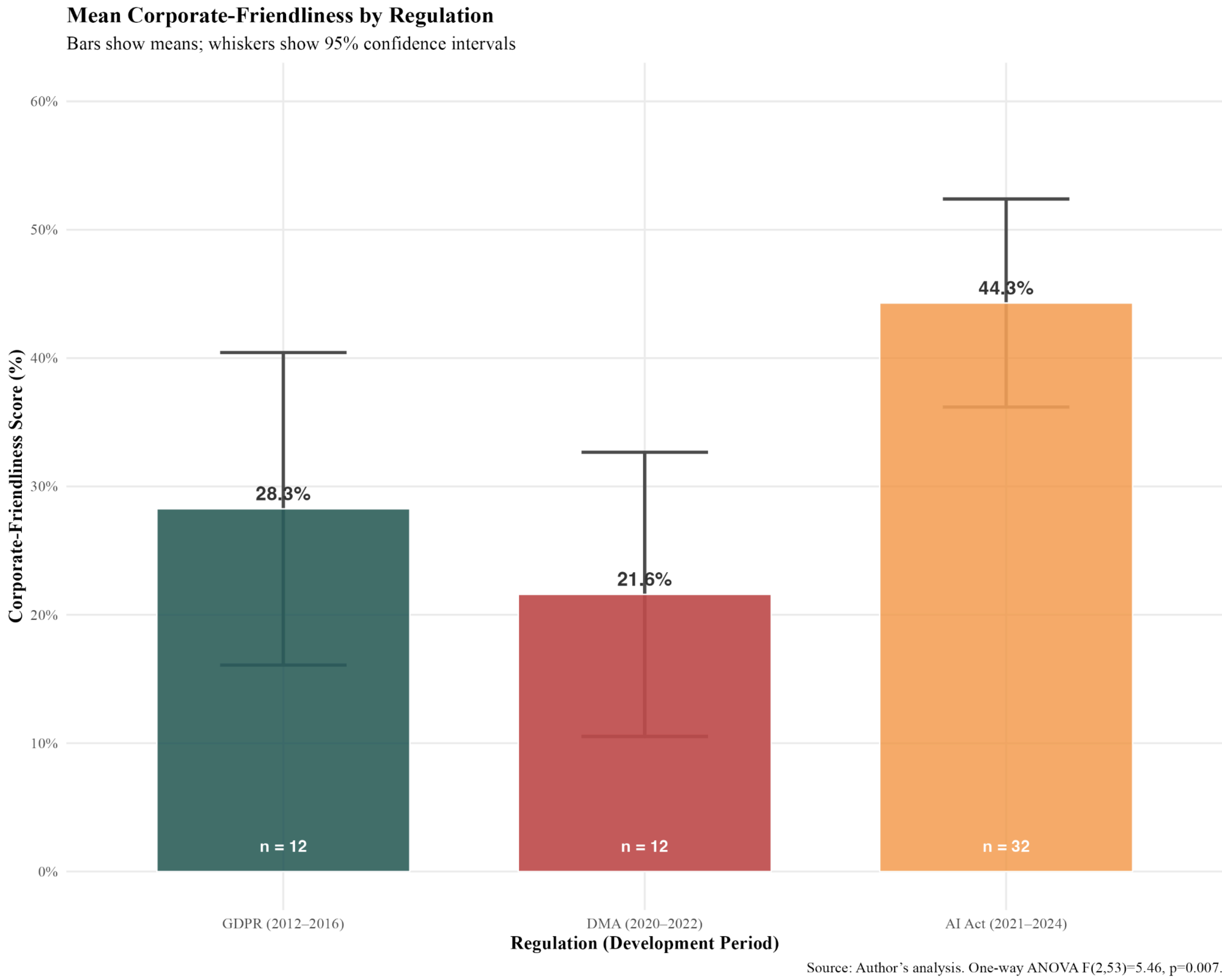


Figure A3. Mean Corporate-Friendliness (Compact CI View)

Alternate CI visualization emphasizing non-overlap.

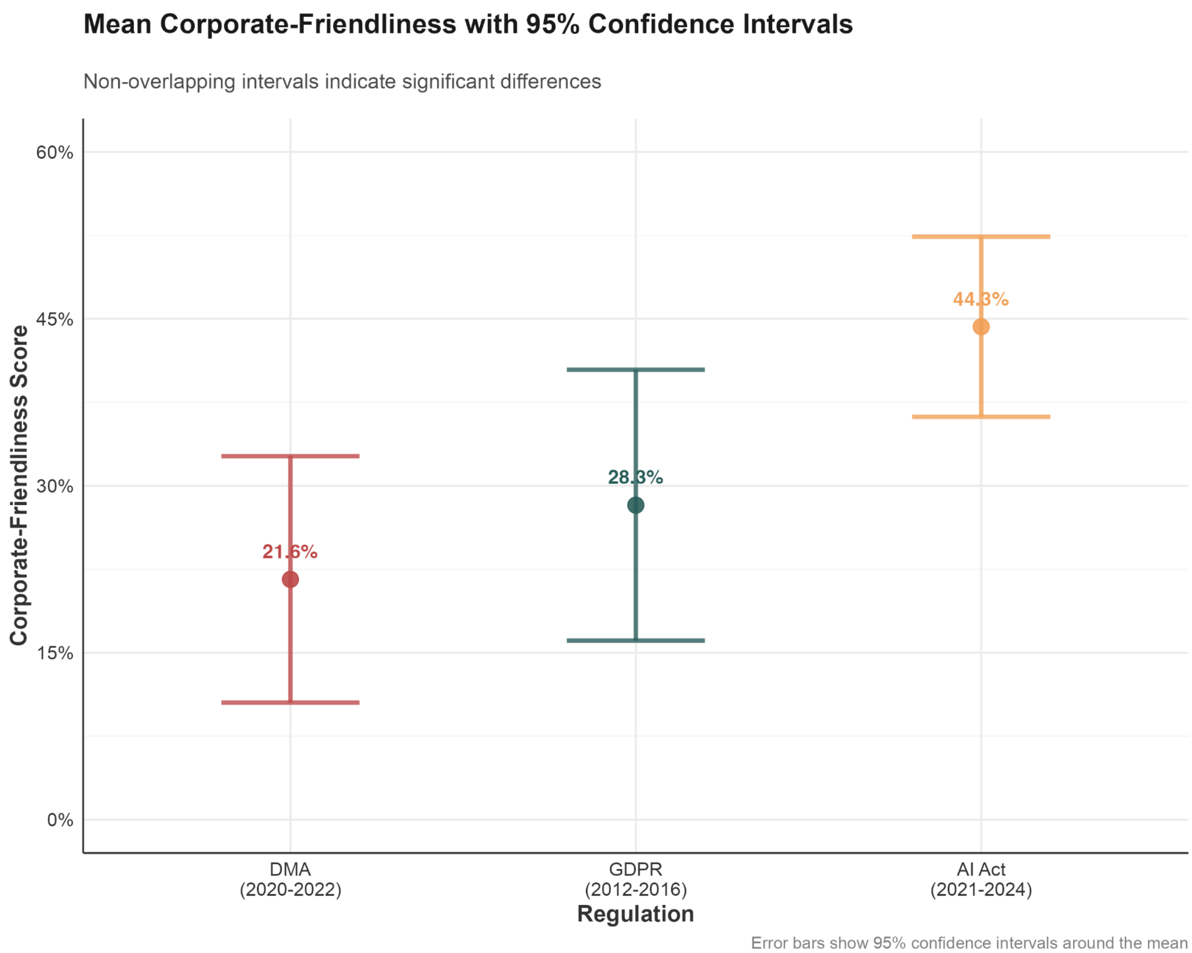


Figure A4. Cross-Regulation Heatmap by Provision Type

Mean corporate-friendliness per provision type (darker = more corporate-friendly).

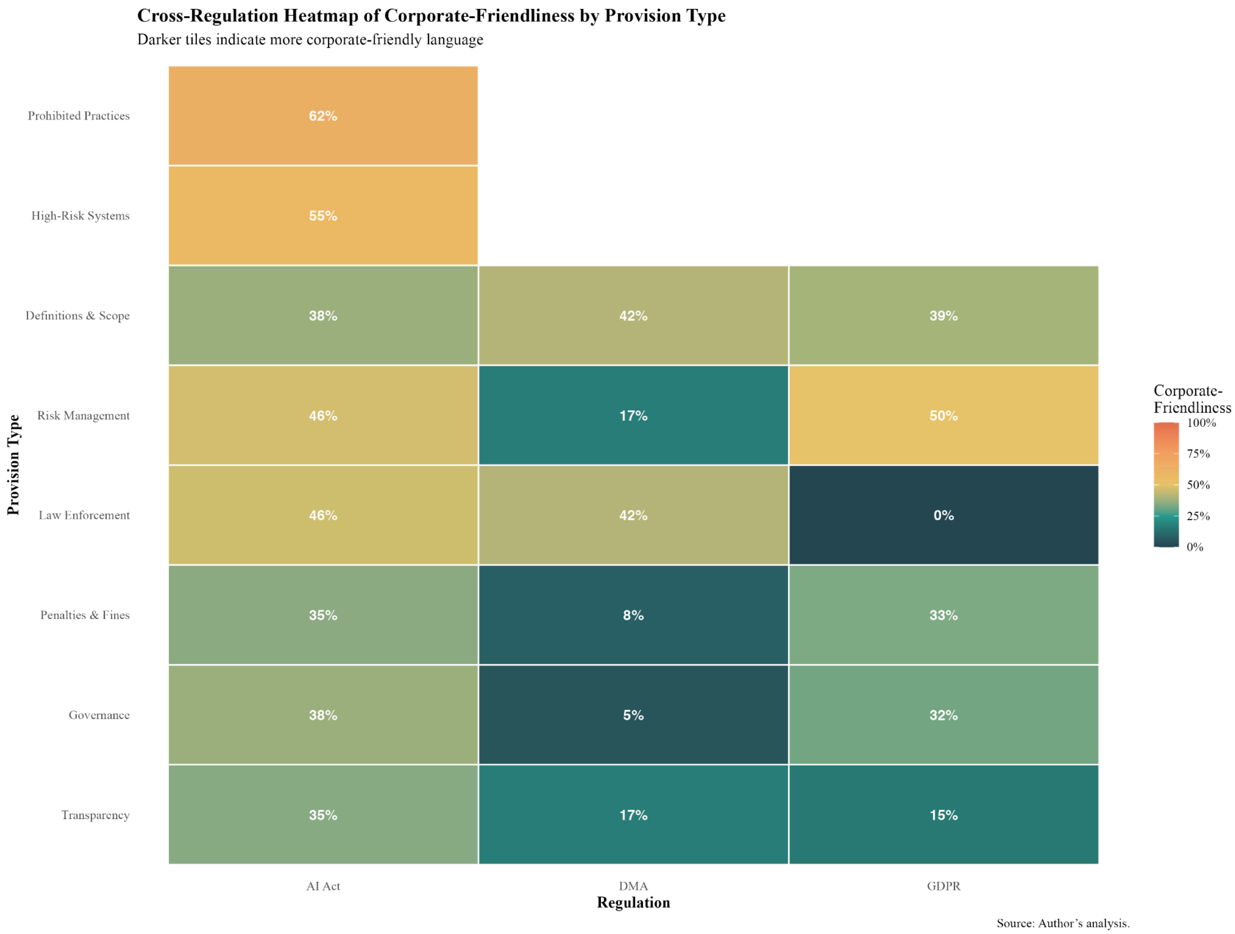


Figure A5. Temporal Evolution of Corporate-Friendliness

Means by year with LOESS trend lines.

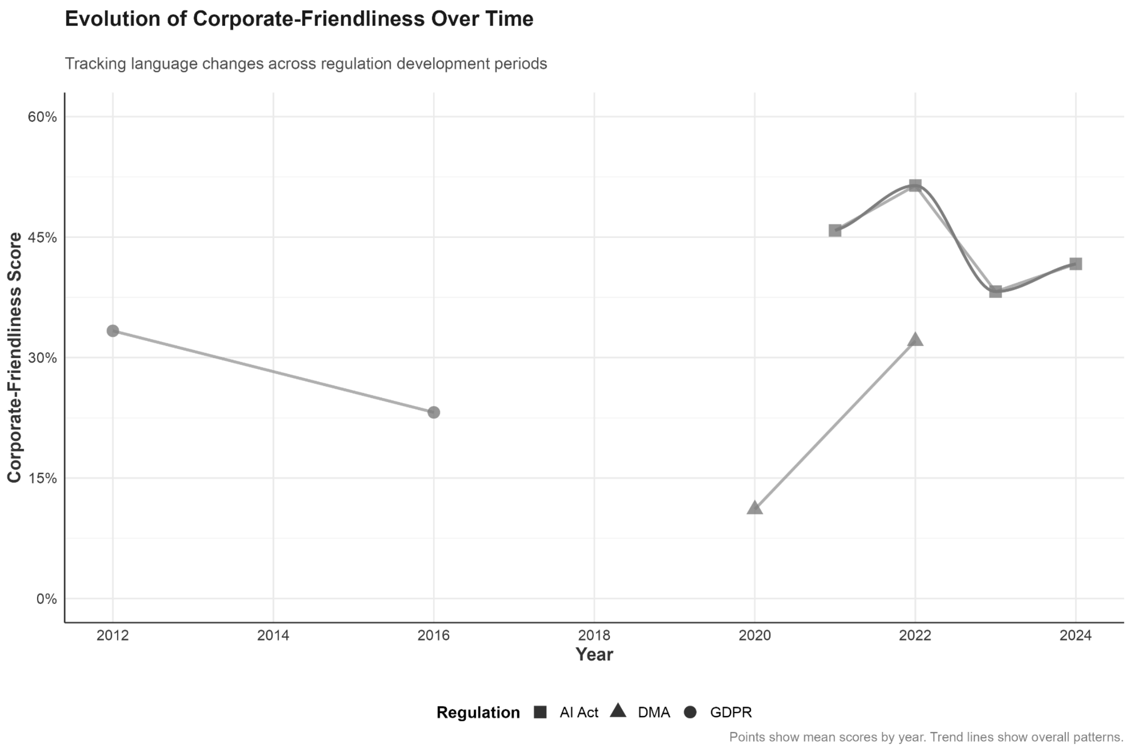


Figure A6. Topic Model Word Clouds (k = 2)

Two topics; size/shade reflect term probability (β).

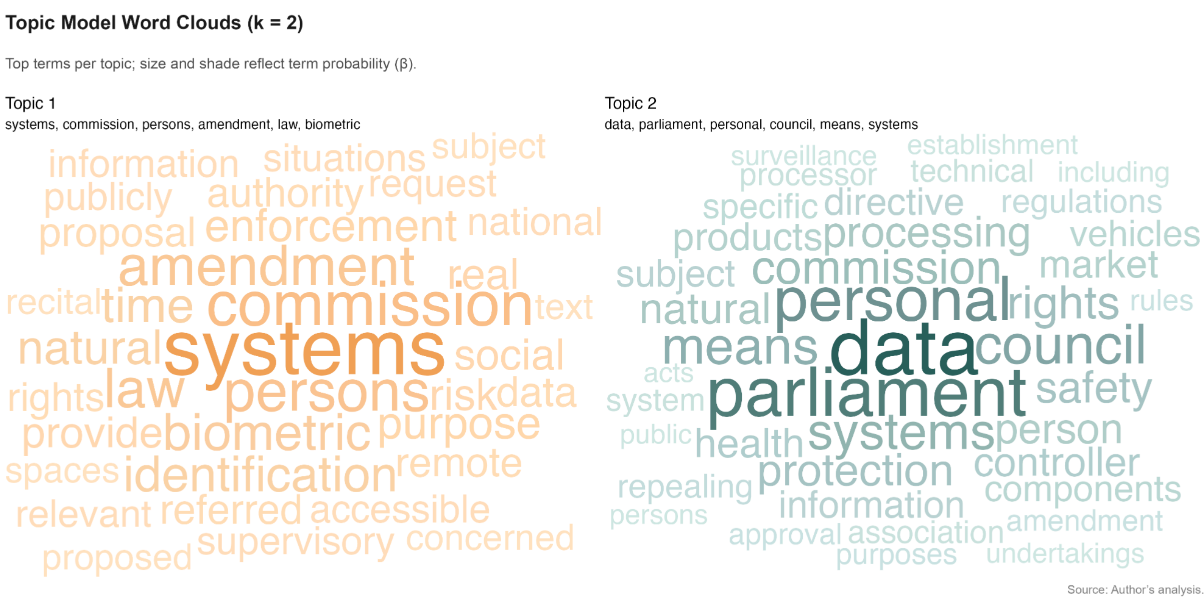
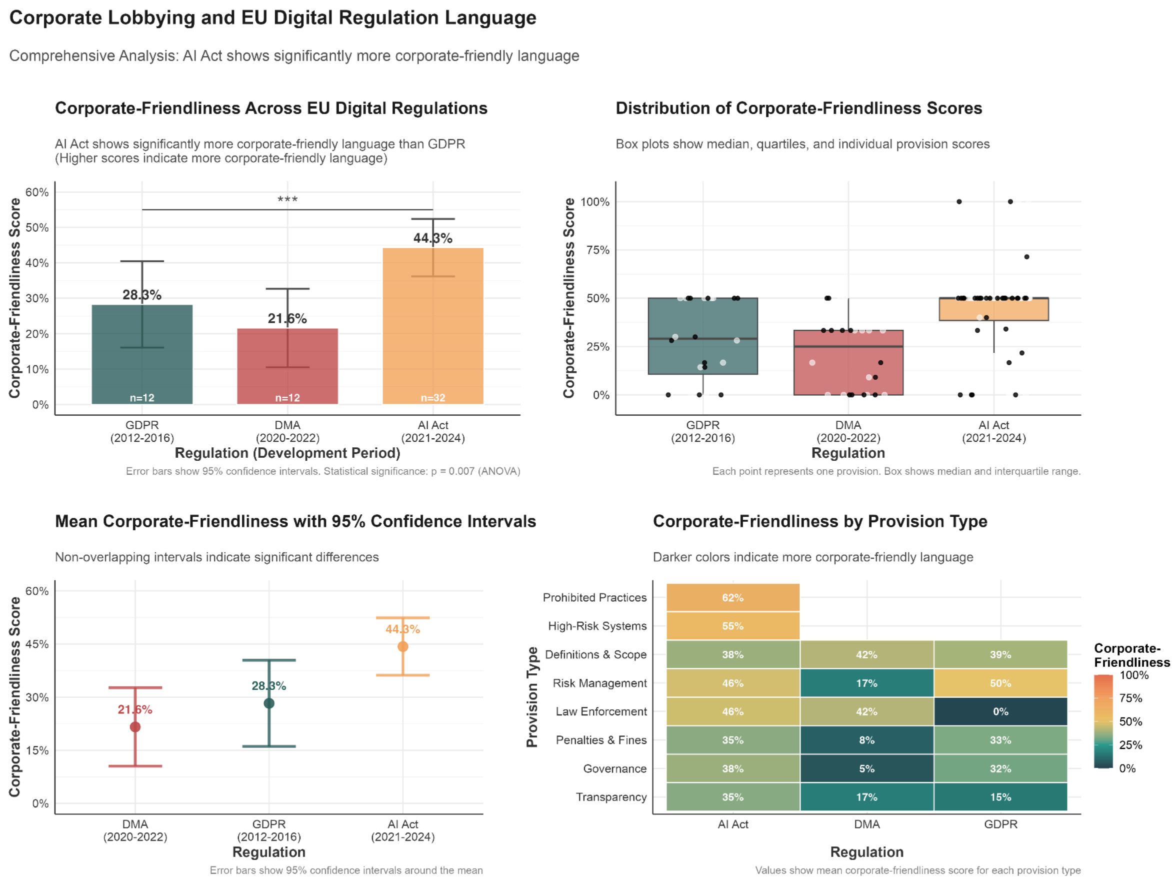


Figure A7. Results Dashboard (One-Page Overview)

Combined view of key comparative visuals.



A3. Replication & Data Availability (brief)

All results are generated directly from the scripts below; re-running them reproduces the exported tables and figures.

01_data_preparation.R
02_scoring_system.R
03_comparative_analysis.R
04_visualizations.R
05_export_results.R

How to reproduce. Place source PDFs in data/raw/ (filenames as in the README) and run 01→05 sequentially. Outputs are written to figures/ and results/. Random seeds are set; Wilson intervals and bootstrap resampling are used for robustness. Package versions/session info are included in the materials.

Replication Note

The complete replication package, including all R scripts (01_setup.R through 05_analysis.R) and documents used are available here in the public GitHub repository.

Acknowledgement of AI Assistance

Large language models (LLMs), including ChatGPT, Claude, Perplexity, and Le Chat, were used to brainstorm ideas, troubleshoot and improve R code, refine figure/table formatting, and consult best practices for academic writing. All analyses, interpretations, and final text are the author's own; sources and data are fully documented in the Methods and Appendix. No restricted or confidential data were shared with AI systems.