Words that Work: Using Language to Generate Hypotheses

Summarized by： Weihao Zhao

03.28.2025

1 Research questions

How to use text data to automatically generate, sort, and screen language features or marketing hypos that affect consumer engagement?

2 Why are the research questions interesting?

• Language is not only an important clue for understanding consumer decision-making processes, but also a valuable source of data for gaining insights into consumer behavior.
• Existing research only focuses on specific text features, making it difficult to systematically capture multidimensional info.
• The generation of assumptions mainly relies on human intuition and experience(time-consuming and biased).
• LLM and ML compensate for shortcomings of existing methods in terms of systematicity and interpretability.

3 What is the paper's contribution?

***(1)* Literature on consumer psychology and language effects on engagement**
Prior: Utilizing text-as-data and NLP, lack of interpretability and generalizability.
This: Introducing a framework to convert unstructured text into marketing insights.

***(2)* Literature studying how language affects engagement**
Prior: Examined how specific textual features capture consumer attention and sustain engagement.
This: Generating and testing actual marketing hypotheses.

***(3)* Literature on organizational learning**
Prior: Organizations today continuously run A/B tests to learn how various messages affect consumers' behavior.
This: Aggregating insights from tons of A/B tests in the form of specific hypos that others can test.

***(4)* Literature on data-driven discovery and hypothesis generation**
Prior: Explored automated hypos generation using structured datasets or experimental simulations.
This: Pioneers a scalable method applies LLMs to generate hypos from A/B test data, with ML-based models.

4 What hypotheses are tested in the paper?

H1: Automated methods can generate high-quality hypotheses.
H2: Language features significantly affect click through rates.
H3: The hypos of automated generation is superior to traditional methods.

a) Do these hypotheses follow from and answer the research questions?
• Yes, these three hypotheses have been properly validated and answered in the paper.
Do these hypotheses follow from theory or are they otherwise adequately developed?
• H2: Consumer psychology, language influence consumer behavior by triggering emotions, curiosity, etc.
• Text as Data theory: Text is also a type of structured data that can extract features through ML.

5 Sample: comment on the appropriateness of sample selection procedures.

Dividing data into multiple sets, the training, tuning, and validation processes do not interfere with each other.

6 Dependent and independent variables: comment on the appropriateness.

CTR, as an indicator of actual behavior; PTE provide basis for evaluating effectiveness of language modifications.

7 Regression model specification: comment on the appropriateness.

The regression design of this study is rigorous, systematic, and scientific.

8 What difficulties arise in drawing inferences from the empirical work?

The data source is relatively single, and the results may reflect more the marketing strategy effects of specific platforms and periods, rather than necessarily applicable to other fields.

9 Describe at least one publishable and feasible extension of this research.

• Expand data sources and use cross platform and multi domain data to verify the universality and stability of the automated hypothesis generation process.
• Adopting a more rigorous experimental design, adding an external control group, or using a structural causal model to decompose the impact path, further verifying the causal effects of the hypothesis.

10 Summarize the similarities, differences, and correlations between literature.

三篇文献共同描绘了 LLMs 在学术研究中的双刃剑效应：

• 效率革命：自动化生成假设、论文或实验数据，极大提升研究速度与规模。
• 伦理危机：HARKing、虚构引用、黑箱模型等问题威胁学术可信度。

差异化定位：

• 本文：作为应用导向研究，为营销实践提供可解释框架，弥合数据科学与消费者心理学的鸿沟。
•Manning 等：作为方法论创新，推动社会科学实验的自动化与因果推断标准化。
•Novy-Marx 等：作为伦理警示，揭露 AI 对学术生态的系统性风险，呼吁制度性改革。

表 1: 三篇文献对比分析

对比维度	市场营销文献	社会科学文献 (Manning 等)	金融文献 (Novy-Marx 等)
研究领域	消费者语言与参与度	社会互动行为模拟	股票收益预测与学术伦理
核心问题	如何从非结构化文本生成可解释的营销假设	如何自动化生成并验证社会科学因果假设	如何工业化生成论文及其对学术生态的影响
方法论	– LLM 生成假设 – ML 排序 + 人类验证 – 跨平台测试	– 结构因果模型 (SCMs) – LLM 代理模拟互动 – 拍卖理论验证	– ”Assaying Anomalies” 协议 – 批量生成论文 – 虚构引用检测
技术贡献	– 文本 → 假设转化框架 – A/B 测试洞见提取	– 自动化社会科学系统 – 社会互动实验标准化	– 论文生成工业化流程 – 学术伦理风险量化
潜在挑战	– ML 黑箱问题 – 假设泛化性限制	– 效应幅度误差 – 对话终止规则缺失	– 虚构引用网络 – 同行评审过载风险
未来方向	– 非文本数据扩展 – 组织因果学习	– 复杂交互建模 – 持续实验循环	– AI 参与披露制度 – 引用追踪区块链