

Summary of Uncovering Information: Can AI Tell Us Where to Look?

Rafael M. Batista and James Ross

Working Paper

1. **What are the research questions?**
 - Whether LLM can extract new information in textual data and explain market response?
2. **Why are the research questions interesting?**
 - There is a sea of narrative text released by a firm which is necessary but hard to tackle.
 - Prior studies rely on researcher's subjective choice and ignore surprise of a disclosure.
 - This paper uses LLM to extract the new information of a specific disclosure.
3. **What is the paper's contribution?**
 1. Literature on textual analysis.
 - **Past:** mainly focus on annual and quarterly reports.
 - **Expand:** use LLM to extract new information from other forms of reports.
 2. Literature that uses LLM to predict returns.
 - **Past:** an LLM is used to transform a source document into another representation.
 - **Expand:** preserve the source document and use LLM to overlay a measure.
 3. Literature on costly information processing.
 - **Past:** measure the similarity of text across these documents.
 - **Expand:** extend this idea by forming priors over all disclosures
4. **What hypotheses are tested in the paper?**
 - H: LLM can measure disclosure information and predict market response and return.
 - a) **Do these hypotheses follow from and answer the research questions?** Yes.
 - b) **Do these hypotheses follow from theory? Explain logic of the hypotheses.**
 - Yes, information in firm disclosure is correlated with its future performance.
5. **Sample: comment on the appropriateness of the sample selection procedures.**
 - The procedure is appropriate as it ensures a diverse and representative set of firms.
6. **Comment on the appropriateness of variable definition and measurement.**
 - It's proper since it uses information theory principles and LLM to quantify information.
7. **Comment on the appropriateness of the regress/predict model specification.**
 - model captures the relationship between information content and market reactions.
8. **What difficulties arise in drawing inferences from the empirical work?**
 - LLM may wrongly extract the new information.
9. **Describe at least one publishable and feasible extension of this research.**
 - this research could involve exploring the impact of regulatory changes on the information content and market reactions to corporate disclosures.
10. **Describe the relationship between papers.**
 - They all use LLM to solve financial questions.