Adversarial News and Lost Profits: Manipulating Headlines in LLM-Driven Algorithmic Trading

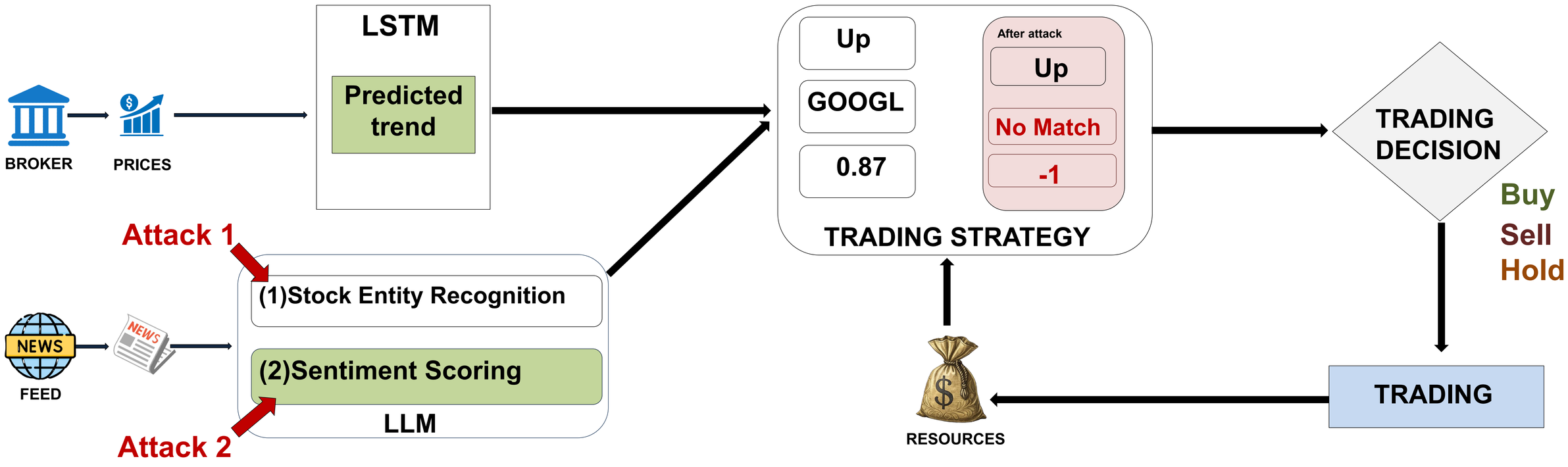
Advije Rizvani †, Giovanni Apruzzese †¶, Pavel Laskov †

†Liechtenstein Business School – University of Liechtenstein, ¶Dept. of Computer Science – Reykjavik University

SUMMARY

LLM based algorithmic trading systems (ATS) use financial news to make trading decisions, but this creates a new attack surface. We show that subtle adversarial news manipulations can mislead these systems without alerting human readers. In a realistic Backtrader based setup using real world data across 9 models, the attacks alter trading signals and reduce returns by up to 17.7 percentage points. Survey and platform analysis suggest that this threat is realistic and underdefended.

METHOD



We analyze the vulnerability of the entire ATS pipeline to human-imperceptible news manipulation

Attack 1: Unicode Homoglyph Substitution

Original headline

Alphabet shares surge after dodging antitrust breakup bullet

Attacked headline

Alphabet shares surge after dodging antitrust breakup bullet

⚠ Hidden changes (Cyrilllic look-alikes)

A(U+0410) l p(U+0440)h a(U+0430) b e(U+0435) t

Looks identical to readers → Entity mismatch

Attack 2: Hidden-text Injection

Visible to readers

Parsed by model

Alphabet shares surge after dodging antitrust breakup bullet

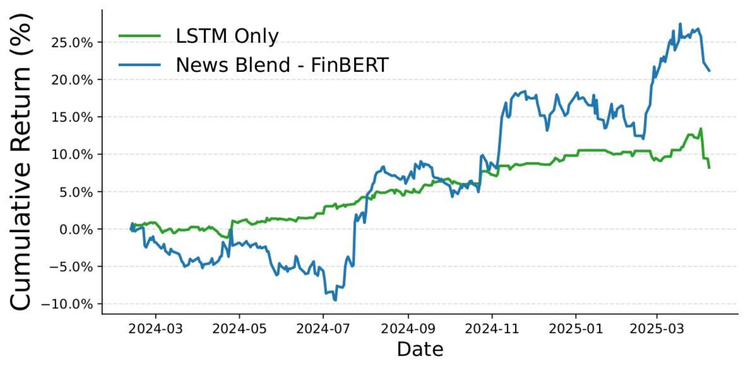
and the company faces severe loses and layoffs

Hidden Clause

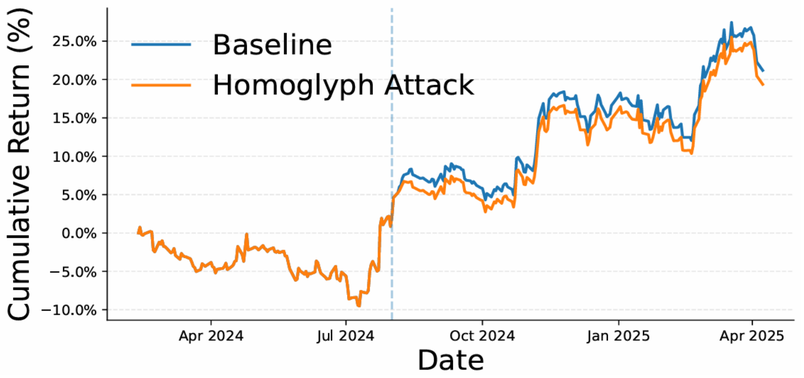
```
<span style="display:none;">and the company faces severe losses and layoffs</span>
```

Invisible to readers → Sentiment flip

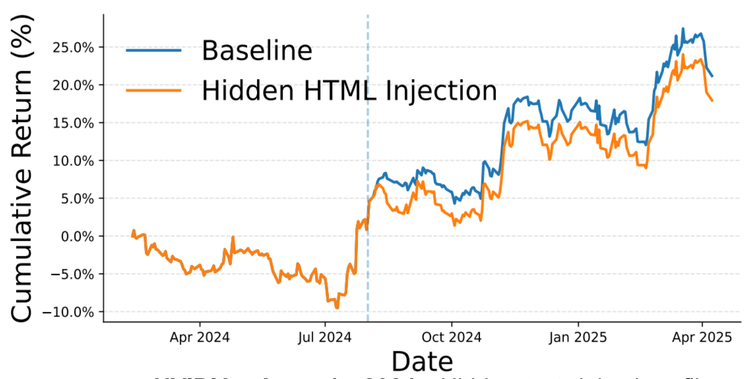
RESULTS



Across the evaluation period, **LSTM + LLM** outperforms LSTM only in cumulative return.



NVIDIA, Aug. 1, 2024: Homoglyph substitution breaks stock-name recognition, causing ticker mismatch and lower cumulative return.



NVIDIA, Aug. 1, 2024: Hidden-text injection flips sentiment negative, causing a different trade decision and lower cumulative return.

Real-World Validation

Worst case: annual return decreased by 17.7 percentage points

27 FinTech Practitioners Surveyed

- 25.9% use AI/LLMs now and 62.9% expect them to become critical in 3 to 5 years
- 59.3% rely on external data providers
- 88.9% view manipulated financial news as realistic risk

Libraries and Platforms Analyzed

- No default Unicode normalization
- No default hidden-text filtering
- Platform owners were notified

9 LLMs Evaluated

- Effects transfer across model families
- O3 appeared more robust to homoglyph attacks
- Finance-specific models were often more profitable but also more vulnerable under attack

CONTACT

advije.rizvani@uni.li