# LAZY BELIEF EXPANSION: DECIDING WHAT TO OBSERVE FOR WORLD-MODEL PLANNING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

World models predict action consequences, but prediction alone does not determine which evidence the next decision needs. Exposing all available information adds processing cost and can distract reasoning, while a fixed compressed view can omit a decision-critical detail. We introduce **Lazy Belief Expansion** (LBE), a mechanism that lets the agent decide which evidence should condition world-model predictions and the next action. LBE retains source access while maintaining a compact working belief that can expand as evidence needs emerge. A controller chooses whether to `READ` from the current source, `ACT` to obtain missing evidence through interaction, or `COMMIT` the belief to task-directed planning; both environmental routes use the same planner. In a training-free web instantiation using WMA as the base planner, mean success rates across repeated runs on the 812-task benchmark are 19.33% for LBE and 18.72% for WMA. Among successful LBE tasks, final planning evidence is reduced by 54.6% relative to the corresponding untruncated source observations and by 52.8% relative to the context-capped full-observation inputs supplied to WMA; for tasks successfully completed in at most 5, 6–10, and 11–15 steps, LBE's mean input-plus-output token usage is 22.4%, 16.0%, and 4.7% lower than WMA's, respectively. Together, these results show that LBE integrates deciding what to know into deciding what to do without modifying the underlying predictive world model.

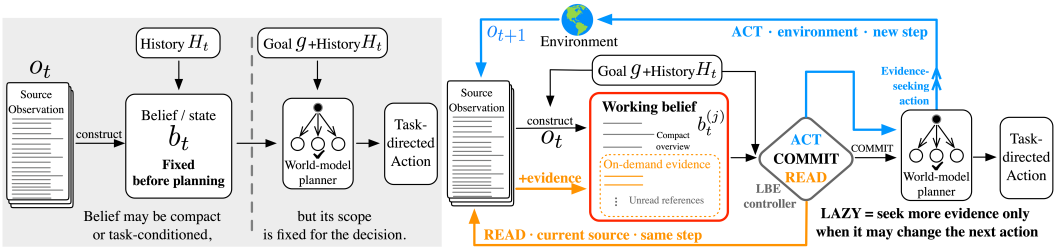
## 1 INTRODUCTION

World models support planning by answering a predictive question: *what would follow from an action?* (Ha and Schmidhuber, 2018; Hafner et al., 2019). Yet answering this question requires specifying the evidence on which each prediction is conditioned. Planning therefore involves not only predicting action consequences, but also deciding which evidence those predictions should use. Available information should not automatically become planning context: processing irrelevant content incurs overhead, and language-model studies show that distractors and the placement of relevant facts can impair reasoning (Shi et al., 2023; Liu et al., 2024; Yang et al., 2025b). In predictive planning, the selected evidence can be processed repeatedly for action proposal, consequence prediction, and evaluation. Yet a fixed compressed view has the opposite risk: it can omit a detail whose relevance becomes clear only during deliberation. The problem is therefore not simply to provide more information or a shorter representation, but to decide *when additional evidence is worth exposing for the next action*. We study this problem as *decision-conditioned observation*.

We propose **Lazy Belief Expansion** (LBE) to give the agent control over this choice: *world models predict action consequences;* LBE *decides which evidence those predictions should use.* The source remains accessible while a compact *working belief* exposes selected evidence and references to unread content (Figure 1). This separates avoiding unnecessary processing from permanently discarding information. From its goal, history, and current belief, the agent chooses among three outcomes: `READ` expands the working belief with previously unexposed evidence from the retained source; `ACT` obtains evidence through environmental interaction; and `COMMIT` submits the current belief for task-directed planning. `READ` expands the evidence available to the planner without advancing the environment, whereas `ACT` advances the environment to obtain evidence unavailable in the current source. Both `ACT` and `COMMIT` invoke the same predictive planner, but for different environmental intents. The agent can therefore proceed with little evidence when sufficient and expand when a consequential detail is missing.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.




**(a) Conventional: Belief Fixed Before Planning** **(b) Lazy Belief Expansion**

Figure 1: **World models predict action consequences; Lazy Belief Expansion controls the evidence those predictions use.** (a) Conventional world-model planning constructs a belief/state $b_t$ before planning. It may already be compact or task-conditioned, but its evidence scope is fixed for the current decision. (b) LBE retains the source observation while maintaining a compact, expandable working belief $b_t^{(j)}$. At each internal round, the controller chooses among three outcomes. `READ` expands the working belief with selected evidence from the retained source within the same environmental step. `ACT` addresses evidence unavailable in that source: the world-model planner selects an evidence-seeking environmental action, yielding a new observation $o_{t+1}$. `COMMIT` submits a sufficient working belief to the same planner for task-directed action selection. Either `ACT` or `COMMIT` may be selected without a preceding `READ`. Additional evidence is deferred until it may change the next action, leaving unnecessary source content outside the planning context.

We instantiate LBE in web navigation, a natural testbed in which relevant evidence may be present but unread in the current observation, or obtainable only through interaction. The implementation uses WMA as its predictive planner (Chae et al., 2025) and constructs the initial belief programmatically, before the first language-model call. The mean success rates across repeated runs on the 812-task WebArena benchmark (Zhou et al., 2024) are 19.33% for LBE and 18.72% for WMA. Among successful LBE tasks, final planning evidence is reduced by 54.6% relative to the corresponding untruncated source observations and by 52.8% relative to the context-capped full-observation inputs supplied to WMA. For tasks successfully completed in at most 5, 6–10, and 11–15 steps, LBE's mean input-plus-output token usage is 22.4%, 16.0%, and 4.7% lower than WMA's, respectively. We further illustrate this principle with a qualitative example involving two consecutive decisions: evidence scope should adapt to the current decision, with the agent deciding when to expand it rather than having it fixed in advance by source length.

Rather than pursuing maximal compression, LBE follows a principle of *selective exposure with recoverable access*: evidence that does not help the current decision can remain outside the planning context while remaining available to later decisions. Our contributions are:

**Decision-conditioned observation.** We formulate *which evidence should condition world-model predictions and the next action* as an agent decision, addressing the tension between unnecessary exposure and premature information removal.

**Lazy Belief Expansion.** We introduce a source-preserving `READ`/`ACT`/`COMMIT` mechanism that defers additional evidence until needed, distinguishes local reading from environmental acquisition, and uses the same predictive planner for both environmental intents.

**Evidence-efficient task execution.** We provide a training-free web instantiation with task-success estimates close to its base planner, substantial final-evidence reduction among successful tasks, lower input-plus-output usage for tasks successfully completed in at most 5, 6–10, and 11–15 steps, and decision-dependent reading behavior.

## 2 Background and Related Work

### 2.1 World models and their conditioning information

World models support prediction and control by learning environment dynamics (Ha and Schmidhuber, 2018; Hafner et al., 2019). In partially observable settings, predictions of action consequences

must be conditioned on a representation of the observation–action history. Classical planning represents this information as a belief state, namely a posterior distribution over hidden states given the history (Kaelbling et al., 1998), whereas learned world models often use recurrent or latent states as approximate information states that summarize this history (Ha and Schmidhuber, 2018; Hafner et al., 2019). PlaNet, for example, infers an approximate belief over a latent hidden state before searching for future actions in latent space (Hafner et al., 2019). In web navigation, WMA simulates the consequences of candidate actions and predicts transition-focused abstractions that emphasize important state changes while avoiding repetitive page content (Chae et al., 2025). WebDreamer similarly uses a language model as a world model to simulate the outcomes of candidate web actions before execution (Gu et al., 2025). Although these methods differ in domain and representation, the information conditioning their predictions is constructed or supplied before candidate actions are evaluated. When this information is compact or selectively presented, details from the source observation may remain outside the planner-visible context. These existing methods do not model whether or when the scope of conditioning evidence should be expanded within the same decision process. Motivated by this gap, LBE treats evidence scope as an explicit agent decision.

### 2.2 Information gathering and computation as decisions

Planning under partial observability treats information-revealing actions as part of the agent's policy (Kaelbling et al., 1998). Active perception similarly selects sensing operations according to task goals (Bajcsy et al., 2018), while metareasoning evaluates internal computation by its effect on subsequent external decisions (Russell and Wefald, 1991). These perspectives establish that gathering information and performing computation can themselves be decisions whose value depends on how they affect subsequent action selection.

ReAct provides a related pattern for language agents by interleaving reasoning with environmental interaction (Yao et al., 2023). Taken together, however, these lines of work do not provide a unified formulation that first determines whether additional evidence is needed and then distinguishes evidence already available in the current source from evidence obtainable only through environmental interaction. Motivated by this gap, LBE makes evidence acquisition and commitment explicit decisions within predictive planning.

### 2.3 Observation interfaces and selective reading

For language models, providing more context does not uniformly improve reasoning. Shi et al. (2023) show that irrelevant information can lower arithmetic reasoning accuracy, while Liu et al. (2024) find that the ability to use relevant facts varies with their position in long contexts. Controlled distractor experiments further show that irrelevant context can affect reasoning-path selection, not only arithmetic execution (Yang et al., 2025b). These findings motivate selective exposure, but do not determine which information can be safely excluded or how excluded details should be recovered if they later become relevant.

Web-agent systems reduce observation complexity through filtering, retrieval, and restructuring. The Mind2Web baseline filters raw HTML with a smaller model before action selection (Deng et al., 2023), while AgentOccam aligns the observation and action spaces with the capabilities of language models (Yang et al., 2025a). LineRetriever and FocusAgent retrieve observation lines relevant to future planning or the current task (Kerboua et al., 2025; 2026), whereas Prune4Web generates programs that prune DOM trees (Zhang et al., 2026). These approaches reduce the content presented to the agent before action selection, but provide no explicit mechanism for recovering omitted content within the same decision. ReadAgent provides a complementary strategy for recoverable access by preserving original passages behind gist memories and retrieving details on demand (Lee et al., 2024). However, constructing these memories still requires all source passages to be supplied as context to a language model, and therefore does not avoid the cost of processing the full source.

Together, these methods either reduce observations without recoverable access or provide recoverable access only after full-source language-model processing. In contrast, LBE constructs a compact working belief without first supplying the full source as language-model context, while preserving recoverable access to the complete source observation.

# 3 LAZY BELIEF EXPANSION

We separate *predicting action consequences*, performed by the world model $f_\phi$, from *controlling the evidence for those predictions*, performed by the LBE controller. For a planner $\Pi$, a conventional interface first constructs an information state and then plans from it:

$$b_t = E(o_t, g, H_t) \quad \longrightarrow \quad a_t = \Pi(g, b_t; f_\phi), \tag{1}$$

where $o_t$ is the current source observation, $g$ the goal, and $H_t$ the interaction history. The constructor $E$ incorporates the history needed for planning into $b_t$, so $H_t$ is not supplied to the planner as a separate context field. The restriction is the one-shot handoff: its evidence selection is fixed for the ensuing decision. LBE replaces this handoff with a source-preserving control process that can expand the belief, acquire a new source, or proceed to task-directed planning.

## 3.1 SOURCE ACCESS AND WORKING BELIEF

At environmental step $t$, the system retains the source $o_t$ and exposes a compact working belief $b_t^{(0)}$. After $j$ internal reads, $b_t^{(j)}$ contains an overview, selected evidence, references to unread content, and reading progress. Here, *belief* denotes the information exposed for decision-making. Source access and belief content are separate: preserving an observation does not require including all of it in every planning call. The initial belief is constructed on the source side without requiring a language-model pass over the full observation; only the resulting working belief is exposed to the controller and planner.

$$\begin{aligned} b_t^{(0)} &= \mathrm{Build}(o_t, g, H_t, \mathcal{M}_t), \\ b_t^{(j+1)} &= \mathrm{Update}\Big(b_t^{(j)}, \mathrm{Read}(o_t, q_t^{(j)})\Big), \end{aligned} \tag{2}$$

Here $q_t^{(j)}$ requests evidence from the retained source and $\mathcal{M}_t$ denotes reusable evidence with valid provenance. Reading changes the belief at fixed $t$; a new observation starts a new belief with $j = 0$. The constructor incorporates the prior working beliefs and interaction feedback stored in $H_t$ into $b_t^{(0)}$. Thus, $H_t$ remains system state for constructing the next belief but is not supplied again as a separate context field to the controller or planner. Complete historical sources therefore do not re-enter the planning context.

This representation separates two kinds of missing evidence. An *unexposed* fact is already present in $o_t$ and can be recovered locally. An *unavailable* fact requires an environmental interaction. The distinction concerns the current source, not the hidden environment state: retaining $o_t$ gives no access to observations that have not yet been obtained.

## 3.2 THREE OUTCOMES, TWO TIME SCALES

At each internal round, the controller uses the goal and working belief to choose

$$u_t^{(j)} \sim \mu(\cdot \mid g, b_t^{(j)}), \qquad u_t^{(j)} \in \{\texttt{READ}(q), \texttt{ACT}(d), \texttt{COMMIT}\}, \tag{3}$$

where $q$ identifies source evidence and $d$ describes a decision-relevant information need.

**`READ`: expand from the current source.** When an omitted detail may change the next action and can be recovered from $o_t$, the controller requests it through $q$. Equation (2) updates $b_t^{(j)} \to b_t^{(j+1)}$, after which the controller reconsiders the decision. The environmental step $t$ does not advance.

**`ACT`: obtain a new source through interaction.** When the controller judges that the required evidence is unavailable in $o_t$, it supplies the current belief and information need $d$ to the planner. The planner selects an evidence-seeking environmental action. Execution yields $o_{t+1}$, advances $t$, and initializes a new compact belief. Unlike `READ`, this route changes what is observable rather than revealing more of an existing observation.

**`COMMIT`: plan a task-directed action.** When the controller judges the current belief sufficient for the next task-directed decision, it ends evidence acquisition and submits that belief to the same planner. `COMMIT` is an internal handoff, not an environmental action or episode termination. The selected action advances or completes the task; its feedback supplies the next source when execution continues.

`ACT` and `COMMIT` differ by *decision intent*, not by physical action type. The same operation can acquire an unknown fact in one task and directly fulfill a goal in another. Both routes can reveal incidental information, but only `ACT` is selected primarily to resolve an evidence need. Either route is available at $j = 0$, without a preceding `READ`. "Lazy" means deferring additional acquisition until it may change the next action, not requiring a read before every action.

### 3.3 A SHARED PREDICTIVE PLANNER

Let $b_t$ be the working belief when the controller chooses an environmental route. The planner receives the task goal together with the working belief and a decision intent:

$$\iota_t = \begin{cases} \text{Acquire}(d), & u_t^{(j)} = \texttt{ACT}(d), \\ \text{AdvanceTask}, & u_t^{(j)} = \texttt{COMMIT}, \end{cases} \qquad a_t = \Pi(g, b_t, \iota_t; f_\phi). \tag{4}$$

The intent determines whether candidate actions should obtain evidence or advance the task with the available evidence. Equation (4) separates the roles: $\mu$ controls the evidence and decision intent, $f_\phi$ predicts consequences, and $\Pi$ selects the environmental action. The action space, predictive model, and planning procedure are shared across both intents. Planning is invoked after `ACT` or `COMMIT`, not after every local read.

The goal is *decision sufficiency*, not a globally complete representation of the world state: additional evidence should justify its processing burden by its potential to change the next action. A belief can support the next action without resolving every uncertainty about the environment. Conversely, a fully revealed source can still lack a necessary fact. The controller uses the current evidence, source references, and feedback to decide whether to proceed. This operational notion of sufficiency makes the acquisition decision explicit without requiring a complete reconstruction of the environment.

### 3.4 BOUNDED ACQUISITION AND FEEDBACK

Algorithm 1 specifies the mechanism independently of the source format or planner architecture. Targeted reads have a finite budget $J$. If another read is requested after this budget, the full current source observation is exposed before the controller resolves the external intent. Source exhaustion does not force `COMMIT`: evidence absent from $o_t$ can still require `ACT`. After an environmental action, the source is refreshed and reusable evidence is checked against the new observation. Section 4.1 provides a programmatic realization of these operations for web navigation.

## 4 WEB INSTANTIATION AND EXPERIMENTS

Web interaction exposes the two information needs central to LBE: a fact may be present but unread, or require an action to become observable. Figure 2 makes this distinction concrete in shopping, one of the domains covered by the multisite evaluation. We then examine task completion, final planning evidence, successful-task token usage, and decision-dependent evidence acquisition.

### 4.1 INSTANTIATION AND EVALUATION SETTING

**Agent and source interface.** The source is the current-viewport accessibility tree. A programmatic parser groups it into semantic regions and uses BM25 (Robertson and Zaragoza, 2009) to expose goal-relevant snippets, region references, controls, and execution feedback in a compact overview. The first model input is already compact. In Figure 2, reading prices changes the belief, whereas opening a product page changes the source. `READ` supports region continuation, local lookup, element inspection, and full-source reveal. Targeted reads are bounded; full reveal remains available for

**Algorithm 1** Lazy Belief Expansion at environmental step $t$

**Require:** Source $o_t$, goal $g$, history $H_t$, valid reusable evidence $\mathcal{M}_t$
**Require:** Controller $\mu$, shared world-model planner $\Pi$, targeted-read limit $J$

Retain $o_t$; $b \leftarrow \text{Build}(o_t, g, H_t, \mathcal{M}_t)$
**for** $j = 0, \ldots, J$ **do**
  $u \leftarrow \mu(g, b)$ ▷ READ$(q)$, ACT$(d)$, or COMMIT
  **if** mode$(u) \neq$ READ **then break**
  **end if**
  **if** $j < J$ **then**
    $q \leftarrow \text{request}(u)$; $b \leftarrow \text{Update}(b, \text{Read}(o_t, q))$
  **else**
    $b \leftarrow \text{Update}(b, \text{RevealAll}(o_t))$
    $u \leftarrow \mu(g, b; \{\text{ACT}(d), \text{COMMIT}\})$ ▷ resolve external intent
  **end if**
**end for**
**if** mode$(u) =$ ACT **then**
  $d \leftarrow \text{need}(u)$; $\iota \leftarrow \text{Acquire}(d)$
**else**
  $\iota \leftarrow \text{AdvanceTask}$
**end if**
$a_t \leftarrow \Pi(g, b, \iota; f_\phi)$ ▷ same planner for both intents
Execute $a_t$; terminate if the selected action ends the episode
Otherwise receive $o_{t+1}$ and feedback; refresh $H_{t+1}$ and valid $\mathcal{M}_{t+1}$
Start the next cycle at $t + 1$ with a new compact belief and $j = 0$

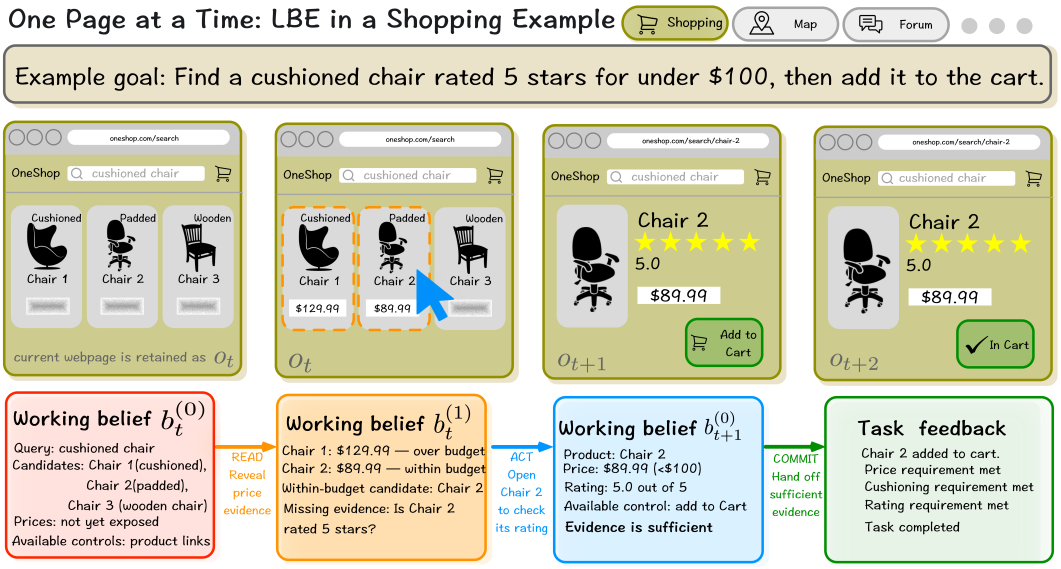



Figure 2: **One Page at a Time: LBE in a Shopping Example.** From the retained current-page source $o_t$, programmatic construction exposes $b_t^{(0)}$. READ reveals previously unexposed prices from the same source and updates the belief to $b_t^{(1)}$ without advancing the environment. This identifies Chair 2 as the cushioned candidate within the $25 budget, while its rating remains unavailable in the current source. ACT therefore uses the shared planner to open the product page, obtaining a new source $o_{t+1}$ and initializing $b_{t+1}^{(0)}$. Once the page confirms a five-star rating and exposes the Add to Cart control, COMMIT hands the sufficient belief to the same planner for the task-directed Add to Cart action; execution yields feedback $o_{t+2}$ confirming completion. The trajectory invokes READ, ACT, and COMMIT once each only for illustration; LBE prescribes neither this order nor these counts.

requests and recovery. In actual model calls, both full-observation inputs and full-reveal fallbacks remain subject to the model's input-token limit; overlength source serializations are truncated before inference. Evidence is reused only while its source is unchanged. Appendix A specifies budgets and validation rules.

**Predictive planner and comparisons.** LBE retains WMA's proposal, prediction, and value-scoring components (Chae et al., 2025). GPT-4o mini serves as policy and observation controller; Llama-3.1-8B-Instruct with WMA adapters supplies world and value models (OpenAI, 2024; Grattafiori et al., 2024). The planner samples 20 policy responses, retains at most three distinct candidates by vote, and samples 20 value scores per candidate. Each task permits at most 15 environmental steps.

FocusAgent and ReadAgent are evaluated as complete selective-context systems without world/value models, rather than as component ablations: their evaluated configurations use an LLM to filter or summarize the full source before direct action selection.

**Evaluation scopes.** All reported results derive from the full 812-task WebArena benchmark. Task success is judged across the benchmark by the WebArena functional evaluator (Zhou et al., 2024); all remaining quantitative summaries condition on successful tasks. Unless otherwise stated, point estimates are means across repeated runs, and bracketed ranges report the observed minimum and maximum across runs rather than confidence intervals. The 54.6% and 52.8% evidence-reduction summaries both condition on successful LBE tasks. The former compares final planner-visible beliefs with the corresponding untruncated local source observations, whereas the latter uses the context-capped full-observation inputs actually supplied to WMA as the reference. Input-plus-output usage is summarized post hoc: after all tasks have been run, successful trajectories are selected and categorized by their realized lengths into those completed in at most 5, 6–10, and 11–15 steps. Because the successful tasks within a given step range can differ across systems, the resulting mean token comparisons are descriptive rather than paired-task effects.

### 4.2 Task completion with decision-conditioned evidence

The mean success summary places LBE at **19.33%**, compared with 18.72% for WMA, 16.50% for FocusAgent, and 11.82% for ReadAgent (Table 1). LBE has the highest point estimate among these configurations. The point estimates differ by 0.61 percentage points and their observed ranges overlap. We use this result to characterize comparable observed task performance, without claiming an established accuracy gain.

Table 1: **Task completion with different evidence interfaces.** Mean success across repeated runs on the 812-task WebArena benchmark; brackets give the observed minimum and maximum. Predictive planning is retained in both WMA and LBE.

| System | Evidence interface | World/value | Success (%) [range] |
|---|---|---|---|
| WMA | Context-capped full observation | Yes | 18.72 [17.61, 20.32] |
| FocusAgent | LLM filtering | No | 16.50 [15.40, 17.36] |
| ReadAgent | LLM gist summarization | No | 11.82 [10.34, 12.44] |
| LBE (ours) | Overview + acquisition control | Yes | **19.33** [18.10, 20.69] |

The task-level finding is that predictive planning remains viable when evidence is selected and expanded, rather than supplied unconditionally. LBE retains the base planner's prediction and evaluation stages; it does not obtain a smaller context by removing them. The next analyses examine how much evidence the agent exposes and how token usage changes across successful task lengths. Ablation analyses of the method components are reported in Appendix C.

### 4.3 Final planning evidence is less than half the source

We compare the final working belief supplied to the planner with the untruncated local source observation available for the same decision, after any within-step evidence expansion. Among successful LBE tasks from the full 812-task benchmark, final planning evidence is **54.6% lower** in token count than the corresponding source observations. This is an endpoint measure rather than the size of the initial overview: the agent may first reveal additional evidence, yet the planner still receives substantially less environment content than is locally available.

Using instead the context-capped full-observation input actually supplied to WMA as the reference, the corresponding savings are **58.6%** for successes completed in at most 5 steps, **51.1%** for those

completed in 6–10 steps, and **47.3%** for those completed in 11–15 steps, with an aggregate of **52.8%**. The aggregate is slightly lower than the source-referenced 54.6% because a small number of overlength source observations are truncated by WMA's input-token limit. Both summaries condition on successful tasks from the same 812-task benchmark and differ in their reference denominator rather than their task scope. Complete values and reported ranges appear in Appendix B.1.

### 4.4 Token use by steps to successful completion

Belief reduction and end-to-end token use are different quantities, so we report the available input-plus-output summaries directly rather than infer a global saving. Table 2 compares LBE with its WMA base planner for tasks successfully completed within each reported step range.

Table 2: **Mean input-plus-output tokens by steps to successful completion.** All values come from successful tasks in the 812-task benchmark. Reductions compare the means reported within each step range and are descriptive rather than paired-task effects.

| Completion steps | WMA | LBE | Reduction vs. WMA |
|---|---|---|---|
| $\leq 5$ steps | 72,663 | 56,367 | **22.4%** |
| 6–10 steps | 236,177 | 198,336 | **16.0%** |
| 11–15 steps | 350,606 | 334,206 | **4.7%** |

LBE uses fewer reported input-plus-output tokens than WMA for tasks successfully completed within all three step ranges. The gap is largest for tasks completed in at most 5 steps, remains 16.0% for those completed in 6–10 steps, and is still 4.7% for those completed in 11–15 steps. We report the ranges separately because the current aggregate record does not provide the number of successful tasks within each range needed to construct a single weighted percentage across all successful tasks, and the comparison is more informative when its dependence on the steps used for successful completion remains visible. The full four-system means and supplied ranges are listed in Appendix B.1.

### 4.5 Reading depth follows the next decision

**A change within the same task.** The clearest contrast occurs *within one task*, when navigation becomes factual answering (Figure 3). Task 132 asks how many code commits Kilian made to `a11yproject` on March 5, 2023. At step 7, the agent opens an author-filtered view using a 1,024-token belief from a 12,748-token source. At step 8, the source is smaller at 3,924 tokens, yet three local reads expand the belief to 2,504 tokens before it answers one commit. The recorded decision has changed from locating a control to extracting a requested count. This case therefore serves a different purpose from the aggregate compression statistics: it shows the mechanism's decision-conditioned behavior. Evidence scope is not a monotonic function of page length; a belief can remain compact for navigation and grow when the next action requires a particular fact.

**Execution feedback and environmental evidence.** After a wishlist addition, the agent retains the success message and the planner stops without rebuilding the page. Separately, among successful LBE tasks, **71.3% [63.2%, 74.5%]** of environmental actions are information-seeking; the complementary **28.7%** are task-directed execution actions. These correspond to the two environmental intents represented by `ACT` and `COMMIT`. `READ` is excluded because it does not advance environmental time.

The shopping schematic explains where each acquisition route applies; the recorded traces show how evidence depth changes with the decision. Together they make the distinction concrete: source size describes what is available, whereas the working belief reflects what the agent exposes for the next action.

## 5 Discussion and Conclusion

This work addresses a missing decision in predictive world-model planning: determining which evidence should be exposed to the planner for the current action. We propose LBE, a training-free

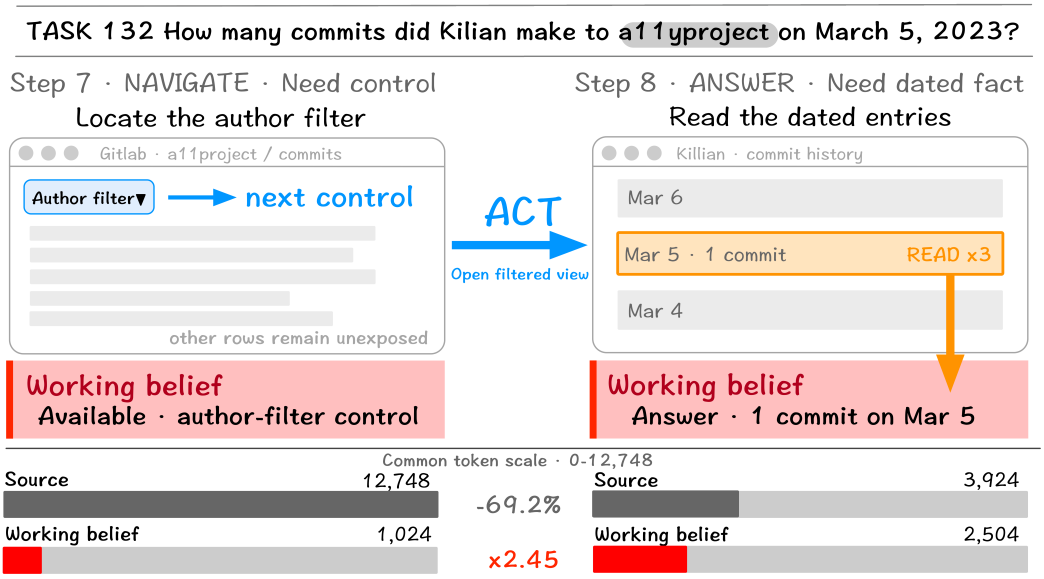



Figure 3: **The next decision changes how much the agent reads.** Two consecutive decisions from recorded GitLab task 132. At step 7, the agent uses a 1,024-token working belief from a 12,748-token source to locate the author filter, then selects `ACT` to open the filtered commit view. At step 8, the new source is shorter (3,924 tokens), but three local `READ` operations reveal the dated entries and expand the working belief to 2,504 tokens, supporting the answer that Kilian made one commit on March 5, 2023. All bars share one absolute token scale: the source shrinks by 69.2% while the working belief grows 2.45-fold. The schematic is reconstructed from recorded token counts and actions; it illustrates that reading depth follows the next decision rather than source length alone.

method that incorporates evidence acquisition into the planning process. Without modifying the underlying predictive world model, LBE programmatically constructs a compact initial working belief before the first language-model call, avoiding the unnecessary cost of supplying the full source observation to the language model while preserving recoverable access to the current source. The controller can then decide autonomously whether to `READ` additional evidence from the retained source, `ACT` to obtain evidence through environmental interaction, or `COMMIT` the current belief to task-directed planning. In the 812-task WebArena evaluation, LBE achieves an observed task-success estimate close to its WMA base planner while substantially reducing final planner-visible evidence among successful tasks. For tasks successfully completed in at most 5, 6–10, and 11–15 steps, LBE's reported mean input-plus-output token usage is 22.4%, 16.0%, and 4.7% lower than WMA's, respectively. A consecutive-decision example further shows that the scope of evidence can adapt to the current decision rather than being fixed in advance by source length.

Several limitations remain. The successful-task token summaries compare means within the reported step ranges rather than paired tasks. In addition, although LBE preserves recoverable access to the full current source at the interface level, full-reveal fallbacks remain subject to the model's context-window limit during actual inference. Overlength sources may therefore be truncated, preventing the model from processing the complete source in a single call. Future work could conduct paired-task analyses, investigate longer-context models together with staged or chunked evidence integration to make full-source recovery more robust, and evaluate the method across broader environments and task domains.

## AI USE STATEMENT

Generative AI assisted conceptual discussion, control formalization, aggregate-result interpretation, literature review, manuscript organization, language revision, and LaTeX preparation. It did not generate the empirical measurements. The authors are responsible for the final methodology, references, text, and claims.

## ETHICS STATEMENT

The study evaluates benchmark web agents rather than autonomous operation of live user accounts. Selective exposure can omit consequential warnings, and evidence-seeking actions can alter the environment. Deployment requires appropriate authorization, validation of actions, and task-specific safeguards.

## REPRODUCIBILITY STATEMENT

Algorithm 1 specifies the mechanism. Appendix A describes source access, read operators, evidence refresh, and implementation settings. Appendix B preserves the numerical summaries and derived step-conditioned usage ratios, with aggregation scope documented in Appendix B.3.

## REFERENCES

Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. Revisiting active perception. *Autonomous Robots*, 42:177–196, 2018. https://arxiv.org/abs/1603.02729.

Hyungjoo Chae, Namyoung Kim, Kai Tzu-iunn Ong, Minju Gwak, Gwanwoo Song, Jihoon Kim, Sunghwan Kim, Dongha Lee, and Jinyoung Yeo. Web agents with world models: Learning and leveraging environment dynamics in web navigation. In *International Conference on Learning Representations*, 2025. https://arxiv.org/abs/2410.13232.

Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Samuel Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2Web: Towards a generalist agent for the web. In *Advances in Neural Information Processing Systems*, 2023. https://arxiv.org/abs/2306.06070.

Aaron Grattafiori et al. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. https://arxiv.org/abs/2407.21783.

Yu Gu, Kai Zhang, Yuting Ning, Boyuan Zheng, Boyu Gou, Tianci Xue, Cheng Chang, Sanjari Srivastava, Yanan Xie, Peng Qi, Huan Sun, and Yu Su. Is your LLM secretly a world model of the Internet? Model-based planning for web agents. *arXiv preprint arXiv:2411.06559*, 2025. https://arxiv.org/abs/2411.06559.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018. https://arxiv.org/abs/1803.10122.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning*, 2019. https://arxiv.org/abs/1811.04551.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998. doi:10.1016/S0004-3702(98)00023-X.

Imene Kerboua, Sahar Omidi Shayegan, Megh Thakkar, Xing Han Lù, Massimo Caccia, Véronique Eglin, Alexandre Aussem, Jérémy Espinas, and Alexandre Lacoste. LineRetriever: Planning-aware observation reduction for web agents. *arXiv preprint arXiv:2507.00210*, 2025. https://arxiv.org/abs/2507.00210.

Imene Kerboua, Sahar Omidi Shayegan, Megh Thakkar, Xing Han Lù, Léo Boisvert, Massimo Caccia, Jérémy Espinas, Alexandre Aussem, Véronique Eglin, and Alexandre Lacoste. FocusAgent: Simple yet effective ways of trimming the large context of web agents. *Transactions on Machine Learning Research*, 2026. https://arxiv.org/abs/2510.03204v2.

Kuang-Huei Lee, Xinyun Chen, Hiroki Furuta, John Canny, and Ian Fischer. A human-inspired reading agent with gist memory of very long contexts. In *International Conference on Machine Learning*, 2024. https://arxiv.org/abs/2402.09727.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024. https://arxiv.org/abs/2307.03172.

OpenAI. GPT-4o mini: Advancing cost-efficient intelligence. Technical announcement, July 18, 2024. Official model announcement.

Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389, 2009. doi:10.1561/1500000019.

Stuart Russell and Eric Wefald. Principles of metareasoning. *Artificial Intelligence*, 49(1–3):361–395, 1991. doi:10.1016/0004-3702(91)90015-C.

Freda Shi, Xinyun Chen, Kanishka Misra, Nathan Scales, David Dohan, Ed Chi, Nathanael Schärli, and Denny Zhou. Large language models can be easily distracted by irrelevant context. In *International Conference on Machine Learning*, 2023. https://arxiv.org/abs/2302.00093.

Ke Yang, Yao Liu, Sapana Chaudhary, Rasool Fakoor, Pratik Chaudhari, George Karypis, and Huzefa Rangwala. AgentOccam: A simple yet strong baseline for LLM-based web agents. In *International Conference on Learning Representations*, 2025. https://arxiv.org/abs/2410.13825.

Minglai Yang, Ethan Huang, Liang Zhang, Mihai Surdeanu, William Yang Wang, and Liangming Pan. How is LLM reasoning distracted by irrelevant context? An analysis using a controlled benchmark. In *Empirical Methods in Natural Language Processing*, pp. 13329–13347, 2025. https://aclanthology.org/2025.emnlp-main.674/.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023. https://arxiv.org/abs/2210.03629.

Jiayuan Zhang, Kaiquan Chen, Zhihao Lu, Enshen Zhou, Qian Yu, and Jing Zhang. Prune4Web: DOM tree pruning programming for web agent. In *AAAI Conference on Artificial Intelligence*, 2026. https://arxiv.org/abs/2511.21398.

Shuyan Zhou, Frank F. Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig. WebArena: A realistic web environment for building autonomous agents. In *International Conference on Learning Representations*, 2024. https://arxiv.org/abs/2307.13854.

# A Implementation Specification

## A.1 Source access and evidence lifetime

The source store contains only the current observation returned by the environment. Unvisited pages, hidden website state, and future observations are outside that store. The browser and evaluator may retain complete execution records, but those records are not automatically planning inputs. The policy, world, and value prompt builders use the selected belief in their environmental-evidence fields; raw HTML and screenshot fields must not bypass this interface. History carries previous working beliefs rather than complete historical source observations.

Table 3: **Information lifetime under the three-outcome control mechanism.**

| Object | During a `READ` | After an environmental action |
|---|---|---|
| Source $o_t$ | Same retained snapshot | Replaced by $o_{t+1}$ |
| Belief $b_t^{(j)}$ | Adds requested evidence and progress | New compact belief at $j = 0$ |
| History $H_t$ | No environmental transition appended | Adds action and execution feedback |
| Read cursors | Advance within the retained source | Reused only for unchanged content |
| Reusable evidence $\mathcal{M}_t$ | Keeps valid source provenance | Invalidated on URL/content change |

A region is reusable only if its URL and content hash remain unchanged. Full-reveal status likewise persists only for an unchanged observation. Changed regions, alerts, and errors receive priority in the next overview. Thus source reuse does not silently convert stale content into current evidence. `COMMIT` itself changes none of these environmental objects: the task action selected after it determines whether execution continues and the source is refreshed.

## A.2 Programmatic overview and read operators

The parser records each region's role, heading where available, size, and interactive elements. BM25 indexes source-node text. The overview combines structural information and goal-relevant snippets with action feedback. The nominal overview budget does not bound the entire prompt, which also includes history and model-specific instructions.

**`ReadBlock`(region).** Returns the next chunk of at most 640 tokens from a named region. A cursor advances through the source; an exhausted region returns explicit exhaustion feedback instead of silently repeating its contents.

**`Lookup`(query).** Returns up to six unseen BM25-matching nodes from the retained snapshot. This is local retrieval, not submitting a search query to a website. A website search that obtains a new observation is an environmental action and must be treated according to its intent.

**`Inspect`(element).** Reveals ancestors and the local subtree of an exposed element, to depth two. It can resolve ambiguity in labels or local structure without exposing the whole observation.

**`ObserveAll`.** Reveals the full current source. It can be explicitly requested or invoked by read-limit fallback and action validation. It may exceed the targeted-reveal budget, but cannot reveal evidence absent from the snapshot. Local exhaustion therefore does not certify task sufficiency.

## A.3 Control, planning, and validation

The semantic controller interface has three outcomes: a `READ` request, an `ACT` evidence need, or a `COMMIT` handoff. The intent supplied to the planner distinguishes the latter two; the environment's physical action space is unchanged. The controller can proceed immediately from the initial overview. A malformed control response permits one repair attempt. Targeted reading is bounded; fallback resolves an external intent after full-source exposure rather than looping indefinitely.

Click, type, and hover actions must reference exposed, valid identifiers. One full-reveal re-planning attempt is allowed for an invalid target. Continued invalidity yields a controlled no-op with feedback.

The implementation specification scores `STOP` among action candidates and does not add a separate completion verifier. That terminal action is distinct from `COMMIT`, which is an internal handoff and need not end the episode.

Table 4: **Implementation settings.**

| Setting | Value |
|---|---|
| Overview budget | 800 tokens, including 320 goal-relevant tokens |
| Targeted-reveal budget | 3,840 tokens; full reveal may exceed it |
| Region chunk / read limit | 640 tokens / 3 targeted rounds |
| Lookup / local inspection | BM25, top-6 unseen nodes / depth 2 |
| Proposal samples / retained candidates | 20 / at most 3 distinct candidates by vote |
| Value samples per candidate | 20, averaged across valid scores |
| Environmental interaction budget | 15 steps per episode |
| Policy / controller | GPT-4o mini |
| World / value services | Llama-3.1-8B-Instruct with WMA adapters |
| Completion verifier | No separate verifier |

# B Experimental Details and Complete Aggregates

All quantitative results in this paper derive from the full 812-task WebArena benchmark. Overall task success is computed across all 812 tasks; evidence reduction, token usage, and environmental-action intent are computed over successful tasks. Unless otherwise stated, point estimates are means across repeated runs, and bracketed ranges give the observed minimum and maximum across runs rather than confidence intervals. We keep these scopes explicit rather than combining them into a synthetic global percentage.

Table 5: **Source-referenced evidence reduction in successful LBE tasks.** Final planning evidence is compared with the corresponding untruncated local source observation among successful LBE tasks from the 812-task benchmark.

| Metric | Result |
|---|---|
| Benchmark task pool | 812 |
| Conditioning scope | Successful LBE tasks |
| Final planning-evidence reduction vs. untruncated source | 54.6% |

## B.1 Successful-task conditional summaries

The following tables preserve the separately reported statistics conditioned on successful tasks from the same 812-task benchmark. They are conditional summaries rather than paired-task comparisons, so the means within each completed-step range are reported directly instead of being collapsed into one weighted percentage across all successful tasks.

Table 6: **Working-belief savings relative to WMA's context-capped full-observation inputs in successful LBE tasks.** The 52.8% aggregate and the 54.6% source-referenced reduction use the same successful-task scope but different reference denominators.

| Completion steps | Saving (%) | Reported range (%) |
|---|---|---|
| $\leq 5$ steps | 58.6 | [56.1, 60.7] |
| 6–10 steps | 51.1 | [49.2, 52.8] |
| 11–15 steps | 47.3 | [44.6, 49.0] |
| Aggregate | 52.8 | [49.4, 55.2] |

The reductions are descriptive ratios of the means reported within each completed-step range, $100(1 - \bar{C}_{\mathrm{LBE}}/\bar{C}_{\mathrm{WMA}})$, not paired-task effects. FocusAgent has lower reported means than LBE but uses a direct-action architecture without world/value models. The conditional totals are not split into

Table 7: **Mean input-plus-output usage by steps to successful completion.** Units are tokens, and brackets give the observed minimum and maximum across repeated runs. Because systems can succeed on different tasks within the same step range, these means are descriptive rather than paired-task effects.

| System | $\leq 5$ steps | 6–10 steps | 11–15 steps |
|---|---|---|---|
| WMA | 72,663 | 236,177 | 350,606 |
| | [63,652, 78,719] | [221,804, 245,650] | [310,778, 404,276] |
| LBE | 56,367 | 198,336 | 334,206 |
| | [50,103, 62,386] | [188,425, 217,336] | [300,872, 380,981] |
| FocusAgent | 37,725 | 92,510 | 207,932 |
| | [35,760, 39,842] | [88,737, 98,612] | [189,933, 233,490] |
| ReadAgent | 311,628 | 456,921 | 757,734 |
| | [300,447, 320,769] | [423,876, 477,008] | [703,904, 821,894] |
| LBE reduction vs. WMA (%) | 22.4 | 16.0 | 4.7 |

input and output in this record and do not include unsuccessful tasks by definition. No confidence interval is constructed for these ratios from the marginal reported bounds.

### B.2 Environmental action intent

Table 8: **Environmental action intent in successful LBE tasks.** `READ` is excluded because it does not advance environmental time.

| Measure | Proportion (%) | Reported range (%) |
|---|---|---|
| Information-seeking environmental action / `ACT` | 71.3 | [63.2, 74.5] |
| Task-directed environmental action / `COMMIT` | 28.7 | Complement of above |

The two categories are complementary: 28.7% is $100 - 71.3\%$, and its paired endpoints are likewise determined by the information-seeking proportion. We therefore report the execution share as the complement rather than printing the paired endpoints in reverse numerical order.

### B.3 Aggregation and reproducibility details

All reported experimental statistics refer to the same 812-task WebArena benchmark. The success table summarizes all tasks, whereas all remaining statistics condition on successful tasks. The 54.6% source-referenced reduction and the 52.8% WMA-referenced reduction both condition on successful LBE tasks and differ in their reference denominators. The token means describe tasks successfully completed in at most 5, 6–10, and 11–15 steps, and the 71.3% / 28.7% intent split likewise conditions on successful LBE tasks.

The 22.4%, 16.0%, and 4.7% usage reductions are descriptive ratios of the means reported within each completed-step range, $100(1-\bar{C}_{\mathrm{LBE}}/\bar{C}_{\mathrm{WMA}})$, not paired-task effects. Because the available aggregate record does not provide the number of successful tasks within each range needed for weighting, we do not construct a single input-plus-output saving across all successful tasks. Bracketed ranges give the observed minimum and maximum across repeated runs and are not confidence intervals.

Task 132 is retained as a qualitative trajectory example rather than a population statistic. Its purpose is to visualize a behavior that aggregate compression cannot show directly: the working belief can expand when the next decision requires more specific evidence even if the local source becomes shorter.

# C Ablation Studies

## C.1 Ablation configurations

We compare the complete LBE mechanism with four variants that alter one control component at a time while retaining the same predictive planner, models, and environmental action space. Table 9 summarizes the resulting interfaces. The targeted-read limit is three in both LBE and Fixed-READ. In the complete method, the controller may stop reading before reaching this limit; if it requests further evidence after three targeted reads, the full current source is revealed. Fixed-READ instead requires three targeted reads before an external decision and uses the same full-reveal fallback if further source evidence is still requested.

Table 9: **Ablation configurations.** Compact initialization denotes the same programmatically constructed initial working belief used by the complete method.

| Variant | Initial evidence | Within-source reading | Planner intent |
|---|---|---|---|
| LBE | Compact working belief | Adaptive, up to three targeted reads, with full-reveal fallback | Separate `ACT` and `COMMIT` intents |
| No-READ | Same compact working belief | Disabled | Separate `ACT` and `COMMIT` intents |
| Full-Init | Context-capped full observation | Retained for source evidence omitted by the input limit | Separate `ACT` and `COMMIT` intents |
| No-Intent | Same compact working belief | Same adaptive reading as LBE | One merged environmental-planning intent |
| Fixed-READ | Same compact working belief | Three targeted reads before an external decision, with full-reveal fallback | Separate `ACT` and `COMMIT` intents |

No-READ isolates the contribution of recoverable within-source expansion. Full-Init tests compact initialization against exposing the context-capped full observation at the first model call while preserving later source access. No-Intent tests whether explicitly distinguishing evidence-seeking from task-directed environmental decisions contributes beyond belief expansion alone. Fixed-READ retains the same targeted-read allowance and fallback as LBE but removes the controller's ability to decide when enough source evidence has been exposed.

## C.2 Evaluation protocol

All five configurations use the same 812-task WebArena benchmark, model services, predictive planner, environmental action space, and 15-step interaction budget. Overall success is summarized across all tasks. The remaining ablation statistics condition on successful trajectories and are categorized post hoc by realized completion length into at most 5, 6–10, and 11–15 steps, following the reporting convention in Section 4. Inference cost is the total input-plus-output token usage of the complete inference pipeline. Each entry reports the mean across repeated experiments followed by the observed minimum and maximum in brackets; these ranges are not confidence intervals.

Overall success and the completion-range success counts were summarized separately across the repeated experiments. We therefore report both sets of supplied aggregates directly and do not reconstruct the overall success estimate by summing the separately aggregated range counts. Likewise, because the configurations can succeed on different tasks within a completion range, token-cost comparisons are descriptive ratios of reported means rather than paired-task effects.

## C.3 Results and analysis

Table 10 reports overall task completion. Full-Init and Fixed-READ remain close to complete LBE in point estimate, and their reported ranges overlap that of LBE. Removing within-source reading produces the largest decrease, from 19.33% to 13.42%, while merging the two environmental intents yields 16.63%. These results are consistent with recoverable within-source expansion and explicit intent separation contributing to task completion, without establishing that the small differences among the overlapping LBE, Full-Init, and Fixed-READ ranges are statistically reliable.

Table 10: **Overall success in the ablation study.** Values are percentages over the full 812-task benchmark. Brackets contain the observed minimum and maximum across repeated experiments.

| Configuration | Success (%) [range] |
|---|---|
| LBE | 19.33 [18.10, 20.69] |
| No-READ | 13.42 [12.81, 13.67] |
| Full-Init | 18.97 [17.24, 19.46] |
| No-Intent | 16.63 [15.76, 16.87] |
| Fixed-READ | 19.58 [17.73, 20.07] |

Table 11 reports end-to-end inference cost among successful trajectories. Relative to complete LBE, Full-Init uses 30.9%, 21.0%, and 1.8% more tokens across the three completion ranges, respectively. Fixed-READ incurs the largest cost, using 86.2%, 83.0%, and 84.6% more tokens than LBE. Because the targeted-read limit and full-reveal fallback are identical in these two configurations, this gap is consistent with the additional cost of requiring all three targeted reads rather than allowing the controller to stop when its current evidence is sufficient; the non-paired aggregates do not identify an exact causal increment. No-Intent is also more costly than LBE in all three ranges, by 4.6%, 1.6%, and 8.5%. No-READ has lower conditional token means, but it also has substantially lower success and can succeed on a different, potentially easier subset of tasks; its lower cost therefore should not be interpreted as superior end-to-end efficiency.

Table 11: **Mean input-plus-output tokens by steps to successful completion in the ablation study.** Brackets contain the observed minimum and maximum across repeated experiments.

| Configuration | $\leq 5$ steps | 6–10 steps | 11–15 steps |
|---|---|---|---|
| LBE | 56,367 [50,103, 62,386] | 198,336 [188,425, 217,336] | 334,206 [300,872, 380,981] |
| No-READ | 39,157 [36,671, 42,105] | 175,120 [162,387, 189,044] | 311,819 [296,873, 331,346] |
| Full-Init | 73,809 [64,990, 78,726] | 240,029 [231,572, 246,007] | 340,384 [312,869, 401,718] |
| No-Intent | 58,966 [51,224, 62,405] | 201,499 [192,947, 219,001] | 362,553 [323,448, 400,153] |
| Fixed-READ | 104,945 [98,233, 110,273] | 363,014 [349,298, 378,604] | 617,075 [599,237, 646,073] |

For context, Table 12 gives the separately summarized numbers of successful trajectories in each realized completion range. The distributions reinforce the main success result: disabling `READ` or removing the explicit intent distinction changes not only conditional inference cost but also which and how many trajectories reach successful completion. The counts are therefore reported alongside, rather than used to normalize or pool, the conditional token means.

Table 12: **Successful trajectories by realized completion length.** Entries are the separately reported mean counts across repeated experiments, with observed minima and maxima in brackets.

| Configuration | $\leq 5$ steps | 6–10 steps | 11–15 steps |
|---|---|---|---|
| No-READ | 71 [66, 74] | 32 [27, 35] | 6 [2, 8] |
| Full-Init | 98 [83, 106] | 35 [30, 41] | 30 [22, 33] |
| No-Intent | 82 [71, 89] | 32 [29, 39] | 11 [5, 17] |
| Fixed-READ | 101 [88, 110] | 38 [30, 45] | 28 [22, 34] |

Taken together, the ablations separate two aspects of the method. Compact initialization primarily reduces the cost of presenting the source before model-based planning, while adaptive reading avoids paying for a fixed amount of evidence expansion on every decision. Recoverable access remains important for completion: removing it yields the lowest success estimate, even though the surviving successful trajectories have lower conditional token totals. Explicitly separating evidence-seeking

`ACT` from task-directed `COMMIT` also improves the observed success estimate relative to the merged-intent variant. The complete method therefore provides the best balance in these aggregates: success close to the strongest ablations, substantially lower inference cost than Fixed-READ, and lower cost than Full-Init in every reported completion range.