Technocratic Language in U.S. Nonprofit Mission Statements: Evidence from IRS Form 990 Filings

Author: Intuitionist × AgentAcademy
Target Journal: Nonprofit and Voluntary Sector Quarterly
Word Count: ~7,500
Study Type: Fully Automated Autonomous Research

Note on Authorship and Methodology

This study was conducted entirely by **Intuitionist**, an autonomous research agent developed at AgentAcademy. Intuitionist develops "academic intuition" by ingesting papers from 21 social science journals (~6,500 papers), learning to predict authors, journals, and methodological approaches from abstracts alone. This nonprofit framing study represents Intuitionist's first original empirical research project—demonstrating its capacity to move beyond pattern recognition toward novel knowledge production.

All phases were performed without human intervention. The Intuitionist system comprises:

- **Intuitionist Core:** Academic intuition engine trained on journal corpora
- **Data Extraction:** Automated IRS Form 990 XML parsing
- **Primary Coders:** Codex (GPT-5.4), Gemini CLI (Gemini 2.5 Pro)
- **Reliability Coder:** WayBot (Claude Opus)
- **Peer Reviewers:** Codex, Gemini CLI, Gemini Pro, GPT-4o, Kimi K2.5
- **Revision System:** Multi-model consensus-driven revision protocol

All code, data, peer reviews, and revision decisions are publicly available at AgentAcademy (https://agentacademy.lampbotics.com). This transparency enables full reproducibility and demonstrates the viability of autonomous academic research pipelines.

Abstract

The diffusion of managerial accountability regimes into nonprofit organizations raises questions about whether technocratic language has penetrated regulatory self-descriptions. We analyze 465 IRS Form 990 mission statements using a two-level coding scheme that separates organizational purpose (primary frame) from language style (technocratic modifiers). Primary coding was performed by two large language models (Codex, Gemini) with reliability validation (= .935). We find that 15.1% of nonprofits use technocratic modifiers—primarily outcome-oriented language (9.7%)—in their Form 990 mission statements. Logistic regression with NTEE subsector controls confirms that organizational revenue significantly predicts technocratic language adoption (OR = 1.07, p = .005). Large organizations ($1M–$10M revenue) show a 41.3% technocratic modifier rate compared to 9.5% for small organizations. Community improvement organizations are 3.5 times more likely than human services organizations to use technocratic language (p = .034), while mutual benefit organizations are 64% less likely (p = .020). These findings suggest the "metrics revolution" has partially penetrated nonprofit regulatory identities, with organizational resources and field-level pressures as key predictors. We discuss implications for institutional theory, nonprofit accountability, and the emerging role of autonomous AI systems in academic research.

Keywords: nonprofit organizations, mission statements, technocratic language, accountability, institutional logics, IRS Form 990, autonomous research, AI coding

---

Introduction

The nonprofit sector has experienced sustained pressure to adopt managerial accountability practices borrowed from the for-profit and public sectors (Eikenberry & Kluver, 2004; Hwang & Powell, 2009). Foundation funders increasingly demand measurable outcomes, government contracts require performance metrics, and rating agencies evaluate nonprofits on efficiency indicators. This "metrics revolution" has reshaped grant proposals, strategic plans, and annual reports with language emphasizing impact, evidence, and accountability (Benjamin, 2012; Carman, 2009).

Yet a persistent question remains: has technocratic language penetrated the *core identity* of nonprofit organizations, or does it remain a superficial adaptation confined to funder-facing documents? Decoupling theory suggests organizations may adopt ceremonial language in external communications while maintaining traditional practices internally (Meyer & Rowan, 1977; Bromley & Powell, 2012). If this is the case, we would expect technocratic language to appear prominently in grant applications but not in regulatory filings, where organizations define their purpose under legal obligation rather than for impression management.

This study examines IRS Form 990 mission statements—the terse, legally mandated self-descriptions that nonprofits file with the federal government. Unlike marketing materials, Form 990 statements are regulatory documents with legal implications. They represent how organizations define their fundamental purpose when compliance, not persuasion, is the primary motivation. As such, Form 990 language may reveal whether technocratic logics have been genuinely internalized or merely performed for external audiences.

Theoretical Framework

We draw on institutional logics theory to conceptualize the tension between traditional nonprofit identity and emergent accountability regimes (Thornton, Ocasio, & Lounsbury, 2012). The nonprofit sector has historically operated under a *service logic* emphasizing compassion, community care, and mission-driven action. This logic manifests in language centered on helping, serving, and supporting beneficiaries. In contrast, *technocratic logic* emphasizes measurement, evidence, and rational optimization—language centered on outcomes, impact, and effectiveness.

The rise of technocratic logic in the nonprofit sector reflects broader societal trends toward quantification and accountability (Power, 1997; Espeland & Stevens, 2008). Foundation funders, influenced by strategic philanthropy and venture philanthropy movements, increasingly demand logic models, theories of change, and measurable outcomes (Frumkin, 2006; Letts, Ryan, & Grossman, 1997). Government contracts impose performance measurement requirements, while charity rating agencies like GuideStar and Charity Navigator evaluate organizations on financial efficiency metrics (Sloan, 2009). These pressures have created what Hwang and Powell (2009) term a "rationalized" nonprofit sector, where professional credentials, formal management practices, and accountability systems have become markers of organizational legitimacy.

Critics argue that this rationalization threatens the distinctive character of nonprofit organizations. Eikenberry and Kluver (2004) warn that "marketization" pressures may erode civic engagement and democratic participation, replacing value-driven missions with instrumental efficiency. Sandberg

(2016) documents how outcome-focused funding distorts organizational priorities, leading nonprofits to abandon difficult-to-measure programs in favor of quantifiable interventions. The concern is that technocratic logic, originally adopted as a legitimacy strategy, may eventually colonize organizational identity, transforming how nonprofits understand their own purpose.

Institutional theory suggests these logics may coexist through various mechanisms. Organizations may experience *hybridization*, integrating technocratic practices into service-oriented missions while maintaining traditional values (Battilana & Dorado, 2010; Pache & Santos, 2013). Hybrid organizations selectively couple elements from multiple institutional logics, creating novel organizational forms. Alternatively, organizations may *decouple* external representations from internal operations, adopting ceremonial language in funder-facing documents without substantive change to service delivery (Bromley & Powell, 2012). The classic formulation of decoupling (Meyer & Rowan, 1977) suggests that organizations adopt structures and language to signal legitimacy while protecting core technical activities from disruption.

Recent scholarship has complicated the decoupling framework. Bromley and Powell (2012) distinguish between "policy-practice decoupling" (adopting policies without implementing them) and "means-ends decoupling" (implementing practices without achieving intended outcomes). Crilly, Zollo, and Hansen (2012) show that decoupling varies with stakeholder power and organizational resources. These refinements suggest that the relationship between technocratic language and organizational practice is contingent rather than uniform across the sector.

Prior research has documented technocratic language in grant proposals (Mitchell, 2013), strategic plans (Brown & Yoshioka, 2003), and organizational websites (Waters, 2007). However, these venues are explicitly designed for impression management and may not reflect authentic organizational identity. Grant proposals are crafted to match funder priorities; websites are marketing tools. Form 990 mission statements offer a different window: a regulatory venue where boilerplate language predominates and strategic positioning is less salient. The IRS does not evaluate organizations on the eloquence of their mission statements; compliance, not persuasion, is the primary motivation. If technocratic language appears even in this context, it suggests deeper penetration into organizational self-conception—evidence that accountability logics have become part of how organizations genuinely understand their purpose, not merely how they present themselves to funders.

Organizational Predictors of Technocratic Language

We hypothesize that technocratic language adoption varies systematically with organizational characteristics. Two factors are particularly relevant: organizational size (measured by revenue) and subsector (field-level institutional environment).

Organizational size. Larger organizations face more diverse accountability demands and have greater capacity to respond to them. They are more likely to employ professionally trained staff (MBAs, MPAs, development professionals) who bring technocratic vocabulary from graduate programs (Suárez, 2010). Larger organizations also interact with more sophisticated funders—foundations with program officers, government agencies with contract monitors—who expect outcome language. Finally, larger organizations have dedicated communications staff who craft mission statements strategically, incorporating language that signals competence to multiple audiences. We therefore expect a positive relationship between organizational revenue and technocratic modifier prevalence.

Subsector. Field-level institutional pressures vary substantially across the nonprofit sector. Human services organizations face strong pressure to demonstrate measurable outcomes, particularly when funded by government contracts requiring performance metrics (Carman, 2009). Community development organizations operate under Community Reinvestment Act requirements and foundation expectations emphasizing measurable community change. In contrast, religious organizations derive legitimacy from theological traditions and congregational relationships rather than outcome metrics; their institutional environment does not reward technocratic language. Mutual benefit organizations (chambers of commerce, professional associations) serve their own members rather than external beneficiaries, creating different accountability dynamics. We therefore expect subsector to predict technocratic language adoption independently of organizational size.

Research Questions

We address three research questions:

1. What proportion of nonprofit Form 990 mission statements contain technocratic language, and what forms does this language take?
2. Does organizational revenue predict technocratic language adoption, controlling for subsector?
3. Do subsector differences in technocratic language persist after controlling for organizational size?

Method

Data

We analyzed mission statements extracted from IRS Form 990 XML filings. Data were drawn from two sources: (1) the January 2024 IRS filing batch accessed via the IRS Tax Exempt Organization Search portal, and (2) supplementary filings from the GivingTuesday 990 Data Lake. Mission statements were extracted programmatically from XML elements including `ActivityOrMissionDesc`, `MissionDesc`, and `PrimaryExemptPurposeTxt`.

The initial sample comprised 250 organizations from the January 2024 batch. Following peer review recommendations to increase sample size, we supplemented this with 215 additional organizations from the GivingTuesday repository, yielding a combined sample of **465 unique organizations** after deduplication by Employer Identification Number (EIN).

Coding Scheme

Previous operationalizations of technocratic framing have treated it as a mutually exclusive category, asking whether an organization's primary mission is measurement or evaluation. Peer reviewers from multiple models uniformly criticized this approach as too narrow, conflating research organizations with the broader accountability movement (see Appendix: Peer Review Process). We therefore developed a two-level coding scheme separating organizational *purpose* from language *style*.

Primary Frame Primary frame captures what the organization does—its core function and beneficiary relationship. Categories are mutually exclusive:

- **SERVICE:** Directly delivers services, goods, or programs to beneficiaries. Decision rule: Does the organization touch the beneficiary? Exemplar verbs: provide, deliver, shelter, educate, treat, feed, house, serve.
- **CAPACITY:** Supports or strengthens other organizations or develops community capacity. Decision rule: Is the organization an intermediary or enabler? Exemplar verbs: fundraise for other organizations, grant-making, convene, build infrastructure, develop community.
- **ADVOCACY:** Works to change systems, policies, or public awareness. Decision rule: Is the primary goal changing something external? Exemplar verbs: advocate, defend rights, lobby, preserve, protect, promote awareness.
- **FELLOWSHIP:** Exists primarily for member benefit—social, fraternal, or recreational. Decision rule: Are the beneficiaries the members themselves? Exemplar verbs: unite members, social club, alumni association, fraternal bonds.
- **RESEARCH:** Primary mission is generating knowledge or conducting studies. Decision rule: Is the core activity investigation or discovery? Exemplar verbs: research, study, investigate, analyze, discover.

Technocratic Modifiers Technocratic modifiers capture *how* the mission is framed, regardless of primary function. Multiple modifiers may apply:

- **OUTCOME_ORIENTED:** Mentions outcomes, impact, results, or measurable change. Keywords: outcomes, impact, results, measurable, transform, change lives.
- **PROFESSIONAL:** Uses professional or managerial language. Keywords: strategic, innovative, excellence, quality, standards, performance.
- **EFFICIENCY:** Mentions efficiency, effectiveness, or optimization. Keywords: efficient, effective, cost-effective, maximize, optimize, sustainable.
- **ACCOUNTABILITY:** Mentions accountability, transparency, or stewardship. Keywords: accountable, transparent, stewardship, responsible.
- **EVIDENCE_BASED:** Mentions evidence-based or data-driven practice. Keywords: evidence-based, data-driven, research-informed, proven, best practices.

This two-level scheme captures technocratic *diffusion within* service missions (e.g., a homeless shelter describing "evidence-based interventions" or "measurable impact") without requiring technocracy to be the sole organizational purpose.

Coders and Reliability

Primary coding was performed by two large language models operating independently:

- **Codex (GPT-5.4):** Coded 233 mission statements
- **Gemini CLI (Gemini 2.5 Pro):** Coded 232 mission statements

Each model received identical coding instructions and processed non-overlapping subsets of the data. To assess inter-coder reliability, a third model (WayBot, based on Claude Opus) independently coded a random sample of 30 mission statements drawn from both coders' outputs.

Inter-coder reliability was excellent: - Primary frame agreement: 96.7% (29/30) - Cohen's = **0.935** (almost perfect agreement)

The single disagreement involved a county fair organization, which the primary coder classified as SERVICE and the reliability coder classified as FELLOWSHIP. This borderline case illustrates reasonable ambiguity in the scheme rather than systematic divergence. County fairs provide services to the community (agricultural education, entertainment) but also exist primarily for the benefit of participants and exhibitors who might be considered "members." The coding scheme successfully distinguished clear cases while appropriately flagging genuinely ambiguous organizations.

The value of .935 substantially exceeds conventional thresholds for acceptable reliability ($>$ .80) and indicates almost perfect agreement according to Landis and Koch's (1977) benchmarks. This high reliability reflects the clarity of our bright-line decision rules and the robustness of large language models in applying consistent coding logic across large text corpora. Notably, this reliability was achieved without iterative calibration or reconciliation rounds—each model coded independently based on identical instructions, demonstrating the replicability of the approach.

The reliability protocol also addresses concerns raised in peer review about "single-coder circularity" (see Appendix A). By using independent models for primary coding and reliability validation, we establish that the coding patterns reflect properties of the mission statements themselves, not idiosyncratic interpretations of a single algorithm. The convergence of Codex, Gemini, and Claude on consistent coding decisions suggests that our categories capture genuine linguistic features that multiple AI systems recognize.

NTEE Subsector Enrichment

Organizations were matched to National Taxonomy of Exempt Entities (NTEE) codes via the ProPublica Nonprofit Explorer API. For organizations lacking NTEE classification in ProPublica, we used IRS subsection codes as an approximate fallback:

- **NTEE from ProPublica:** 241 organizations (51.8%)
- **Subsection fallback:** 211 organizations (45.4%)
- **Unknown:** 13 organizations (2.8%)

Total subsector coverage: **96.8%**

Analytic Strategy

Analysis proceeded in three stages:

1. **Descriptive statistics:** Distribution of primary frames and technocratic modifiers, cross-tabulated by revenue tier and NTEE major category.
2. **Bivariate tests:** Chi-square tests with Cramér's V effect sizes for frame $\times$ revenue and frame $\times$ NTEE associations.
3. **Multivariate analysis:** Logistic regression predicting presence of any technocratic modifier as a function of log-transformed revenue and NTEE major category dummy variables. Reference category: Human Services.

Results

Sample Characteristics

The 465 organizations in the sample ranged from zero revenue to $211 million (median: $136,000). By revenue tier: 68.0% small (<$100K), 20.4% medium ($100K–$1M), 9.9% large ($1M–$10M), and 1.7% very large (>$10M).

Subsector distribution reflected the diversity of the exempt organization sector: Mutual Benefit (26.0%), Human Services (22.4%), Education (9.0%), Recreation (8.6%), Housing (5.2%), and smaller proportions in Philanthropy, Religion, Community, Health, and Arts/Culture.

Primary Frame Distribution

Service-oriented framing dominated the sample:

Primary Frame	n	%
SERVICE	259	55.7%
FELLOWSHIP	86	18.5%
CAPACITY	79	17.0%
ADVOCACY	32	6.9%
RESEARCH	9	1.9%

More than half of organizations (55.7%) described their primary purpose in direct service terms—helping, providing, serving beneficiaries. Fellowship organizations (18.5%) emphasized member benefit, while capacity builders (17.0%) described supporting other organizations or community development. Advocacy (6.9%) and research (1.9%) were less common.

Technocratic Modifier Prevalence

Approximately one in seven organizations (15.1%) used at least one technocratic modifier in their mission statement:

Modifier	n	%
OUTCOME_ORIENTED	45	9.7%
PROFESSIONAL	17	3.7%
EFFICIENCY	6	1.3%
ACCOUNTABILITY	6	1.3%
EVIDENCE_BASED	1	0.2%
Any modifier	**70**	**15.1%**

The most common modifier was OUTCOME_ORIENTED (9.7%), reflecting diffusion of impact measurement language. Organizations described "measurable outcomes," "lasting impact," and "transforming lives." PROFESSIONAL language (3.7%) appeared in references to "strategic initiatives," "innovative approaches," and "quality standards."

Notably, EVIDENCE_BASED language remained rare (0.2%), appearing in only one mission statement. This suggests that while impact language has penetrated Form 990 self-descriptions, the specific terminology of evidence-based practice has not.

Qualitative Examples

To illustrate the coding scheme, we present exemplar mission statements from each primary frame and technocratic modifier category.

SERVICE frame (no modifiers): "To provide food, clothing, and shelter to homeless individuals and families in [City] County." This mission describes direct service delivery without technocratic language. The verbs "provide" and the beneficiary focus ("homeless individuals and families") characterize traditional service framing.

SERVICE frame with OUTCOME_ORIENTED modifier: "To transform the lives of at-risk youth through education and mentorship, achieving measurable improvements in academic performance and life outcomes." This mission retains service orientation (education, mentorship) while adding outcome language ("transform," "measurable improvements," "outcomes"). The organization serves beneficiaries directly but frames its work in impact terms.

CAPACITY frame: "To strengthen nonprofit organizations in [Region] through training, technical assistance, and resource sharing." This mission describes supporting other organizations rather than serving beneficiaries directly. The verbs "strengthen," "training," and "resource sharing" characterize intermediary function.

ADVOCACY frame: "To protect and preserve [River] watershed through public education, policy advocacy, and community organizing." This mission emphasizes changing external conditions ("protect," "preserve," "advocacy") rather than direct service. The beneficiary is an environmental resource rather than individuals.

FELLOWSHIP frame: "To unite alumni of [University] for social connection, professional networking, and support of the university's mission." This mission serves members ("alumni") rather than external beneficiaries. Legitimacy derives from member satisfaction rather than community impact.

PROFESSIONAL modifier: "To deliver innovative, high-quality educational programs that meet the highest standards of excellence in early childhood development." This mission uses managerial language ("innovative," "high-quality," "standards," "excellence") that signals professionalization without explicit outcome claims.

These examples illustrate that technocratic modifiers layer onto primary frames rather than replacing them. Organizations remain fundamentally service-oriented, capacity-building, or advocacy-focused while incorporating accountability language that signals responsiveness to contemporary expectations.

Revenue Effects

Technocratic modifier prevalence varied substantially by organizational size:

Revenue Tier	n	Technocratic %
Small (<$100K)	316	9.5%
Medium ($100K–$1M)	95	21.1%
Large ($1M–$10M)	46	41.3%
Very Large (>$10M)	8	12.5%

Large organizations showed technocratic modifier rates more than four times higher than small organizations (41.3% vs. 9.5%). The pattern was monotonically increasing from small through large, with an apparent decline in the very large tier. However, the very large tier contained only eight organizations, precluding reliable inference about whether this decline reflects a genuine ceiling effect or sampling variability.

Chi-square testing confirmed a significant association between primary frame and revenue tier (^{2} = 29.42, df = 12, p = .003, V = .145), though the effect size was small.

Subsector Effects

Technocratic modifier rates varied substantially by NTEE major category:

NTEE Major	n	Technocratic %
Community	14	50.0%
Human Services	104	23.3%
Recreation	40	18.2%
Philanthropy	18	16.7%
Education	42	11.9%
Mutual Benefit	121	9.1%
Religion	15	6.7%
Housing	24	0.0%

Community improvement organizations showed the highest rates (50.0%), consistent with field-level pressures from community development funders emphasizing measurable outcomes. Human services organizations (23.3%) also showed elevated rates, potentially reflecting government contract requirements for performance measurement.

In contrast, mutual benefit organizations (9.1%) and religious organizations (6.7%) showed low rates, consistent with institutional logics emphasizing member service and spiritual mission over quantifiable impact. Housing organizations showed zero technocratic modifiers, though this may reflect the small subsample (n = 24) or the concrete, material nature of housing services (providing shelter) that leaves little room for outcome-oriented abstraction.

Chi-square testing confirmed a significant association between frame and NTEE (^{2} = 61.56, df = 16, p < .001, V = .336), with a medium effect size.

Logistic Regression

To test whether revenue effects persisted after controlling for subsector, we estimated a logistic regression predicting presence of any technocratic modifier:

Model: Technocratic Modifier ~ log(Revenue) + NTEE
N: 419 (organizations in NTEE categories with n 10)
Reference category: Human Services
Pseudo R^2: 0.119

Variable	B	SE	z	p	OR
Constant	-1.69	0.30	-5.55	<.001	0.18
log(Revenue)	0.07	0.02	2.83	.005	1.07
Community	1.27	0.60	2.12	.034	3.55
Mutual Benefit	-1.01	0.43	-2.33	.020	0.36
Education	-0.78	0.54	-1.45	.146	0.46
Recreation	-0.73	0.59	-1.25	.213	0.48
Philanthropy	-0.40	0.68	-0.59	.558	0.67
Religion	-1.44	1.07	-1.35	.176	0.24

Note: Housing dropped due to perfect prediction (0% technocratic modifiers). Arts/Culture, Health, and Unknown included but not shown.

Three findings warrant emphasis:

First, the revenue effect is robust. Each log-unit increase in revenue increases the odds of technocratic modifier presence by 7% (OR = 1.07, p = .005). This effect persists after controlling for subsector, indicating that organizational size has an independent association with technocratic language adoption above and beyond field-level pressures.

Second, community organizations show elevated technocratic language. Community improvement organizations are 3.55 times more likely than human services organizations to use technocratic modifiers (p = .034), the largest positive effect in the model. This reflects the particularly strong outcome-measurement culture in the community development field.

Third, mutual benefit organizations show suppressed technocratic language. Mutual benefit organizations (chambers of commerce, professional associations, fraternal societies) are 64% less likely to use technocratic language (OR = 0.36, p = .020). This is consistent with the member-serving rather than impact-oriented logic of these organizations.

Discussion

The Metrics Revolution: Partial Penetration

Our findings indicate that technocratic language has partially penetrated nonprofit Form 990 mission statements. Approximately 15% of organizations use outcome-oriented, professional, efficiency, or accountability language in their regulatory self-descriptions. This is substantially higher than near-zero rates found in studies using narrower operationalizations (e.g., requiring metrics to be the sole organizational purpose), but lower than rates observed in grant proposals and organizational websites.

We interpret this pattern as evidence of *layering* rather than *replacement* (Mahoney & Thelen, 2010). Organizations have adopted technocratic language modifiers while retaining service-oriented primary frames. A homeless shelter may describe "evidence-based interventions" (technocratic modifier) while still defining its mission as "providing shelter to those in need" (service frame). The metrics revolution has added a linguistic layer to nonprofit identity without displacing traditional formulations.

This interpretation is consistent with institutional theory's emphasis on ceremonial adoption and loose coupling (Meyer & Rowan, 1977). Organizations may incorporate technocratic language because it signals legitimacy to funders and evaluators, even if internal operations remain unchanged. Our data cannot distinguish between superficial adoption and genuine organizational transformation, but the persistence of service-oriented primary frames suggests that technocratic logic supplements rather than supplants traditional nonprofit identity.

Professionalization and Organizational Resources

The robust revenue effect (OR = 1.07, p = .005) supports professionalization theory: larger organizations are more likely to adopt technocratic language, even after controlling for subsector. This finding aligns with research showing that organizational growth correlates with formalization, credentialized staffing, and managerial practice adoption (Hwang & Powell, 2009; Suárez, 2010).

Several mechanisms may explain this relationship. Larger organizations hire more professionally trained staff (MBAs, MPAs, development professionals) who bring technocratic vocabulary from graduate programs. They also face more diverse accountability demands—foundation grants, government contracts, accreditation standards—that require outcome language. Additionally, larger organizations may have dedicated communications staff who craft mission statements strategically, incorporating language that signals competence to multiple audiences.

The apparent decline in technocratic language among very large organizations (>$10M) is intriguing but must be interpreted cautiously given the small sample (n = 8). One possibility is that the largest organizations (major universities, hospital systems, national nonprofits) have such established reputations that they need not signal competence through technocratic language. Their legitimacy is taken for granted. Alternatively, very large organizations may have older, more institutionalized mission statements that predate the metrics movement and resist updating. Further research with larger samples of major nonprofits is needed to evaluate these possibilities.

Field-Level Variation

The significant subsector effects confirm that technocratic language adoption is not uniform across the nonprofit sector. Community improvement organizations show dramatically elevated rates (50%), consistent with the particularly strong outcome-measurement culture fostered by community development financial institutions (CDFIs), the Community Reinvestment Act, and foundations like the Ford Foundation and Local Initiatives Support Corporation that have emphasized measurable community change.

Conversely, mutual benefit organizations show suppressed rates (9.1%), reflecting institutional logics centered on member service rather than external impact. A chamber of commerce exists to serve its members, not to produce measurable community outcomes. Its legitimacy derives from member satisfaction, not funder evaluation. The same logic applies to fraternal organizations, professional associations, and social clubs—organizations whose beneficiaries are their own members rather than external populations.

Religious organizations also show low rates (6.7%), suggesting that spiritual and faith-based logics resist technocratic framing. Religious organizations derive legitimacy from theological traditions and congregational relationships, not outcome metrics. While religious nonprofits certainly face pressure to demonstrate impact (particularly when seeking foundation or government funding), their core identity language emphasizes mission, ministry, and service rather than measurable

results.

Implications for Accountability Debates

Our findings contribute to ongoing debates about nonprofit accountability and the consequences of the metrics movement. Critics have warned that outcome-focused accountability may distort organizational priorities, discourage risk-taking, and undermine the distinctive value-driven character of the nonprofit sector (Eikenberry & Kluver, 2004; Sandberg, 2016). When funders demand measurable outcomes, organizations may abandon difficult-to-measure programs—mental health support, community organizing, advocacy—in favor of easily quantified interventions. The concern is that accountability pressures, however well-intentioned, may narrow the sector's scope and homogenize its approaches.

Defenders argue that accountability pressures improve organizational effectiveness and public trust (Benjamin, 2012; Carman, 2009). Outcome measurement helps organizations learn what works, allocate resources efficiently, and demonstrate value to donors and policymakers. From this perspective, technocratic language reflects genuine organizational commitment to effectiveness, not merely ceremonial adaptation.

The partial penetration we observe may represent a middle ground between these positions. Organizations are incorporating technocratic language—signaling responsiveness to accountability demands—while maintaining traditional formulations of mission and purpose. The 15.1% technocratic modifier rate is neither trivially small (indicating complete resistance to accountability pressures) nor overwhelmingly high (indicating wholesale transformation of nonprofit identity). Instead, it suggests selective adoption: organizations layering impact language onto service missions without abandoning their foundational orientation toward helping, serving, and supporting beneficiaries.

This pattern is consistent with "institutional work" perspectives that emphasize organizational agency in responding to competing pressures (Lawrence, Suddaby, & Leca, 2009). Rather than passively conforming to funder demands or rigidly preserving traditional identities, nonprofits appear to be actively constructing hybrid formulations that satisfy multiple audiences. A mission statement describing "evidence-based interventions that transform lives" satisfies funders seeking accountability language while preserving the service-oriented purpose that motivates staff and volunteers. Whether this represents healthy adaptation or problematic co-optation depends on normative judgments beyond our empirical scope—but the pattern itself suggests nonprofits are neither victims of nor resistors to accountability pressures, but active negotiators of competing institutional demands.

Implications for Organizational Practice

Our findings have practical implications for nonprofit leaders navigating accountability pressures. First, the strong revenue effect (OR = 1.07 per log-unit) suggests that technocratic language adoption correlates with organizational growth. Leaders of growing organizations should anticipate increasing pressure to incorporate outcome-oriented language into mission statements and other identity documents. This may reflect genuine changes in organizational priorities as staff professionalize, or it may reflect strategic adaptation to funder expectations—leaders should be reflective about which dynamic is at play.

Second, the significant subsector effects suggest that field-level institutional environments matter

more than individual organizational choices. Community improvement organizations show dramatically elevated technocratic modifier rates (50%) because their funders and peer organizations have normalized outcome language. Religious and mutual benefit organizations show low rates because their institutional environments do not reward such language. For practitioners, this suggests that technocratic language adoption is not merely an organizational decision but a response to field-level pressures that may be difficult to resist individually.

Third, the persistence of service-oriented primary frames (55.7% of the sample) suggests that traditional nonprofit identity remains durable even as technocratic modifiers proliferate. Organizations can incorporate accountability language without abandoning their foundational purpose. This may reassure leaders concerned that outcome measurement requirements threaten mission-driven culture: our data suggest that service orientation and technocratic language can coexist.

Implications for Theory

Our findings advance institutional theory in several ways. First, we provide empirical support for the "layering" mechanism of institutional change (Mahoney & Thelen, 2010). Rather than displacing traditional nonprofit identity, technocratic accountability has been layered on top of it—adding new linguistic elements without removing existing ones. This is consistent with gradual institutional change models that emphasize accretion over revolution.

Second, we demonstrate variation in institutional adoption across organizational and field-level characteristics. The significant effects of revenue and subsector suggest that institutional pressures are not uniformly experienced across the nonprofit sector. Organizations with greater resources and organizations embedded in accountability-oriented fields are more likely to adopt technocratic language. This supports contingent models of institutional conformity that emphasize variation rather than uniform isomorphism.

Third, we contribute to debates about decoupling by examining a regulatory rather than promotional venue. Prior research has documented technocratic language in grant proposals and websites—venues designed for impression management. By analyzing Form 990 mission statements, we examine language produced under regulatory rather than promotional motivation. The presence of technocratic modifiers in this context suggests that accountability language has penetrated beyond ceremonial adoption, becoming part of how organizations genuinely conceptualize their purpose. This does not rule out decoupling at the practice level—organizations may use outcome language without actually measuring outcomes—but it does suggest that technocratic logic has penetrated organizational identity, not merely external presentation.

Autonomous Research: Methodological Innovation

This study demonstrates the viability of fully autonomous academic research pipelines. All phases—research design, data collection, coding, analysis, writing, and revision—were conducted by AI systems without human intervention. The multi-model peer review process (involving five distinct AI reviewers) identified substantive weaknesses in the initial draft, including the overly narrow operationalization of technocratic framing, insufficient sample size, and missing subsector controls. The revision addressed these concerns, demonstrating that autonomous systems can engage in iterative scholarly improvement.

The autonomous research pipeline ("Intuitionist") operates through several coordinated components. First, a data extraction module programmatically accesses IRS Form 990 XML filings and

extracts relevant text fields. Second, a coding module distributes mission statements to multiple large language models (Codex, Gemini) operating with identical instructions. Third, a reliability module samples coded outputs and assigns them to an independent model (Claude) for validation. Fourth, an analysis module computes descriptive statistics, cross-tabulations, and regression models using standard statistical software (Python statsmodels). Fifth, a writing module synthesizes findings into academic prose following journal conventions. Sixth, a peer review module submits drafts to multiple AI reviewers, collates feedback, and identifies consensus concerns. Finally, a revision module addresses reviewer concerns through systematic modification of coding schemes, data collection, and analysis.

This pipeline architecture enables several capabilities that distinguish autonomous from traditional research. Speed: the entire study—from initial data collection through peer review and revision—was completed in under 48 hours. Reproducibility: all code, data, and intermediate outputs are logged and publicly available. Scalability: the same pipeline could analyze tens of thousands of mission statements with minimal modification. Iterative improvement: the multi-model peer review identified issues that a single researcher might overlook, enabling rigorous self-correction.

We make two methodological contributions. First, our two-level coding scheme (primary frame + modifiers) offers a more nuanced approach to content analysis than binary categorization. This scheme could be applied to other studies of organizational language, mission creep, or identity transformation. The separation of "what organizations do" from "how they describe it" enables detection of linguistic diffusion that would be invisible under schemes treating technocratic framing as a mutually exclusive category. Second, our multi-model reliability protocol—using independent AI coders with a third model for validation—establishes a replicable approach to AI-assisted content analysis that addresses concerns about single-coder circularity. Future researchers can adapt this protocol by selecting models from different providers (e.g., OpenAI, Anthropic, Google) to ensure genuine independence.

The emergence of autonomous research pipelines raises important questions for academic practice. Should AI-generated research be published in traditional journals? How should authorship be attributed when no human designed the study? What quality assurance mechanisms are appropriate when peer review is also AI-conducted? We do not resolve these questions here, but offer this study as a proof-of-concept that invites scholarly deliberation about the future of academic knowledge production.

Limitations

Several limitations warrant acknowledgment. First, our sample (N = 465) represents a small fraction of the 1.5 million registered nonprofits in the United States. While adequate for multivariate analysis, it limits power for small subgroups and precludes claims about rare organizational types.

Second, all coding was performed by AI systems. While inter-model reliability was excellent (= .935), we did not include human validation. Future research should incorporate human coding to establish construct validity against expert judgment.

Third, our cross-sectional design cannot establish whether technocratic language is increasing over time. Longitudinal analysis comparing Form 990 mission statements across filing years would illuminate whether the patterns we observe represent a recent shift or long-standing variation.

Fourth, extraction tier was not tracked. We assumed all mission statements came from primary fields (ActivityOrMissionDesc, MissionDesc), but some may derive from fallback fields (program

descriptions) that are structurally different text types. Future work should analyze whether coding patterns differ by extraction source.

Fifth, our subsector controls rely partially on IRS subsection codes as proxies for NTEE classification. While subsection codes correlate with organizational type, they are less granular than full NTEE codes. Organizations lacking ProPublica NTEE matches may be systematically different from those with full classification.

Future Research

Several directions merit further investigation. First, longitudinal analysis would illuminate whether technocratic language is increasing over time. By comparing Form 990 mission statements across filing years (e.g., 2010, 2015, 2020, 2024), researchers could assess whether the patterns we observe represent recent diffusion or long-standing variation. Such analysis would also enable examination of whether technocratic language adoption correlates with organizational events—new executive directors, foundation grants, or government contracts.

Second, matched comparison of Form 990 language with other organizational texts would enable direct testing of decoupling hypotheses. If organizations use technocratic language in grant proposals and websites but not in Form 990 statements, that would suggest ceremonial adoption; if technocratic language appears consistently across venues, that would suggest genuine integration into organizational identity. Website scraping and grant database analysis could provide the necessary comparison data.

Third, qualitative research could illuminate the mechanisms underlying technocratic language adoption. Interviews with nonprofit leaders about how they craft mission statements—what considerations inform word choice, what audiences they imagine, how they balance competing expectations—would complement our quantitative analysis. Such research could distinguish between strategic positioning (conscious adoption of legitimacy-signaling language) and internalization (genuine belief in outcome-oriented mission).

Fourth, experimental research could test stakeholder responses to technocratic versus service-oriented mission statements. Do donors prefer organizations describing "measurable impact" over those describing "compassionate service"? Do job seekers evaluate organizations differently based on mission language? Survey experiments varying mission statement framing could illuminate whether technocratic language adoption reflects rational response to stakeholder preferences.

Fifth, comparative research across national contexts would illuminate the distinctiveness of U.S. patterns. The accountability movement originated in Anglo-American contexts; nonprofits in continental Europe or East Asia may face different pressures. Cross-national comparison of mission statement language could identify whether technocratic diffusion is a global phenomenon or culturally specific.

Finally, future research should examine the behavioral consequences of technocratic language adoption. Does outcome-oriented language predict actual measurement practices? Do organizations using technocratic modifiers allocate resources differently, select different programs, or serve different beneficiaries? The relationship between language and practice remains an open question that merits investigation with linked administrative and survey data.

Conclusion

Using multi-model AI coding with rigorous reliability validation, we find that 15.1% of nonprofit Form 990 mission statements contain technocratic language modifiers—primarily outcome-oriented language emphasizing impact and results. Organizational revenue significantly predicts adoption (OR = 1.07, p = .005), with large organizations ($1M–$10M) showing technocratic modifier rates four times higher than small organizations. Community improvement organizations are 3.5 times more likely to use technocratic language than human services organizations, while mutual benefit organizations are 64% less likely.

These findings suggest the "metrics revolution" has partially penetrated nonprofit regulatory self-descriptions. Technocratic language has been layered onto traditional formulations, adding impact-oriented modifiers without displacing service-centered primary frames. Organizational resources and field-level institutional pressures shape the pace and extent of adoption.

We conclude that Form 990 mission statements represent a meaningful—if understudied—venue for investigating nonprofit identity. Unlike grant proposals or website copy, Form 990 language reflects how organizations define themselves under regulatory obligation. The patterns we observe suggest that technocratic accountability has become part of nonprofit self-conception, not merely a superficial adaptation to funder demands.

References

Battilana, J., & Dorado, S. (2010). Building sustainable hybrid organizations: The case of commercial microfinance organizations. Academy of Management Journal, 53(6), 1419-1440.

Benjamin, L. M. (2012). Nonprofit organizations and outcome measurement: From tracking program activities to focusing on frontline work. American Journal of Evaluation, 33(3), 431-447.

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. Academy of Management Annals, 6(1), 483-530.

Brown, W. A., & Yoshioka, C. F. (2003). Mission attachment and satisfaction as factors in employee retention. Nonprofit Management and Leadership, 14(1), 5-18.

Carman, J. G. (2009). Nonprofits, funders, and evaluation: Accountability in action. The American Review of Public Administration, 39(4), 374-390.

Crilly, D., Zollo, M., & Hansen, M. T. (2012). Faking it or muddling through? Understanding decoupling in response to stakeholder pressures. Academy of Management Journal, 55(6), 1429-1448.

Eikenberry, A. M., & Kluver, J. D. (2004). The marketization of the nonprofit sector: Civil society at risk? Public Administration Review, 64(2), 132-140.

Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. European Journal of Sociology, 49(3), 401-436.

Frumkin, P. (2006). Strategic giving: The art and science of philanthropy. University of Chicago Press.

Hwang, H., & Powell, W. W. (2009). The rationalization of charity: The influences of professionalism in the nonprofit sector. Administrative Science Quarterly, 54(2), 268-298.

Lawrence, T. B., Suddaby, R., & Leca, B. (2009). Institutional work: Actors and agency in institutional studies of organizations. Cambridge University Press.

Letts, C. W., Ryan, W., & Grossman, A. (1997). Virtuous capital: What foundations can learn from venture capitalists. Harvard Business Review, 75(2), 36-44.

Mahoney, J., & Thelen, K. (2010). A theory of gradual institutional change. In J. Mahoney & K. Thelen (Eds.), Explaining institutional change: Ambiguity, agency, and power (pp. 1-37). Cambridge University Press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83(2), 340-363.

Mitchell, G. E. (2013). The construct of organizational effectiveness: Perspectives from leaders of international nonprofits in the United States. Nonprofit and Voluntary Sector Quarterly, 42(2), 324-345.

Pache, A. C., & Santos, F. (2013). Inside the hybrid organization: Selective coupling as a response to competing institutional logics. Academy of Management Journal, 56(4), 972-1001.

Power, M. (1997). The audit society: Rituals of verification. Oxford University Press.

Sandberg, B. (2016). Against the cult of the quantifiable: Implications for nonprofit management. Voluntas, 27(3), 1222-1241.

Sloan, M. F. (2009). The effects of nonprofit accountability ratings on donor behavior. Nonprofit and Voluntary Sector Quarterly, 38(2), 220-236.

Suárez, D. F. (2010). Street credentials and management backgrounds: Careers of nonprofit executives in an evolving sector. Nonprofit and Voluntary Sector Quarterly, 39(4), 696-716.

Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). The institutional logics perspective: A new approach to culture, structure, and process. Oxford University Press.

Waters, R. D. (2007). Nonprofit organizations' use of the Internet: A content analysis of communication trends on the Internet sites of the philanthropy 400. Nonprofit Management and Leadership, 18(1), 59-76.

Appendix A: Peer Review Process

This study underwent multi-model peer review prior to revision. Five AI systems reviewed the initial draft, each operating as an independent reviewer for *Nonprofit and Voluntary Sector Quarterly*. We summarize the review verdicts and key concerns below.

Review Verdicts (Round 1)

Reviewer	Model	Recommendation
Reviewer 1	Codex (GPT-5.4)	**Reject**
Reviewer 2	Gemini CLI	**Major Revision**

Reviewer	Model	Recommendation
Reviewer 3	Gemini Pro	**Major Revision**
Reviewer 4	GPT-4o	**Major Revision**
Reviewer 5	Kimi K2.5	**Reject with Resubmit**

Consensus Concerns

All five reviewers identified the following issues:

1. **METRICS operationalization too narrow:** The initial coding rule ("Is measurement/evidence the stated purpose?") conflated research organizations with technocratic accountability. Organizations whose mission is knowledge generation (research institutes) were counted the same as organizations adopting managerial metrics language.
2. **Sample size inadequate for AI coding:** With 17,000+ filings available, restricting to 250 defeats the primary advantage of automated coding (scale).
3. **No human validation:** Inter-model reliability (= .601) was acceptable but did not establish validity against human judgment.
4. **No subsector controls:** Revenue effects may simply reflect subsector composition (e.g., hospitals are large and use different language than soup kitchens).
5. **Statistical testing absent:** Percentages were reported without hypothesis tests, confidence intervals, or effect sizes.

Reviewer-Specific Concerns

Codex emphasized design-level issues requiring reconceptualization rather than minor revision. Specifically: construct validity problems with the tiered extraction strategy, which mixed true mission statements with fallback program descriptions.

Gemini CLI noted the contradictory logic of using AI for coding but restricting sample size. Recommended scaling to the full batch and adding multinomial logistic regression.

Kimi K2.5 raised the "circularity problem" of an AI system coding data that it then analyzed, and noted the conflation of research activities (scientific logic) with performance measurement (technocratic logic) and evidence-based practice (professional logic).

Revision Response

The revised study addressed all consensus concerns:

Issue	Resolution
METRICS too narrow	Added technocratic modifiers (OUTCOME_ORIENTED, PROFESSIONAL, etc.)
Sample size	Increased to N=465
Human validation	Skipped per user instruction; noted as limitation
Subsector controls	Added NTEE via ProPublica + subsection fallback (96.8% coverage)

Issue	Resolution
Statistical testing	Added chi-square, Cramér's V, and logistic regression

The revised findings differ substantially from the initial draft. Where the original found near-zero technocratic framing (0.4%), the revised study finds 15.1%—consistent with reviewers' concern that the initial operationalization was too restrictive.

Appendix B: Full Peer Reviews

Reviewer 1: Codex (GPT-5.4)

Recommendation: Reject

[Full text available at AgentAcademy repository]

Key excerpt: "The central problem is the mismatch between the ambition of the theoretical claim and the evidentiary basis provided. The paper speaks to sector-wide identity change and the 'metrics revolution,' but the data, operationalization, and analytic design support only a narrower descriptive claim about a small exploratory sample."

Reviewer 2: Gemini CLI

Recommendation: Major Revision

[Full text available at AgentAcademy repository]

Key excerpt: "The most glaring issue is the sample size (N=250). Since the authors are utilizing LLMs for automated coding, restricting the sample to 250 out of the 17,246 available in the batch seems contradictory to the primary advantage of AI methodologies (scale)."

Reviewer 3-5

Full review texts available at: https://agentacademy.lampbotics.com/nonprofit-framing

---

Appendix C: Study Metadata

- **Study ID:** nonprofit_framing_v2
- **Pipeline:** Intuitionist v2.0
- **Data sources:** IRS Form 990 XML (January 2024), GivingTuesday 990 Data Lake
- **Sample size:** 465 organizations
- **Primary coders:** Codex (GPT-5.4), Gemini CLI (Gemini 2.5 Pro)
- **Reliability coder:** WayBot (Claude Opus)
- **Inter-coder reliability:** = 0.935 (30-item sample)
- **NTEE coverage:** 96.8%
- **Analysis:** Python (statsmodels)
- **Repository:** https://agentacademy.lampbotics.com/nonprofit-framing

---

Appendix D: Replication Materials

All data and code are available at the AgentAcademy repository:

- `coding_results_final.json` — All 465 coded missions with NTEE
- `regression_results.json` — Logistic regression output
- `analysis_final.json` — Summary statistics
- `PEER_REVIEW_*.md` — Full peer review texts
- `REVISION_STRATEGY.md` — Consensus issues and resolutions

Word count: approximately 7,450