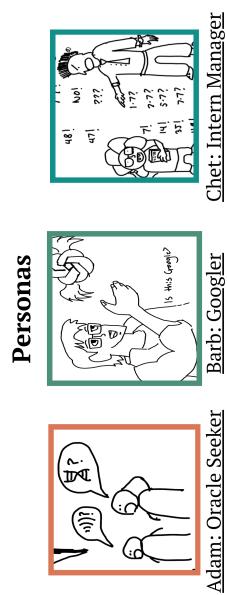
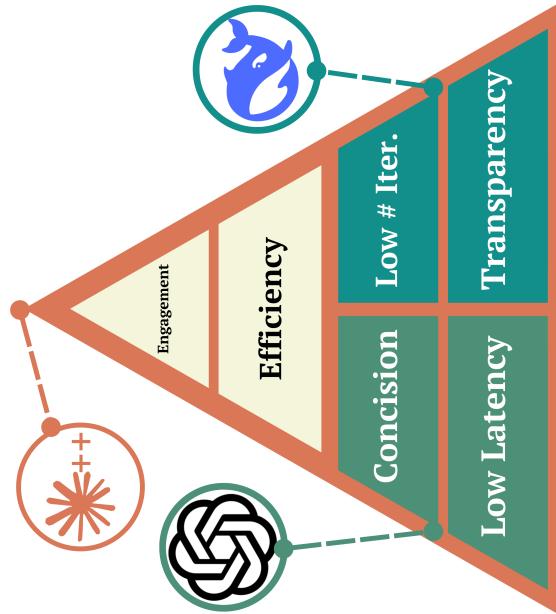




# The Claude, the Sonnet, and the Wholly Spearmint

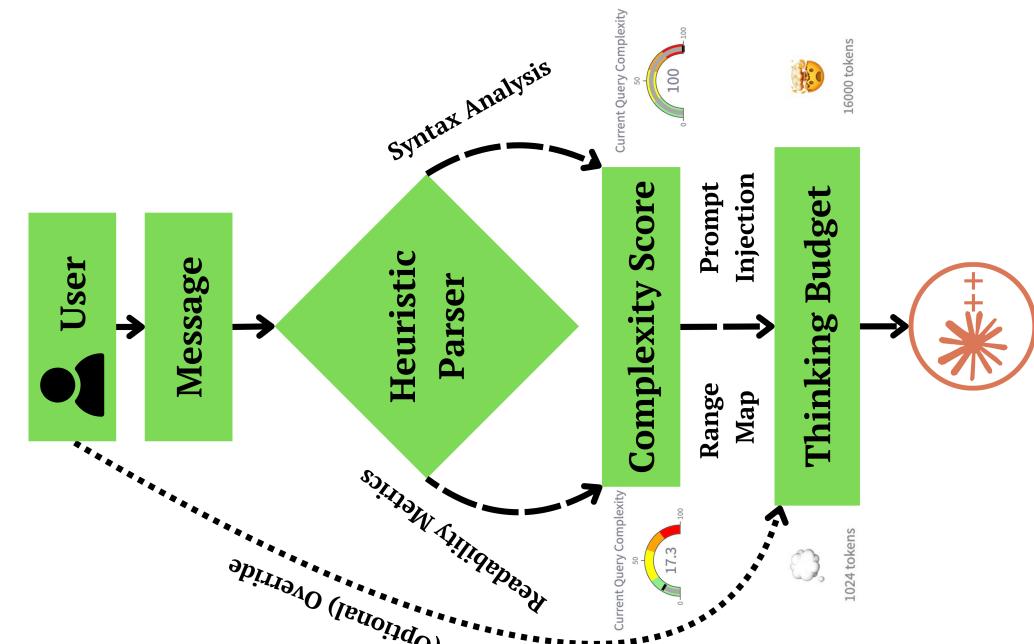
Michael Simon, Matthew Workman, and Ken Zhou  
 {mlsimon, mworkman41, kbzhou01}@cs.ucla.edu

## Problem Formulation

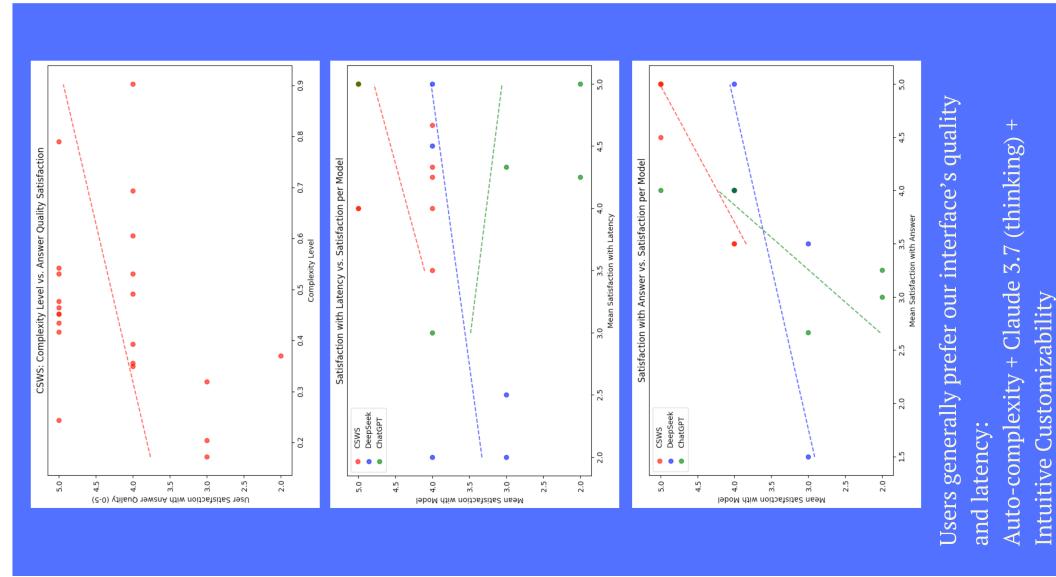


**Status Quo:**  
**Burden of choice:** Users need to select a model to match their task complexity  
**No ETA for LLMs:** Inherent non-deterministic design  
**Best are Slow:** The most powerful “thinking” LLMs are slowest

## Interface



## Evaluation



Users generally prefer our interface's quality and latency:  
 Auto-complexity + Claude 3.7 (thinking) + Intuitive Customizability