Discussion of “AI Washing” by Boyuan Li

2025 Global AI Finance Research Conference

Discussant: Henry Zhang

December 15, 2025

Big Picture

The paper defines firm-quarter "AI talk" from earnings calls' transcripts and "AI walk" from employees' resumes similarly to Babina et al (JFE 2024) but with a different dataset (Revelio vs Cognism). Sample period 2016Q1 to 2024Q2.

- Talk = what firms want investors to believe (forward-looking)
- Walk = firms' true actions (backward-looking)
- Washing = firms talk but do not follow through with AI workers. Contribution to disclosure and innovation with analogy to greenwashing.
- Investors reward walk in long-run, although talk has short-run returns

Overall comment: intuitive story documenting a novel pattern on a timely and important topic, but focus on improving measurement of "walk"

Key Evidence

- Talk surges after 2019. Walk grows slower.
- AI walk predicts future AI innovations. AI talk does not, conditional on walk.
  - Previous quarter walk predicts 17% more AI patents and 25% higher AI patent value, but talk is not predictive.
- Talk spikes during seasoned equity offering quarters, with higher apparent return to hype
  - 7 times higher return (1.6 pct pts cumulative abnormal return) to talk vs typical. Walk stays flat.

Comment 1: Resume Revision Bias

Measurement issue with "walk" is that resume snapshots from 2024 capture workers' 2024 *rewritten* historical job descriptions, often with AI.

With rapid takeup of ChatGPT since 2022, workers standardized AI phrasing, especially for more recent jobs that recruiters read first:

- Machine learning, coding tools $\rightarrow$ generative AI, LLMs, copilot

Concern: Walk in later years mechanically inflated. Potential solutions:

- Obtain pre-2022 snapshots, e.g. Crunchbase / Zoominfo / Snowflake + Wayback. Compare to phrasing in 2024 snapshot for same employees.
- Freeze AI keyword dictionary to pre-2022 vocabulary
- Restrict analysis to resume descriptions were *written* before ChatGPT, if timestamps exist

I think your year-by-year separate training on conference calls does not address the resume revision concern, and the paper's robustness focuses on missingness and job titles

Comment 2: Manufacturing and Industry Heterogeneity

Manufacturing has over half of AI washing firms, but manufacturing firms probably have low resume coverage

Hypothetical: A firm hires experienced robotics engineers to improve the automation of tuning machines using AI embedded systems. Less likely to use LinkedIn vs software programmers and younger employees. Work on factory floors, not only in cubicles, with lower rate of job searching.

Concern: "Walk = 0" might mean either (a) the firm didn't hire AI workers, or (b) it did, but we can't see them in the data. Potential solutions:

- Show industry-specific resume coverage, and control for employee observables (age, degree, years of experience)
- Re-estimate washing using only high-coverage industries (e.g. professional services)

Main Comment 3: Walk Persistence and Reputation

$$\text{AI Walk}_{i,t} = \alpha + \sum_{h=1}^{8} \beta_h \text{AI Talk}_{i,t-h} + \gamma X_{i,t-1} + \alpha_{\text{firm}} + \lambda_{\text{industry-quarter}} + \varepsilon_{i,t}$$

Table 4 shows that R^2 with firm FE exceeds 0.93, meaning that within-firm walk barely moves at the quarterly frequency, but R^2 without firm FE is below 0.3 (predicting on log talk)

Intuitive mechanism: **reputation as an asset**. Conditional on past, firms expecting weaker future talk more, drawing down on reputation. Damage control as much as opportunism.

- Does talk spike *after* negative fundamentals (lower profit margins, market share loss)?
  - You have firm characteristic controls and industry FE, but do these capture firm-specific forecasts?
- Would a first-difference separate adjustment lags from reputation-driven signal inflation?

Additional Small Concerns

Innovation validation: Patents truncated mid-2021, yet washing peaks afterwards 2024. Consider alternative proxies (GitHub activity, analyst coverage) to validate walk post-ChatGPT.

Cumulative abnormal return: Table 7 shows talk predicts short-run CARs but lacks controls for general call sentiment (tone, length, analyst engagement). Is sentiment correlated with AI talk?

Institutional investor weights: Fund count is ok, but AUM-weighted results would be more convincing than raw counts.

Side question: In Table 4, is it reasonable for me to infer that firms' AI headcount is highly persistent = hard to scale up quickly? As a benchmark, why not run the opposite regression, talk on lagged walk?

Conclusion

This paper delivers a clear and timely result: firms' AI rhetoric and real AI investment have diverged, and markets do not fully price that distinction in the short run

My comments focus on ways to tighten your story:

- separating genuine changes in firms' AI investment from workers' resume-revision effects,
- understanding how sample coverage differs across industries, especially manufacturing,
- and sharpening the interpretation of the talk–walk null result given highly persistent AI hiring and reputational incentives.

Ideally show the mechanisms for *why* talk and walk diverge, not merely *whether* they diverge