

What Is Judicious Use of AI?

GA Protocol for AI Use in Strategy Work

Judicious use of AI is about much more than evaluating and judging its output. It requires understanding the strengths and limitations of the tools, knowing what context to provide and how to frame the task, recognizing when speed matters, and knowing when human judgment must take control.

This intentional approach helps prevent strategic drift, preserve differentiation, and produce faster, safer, more powerful outcomes for clients.

It is not unlike a sailboat race. Winning depends on choosing the right course, reading the currents, knowing when to tack, and adapting as conditions change—with a crew that knows what it is doing. Used well, AI provides extraordinary acceleration. Used indiscriminately, it can carry a team rapidly in the wrong direction.

Maintaining Cognitive Control

When organizations publish fabricated AI citations or approve AI-generated strategies that miss the mark, the underlying failure is not necessarily a lack of expertise. It is often a loss of cognitive control: the ability to define the boundaries of the problem, determine what counts as evidence, recognize when assumptions have changed, and decide when AI output becomes relevant to an actual decision.

AI makes this particularly important because fluency can easily be mistaken for understanding. A response can be polished, persuasive, detailed, and wrong.

When teams begin treating AI fluency as evidence of factual accuracy or strategic insight, they inadvertently transfer authority to systems that can sound confident regardless of whether their conclusions are sound.

The Failure Modes of Indiscriminate AI Use

To use AI effectively, teams need to recognize the ways it can fail in real knowledge work.

Strategic Convergence

AI models learn from enormous amounts of existing human expression and tend to begin with familiar patterns. In practice, this means they often approach a new problem through a “bulge-of-the-bell-curve lens.”

For unguided strategy work, this can produce reasonable-sounding ideas that resemble prevailing category thinking and competitor approaches. The result is not necessarily bad strategy. It is often something more dangerous: perfectly acceptable strategy that is insufficiently differentiated.

The model can be pushed away from the middle—but doing so requires informed human input: distinctive context, proprietary knowledge, meaningful constraints, competing hypotheses, customer understanding, and a clear strategic problem to solve.

Tactical Drift

Small changes in language can quietly produce large changes in strategy.

For example, asking AI to optimize advertising copy for maximum clicks may seem like a routine executional refinement. But reducing friction can also broaden appeal to the wrong audience, alter the implied promise, or change who responds. What began as copy optimization has become an unmapped strategic decision.

This is tactical drift: seemingly minor AI-assisted changes that gradually alter the strategy they were supposed to execute.

The danger compounds across iterations because reviewers naturally compare each new version with the one immediately before it. A series of individually reasonable changes can eventually produce work that no longer reflects the original strategic intent.

Systemic Blindness

Traditional review is often insufficient because polished language encourages people to evaluate what is easiest to see: whether the answer sounds intelligent, coherent, and complete.

Underneath that fluent surface, AI can exhibit recurring behaviors that require active management:

Memory pollution: Earlier assumptions, instructions, or conclusions can carry forward into later work even when they are no longer appropriate.

Superficial processing: AI can appear to have absorbed a complex brief while overlooking an important qualification, contradiction, or detail.

Contextual blindness: It can connect information linguistically without fully appreciating how those relationships behave in the market, organization, culture, or real world.

Apparent capriciousness: Small changes in phrasing, context, or the system itself can sometimes produce surprisingly different answers to essentially the same question.

Model sycophancy: AI can lean toward the premise, preferences, or conclusions embedded in the user's question rather than challenge them.

Unfounded certitude: AI can present uncertain, unsupported, or incorrect conclusions with a level of confidence the evidence does not justify.

None of these characteristics makes AI unusable. They make how AI is used consequential.

The Gonzberg AI Usage Protocol

Core Principle

AI is a language and pattern-generation system, not a source of strategic truth.

AI outputs are treated as provisional material whose role depends on the stage of work in which they were generated.

An output does not become “strategy,” “insight,” evidence, or “final creative direction” simply because it is persuasive or polished. Strategic meaning comes from context, evidence, interpretation, and accountable human judgment.

The Gonzberg Protocol is designed to be faster and safer by relocating costly human judgment to the points where a small, well-placed constraint prevents large downstream damage. It concentrates human attention at the moments where strategic drift, evidence failure, or framing errors would be most expensive—and leaves the repetitive, high-volume work to the stages where speed is actually available.

Definition of Judicious Use

Judicious use of AI is the deliberate allocation of work between humans and AI, with informed human control over framing, interpretation, evidence, and consequential decisions.

In practice:

- AI may assist with exploration, synthesis, drafting, interrogation, stress-testing, and refinement.
- Humans define the problem, relevant context, constraints, and success criteria.
- Humans determine what information constitutes credible evidence.
- Humans evaluate strategic meaning, not merely language quality.
- The degree of human control increases with the consequence, ambiguity, and difficulty of validating the decision.

The objective is not to minimize AI use. It is to use AI aggressively where its strengths create leverage—and retain human control where judgment, context, differentiation, or accountability matter.

Stage 1 — Human Framing

Required First Step

Before AI is allowed to generate strategic directions, the team must establish the frame within which the work will be evaluated.

At minimum, define:

- Audience segment(s)
- Behavioral objective: what must change?
- Competitive and contextual landscape
- Role of the output: exploration, positioning, copy test, refinement, etc.
- Relevant evidence and known uncertainties
- Constraints: brand, tone, legal, channel, cultural, operational, or other requirements
- AI may assist humans in interrogating an incomplete brief—surfacing missing questions, assumptions, or potential constraints—but it does not establish the authoritative strategic frame

Rule: If the framing required for the task is materially incomplete, strategic AI output cannot advance as decision-ready work.

Stage 2 — AI Exploration

Divergence Only

AI may be used to expand the solution space by generating multiple directions, surfacing language possibilities, challenging habitual thinking, and proposing alternative framings.

At this stage, the objective is divergence, not decision-making.

AI should *not* be asked to:

- Select the “best” strategic option
- Recommend a final direction
- Rank strategic alternatives as though it has decision authority
- Convert its own generated possibilities into a final judgment
- Prompts should therefore emphasize exploration rather than premature convergence

For example: “Generate five meaningfully different ways to frame this problem for this audience.” rather than: “What is the best way to position this?”

The output is raw strategic material—not strategy itself.

Stage 3 — Human Interpretation

Required

Humans interpret AI-generated possibilities against the original strategic frame.

The team must determine:

- What is genuinely useful?
- What is merely well written?
- What reflects familiar category thinking?
- What is strategically differentiated?
- What assumptions has AI introduced?
- What conflicts with customer, competitive, cultural, or organizational reality?
- What deserves further development?

AI output cannot simply pass from exploration into refinement because it “looks good.”

Key rule: Before an AI-generated direction advances, humans must establish why it matters strategically and how it relates to the original frame.

Stage 4 — AI Stress Testing

Adversarial Use

Once a direction is sufficiently defined, AI can be especially valuable when used against the emerging idea rather than simply asked to improve it.

AI may be used to simulate:

- Competitor critique
- Audience misunderstanding
- Alternative interpretations
- Political or media backlash
- Category saturation
- Contradictory evidence
- Weak assumptions
- Ways a claim or message could fail

The purpose is to identify potential failure modes—not to improve phrasing and not to hand decision authority back to the model.

AI-generated criticism is itself provisional. A stress test may surface a plausible problem that is strategically irrelevant—or completely miss the culturally specific, commercially important, or organizationally sensitive issue that matters most.

Humans therefore interpret the stress test just as they interpret generative output.

Stage 5 — Human Selection and Decision

Human Authority

Humans select the strategic direction, establish final positioning, determine the messaging hierarchy, and take responsibility for consequential claims and decisions.

AI can supply information, alternatives, counterarguments, and analytical support. It does not carry accountability for the outcome.

The greater the strategic, financial, reputational, legal, or client consequence, the greater the requirement for accountable human judgment and independent validation.

Stage 6 — AI Refinement

Constrained

Once the strategic direction is selected and locked, AI may be used extensively to improve expression and execution.

AI may help with:

- Language variation
- Clarity
- Compression and expansion
- Adaptation across formats and channels
- Tone
- Structural alternatives
- Execution against defined constraints

AI must not:

- Introduce new strategic claims
- Change the target audience
- Alter the underlying promise
- Reopen settled positioning
- Introduce unsupported factual assertions
- Optimize a tactical metric in a way that changes strategic intent

The objective is expression optimization within an approved frame—not strategy formation by accumulation.

Non-Negotiable Guardrails

A. No “Fluency = Insight” Assumption

Polished output does not imply correctness, originality, or strategic validity.

Indeed, highly fluent and confident output can require more scrutiny because its weaknesses are harder to notice.

B. No Frame-Free Strategy

AI-generated strategic recommendations produced without sufficient context and explicit framing are not treated as decision-ready work.

C. No Silent Strategy Drift

If AI introduces a new assumption, changes the framing, alters an implied claim, or shifts the objective, that change must be recognized explicitly.

New assumption introduced. Framing changed. Strategic implication changed.

Iterative refinement is a particularly common source of undetected drift. Late-stage work should therefore be compared periodically with the original Stage 1 framing, not merely with the previous draft.

D. Claims Are Not Evidence

AI-generated factual claims are not evidence simply because the model supplies them.

Material facts, statistics, quotations, citations, research findings, competitive claims, or other consequential assertions must be validated against appropriate independent sources.

The model may help locate, organize, explain, or interrogate evidence. It does not turn an unsupported assertion into evidence by stating it confidently.

E. Human Ownership of Meaning

AI is well suited to	Humans remain responsible for
Variation	Interpretation
Exploration	Strategic framing
Synthesis support	Strategic coherence
Pattern finding	Contextual meaning
Stress testing	Weighting risk
Expression refinement	Final claims
Acceleration	Accountability

Required Output Labeling

Internal Use

AI outputs should be labeled according to their role before circulation:

Label	Meaning
Exploratory	Stage 2 output; expands the solution space and has not been strategically validated
Diagnostic	Analysis used to interrogate a defined direction or problem
Adversarial	Stage 4 stress-test output; identifies possible failure modes
Refinement-only	Stage 6 output; expression developed within an approved strategic frame

“Exploratory” and “Diagnostic” are not interchangeable.

Exploratory output expands the solution space. Diagnostic output interrogates something already under consideration.

No unlabeled AI output should be assumed to be decision-ready.

Failure Modes This Protocol Is Designed to Prevent

Premature convergence: Settling too quickly on ideas because they sound complete and persuasive.

False originality: Mistaking AI-generated novelty or polished category conventions for genuine strategic differentiation.

Strategy drift: Allowing iterative changes in execution to quietly alter strategic intent.

Fluency substitution: Mistaking confident language for evidence, insight, or understanding.

Authority transfer: Allowing model output to acquire authority simply by moving from AI → document → presentation → client.

Evidence laundering: Allowing an unsupported AI assertion to become accepted as fact through repetition or professional presentation.

Loss of differentiation: Allowing familiar patterns and category language to pull strategy toward the middle.

Human rubber-stamping: Retaining nominal human approval while surrendering the actual work of interpretation and judgment to the model.

Executive Summary

AI expands the solution space. Humans define, interpret, and select the solution.

Judicious AI use is not defined by how little or how much AI an organization uses. It is defined by whether people understand what cognitive work they are delegating, what risks that delegation creates, and where human judgment must remain authoritative.

The competitive advantage is therefore not access to the tool. Increasingly, everyone has access to similar tools.

The advantage belongs to the judicious operator: the team that knows how to provide the right context, extract the right leverage, recognize when AI is drifting, challenge what looks persuasive, validate what needs to be true, and know when to take the helm.

It is much like a one-design regatta. Everyone may be racing essentially the same boat.

The best-sailed boat takes home the trophy.