Explanation Of Revision

- Reframing the Core Contribution: We have fundamentally restructured the paper's narrative. The Abstract, Introduction, and Discussion now foreground our **modular evaluation framework** as the primary, generalizable contribution. The deep analysis of the SeeAct agent is positioned as a motivating **case study** that demonstrates the framework's power to uncover specific, actionable insights (e.g., pipeline bottlenecks and architectural trade-offs) that are missed by end-to-end metrics. This reframing clarifies our contribution's broader relevance to the agent evaluation community.
- Strengthening Comparison to Related Work: We have expanded the Related Work section to explicitly discuss other evaluation frameworks, clearly differentiating our approach and its unique contributions.
- Methodological and Presentation Refinements: We have clarified methodological details, improved figure/table captions, and performed a thorough proofread to address all minor concerns regarding clarity and presentation.
- On Correlated Metrics (RE Acc. and AP Acc.): We have clarified their distinct roles in the metric definitions (Sec 3.1). We explain that Relevant Element measures the quality of the *candidate set* itself (a prerequisite for success), while Action Prediction Accuracy measures the agent's ability to select correctly *within* that set. Both are necessary for a complete diagnosis.
- Complete Restructuring of Results and Analysis: We have merged and rewritten the entire Results and Analysis section (now Sec 5) into a single, cohesive narrative. This new structure logically flows from identifying high-level bottlenecks to a deep-dive analysis of their causes, making the paper's core findings much easier to follow.
- Discussion (Sec 6): The entire Discussion section has been rewritten to focus on broad implications for the design of *all* web agents (e.g., the need for global context in parallel architectures, the tension between structured and holistic reasoning, and the need to rethink benchmarking).
- On Typos and Formatting: We have performed a full proofread of the paper to correct all noted typos, spacing issues, and ensure consistent formatting for abbreviations like TG, VC, etc.

We are confident that these substantial revisions have addressed the reviewers' concerns, resulting in a much stronger, clearer, and more impactful paper.