# Summary of Words that Work: Using Language to Generate Hypotheses

*Rafael M. Batista and James Ross*

*Working Paper*

1. **What are the research questions?**
   - Can LLMs and ML systematically generate hypotheses about how specific features influence consumer behavior?

2. **Why are the research questions interesting?**
   - Manual hypothesis generation in language studies is slow, context-dependent;
   - LLMs and ML can automate pattern detection in unstructured text.
   - This paper tries to use LLM and ML in hypothesis generation.

3. **What is the paper's contribution?**
   1. Exploring how researchers can use text to study consumer behavior.
      - **Past:** just study text data.
      - **Expand:** introduced a framework to convert unstructured text.
   2. Studying how language affects engagement.
      - **Expand:** uncover new insights.

4. **What hypotheses are tested in the paper?**
   - H: LLM and ML can help us generate hypotheses about language and human behavior.
   a) **Do these hypotheses follow from and answer the research questions?** Yes.
   b) **Do these hypotheses follow from theory?Explain logic of the hypotheses.**
      - theory: LLM and ML could treat unstructured data.

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

   - It is proper as it leverages randomized A/B test data and diverse social media posts.

6. **Comment on the appropriateness of variable definition and measurement.**

   - They are proper as engagement is measured via CTRs, linguistic features are validated.

7. **Comment on the appropriateness of the regress/predict model specification.**

   - It is proper as it combines OLS and ML to balance interpretability and predictive power.

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

   - Difficulty: subjective human/GPT labeling biases.

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

   - Test the framework on diverse text domains.

10. **Describe the relationship between papers.**
    - They all use LLM in the process of studying.