OCT 26, 2026

The Agent Land Grab

The Race to Turn Subsidized Intelligence Into a Lasting Moat



By Jay Zhao, Liang Wu, and Jenny Xiao

Leonis Capital · leoniscap.com/research/the-agent-land-grab/



On September 28, Instinct announced a $1 billion round at a $10 billion valuation, just a month after a reported $2.5 billion valuation. Its personal agent was still free for invite-only users. Investors committed another billion dollars to a product that had yet to charge the people using it.

Instinct is the largest bet in a crowded field. Town raised $108 million in its latest round to build a personal assistant that learns how people work, while Viktor raised $75 million for an AI coworker embedded in teams' communication tools. The largest platforms have entered as well, with Meta launching Muse, OpenAI launching Dots, xAI offering Grok Bot to its subscribers, and Anthropic previewing Claude Duet, an agent built for work. Across personal life and work, an agent land grab is underway.

The consensus view is that these rounds are hype, with investors overpaying for interchangeable products built on the same rented models. Whether they are depends on where the capital goes. Agent companies are running a familiar internet growth playbook, in which they subsidize the product, build a habit, and monetize later. **The difference between agent companies and internet-era companies is that the subsidy now pays for the agent's work itself, through compute.** That makes every round a bet on token economics. Agents burn far more tokens than chat and keep moving to pricier frontier models, and their backers are counting on token prices to fall faster than usage grows. Until prices catch up, the question is what the subsidy buys in the meantime.

Capital spent on the agent's work builds trust, and trust compounds. Spent only on attention, the capital buys a growth chart of vanity metrics. Personal and team agents then collect on that trust in different ways. Personal agents earn a share of the commerce their users route through them, while team agents charge the business directly.

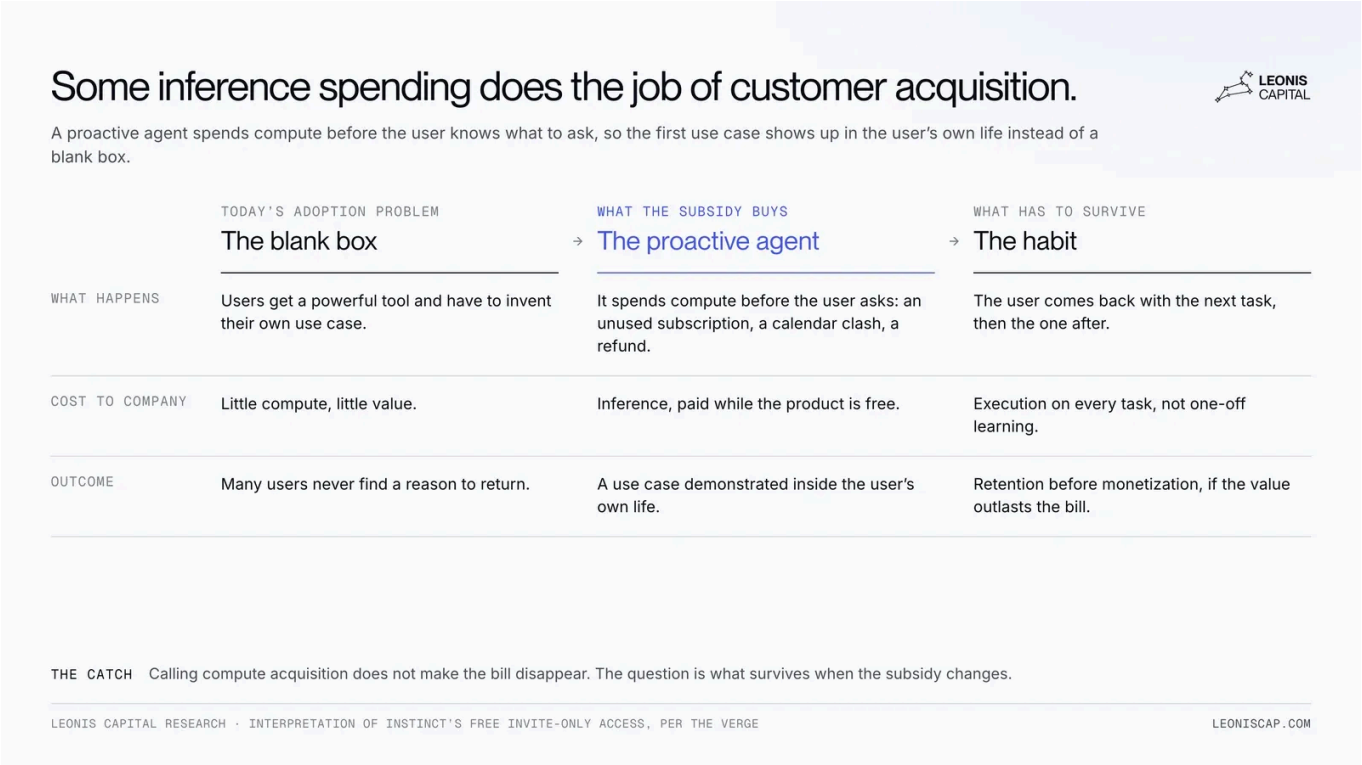
The race is to earn that trust before competitors do. An agent company that earns it first raises the amount of work a later entrant has to do, because the challenger has to win back responsibilities that already run smoothly. In this essay, we look at how token economics shapes the subsidy, what it builds when it is spent well, what that position is worth, and what it takes to win.



Agent Subsidies Are a Bet on Token Economics

Most internet-era subsidies sat outside the product, in free access, paid marketing, and referral bonuses. Occasionally it sat inside. When Google launched Gmail in 2004, it gave every free account one gigabyte of storage, at a time when Yahoo offered four megabytes. Gmail was also invite-only, like Instinct today, and the storage helped Google win the inbox.

Agents put the subsidy inside the product, and what they subsidize is work, paid for in tokens. A proactive agent spends compute finding something useful in the context a user has shared and doing it before the user knows to ask, such as catching an overlooked refund or preparing for an upcoming appointment. That is a better pitch than the blank chat box most AI products still open on, which asks users to invent their own use case.



Gmail was a bet that storage would keep getting cheaper, and agent companies are making the same bet on compute. The price of a given level of model performance is falling fast, the Stanford AI Index highlighted that inference at GPT-3.5's level became more than 280 times cheaper in about 18 months. The catch is that an agent's bill is not a fixed allowance like Gmail's gigabyte of storage, it grows with every task. Agents used about 4x as many tokens as chat, and multi-agent systems about 15x. The bet is that collapsing token prices more than cover the jump, so cost per user stays flat or even falls as the agent takes on more work. But that holds only if prices fall faster than usage grows and the cost of new model increases.

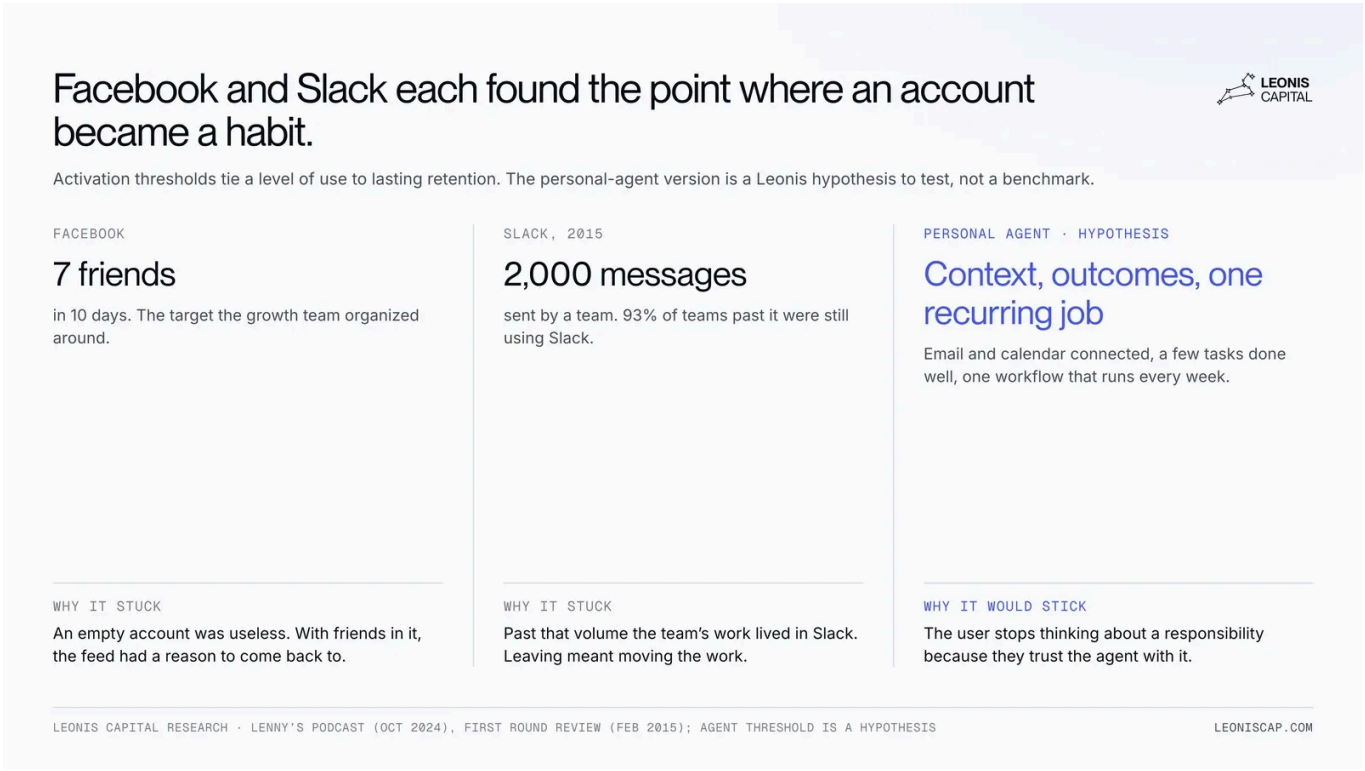
Team agents have a second lever that personal agents mostly lack. They launch on the best frontier models to earn a team's trust, then move repeatable work, such as triaging tickets or drafting routine updates, onto cheaper and open-source models once quality holds. Work inside a company is more repetitive and easier to measure than a person's errands, so a team agent can shrink its own bill instead of waiting for token prices to fall. Personal agents depend more on the market doing that for them.

Venture capital raised serves as a bridge. It pays for the agent's work until compute is cheap enough for the economics to close, and what matters is what the company has built by the time it gets there.

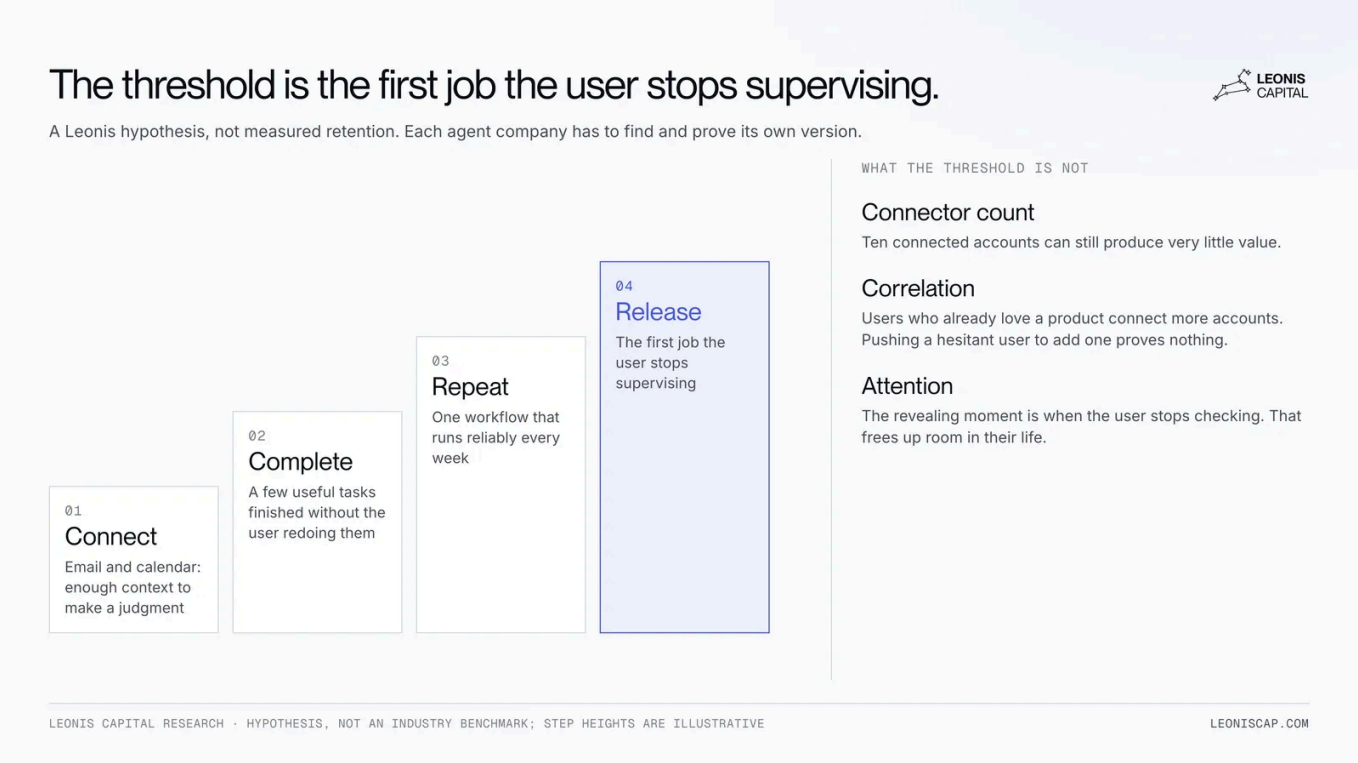
How capital raised is spent becomes the central question for every company in this land grab. Money raised can go into the "front end" of go-to-market, into paid social and other channels that keep getting more expensive. It can also go to the "back end", into the discovery work that turns a trial user into someone who delegates more work to their agent. That is inference spending working as customer acquisition, except the user gets the work. **Front-end spending buys sign-ups that last only as long as the spending does, while back-end spending compounds into a moat and network effects.** A company that raises large rounds of capital and spends mostly on the front end looks healthy until the market corrects or the cash runs low.

Trust Compounds Through Working Together

Every consumer platform has had to find the point at which a new account becomes a habit. Facebook and Slack turned that point into a concrete growth objective by identifying an activation threshold, a level of participation associated with lasting use: seven friends in ten days for Facebook and 2,000 messages for Slack. In our OpenClaw essay, we described the technical threshold at which model capability makes a new class of product possible. **Personal agents now face an activation threshold**, the point at which a particular user moves from trying the product to trusting it with work.



A personal agent reaches its activation threshold when it has enough context to do useful work and earns a recurring responsibility. The most compelling moments come when **an agent makes a judgment the user would otherwise have needed to explain**. Say a delayed flight puts two meetings at risk, and the user would happily move an internal check-in to preserve a customer conversation. On day one, the agent needs that distinction explained. By day 100, having learned from similar decisions and corrections, it knows which meeting to protect, and each judgment it gets right earns it more responsibility and less supervision.

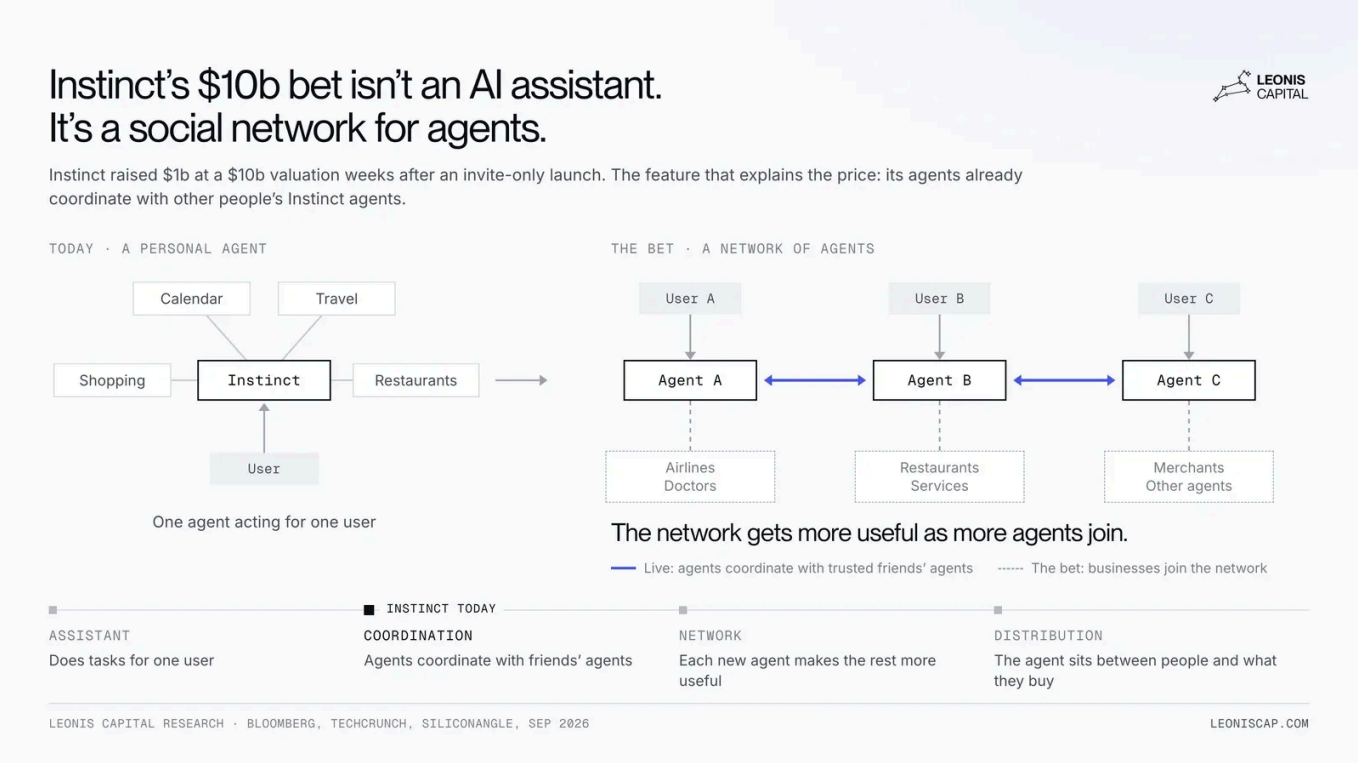


That earned trust is the moat. A rival can connect to the same inbox, calendar, and accounts, and even import a user's preferences, but connectors only supply context. **What it cannot quickly reproduce is the record** of successful outcomes an agent earns by working with one person. Switching means teaching a new agent and deciding again whether to trust it, **which is why two agents built on similar models diverge** on how well they understand the user and how much they are allowed to do. At work, that trust is organizational as well as personal. Admins grant permissions, teams build workflows around the agent, and replacing it means retraining a whole team rather than one person. That higher switching cost drives businesses to keep paying for an agent they already rely on.

Agent Network Effects Widen the Moat

The moat widens once agents start working with each other. Scheduling, planning, and shared decisions require reconciling several people's priorities, and as more people delegate them, their respective agents need to coordinate. In 2023, our research predicted that multi-agent systems, with specialized agents working together, would outperform single agents.

Instinct has now built coordination into its pitch through its Trusted Person Network, which lets one user's agent coordinate directly with another's. Each additional participant gives existing users another person whose agent they can work with.



Agents from different providers can talk to one another, but coordination works best between agents on the same platform. Two agents on the same service settle a meeting by checking availability and preferences directly, under permissions both users have already granted, while an agent coordinating with an agent on a competitor's platform falls back to the back-and-forth people send today. **People will cluster with the agent companies** their colleagues, friends, and households already use, and the network compounds the individual moat.

How the Subsidy Gets Paid Back

What the subsidy ultimately buys is becoming the default. For a category of decisions, one agent receives the request first, already has the context, and holds permission to act. No single agent is likely to handle everything, since work, personal administration, and sensitive decisions will stay with different providers, but the default within one important domain is valuable on its own.

For personal agents, becoming the default opens up a business model **built on take rates instead of subscriptions.** An agent handling travel, household services, or routine shopping repeatedly chooses which businesses get paid, and the agent company earns a share of the commerce that flows through it. That creates a **virtuous cycle in which the agent stays free for the user**, more delegation routes more transactions through the agent, and those transactions fund the work that earns further trust.

However, that payoff is not automatic. Transaction revenue grows with the spending users route through the agent, while the cost of serving them grows with everything the agent does. At today's take rates, transaction fees can pay for a light agent. A heavy agent pays for itself only when its user routes a large share of their spending through it.



The agent's influence over spending also only lasts while users believe their agent is prioritizing them versus any merchant. If merchant payments distort its recommendations, users go back to checking its work, and the moat that made the influence possible starts to erode.

Team agents get paid back differently. The business pays for the agent directly, and the agent bundles model access into one product, so the customer buys finished work rather than choosing models. Once a team has built its workflows around the agent, it keeps paying, and spends more as it delegates more, on frontier and open-source usage alike. The margin comes from the spread between what the customer pays and what the work costs to serve, and that spread widens as routine work moves to cheaper models. The risk mirrors the consumer one. If quality slips when the agent switches models to protect its margin, the team notices before the margin does.

What It Takes to Win the Agent Land Grab

The largest platforms start closer to the activation threshold. Google, Microsoft, and Apple already hold the inbox, the calendar, and the device, and Meta made Muse free for most

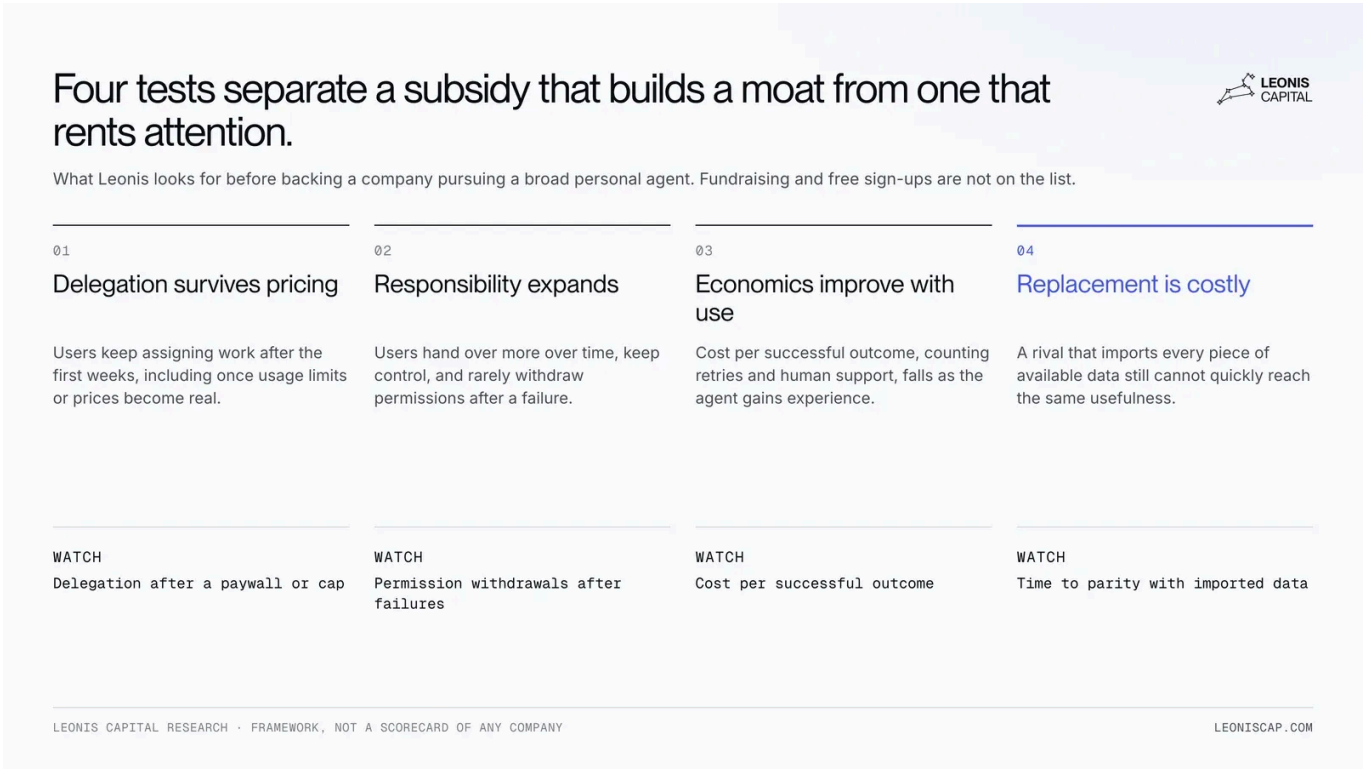
everyday use when it launched. An incumbent can supply the context for an agent's first useful judgment without asking anyone to connect an account, and it can fund the subsidy from existing profits. At work, Microsoft, Google, and Slack can bundle agents into seats their customers already pay for.

A startup can still win the default in a domain by executing reliably within it, accumulating corrections and responsibility faster than a general assistant spread across billions of users. Alignment is another opening, since a company whose core business is advertising has more to prove when it asks users to believe its agent's purchases are made purely on their behalf.

Fundraising alone will not decide which of them wins. Neither will the usual engagement metrics. On a social app, more messages and more time spent mean the product is working, but with an agent they often mean the opposite, since a user who trusts their agent stops checking in while a user correcting a weak one keeps messaging it.

Every subsidy in this race ends one of two ways: token prices fall far enough for the economics to make sense, or the cash runs out first. What separates the winners is how users behave when that moment comes.

In the winning companies, users keep delegating after the first weeks, including once usage limits or prices become real, and they hand over more responsibility over time without withdrawing permissions after a failure. The economics improve with use, as cost per successful outcome falls as the agent gains experience and tokens get cheaper, even while users hand it more work. And a rival that imports every piece of available data still cannot quickly reach the same usefulness, which is the clearest evidence that the trust was earned and the moat is real.



Some companies in this land grab will build relationships that pass these tests. Others will discover that their relationship with the user ended at the subsidy.

A billion dollars can fund a lot of inference, but the agent still has to earn a place in someone's life.