# Entity Neutering

Joseph Engelberg[1] Asaf Manela[2] William Mullins[1] Luka Vulicevic[1]

[1]University of California, San Diego
[2]Washington University in St. Louis

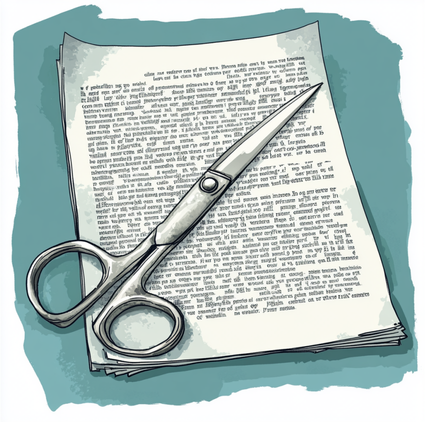
March 17, 2025

# Motivation

- LLMs extract signals from financial text.
- Training on recent data risks look-ahead bias.

# Key Idea

- Entity Neutering removes identifying clues.
- Instructs the LLM to replace names, dates, products, etc., with generic tokens.
- Prevents inference of firm identity or timing.
- Maintains semantic content for sentiment extraction.



# The Neutering Approach

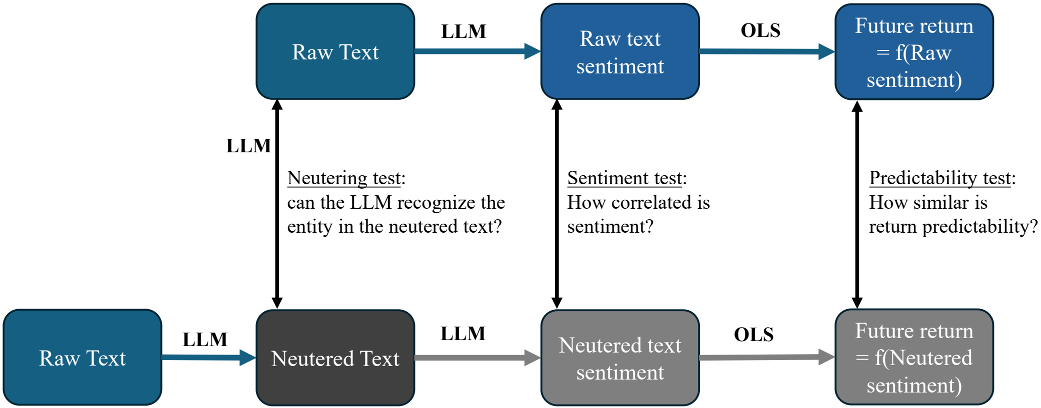
- Detailed prompt replaces:
  - Company names → `Company_x`
  - Brokers, products, dates, numbers → generic tokens
- Example: "General Mills" becomes `Company_1`.
- Applied to over 900K Dow Jones articles.

# Econometric Framework Snapshot

- Look-ahead bias is framed as training leakage.
- Define a neutering function $n(x; \tau)$ to remove leakage.
- Under generalization neutrality, the loss increase bounds the bias.

# Neutering Process Overview

## Procedure



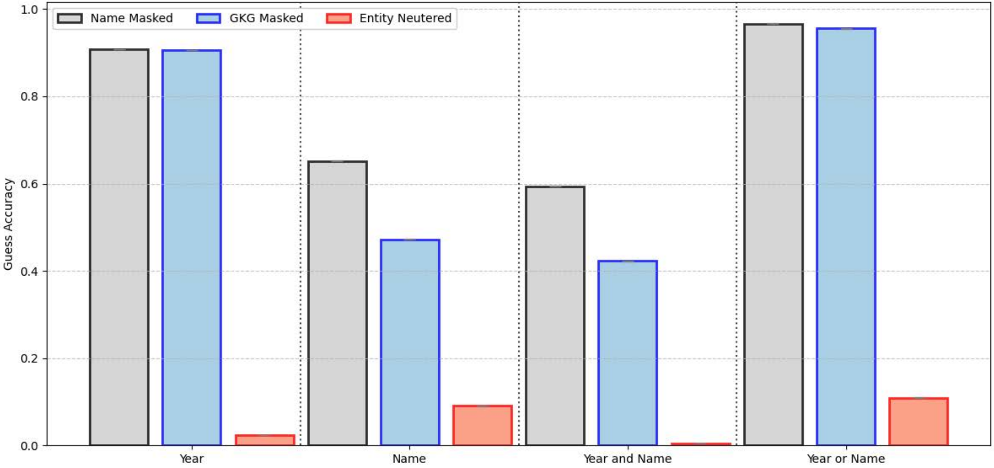
# Neutering Test: Neutering Reduces Firm Recognition to 9%

## Firm identification rate by LLM

| Firm | Model | # parameters | identification rate | # obs. |
|---|---|---|---|---|
| OpenAI | GPT-4o-mini | unknown | 9.1% | 911,283 |
| Meta | Llama 3.2 | 1B | 7.2% | 5,000 |
| Meta | Llama 3.2 | 3B | 7.2% | 5,000 |
| DeepSeek | R1 | 1.5B | 7.2% | 5,000 |

# Low Entity Recognition via Neutering

## LLM guessing accuracy by de-identification method



# Sentiment Test: Preservation of Information

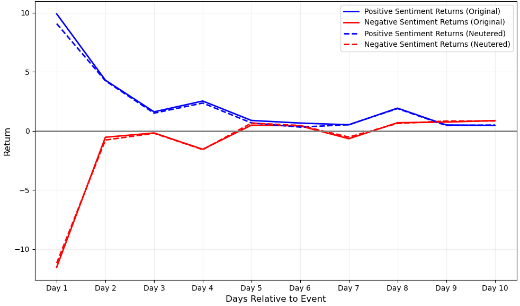
- Compare sentiment from raw vs. neutered text.
- Over 90% of sentiment signals (bullish/bearish) match.
- Magnitude rank correlation is approximately 72%.
- Indicates minimal loss of semantic content.

# Return Predictability Tests

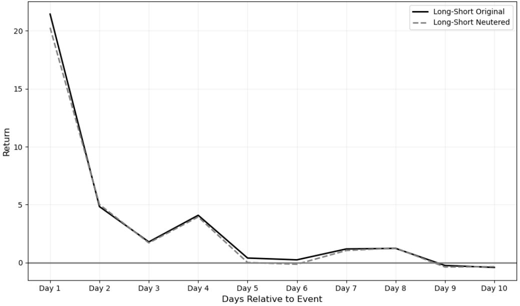
- Extract sentiment from both raw and neutered texts.
- Use sentiment to forecast next-day abnormal returns.
- Overall predictive performance is nearly identical.

# Return Patterns from Neutered Text Mirror Raw Text

## Event study returns by news sentiment



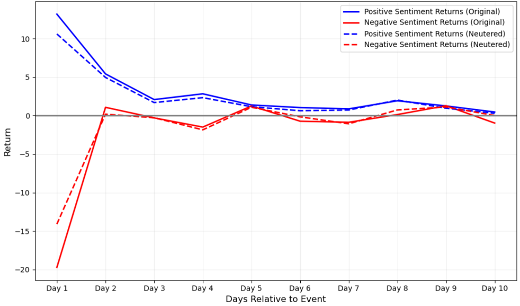

(a) Positive and negative



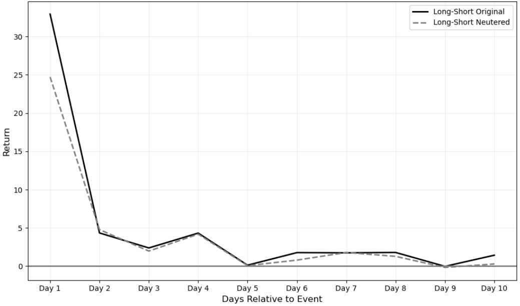

(b) Positive−negative

# Under high sentiment intensity, small differences emerge (17–38%)

## Event study returns by news sentiment




(a) Strong positive and strong negative




(b) Strong positive−strong negative

# Raw and Neutered Text Yield Similar Return Predictability

## Return predictability – raw vs. neutered text

| | Dep. var.: $DGTW_{i,t+1}$ | | | | Dep. var.: $DGTW_{i,(t+2) \to (t+5)}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) Raw | (2) Neutered | (3) Raw | (4) Neutered | (5) Raw | (6) Neutered | (7) Raw | (8) Neutered |
| Positive sentiment$_{i,t}$ | 7.18*** | 5.69*** | | | 7.31*** | 5.17*** | | |
| | [5.66] | [4.76] | | | [3.65] | [2.84] | | |
| Negative sentiment$_{i,t}$ | -10.48*** | -10.80*** | | | -13.96*** | -14.94*** | | |
| | [-6.38] | [-5.95] | | | [-5.79] | [-6.89] | | |
| Strong positive$_{i,t}$ | | | 10.14*** | 7.07*** | | | 11.20*** | 6.96*** |
| | | | [6.81] | [5.62] | | | [4.96] | [3.61] |
| Weak positive$_{i,t}$ | | | 1.06 | -0.49 | | | -0.62 | -2.85 |
| | | | [0.79] | [-0.32] | | | [-0.28] | [-1.13] |
| Weak negative$_{i,t}$ | | | -1.40 | -3.16 | | | -4.37* | -6.06** |
| | | | [-0.69] | [-1.22] | | | [-1.67] | [-2.16] |
| Strong negative$_{i,t}$ | | | -20.49*** | -14.11*** | | | -24.54*** | -18.78*** |
| | | | [-9.62] | [-7.72] | | | [-5.93] | [-6.99] |

# Comparable Trading Strategy Returns

## Trading strategy returns

| | Dep. var.: Long-short$_{t+1}$ | | | | | |
|---|---|---|---|---|---|---|
| | (1) Raw | (2) Neutered | (3) Difference | (4) Raw & strong | (5) Neutered & strong | (6) Difference & strong |
| $\alpha$ | 25.69*** | 24.72*** | 0.97 | 37.66*** | 29.69*** | 7.97*** |
| | [9.87] | [8.98] | [1.38] | [10.43] | [9.61] | [6.97] |
| Market | -0.04* | -0.03 | -0.01 | -0.07* | -0.04 | -0.03* |
| | [-1.65] | [-1.31] | [-0.68] | [-1.76] | [-1.24] | [-1.88] |
| HML | -0.10 | -0.10 | -0.00 | -0.14 | -0.13 | -0.00 |
| | [-1.51] | [-1.41] | [-0.20] | [-1.27] | [-1.45] | [-0.11] |
| SMB | -0.02 | -0.03 | 0.01 | -0.06 | -0.05 | -0.02 |
| | [-0.28] | [-0.42] | [0.53] | [-0.69] | [-0.56] | [-0.63] |
| UMD | -0.01 | -0.00 | -0.01 | 0.04 | 0.01 | 0.02 |
| | [-0.40] | [-0.09] | [-0.95] | [0.77] | [0.32] | [1.51] |
| RMW | -0.02 | 0.00 | -0.02 | -0.00 | 0.02 | -0.03 |
| | [-0.33] | [0.05] | [-1.11] | [-0.01] | [0.44] | [-0.97] |
| CMA | -0.02 | -0.05 | 0.03 | -0.07 | -0.04 | -0.04 |
| | [-0.26] | [-0.56] | [1.22] | [-0.59] | [-0.34] | [-1.04] |
| # observations | 2,219 | 2,219 | 2,219 | 2,219 | 2,219 | 2,219 |
| Adj. $R^2$ | 0.002 | 0.001 | 0.001 | 0.005 | 0.003 | 0.007 |

# Takeaways

- Entity neutering effectively anonymizes text and mitigates training leakage.
- Semantic content is largely preserved.
- Return predictability is nearly identical, bounding look-ahead bias.
- LLMs serve both as a source of bias and a tool for mitigation.

# Future Directions

- Enhance neutering methods as LLMs evolve.
- Extend the approach to other domains with textual bias.
- Refine econometric bounds on training leakage.