

find out sensitivity of quality of workflow based on prompt  
overall how variation based on successive prompt quality  
how is overall performance changing based on prompt

write abstract, submit by eod:

for example:

in an automated agentic world how the quality of prompt impacts the quality of response

my prompt - 1

create gemini prompt - 0.9

Multiplicative property will reduce the response quality as the number of agent increase  
(cascade effect)

how response quality varies

## ABSTRACT

In an automated agent world where agentic AI is fast becoming the fore-runner in the field of cybernation, an important aspect that must be taken into account is the variability of the quality of the prompt utilized by the agents, and how it impacts the quality of the response produced. This paper will highlight the findings that came about when the quality of workflow based on prompt was evaluated.

Formalized version:

The rapid integration of agentic ai systems in autonomous (automated?) workflows requires (syn?) a rigorous eval of input prompt quality when assessing its dependability with regards to the output generation. This paper presents the findings of the analysis done on (experiments conducted to determine?) the prompt sensitivity across various agents (are we? How many?), examining how variations in the prompt - including clarity, specificity, comprehensiveness, and structure - directly influences response accuracy, reliability, and deviation from prompt. Our findings bolster our initial assumption that <> (initial assumption?)

Findings indicate that <>

(what is the conclusion? What did we arrive at?)

What did we do?

We analyzed how prompt quality (calculated using an agentic ai toolkit/model we created and trained on x data points, with memory = True to save and make better), when decremented unit-by-unit from 1.0 to 0.7, resulted in the workflow success rate dropping from A% to B% after P agent interactions (multiplicative property)

What all to create to test this?

- Prompt generator - simple enough, maybe?
- Prompt evaluator (scale of 0.00 to 1.00) - have to give this data to train on

- Guinea pig agentic ai workflow
- Response quality evaluator (0% to 100%) - have to give this data to train on
  - How to calculate performance decay?
    - ???
- Do MANUAL response quality check

This is the description