KU LEUVEN

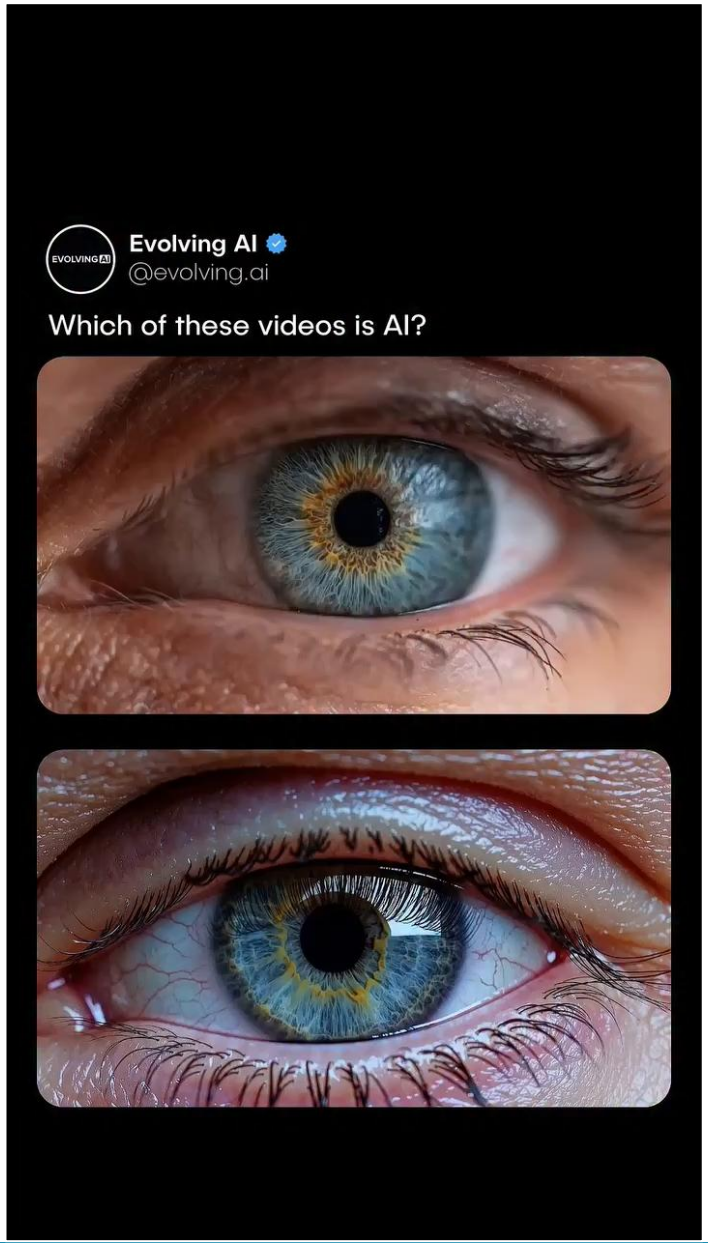
From Choice to Mandate: Artificial Intelligence Disclosure as a Pseudo-Certification Scheme

Marcos Balmaceda, KU Leuven
Prof. Dr. Michael Mödl, KU Leuven

85th Annual Meeting of the Academy of Management
17th July 2025

fwo Research Foundation Flanders
Opening new horizons

What is real?



Home Reporting Research Fellowships

Home > News

AI-generated slop is quietly conquering the internet. Is it a threat to journalism or a problem that will fix itself?

Three experts discuss the rise of low-quality content and its implications for the profession, the news industry and the public sphere

Artificial Intelligence

The Dark Side of AI-Generated Content

AI-generated content carries risks like misinformation, bias, and security threats. Responsible AI practices are crucial

Pros and cons of AI-generated content

Artificial intelligence is growing rapidly in the tech sector and assisting with various tasks such as creating images and generating content. However, it cannot do everything.

AI, Ethics & Misinformation: The Truth About AI-Generated Content

AI-generated content is everywhere - from blogs to news, but it raises serious ethical concerns. Can you trust AI-created information? Here's a quick breakdown:

- **Trust Issues**: AI content often mimics human writing but lacks real expertise and can spread false or outdated information.
- **Transparency**: Readers demand disclosure when AI is used, but many platforms fail to make this clear.
- **Copyright Challenges**: U.S. law protects only human-created work, leaving AI-generated content in a legal gray area.
- **Bias Problems**: AI systems often reflect biases in gender, race, and age, impacting fairness in representation.
- **Misinformation Risks**: AI is increasingly used to spread false information, eroding trust in media and education.

Source: https://rewriterpro.ai/blog/ai-ethics-and-misinformation

Motivation

- Business leaders and experts show overwhelming support for responsible use of artificial intelligence, with 84% favoring *mandatory AI transparency policies* (Renieris et al., 2024).

- Major internet platforms have implemented AI disclosure policies:
  - **Meta:** Mandates labeling AI-generated images, videos, and audio content
  - **YouTube:** Requires creators to disclose realistic altered or synthetic content
  - **Amazon:** Requires disclosure of AI-generated book content and reviews
  - **Kickstarter:** Requires disclosure of AI use in creative projects with human moderation team verification

- Potential views of AI:
  - *"Algorithm aversion"* suggests general public skepticism towards AI (Dietvorst et al., 2015; Glikson & Williams Woolley, 2020).
  - AI as *“innovation marker”* (Taeuscher & Rothe, 2021).

- Question: **Effects of mandatory AI disclosure on platform success?**

Background literature

Information disclosure

- Disclosure literature shapes information sharing incentives:
  - **Voluntary:** firms control what they reveal (Grossman, 1981; Verrecchia, 1983)
    - Often strategic, selective, and can be insufficient, especially with complex technologies (Verrecchia, 1983; Foss & Weber, 2016).
    - Leads to persistent information asymmetry; positive information emphasized.
  - **Mandatory:** when information is complex, mandatory disclosure is more effective (Fishman & Hagerty, 2003)
    - Even coarse disclosure can increase market information despite simplicity by encouraging broader participation from creators who might otherwise opt out of detailed disclosure systems (Asseyer & Weksler, 2024; Harbaugh & Rasmusen, 2018).
- Trade-off: Potential skepticism towards disclosed information vs. increased transparency (trust)

Theory

Certification Theory

- **Traditional Certification:** Independent third-party validation reduces information asymmetry and increases credibility, i.e., signals quality (Booth & Smith, 1986; Lizzeri, 1999; Stahl & Strausz, 2017).
- **Pseudo-Certification:**
  - Non-traditional entities (e.g., platforms) can act as pseudo-certifiers, i.e., assume certification-like roles without formal independence (Gross et al., 2005; Mogyoros, 2021, 2023).
  - (Crowdfunding) platforms, as intermediaries, can inherently build credibility and verify information (Drover et al., 2017).
- **In our context:**
  - Digital platforms (e.g., **Kickstarter**, Amazon, Meta, YouTube) increasingly mandate AI use disclosure.
  - Kickstarter's moderation & enforcement of AI disclosure aligns with this concept (see examples)

Hypotheses Development (1)

Pre-Policy: Informational effect of AI disclosure under voluntary conditions

- AI disclosure acts as an "innovation marker" in a context of high information asymmetry (Agrawal et al., 2014; Taeuscher & Rothe, 2021).
- Appeals to backers seeking novelty, potentially increasing funding (Chan & Parhankangas, 2017).
- However, credibility is limited by strategic, selective disclosure and the potential for unverified claims ("cheap talk") (Albano & Lizzeri, 2001; Verrecchia, 1983; Foss & Weber, 2016).

Hypothesis 1a: *Under voluntary disclosure, projects that disclose AI use increase their probability of funding success relative to non-disclosing projects.*

Hypothesis 1b: *Under voluntary disclosure, projects that disclose AI use receive a higher average funding amount compared to non-disclosing projects.*

Hypotheses Development (2)

Post-Policy: Amplification of informational effect (pseudo-certification)

- Mandatory disclosure policy transforms AI disclosure from a simple claim to a platform-warranted cue of authenticity/innovation (cf. Drover et al., 2017; Chan & Parhankangas, 2017).
- Verified (even if coarse) disclosure enhances credibility and informs backers (Fishman & Hagerty, 2003; Harbaugh & Rasmusen, 2018; Asseyer & Weksler, 2024).
- Moves beyond "algorithm aversion" (Dietvorst et al., 2015) by providing a trusted cue; platform enforcement discourages misleading claims (cf. Almazan et al., 2008; Chakraborty & Harbaugh, 2007).
- The verified platform-enforced transparency mechanism (pseudo-certification) makes AI disclosure a more potent and credible asset (Gross et al., 2005; Mogyoros, 2021, 2023).

Hypothesis H2a: *A mandatory AI disclosure policy increases funding success probability for AI-disclosing projects compared to the voluntary regime.*

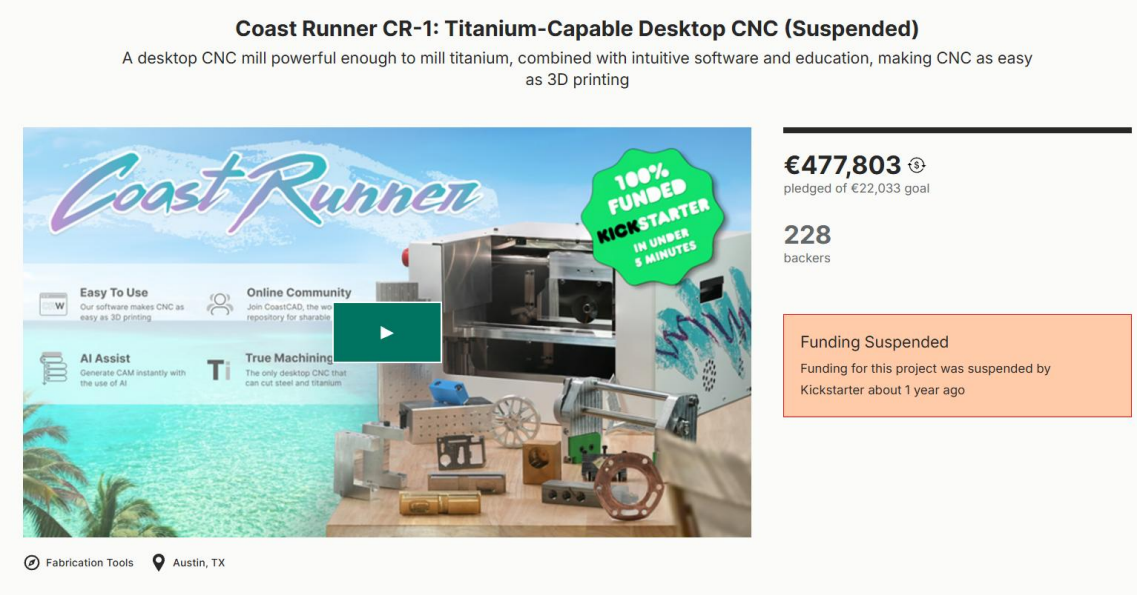
Hypothesis H2b: *A mandatory AI disclosure policy widens the funding amount gap between AI-disclosing and non-disclosing projects.*

Empirical Setting

- **Kickstarter** is the largest (reward-based) **crowdfunding platform:**
  - Crowdfunding = open call by creators (e.g., entrepreneurs) through an online platform for the provision of financial resources by many small contributors (Belleflamme et al., 2014; Mollick 2014).
  - $ 8.66 billion for 276,000+ projects since 2009
  - Time Magazine's "100 Most Influential Companies of 2023”
- On August 29, 2023, Kickstarter implemented a **mandatory AI disclosure policy:**
  - Requires creators to disclose and specify how AI is used.
  - Creators must detail which elements of their project are wholly original (human) work versus AI-generated; and projects developing AI technology must disclose information about databases and data sources used, including how these sources handle consent and credit.
  - These AI disclosures become public information within a dedicated "Use of AI" section on the project page.
- Active and strict **policy enforcement:**
  - A human moderation team reviews AI disclosures during project submission, using targeted questions.
  - Platform moderation continues post-launch specifically to identify non-compliant submissions.
  - Projects may face suspension, and misrepresentation of AI use can lead to restrictions on future platform access.

Strict disclosure policy enforcement

Example: Coast Runner (February 2024)
Terminated after raising $542,123 in a few hours (goal: $24,999) for non-compliance with Kickstarter's AI policy



→ demonstrates Kickstarter's strategic prioritization of AI transparency over immediate project funding success and reinforcing the policy's impact.

Current federal antitrust lawsuit from Coast Runner against Kickstarter in Texas

Case 7:24-cv-00326 Document 1 Filed 12/09/24 Page 1 of 38

UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF TEXAS
MIDLAND DIVISION

COAST RUNNER, INC.,

Plaintiff,

v.

KICKSTARTER, PBC; INDIEGOGO, INC.; LAUNCHBOOM, INC.; JOHN DOES #1 AND #2,

Defendants.

Civil Action No. 7:24-cv-326

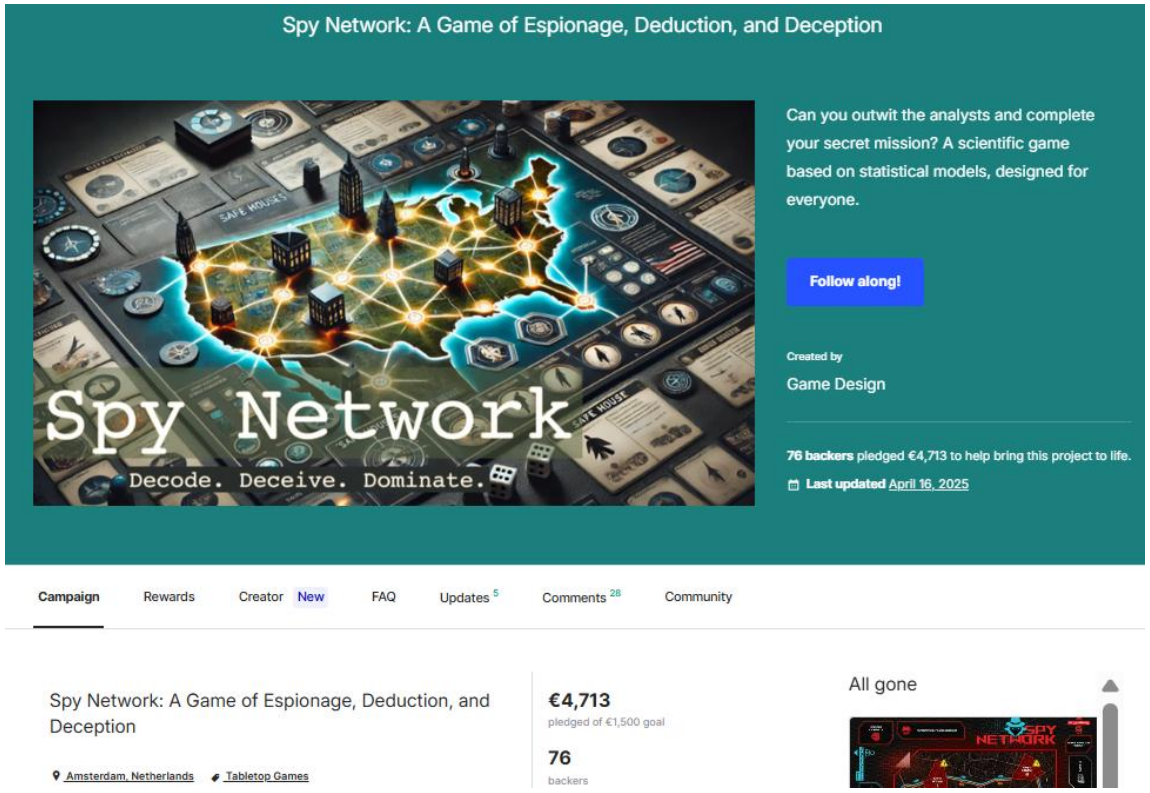
JURY TRIAL DEMANDED

PLAINTIFF'S ORIGINAL COMPLAINT

1. Turning an idea into a commercial success is a challenge. At some point, every entrepreneur must make difficult choices about how to raise money to grow his business. Some

Active monitoring

Spy Network campaign's initial suspension due to inadequate AI disclosure



Enforcement timeline: Successful appeal process following AI disclosure policy compliance

December 2024:
Kickstarter Launch – Launching the campaign and driving excitement for our backers.

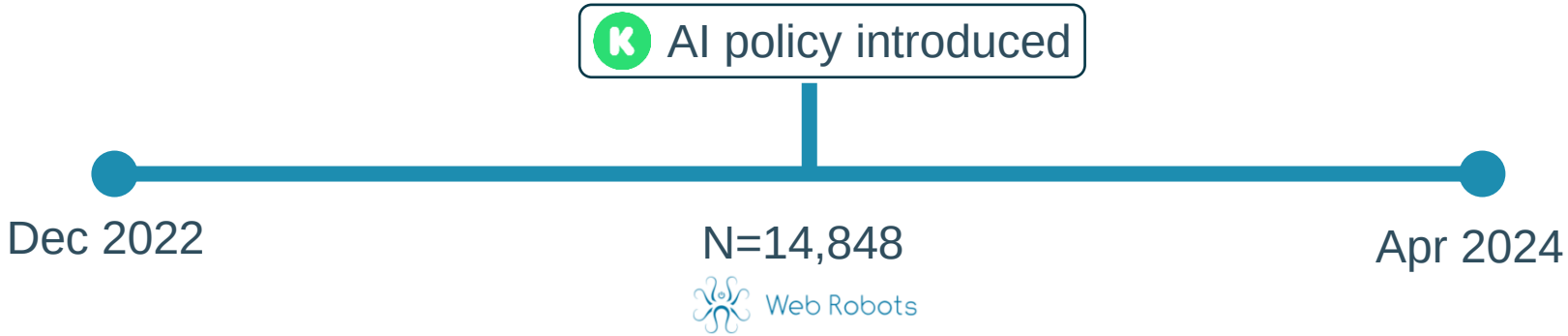
January 2025:
Campaign Suspension – On January 15, our campaign was suspended due to a misunderstanding related to AI disclosure, as that was our first Kickstarter project. After resolving the issue with Kickstarter, they allowed us to re-launch.
Campaign Re-launch – Re-launching the campaign on January 22.

→ Demonstrates active monitoring by Kickstarter's human moderation team

- Project initially passed pre-launch review but was later flagged
- Shows that moderation is ongoing throughout a campaign's lifecycle
- Proves that Kickstarter actively enforces its AI policy, not just at submission

Data & Variables

- **Data Source:**



- **Key Variables:**
  - **AI disclosure:** Identified via text mining (extended Rezazadegan et al., 2024 lexicon).
  - **Generative AI:** Text-based classification (Feyzollahi & Rafizadeh, 2025).
  - **Outcomes:** Funding Success (binary), Pledged Amount (log-transformed).
  - **Moderators** (for heterogeneity):
    - Creator attributes (gender/ethnicity inferred),
    - Risky Goals (90th percentile funding goal),
    - Innovative Creators (tech-intensive sectors).

Empirical Strategy

Quasi-Natural Experiment: Kickstarter's AI Disclosure Policy (August 2023)

Data & Analysis Periods:

- **Core Regressions:** ±18 weeks around policy implementation (Pre-policy: ~3,870; Post-policy: ~2,940), Dec22-Apr24.
- **CDiD:** Across 3 focused time spans (Mar23-Dec23 (immediate), Mar23-Mar24 (medium), Dec22-Apr24 (full)).

Estimation Strategy:

1. Core Regressions (Pre vs. Post): Logit models (success) and OLS (log-transformed pledged amounts)
2. Conditional Difference-in-Differences (CDiD) – Logit and LPM (success)
   - Treatment: AI-disclosing projects after policy implementation.

- Entropy balancing to address selection bias (results robust without reweighting)
- Controls (project features, creator demographics, AI market trends), fixed effects (subcategory, time), robust standard errors.
- Parallel trends verified across all CDiD models and confirmed via stacked event study.

Descriptive Statistics

Panel B: After AI Policy Implementation (+18 weeks)

	N	Mean	Median	SD	P10	P90
Continuous variables:						
Pledged	5,673	24,716	3,386	142,679	65	36,790
Goal	5,673	38,543	3,291	1,369,476	368	25,000
Backers	5,673	205	45	810	3	383
Blurb Length	5,673	16	16	6	7	23
	Yes (%)	No (%)				
Discrete variables:						
US located	56	44				
Female	19.02	80.98				
People of Color (PoC)	73.54	26.46				
AI Disclosed	2.33	97.67				
Staff Pick	20.68	79.32				
Successful	73.17	26.83				

Panel C: Before AI Policy Implementation (-18 weeks)

	N	Mean	Median	SD	P10	P90
Continuous variables:						
Pledged	7,443	24,175	3,511	128,069	82	37,377
Goal	7,443	28,824	3,500	737,768	310	25,000
Backers	7,443	223	54	757	3	443
Blurb Length	7,443	16	16	6	8	23
	Yes (%)	No (%)				
Discrete variables:						
US located	57.33	42.67				
Female	18.16	81.84				
People of Color (PoC)	73.32	26.68				
AI Disclosed	1.63	98.37				
Staff Pick	18.77	81.23				
Successful	75.18	24.82				

Main Results

TABLE 3: Disclosure and project performance

Regression results (reweighted)				
	Before AI Policy (-18 weeks)		After AI Policy (+18 weeks)	
Dependent variables	Logit (1)	OLS (2)	Logit (1)	OLS (2)
ln(pledged+1)		X		X
Success	X		X	
Independent Variables				
AI Disclosed	0.077	0.634*	2.032***	0.954***
	(0.27)	(2.53)	(5.45)	(3.70)
ln(Goal+1)	-0.988***	-0.067	-0.697***	0.063
	(-9.10)	(-1.09)	(-8.08)	-0.85
ln(Duration+1)	0.551***	0.571***	0.575***	0.515***
	(6.58)	(6.86)	(5.39)	(5.26)
ln(Campaign Days+1)	-0.07	-0.212	-1.102***	-0.011
	(-0.22)	(-0.89)	(-3.46)	(-0.03)
Staff Pick	3.628***	3.034***	3.100***	2.609***
	(13.11)	(13.25)	(9.90)	(10.74)
US located	0.106	0.273	-0.012	0.046
	(0.38)	(1.05)	(-0.04)	(0.18)
Demographic controls	Yes	Yes	Yes	Yes
Trends and characteristics controls	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	3,869	3,869	2,939	2,939
Pseudo-R^2	0.558		0.319	
Adjusted-R^2		0.59		0.508

TABLE 5: Conditional difference-in-differences

CDiD (reweighted)	Logit	LPM	Logit	LPM	Logit	LPM
Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Difference-in-differences estimates						
DiD estimate (AI x Post)	0.226***	0.173***	0.143***	0.105***	0.125***	0.094***
	(0.043)	(0.036)	(0.038)	(0.032)	(0.033)	(0.035)
Panel B: Treatment effects by period						
AI projects (pre-period)	-0.008	0.006	-0.071**	-0.055	-0.056**	-0.047
	(0.036)	(0.033)	(0.036)	(0.028)	(0.031)	(0.024)
AI projects (post-period)	0.217***	0.179***	0.072***	0.050*	0.069***	0.047*
	(0.037)	(0.025)	(0.028)	(0.023)	(0.025)	(0.029)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Trends and characteristics controls	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,561	7,561	9,948	9,948	13,038	13,038
Pseudo-R^2	0.515		0.462		0.458	
Adjusted-R^2		0.512		0.481		0.477

Heterogeneous Effects

TABLE 4 - Heterogeneous effects

Regression results (reweighted)	Before AI Policy (-18 weeks)						After AI Policy (+18 weeks)					
	GenAI		Risky goals		Innovative creators		GenAI		Risky goals		Innovative creators	
	Logit (1)	OLS (2)	Logit (5)	OLS (6)	Logit (7)	OLS (8)	Logit (1)	OLS (2)	Logit (5)	OLS (6)	Logit (7)	OLS (8)
AI Disc.	0.123	0.609*	0.109	0.527*	0.134	0.880*	2.045***	0.970***	2.067***	0.968***	1.690*	0.371
	(0.42)	(2.42)	(0.37)	(2.05)	(0.20)	(2.49)	(5.48)	(3.74)	(5.47)	(3.63)	(2.54)	(1.00)
GenAI	0.037	-0.823*					-0.621	0.362				
	(0.07)	(-2.24)					(-0.87)	(0.70)				
AI Disc. x GenAI	-2.963*	0.978					0	0				
	(-2.55)	(1.18)					(.)	(.)				
Risky Goal			-0.425	-0.827*					0.676	-0.595		
			(-0.83)	(-2.27)					(1.31)	(-1.45)		
AI Disc. x Risky Goal			0	0					0.246	-0.348		
			(.)	(.)					(0.19)	(-0.34)		
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trends and characteristics controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,869	3,869	3,856	3,856	538	538	2,938	2,938	2,939	2,939	457	457
Pseudo-R²	0.565		0.536		0.729		0.510		0.511		0.632	
Adjusted-R²		0.591		0.588		0.473		0.508		0.511		0.513

1. **GenAI Projects - Pre: AI Disc. X GenAI (β = -2.963*) and GenAI overall (β = -0.823*) → negative penalty, Post: AI Disc. X GenAI (β = 0) and GenAI (β = 0.362) → no penalty**
2. **Risky Goals - Pre: Risky (β = -0.827*) → negative funding premium overall, Post: Risky → not significant**
3. **Innovative Creators – Pre: AI Disc. (β = 0.880*) → funding boost, Post: AI Disc. (β = 1.690*) → success boost**

- **Gender and PoC Effects - Both periods: AI Disc. x Female and AI Disc. x PoC (not significant)**

Contribution and Implications

Summary:

- **Key Finding:** Mandatory AI disclosure policy increased funding success by 17 percentage points for AI-disclosing projects immediately after implementation (9 points over longer periods).
- **Heterogeneous Effects:** Policy eliminated previous disadvantages for GenAI projects (β = -0.823* → 0.362) and especially for AI projects using GenAI (AI Disc. X GenAI: β = -2.963* → 0), reduced penalties for high-risk goals (β = -0.827* → -0.595), and shifted benefits for innovative creators from funding amounts to success rates (β = 0.880* → 1.690*).
- **Temporal Dynamics:** Effects strongest immediately post-policy, moderating over time as markets adapt to new information environment.

Theoretic contribution

- Platform policies shape entrepreneurial resource acquisition for emerging technologies (Nambisan et al., 2019; Yoo et al., 2012).
- Mandatory disclosure functions as pseudo-certification in contexts with high information asymmetry (e.g. Fishman & Hagerty, 2003; Gross et al., 2005; Mogyoros, 2021, 2023).
- Technological disclosure affects backer behavior and funding outcomes (e.g. Ahlers et al., 2015; Chan & Parhankangas, 2017).
- Empirical evidence on how disclosure policies shape market responses to AI (Acemoglu & Lensman, 2024).

Managerial Implications

- **Platforms:** Enforced transparency policies can improve information quality & foster tech adoption.
- **Entrepreneurs:** Mandated, verified disclosure of complex tech (AI) becomes a strategic asset that enhances funding prospects.
- **Strategic Value:** Policy creates credible verification mechanism for technological claims without requiring prohibitive verification costs.