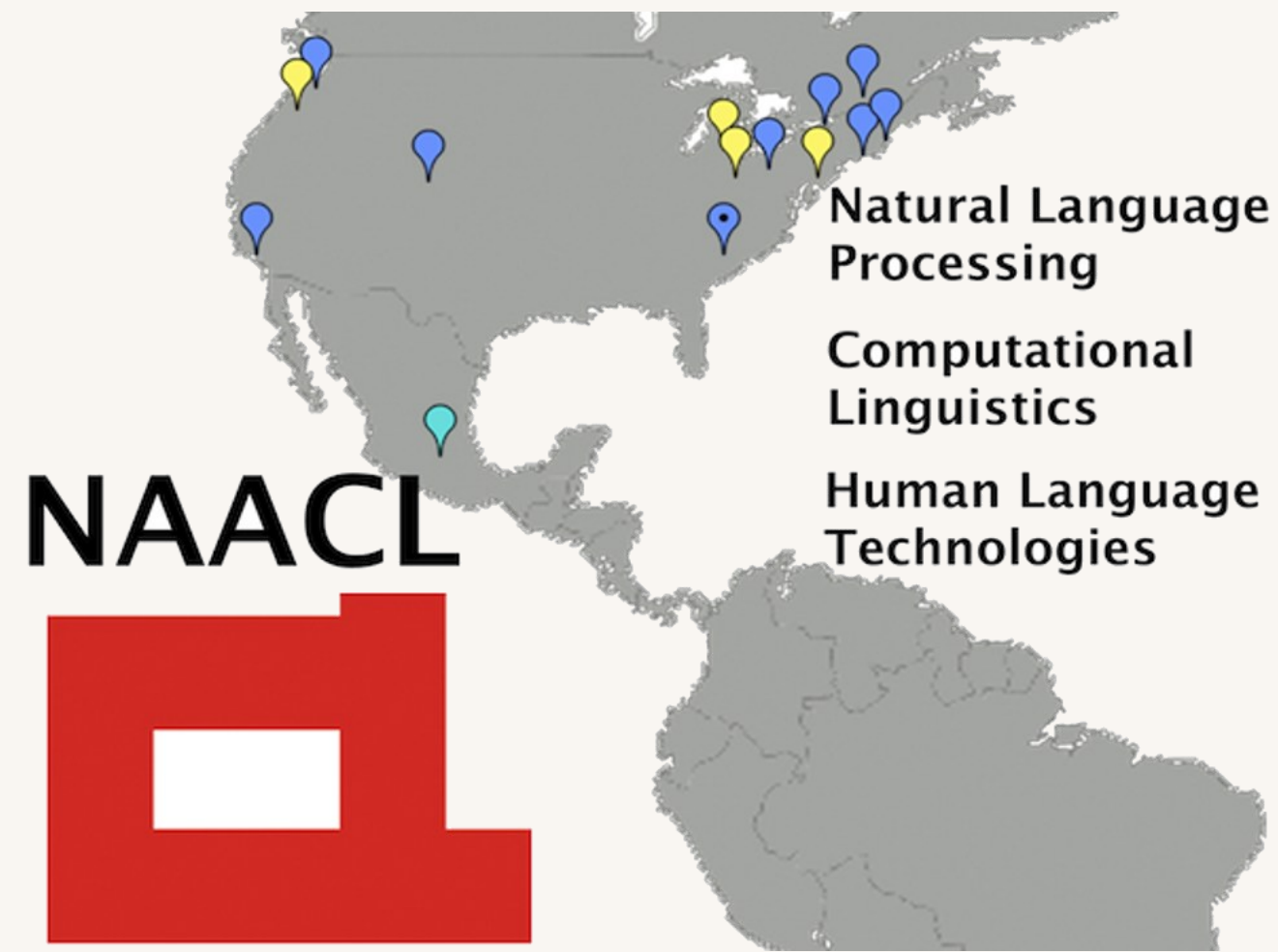
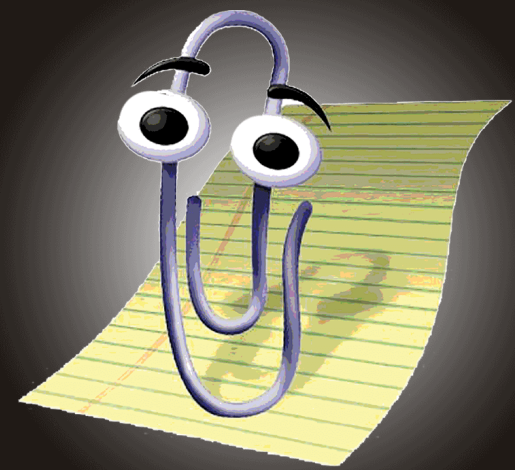


Conversational AI Beyond Chat

Task-oriented, goal-oriented, and proactive LLMs: how much can you leave unsaid?



2026 NAACL-Sponsored Summer School on Speech and Language Technology

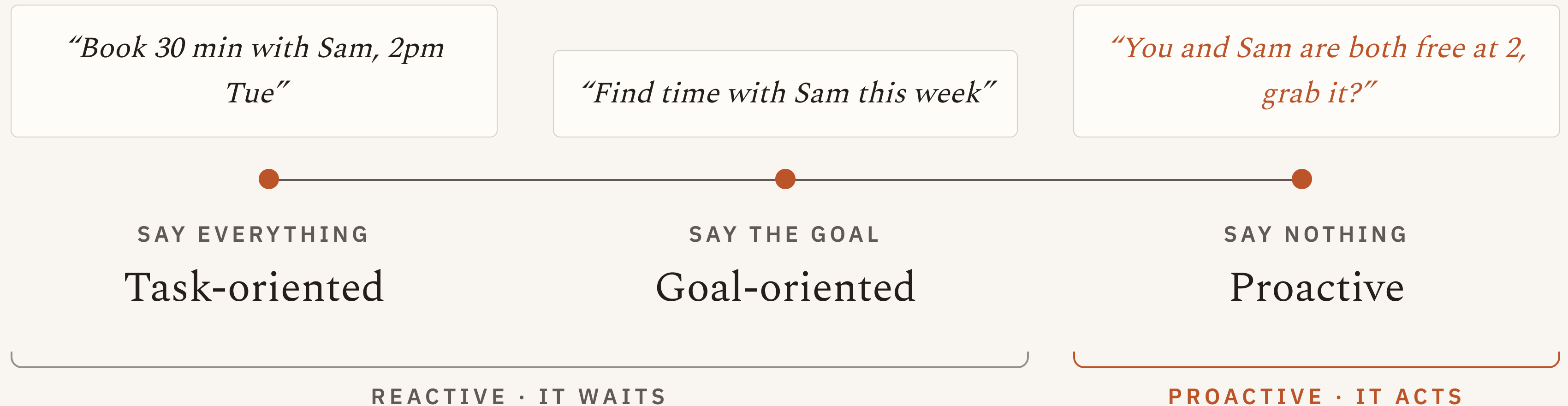


“It looks like you're writing a presentation. Would you like help?”

*When is it okay for a machine to act on what you **didn't** say?*

One scale: how much you leave unsaid

From **passive**, it waits for you, to **active**, it moves first.



The same move every time: the human says less, and the machine fills the gap.

Same trip, three assistants

One trip to plan. Three assistants, the difference is just how much you spell out.

01

“Book the 9am to Lisbon, window seat.”

Does exactly that, then waits for the next instruction.

TASK-ORIENTED · PASSIVE

02

“Plan my Lisbon trip.”

Breaks it into flights, hotel and itinerary, and works through them.

GOAL-ORIENTED · PASSIVE

03

(you say nothing)

Knows you've wanted Lisbon, sees you're free this weekend, spots a cheap flight, and messages you first.

PROACTIVE · ACTIVE

The assistant that waits

CONCEPT 01 · TASK-ORIENTED

Finish a bounded, well-defined task through conversation, and talk to a database to do it. You lead; it runs the exchange.

RESTAURANT BOOKING

Party size?

2 people

Date / time?

Fri, 8pm

Cuisine?

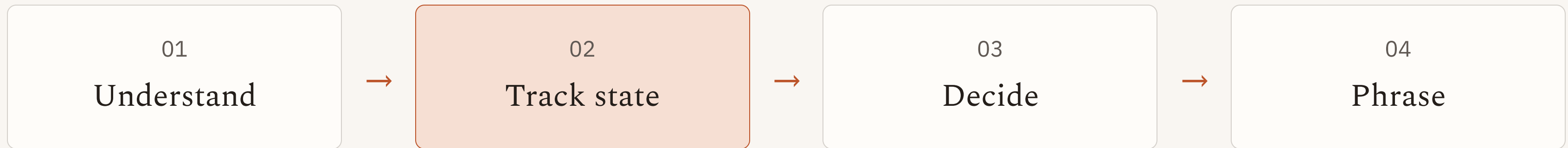
Italian

● Task-oriented · say everything

○ Goal-oriented

○ Proactive

Under the hood: understand, track, decide, phrase



STATE IS A BELIEF

“A cheap hotel” → did you mean £40 or £90 a night?
The system holds a *distribution* over what you meant, not a single guess.

THE BENCHMARK

MultiWOZ, the field's standard benchmark, is a large set of multi-domain task-oriented dialogues, exactly the pipeline above. The *WOZ* is Wizard-of-Oz: it was collected by paying humans to secretly play the bot behind a curtain.

The pipeline collapses into one model



Arenas: **MultiWOZ**, **Schema-Guided Dialogue**

*The honest twist: prompting a big LLM still hasn't reliably
beaten fine-tuning a small one here.*

Give the destination, drop the directions

CONCEPT 02 · GOAL-ORIENTED

Hand over a high-level objective. The system decomposes it into sub-goals and works through them on its own.

Heads up: “goal-oriented” gets used three ways, which is why the words blur.

“Plan my trip to Lisbon”

Book the flights

Find a hotel

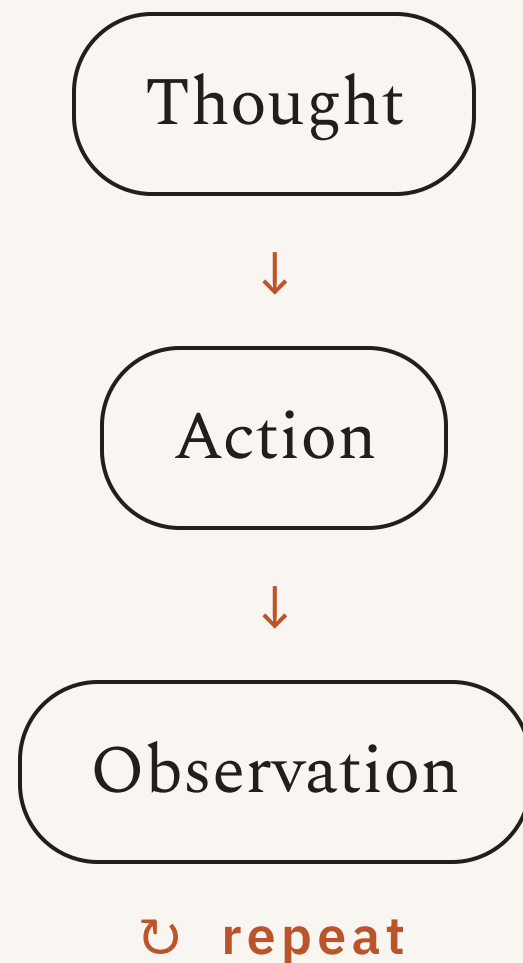
Build a day-by-day plan

○
Task-oriented

●
Goal-oriented · say the goal

○
Proactive

ReAct: think, act, observe, repeat



THOUGHT

wants a central hotel under £100 for Friday

ACTION

search hotels, filter by price and area

OBSERVATION

2 matches, both above £100

THOUGHT

widen the radius by 1 km and try again

The loop that turns a chatbot into an agent, and the founding paper got it working with one or two examples.

AutoGPT: the dream goes mainstream

Spring 2023 · the fastest-starred repo in GitHub history

160k

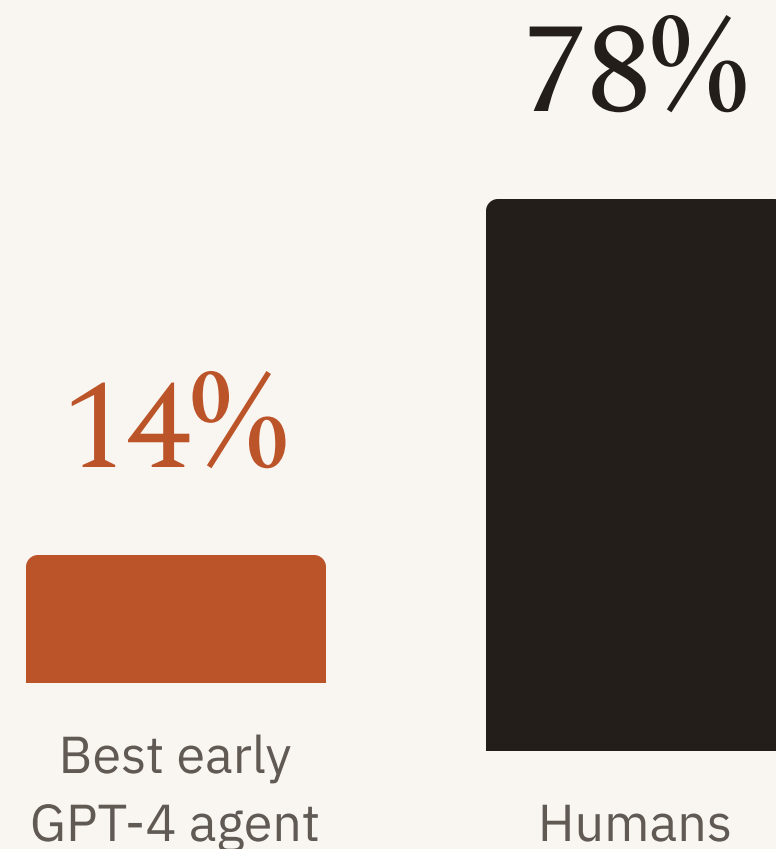
★ stars

- › goal: launch the new product
- › creating task list...
- › executing task 1: research competitors
- › executing task 2: draft landing page
- › thinking... thinking... thinking...
- › re-planning task 1... thinking...

The first public glimpse of the gap between the demo and the reality.

Do they actually get there?

Evaluation shifted to “did the task get done”: **WebArena**, **GAIA**, **SWE-bench** on real GitHub issues.



Open-ended help has *no* clean benchmark like this.

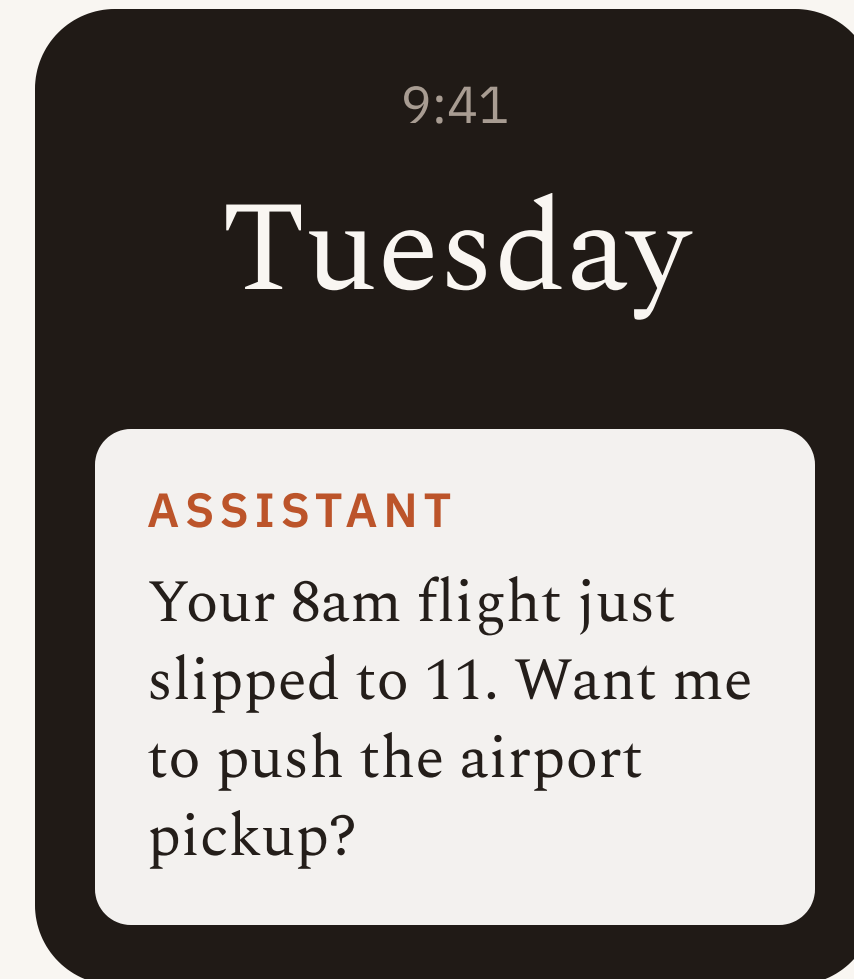
When success is progress over weeks, not one task finishing, there's nothing clean to score, the territory Part 2 walks into.

The agent moves first

CONCEPT 03 • PROACTIVE

It takes initiative: acting before you ask, or asking before you finish, anticipating what you need.

An assistant that only ever responds is passive; proactivity is what makes it active, but now it's guessing intent and timing.



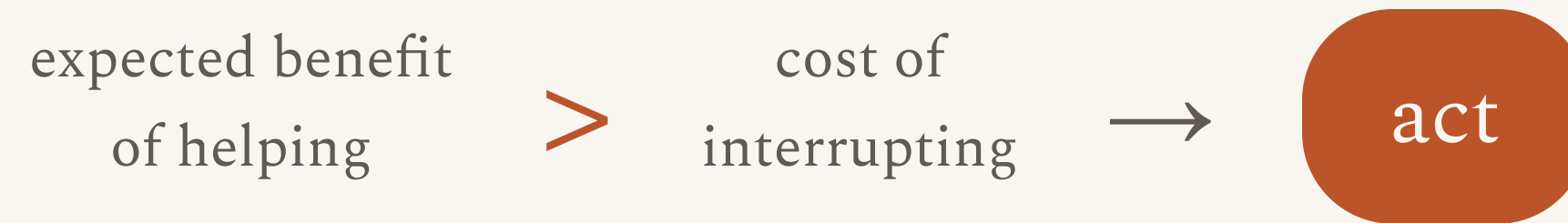
○ — Task-oriented

○ — Goal-oriented

● Proactive • active • say nothing

The old idea, done with math

Horvitz, *Principles of Mixed-Initiative User Interfaces*, 1999.



Clippy never did this arithmetic, which is exactly why it failed.

Act too eagerly → annoying

Horvitz, Mixed-Initiative UIs, CHI 1999

Wait too long → useless

Ask, don't guess

Two replies to *“Plan me a holiday.”*

GUESS

“Here's a generic 5-day Paris itinerary:
Louvre, Eiffel Tower...”

CLARIFY FIRST

“Beach, city, or mountains, and what's
your budget?”

The catch: they fixate on a vague word, then quietly guess on the things that change the answer.

Clippy 2.0

OFFERS MORE · READS YOUR WEEK

“You've cleared your inbox by 9 every day this week. Want me to start drafting replies?”

EASES OFF · KNOWS YOUR CALENDAR

“You're traveling Thursday and Friday. Want me to hold non-urgent pings till Monday?”

Agents that read your context and offer help before you ask. **ProactiveBench** tests exactly this.

A small fine-tuned model edged out GPT-4o. What Clippy wanted to be, occasionally getting it right.

Same message, two outcomes: genuinely helpful, or genuinely irritating.

THE TURN

Capable is not reliable

On **τ-bench**'s airline tasks, a strong model solves about a third.

$\approx 1/3$

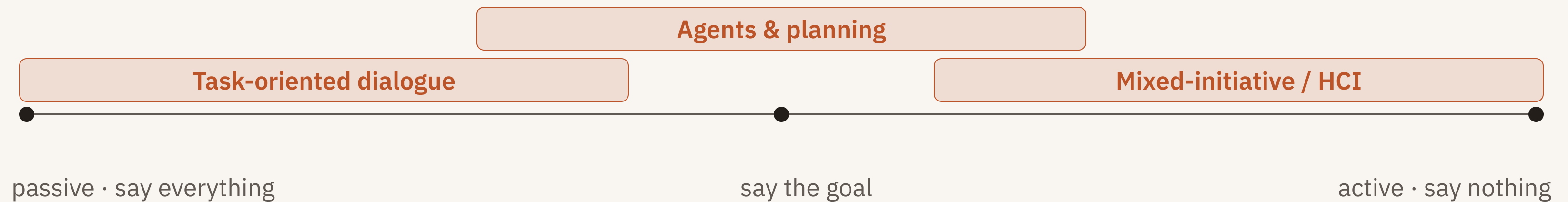
of airline tasks solved

SAME TASK, 8 TRIES



One try in three is a demo, not a product.

Three communities, one scale



Task and goal are the same story at different zoom; proactive is the plot twist where the machine stops waiting. Slide all the way to the active end and you have an **active conversational agent**, where Part 2 begins. **τ -bench** is where all three meet: a bounded job, run by a planning agent, that only works if it draws missing details out of the user.

Why this is genuinely hard

Leave things unsaid, and the machine must *infer* intent. Inference fails three ways.

WRONG CONTENT

Guessing wrong

Clippy recommends the wrong thing.

“Booked the 6am flight,” the morning after you said you hate early starts.

WRONG CONSISTENCY

Guessing unreliably

Advice that flips between sessions.

*Monday: “go with the cheap hotel.”
Thursday: “splurge on the nice one.”*

WRONG TIMING

Guessing at the wrong moment

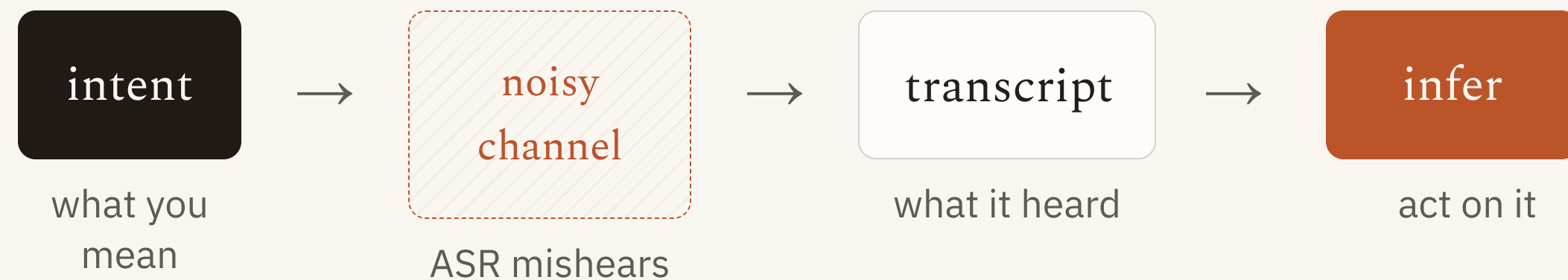
A 2am nudge about tomorrow's trip.
“Ready to plan?” pinged at 2am.

Current mitigation: a human in the loop, a person checks the agent before it reaches the user.

The more you leave unsaid, the harder all three bite.

Now add speech: a noisy channel

Everything so far assumed clean text. Make the input a misheard waveform, in real time, with no chance to reread.



You say “I can't make the
8am.”

→

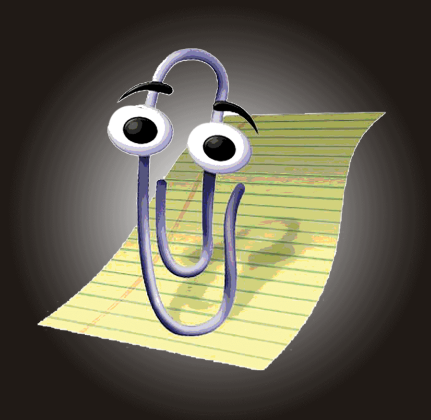
heard as “I can make the
8am.”

→

*one dropped word flips the whole
plan.*

Saying less is hardest exactly where this workshop lives.

THE TAKEAWAY



So: how much do you want to **leave
unsaid?**

Slide to the active end and you've built an active conversational agent, which is where Part 2 begins.

From chatbots to active conversational agents

From answering questions to guiding structured, adaptive conversations.

PASSIVE CHATBOT

Answers when asked



Q: What is task-based language teaching?



Task-based language teaching (TBLT) is an approach that uses meaningful tasks as the basis for language learning.

One-off question → one-off answer

respond

remember

guide



ACTIVE CONVERSATIONAL AGENT

Guides a structured conversation

1

What is your teaching context (e.g., level, setting)?

University EFL, intermediate.

2

What are your main goals this term?

Improve speaking fluency.

3

Which skills do you want to prioritize?

Speaking and listening.

4

Here are task ideas and a plan. Refine anything?

Yes, adjust the assessment.



Great! Here's your tailored teaching plan.



Multi-step dialogue → goal-directed support

🎯 The shift is from isolated replies to **goal-directed conversational support**.

Why is health coaching a useful case study?

A demanding domain for active conversational agents.

Health coaching is a collaborative, client-centered process where a trained professional guides, empowers, and motivates you to make sustainable lifestyle changes.

Unlike traditional medical advice where a doctor tells you what to do, health coaches partner with you to achieve self-identified wellness, fitness, and chronic disease management goals.

WHY IT STRESSES CONVERSATIONAL AGENTS

Memory & continuity

Remember history, track goals, detect trends over time.

Adaptive dialogue

Adjust questions, depth, and tone to the person and moment.

Uncertainty handling

Manage incomplete, ambiguous, or noisy information safely.

Structured guidance

Offer evidence-based steps, plans, and decision support.

 Health coaching combines structure, personalization, continuity, and safety, making it a **powerful test case** for conversational agents.

Health coaching as **active** conversation

PASSIVE CHAT (REACTIVE)

The agent waits for the client to ask.

CLIENT



How can I eat healthier?

AGENT

Here are some tips for eating healthy...



CLIENT



Okay, thanks.

 No agenda or structure

 Limited exploration of goals and barriers

 Missed opportunities to motivate

 Harder to drive progress

ACTIVE COACHING (PROACTIVE)

The agent takes initiative toward progress.

COACHING AGENT

Welcome back! Last time your goal was to exercise 3 times a week. How has it been?



CLIENT



Honestly, not great. I've been really busy and tired.

COACHING AGENT

Thanks for sharing. What were the biggest challenges? What could help this week?



 Agent guides the agenda

 Explores barriers and context

 Elicits motivation and commitment

 Drives progress toward the goal

ACTIVE COACHING FLOW

 **1. Set the focus**
Clarify the current goal and session focus.

 **2. Explore**
Ask about progress, barriers, context, and feelings.

 **3. Elicit motivation**
Reflect, affirm, and strengthen reasons for change.

 **4. Collaborate**
Brainstorm strategies and plan realistic next steps.

 **5. Confirm & commit**
Summarize, confirm the plan, and elicit commitment.

Active = guided, purposeful, collaborative, client-centered.

 An active agent **leads the conversation**, so the client doesn't get stuck, drift, or give up.

The end goal

From labour-intensive monthly coaching toward AI-supported, continuously supervised habit building, with human oversight.

TODAY: TRADITIONAL COACHING

-  **Periodic (e.g., monthly)**
Check-ins happen infrequently.
-  **Labour intensive**
Coaches spend significant time on each client.
-  **Manual tracking**
Notes and progress tracking are fragmented.
-  **Limited scalability**
Hard to support many clients consistently.
-  **Coach-centric**
Clients depend on coach availability and capacity.



FUTURE: AI-SUPPORTED, CONTINUOUSLY SUPERVISED



🎯 **Goal:** help every client build healthier habits and achieve progress everyday.

🎯 Human expertise + AI support + continuous interaction = better health coaching at scale.

What do LLMs already do well?

Open-ended dialogue, summarization, personalization, memory, and adaptive generation, now feasible.

Open-ended dialogue

Natural language across many topics and phrasings.

Impact: more human-like conversations.

Personalization

Adapt tone, style, and level to the individual.

Impact: more relevant and supportive.

Adaptive responses

Adjust to the client's state and evolving needs.

Impact: timely and flexible.

Summarization

Condense long sessions into goals and action items.

Impact: keep the big picture.

Memory use

Reuse history, preferences, and past progress.

Impact: continuity and context.

Knowledge integration

Blend general knowledge with coaching principles.

Impact: evidence-aligned guidance.

EXAMPLE: GOAL CHECK-IN

Before · rule-based

User: "I've been struggling to stick to my plan lately..."

System: "Please describe your difficulty in more detail."

With LLMs

User: "I've been struggling to stick to my plan lately..."

LLM: "It sounds challenging to stay on track. Let's explore what's been getting in the way, what's still working, and how to adjust your plan. Where would you like to start?"

 LLMs expand what conversational agents can do: From rigid scripts to **flexible, adaptive, personalized** health coaching.

Why is generic LLM chat not enough?

THE 5 BUILDING BLOCKS

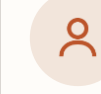
-  **1. Goal-oriented**
Keep the goal clear, specific, and measurable.
-  **2. Progress focused**
Celebrate small wins and track meaningful progress.
-  **3. Longitudinal memory**
Remember history, patterns, preferences, and barriers.
-  **4. Personalized**
Adapt to the person's goals, context, readiness, and style.
-  **5. Uncertainty aware**
Embrace uncertainty, plan flexibly, and adapt.

Combined, these create conversations that inspire action and sustain change.

FULL EXAMPLE: PUTTING IT ALL TOGETHER

-  **Client**
I want to exercise more, but I'm overwhelmed this week. Work is busy and I'm tired.
- Coaching Agent** 
I hear you. Tight deadlines this week (remembers history). Let's focus on a small step that fits your routine. What feels doable?
-  **Client**
Maybe a 10-minute walk after lunch.
- Coaching Agent** 
Great choice! That supports your goal to feel more active. Let's plan it Mon–Fri after lunch.
-  **Client**
On Wednesday I missed it. I felt discouraged.
- Coaching Agent** 
Missing a day happens (uncertainty is normal). What could help you get back on track tomorrow?
-  **Client**
I'll walk with a colleague after lunch.
- Coaching Agent** 
Perfect — that adds support. Last week you did 3 short walks, great progress! Let's keep building.

WHAT MAKES THIS GREAT HC

-  Keeps the goal visible and meaningful
-  Encourages action and tracks progress
-  Uses memory for relevant support
-  Personalizes tone, ideas, and strategies
-  Adapts to uncertainty and changing plans
-  Builds trust, motivation, accountability

OUTCOMES FOR THE CLIENT

- ✓ More confidence
- ✓ Healthier habits
- ✓ Better well-being
- ✓ Long-term success

 Great coaching = the **right conversation, at the right time, in the right way** turning goals into action, plans into habits, and habits into a better life.

Health coaching as goal-oriented conversation

KEY IDEA

Every coaching conversation should connect to a **specific behavior goal**.



- 🎯 Define the goal clearly
- 📈 Track progress over time
- 📋 Address barriers
- 💡 Plan actionable next steps
- 🚩 Reinforce commitment

FROM GENERIC TALK → GOAL PROGRESS

Generic Chat

Client

I had a stressful week.

Bot

I'm sorry to hear that.
Take care.

Goal-Oriented Coaching

Client

I skipped workouts 3 times.

Bot

That sounds tough. How did it affect your exercise goal?

Bot

What got in the way? What could help this week?

Client

I was too tired. Maybe morning walks.

Bot

Great idea! Let's plan 3 short morning walks. How confident are you?

WHAT COUNTS AS SUCCESS?

- 🎯 **Goal clarity**
The client and agent share a clear, realistic goal.
- 📈 **Measurable progress**
Small wins are tracked and recognized.
- 📋 **Barrier resolution**
Obstacles are identified and addressed.
- ☰ **Actionable next steps**
Concrete, doable plans are created.
- 🔄 **Sustained motivation**
Commitment is built and renewed.

Goal-oriented = Every turn has a purpose and brings the client closer to change.

🎯 A good answer is helpful. **Goal progress is what changes lives.**

Health coaching as longitudinal conversation

WHY LONGITUDINAL MEMORY MATTERS

-  **Builds continuity**
Remembers past goals, discussions, and outcomes across sessions.
-  **Understands patterns**
Recognizes recurring barriers, triggers, and facilitators over time.
-  **Personalizes effectively**
Learns preferences, routines, values, and communication style.
-  **Tracks progress**
Keeps track of small wins, setbacks, and overall trajectory.
-  **Supports long-term change**
Behavior change happens over weeks and months, not one chat.

EXAMPLE: REMEMBERING OVER TIME

-  **Session 1 (May 5)**
Goal: walk 30 min, 5 days/week
Barrier: busy evenings
Plan: walk after dinner on weekdays
-  **Session 4 (May 26)**
Progress: 3–4 days/week
Insight: prefers morning walks
Plan: try morning walks Tue–Thu
-  **Session 8 (June 16)**
Progress: 4–5 days/week
New barrier: work deadline stress
Plan: short 15-min walks + stretch

The agent uses this history to provide context-aware, relevant, and empathetic coaching at every turn.

EXAMPLE: USING LONGITUDINAL MEMORY

-  **Client (today)**
I haven't been walking much this week. I'm so exhausted after work.
-  **Coaching Agent (uses history)**
I remember work deadlines were a challenge before. Last time you tried short 15-minute walks on stressful days and said it helped. Worth trying again this week?
-  **Client**
Yes, that could work. 15 minutes feels doable.
-  **Coaching Agent**
Great! Let's plan 15-minute walks on heavy-deadline days. We'll review next time.

 Healthy habits are built over time. Longitudinal memory enables **consistent, personalized coaching** that grows with the client.

Health coaching as **personalized** conversation



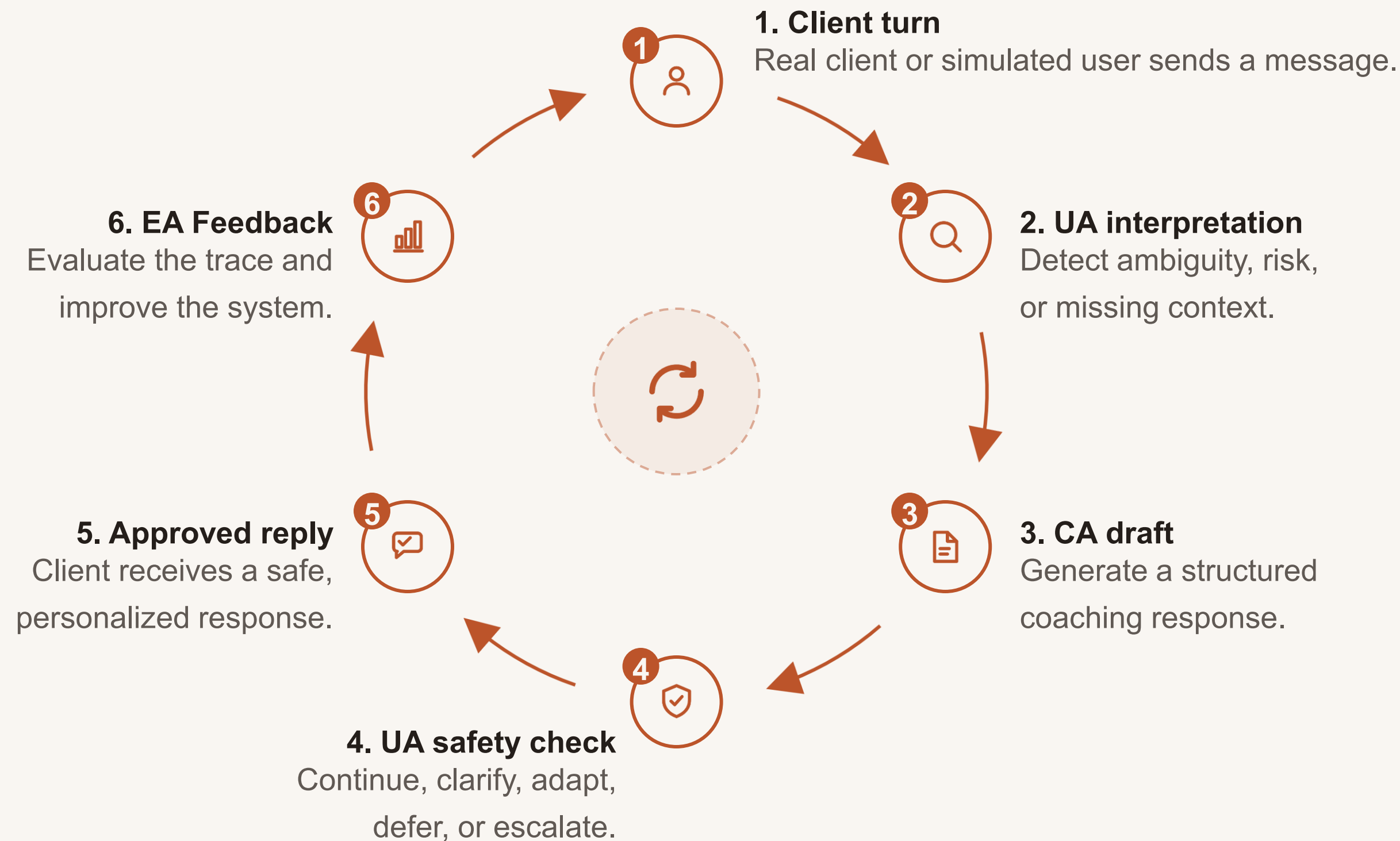
 Personalized coaching meets the client **where they are** and helps them move toward **where they want to be**.

Health coaching as **uncertainty-aware** conversation

UNCERTAINTY IS NORMAL	HOW THE AGENT RESPONDS	EXAMPLE: AN UNCERTAIN WEEK
Clients face many unknowns: ? Will I have time this week? ☁️ What if something unexpected comes up? 📅 Will I have enough motivation tomorrow? 👥 How will my environment influence me? ↑ What if progress is slower than expected?	🎯 1. Acknowledge uncertainty Validate that setbacks and unknowns are a natural part of change. 🕒 2. Focus on what can be controlled Shift attention to actions and routines the client can influence. ➡️ 3. Plan for different scenarios Help the client think ahead and prepare “if-then” plans. 🔄 4. Review and adapt Check what happened, learn from it, and adjust the plan. 📈 5. Keep the long-term goal in sight Reconnect to purpose and values when the path feels unclear.	👤 Client I planned to exercise 4 times this week, but overtime happened twice. I only did it once. I feel like I’m falling behind. 🗣️ Coaching Agent Thanks for sharing. Unexpected things happen. Let’s look at what you can control and adjust the plan. 👤 Client I can do shorter workouts at home on busy days. 🗣️ Coaching Agent Great idea! What’s a realistic plan for the rest of the week? 👤 Client 2 short workouts after work and one longer session on the weekend. 🗣️ Coaching Agent Perfect. You adapted and kept moving. That’s progress!

🎯 The agent’s job is not to predict the future, but to help the client **navigate it with confidence.**

The live health coaching loop



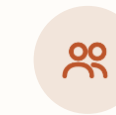
WHY THIS LOOP MATTERS



Keeps the conversation purposeful.



Handles uncertainty before harm.



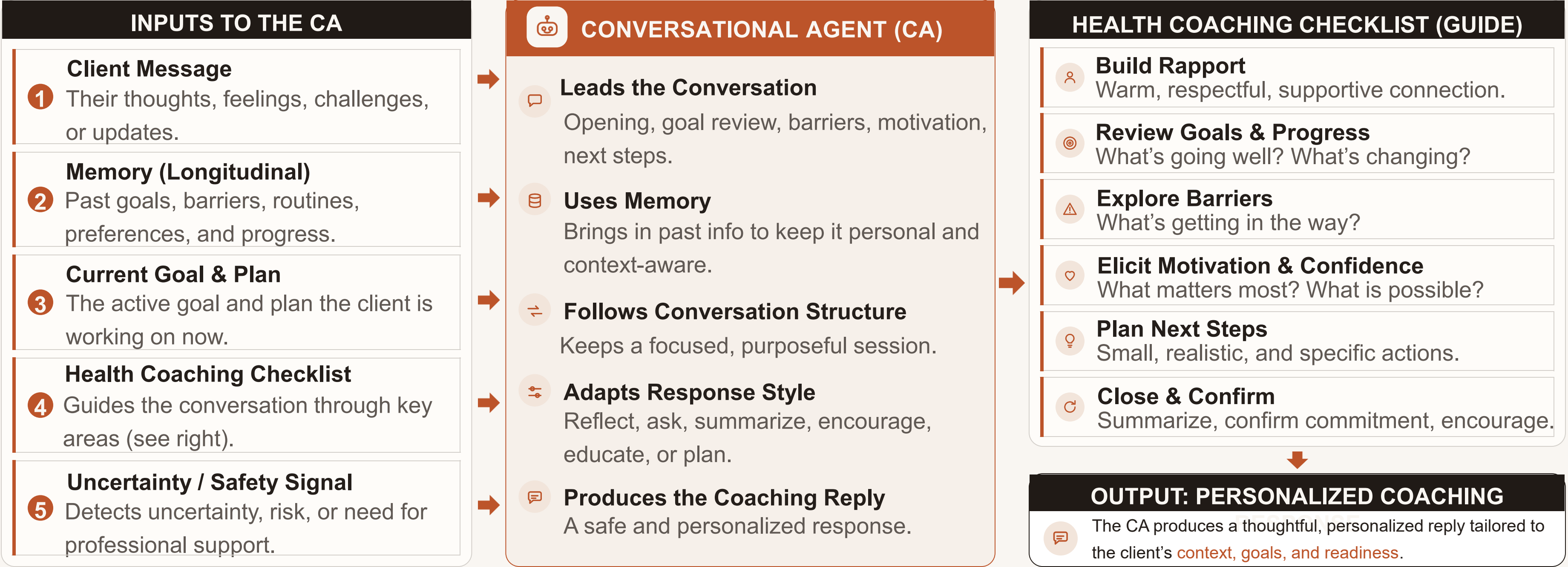
Supports human oversight.



Creates evaluable interaction traces.

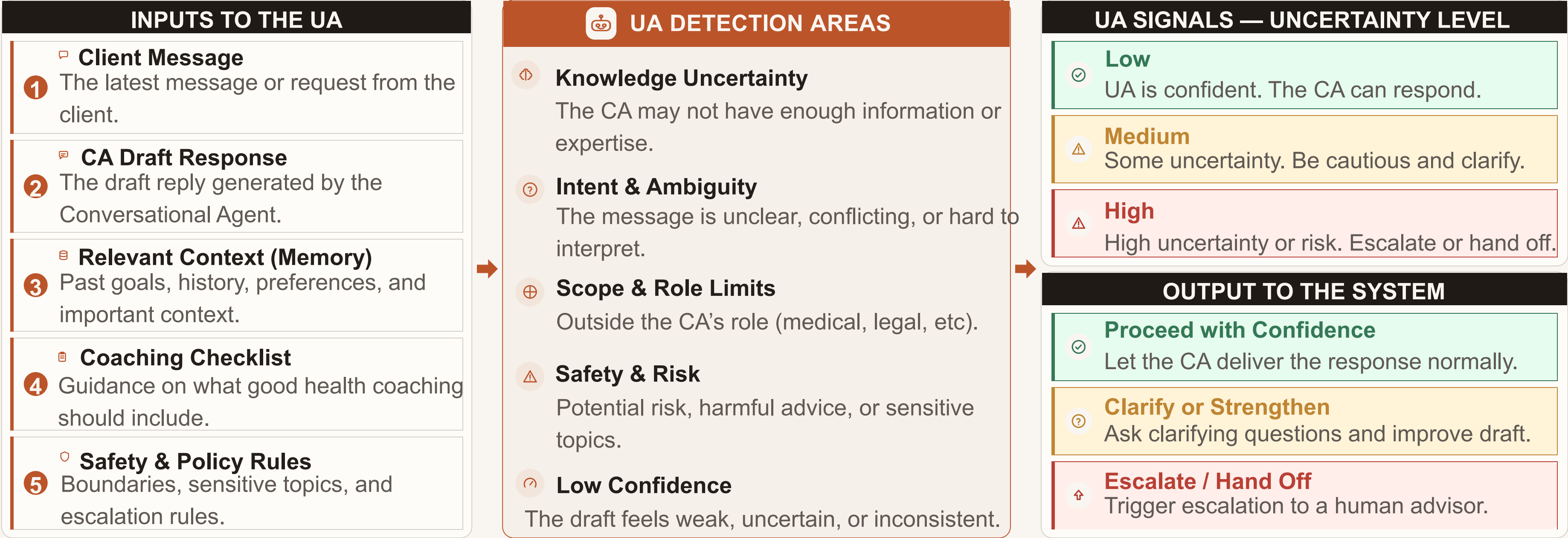
🎯 The system is not a single reply generator — it is a monitored coaching loop built for **safe, adaptive, and measurable** support.

Conversational Agent (CA)



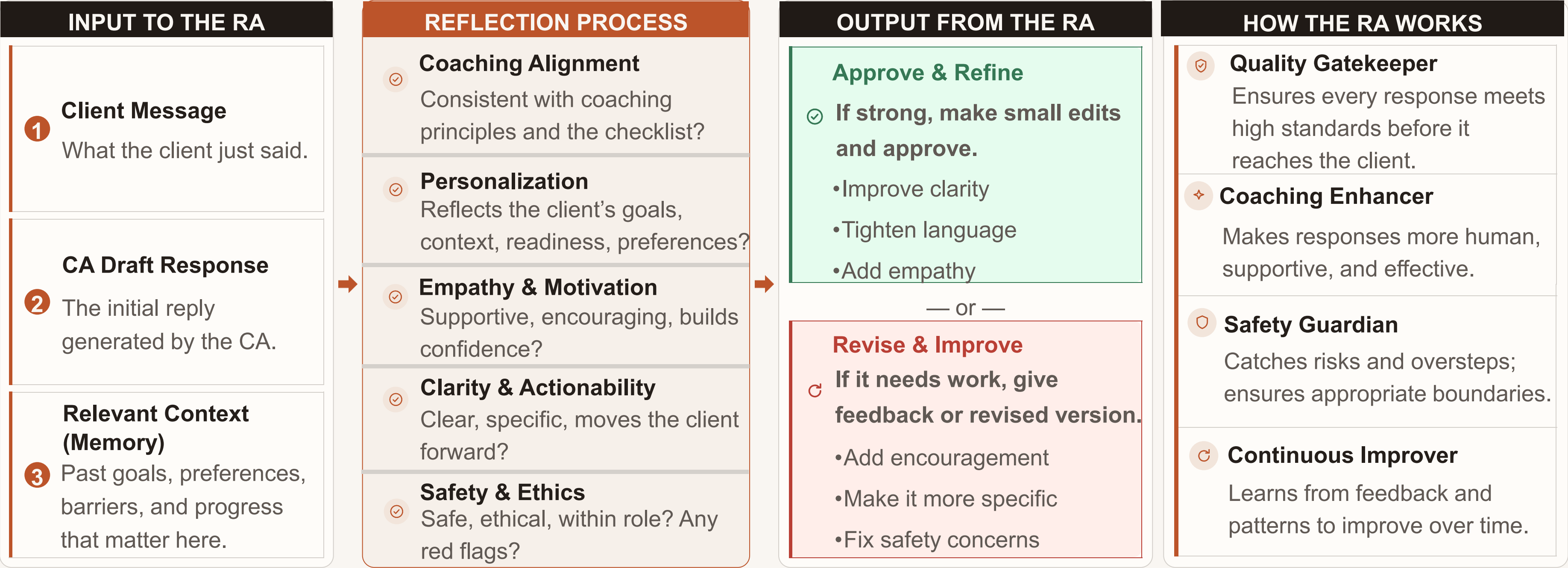
The CA turns information and guidance into a **helpful, relevant, and safe** coaching conversation that empowers clients to create healthier lives.

Uncertainty Agent (UA)



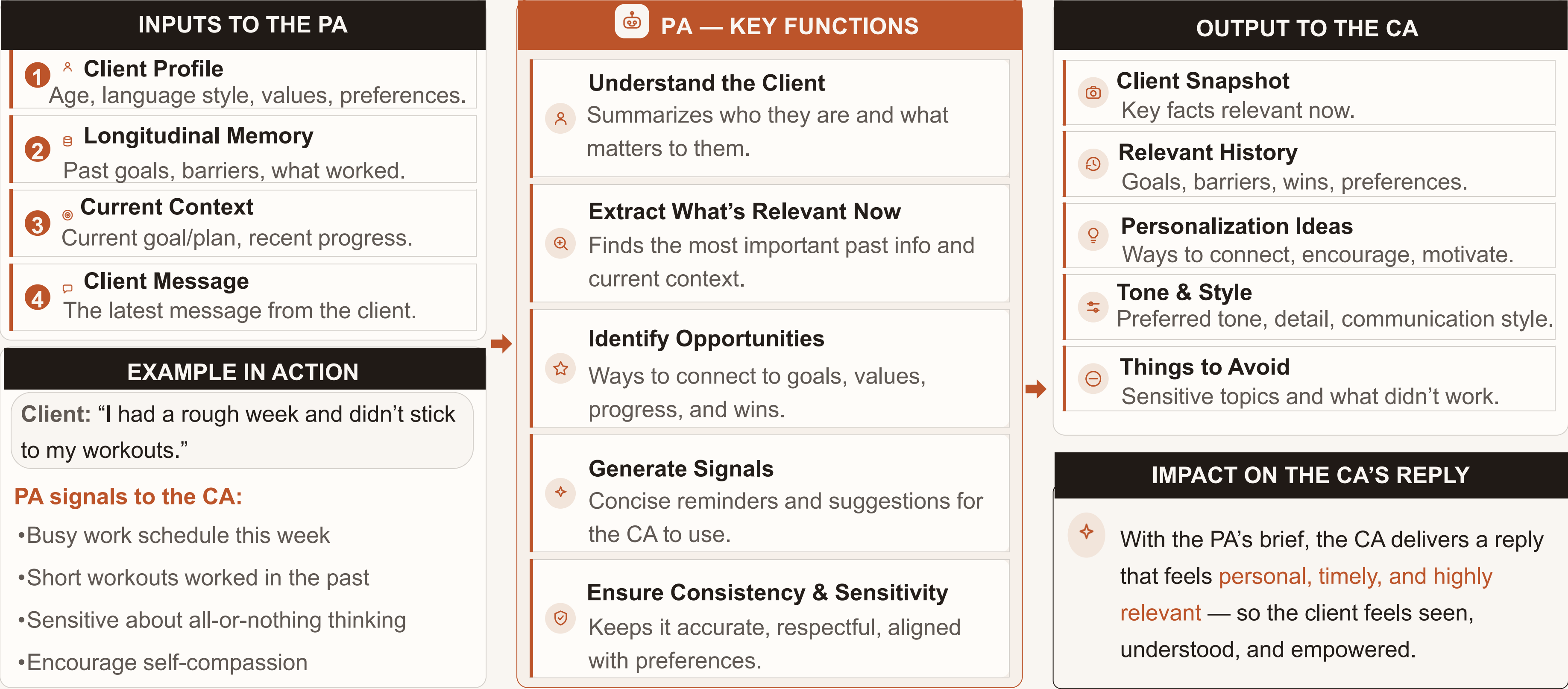
🎯 The UA ensures the CA knows when to **answer**, when to **clarify**, and when to **escalate** — so every client gets safe, accurate, and responsible support.

Reflection Agent (RA)

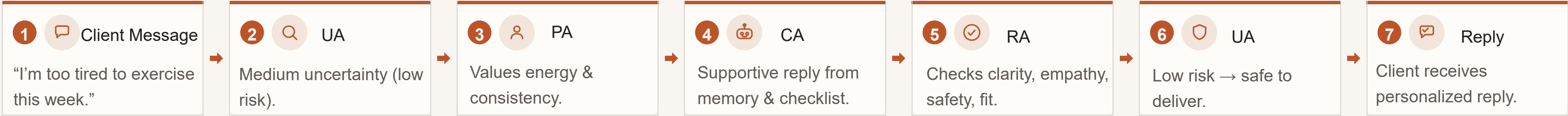


🎯 The RA is the quality partner of the CA — ensuring every message is **clear, personalized, supportive, safe**, and truly helpful for the client.

Personalization Agent (PA)



End-to-end conversation example



WHAT EACH AGENT CONTRIBUTES

	CA Leads the conversation and drafts the reply.
	RA Polishes the draft for quality and empathy.
	PA Injects personal context and style.
	UA Detects uncertainty and risk.
	Safety Layer Applies policies and boundaries.
	EA Learns from feedback and outcomes.

ONE TURN IN ACTION

AGENT	WHAT IT DOES	EXAMPLE OUTPUT / DECISION
 UA	Detects uncertainty level	Medium — tiredness could be stress, illness, or low motivation.
 PA	Provides personalization signals	Values consistency · prefers gentle approaches · busy schedule.
 CA	Drafts the reply	"I hear you're feeling really tired. Let's find a small step that still fits."
 RA	Refines empathy & clarity	Adds empathy + choice + normalization.
 Safety	Final safety check	Low risk → OK to deliver.
 Reply	Approved response is sent	"How about a 10-minute walk after dinner? You decide."
 EA	Logs outcome & feedback	Client reply used to improve future conversations.

CONVERSATION SNAPSHOT

 Client	"I'm too tired to exercise this week."
 UA	Medium uncertainty (low risk).
 PA	Values consistency; low energy this week.
 CA	Drafts a small, achievable step.
 RA	Refined for empathy and clarity.
 Safety	Low risk → approved.
 Reply	"How about a 10-minute walk after dinner?"
 Client	"Okay, I can try that." Feedback logged.

🎯 All agents work together in every turn to ensure **safe, relevant, empathetic, and effective** health coaching.

Build systems that help people move forward

Conversational AI is not just about better answers, but better support.

THE RESEARCH OPPORTUNITY

-  Move from answers to guidance.
-  Design for goals, memory, and adaptation.
-  Treat uncertainty as a first-class problem.
-  Build systems that know when to ask, act, or hand off.

A FINAL THOUGHT

The best conversational systems do not just talk well. They help people move forward — safely, purposefully, and at the right moment.

How much can you leave unsaid?

FINDING YOUR PATH

-  1. Ask yourself what you are passionate about.
-  2. Building foundational models matters.
-  3. Building applications that improve people's lives matters too.
-  4. Both paths are equally valuable.
-  5. The best research connects curiosity with human impact.

 The most exciting research begins where systems must **remember, adapt, and act responsibly.**