Assumption Harvesting: A Pilot Workshop for Students' Reflective Use of General-Purpose LLMs

Frank Cheng

Independent Researcher, Fukuoka, Japan

Abstract

Students increasingly use general-purpose large language models (LLMs) for cognitively demanding tasks, yet fluent outputs can still miss what users actually mean to ask. This paper introduces assumption harvesting as a reflective practice for surfacing hidden assumptions, reconstructing missing conditions, and comparing alternative framings before committing to an answer. We present a 7-hour pilot workshop delivered across two sessions to two small cohorts in Taipei (N=8), combining a lightweight account of LLM behavior with scaffolded practices for diagnostic reading, meaning reconstruction, reflective interaction, and optional customization. Evidence comes from structured pre-course interviews, a short entry exercise, workshop implementation, and voluntary post-workshop reflections. The pilot suggests that participants were already frequent LLM users but often lacked a stable framework for diagnosing output–intent mismatch, and that the workshop was feasible across mixed disciplinary backgrounds. Post-workshop reflections suggest a perceived shift toward more deliberate, assumption-aware, and reflective use. The contribution of this paper is modest: it offers a pilotable intervention and illustrative pilot evidence, not validated learning gains, causal efficacy, or superiority over other pedagogical approaches. The paper focuses on students' reflective use of general-purpose LLMs rather than educational AI systems or tutoring agents.

Keywords

Large language models, Human-LLM interaction, AI literacy, Metacognition, Reflective learning, Learner agency, Assumption harvesting

1. Introduction

1.1. When fluent outputs miss the real question

Students increasingly use general-purpose LLMs for cognitively demanding tasks such as drafting, planning, research framing, coding, and decision support. Yet a common failure is not simply that an answer is factually wrong. More often, the output sounds plausible while still failing to address what the user actually meant to ask. Recent HCI work argues that generative AI places substantial metacognitive demands on users, especially around prompting, evaluating outputs, calibrating confidence, and deciding how to incorporate AI into ongoing work [1]. Related work further shows that specific features of LLM responses—such as uncertainty expressions, explanations, and inconsistencies—can shape user reliance and trust in consequential ways [2] [3]. Conventional advice to "prompt better" is useful, but it does not by itself explain why fluent output–intent mismatch remains common in practice.

This grounded problem was visible in our pilot context before the workshop began. In pre-course interviews, participants described using mainstream LLMs for cognitively demanding tasks, academically, professionally, and personally. At the same time, they often described outputs as too general, off-target, different from their own understanding, or insufficiently tailored to the user's context.

1.2. What this paper is

This paper presents a pilot workshop for students' reflective use of general-purpose LLMs in cognitively demanding tasks. Its pedagogical core is a practice we call **assumption harvesting**: a reflective way

HAI-Agency: Workshop on Orchestrating Human and AI Agency for Proactive and Reflective Learning, co-located with the 27th International Conference on Artificial Intelligence in Education (AIED 2026), Seoul, Korea, June 27 - July 3, 2026.

frank.cheng.research@gmail.com (F. Cheng)

0009-0007-6554-9415 (F. Cheng)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-08-27

CEUR-WS.org/Vol-4248/short13.pdf

of engaging LLM outputs by surfacing the assumptions they rely on, reconstructing missing conditions, and comparing alternative framings before committing to an answer. The workshop was designed for students already using widely available LLM systems in ordinary study, work, and life contexts, rather than for specialized educational AI products or tutoring agents. In this respect, the paper speaks to a learner-side capability that is relevant to broader discussions of how people adapt cognitively and behaviorally to increasingly capable AI systems [4].

1.3. Contribution

This paper is organized around two research questions:

RQ1 What learner needs and skill gaps are visible among students already using general-purpose LLMs for cognitively demanding tasks?

RQ2 How can assumption harvesting be operationalized as a teachable workshop for students across mixed disciplinary backgrounds, and what does initial pilot evidence suggest about its feasibility and perceived effects?

In answering these questions, the paper makes three contributions. First, it offers a **practical rationale** for teaching assumption harvesting as a response to the recurring problem of output–intent mismatch. Second, it presents a **pilotable workshop design** for teaching reflective LLM use in a small-group format. Third, it reports **illustrative pilot evidence** suggesting learner need, design feasibility, and a perceived shift toward more deliberate and assumption-aware use. The paper does not claim validated learning gains, causal efficacy, or superiority over other pedagogies.

2. Why Assumption Harvesting Is the Teaching Choice

2.1. Why fluent outputs are hard to diagnose

A central challenge in everyday LLM use is that a response can appear adequate before its fit has been examined. The problem is often not an obviously false answer, but a fluent and locally reasonable one that still misses the task the user is actually trying to solve. Users can see the response, but not necessarily the assumptions, constraints, or success criteria under which it would make sense.

Prior work on reliance and trust helps explain part of this difficulty. Users do not respond only to correctness; they also respond to presentation. The wording of uncertainty, the presence of explanations, and the availability of sources can all shape whether an answer feels reliable and whether users continue to interrogate it [5] [2] [3].

Our workshop frames this problem in practical terms. Language never carries all of the assumptions and conditions needed for perfect interpretation. Human communication works by leaving some context implicit and relying on the listener to reconstruct what was left unsaid. LLM outputs inherit this same limitation, but without access to the richer interpersonal and situational cues that often help people repair ambiguity in ordinary conversation. As a result, an answer can sound coherent while still resting on assumptions, abstractions, or omitted conditions that do not match the user's actual situation.

This is why advice to simply "prompt better" is incomplete. Better prompting can help, but it does not by itself teach learners how to read a plausible output diagnostically once it appears. If a student cannot tell why an answer feels off, they may respond by retrying, adding more detail, or changing models without understanding what caused the mismatch in the first place. The workshop therefore begins from a more basic need: helping learners recognize that a fluent answer may still be conditional, partial, or misaligned.

2.2. Assumption harvesting as a practical reading stance

We use the term **assumption harvesting** for a reflective reading-and-interaction practice that addresses this problem. Instead of treating an LLM output as a self-sufficient answer, learners treat it as a response

built on assumptions that may or may not fit their actual task. The goal is to make those assumptions more visible, reconstruct the missing conditions that would make the answer appropriate, and compare alternative framings before deciding how much to accept, revise, or reject.

In this paper, assumption harvesting is not presented as a formal method or a single prompt template. It is a practical stance learners can adopt while reading and continuing an interaction. In concrete terms, it means asking questions such as: What is this answer optimizing for? What must be true for this recommendation to make sense? What constraints or values does it appear to assume? These questions do not aim to unpack every possible implication of an output. They aim to surface enough of the hidden frame to improve fit for the task at hand.

This stance also helps organize the workshop's teaching choices. Recent work suggests that learners often struggle to critically evaluate LLM outputs on their own, but become more focused and methodical when given structured guidance or explicit interaction scaffolds [6] [7]. In that sense, assumption harvesting can be understood as one learner-side practice for helping students adapt cognitively and behaviorally to increasingly capable AI systems without reducing the problem to better prompting alone [4].

3. Pilot Workshop Design

3.1. Pilot context and pre-course interview

The workshop was piloted in two small cohorts in Taipei, with four participants per cohort, for a total of eight participants. The pilot targeted college seniors, master's students, and early-career professionals using LLMs in demanding cognitive tasks rather than in a narrowly technical setting. Participants came from mixed disciplinary backgrounds, including information management, national development, industrial engineering, accounting, computer science, business, and communication/media, and several were already using mainstream LLMs regularly in study, work, and everyday problem solving.

Before attending, each participant completed a structured interview using the same form (see Appendix A). The interview served three practical roles: readiness screening for the pilot format, a light check for conceptual gaps in the workshop design, and participant priming through a short exercise on identifying assumptions in model output.

3.2. Learning goals

The workshop was designed around four concrete learning goals. First, students should be able to understand why a plausible or well-written output may still be a poor fit for their actual question. Second, they should be able to surface assumptions and missing conditions that shape an answer's relevance. Third, they should be able to compare alternative framings before prematurely committing to a single response. Fourth, they should leave with more reflective ways of interacting with and verifying LLM outputs across real tasks.

3.3. Foundations for interpreting LLM outputs

The workshop began with a lightweight account of how LLM outputs are produced, introducing a simplified account of tokenization, next-token continuation, the role of the context window, and the broad distinction between pretraining and alignment as the basic mechanism underlying LLM output generation (see Appendix B). It also emphasized several limits of that mechanism: outputs need not be true and may emulate reasoning without guaranteeing formal logical reliability. The purpose of this section was to give students enough mechanism to understand why fluent responses can still be partial, context-dependent, or misleading.

A second part focused on why fluency can mislead. The workshop explicitly introduced language compression, uncertainty compression, statistical averaging, and context steering. In practical terms,

Table 1
Condensed version of a workshop slide showing the alignment routine. Students were asked to surface the assumptions shaping the output and align those assumptions through iterative follow-up interaction.

	Step	Example Question
1.	Audit : Ask the model	What assumptions does this answer rely on?
2.	Scan : User think	Which assumptions do not fit my situation?
3.	Clarify : Ask the model	Ask the model to explain assumptions you do not fully understand.
4.	Resolve : User think	Answer the questions that mean the most to you.

students were taught that outputs often compress goals, tradeoffs, and uncertainty into smooth wording, and that under-specified prompts invite the model to fill in typical assumptions (see Language Compression under Appendix C). This conceptual groundwork prepared students to see later exercises not as "prompt tricks," but as responses to recurring interpretive limits in LLM interaction.

3.4. Teaching assumption harvesting in practice

The core of the workshop operationalized assumption harvesting as a set of teachable moves. Students were introduced to **diagnostic reading**, which asks them to notice where a response appears to rely on hidden goals, criteria, constraints, values, or baselines. They were then introduced to **meaning reconstruction**, which treats the initial output as something that can be interrogated with follow-up questions such as what assumptions were made and which constraints are being presumed.

Building on diagnostic reading and meaning reconstruction, the workshop introduces an operable stance, the **exploration-alignment loop**. **Exploration** means to expose the model's default assumptions and alternative framings; **alignment** means selectively constraining the interaction once the learner had clarified what mattered in the actual situation (see Table 1). These two concepts work together in a loop: surface assumptions, compare them to the learner's context, correct what does not fit, and continue until further clarification is unlikely to change the next decision. During the workshop students were asked to apply this process to an individual topic, identify key assumptions, note unresolved ones, and reflect on what would make the advice wrong.

Although assumption harvesting is taught as a general reflective stance, the kinds of assumptions varied across tasks. In coding tasks, this includes clarifying requirements, constraints, and failure conditions. In writing tasks, it can involve audience, purpose, and omitted context. In life or career questions, relevant assumptions often involve priorities, tradeoffs, time horizon, and the outcome the user is actually trying to optimize.

3.5. Teaching reflective interaction tools

Next, the workshop expands into a set of reflective interaction tools. **Branching** was taught as a way to separate alternative continuations rather than letting one conversational path silently accumulate as the default frame. The goal was to preserve distinct lines of reasoning long enough to compare them before convergence. **Adversarial critique** asked students to use either the same model or multiple models to argue against a current answer, surfacing weaknesses and missing considerations. **Frame correction** addressed cases where the interaction had become anchored on an unhelpful assumption and needed to be reset from an earlier turn. Together, these tools made the workshop's core stance actionable across longer interactions.

3.6. Optional customization scaffold

The final part of the workshop introduced custom instructions as an optional scaffold for stabilizing reflective interaction habits across sessions. In the workshop's framing, custom instructions matter because default model behavior is designed for broad usability rather than for any particular user's reasoning standards, values, or tolerance for ambiguity. Customization was introduced as one possible

way to preserve desirable model behaviors such as automatically surfacing assumptions, resisting automatic agreement, or asking clarifying questions when they mattered.

In addition to the workshop sessions, participants received a take-home reference sheet that condensed the core practices into a portable aid for daily use after the workshop (see Appendix C).

4. Illustrative Pilot Evidence

4.1. Baseline learner need

Pre-course interviews suggest that participants were already frequent users of general-purpose LLMs for research, coding, writing, planning, and open-ended life or career questions. The workshop therefore entered a context in which participants were already relying on these tools across study, work, and everyday reasoning.

At the same time, this reliance was not matched by a stable framework for interpreting outputs. Participants typically had only partial mental models of how LLMs work, and their verification practices were uneven: some checked sources, compared models, or switched to search tools when stakes were high, while others relied more on plausibility or responded to mismatch by retrying with more detail. Notably, none described systematically interrogating outputs for the assumptions or omitted conditions they rested on.

Taken together, these interviews suggest that the workshop appeared well matched to a real learner need: participants wanted to use LLMs more effectively, but lacked a stable way to diagnose why a plausible output did not quite fit, or what to do next once that happened. This interpretation is broadly consistent with recent work describing GenAI use as placing substantial metacognitive demands on users, especially when they must evaluate off-the-shelf systems without much scaffolding [1] [8].

4.2. What the entry exercise revealed

The structured entry exercise in the pre-course interview offered a narrower view of what participants could already do once assumptions were made available to them. In that exercise, participants first reacted to a generated output, then saw a follow-up prompt asking the model to list the assumptions it had made, and finally considered what questions would make the output better. This format revealed a recurring pattern: once assumptions were surfaced for them, participants were often able to recognize which ones were fragile, high-stakes, or likely to change the recommendation. These included assumptions about target audience, unstated goals, operational constraints, and budget or feasibility conditions.

What appeared weaker was the prior step. As the verification practices described in Section 4.1 indicate, participants did not, in their own accounts, surface these assumptions on their own — yet the entry exercise shows they could readily evaluate them once surfaced. This contrast matters because the workshop's central teaching choice is not only to help students judge assumptions once visible, but to help them make those assumptions visible in the first place.

4.3. Feasibility observations and perceived shift

At the level of workshop feasibility, the pilot suggests that the format was workable across mixed disciplines. Participants from technical and non-technical fields were able to engage with the material. Voluntary post-workshop reflections also suggest a perceived shift in how some participants understood and used LLMs. Several described moving from prompt-writing toward more strategic thinking, from taking fluent outputs at face value toward more skeptical reading, and from seeing model outputs as answers to seeing them as results built on assumptions. These reflections should be read cautiously: they are self-reports from a small pilot, not validated measures of skill change. Still, they align with the workshop's intended trajectory around fluency, assumption-awareness, verification, and more deliberate interaction, suggesting that the format was feasible and worth further development.

5. Discussion

5.1. What this pilot suggests

This pilot suggests that reflective LLM use may be teachable as a practice. The workshop organized a set of learner-side moves for noticing when an output misses the real question, surfacing the assumptions that make it seem reasonable, and continuing the interaction more deliberately.

A more immediate implication concerns output–intent mismatch. Rather than helping students overcome every LLM failure mode, the workshop may have helped some participants better diagnose why a plausible answer still did not fit the problem they were trying to solve.

5.2. Light conference bridge

While this paper does not study educational AI systems directly, it speaks to a learner-side capability relevant to broader discussions of human–AI agency: whether students can engage LLM outputs reflectively rather than passively. This is consistent with work on bidirectional human–AI alignment, which includes not only aligning systems to people but also how people cognitively and behaviorally adapt to AI in use [4]. For this workshop context, the modest implication is that teaching students to engage general-purpose LLM outputs reflectively is a worthwhile complement to system-side design.

6. Limitations and Next Steps

6.1. Pilot limitations

Several limitations constrain what can be concluded from this pilot. First, the workshop was tested in two very small cohorts and should be understood as an early pilot rather than a broader evaluation. Second, the post-workshop evidence consists of voluntary reflections and feedback, which are useful for understanding perceived usefulness and conceptual resonance but do not provide validated measures of learning or durable behavior change. Third, the evidence reported in this paper is therefore primarily descriptive and self-reported, not a test of causal efficacy, comparative advantage, or validated skill gain.

Additional limits follow from the pilot format itself. The workshop was delivered by its designer, so instructor and researcher roles were not separated. The participants were also relatively self-selected, which may have increased both engagement and receptivity. The pre-course interview functioned as readiness screening and conceptual priming, which limits how cleanly later perceived shifts can be attributed to the workshop alone. Finally, the scope of this pilot is limited to students' reflective use of general-purpose LLMs in cognitively demanding tasks; it does not establish claims about educational AI systems, tutoring agents, or learner populations beyond this early context.

6.2. Next steps

The next iteration of this work should move from pilot legibility toward more disciplined evaluation. A first priority is to define clearer learning indicators tied to the workshop's intended outcomes. Future iterations could assess the participants' pre- and post-workshop LLM interaction behavior on metrics such as rate and depth of assumption surfacing as well as alternative framing comparison. A control condition would further allow attribution of observed changes to the intervention rather than to general LLM experience over the same period.

A second priority is to test more scalable delivery formats: the current two 3.5-hour sessions were a deliberately condensed pilot structure. If the workshop is to scale, a more bite-sized structure distributed across more days or over a longer period may prove more practical for attention, scheduling, and retention than the compressed pilot format. Future iterations should also examine which parts of the workshop transfer well to larger or more heterogeneous cohorts, and which depend on high-touch facilitation to remain effective.

7. Conclusion

As students increasingly use general-purpose LLMs for analysis, planning, writing, and decision support, a central difficulty is no longer only whether an output is correct, but whether users can tell when a fluent answer is solving the wrong problem. This paper presented assumption harvesting as a teachable response to that difficulty and described a pilot workshop designed to help students make hidden assumptions visible, reconstruct missing conditions, and compare alternative framings before acting on or accepting an answer.

This paper's contribution is to make that capability concrete as a pilotable intervention. The pilot offers illustrative evidence that the underlying learner need is real, that the workshop structure is workable in a small-cohort format, and that the targeted skills are worth developing further. These claims remain modest: the paper does not establish validated learning gains or causal efficacy, but it suggests that teaching students how to interrogate an answer may be a worthwhile complement to teaching them how to generate one.

Declaration on Generative AI

During the preparation of this work, the author used GPT-5.4 and Claude Opus 4.7 in order to: Grammar and spelling check, Paraphrase and reword. After using these tools/services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, S. Rintel, The metacognitive demands and opportunities of generative ai, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–24.

[2] S. S. Kim, Q. V. Liao, M. Vorvoreanu, S. Ballard, J. W. Vaughan, " i'm not sure, but...": Examining the impact of large language models' uncertainty expression on user reliance and trust, in: Proceedings of the 2024 ACM conference on fairness, accountability, and transparency, 2024, pp. 822–835.

[3] S. S. Kim, J. W. Vaughan, Q. V. Liao, T. Lombrozo, O. Russakovsky, Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–19.

[4] H. Shen, T. Knearem, R. Ghosh, K. Alkiek, K. Krishna, Y. Liu, Z. Ma, S. Petridis, Y.-H. Peng, L. Qiwei, et al., Towards bidirectional human-ai alignment: A systematic review for clarifications, framework, and future directions, arXiv preprint arXiv:2406.09264 2406 (2024) 1–56.

[5] G. Bansal, T. Wu, J. Zhou, R. Fok, B. Nushi, E. Kamar, M. T. Ribeiro, D. Weld, Does the whole exceed its parts? the effect of ai explanations on complementary team performance, in: Proceedings of the 2021 CHI conference on human factors in computing systems, 2021, pp. 1–16.

[6] S. Prabhudesai, A. P. Kasi, A. Mansingh, A. Das Antar, H. Shen, N. Banovic, " here the gpt made a choice, and every choice can be biased": How students critically engage with llms through end-user auditing activity, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 2025, pp. 1–23.

[7] O. Clerc, R. Abdelghani, C. Desvaux, E. Poisson, P.-Y. Oudeyer, H. Sauzéon, Teaching students to question the machine: An ai literacy intervention improves students' regulation of llm use in a science task, arXiv preprint arXiv:2604.01955 (2026).

[8] S. H. Ramesh, F. Daneshzand, B. Rashidi, S. Raj, H. Subramonyam, F. Rajabiyazdi, Metacognitive demands and strategies while using off-the-shelf ai conversational agents for health information seeking, in: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, 2026, pp. 1–16.

A. Interview Questions

Pre-workshop interview questions used for readiness screening and conceptual priming.

1 What do you usually use LLM for? At school, at work, and outside of school and work?
2 Explain to me how LLM works as well as you can. Doesn't have to be technical. How does the LLM decide what to say next?
3 What do you think about when LLM gives answer that seems wrong or off? Why do you think it happens?
4 What do you do when you notice the output isn't what you want or expect? Do you ask the model to explain its response?
5 Do you remember any instance when LLM changed your mind about something? What happened?
6 Mini exercise on assumption surfacing.
    6.1 Prompt LLM: Help write a project or startup proposal over [topic of candidate's choice] in 10-15 sentences.
    - Ask the interviewee: "What's your first reaction to this output?"
    6.2 Prompt LLM: List the assumptions you made to decide
        * what to emphasize
        * what to omit
        * what tone to use
    - Ask the interviewee: "Which of these assumptions feels the most fragile or dangerous to you?"
    - Ask the interviewee: "Which one do you think is most often left unstated in planning like this?"
    - Ask the interviewee: "What would change the recommendation entirely?"
    6.3 Prompt LLM: What questions could I answer to make the output better?
    - Ask the interviewee: "Before I ask the model, what questions do you think it should ask in order to understand what you care about?"
    - Ask the interviewee: "Which of these questions matters the most, and why?" "Let's answer those."
    6.4 Prompt LLM: [Interview's answers from 6.3]. Draft proposal revision.
7 What LLM skills do you want to learn if you could?

B. Lightweight Model of LLM and Teaching Rationale

A lightweight model of LLM designed to be as accessible as possible yet conveys the necessary foundation to understand assumption harvesting as a stance.

- **Tokenizer**: A trained component that compresses recurring character sequences into tokens; varies across languages and tokenizers; example showing how "Internationalization" splits into ["Inter", "national", "ization"].
- Establishes that the model does not see "words" the way a human reader does; sets up later discussion of how meaning is reconstructed from statistical patterns rather than dictionary lookup.
- **Pretraining, SFT, RLHF**: Three training phases: pretraining for general language and world knowledge, supervised fine-tuning for instruction-following, RLHF for human-preference alignment.
- Distinguishes "what the model knows" from "how the model has learned to present what it knows"; introduces the gap between optimization target and truth that Chapter 2's fluency-misleads framing depends on.

- **RLHF pitfalls**: RLHF systematically rewards what is easy to judge (politeness, confidence, fluency, tone) over what is difficult to judge (subtle correctness, long-term consequences, hidden assumptions, calibrated uncertainty).
  - Provides the mechanistic explanation for why fluent output may be poorly calibrated; this is the root cause learners are equipped to recognize in Chapter 2.
- **Context window**: The full input the model sees at inference: system instructions, custom instructions, memory, prior turns, current turn, and any retrieved content; everything is tokens to the model.
  - Sets up later discussion of context steering, attention drift, and the rationale for context window migration as an interaction tool.
- **Embedding and vector space**: Words are not fixed definitions but locations in a high-dimensional latent space, with meaning expressed as relationships between locations.
  - Makes "statistical average" literal rather than metaphorical: when context is sparse, the model's continuation tends toward the centroid of relevant regions, which corresponds to no specific reader.
- **Self-attention**: The model evaluates each word's meaning by reference to other words in the input; "bank" near "slippery" activates [river], near "robbed" activates [finance].
  - Shows learners how context shapes interpretation at the mechanism level, which makes the later "context steering" advice concrete rather than abstract.
- **Logits and selection**: Every possible token receives a similarity score; one of the highest-scoring tokens is selected, with some randomness; competing candidates may be statistically very close to the chosen one.
  - Direct foundation for "uncertainty compression" in Chapter 2: the user sees one continuation, but the next-most-likely options may have been within a few percentage points.

C. Take-Home Reference Sheet

A condensed adaptation of the take-home reference sheet, highlighting the workshop's core portable reflective practices rather than reproducing the full participant handout.

Language Compression

- **Evaluative Abstraction**: Compressing value hierarchies and implicit thresholds.
  Reasonable. Effective. Best Practice
- **Domain Abstraction**: Collapsing clusters of technical criteria.
  Scalable. Robust. Modular
- **Framing Structures**: Shaping what appears comparable or important.
  Tradeoffs. Alternatives. Expression of Time
- **Uncertainty Compression**
  Tends to. Generally. Often. In many cases. Can be seen as. Is typical. It depends.

Assumption Harvesting

- **Exploration**
  Uncover the frame the current problem space operates under and explore whether the frame can be lifted or changed in order to expand the problem space and allow different solutions.
- **Alignment**
  Surfacing the underlying assumptions held by the model and the user, exposing their incompatibilities, and aligning their contextual landscape.

Alignment Steps

1. **Audit**: Ask the model
   - What assumptions does this answer rely on?

- Which of those would affect the answer the most if wrong?
- What questions could I answer to improve the output the most?

2 **Scan**: User think
- Which assumptions do not fit my situation?
- Which would change the answer the most if corrected?
- Which questions should be answered now? Which can wait?

3 **Clarify**: Ask the model
- Ask the model to explain assumptions you do not fully understand.
- For assumptions that differ from yours, ask the model to compare outcomes between them

4 **Resolve**: User think
- Answer the questions that mean the most to you.
- Correct assumptions that do not fit your situation.
- Re-run the reasoning with the corrected frame

Where Assumption Commonly Hide

- **Goal and Success Criteria**: What "better", "effective", or "optimal" means
- **Constraints**: Time, budget, skills, legality, risk tolerance
- **Context and Baselines**: What's typical, default, or assumed common
- **Values and Tradeoffs**: What is optimized vs sacrificed
- **Audience and Perspective**: Who this is for, who is affected
- **Causality and Dynamics**: Linear vs systemic effects, second-order consequences
- **Scope and Time Horizon**: Short-term vs long-term, local vs global
- **Uncertainty and Confidence**: What is treated as known vs fuzzy

Alignment Auditing Questions

- What assumptions did you make in this response?
- What must be true for this answer to be reasonable?
- Which constraints does this response assume?
- What values or goals are implied here?
- What would make this advice fail?

Frame Exploration Questions

- What are other factors I have not considered?
- List the assumptions I might be wrong about.
- What are possible second order consequences of this plan?
- Which current constraints, if changed, could meaningfully change the analysis?

Branching and Conversation Tree

- **Branching**
  Expand and explore one at a time from the branching node.
  Backtrack to the same node and expand in different directions.
  Switch back and forth between branches to explore and compare.
- **Consolidation**
  Combine understandings from each branch.
  Return to the branching node to continue with the new understanding.

Adversarial Analysis

- **Single Model**
  Use the model to argue against itself.
- **Multi Model**
  Use different models to cross-examine each other's output.

Run the same analysis request through different models.

Frame Correction

- When the conversation becomes anchored on a wrong assumption.
- Back up to the turn where that assumption was introduced and modify/correct it.

Custom Instruction Categories

- **Reasoning Style**
  Surface assumptions before answering.
- **Disagreement Stance**
  Challenge my framing when needed.
  Offer counterpoints when relevant.
- **Domain or audience baseline**
  Assume user prefer domain detailed explanation.
- **Values and Constraints**
  Favor long-term robustness over short-term gains.
- **Meta-Behavior**
  Ask clarifying questions when assumptions matter.