AGENTIC AI · IPSOS CLIENT WORKSHOP

The agent has entered the building.

What autonomous AI agents are, why they spread faster than any open-source software before them, and what a research and marketing leader should do about it. This brief carries everything from the workshop, sources included.

PREPARED BY

BASIC

DATE

August 2026

WORKSHOP

aiagent.research.my

Prepared for attendees of the BASIC agentic AI workshop series, hosted for Ipsos clients. Please share within your organisation.

OVERVIEW

THE THIRTY MINUTES, ON PAPER

A chatbot answers. An agent acts.

A chatbot waits for a question, answers it, and forgets. An AI agent is a model given tools, memory and permission to act toward a goal without step-by-step supervision. It plans its steps, operates real software (browser, files, email, databases), checks its own output, and keeps working while you sleep. That difference, follow-through, is what changed between late 2025 and today.

The evidence: adoption without permission

Two open-source personal agents, built outside every major AI lab, produced the fastest adoption curves in recent software memory. OpenClaw shipped on 24 November 2025 as one developer's weekend project and counted 386,400 GitHub stars by 16 August 2026. Hermes Agent followed in February 2026 and by May had overtaken OpenClaw to become the most-used agent in the world; in the 30 days to 16 August 2026 it processed 36.4 trillion tokens on OpenRouter, holding the number one global rank among all AI applications. Around February 2026, machine-driven agent traffic passed human traffic on that platform altogether. The whole category is under nine months old. Chapter 02 carries every source.

386K	36.4T	Feb 26	<9
OpenClaw GitHub stars, 16 August 2026	Hermes tokens on OpenRouter in 30 days	Machine traffic passed human traffic	Months old, the whole category

The delivery mechanism matters as much as the numbers. Previous enterprise software arrived top-down through procurement. Agents arrive bottom-up: free, self-hosted, installed on a laptop in an evening, talking to their owner on WhatsApp, which already reaches 90.7 percent of Malaysian internet users. If your team includes anyone technical, the realistic assumption is that an agent is already inside the building.

What it means for insight and marketing work

Three scenarios sized for this audience, each a full chapter in this brief.

SCENARIO 01

Brand intelligence

Always-on competitive coverage

An agent sweeps competitor pricing, campaigns and ratings nightly and delivers a sourced digest before the Monday meeting. The analyst becomes the editor.

SCENARIO 02

The production layer, compressed

Open-end coding, transcript synthesis, questionnaire QA and first-draft reporting move from days to hours, with humans reviewing instead of producing.

SCENARIO 03

Chief of staff

The personal starting point

Inbox triage, meeting prep and brief drafting from a personal agent on WhatsApp: the lowest-risk way for a leader to learn what agents really do.

The workshop's hands-on segment put two live agents in every attendee's pocket: GAJAH on a frontier model and KANCIL on a budget model, both sandboxed, both querying structured Malaysian data (612,624 official statistics rows, income and poverty measures for 999 geographies, and a census of 17.2 million Southeast Asian businesses, 1.9 million of them Malaysian). The lesson of the comparison is a management lesson: match the model to the job, because budget models now run volume work at 1 to 3 percent of frontier token prices.

The honest part, and the playbook

Everything that makes an agent useful, the tools, the memory, the autonomy, is also the attack surface. The documented record includes a zero-click prompt injection against a production enterprise copilot, 40,214 self-hosted agent instances found exposed on the open internet in February 2026, and IBM's finding that one in five breached organisations traced the breach to unsanctioned AI at an average of USD 670,000 in extra cost. None of it argues for a ban. It argues for governance, because a ban does not remove the agents, it keeps them invisible.

What good looks like is six rules: sandboxed, least privilege, human approval on irreversible actions, everything logged, right-sized models, and data kept in your jurisdiction. The first ninety days are three moves: see (inventory and a two-page policy), pilot (one governed workflow on real data), decide (measure, then scale or stop with evidence). The goal of the first ninety days is not transformation. It is an informed decision.

The takeaway. Do not ask whether your company will use agents. Ask who is governing the ones it already has.

CHAPTER 01

FOUNDATIONS · THE SHIFT

What an AI agent actually is

A chatbot answers. An agent acts. The difference is not a smarter model, it is what the model is given: tools that touch real software, memory that survives the session, and permission to pursue a goal without step-by-step supervision. This page is the conceptual foundation behind slides 2, 3 and 6.

The chatbot and the agent

A chatbot is a conversation. You ask, it answers, and the exchange ends there. If the answer implies work, the work is still yours: open the spreadsheet, send the email, chase the supplier. An agent takes the goal instead of the question, and keeps going after you stop typing.

Chatbot	Agent
Waits for the next prompt	Works towards a goal between prompts
Forgets when the window closes	Remembers you, your projects and past sessions
Describes what you should do	Does it: opens the file, runs the task, sends the update

> THE ONE-LINE DEFINITION
>
> An agent is a model given tools, memory and permission to pursue a goal on your behalf, checking its own work as it goes.

The six shifts that made it real

None of this is a single breakthrough. Six changes landed close together, and each removed a reason agents used to fail.

- **Tool use:** models now operate real software, not simulations of it. OpenClaw's official site describes it reading and writing files, running shell commands and scripts, browsing the web and filling forms, and managing Gmail, calendar and GitHub (source: openclaw.ai).
- **Memory:** agents keep durable, curated context. Hermes Agent pairs two core files injected into the system prompt, MEMORY.md and USER.md, with every session stored in SQLite with full-text search for cross-session recall (source: the Hermes Agent official docs on GitHub).
- **Long-running loops:** the agent improves between conversations. In Hermes, a background self-improvement review runs after each turn, extracting durable lessons into memory or reusable skills (source: Hermes docs).
- **Messaging as the interface:** no new app to learn. OpenClaw answers on 29 channels, including WhatsApp, Telegram, Slack, Discord, Signal and iMessage (source: openclaw.ai).

- **Price collapse:** OpenAI's flagship GPT-5.6 Sol costs $5.00 per million input tokens and $30.00 per million output tokens, while DeepSeek v4 Flash costs $0.14 and $0.28 (sources: OpenAI and DeepSeek official pricing docs, August 2026). That puts budget models at roughly 1 to 3 percent of frontier prices, which turns always-on agents from a luxury into a line item.
- **Open source:** complete agent stacks are free on GitHub, bring your own model. OpenClaw runs on Mac, Windows or Linux, and its site is explicit that "state lives on your machine, not a vendor cloud" (openclaw.ai).

The loop

Under the hood, every agent runs a version of the same cycle. It takes a **goal**, drafts a **plan**, picks the **tools** the plan needs, **acts**, **checks** the result against the goal, and **remembers** what it learned. Then it loops: if the check fails, it revises the plan and tries again, without waiting for a human to notice.

For an executive, the useful mental model is a capable junior with a to-do list, not an oracle. It is diligent, fast and tireless. It is also literal: it will pursue the goal you actually gave it, not the one you meant.

THE DOUBLE EDGE

The same autonomy that finishes the job unattended also makes mistakes at speed. An agent that can send fifty emails overnight can send fifty wrong ones. This is why the loop matters as much as the model, and why the later pages on governance exist.

How big is this already?

One marker of scale: OpenRouter, a routing layer that serves many models, reports on its official blog that machine-originated agentic traffic passed human traffic on its platform around 1 February 2026, at roughly 15 times the tokens per request. Software talking to software is now the majority of what these models do there.

29
Channels OpenClaw answers on (openclaw.ai)

15×
Tokens per request, agentic vs human traffic (OpenRouter)

$0.14
DeepSeek v4 Flash, per million input tokens

1 Feb 2026
Agentic traffic passes human on OpenRouter

Sources

Source	Supports	Link
OpenClaw official site	Tool use, 29 channels, platforms, local state quote	openclaw.ai
Hermes Agent official docs	MEMORY.md and USER.md memory files, SQLite recall, self-improvement loop	github.com/NousResearch/hermes-agent
OpenAI pricing docs	GPT-5.6 Sol at $5.00 and $30.00 per million tokens	developers.openai.com/api/docs/pricing
DeepSeek pricing docs	DeepSeek v4 Flash at $0.14 and $0.28 per million tokens	api-docs.deepseek.com/quick_start/pricing
OpenRouter official blog	Agentic traffic crossover, tokens per request	openrouter.ai/blog/insights/deepseek-v4-adoption

CHAPTER 02

EVIDENCE · ADOPTION

The numbers behind the fastest software wave in memory

Slides 4 and 5 make a large claim: that personal AI agents are the fastest-adopted software category in recent memory. This page is the footnote behind that claim. It follows two open-source agents, OpenClaw and Hermes Agent, both built outside every major AI lab, and sets their adoption curves against what the analyst houses say a boardroom should actually do about it. Every figure is sourced inline or in the table at the end.

OpenClaw: from weekend relay to 386,000 stars

OpenClaw did not start as a platform. It was first published on 24 November 2025 as Warelay, a small WhatsApp-to-Claude relay by Austrian developer Peter Steinberger, and was soon renamed Clawdbot (Wikipedia). On 27 January 2026 it became Moltbot after a trademark request from Anthropic over the name's similarity to Claude, the team quipping "Same lobster soul, new shell" (Laravel News, Forbes). Three days later, on 30 January 2026, it was renamed again to OpenClaw (Wikipedia). Three names in ten weeks is the tell: the project was moving faster than its own branding.

- **100,000 GitHub stars** in about two months from first publication (TechCrunch, 30 January 2026).
- **247,000 stars by 2 March 2026,** about five weeks after its late-January viral breakout (Wikipedia).
- **386,400 stars and 81,200 forks on 16 August 2026** (GitHub, fetched directly for this workshop).

That trajectory makes it one of the fastest-growing open-source projects in history. The stars are a proxy, not a revenue line, but as a measure of developer attention the curve has few precedents.

> COLOUR, WITH A CAVEAT
>
> In late January 2026, OpenClaw agents flooded onto Moltbook, a Reddit-style social network where only AI agents can post, launched by developer Matt Schlicht. The agents self-organised into communities, and Andrej Karpathy called it "genuinely the most incredible sci-fi takeoff-adjacent thing" he had seen recently (TechCrunch). The nuance matters: humans built the network, humans deployed the agents, and humans wrote the prompts. It is a demonstration of scale, not of autonomy.

Hermes Agent: the challenger that took the top rank

Hermes Agent was released on 25 February 2026 by Nous Research under the MIT licence (press reports). It arrived three months after OpenClaw and closed the gap in weeks, not years.

- **231,300 GitHub stars by 16 August 2026** (GitHub).
- **Overtook OpenClaw** on OpenRouter's global daily rankings as reported on 10 May 2026, at 224 billion daily tokens versus OpenClaw's 186 billion (MarkTechPost).

- **Number one daily global rank on OpenRouter** as of 16 August 2026, with 36.4 trillion tokens processed in the previous 30 days across 422 models (OpenRouter app page, fetched directly).
- **Nous Research reported in talks** to raise at least $75 million at a $1.5 billion valuation (TechCrunch, 13 July 2026).

386,400	231,300	36.4T	#1
OpenClaw GitHub stars, 16 Aug 2026	Hermes Agent GitHub stars, 16 Aug 2026	Hermes tokens on OpenRouter, prior 30 days	Hermes daily global rank on OpenRouter

Two projects, neither from a major lab, now sit at the top of a global model-routing marketplace. The lesson for a boardroom is not which agent wins. It is that the category moved from hobby code to infrastructure-scale traffic inside nine months.

The machine traffic crossover, and the price story

The single most important structural datapoint is the crossover. OpenRouter reports that agentic workloads passed human usage on its platform around 1 February 2026, with agentic requests consuming roughly 15 times the tokens per request of human ones (OpenRouter blog). Most traffic through that marketplace is now machines talking to models on behalf of people, not people typing.

Price pressure is part of the same story. DeepSeek roughly doubled its share of OpenRouter weekly token flow, from about 9 percent in January 2026 to about 18 percent by June 2026 (OpenRouter blog). Agents are heavy, repeat consumers of tokens, and heavy consumers shop on price. Cheaper capable models make agents cheaper to run, which creates more agents, which creates more demand for cheap tokens. The flywheel is already turning.

What the analysts say: the sober counterweight

The adoption curves above are real, and so is the friction. The analyst houses are worth quoting precisely because they cut both ways.

- **Gartner, on speed:** 40 percent of enterprise applications will feature task-specific agents by end 2026, up from under 5 percent in 2025 (Gartner press release).
- **Gartner, on failure:** separately, over 40 percent of agentic AI projects will be cancelled by end 2027, on cost, unclear business value and inadequate risk controls (Gartner press release).
- **McKinsey, State of AI (November 2025):** 23 percent of organisations are scaling agentic AI and 39 percent are experimenting, with no single business function above 10 percent scaled.
- **IDC:** agentic AI is forecast to reach $1.3 trillion, over 26 percent of worldwide IT spending, by 2029.

THE FRAME FOR SLIDES 4 AND 5

Read together, the numbers say one thing: adoption is outrunning governance. The same houses forecasting a trillion-dollar category are forecasting mass project cancellation for want of value cases and risk controls. That gap between adoption speed and governance maturity is the story of this workshop, not the star counts.

Sources

Source	Supports	URL
Wikipedia: OpenClaw	Warelay origin (24 Nov 2025), rename dates, 247,000 stars by 2 March 2026	en.wikipedia.org/wiki/OpenClaw
Laravel News	Clawdbot to Moltbot rename after Anthropic trademark request, "lobster soul" quote	laravel-news.com/clawdbot-rebrands-to-moltbot...
TechCrunch, 30 Jan 2026	100,000 stars in about two months; Moltbook and the Karpathy quote	techcrunch.com/2026/01/30/openclaws-ai-assistants...
GitHub: openclaw/openclaw	386,400 stars, 81,200 forks on 16 Aug 2026 (fetched directly)	github.com/openclaw/openclaw
GitHub: NousResearch/hermes-agent	Release under MIT licence; 231,300 stars on 16 Aug 2026	github.com/NousResearch/hermes-agent
MarkTechPost, 10 May 2026	Hermes overtaking OpenClaw: 224B versus 186B daily tokens	marktechpost.com/2026/05/10/openclaw-vs-hermes-agent...
OpenRouter app page	Hermes number one daily rank; 36.4T tokens over 30 days across 422 models (fetched directly)	openrouter.ai/apps/hermes-agent
TechCrunch, 13 Jul 2026	Nous Research in talks for at least $75M at a $1.5B valuation	techcrunch.com/2026/07/13/hermes-agent-maker-nous-research...
OpenRouter blog	Agentic traffic passing human usage around 1 Feb 2026 at roughly 15x tokens per request; DeepSeek share from about 9% to about 18%	openrouter.ai/blog/insights/deepseek-v4-adoption
Gartner press release, 26 Aug 2025	40% of enterprise apps with task-specific	gartner.com/en/newsroom/press-releases/2025-08-26...

Source	Supports	URL
	agents by 2026, up from under 5% in 2025	
Gartner press release, 25 Jun 2025	Over 40% of agentic AI projects cancelled by end 2027	gartner.com/en/newsroom/press-releases/2025-06-25...
McKinsey, State of AI (Nov 2025)	23% scaling agentic AI, 39% experimenting, no function above 10% scaled	mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
IDC FutureScape 2026	Agentic AI reaching $1.3T, over 26% of worldwide IT spending, by 2029	idc.com/resource-center/blog/futurescape-2026...

CHAPTER 03

SCENARIO 01 · MARKETING

Always-on brand and competitor intelligence

Most teams run competitor intelligence as a Monday scramble: someone opens the same set of tabs, skims what changed, and pastes fragments into a chat thread. An agent replaces the scramble with a nightly sweep that never skips a source, never forgets where it read something, and hands the team a drafted digest before the day starts. Every claim in that digest carries its source link.

The workflow, today and with an agent

Today the work is manual and partial. An analyst checks what there is time to check, and the sources that get skipped are exactly the ones nobody is watching. Coverage depends on who is in the office and what else landed that week.

With an agent, the sweep runs every night against a curated list. The team decides what matters; the agent does the walking.

- **Competitor pricing pages:** current prices and promotions, compared against what the agent recorded the night before.
- **Campaign launches:** new creative, landing pages and offers appearing on competitor sites and channels.
- **Ratings and reviews:** movement in scores and the themes surfacing in recent review text.
- **Store listings:** changes to product availability, descriptions and positioning on marketplace and delivery platforms.
- **Social chatter:** what is being said about the brand and its competitors on the monitored public channels.

The agent drafts a digest, flags what moved since the previous sweep, and delivers it to the team WhatsApp before 9am. Anything that did not change is noted as unchanged, not repeated.

> THE CORE DISCIPLINE
>
> Every claim carries its source link. If the digest says a competitor cut a price, the line links to the page where the price was read. The team never has to take the agent's word for anything.

How it is built, in plain words

There is no exotic machinery here. The build is five parts, and only one of them is software the team has not seen before.

- **A goal prompt:** a written brief describing what the agent watches for and what counts as worth flagging. It reads like an instruction to a junior analyst, because that is what it is.

- **A source list the team curates:** the exact pages, listings and channels the agent is allowed to visit. Nothing off the list gets swept.
- **Read-only browsing tools:** the agent can open pages and read them. It cannot post, buy, log in as the brand, or touch anything.
- **A delivery channel:** the digest arrives where the team already works, in this case WhatsApp.
- **A human editor:** a named person who owns the final digest, cuts what is noise, and adds the judgment the agent cannot.

A realistic weekly rhythm

Cadence	What runs	Who acts
Nightly	Full sweep of the source list; changes logged against the previous sweep	Agent only; run is logged
Each morning	Drafted digest with flags on what moved, delivered before 9am	Editor reviews, trims, releases to the team
Weekly	Deep-dive synthesis: what the week's movements add up to	Editor leads; agent drafts the synthesis
Monthly	Source-list review: what to add, retire or reweight	Team decision; the agent does not choose its own sources

What changes for the team

The analyst does not disappear. The analyst becomes the editor: the person who decides what the digest says, not the person who gathers what it says. That is a promotion in the work, not a replacement of it.

Coverage changes shape. It goes from "what one person could check" to "everything on the list, nightly". A beverage brand watching a price war no longer finds out about a competitor's move when a distributor mentions it; the move is in the morning digest with a link. A QSR chain tracking delivery promotions sees every platform listing change overnight, not the ones someone happened to open.

The judgment work is untouched and becomes more valuable. "So what" and "so now what" are still human questions. The difference is that the human answering them starts from a complete picture instead of a partial one, and spends the morning on the answer rather than the gathering.

Governance notes for this scenario

This scenario is deliberately low-risk, which is why it makes a good first deployment. The controls are simple and absolute.

- **Read-only tools:** the agent browses and reads. It has no posting or purchasing rights of any kind.
- **Source whitelist:** the agent visits only the pages the team has approved, and the list is reviewed monthly by people.

- **Logged runs:** every sweep is recorded, so any line in any digest can be traced to a run, a page and a timestamp.
- **A human before anything leaves:** nothing reaches anyone outside the team until the editor has released it.

ON COST

The costs here are volume work, and they are shaped accordingly. A budget model does the nightly sweeping, which is high-volume reading with low judgment. A frontier model writes the weekly synthesis, which is low-volume and judgment-heavy. Paying frontier rates for page-reading is the most common way this pattern gets needlessly expensive.

CHAPTER 04

SCENARIO 02 · RESEARCH

The research production layer, compressed

Agents do not replace insight. They compress the hours between fieldwork and insight. The craft of asking the right question, reading a market and standing behind a finding stays human. What changes is the production layer underneath it: the coding, the synthesis, the checking and the first draft, which today consume most of a study's calendar.

Four production tasks

These are the tasks that sit between a completed field and a delivered finding. Each one is real work today, and each one is where an agent earns its place.

Open-end coding

Thousands of verbatims are themed and counted overnight. The human role inverts: instead of building the codeframe from a blank page, the coder reviews a proposed codeframe against the data and corrects it. Quality assurance comes from disagreement sampling, where a human codes a sample blind and the gap between human and agent is measured, not assumed.

Transcript synthesis

Eight groups in, one findings document out. The discipline that makes this trustworthy is traceability: every claim in the synthesis links back to the transcript line it came from, so a moderator can audit any sentence in seconds rather than re-reading the night's tapes.

Questionnaire QA

Routing logic, translations and quota maths are checked without fatigue. An agent does not tire deep into a long script, and it applies the same scrutiny to the last translation as the first. Humans still own the questionnaire; the agent owns the tedium of verifying it.

First-draft reporting

The agent drafts the data story overnight: the structure, the charts described, the movements flagged. Seniors then spend their hours where they are irreplaceable, on the "so what" that turns a data story into a recommendation.

> THE PATTERN
>
> In every task the agent takes the production pass and the human takes the judgment pass. The deliverable is still authored by the researcher; it simply arrives at their desk further along.

The honest limits

Research houses trade on trust, so the limits matter as much as the capability. Four of them are non-negotiable.

- **Hallucination is real:** agents invent when unchecked, so traceability is not a nice-to-have, it is the design requirement. If a claim cannot point to its source line, it does not ship.
- **Codeframes need human judgment:** nuance, sarcasm and cultural register still defeat automated coding. A Malaysian respondent's "okay lah" is not a neutral code, and a human has to say so.
- **Confidentiality comes first:** nothing client-confidential goes to a model without a data agreement in place. The procurement conversation precedes the pilot, not the other way around.
- **A human signs every deliverable:** accountability does not delegate. The name on the report is a researcher's, and that researcher has reviewed what they are signing.

WHY IT MATTERS

A research house that automates production without these guardrails is not faster, it is unaccountable. The guardrails are what make the speed sellable.

Where the value moves

When production compresses, it does not disappear; its value migrates. Judgment and design appreciate. The questionnaire architect, the moderator who hears what was not said, the analyst who knows which movement is noise: these roles become more valuable, not less, because their hours are no longer spent on mechanical work. The teams that win will be the ones whose analysts direct fleets of agents rather than compete with them.

The Malaysian context sharpens the timing. AI adoption among Malaysian businesses reached 27 percent in 2025, according to a vendor survey by AWS and Strand Partners of 1,000 business leaders. But depth is shallow: 73 percent of adopters remain at basic use, and only 10 percent have reached advanced integration. That gap between broad adoption and shallow depth is the opportunity. Research teams that build real agentic depth early will be operating in a market where most of their competitors, and most of their clients, have not.

27%
Malaysian businesses adopting AI, 2025

73%
Of adopters remain at basic use

10%
Have reached advanced integration

1,000
Business leaders surveyed, AWS and Strand Partners

Sources

Source	Supports	Link
TechNode Global, 5 November 2025, reporting the AWS and Strand Partners survey	Malaysian AI adoption figures: 27 percent adoption, 73 percent basic use, 10 percent advanced integration, 1,000 business leaders surveyed	technode.global

CHAPTER 05

SCENARIO 03 · PERSONAL

A chief of staff in your pocket

Of the three scenarios in this workshop, this is the one every leader in the room can start this quarter. A personal agent that lives in WhatsApp, reads what you let it read, and handles the connective tissue of a senior job: triage, preparation, capture and follow-through. It is also the lowest-risk way for a leader to learn, personally, what agents can and cannot do before any organisation-wide decision has to be made.

A day with it

The value is not one dramatic capability. It is the steady removal of small coordination work, hour after hour, in the channel you already live in.

- **Morning triage:** before you open your inbox, the agent has sorted it into three piles: needs-you, drafted-for-you, and noise. You read the first pile, review and release the second, and never see the third.
- **Meeting preparation:** for each meeting on today's calendar, a one-page prep: who is in the room, what was agreed last time, what is still open, and what you said you would bring.
- **Capture on the move:** a voice note recorded in the car becomes a structured brief by the time you park: context, decision needed, options, next steps, ready to forward or file.
- **Follow-through:** the agent keeps a ledger of promises made, yours and other people's, and nags. "You told the finance lead you would respond by Thursday" is a message you get on Wednesday.

> THE PATTERN
>
> Every item above is the same loop: the agent reads sources you have granted, produces a draft or a reminder, and you stay the decision point. Nothing goes out that you did not release.

What it needs, and what that means

To do the job described above, the agent needs real access: your email, your calendar, your files, and a messaging channel such as WhatsApp to reach you. That access is precisely why the setup must be deliberate. A chief of staff is useful because it sees everything; that is also the risk to manage.

What it needs	Why it needs it	What that obliges you to decide
Email access	Triage, drafting, promise tracking	Read-only first; sending rights are a separate, later grant
Calendar access	Meeting preps, scheduling awareness	Whose calendars it may see beyond your own
File access	Context for briefs and preparation	Which folders are in scope, which are excluded
Messaging channel	The interface you actually use	Which conversations, if any, it may read versus only write to you

Behind those grants sit three setup decisions that should never be defaults. Use your own API key or a governed company instance, not a free consumer account, so usage is attributable and terms are known. Know what data leaves the device, and to which model provider it goes, before the first message is processed. And decide where the agent runs: your own instance, or one your organisation governs.

THE UNCOMFORTABLE TRUTH

An agent with your inbox is the most privileged assistant you have ever hired, and it was never interviewed. Deliberate setup is the interview.

Starting safely as an individual

The staircase for an individual leader is short and conservative. Each step earns the next.

- **Start read-only.** Triage, preps and briefs need no sending rights at all. Weeks of value are available before the agent can act on anything.
- **Add sending rights only with per-message approval.** The agent drafts; you release. Autonomy over outbound messages is a later decision, if it is ever taken.
- **Keep client-confidential material out** until your organisation has a policy that covers it. Your own diary and drafts are yours to risk; a client's data is not.
- **Prefer a governed setup over a laptop experiment.** A personal instance running unmanaged on a laptop is how shadow IT starts. A governed instance gives you the same learning with an audit trail.

Why leaders should go first, personally

The organisational decisions ahead, which workflows to automate, what autonomy to grant, what governance to require, will be made by people in this room. A leader who has lived with an agent for a quarter makes those calls from experience: they know where it saved an hour, where it confidently got something wrong, and what supervision actually feels like. A leader who has not is deciding from vendor slides. The personal chief of staff is not just a productivity tool; it is the cheapest management education on agents available, and it is available now.

THIS QUARTER

Read-only, own key, no client data, per-message approval on anything outbound. That configuration is safe enough to start now and instructive enough to change how you decide later.

CHAPTER 06

HANDS ON · LIVE DEMO

Two agents, one experiment: feel the difference

Two live agents are running in a sandbox right now, both reachable on WhatsApp. They have the same tools and the same Malaysian datasets. The only difference is the brain: one runs a frontier-class model, the other a budget model. Message both, ask the same question, and compare what comes back. This page is your guide while the demo runs.

The two agents

Agent GAJAH is the heavyweight. It runs a frontier-class model (OpenAI GPT-5.6 class), priced by OpenAI at $5.00 per million input tokens and $30.00 per million output tokens.

Agent KANCIL is the lightweight. It runs a budget model (DeepSeek v4 Flash), priced by DeepSeek at $0.14 per million input tokens and $0.28 per million output tokens.

Agent	Model	Input, per million tokens	Output, per million tokens
GAJAH	Frontier-class (OpenAI GPT-5.6 class)	$5.00	$30.00
KANCIL	Budget (DeepSeek v4 Flash)	$0.14	$0.28

Both agents carry identical tools and identical data. Only the brain differs. That is the whole point of the experiment: any difference you feel in the replies is the model, not the plumbing.

> WHERE TO MESSAGE THEM
>
> GAJAH: aiagent.research.my/wa1 • KANCIL: aiagent.research.my/wa2. Prefer Telegram? The same agents are mirrored at /tg1 and /tg2.

What to try

Ask anything about the Malaysian market. If you want a starting point, these questions exercise the datasets well:

- Compare median household income in Johor Bahru and Shah Alam.
- How many cafes are there in Penang?
- Which district in Selangor has the highest spending power?
- Profile the businesses within 2km of our venue.
- Where in Malaysia are pharmacies growing fastest?

The most useful move: send the same question to both agents, then compare. Look at three things. Depth: which answer goes further into the data? Caution: which agent flags what it does not know? Speed: which one comes back first? The differences are not subtle.

The rules of the sandbox

These rules are not decoration. They are a small, live example of the governance principles on the governance slide: constrain what an agent can reach, log what it does, and put a time limit on it.

- **Replies take 30 to 60 seconds,** and longer when the room is busy. The agents plan, query and check before answering; that takes time.
- **Both agents are sandboxed** on isolated infrastructure, with read-only access to public Malaysian datasets.
- **They cannot act on the world.** No email, no spending, no reach into any company system.
- **Conversations are logged** for the demonstration and deleted after. Do not send personal or confidential information.
- **The agents stay live for one week** after the workshop, so you can keep testing from your desk.

WHY THE LIMITS MATTER

Every constraint here (isolation, read-only data, no outbound actions, logging, an expiry date) is a control you would want on a production agent too. The sandbox is the governance model in miniature.

What you are feeling

Two gaps at once. The capability gap: the frontier model tends to reason further, hedge more honestly and handle messier questions. The cost gap: the budget model does respectable work at a fraction of the price. Neither agent is simply better. The lesson for an insights or marketing leader is that model choice is a management decision, not a technical one: match the model to the job, and pay for frontier reasoning only where the job demands it.

Sources

Source	Supports	URL
OpenAI pricing	GAJAH model pricing: $5.00 input, $30.00 output, per million tokens	developers.openai.com/api/docs/pricing
DeepSeek pricing	KANCIL model pricing: $0.14 input, $0.28 output, per million tokens	api-docs.deepseek.com/quick_start/pricing

CHAPTER 07

HANDS ON · DATA

What the demo agents actually query

An agent is only as good as the data it can reach. The agents in this demo do not browse the open web and hope. They query structured data assets: normalised, indexed, and hosted where we can vouch for them. That is the same posture an agent inside your business would take with your own data. The figures on this page were measured on the live system, August 2026.

Official statistics, normalised

The first asset is Malaysia's official statistics, drawn from OpenDOSM, the Department of Statistics Malaysia's open-data platform. OpenDOSM catalogues and opens DOSM's data with free API access, which makes it a clean foundation for an agent: authoritative figures, published methodology, and a stable interface to query.

Inside the demo system that foundation has been normalised into query-ready tables. Measured on the live system in August 2026, it holds 612,624 population fact rows, alongside household income, expenditure, poverty and Gini coefficients for 999 geographies. Coverage runs from state level down through district, parliamentary and state assembly (DUN) level, so an agent can answer at the granularity the question demands.

> OPENNESS AS POLICY
>
> OpenDOSM's own tagline puts it well: "If data is the new oil, then openness is the pipeline that maximises its value." The demo takes that pipeline at its word.

A Southeast Asian business census

The second asset is a geocoded census of businesses across the region: 17,166,301 listings across eight markets, measured on the live system, August 2026. Each listing carries categories, coordinates and contact points, and the asset holds 27.9 million contact records in total, including 5.6 million email addresses.

Market	Geocoded business listings
Indonesia	9.1 million
Thailand	2.2 million
Malaysia	1.9 million
Philippines	1.8 million
Vietnam	1.3 million
Singapore	561,000
Myanmar	108,000
Cambodia	106,000
All eight markets	**17,166,301**

For an agent this means a market question ("how many pharmacies within five kilometres of this site") resolves against a real, located universe of businesses rather than a scraped guess.

A live property pipeline

The third asset is not a snapshot but a pipeline: continuous capture of Malaysian property listings, prices and media from public portals, refreshed continuously. Where the census answers "what exists", the pipeline answers "what is on the market right now, and at what price".

Because capture runs continuously, the agent's view of the property market ages in hours, not quarters. That matters for any question where timing is the point: pricing a launch, reading a corridor, or watching how supply responds to an announcement.

Why this matters for the demo, and for you

Structured, owned data beats open-web browsing on the two things that decide whether an agent is usable: reliability and speed. A query against a normalised table returns the same answer every time, in seconds, with a lineage you can inspect. A browse of the open web returns whatever the web happened to serve that day.

612,624
Population fact rows, normalised from OpenDOSM

999
Geographies with income, poverty and Gini figures

17,166,301
Geocoded business listings, eight markets

27.9M
Contact records, including 5.6 million emails

The second point is sovereignty. Everything above is hosted in Malaysia on infrastructure BASIC controls, and these assets are BASIC's own. When you ask the demo a question, that question never leaves the system for a third-party data broker. The same principle applies when the data being

queried is yours: the agent comes to the data, the data does not go to the agent's vendor.

THE TAKEAWAY

The demo is not impressive because the model is clever. It is useful because the data underneath is structured, current, and held on infrastructure the operator controls. Your version of this starts with your data, treated the same way.

Sources

Source	What it supports	URL
OpenDOSM	Official statistics platform behind the normalised population and household tables; all row and geography counts were measured on the live system, August 2026	open.dosm.gov.my

CHAPTER 08

THE HONEST PART · RISK

The downside, documented

Everything that makes an agent useful is also the attack surface. The tools it can call, the memory it keeps, the autonomy it is granted: each one is a capability and a liability at the same time. This page is the record of what has actually gone wrong, with dates and sources, because a risk story without dates is just mood.

Prompt injection: the attack class is real

In June 2025, Aim Security disclosed EchoLeak (CVE-2025-32711, CVSS 9.3), the first known zero-click prompt injection against a production AI agent. A single crafted email could make Microsoft 365 Copilot exfiltrate data from mail, OneDrive, SharePoint and Teams, with no user interaction at all. The victim did not have to click anything. The agent read the email, followed the instructions hidden inside it, and did the rest.

The honest framing matters here. Microsoft patched EchoLeak server-side and found no evidence of exploitation in the wild (The Hacker News). Nobody was breached by it, as far as anyone knows. But it proved the attack class is real: an agent that reads untrusted content and holds real permissions can be steered by that content.

> THE PRINCIPLE
>
> The Five Eyes guidance puts it plainly: system prompts are instructions, not controls. Telling an agent to behave is not a security boundary. The boundary has to live outside the model, in what the agent is permitted to reach.

Exposure at scale: the self-hosted wave

When agent frameworks went mainstream, the exposure went with them. In January 2026, researcher Jamieson O'Reilly found hundreds of self-hosted Clawdbot instances exposed via Shodan, eight of them completely open with no authentication, risking months of private messages, credentials and API keys (The Register, 27 January 2026). By 9 February 2026, SecurityScorecard's STRIKE team counted 40,214 exposed OpenClaw instances across 28,663 IP addresses, of which 12,812 were exploitable via remote code execution (Infosecurity Magazine).

The software itself had a serious hole in the same window. CVE-2026-25253 (CVSS 8.8) allowed one-click remote code execution via a malicious link; it was patched in the 30 January 2026 release (The Hacker News). The ecosystem around the software fared no better. Snyk scanned 3,984 community skills and found 283, about 7.1 percent of the registry, with critical flaws, including instructions that passed API keys and card numbers through the model in plaintext (Snyk). And O'Reilly demonstrated the supply-chain risk directly: he published a deliberately harmless poisoned skill, and developers from seven countries downloaded it (The Register).

40,214
Exposed OpenClaw instances counted by 9 February 2026

12,812
Exploitable via remote code execution

283
Community skills with critical flaws, of 3,984 scanned

7
Countries whose developers downloaded one poisoned test skill

The shadow agent problem

The largest risk in most organisations is not the sanctioned agent. It is the one nobody approved. IBM's Cost of a Data Breach Report 2025 found that one in five breached organisations traced the breach to unsanctioned "shadow AI", at an average of USD 670,000 in extra breach cost. Of the organisations reporting AI-related breaches, 97 percent lacked proper AI access controls, and 63 percent had no AI governance policy at all (reported by Cybersecurity Dive).

The behaviour behind those numbers is well documented. A PagerDuty survey of 1,250 office professionals (June 2026) found that 66 percent had used AI tools at work despite believing it was against policy, and 88 percent shared work information with public AI tools. Policy on paper is not policy in practice.

Two further data points come from vendor surveys and should be read as such. Gravitee's State of AI Agent Security 2026, a vendor survey of 900+ respondents, found 88 percent reported confirmed or suspected agent security incidents, and only 14.4 percent said all their agents went live with full security approval. A Cloud Security Alliance survey commissioned by Zenity, a vendor in this space, found 54 percent of organisations report unsanctioned shadow agents in their environment.

The institutions have noticed

On 1 May 2026, CISA, the NSA and the cyber agencies of the UK, Australia, Canada and New Zealand jointly published "Careful Adoption of Agentic AI Services", the first joint Five Eyes guidance on AI agents. It names five risk categories and urges strict least privilege, low-risk first use cases, and full traceability of every agent action. When six national security agencies co-sign a document about a technology, the technology has stopped being a curiosity.

THE CONCLUSION

None of this argues for a ban. It argues for governance. A ban does not remove the agents from your organisation; it keeps them invisible, unmonitored and unpatched, which is exactly the condition every incident on this page grew out of.

Sources

Source	Supports	Link
The Hacker News	EchoLeak, CVE-2025-32711, CVSS 9.3, zero-click exfiltration, patched with no exploitation found	thehackernews.com
The Register, 27 January 2026	Exposed Clawdbot instances, eight with no authentication; the poisoned-skill demonstration	theregister.com
Infosecurity Magazine	40,214 exposed OpenClaw instances, 28,663 IP addresses, 12,812 exploitable via RCE	infosecurity-magazine.com
The Hacker News	CVE-2026-25253, CVSS 8.8, one-click RCE, patched 30 January 2026	thehackernews.com
Snyk	3,984 community skills scanned, 283 with critical flaws, about 7.1 percent	snyk.io
Cybersecurity Dive, on IBM 2025	One in five breaches traced to shadow AI, USD 670,000 extra cost, 97 percent and 63 percent governance gaps	cybersecuritydive.com
PagerDuty, June 2026	1,250 office professionals: 66 percent used AI against policy, 88 percent shared work information	pagerduty.com
Gravitee (vendor survey)	88 percent reported incidents, 14.4 percent full security approval, 900+ respondents	gravitee.io
Cloud Security Alliance, commissioned by Zenity	54 percent of organisations report unsanctioned shadow agents	cloudsecurityalliance.org
CISA and Five Eyes partners, 1 May 2026	"Careful Adoption of Agentic AI Services", five risk categories, least privilege, traceability	cisa.gov

CHAPTER 09

THE PLAYBOOK · GOVERNANCE

Governing agents: the six rules and the first ninety days

The goal of the first ninety days is not transformation. It is an informed decision, made with your own evidence: what agents are worth to your organisation, on your workflows, under your controls. The six rules below make the pilot safe to run. The ninety-day plan makes the decision honest.

The six rules

Every agent that touches company work runs under all six. They are infrastructure choices, not policy aspirations, which is what makes them enforceable.

1. Sandboxed

Agents run on isolated infrastructure set up for the purpose, never on personal laptops. A sandbox limits what a misbehaving agent can reach, and it makes the environment reproducible when something needs investigating.

2. Least privilege

Agents get read-only access to data and a whitelist of tools, nothing more. The Five Eyes guidance on agentic AI puts it plainly: system prompts are instructions, not controls. The real controls live in permissions and infrastructure, so that is where the boundaries are set.

3. Approval gates

A human approves anything irreversible before it happens: send, publish, pay, delete. The agent can draft, queue and recommend, but the consequential click belongs to a person.

4. Logged

Every message, every tool call and every data access is recorded. Logs are what turn an incident into a lesson rather than a mystery, and they are the evidence base a regulator or a client will eventually ask for.

5. Right-sized models

Frontier brains for judgment, budget brains for volume. Most agent steps are routine and do not need the most capable model. The workshop demo made the cost gap between the two tiers tangible.

6. Data stays home

Company and client data stays in your jurisdiction, on infrastructure you control. Where the model runs and where the data rests are decisions you make deliberately, not defaults you inherit from a vendor.

NOT HYPOTHETICAL

The two agents demonstrated in this workshop run under all six rules. The governance is not a slide; it is the operating condition of everything you watched.

The first ninety days

Three phases, twelve weeks, one decision at the end. Each phase has a concrete output, so progress is visible and the final call is grounded in measurement rather than sentiment.

Phase	Weeks	What happens	Output
See	1 to 2	Inventory what already runs, amnesty included: staff declare the tools they use without penalty. Then write the policy.	A two-page policy: allowed, banned, needs-approval.
Pilot	3 to 8	One governed pilot. One workflow, real data, a task already measured in hours, humans in the loop throughout.	A running agent under all six rules, with logs.
Decide	9 to 12	Measure hours saved, quality and incidents. Scale what worked, kill what did not.	A published internal decision, so shadow agents come into the light.

The amnesty in week one matters more than it looks. People are already using these tools quietly. An inventory that punishes honesty produces a fictional inventory; an amnesty produces the real one, which is the only one worth governing.

The Malaysian context

The government is moving on this. Malaysia's National AI Office launched on 12 December 2024. On 28 July 2026 the Prime Minister launched AI Malaysia in Cyberjaya as the permanent apex body for national AI governance, alongside the National AI Action Plan covering 2026 to 2030, which carries 28 initiatives toward the goal of "AI Nation by 2030". An AI Governance Bill is being drafted, targeted for completion by the end of 2026. All of this is documented in official releases from the Ministry of Digital Malaysia.

THE PRACTICAL READ

A regulatory framework is imminent. Organisations that can already show governed agent use, with a policy, a sandbox and logs, will be ahead of the framework when it lands. Everyone else will be scrambling after it.

Where to start on Monday

Start with the smallest real step: name an owner, run the amnesty inventory, and draft the two-page policy before the month is out. Pick one workflow your team already measures in hours and scope it as the pilot. None of this requires a budget line or a transformation programme; it requires a decision

to look at the question with your own evidence rather than someone else's slides. BASIC builds and hosts governed agent deployments on Malaysian infrastructure and is happy to compare notes.

Sources

Source	Supports	Link
Ministry of Digital Malaysia, official release	National AI Office launch, 12 December 2024	digital.gov.my
Ministry of Digital Malaysia, official release	AI Malaysia launch, National AI Action Plan, AI Governance Bill timeline	digital.gov.my
CISA with US and international partners	Guidance on secure adoption of agentic AI; system prompts as instructions, not controls	cisa.gov

THANK YOU

Keep the agents.
Keep them governed.

The demo agents stay live for one week after the workshop. The deck, this brief and every source live at aiagent.research.my.

THIS WORKSHOP

aiagent.research.my

ENQUIRIES

info@basicinsight.my

WEB

basicinsight.my