RESEARCH PRESENTATION

Measuring CEO Meanness Using NLP

How language patterns reveal a neglected executive trait

Stephen J. Smulowitz · Wake Forest University School of Business

INTRODUCTION

Who I Am

Background

- Wake Forest University School of Business
- Former attorney: Skadden Arps, SEC lawyer
- Research: corporate governance and executive behavior

“The High Cost of Cheap Talk” — *Journal of Management*

With Mike Pfarrer, our organizer. We built a dictionary of disingenuous ethical language from restating firms’ Codes of Ethics; cheap talk in 10-Ks predicts declining corporate social performance.

CEO Triarchic Psychopathy — *current project*

With Michael Holmes (Florida State) & Philip Howard (Wake Forest). Meanness is one of the three triarchic facets, alongside boldness and disinhibition.

Language reveals what managers are actually doing.

The Idea in Three Lines

1. Examine CEO meanness — and how can researchers can study this personality traits in this population?

2. Begin with a small set of CEOs for whom direct assessments of meanness are available, and use these assessments to guide the analysis of CEO language in analyst calls.

3. Language patterns reflect persistent differences in how CEOs engage with others — NLP applied to an executive trait that has received limited attention.

WHY MEANNESS?

We’ve all had a mean boss.

It shapes careers, health, and whether people stay or quit.

Now put that person at the top of a Fortune 500 firm. **If a mean middle manager can wreck a department, what does a mean CEO do to a company?**

Yet narcissism dominates the executive literature. Meanness — the facet most tied to how executives actually treat people — is nearly untouched.

THE CONSTRUCT

Meanness in the Triarchic Model

Boldness

Social dominance, fearlessness, confidence under pressure.

Disinhibition

Impulsivity, poor restraint, weak behavioral control.

Meanness

Callousness, cruelty, disdain for others, excitement-seeking through domination.

Core claim: Meanness leaks into language. How a CEO treats analysts on a call reflects persistent differences in how they engage with people generally.

THE MEASUREMENT PROBLEM

Direct Scores Are Scarce; Language Is Abundant

Direct trait assessments exist for a small anchor sample, while earnings-call speech exists for a much larger executive population.

The NLP task: learn the language pattern associated with assessed traits, then generate comparable CEO scores from text.

Why earnings calls?

- Repeated observations of the same executive
- CEO-attributed speech across prepared remarks and Q&A
- A common setting across firms and time

WAYS TO TEST — APPROACHES 1 & 2

Two Classification Paths

Human-anchored

Researchers read the anchor CEOs' transcripts, sort them into high- and low-meanness samples, and train an algorithm to distinguish the two.

+ Interpretable ground truth

– Small N; coder subjectivity

LLM-anchored

An LLM does the reading and sorting into two samples; the classifier trains on its labels.

+ Scalable and consistent

– LLM judgment must first be validated against the direct assessments

WAYS TO TEST — Our Approach

From Anchor CEOs to the Full Population

1. Anchor set

A small set of CEOs for whom direct assessments of meanness are available.

2. Training corpus

Analyst call Q&A — spontaneous, interpersonal, hard to script (Harrison et al., 2019).

3. Score at scale

Learn what distinguishes high- vs. low-meanness language, then score the full CEO population.

Q&A over prepared remarks: analysts' questions are direct and not easily anticipated — the language is more likely the CEO's own.

Open-Language ML: Our Current Approach

Follows **Harrison, Thurgood, Boivie & Pfarrer (2019, SMJ)**, who built a validated linguistic measure of CEO Big Five traits from earnings calls.

Vectorize

Word2Vec on the full corpus of CEO earnings-call language

Train

Gradient boosting predicts directly assessed trait scores for anchor CEOs

Score

Apply the best model to thousands of CEOs

.62–.67

Correlations between predicted and observed trait scores in Harrison et al. — far outperforming closed-language dictionaries. We apply the same pipeline to meanness.

THE EMPIRICAL BASE

Small Labeled Sample, Massive Language Corpus

Independent CEO assessments supply outcomes; executive speech supplies domain-specific language.

210

CEOs with four complete trait scores

91,140

300-word modeling passages

26.9M

words from the labeled CEOs

1.58B

tokens for word2vec training

The labeled corpus connects language to outcomes. The broader unlabeled corpus teaches word2vec how executives use language.

OPEN-LANGUAGE PIPELINE

From CEO Speech to a Trait Score

Four steps turn text into a scalable measure while retaining independent assessments as the anchor.

1

CEO speech 300-word passages

2

Average the word2vec vectors

3

Fit separate gradient-boosting models

4

Average passage predictions by CEO

The model predicts each passage. Those passage predictions are averaged to produce a CEO-level score for each trait.

OUR RESULTS — CONVERGENT VALIDITY

Predicted vs. Observed CEO Scores

Correlations across 210 CEO-level observed–predicted pairs, after averaging predictions from each CEO's held-out passages.

.839	.841	.832	.821
Psychopathy composite	Boldness	Disinhibition	Meanness

Mean across the four outcomes: .833. Harrison et al. report .62 to .67 for the Big Five. Both use within-CEO text validation — but different constructs and different labeled samples, so not a like-for-like comparison.

The Model Recovers the Facet Structure

	1	2	3	4
1. Psychopathy composite	—	0.15	0.41	0.49
2. Boldness	0.20	—	0.05	0.12
3. Disinhibition	0.51	0.20	—	**0.59**
4. Meanness	0.55	0.06	**0.62**	—

Below the diagonal: correlations among predicted CEO scores. Above the diagonal: correlations among the observed scores. Same 210 CEOs.

- **Boldness stands apart**

  Smallest correlations with the other dimensions in both matrices.

- **Disinhibition ↔ Meanness**

  The strongest pair, and the model reproduces it closely: .62 predicted against .59 observed.

- **Elsewhere**

  Predicted and observed correlations differ by .05 to .15 in absolute value.

.997 split-half reliability across all four outcomes

WHAT WE'RE FINDING

The Bag of Words Carries the Signal

This approach sits in the "bag of words" stream of NLP. The model learns from word and phrase features — not from meaning in context.

We've experimented with adding more context to the features — **with minimal improvement over the bag-of-words baseline.**

That is an interesting result in itself. Two readings:

- Word-level signal carries most of the trait information — at least for this construct.
- Or: our context methods aren't yet capturing what matters.

APPLICATION — M&A

CEO Meanness Is Linked to Lower M&A CARs

Why M&A is revealing

Integration is where acquisition value is realized: trust, coordination, and knowledge sharing determine whether synergies materialize.

Meanness may make that process harder—especially when integration demands are high.

MECHANISM TESTS — PLANNED + SEEKING FEEDBACK

- Same- vs cross-industry acquisitions
- SIC/NAICS or product-market distance
- Knowledge- vs asset-intensive industries
- Dynamic, regulated, or unionized industries
- Cross-border or culturally distant deals
- Target executive/employee departures
- Impairments, divestitures, or delayed synergies

Finding: CEO meanness is associated with lower acquiring-firm CARs in both [-1,+1] and [-2,+2].
Question for the room: Which test would best isolate post-merger integration problems?

THE BIGGER POINT

NLP lets us study executive traits that matter enormously — but are missing from the literature.

CEO meanness on M&A is our first step.

Where this goes:

- Misconduct
- TMT exodus
- Stakeholder outcomes

Questions

- Questions?
- THANK YOU!!!!
- smulows@wfu.edu