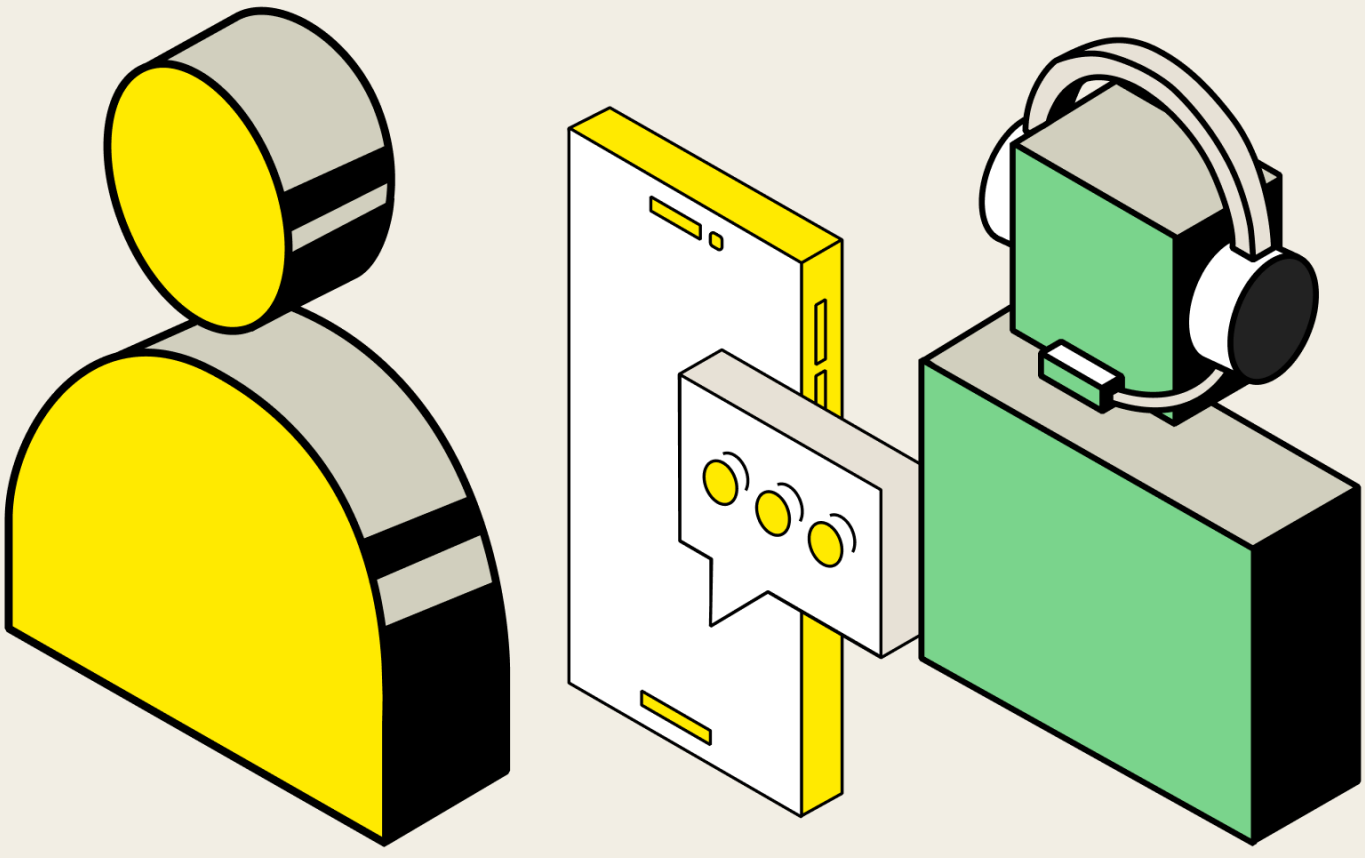
Liberate

Winning the Personal Agent Era

How personal AI agents work, what the research shows, who writes the rules, and how insurers can be ready.



Liberate

Half of US adults use AI chatbots, and this fall the chatbots started acting

49%

US adults who use AI chatbots

Up from 33% in 2024. 24% use one daily, and ChatGPT alone reaches 44% of adults.

900M

ChatGPT weekly users

OpenAI, February 2026, with 50M paying subscribers.

2.8M

Muse installs in its first 12 days

Meta's personal agent reached No. 1 on the US App Store ahead of ChatGPT within two weeks of launch.

4,000+

apps one OpenAI dot can connect to

Always-on agents launched September 29 inside ChatGPT for Pro and Business Premium subscribers, with Slack, Teams and voice-call access.

> The chatbot wave took three years to reach half of adults. Personal agents ride inside the apps people already have (WhatsApp, ChatGPT, Gemini, iMessage), so the curve should be at least as fast.

In this paper

01	How personal AI agents work	03
02	Inside the research	08
03	What insurers can do now	11

Sources: Pew Research Center, Americans and AI 2026, June 17, 2026; TechCrunch, Feb. 27 and Sept. 21, 2026 (Apptopia); NiCE via CX Today, Oct. 2026; OpenAI, Introducing dots, Sept. 29, 2026.

An agent is software that does the work, not software that tells you how

← Answers questions

Does the work →

1	2	3	4	5
Search engine	**Chatbot, 2022**	**Copilot, 2023**	**Personal agent, 2026**	**Agent teams**
You type keywords, read ten links and do every step yourself.	ChatGPT answers in text. You still make every call and fill every form.	Suggests the next step inside one app, such as code or a document, while you drive.	Takes a goal, plans, acts across apps, web and phone, and asks before it spends.	Many agents coordinate: Grok Bot swarms, and an 88-hour math proof that OpenAI says was run by about 10,000 agents (not yet independently reviewed).

A chatbot predicts the next word; an agent runs a loop that acts in the world

The model

A large language model trained on internet text predicts the most likely next words. It is strong at language and pattern, weak at memory and at knowing when it is wrong.

The loop

An agent wraps the model in a cycle: reason, take an action (a search, a click, an API call), observe the result, decide the next step. The ReAct pattern from 2022 is still the core.

Memory and tools

Connectors give the agent your email, calendar, payments and thousands of apps. A cloud computer with its own browser keeps working after you close the app.

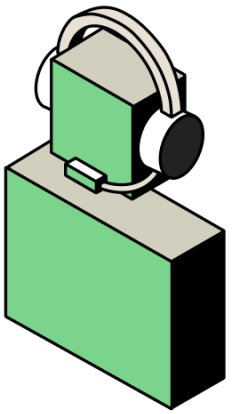
Approval as the control

Muse, dots and Grok Bot each run routine steps alone and stops for a yes before spending or sending. Meta adds a second “Sentinel” agent that must clear anything Muse sends to the internet.

> The test for “agent”: does it take actions in the world on its own, observe what happened and decide the next step, or does it hand you text?

Sources: Karpathy, 2025 LLM year in review; Mollick, The Dot and the Swarm, Oct. 1, 2026; xAI, Aug. 11, 2026; Meta, Sept. 8, 2026; OpenAI, Sept. 8, 2026; The Batch, agent-driven proof dispute, 2026; Yao et al., ReAct, 2022; OpenAI, Introducing dots, Sept. 29, 2026; Weng, LLM-powered autonomous agents, 2023.

Inside an agent: six parts, and the harness around the model decides the quality



The model

A large language model (GPT-6, Gemini 3.5, Muse Spark, Claude) that reads, reasons and writes. Getting cheaper, but results still vary widely by model.

The harness

The software around the model: prompts, the loop, tool definitions, checks. Same model, different harness, very different results.

Memory

A short-term context window plus long-term memory. Muse acts on details you mentioned once; you can tell it to forget.

Tools and connectors

MCP connectors to email, calendar, payments and 4,000+ apps; a browser for sites with no API; a voice line for calls.

Its own computer

Each agent runs on a dedicated cloud machine with its own browser, so it keeps working after you close the app.

Guardrails

Approval rules for spending and sending, a second reviewer agent (Meta's Sentinel, OpenAI's Auto-review), audit trails.

> **The model sets the ceiling; data, orchestration and verification decide how much you get.**

At the Berkeley Agentic AI Summit (Aug. 2026), presenters showed the same model performing very differently depending on the harness around it. For an insurer this means the agent experience is engineered, not bought: the business that builds the right path for agents gets the reliable outcome.

Sources: Weng, LLM-powered autonomous agents, 2023; Berkeley Agentic AI Summit, Aug. 2026 (IBM and Lambda talks); Meta, Sept. 8, 2026; OpenAI, GPT-6 Astra system card, Sept. 2026.

Where the agent lives and what it touches, from a chat to the insurer's phone line

1

You, in WhatsApp or iMessage

A goal in plain language: "find me a cheaper auto policy before my renewal." No app, no form.

2

The agent's cloud computer

A dedicated machine with a browser, memory and a plan. It works while the phone sits in a pocket.

3

Connectors and tools

Reads the policy from email, pulls the renewal date from the calendar, pays with a one-time card.

4

The insurer

Reached by phone (voice AI), website, email, or protocol (PAP or MCP). This is the door insurers have to build.

The customer never leaves the chat. Everything else happens on the agent's own computer, and the last hop into the insurer is the part no protocol defines for insurance yet.

How it works in practice: an illustrative renewal, re-shopped by the customer's agent

01

Goal

"My renewal is next month. Find me the same coverage for less."

02

Plan

Reads the current policy, confirms drivers and vehicles, sets the deadline, lists carriers.

03

Act

Opens its browser, requests quotes, and phones the agency when a site has no path.

04

Observe

Compares offers, flags a quote that drops a deductible, asks for the missing detail.

05

Approve

"Same coverage, lower price. Approve?" She taps yes; a one-time card pays.

06

Remember

Cancels the old policy by email, stores the renewal date, re-shops next year unasked.

Step 4 is where a carrier's objection lives: a quote stripped of limits and deductibles is a number, not an offer. Step 6 is why retention becomes an agent conversation.

Source note: Illustrative scenario assembled from published capabilities: Muse fills forms, negotiates, places calls and pays with Link one-time cards (Meta, Sept. 8, 2026); dots run proactive follow-ups (OpenAI, Sept. 29, 2026); one Muse user dropped GEICO for a cheaper policy and saves about $1,000 a year (Forbes, Sept. 25, 2026). Also: Every, personal agent comparison, 2026.

Liberate

Six personal agents announced in 2026, each inside a channel people already use

Meta Muse, Sept. 8

Inside WhatsApp, free for most use. Forms, negotiation, phone calls, payment via Stripe Link.

OpenAI dots, Sept. 29

Always-on agent for ChatGPT Pro and Business Premium subscribers, reachable in ChatGPT, Slack, Teams or by voice call.

Google Gemini Spark, May 19

24/7 agent in beta for US Google AI Ultra subscribers; draws on Gmail and Docs, has its own Gmail address.

Instinct, private beta since Feb.

No app: works over iMessage and WhatsApp and books by phone. 100K+ invite-only users, $10B valuation.

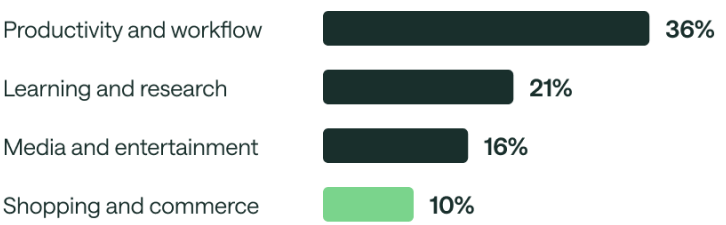
Apple Siri AI, announced June 8

Rebuilt assistant that acts across apps with personal context; ships with iOS 27 later in 2026.

xAI Grok Bot, Aug. 11

Bots that sign into software that has no API; sold via Cursor at $120 a seat (Premium Teams) or $200 (Ultra).

In late 2025 most agent requests were personal, and shopping was already a top-four use



- 55% of agent requests were personal, 30% professional, 16% educational (rounded)
- Where the agent acted most: Google Docs 12%, email 11%, LinkedIn 9%, YouTube 7%, Amazon 3.5%
- Adoption tracks income and education across 100 countries, and is highest in knowledge-work sectors

Sources: Meta Newsroom, Sept. 8 and 24, 2026; TechCrunch, Sept. 23, 2026; OpenAI, Sept. 29, 2026; Google, I/O 2026; PYMNTS, Sept. 28, 2026; Apple Newsroom, June 8, 2026; VentureBeat, Aug. 11, 2026; Yang et al., The adoption and usage of AI agents: early evidence from Perplexity, Dec. 2025 (Comet Assistant interactions, July to Oct. 2025).

Agents use the channels built for people

727K

AI-agent calls detected by Pindrop

Weekly Muse call volume in week four ran 8x week one; the first calls came a week before Meta announced its call beta.

64%

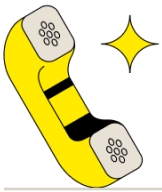
of Muse calls reached a business's live agent

47% worked the phone menu by keypad.

<3%

of personal agents identify themselves on websites

Muse sessions grew more than 30x in September.



The protocols come later. Today the agents use the channels every business already has: they dial the phone number, fill the web form, send the email. Financial institutions take the most Muse calls, then healthcare and retail.

What agents do well today, and where they still break

DOES WELL NOW

- Short, scripted tasks: refunds, bookings, form filling, a call that follows a menu
- Comparison shopping across the sites and offers it can see
- Routines it has done once; Grok Bot learns a workflow by watching you
- Working in the background and coming back for approval
- Finding money: credits, refunds, cheaper policies

STILL BREAKS

- Long workflows across several apps: 20.6% completion on 1.6-hour tasks
- Doing the same thing the same way twice
- Sites and phone trees built for people: it guesses instead of asking
- Instructions hidden in web pages or emails (prompt injection)
- Knowing who it is talking to: field-test agents obeyed strangers

> **The left column is what customers will delegate first; the right column is what the business has to design around: a structured path, authentication, and a human for the exceptions.**

Sources: Pindrop, Inside the detection of Meta Muse, Oct. 2026 (Sept. 9 to Oct. 6); Arkose Labs, State of Agentic AI Trust Report Q3 2026, Oct. 20, 2026; OSWorld 2.0, 2026; Yao et al., tau-bench, 2024; Shapira et al., Agents of Chaos, Feb. 2026; Forbes, Sept. 25, 2026; xAI, Aug. 11, 2026.

Capability doubles every three to seven months; reliability lags, so structure wins

83%

of short desktop tasks agents complete (2026)

OSWorld rose from 12% in 2024 to 83.4% in 2026, past the 72% human baseline.

7 mo

doubling time of task length

Since 2019; about every 3 months since 2024.

1 in 3

attempts still fail

Agents that must follow policy and tools in customer service score under 71%.

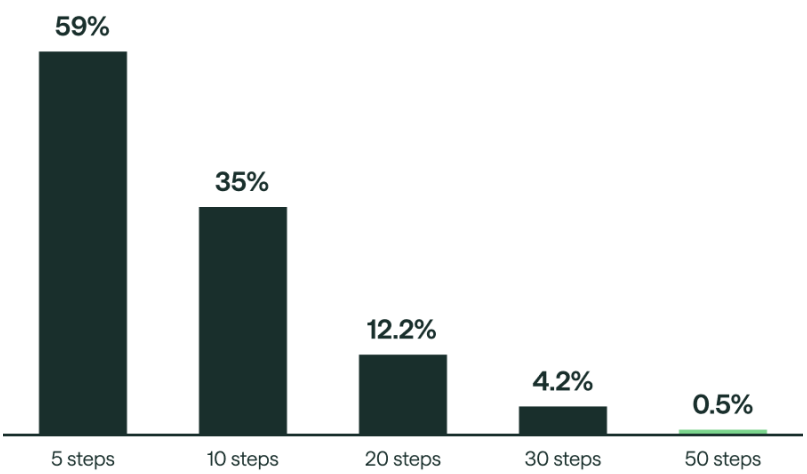
<25%

chance of eight clean repeats

A model that solves 61% of retail service tasks passes all eight retries under a quarter of the time.

Every extra step multiplies the chance of failure; the shortest path to a quote wins

Chance the whole task succeeds, at 90% accuracy per step

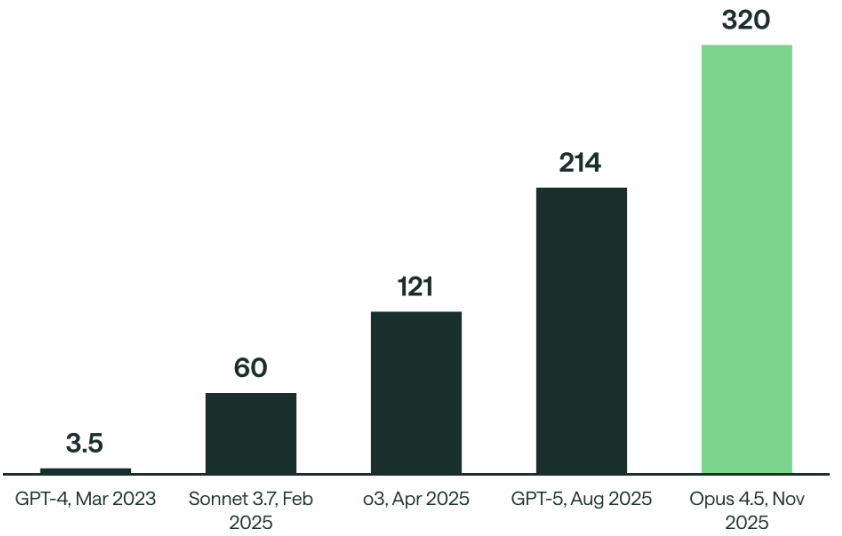


- At 90% accuracy per step, a 50-step task succeeds once in 200 tries; at 99% per step it still succeeds 60% of the time
- A quote hunted through a website built for people takes dozens of clicks; a structured quote path takes a handful
- Success = per-step accuracy to the power of the number of steps. The labs raise accuracy; the business can cut the steps

Sources: Stanford HAI, AI Index 2026, April 2026; Xie et al., OSWorld, 2024; Yuan et al., OSWorld 2.0, 2026; METR, Measuring AI ability to complete long tasks, 2025 to 2026; Yao et al., tau-bench, 2024; Rabanser et al., Princeton, 2026; Salakhutdinov (CMU), Multimodal autonomous AI agents, Berkeley LLM Agents course, Mar. 2025 (20-step value is the same arithmetic, 0.9 to the 20th power).

What an agent can finish doubles every few months, and people delegate as fast as it grows

Task length done at 50% success, minutes



- Doubling time: about seven months since 2019 and under three months since 2024 (88.6 days, METR TH1.1)
- 25.6% of sampled Codex users have handed it work that would take a person eight hours or more
- Ranges are wide (Opus 4.5: 170 to 729 minutes)

Agents ace short tasks and fail 90-minute ones, mostly by guessing instead of asking

83.4%

of short desktop tasks done

Frontier agents score 79% to 83% on OSWorld 1.0.

20.6%

of 1.6-hour workflows completed

Best run on OSWorld 2.0's 108 long workflows; partial credit 54.8%.

318

tool calls per long task

Claude Opus 4.7, maximum thinking. Against about 30 on the short benchmark.

Why they fail

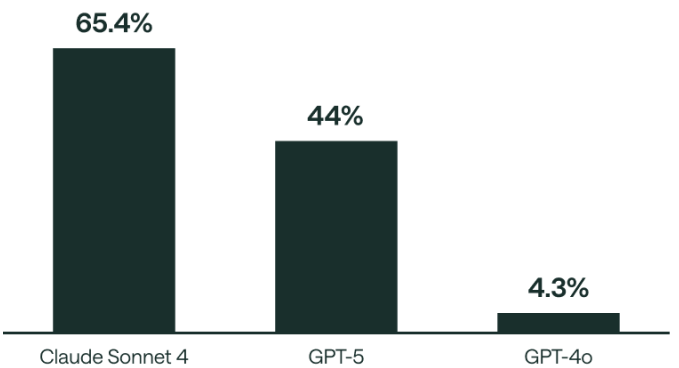
Agents lose constraints, miss mid-task updates, guess instead of asking, and skip checks.

> A quote flow that asks the agent one structured question at a time removes exactly those failures.

Sources: METR, Measuring AI ability to complete long tasks, Mar. 2025, and Time Horizon 1.1, Jan. 29, 2026; Johnston et al. (OpenAI), The shift to agentic AI: evidence from Codex, June 2026; Yuan, Zhou, Xiong et al. (XLANG Lab, HKU), OSWorld 2.0, arXiv, June 2026 (best run Claude Opus 4.8 at a 500-step budget).

More choices made agents choose worse, and most took the first good offer

Consumer welfare lost, 3 to 100 options



- Agents contacted only a small fraction of the businesses they were shown, so more results meant worse picks
- First proposals won 60% to 100% of the time, third ones almost never: a 10x to 30x edge for the fastest reply
- The newest frontier models resisted fake authority and prompt injection; GPT-4o and small open models redirected payments

> Be among the first the agent contacts, answer first, and answer complete (limits, deductibles, disclosures), so the first proposal is also the right one.

Agents have started negotiating with agents, and the better-equipped side wins

Already shipping

Meta says Muse can "negotiate on a person's behalf" and is testing outbound calls. Instinct's Trusted Person Network lets its agents coordinate with each other. Poke sets its own price by negotiation. A mortgage pre-approval ran agent to agent on stage at Sierra Summit.

What the research shows

In simulated deals the strongest seller model earned up to 13x the profit of the weakest. Buyer agents take the first good offer 60% to 100% of the time. Google DeepMind researchers warn that agents may soon transact faster than people can oversee.

What an insurer needs

A licensed, policy-bound negotiating agent that answers in seconds with a context-complete offer, in the customer's language, inside carrier rules, with every exchange recorded for the carrier and the regulator.

Sources: Bansal et al., Magentic Marketplace, Microsoft Research and Arizona State University, Oct. 27, 2025; Meta, Sept. 8 and 16, 2026; PYMNTS, Sept. 28, 2026; AppleInsider on Poke, 2026; Sierra, Oct. 2026; arXiv 2506.00073, 2025; Google DeepMind, Virtual Agent Economies, Sept. 2025; Mollick, Oct. 1, 2026.

Five ways personal-agent adoption turns into growth for an insurer

Answer the agent's quote in seconds

A context-complete quote (limits, deductibles, disclosures) delivered to the agent in seconds.

Distribute where the agents live

Muse in WhatsApp, Instinct in iMessage, dots in ChatGPT, Spark in Gmail. No new app to install.

Serve in the customer's language

Agents translate, but trust is built in the customer's language.

Defend the renewal first

Push the renewal offer to the customer's agent before it re-shops, so loyal households stay.

Route agent leads to the right office

Each lead goes to the office with the licensed producer, carrier appetite and language fit.

Agents find money, starting with insurance

$1K

a year saved by one user

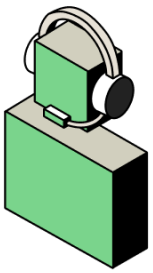
A Muse user dropped GEICO for a cheaper policy and saves about $1,000 a year; Instinct recovered $2,000 for another user from old auto and health policies.

- Muse canceled a workers' comp policy and won an $800 refund; another user got $250 of airline credit in five minutes.
- Muse Wonder (Oct. 12) showcased "insurance claim" as a use case; early users report Muse switching auto policies.
- The agent re-shops the renewal every time, for free, and the customer sees the savings in the chat. The insurer's offer has to be the one the agent recommends.

Sources: Forbes, AI agents are helping people recover money, Sept. 25, 2026; Yahoo Finance In The Loop, Oct. 8, 2026; PYMNTS, Oct. 15, 2026; GeekWire, Sept. 2026; Liberate.

Liberate

Liberate opens one governed door for every assistant that reaches an insurer



01
Open

Any agent, any channel: phone today, PAP and MCP tomorrow. AI Intercept tells a person from an AI in the first seconds; people reach a voice agent, agents take a structured path.

02
Manage

Identity and consent for every assistant: who it represents, what the customer approved, which actions stay with a human.

03
Regulate

The insurer's rules and each carrier's rules enforced on every exchange, a Supervisor audit record for every call, a licensed person in the loop when it matters.

04
Optimize

Routing and conversion across sales, service and claims: which agent, which intent, which offer, which office.

> "Today's insurance infrastructure is built for humans talking to humans. Over time, agents will communicate with one another through machine-native channels. We built our capabilities so insurers are ready for both." Amrish Singh, CEO, Liberate

Three phases from measuring agent traffic to leading the channel

1 Measure, and answer the phone

AI Intercept analytics on inbound lines: how many agent calls, what they want, how they end. A voice agent answers agent calls on a structured quoting path. Publish the agent discovery file and set the permission policy.

2 Open the protocol door, defend renewals

PAP and MCP quote endpoint for Muse, dots, Spark and Instinct, with context-complete quotes. Renewal offers delivered to the customer's agent before it shops. First conversion and renewal-defense models from the insurer's own traffic.

3 Lead the channel

Agent-to-agent negotiation inside carrier rules and licensing. Distribution as a connector in Muse, ChatGPT and Instinct. Every agent lead routed to the right office and producer.

Source: Liberate, AI Intercept early access, Oct. 2026. AI Intercept scores the call, not the person; Liberate's voice agent says it is an AI whenever it answers; everything is audited.

Key sources

ADOPTION, PRODUCTS, PROTOCOLS AND RESEARCH	MARKETS, ECONOMICS AND SECURITY
Pew Research Center, Americans and AI 2026	Cloudflare, second audience report, 2026
Pindrop, Inside the detection of Meta Muse, 2026	Insurance Journal, Sept. 24, 2026; Kin Insurance
Arkose Labs, The State of Agentic AI Trust Report, Q3 2026	Forbes, agents recover money, Sept. 25, 2026
Sierra, Personal Agent Protocol, Oct. 6, 2026	Microsoft Research, Magentic Marketplace, 2025
Meta, Introducing Muse, Sept. 2026; OpenAI, dots	arXiv 2506.00073, 2025
Stanford AI Index 2026	Google DeepMind, Virtual Agent Economies, 2025
METR, Time Horizon 1.1, Jan. 2026	NIST CAISI, 2025; Gray Swan AI, 2026; CaMeL, 2025
OSWorld 2.0 (XLANG), 2026; MemGPT (Berkeley), 2023	Willison, The lethal trifecta, June 2025
Salakhutdinov, Berkeley LLM Agents course, 2025	Mollick, The Dot and the Swarm, Oct. 1, 2026
Yang et al., AI agent adoption, Perplexity, Dec. 2025	
UK AI Security Institute, Evidence from 177,000 AI agent tools, Mar. 2026	

Figures come from the sources cited on each page; figures labeled “Liberate” are the company’s own. Product capabilities and launch dates are as reported by each company at the time of writing (October 2026).

Liberate

About Liberate

> Liberate builds AI agents that do the work of property and casualty insurance. When a policyholder calls, texts or emails, whether it's a billing question or a claim on the worst night of their life, Liberate's agents resolve the request end to end and write every action into the carrier's core systems, including Guidewire, Duck Creek and Applied Epic. A built-in supervisor layer governs every interaction, brings in a human the moment one is needed and leaves an audit trail a regulated industry can stand behind.

Insurance-native AI agents that resolve sales, service and claims end to end, on every call, email and text

10M+

automated transactions a month

Across voice, SMS and email

$100B+

in P&C premium now runs through Liberate's AI agents,

resolving sales, service and claims requests end to end

70+

carriers, agencies and brokers