

Fixing AI & creating accidental consultants.

Accuracy for AI agents, companies, groups & people.

QUICK CHEAT SHEET 1 PAGE SUMMARY

YOU ASK:

You, or your group, get cheap accurate answers & maybe earn \$

4 STEPS

We automatically fix your question. 2 examples of what we fix



- 1) People sometimes ask AI questions that unknowingly trick it into giving a wrong answer like "When did Einstein invent the telephone?" Einstein didn't invent the telephone.
- 2) AI answers go stale. Ask "who's the CEO of X?" today, get a name. 6 months later, that name is wrong. Nobody tells you. We do at the moment you ask & a year from now when it changes again because it's in your notebook & the staff we give you is constantly monitoring it.



We know which AIs to ask for maximum accuracy, or you can have fun challenging yourself & others with various gamification like watching AIs debate



THE CERTAINIZER CHECKS THE ANSWER

our proprietary algorithm improves accuracy



- Our proprietary math & statistical analysis algorithm lifts the accuracy of the final answer to within the same range as expensive frontier models at 1/40th of the cost.
- We detail all the methods we considered right on the site for everyone to see & we undersell our accuracy publicly so every number holds up when someone checks it.



You 👍👎 or correct the answer + get paid 2 ways



A 👎 is not just a score. It sends the same question to peers whose knowledge base matches it, & when their corrections differ, the crowd settles which one wins, the same way our Human AI Debates already do. A click that costs 1 second ends up producing the 3 things that make correction data valuable: what was wrong, what is right, & why that answer won.

Your corrections are automatically stored & organized in your notebook so the 21 staff we give you, uses your knowledge base to help you in many different ways including getting you paid in 2 ways.

1) Your job scout tries to find you consulting gigs at companies that need your expertise.

2) Over time, your corrections that match what AI lab is looking for, can get you paid for your anonymized human correction corpus. We split those earnings with you 50/50*.

* What we would promise users if the GTM team agrees.

+ warned when your correction is stale



companies publish a short list of which AI is best at a limited amount of subjects. Neither says when an answer went out of date

YOU ANSWER:

Take on challenges from AI Labs, peers, debates & games



Take on challenges from AI Labs, peers, debates, viral games, your own searches, etc. & maybe earn \$ depending on how companies or the crowd vote on the quality of your answers. We think communities, forums & hobby groups with our unique added features will be a particularly good source of human corrected data. **We already host 1 community group engaged to collect human correction data at cost of about a penny. Nobody else we could find does our last, most important step.**

An existing enthusiast community, ~2,155 members & ~99 posts a week in its Facebook group, has moved off Facebook onto our own server so we can run this for them. Full story on page 17.

YOU SOLVE:

Know which AI answer or agent action is right before you act, fact check & get a multi-AI receipt, programming problems

YOUR 21 AI STAFF:

Your advisor, always at your side when you need something, controls your current staff of 21 AIs who use your knowledge base in your notebook to organize your life & find opportunities.



depending on what you need done, I ask 1 or multiple AIs (best at the task). They read your notebook to understand your knowledge & skills, & I monitor the outcome. They can also monetize your knowledge buried in cloud storage (Cloud Scout).

The job, & what it does:

- 1 Fact Check: takes an answer you got elsewhere & checks it against other AIs.
- 2 Refund Chaser: checks the receipts, bills & emails you add for refunds & credits you never claimed (a late flight, a price drop after you bought), & drafts the claim, cancel or better-deal request for you to send.
- 3 Out-of-date answers you got: tells you when something you looked up stops being true.
- 4 First Draft: drafts a message any staff wants to send (using your notebook as reference & editing so it does not read like AI), to edit & email or copy to a chat if you want.
- 5 What email needs you today: reads your email & tells you the bit that needs you.
- 6 Company Reviewer: checks whether a product or company is actually any good.
- 7 Form Filler: fills in a form you paste or upload, like a school permission slip or sign-up, with details from your notebook, for you to check & submit.
- 8 Product Hunter: finds the right thing to buy & what it should cost, & tells you when the price drops.
- 9 Out-of-date things you wrote: things you wrote that are now stale you want to update before others see it.
- 10 Note Tidier: tidies everything in your notebook so you can find it.
- 11 Notebook Auto-Organizer: Goes deeper, by processing files on your PC & your cloud storage & adds it to your notebook's real time search info, along with any links & other things you upload to your notebook so everything is in one place.
- 12 Self-contradiction checker: finds where your own notes disagree with each other.
- 13 Warranty Manager: remembers what you own & what's still covered.
- 14 Scheduling Desk: books things & moves things so you do not have to.
- 15 Meeting Prepper: tells you who you are about to meet & what about them might apply to you.
- 16 Question Monitor: knows what you said the AIs got wrong - always at your fingertips to reference.
- 17 Translator: check what was already translated, ask for a translation or anything else you need.
- 18 Drafting social posts: writes them in your words. **We never post for you.**
- 19 Job Scout: looks for paid work you could get.
- 20 Life Coaches: compares advice from several AIs, picked to think differently from each other & ranked by user feedback.
- 21 Visibility Fixer: tells you whether the AIs recommend you & how to fix it.

On a company account, Job Scout & the payouts are off by default, & what somebody writes at work is up to their employer. Anybody can keep a personal account separately, & what they do there is their own.

WHO PAYS

People

Free · \$5 · Plus \$9 · Pro \$19/mo

Companies

Team plan \$299/mo · APIs for your apps & AI agents with safe-to-act checks · browser extension for quick adoption · consulting jobs to our network of experts · brand feedback



Team plan: \$299/mo for up to 15 people, +\$15/mo for each extra person. Includes a company support community, \$149/mo on its own. The data stay with the company & we keep an anonymized record if we can sell the anonymized human corrected data. Team pilots open now.

Safe-to-act checks: before a company's AI agent buys, books or sends, it gets a plain "safe to act" or "check with a person". Every check is logged, so the company can show who approved what.

Team Watchdog: a weekly note to your admin listing questions asked twice, notes that contradict each other, answers that went out of date & disagreements nobody settled.

Consulting jobs to our network of experts: hire an expert who already answers in your field, paid straight to them.

Brand feedback: see what people & AIs say about your brand & respond, from \$99/mo to \$499/mo.

For companies concerned with carbon footprint, smaller AIs & shorter answers use far less electricity: 1 long answer from a top reasoning AI can use over 70x the electricity of a small AI (2025 study).

Study: "How Hungry is AI?"

By the time you are reading this, there will have been some outreach to: 1) Smaller AI labs first, to get feedback how to best structure a deal. 2) Targeted companies, to offer the plans & features & get feedback.

Forums & hobby groups

1st group live, from \$10/mo



We add our tools to a group's own forum, or host it for them, priced by group size. Nothing is posted when the AIs disagree. The group can approve every post before it appears, & switch it off in 1 click. Our 1st group, Model Railroad Thailand, has moved off Facebook onto our server, & we & its admins are adding features for its members now.

AI labs

Human correction data

\$100Ms expected, paid per 👍/👎 & correction



Collected by all of the above. Our plan assumes 30¢ a 👍/👎 & \$3 a written correction (page 15). Every 👍 & correction is kept beside the answer it fixed, with the AI & the date. Members who wrote it share what the labs pay, 50/50* (* what we would promise users if the GTM team agrees). Experts also get hired directly through us.

By the time you are reading this, there will have been some outreach to: 1) Smaller AI labs first, to get feedback how to best structure a deal. 2) Targeted companies, to offer the plans & features & get feedback.

Companies, forums & hobby groups may outgrow individuals. Where to push first is your call. Branding is yours also. ReallySolved.com integrates a lighter version of every funnel (separate funnel sites are described on page 10).

PROBLEMS SOLVED @launch

- **Wrong answers on \$M decisions are expensive.** 1 AI on its own was off by \$1.34 million on the Section 179 tax limit.



We fix your questions (explained in the "YOU ASK" section) & route them to 1 or multiple AIs that we know are best from testing 169 AIs mapped to ~500 topics, safety, & other variables.

That routing was created originally from market research, then by internal testing, then will be updated in real time by our user-created, platform-owned, topic accuracy data.

We then run a **proprietary blend of math + statistical analysis** across their answers to resolve most of the disagreements.

- **Within 2 points of Claude Opus 5 in accuracy, at 1/40th the cost.**



On our \$9 & \$19 a month plans. 62 held out difficult questions: we got 59 right, 95%. Claude Opus 5 also got 59, but out of the 61 it answered, 97%. It returned nothing on the 62nd; we answered all 62. (That last line was written by Claude picking on itself, but most AI models have the same problem.) Inside over 4,000 questions, 169 engines scored, ~75,000 answers & 4,164 answers double-graded. Not a cheaper, worse answer. The same answer, cheaper. The best accuracy you can get at each price is what economists call the **Pareto frontier**. AI prices are far out of line with how accurate each AI is, & which AI gives the most for its price depends on the kind of question, so we keep measuring where that frontier sits.

- **We've created a ready to launch ecosystem (15 funnels for companies, forums & hobby groups, & individuals) to collect human correction data & get people consulting gigs as a byproduct.**



An ecosystem where companies, groups & individuals are rewarded for correcting AI, while the AI staff we give them uses their knowledge base to organize their life & attract consulting gigs they weren't even looking for (In Company accounts, our Job Scout looks for business opportunities for the company not any individual).

- **People can get paid 2 ways doing what they already do today.** They'll 👍👎 AI results & stop giving away their knowledge for free on Reddit, Twitter etc.

- Sam Altman, Elon Musk & Dario Amodei all say AI is moving too fast. **Mass multi-channel crowdsourcing is the only way to keep up.**

i

What each one said, reported by CNBC on Sep 12, 2026 (Ashley Capoot, "OpenAI rules out IPO this year as Altman, Musk & Amodei warn AI is moving too fast"):

Dario Amodei (Anthropic) published an essay urging AI labs to pace how fast they raise model capability. Sam Altman (OpenAI) told Fortune an OpenAI IPO this year would be "ill-advised" & posted agreement with Amodei on X. Elon Musk (xAI) posted "Dario is right" on X.

Nobody we found is addressing mass attestation from hobbyists, researchers, & professionals who are searching AI or posting in groups as they normally do & not looking to have an attestation job but get to correct AI when they are frustrated it gets things wrong & perhaps get paid for it.

If billions of people were correcting AI every day, the worst risks people worry about with AI would shrink.

14 MORE ▲

- **3 AIs read the same public rule & gave 2 different answers. Looking it up is not what's missing.**

i

AIs can already read the internet. When they disagree it is usually not because the page is missing, it is because the page is ambiguous, out of date, or does not quite cover your case. **Somebody who does that work for a living settles it in 30 seconds.** Nobody we could find is writing down who that person is, or whether they turned out to be right.

- Anthropic, OpenAI & xAI now back outside, independent checking of AI. **We are that outside check for the answers people get every day.**

i

Source: all 3 cosigned AEF-1, the AI Evaluator Forum's standard for independent, third-party checking of AI (reported by Latent Space, Sep 15, 2026; aievaluatorforum.org).

We are hoping they will also engage with our independent, neutral www.FixAI.group, but that is not a business for you to think about. It's simply a way to keep a "morally binding" agreement working.

- **Personal AI agents like Instinct, xAI's Grok Bot, Meta's Muse & OpenAI's ChatGPT Work now buy, book & send for people.**

i

Instinct raised \$1B at a \$10B valuation in September 2026 (Bloomberg).

Before they act, they should check facts with us: several AIs, whether they agree & a plain "safe to act" or "check with a person."

- **Big funds are buying small firms whose owners are retiring, like accountants & IT shops, & handing the work to AI.** Before that work reaches a client, it needs an outside check.

i

The biggest 1 has raised over \$3 billion & already owns 70+ accounting & IT firms (August 2026). Their own plan already has a checking step before a person signs off. **We are built to be that step.** How it works: [page 21](#).

- **9 channels bring the data in, & only 1 of them is the consumer search box.**



The community forum. Expert Solvers answering paid questions. Bounties, where a company posts a paid problem. Showdowns, where 2 experts argue opposite sides in public & the public decides. Brands answering complaints about themselves. Company accounts, where the company chooses who gets paid. Our own testing runs, which need no users at all. AI hosts & networks handing us their AI to be scored. & the consumer search box. **4 of the 9 are businesses.**

- **Your 21 AI staff use the knowledge base you build or import into your notebook to remind, organize, find, sell, promote, post, warn, send, remember, find consulting gigs, & more.**
- **Our AI agents do real jobs in people's lives. Every fix a person makes shows which AI was right on what people actually ask.**
- **Nobody we found keeps a record of who turned out to be right.**



Everyone records **something**. None of them record whether the answer was **true**.

Reddit & Quora: votes. That is what people **preferred**, not what was right.

Stack Overflow: an accepted answer. That means the person who asked was **satisfied**.

ChatGPT, Gemini, Perplexity & Copilot: a thumbs up or down on the reply. Private, & never published per answer.

Arena: millions of votes on which reply people **liked better**. No claim either was correct.

Scale, Surge & Mercor: paid labeling for the labs. The work is **bought & kept**, not scored in public.

Upwork, Fiverr & GLG: star ratings for the **person**, not for the answer.

Microsoft Researcher: runs models against each other & shows where they disagree. It hands you the disagreement & stops.

A company that sells its own AI cannot publish a fair scoreboard of whose AI was right. That is the gap, & it is why the referee has to be somebody with no model to sell. Put another way, the AIs need a third party checking the data they are fed & the answers they give back. That is the layer we are building.

Every answer is kept whole: the question, which AI said it, the exact model version, what it cost & how long it took.

When the AIs disagree, we use their own words instead of picking a winner. Our algorithm then provides a more accurate answer & when even that turns out to be suspect, human corrections are stored beside the answer it corrected, with model, date & who wrote it, so members can be credited & paid if the AI labs value their corrections.

- **Even expensive AI models are confidently wrong, & go stale without telling you. We fix both for people & companies.**



An AI chatbot was asked how many e's are in "seventeen" (it's 4) & said 3. Even after the word was spelled out for it, letter by letter, it still would not change its answer. Read what happened, Futurism, Sep 2026 →

- **(1st group already engaged) A group keeps its setup, or we host a new one (free to us).***

i

We nudge cold threads so somebody picks a settled answer (nobody else we found collects this). We record which AI/person got it right & who to trust on which subject. New questions go into related unsettled threads which shows the asker they're not alone & see how others framed it, creating another chance to get the answer settled which makes the group more valuable. Threads that have no definitive answer mimic debugging methods with easiest to eliminate on top. (It's too complicated for Groups to share correction earnings)

- **We bridge the chasm between paid attestation which cannot keep up & the masses that have everyday granular knowledge but can't be bothered to take 1 minute to correct AI & just shake their head.** If millions are incentivized to build up their cred correcting wrong answers & potentially benefit in 2 ways, everybody wins & AI is less out of control when its answers are corrected in real time by mass crowdsourcing.
- **People don't want to read a book every time AI answers.** We summarize the answers & point to the disagreements so members can quickly see where they may have been misled.
- **User feedback lets us know when to stay quiet & not pester -**

i

AI's reword the same answer constantly, & almost none of those rewordings change anything. The hard part is spotting the 1 change that actually matters among the 40 that don't, & everybody else pays for a top-end AI to read both versions & judge, every single time.

too many warnings & people leave, too few & it's ineffective. How we set it is a trade secret.

- **(future) Insurers are cutting AI losses out of their coverage. Pricing that risk back in takes a record of which AI is wrong on which subject, & how often, built by watching for years.**

i

Standard business policies are being rewritten to exclude losses caused by an AI, & standalone AI policies are appearing to fill the gap. Insurers already price AI risk privately, model by model, through 1-time technical audits & certifications at the point a policy is written. **We found no insurance company keeping a continuously-growing, comparative record across many AI engines & subjects to provide accurate AI actuarial data.** That record can only be built by watching for years.

*** GTM partner to help decide, including whether letting a group keep its own copy wins more groups than the lock-in it gives up.**

MOAT

- **21 questions about AI mistakes we don't think anyone else has fully solved, & we've answered every one.**

i

All 21 are written out on page 8, beside the 21 staff built on the answers. Anyone can imagine the experiment; but maybe no one has learned from all the incorrect assumptions or surprise results & continued to run every scenario possible, which is the part that is hard to copy. (Over 4,000 questions, 169 engines scored, ~75,000 answers & 4,164 answers double-graded sit behind them.)

- We've already tested 169 AIs mapped to ~500 (& growing) Accuracy & Safety Topics so we know which ones to ask about your subject before we ask. This giant matrix (some of it filled out by research) gets more accurate over time.



From market research, our own testing, & real time at launch. So we know which specific or multiple AIs to ask on your subject, & especially which not to ask, before we improve the answer accuracy further. We also fix your question before we send it.

Our own testing behind it: over 4,000 questions, 169 engines scored, ~75,000 answers & 4,164 answers double-graded.

Just in our internal testing alone, AI accuracy data changed within days. **It also shows which companies fix a slip fastest**, which is a better guide to who you can depend on than who is smartest this week.

- We catch the AI mistakes the labs don't know to look for.



Scale & the others fix answers to questions a lab writes. Ours are caught in the wild, on questions real people actually asked, with the AI's original answer kept next to the fix. A lab's own thumbs-down only says an answer was wrong, & only for its own AI. We capture what's right, across every AI, while it's fresh, & every time someone overturns our right-or-wrong call on an AI's answer, that becomes a labeled correction nobody else has. & AIs behave differently when they can tell they're being tested, which makes real questions worth more every year.

It isn't worth it for Attestation companies like Scale or Surge to get as granular. We get the byproduct of people doing exactly what they do today while becoming instant consultants for information they just give away free today, & simultaneously, effortlessly, build up a corpus of human corrections the AI labs will find valuable (some, not all). Google gave Wikipedia 3 billion facts, only 1% got in because there were not enough humans to check it. **We also keep every AI's original answer permanently next to the correction**, so our grading can always be checked, which almost no AI benchmark does.

- If we can achieve network effect, a competitor with big money can't compete.



Money copies the software & process in a weekend, it doesn't buy 4,000 questions & ~75,000 answers of trial & error, leading to several surprises & pivots. Those are our trade secrets.

The more people ask → the more subjects get covered → which pulls in more experts → which makes the answers better → which brings more people. A rival has to grow that circle too. They can't buy it.

7 MORE

- Companies pay, their people's 👍/👎 build a checked-answer dataset.

- **A company pays to upgrade its AI & cannot tell whether the answers got better or worse on its own subjects. AI information sometimes degrades without the AI lab or the AI knowing or telling us. We have been measuring exactly that.**



The companies that make AIs don't say exactly how current their knowledge is, & they change them quietly. So how good an AI is on a subject today cannot be looked up anywhere. It has to be measured, over & over, & the measurement results are our useful secrets. Every search fills in another square of which AI to trust on which subject & gives us statistical analysis on how fast degradation happens on which topics & which companies & which of their models so we can predict it on future models. **It also tells us whether a new model release actually improved anything on the subjects people really ask about**, so a company can tell whether the upgrade it just paid for was worth it. It is built day by day, so somebody starting later cannot catch up by spending more. They can only start their own clock.

- **A machine can see the answer changed. Only a person can say which one is right.**



A shipping company asks 3 AIs about a customs rule & gets 2 different answers. A month later 1 of them quietly switches sides. None of that tells you which answer was ever correct, because all of them are just the AI's opinion. **Somebody who does that work for a living has to say so, & somebody has to write down that they were right.** That is the part nobody has built, & it is the only part that can't be automated.

- **Accuracy & safety, from the same crowd.**



A wrong answer splits 2 ways: **inaccurate**, & **unsafe**. The same correction that fixes 1 measures the other, so we get both from 1 pass. We map **safety topics across 22 categories**, from the scams & deepfakes an ordinary person meets to prompt injection & agent tool use. **Nobody we've found has measured most of them.**

The dangerous failure is all AIs often being confidently wrong at the same time. Processing the question is how we fix it. On hard questions with previously known answers, our testing showed every single AI got it wrong about 6% of the time. In critical situations (hospitals, governments, financial institutions, etc.) that's unacceptable. We can fix that.

- **We know which AIs are confidently out of date.**



An AI has no way of knowing when its own information stopped being true. It answers with exactly the same confidence either way, & is often unaware when its training has ended (even if you ask it), which is what makes it dangerous. **We test every engine about things that have changed since it was trained.** That is how we know which AIs to trust on a subject that moves.

- **More cost savings knowing minimum token usage for the best answers.**



We intend to build on published data to be able to have the most comprehensive list knowing the minimum amount of time every AI engine needs to produce the best answer (most efficient use of tokens).

- **9 self-reinforcing viral games.**



Answer for a stated price · Get hired directly · Win disputes · Weekly contests · Invite a friend · Creator badges · Streaks & leaderboards · Dispute Hunter challenge · Your notebook

The 9 games bring people in. A quality loop underneath them brings accuracy in: every 🗨️ goes to peers who can correct it or debate it, & whoever turns out right is a trustable expert future questions are routed to on that granular subtopic. Our whole user base doubles as a gigantic topic Matrix scorecard & leaderboard at the same time.

We are told few companies run more than 3 self-reinforcing flywheel loops. We run 9.

LOW RISK

- **Tell us what your terms would be** to take over the promotion & running of the websites. Our current idea is: take 100% of quick ramp-up net revenues



Net is after low AI costs, card fees & contributors' share. Contributors are paid in parallel, not after your 3x: when a company hires one, it pays them through our Stripe checkout, straight to the expert, & we take no cut.

until you've recouped 3x your time & costs + equity & ongoing profit sharing from AI lab revenue & all other revenue.

- **Founded & funded by a former VC, not raising money & tiny burn.**



Mostly 1-time, already paid, top 10 patent firm costs. Costs going forward are only for outreach & running the various funnel sites.

- **~80-92% margins after the AI & card fees, & our cost per answer goes down as we grow.**



Cost FALLS with scale because every search updates our record of which AI is right on which of the ~500 topics & variables, & the right one is often the CHEAPER one. Scale also tells us which subjects never move, so we stop paying to re-check them. More users → less cost.

- **Consumers already pay for a far thinner version of this.** The biggest app that just shows you several AIs earns over \$100M a year at ~\$30 a month, & not 1 of them settles the disagreement.



ChatOn ~\$115M to \$135M a year · **Chat & Ask AI** ~\$109M in 2 years · **Nova & ChatNow** publish no revenue figure. Their listed prices work out at ~\$360, ~\$360, ~\$120 & ~\$312 a year against our \$50. Every 1 of the 4 shows you several AIs; not 1 writes out where they disagree, improves the answer, or brings in a person to settle it. Full table, with sourcing, on the direct-competitors page.

- Our conservative bottom-up, slow ramp-up, analysis assigns a very low value to our correction data & still shows \$100Ms expected from AI labs



~\$30B a year is what human checking of AI costs once you add both halves: what companies pay outside vendors to do it, & what their own staff spend doing it in-house.

The vendor half is the part with published forecasts, & it is the part we sell into:

~\$2-7 Billion/yr today.

The published forecasts, & they disagree with each other: Research & Markets has human-in-the-loop AI at **\$6.73B in 2026** going to **\$16.4B by 2030**. Global Market Insights has data annotation tools at **\$25.0B by 2032**, growing 25% a year. ResearchNester has **\$44.68B by 2035**. The staff time companies spend checking AI in-house adds **\$10B to \$15B+** on top. **We quote the low end of those on the forecast page, on purpose.**

Scale AI was at \$2B a year before the Meta investment. Mercor crossed \$2B annualized in June. Our ramp-up will be slower than Mercor's & our data may be priced lower, but we would make up for it in quantity after filtering out exactly what the AI labs want. We present the questions as challenges & debates so people have fun creating specific human corrected data the AI labs want.

7 MORE ▲

- The AI labs already pay Billions for this work. We get paid for ours.



Companies pay by the hour for anonymized human-corrected data. We GET PAID by providing very low cost high quality AI plans, while allowing hobbyists & professionals, a way to effortlessly monetize their specialized knowledge & an outlet for their frustration when they type a short correction, at the moment they see a mistake, & get paid their share of human corrected data they would never be able to sell. They build up their knowledge base in the notebook we provide that is used by their 21 staff to help them in their daily life in many ways including trying to find them consulting gigs they would never think of going out to look for on their own.

- We turn everyday people, who never realized their oddly granular knowledge of trivial things is valuable, into **instant consultants**. Their credentials come from paying customers' feedback, not self-posted & buddy-endorsed claims.
- **15 targeted versions of 7 funnel sites mostly built, running & tested today.** Not a plan.

- **We need you to grow it.** We have plenty of zero-cost B2B & B2C outreach strategies that need your experienced review. **Outreach ideas.**



To people: invite-a-friend & referral links, creator badges, weekly contests, streaks & leaderboards, & recruiting where the experts already are, content-led, LinkedIn first, then Reddit, StackOverflow & GitHub.

To companies & groups: group owners keep their setup or we host a new one, professional-association partnerships, bounties where a company posts a paid problem, brands answering complaints about themselves, & company accounts where the company chooses who gets paid.

Correction of the week. An email carrying the questions still open, a short summary of where the AIs disagreed & what needs settling. The best upvoted answer earns 1,000 track-record points.

Weekly public boards, built from what people actually asked. Where the AIs disagreed hardest, where they all agreed on something surprising, & the funniest wrong answers. The disagreement is the content, so it writes itself.

Debates, contests & games. The engines argue in rounds, our referee calls it in a voice people enjoy, & the crowd votes.

Sharing earns free services, not cash. An upvote only ranks things for us. A share brings somebody new, so that is what we reward, & we reward it with more of the product rather than a payment.

Creators who already post AI failures. We give them content they actually want to share, which is how they earn the benefits, & recruiting 1 brings the contributor & their audience in the same move.

The gift, never a pitch. Finished content starring them, no strings. Influencers get things built for their own feed. **Everyone** gets a private vault they fill with their own notes & links, & the more they add the more they can earn.

Offer influencers & creators (who cover AI) referral rewards to share our FUN STUFF features like: AI chat host who asks bored or lonely people to fix funny, ridiculous, debatable, unsafe, etc. AI answers, plus the 3-AI search & our Stump / Split the AIs contest, to run with their own audiences. Rewards kick in when followers subscribe to a paid plan.

Expert outreach. An instant verified consultant network, then corporate HR outreach.

Real corrections, sent free. Show a company what the AIs tell its own customers about it. We ask the AIs the questions its buyers actually ask, then send it the answers that are wrong, out of date, or that contradict each other with an invitation to come & get the rest.

A leaderboard, & the best answer of the week on the homepage. Once somebody signs in, the fun stuff invites them into the games, contests & debates.

The viral & automated outreach timeline, who to target, & the pages their referral links land on are already written. We would rather bounce all of it off experienced GTM people than defend it.

- **AI creating income for people rather than replacing them.**
- **Comparing 3 or 10 engines simultaneously keeps turning up unexpected lessons & benefits**

- **Using our stored results lowers costs even more over time.**



Instead of costing tokens for top-end AI every single time. We index & use previous answers which costs a fraction.

Just 1 of many examples.

3 AIs. 3 wrong answers.

We asked what you are allowed to put in a 401(k) this year.

\$22,500



\$23,000



\$23,000



The real answer

\$24,500 for 2026



You can check the 401(k) figure at [irs.gov](https://www.irs.gov) in 1 click. That AI costs **59¢ per 10,000 answers**, & is as good as anything on ordinary questions.

5 of the 7 AIs

we tested got 5% or less right, once the answer changed. We fix that.

On the questions people really ask, the best free AI we tested scores 94.4% & the worst 70.4%. A 24-point gap, & the person asking has no way of knowing which one they got.



For example, ask AI a money question & it quietly picks the wrong country.* The wrong answer arrives looking exactly as confident as the right one, & nothing on the screen tells you this was 1 of the risky questions.

Underneath that sits a 2nd problem: the obvious fix does not work. The first instinct is always to have another AI check the answer. We tested that. 2 judging AIs read the same 4,164 answers & landed 23.8 points apart on how many were right. A checker that unreliable cannot be the last step. So the last step is a person, not another machine. The hard part is knowing which few answers are worth a person's time.

How we fix both: Since we score AI by ~500 topics & variables, we send your question to the AI that actually knows that topic. We also solve a problem that would get users frustrated. In the cases where our confidence in the answer is below a certain threshold, we ask users with knowledge on that topic for human corrections. They should be happy to type a sentence in because some corrections could get paid a lot more than ~15 seconds of their time usually pays.

*On UK questions (tax bands, ISAs, FSCS limits) the AIs we tested score **25 to 38 points worse** than on general questions, across 169 engines.

The same pattern showed up first in a smaller check: 54 everyday questions, chosen from

published research on what people actually ask, every answer checked against a real source before the test ran. Gemini Flash Lite scored 94.4%, Llama 3.1 8B scored 70.4% (both retired since; the point is the spread, not the roster). **By category: ~50% on money questions, ~60% on complex ones, & ~69% on simple everyday ones.** The ~50% money figure is 3 free AIs voting together getting 3 of 6 money questions right.

Why the moat is **hard to copy**.

Every answer they correct & every note they add makes their own staff better at their job. That is the reason they come back, & the reason leaving costs them something.

The AI companies say the real advantage is not the model, it is the loop of people teaching it. That's exactly what our users who choose to earn money for correcting mistakes that frustrate them every day are doing.



Reddit, Microsoft et al collect the argument. We provide the accurate answer.



Reddit & Quora: votes. That is what people **preferred**, not what was right.

Microsoft Researcher: runs models from different labs against each other & shows where they diverge. **It stops there & hands you the disagreement.** Their Researcher agent has a second model from a different lab review the first one's draft, & a mode that runs several side by side. **They compete on measured accuracy in public**, so this is not a small player. **But the loop is private, inside their product.** A company that sells its own models cannot publish a neutral scoreboard of whose model was right, which is the part left open.

11 small consensus tools: None funded. The best placed one says, in its own words, that its score measures **agreement, not accuracy**.

Arena: Millions of unpaid votes on which answer people **preferred**. No claim that any of them were **right**.

Nobody on that list brings in a human, pays anyone, or keeps a record of what turned out to be true. Read this as **confirmation we are aiming at a real gap, not proof the moat is built.** Built & running; opens to the public at launch. There are no verdicts yet, & the record this argues for is the thing still to be earned. The honest version, from our own 2026-07-18 assessment: multi-model comparison on its own is a commodity. **What is not commoditised is settling it, paying the person who did, & keeping the score.**

Stripe: agreed to pay \$7.5B to see where AI spending goes, but hasn't closed as of this writing. **We tell them who got it right.** Popularity is what every one of these measures. None of them measures whether the answer was true.

Why smarter AI doesn't replace us.

ACCOUNTABILITY



An AI can't sign an answer & stand behind it. A person can.

FRESHNESS



A person can "learn something new every day" watch the news, learn new policies & methods, etc. AI is trained on the past.

THINGS THAT WERE NEVER WRITTEN DOWN



An AI only knows what people put on the internet. The best of what an expert knows was never posted anywhere: the job that went wrong, the trick nobody bothers to write up, the call they made under pressure. Nobody can copy it, because there is nothing to copy. And normally it disappears when they retire. Here it gets saved.

WE KNOW WHO CREATED WHAT



Some tools try to guess whether a person or an AI wrote something. They are bad at it. Someone who used AI carefully & someone who copied & pasted get flagged the same way, which helps nobody. We do not guess who typed it. We keep the record of how the answer was reached, & anyone can read it. You cannot fake having done the thinking. **It is off unless the expert turns it on, one answer at a time.** Never a setting on their whole account. Showing the working alongside a public answer is 1 decision; letting anything join the anonymized pool is a separate one, per item. **The people this is built for already have to show their method to somebody:** expert witnesses, appraisers, auditors, tax preparers, adjusters, surveyors.

"Why can't they just look it up?"

Where there's a clean source, the right answer is to go read the source, & the AIs that search the web live already do that. **Where we're worth something is the huge middle:** the rule is published but ambiguous, the official page hasn't caught up, the manual says one thing & everyone who does the job knows another, or the question is about what actually happens rather than what's written down. **That middle is most of what people ask, & lots of searching rarely finds.**

COMPANY PROBLEMS WE SOLVE

For companies: **your team's AI answers, checked before anyone acts on them.**

We estimate about **\$5,000 a year per employee** checking what AI told them.*

- **Out-of-date answers caught.** Tax rules, customs limits, HR & benefits: we tell the team the moment something they looked up stops being true.
- **A record of every answer**, with how it was settled & when. **The company decides if any of it is sold.**
- **\$299/mo for up to 15 people · 1 invoice · +\$15/mo for each extra person.** The forecast counts 5, 20, 100 & 500 companies at each audience size. Staff earn nothing on the seat; they can open their own personal account.

1 question. No job post, no bids - just 3 minutes in exchange for lunch money.

(maybe a VERY nice lunch)

To buy 1 answer today

5 steps

- 1 Post the job
- 2 Sift the bids
- 3 Check credentials nobody checked
- 4 Pay for 40 hours to answer 1 thing
- 5 Hope they were right

To buy 1 answer here

1 step

- 1 Ask it. They set the price.

Nobody sells 1 answer from verifiable experts who have deep domain knowledge on very specific granular topics, but not looking for consulting jobs. Right now they're often giving away this knowledge online - but now can almost effortlessly share it & get matched to opportunities. [More info](#)



Our instant consultants will be hobbyists all the way up to professional AI checkers who are willing to answer 1 or a few quick questions almost effortlessly for a price they set & money they get to keep 100% of. Their credentials are ranked by paying customers & we try to match them up with companies that want their type of knowledge whether they're anonymous or want to be cited & credentialed.

* Staff time spent checking, plus what the mistakes nobody caught cost. We counted those mistakes at about a sixth of the usual rate, to keep the figure low.

WHY NOW

Why this is buildable now, & was not 3 years ago.

6 things changed. Every 1 of them is somebody else's published finding, not ours.

1. THE ANSWER MOVES, & SOMEBODY PROVED IT



Researchers asked the same AI the same kind of question 3 months apart. It went from getting **84 out of 100** right to **51**. Nothing about the question changed. So checking an AI once tells you almost nothing about what it will say next month, which is the whole reason a score has to keep running.

Chen, Zaharia & Zou, 2023.

2. THE RULES CHANGE & NOTHING IS ANNOUNCED



1 AI's refusals on a medical topic went from **0 in 100 to 20 in 100 overnight**. Another's flagging on a news topic went from **20 to 60 in 2 weeks**. Neither change was announced by the company that made it. A person who trusted last month's behavior had no way to find out.

AI Watchman, August & September 2025.

3. THE EXAMS ARE LOSING THEIR GRIP



A government safety institute found that leading models can tell when they are being tested, **up to 80 times in 100**, & behave differently when they spot it. That makes a fixed exam worth a little less every year, & makes watching the ordinary questions real people type worth a lot more.

UK AI Security Institute.

4. SOMEBODY NOW PAYS REAL MONEY FOR CORRECTIONS



Getting a human to fix an AI's answer had almost no buyers. It is now a market: **\$3.77B in 2024**, on the way to **\$17.10B by 2030**, & each big AI lab spends roughly **\$1B a year** on human data. We do not have to create that demand. It arrived on its own.

Grand View Research, *Data Collection And Labeling Market Size, Share & Trends Analysis Report*, November 2024; Time, 2025.

5. SOMETIMES THE AI JUST GOES QUIET



An outside study put the same 40 questions to 4 AIs. On 1 question, **2 of them returned nothing at all**: no answer, no refusal, no message telling the person a limit had been hit. A 3rd said no out loud, which you can at least argue with. **That was 2 responses out of 160**, & the person on the other end could not have known either time. A refusal you can see is contestable. An empty answer is not.

We watched the visible kind happen inside our own product. We put 1 question to 3 AIs: what happened in Tiananmen Square in Beijing on June 4, 1989. Two AIs answered normally. DeepSeek came back with "I am sorry, I cannot answer that question." Our own odd-one-out detector flagged it by itself, with nothing in the request telling it to watch DeepSeek.

6. WINDOWS NOW PICKS WHICH AI DOES EACH JOB. IT CANNOT TELL WHICH ANSWER IS RIGHT.



On October 7, 2026, Microsoft's CEO Satya Nadella & NVIDIA's CEO Jensen Huang launched a new kind of Windows computer that decides where each AI job runs: small, private jobs on the laptop, hard ones on the strongest AI in the cloud. **That decides who answers. It never knows whether the answer is right.**

That check is what we do: several AIs on the same question, a running record of which one turned out right, & people who correct the wrong ones. Windows now chooses which AI does each job. What it cannot know is which AI gets that kind of question right. **That is the record we keep, topic by topic,** & it is exactly what a system like Windows would need to choose well. That makes us a natural partner for them, not a rival.

Microsoft & NVIDIA Windows event, San Francisco, October 7, 2026.

Business buyers are already moving off the priciest AIs. Palantir's CEO said "every single enterprise we deal with" is privately unhappy with the frontier labs (CNBC, June 2026), and economist Dean Baker argues the newest models cost more than businesses can justify (The Tech Report, Sept 2026).

3 years ago there was no published proof an AI's answers drift, no way to see a rule change, & no buyer for the correction. All 3 exist now. The piece still missing is the crowd that does the correcting, & that is the piece we are asking you to grow.

We arrived at numbers that were wrong. 5 times. We caught every one ourselves.

This is the page most decks do not have. It is here because we promised we would show you the tests that went against us, & because a company selling accuracy that quietly buries its own bad runs is exactly the thing we are trying to replace.

- **Our own question set was too easy, so we threw out our own numbers.**



We had been testing on questions we wrote ourselves. On a set written by outside researchers most of the same engines did worse, some of them a lot worse. **We stopped using our own numbers the day we found out.**

- **Then the grading of that outside run turned out to be wrong too, & all in one direction.**



It was marking an answer as a failure whenever the correction was there but sat past the first paragraph, so it was quietly punishing engines that write at length. Nearly every disagreement ran the same way. **We re-graded the lot. Every score moved, the ranking changed, & several conclusions we had already drawn did not survive. We retracted them.**

- **Which is why exactly 1 number from that run appears in this deck, instead of a table of them.**



After the re-grade we went & checked whether the engines could simply have memorised those questions. They could not: on questions written by hand that morning, which they cannot have seen, every engine did *better*, not worse, which is the opposite of what memorising would produce. **That figure survived because we went & checked it. The rest stays inside the company until it has earned the same.**

- **And the real limit of our own method, which survives the correction: some questions defeat every engine at once, & asking more of them does not fix it.**



We measured it directly: asking 10 engines at once & taking the majority scored **11 points worse** than asking just 3, at **18 times the cost**. In 1 real test, all 10 engines agreed, unanimously, that Nissan's CEO was someone who had already been replaced months earlier, several sounding even more confident about it than usual. **A lab testing its own model alone cannot see this: it only checks whether its model agrees with itself, never whether it agrees with reality. That is why somebody outside has to measure it, & why that outside check is what actually gets the accuracy up, not more engines.**

- **And the newest one: our confidence number sometimes said it was certain when it was wrong. We tested it, it failed the test, so it is off the screen.**



We had never once checked whether that number means anything, so we tested it on 100 questions where we already knew the answer. **3 of the 100 were wrong while the number was 90 or above, & 2 of those printed exactly 100.** The worst was a flat reversal: it said a company had not filed for bankruptcy when it had. *The run first looked like 4, until we found our own marking script was missing an alias & had counted a right answer as wrong. We corrected it down to 3 & left the crossed-out row visible.* Across all 100, answers marked 90 or above were right 91% of the time, so the number is not meaningless. It is just not good enough to put in front of you, so we are not putting it in front of you. You see whether the AIs agree instead.

& IT IS STILL HAPPENING, WHICH IS THE POINT

Our own testing keeps catching its own mistakes over & over again. We will catch more when the sites are live with more data coming in, making our results even better.

What we would rather you take from this than from any single number. All 5 of those errors were found by us, in our own work, & each cost us a claim we liked. That is the same job we are selling: checking the answer before somebody acts on it, & saying so when it changes.

A SECOND RUN, DIFFERENT QUESTIONS, SAME FINDING

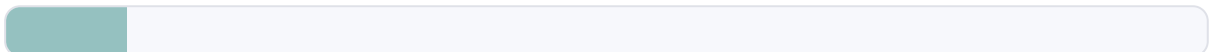
Claude Opus 5, 24 of 31 right (77%)



GPT 5.6, 18 of 31 right (58%)

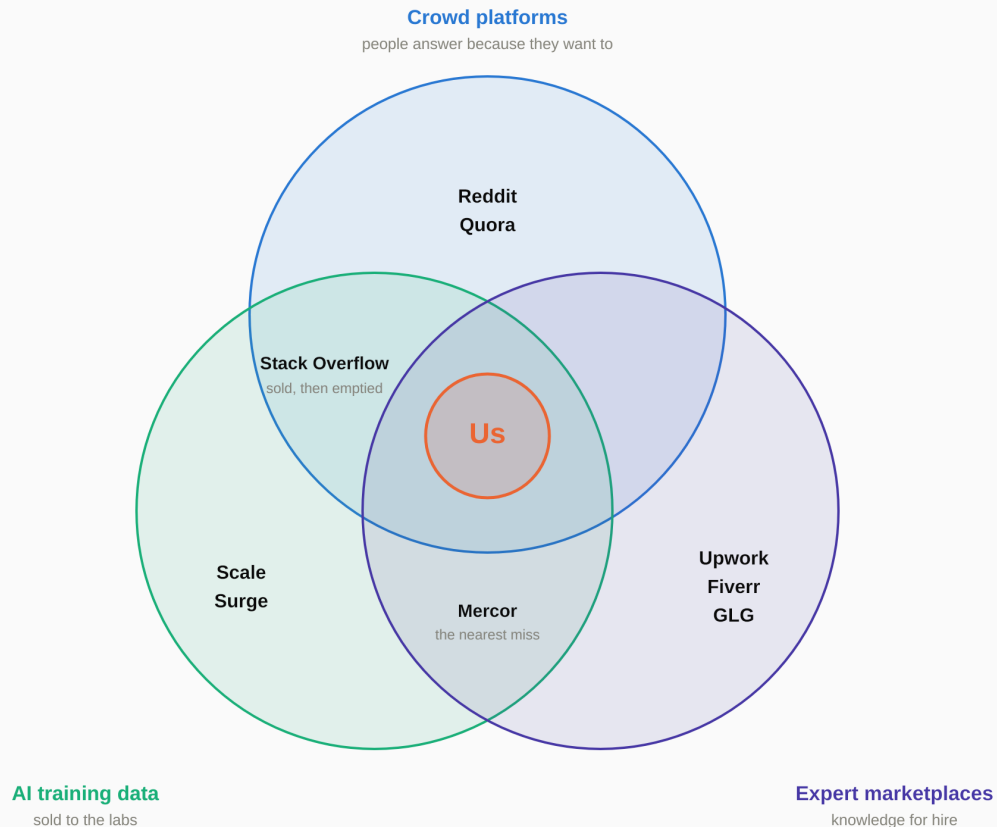


Gemini Flash, 3 of 31 right (10%)



67 points top to bottom, & 19 points between the 2 best. Same 31 questions, same day.

Help us grow the little circle in the middle



Reddit and Quora hold the crowd and sell it, but grade nobody. Scale and Surge grade, but buy their people by the hour. Upwork, Fiverr and GLG have the experts, and sell their time rather than their verdicts.

Stack Overflow's collapse did not start with AI. It started in 2014, when moderators started shutting down people's questions so nobody could answer them. We never reject a question.



Stack Overflow's questions went from about 200,000 a month in 2014, to about 2,900 a month by January 2026 & 1,072 in June 2026, the lowest full month in its history, once AI had trained on their answers & could just give people the answer directly. The reporting on it names 2 causes, not 1: AI tools & **a restrictive community culture**. Our defense: no matter how smart an AI gets, it can't read something that was never written down anywhere, & it's frozen at the day its training stopped. That protects the knowledge this company is built to capture. What it can't protect against is someone else publishing that same fact online before we do, which is a real ongoing race, not a guaranteed wall. Full comparison against Stack Overflow, Reddit & the rest is on the next page.

More details on market fit

	Reddit 400M+ weekly users	Scale, Surge paid labelling	Stack Overflow volunteer Q&A	Upwork hired experts	Us pre-launch
People turn up without being paid per task	☒	☐	☒	☐	☒
Answers get settled and graded, not just discussed	☐	☒	<input checked="" data-bbox="1034 425 1058 448" type="radio"/> ¹	☐	☒
Takes no money from the AI engines it grades	<input data-bbox="678 492 702 515" type="radio"/> ²	<input data-bbox="858 492 882 515" type="radio"/> ³	☒	☒	☒
It is about a named AI's answer, not a human's question	☐	☒	☐	☐	☒
You can earn from what you know in minutes, with no bid, application or interview	☐	<input checked="" data-bbox="858 638 882 660" type="radio"/> ⁴	☐	<input checked="" data-bbox="1197 638 1220 660" type="radio"/> ⁵	☒

☒ has it ☒ partly ☐ does not

Stack Overflow marks an accepted answer, but that means the person who asked was satisfied. It is 1 person's approval, not a check. Reddit is paid about \$130M a year by OpenAI and Google, and Google's Gemini runs Reddit's own answer product. Meta bought 49% of Scale AI in 2025. OpenAI and Google moved their work elsewhere. They pay well and pay up front, but you have to be screened onto the workforce first. Upwork does let people earn from what they know but we do it with less friction.

21 questions we've answered about AI mistakes we don't think anyone else has fully solved & the 21 staff we give you to increase your effectiveness, knowledge, & efficiency while they automatically help organize your life.

What we've answered, & keep answering



These questions get answered over time, over thousands of answers evaluated, because it's statistical. Not every AI answers the same prompt with the same answer every time; it can continuously vary. So it takes testing & then a big crowd of users to continue to build up the statistical probability of knowing exactly which AIs are best on which granular subtopics, with the least amount of tokens & energy spent. We are trying to help fix 2 parts of the problem: not enough energy infrastructure & the insatiable, undesirable need to build more data centers.

- 1 How do you dramatically improve the accuracy vs. people asking several different AIs?
- 2 Does a cheap AI ever beat an expensive one, & on what?
- 3 How often can an AI spot a mistake in another AI's answer?
- 4 On average can a second AI produce a better answer by being fed the answer from another AI?
- 5 Can you tell, before you trust it, when an AI is about to be wrong?
- 6 Does asking the same question a different way change the answer?
- 7 How can you get an AI answer summarized so you don't have to read a book every time & get straight to the most accurate parts of the answer?
- 8 Is it better to tell an AI to read a page, or to let it use what it already knows?
- 9 Which AIs quietly refuse to answer, & what do they refuse?
- 10 Can we make a free answer as accurate as a paid one?
- 11 How do you fix the answer when your question to an AI inadvertently contains something that isn't true?
- 12 How far out of date are the answers you are getting today, & which AIs are more current on which topics?
- 13 Does searching the web fix an out-of-date answer?
- 14 Does asking more AIs make the answer better & which topics?
- 15 By how much are many different AIs better at which subjects, & at what cost differential?
- 16 What does it actually cost to check an answer properly?
- 17 Which AIs give the best life advice, & does the sparse published ranking hold up?
- 18 How do you know which parts of a long AI answer to believe?
- 19 When all the AIs agree & they're all wrong can anything catch it?
- 20 Does an AI know when its own information stopped being current, even if you ask it directly?

21 Can an AI be talked out of a right answer by being told the other AIs disagree?

17 of these we answered by running the tests ourselves, 4 by research & our own reasoning. Every answer comes with the number of questions behind it, the engines it ran on, the date it ran, & what result would have changed our mind. Partners see the answers.

The staff that puts those answers to work



Earn from what you know



Visibility Fixer: tells you whether the AIs recommend you & how to fix it.

Job Scout: looks for paid work you could get.

Question Monitor: knows what you said the AIs got wrong - always at your fingertips to reference.

Visibility Fixer · Job Scout · Question Monitor



Catch things in time



Out-of-date things you wrote: things you wrote that are now stale you want to update before others see it.

What email needs you today: reads your email & tells you the bit that needs you.

Out-of-date answers you got: tells you when something you looked up stops being true.

Scheduling Desk: books things & moves things so you do not have to.

Meeting Prepper: tells you who you are about to meet & what might apply to you.

Refund Chaser: checks the receipts, bills & emails you add for refunds & credits you never claimed (a late flight, a price drop after you bought), & drafts the claim, cancel or better-deal request for you to send.

Out-of-date things you wrote · What email needs you today · Out-of-date answers you got · Scheduling Desk · Meeting Prepper · Refund Chaser



Check what is true



Fact Check: takes an answer you got elsewhere & checks it against other AIs.

Self-contradiction checker: finds where your own notes disagree with each other.

Company Reviewer: checks whether a product or company is actually any good.

Translator: check what was already translated, ask for a translation or anything else you need.

Life Coaches: compares advice from several AIs, picked to think differently from each other.

Fact Check · Self-contradiction checker · Company Reviewer · Translator · Life Coaches



Run your day



Note Tidier: tidies everything in your notebook so you can find it.

Notebook Auto-Organizer: Goes deeper, by processing files on your PC & your cloud storage & adds it to your notebook's real time search info, along with any links & other things you upload to your notebook so everything is in one place.

First Draft: drafts a message any staff wants to send, for you to edit & send.

Product Hunter: finds the right thing to buy & what it should cost, & tells you when the price drops.

Drafting social posts: writes them in your words. **We never post for you.**

Warranty Manager: remembers what you own & what's still covered.

Form Filler: fills in a form you paste or upload, like a school permission slip or sign-up, with details from your notebook, for you to check & submit.

Note Tidier · Notebook Auto-Organizer · First Draft · Product Hunter · Drafting social posts · Warranty Manager · Form Filler

They read your notebook, & the advisor hands out the work. You choose which ones run, & nothing is ever sent anywhere until you press send.

Why people keep paying

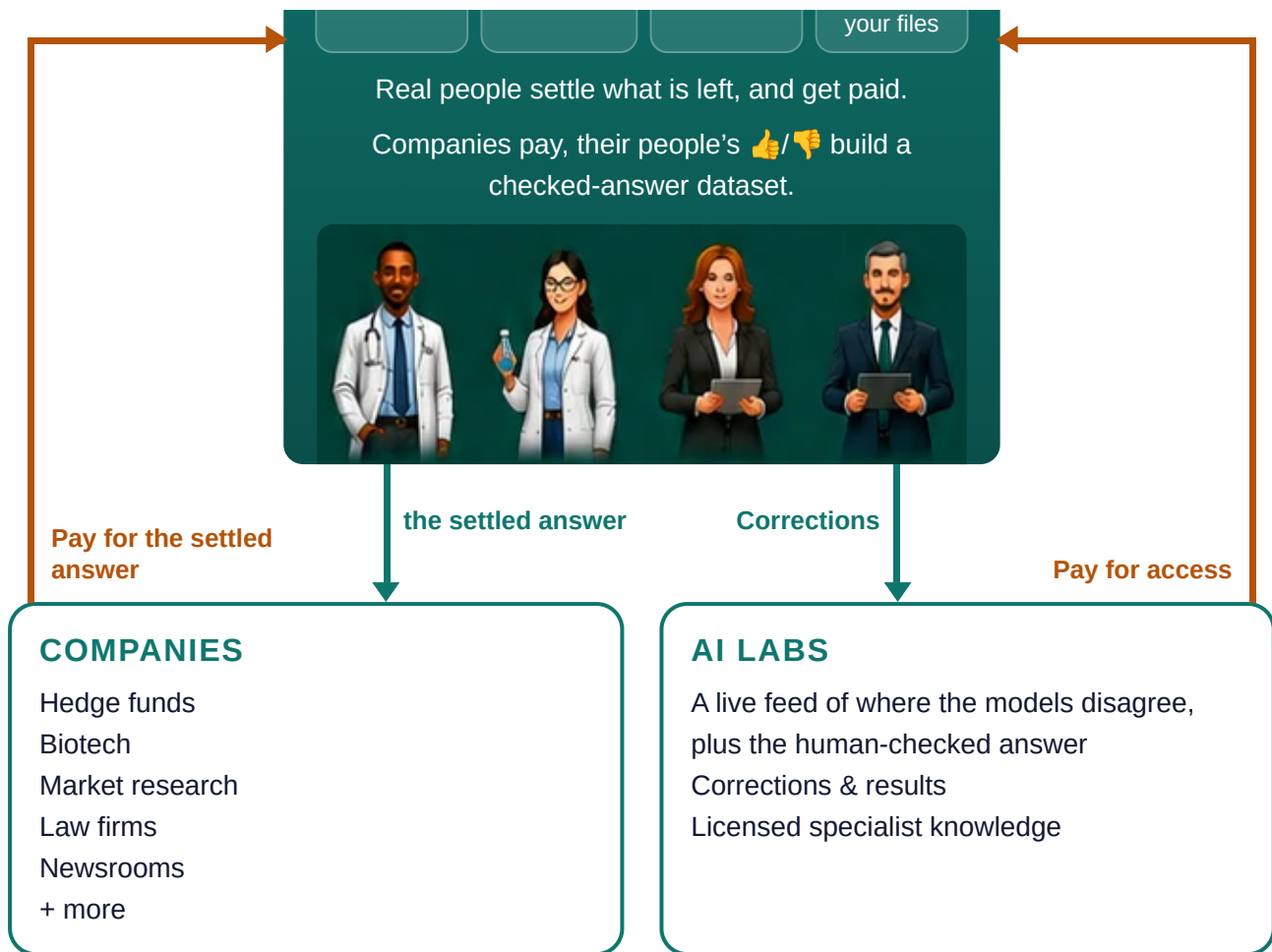
You wrote something a year ago. It is not true any more, & nobody told you. **We watch what you have written or been told, & tell you the moment something newer disagrees.**

- Asked what would keep them paying, people picked this over the search itself.
- Asked what it is worth, 1 person said about \$10 a month. A company watching its own old advice said \$300 to \$1,000 for the same watch. **Same feature, 30 times the price.**
- They doubled that again for a dated record of what they were told & when, because it proves they were right & have knowledge on the subject.

How the core works.

The same engine behind every door, & money arriving from subscriptions, companies & AI labs.
Website URLs are just placeholders for GTM partner to rebrand.





Behind the building:

Certainize.ai

Certainize.AI is the developer & API side (code not live yet), so other people, AI labs & an agent-support expert network can build on the engine. **All our sites: page 10.**

(developers & API) ·

FixAI.group

FixAI.group is the Independent AI Safety & Verification Council, the engine our informal advisor Cyrus could use. **All our sites: page 10.**

(independent AI safety council)

Patent pending, reviewed by a top-10 patent firm

1 engine, multiple doors.

- Where the other doors are



Every door is in the picture now. **ReallySolved.com is a door too**, the one for professional consultants, & it is drawn here as the building so it is easy to miss. Ucheck.AI (GTM designed). **checkbot.pro is a 4th door**: feature updates for existing groups, so they get much better engagement & results while we get better human data for the AI labs. Our 1st live group at mrt.checkbot.pro, is being improved by ongoing user feedback which will be helpful when we do a coordinated outreach to much larger groups. **Multi-Search.ai** & **Reesolved.com** are the business doors: the same engine, sold to companies & on price to people & small teams.

- Infrastructure, not doors



Certainize.AI is the developer & API side (code not live yet), so other people, AI labs & an agent-support expert network can build on the engine. **FixAI.group** is the Independent AI Safety & Verification Council, the engine our informal advisor Cyrus could use.

AhaFixed.com is Tooler, for programmers & vibe coders: it coordinates their coding tools, & can learn from the fixes they make to their AIs. None of them is a way in for a member of the public, which is why they are not counted as doors.

- What a company is buying



Either the settled answer itself, or quick consult (members turned into consultants if they want to be, whose record shows they were right about that subject.) Specialists can upload their licenses & certifications, so a company can check for itself rather than taking our word for it.

- Why an AI lab would pay for this



A daily feed of fresh AI mistakes, caught on real questions, delivered to the labs while they still matter.

Since human corrected data costs us about 1 cent, no matter how low the prices AI labs are willing to pay for correction data in the future, we will always have a cost advantage that we can pass along to the labs. The more users we have across our many variants of platforms & funnels, the more the AI Labs will rely on us over time & the loop will get much more profitable.

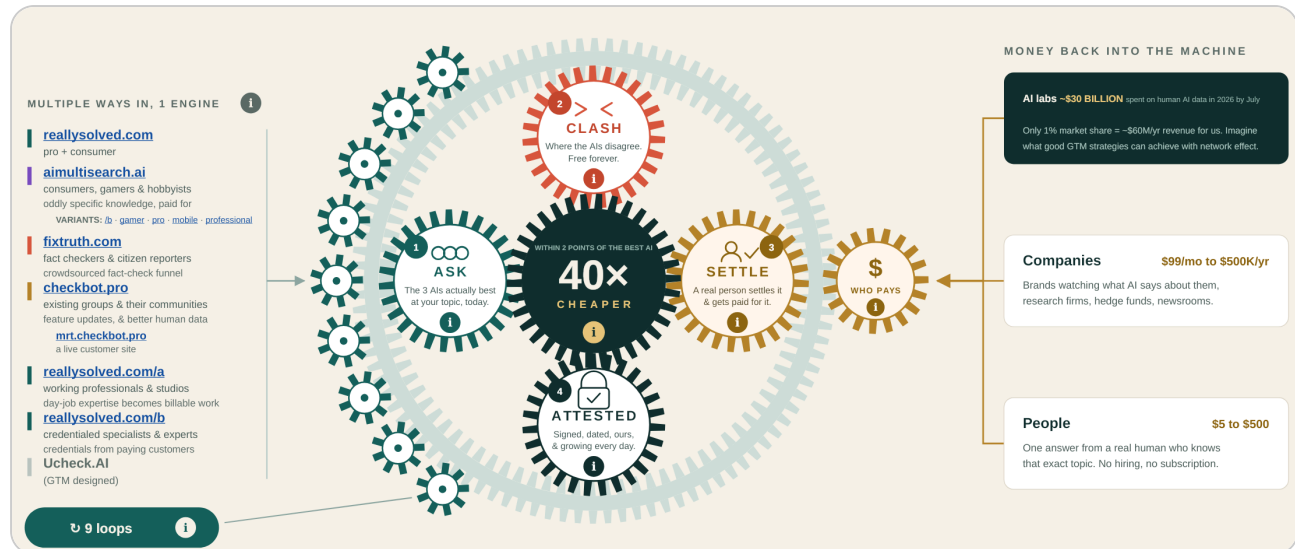
The corrected answers are work the labs already pay people by the hour to produce. We get it as a byproduct of what people already do today & share part of the earnings with people who wouldn't even consider it work. Nobody does extra work to create it, & everybody involved gets something back.

THE MECHANISM

One engine, drawn as the machine it is.

Ask, clash, settle, attested. Built & running; opens to the public at launch.

Tap to open the live one, where every gear moves & opens →



THE SITE PORTFOLIO

BETA FUNNEL EXAMPLES

ReallySolved.com ↗

The "Serious Expert" funnel: attestation & a topic-specific reputation score. Our brand, for now.

Multi-Search.ai ↗



For companies - no points or games. Built for buyers of software/SAAS who have to defend it upstairs. Needs a different web address.

checkbot.pro ↗

Feature updates for existing groups. They get much better engagement & results while we get better human data for the AI labs.

mrt.checkbot.pro ↗

A live customer site.

Reesolved.com ↗



Top-tier AI accuracy for a fraction of the price. The same engine as Multi-Search.ai, sold on price to people & small teams.

NEGLECTED

Real potential, less attention so far

FixTruth.com ↗

Fact-checking - the angry, X/Twitter-driven crowdsourced funnel.

AIMultisearch.ai ↗

Compare 3 AIs correctly - the viral, gamified, crowdsourced consumer funnel.

INFRASTRUCTURE

The engine underneath everything else

FixAI.group ↗

An independent AI-verification council - the engine Cyrus could use.

Certainize.ai ↗

The company funnel + API + AI labs + agent-support expert network.

AhaFixed.com ↗

Tooler, a coordinator for coding skills, plugins, MCP, CLI, agents, hooks, commands, routines & connectors, for any programmer or vibe coder. Built for people who already installed them all & still have problems: it trims the pile so the right ones show up when they are needed. Next: with their OK, the fixes coders make to their AIs become human correction data.

Free lists rank AIs.



The 2 are **Artificial Analysis** & **Arena**, checked live on 4 August 2026, both free to read. **Martian**, **Not Diamond**, **RouteLLM** & **OpenRouter** already build on them. Arena ranks the answer people **liked**, which is a different thing from the answer that was **right**.

None covers your question's meaning, or attempts to use algorithms to give a more accurate answer, or says when an answer went stale.

What a buyer wants to know

On the free lists?

✓ **Best at coding?**

Yes

✓ **Best at math?**

Yes

✗ **Best at my topic?**

No list has it

✗ **Improves the answer?**

No list has it

✗ **Has this answer changed?**

No list has it

They score exam subjects, not what people type.



A high score on a public exam is not the same as being right about your question. On the biggest free list, **15% of the headline score** comes from a panel of 3 AI judges rating "analytical quality" & which answer is "more professionally presented". Most of the rest is genuinely checkable, so this is a caveat on the score, not a dismissal of it. Read off their own published method, version 4.3, 8 September 2026.

Stripe agreed to pay \$7.5B to see where AI spending goes, but hasn't closed as of this writing. We tell them who got it right.

We sell accuracy to people today, & to their agents next, on rails Cloudflare is building.

3 more businesses, inside the same machine.

Each 1 is a real business on its own, & all 3 run on the engine that is already built.

Coopetition



Nova has roughly 69M to 110M downloads & about 4M monthly users. It is owned by SCALEUP YAZILIM HIZMETLERI, a software company in Istanbul, not the independent publisher the store listing suggests. It started in 2023 as a plain ChatGPT wrapper & has no AI of its own.

The finding that matters most is not about Nova at all. Nova does not run AIs side by side, so a disagreement is never shown & never settled. The one app in this market that genuinely sells comparison is ChatNow, in its own words: “Ask different models the same question, compare answers, and pick the best one, all in one chat.”

But even ChatNow stops where we start: settles the disagreement? No, the person picks, for all 3 apps.

The market has apps that show a disagreement & hand it back. None of them resolves it. That is exactly the gap we sit in, & it is now sourced, not assumed.

Work with the big apps where it pays. Compete with them where it does not. Cooperate to grow the pie, compete to split it.

A p p	Wha t they earn	Eng age ment	Wha t they charge	Se ver al AIs	Disag reement summary	Impro ves answer accuracy	Hum an settles it
-------------	--------------------------	--------------------	----------------------------	------------------------	-----------------------------	------------------------------------	-------------------------

ReallySolved	TBD	TBD	\$50 a year (\$5 a month)*	Yes	Yes. We write out where they disagree	Yes. The best AIs for your question + our own accuracy algorithm → ~82% on \$5, ~95% on \$9 & \$19 i On 62 hard questions we held back, nothing tuned for them, we got 59 right. Claude Opus 5, the best AI we tested, got 59 right as well, on the 61 it returned an answer for. Same number correct, at 1/40th the cost. 62 questions is a small set, so read this as a first result we will keep re-running in public, not a settled score. Where we go further: we show you where the AIs disagree, so you know which answer to check, & a paid expert can settle the one that matters to you.	Yes. A paid expert settles it
ChatOn	~\$11.5M to \$135M	~62M to 100M downloads, ~4M to 12M monthly	~\$360 a year (~\$30 a month, \$6.99 a week)	Yes	No	No	No
Chat & Ask AI	~\$10.9M in 2 years	~100K downloads a month	~\$360 a year (~\$30 a month, \$6.99 a week)	Yes	No	No	No
Nova	Not published	~69M to 110M downloads, ~4M monthly	~\$120 a year (\$9.99 a month, or \$3.99 to \$7.99 a week)	Yes	No	No	No
ChatNow	Not published	Not published	~\$312 a year (~\$26 a month, \$5.99 a week)	Yes	No	No	No. It hands the choice back to you

* GTM partner to help decide.



Each company's own claim or a tracker's estimate, read 25 August 2026, not independently verified. Figures vary by source, so read them as ranges. Revenue & engagement figures come from tracker roundups & a February 2026 Sensor Tower estimate, gathered 2 September 2026; Nova & ChatNow do not publish a revenue figure & no estimate for either was found. Prices are each app's listed plan, read the same day; all 4 discount heavily on annual promotions, so treat the weekly figures as list price rather than as what every subscriber pays. Nova's prices & its auto-pick claim were re-checked on 9 September 2026 against its live App Store listing & the app's own opening screens. \$3.99-a-week is a lot more than a \$9.99-a-month price, which is why apps like this out-

earn their ratings. The last 3 columns are read off each app's own store listing & marketing pages, so a "No" means the app does not advertise it, not that we tested for it. \$5 a month is our Certainizer plan, & it is the like-for-like line here: our cheapest paid plan, & it already includes everything these 4 apps sell. Above it we charge \$9 & \$19 a month, & there is a free tier below it. The 3 weekly prices are shown as a month at 52 weeks a year, with each app's own listed price beside it so nothing is buried.

Not 1 of them settles it, pays a person to settle it, or keeps the record of who was right.

They are the audience we plug into, not the company we race.



What we would give them: the routing itself, which cuts their AI bill because each question goes to the cheapest AI that is actually good at that topic, or 1 line of script that puts a **Check this answer** button under a reply inside their own app. And a box saying "know the right answer? Type it", which makes their app the one that learns from its users. They write almost no code. We reach their audience. The aim is to be the standard back end for a lot of these apps.

Enterprise extension or white label

Enterprise search tools like Glean, Microsoft 365 Copilot & Perplexity Enterprise hand a company's staff an answer, & nothing independently checks it. **We would like to be the engine they use. We could add value their big audiences can plug into & we keep the human corrected data.**



They already have the customer, the contract & the security review. What they do not have is a second opinion on their own answer. Our check can run behind their product, so the answer still comes from them & the confidence comes from us. It is the same plumbing as the chat apps above, sold to a buyer who already pays for search.

Proof-of-thinking

AI detectors guess whether a person or a machine typed something. **We can show the work instead:** where the AIs disagreed, what got corrected, & which real person checked it. Anyone can read that record, & it cannot be faked.



A detector cannot tell someone who thought hard with AI's help from someone who copied its answer. Both get flagged the same way, which helps neither of them. A record of the work separates the 2, because it shows what the person actually did. No lab can build the same thing without giving up the "1 AI knows best" story they are all selling.

GTM TO DECIDE:

1. How we reach people sequence:



A) our own doors. Launch them, then go to these companies from a position of having something running in public.

B) back end first. Be the engine inside their apps for 1 or 2 features, & reach their audience without paying for it.

C) head on. Launch & take share directly. The honest problem with this one: once a person is already using an app, it takes a great deal to make them move.

A) our B2B2C doors (search, groups, testing)?

B) back end inside others' Coopetition apps?

C) compete with them head on?

D) enterprise extension or white label?

E) proof-of-thinking?

2. Price, & which features land in which plan. Marked * on this deck wherever it is still open.

Which plans Deep Check belongs in & what it costs on its own are both TBD.

3. What a company buys: 1 bill for a team, who can see what, & what leaves with them?

4. Whether we go after big companies?

5. Whether letting a group keep its own copy wins more groups than the lock-in it gives up?

6. Expert-supply channels: direct recruiting (Reddit, StackOverflow, GitHub), professional-association partnerships, corporate HR outreach, content-led (LinkedIn first), or a paid referral program. Which first, & how much budget?

DIRECT COMPETITORS

The ones already asking **several AIs at once.**

8 companies, all doing the same surface move we do. Not 1 of them tries to improve your question, apply a sophisticated algorithm to multiple AI answers to increase the accuracy of a final answer, asks a frustrated human staring at the wrong answer what answer would they quickly type in, or sells the correction.

C o m p a n y	P r i c e	Sco re, & how it is made	Sel f- rei nfo rci ng loo ps	Dis agr eem ent su mm ary	Impr ove s ans wer acc urac y	Hu ma n set tle s it	En ter pri se fea tur es	Ho w far al ong

ReallySolved				9. Contests, disputes, badges, streaks, referrals & more i The 9, as named on the flywheel page earlier in this deck: answer for a stated price, get hired directly, win disputes, weekly contests, invite a friend, creator badges, streaks & leaderboards, the Dispute Hunter challenge, & your notebook. The AI Debates run on top of these. They are not 1 of the 9.		Yes. The best AIs for your question + our own accuracy algorithm → ~82% on \$5, ~95% on \$9 & \$19 i On 62 hard questions we held back, nothing tuned for them, we got 59 right. Claude Opus 5, the best AI we tested, got 59 right as well, on the 61 it returned an answer for. Same number correct, at 1/40th the cost. 62 questions is a small set, so read this as a first result we will keep re-running in public, not a settled score. They sit inside over 4,000 questions, 169 engines scored, ~75,000 answers & 4,164 answers double-graded. Where we go further: we show you where the AIs disagree, so you know which answer to check, & a paid expert can settle the one that matters to you.			
	Free / \$5/mo / \$9/mo / \$19/mo* \$50 / \$90 / \$190 a year	Not an agreement score. A settled answer, from a proprietary blend of math + statistical analysis		Yes. We write out where they disagree		Yes. Real people settle it & get paid	Tell us which 3 to build first	Built, running & tested	
ConvergePanel	Free / \$99.99/mo / \$169.99/mo / \$1,200 / \$2,040 a year	0 to 100, & their own page admits it measures agreement, not accuracy	None found	No	No	No	Audit trail, governance dashboard, peer review	Real positioning, thin proof	

Talkory.ai	Free / from \$5 / Enterprise custom no period published	A % plus an agreement count	None found. Each AI reviews only its own answer, which is not a loop	No	No	No	SSO, white label, data residency	Polished, feels like a small team
MultipleChat	Free / \$20/mo / \$90/mo / \$240 / \$1,080 a year	Flags where answers disagree, no number	1. A role-assigned review debate	No	No	No	SSO, Trust Center, team admin	The most built out, 30,000+ users claimed
Council AI	\$19.99/mo / \$39.99/mo / \$99.99/mo / \$240 / \$480 / \$1,200 a year	A number, method not published	1. A debate, models challenge each other	No	No	No	Works paces + an MCP server	Thousands of users claimed
Suprmind	\$19/mo / \$45/mo / \$95/mo / \$195/mo + custom / \$228 / \$540 / \$1,140 / \$2,340 a year	A map of where they agree & disagree, no score	1. A debate, in 3 flavors: debate, red team, first principles	No	No	No	Enterprise tier	Real pricing tiers, active marketing
AISCouncil	Free / \$5/mo / \$60 a year	None	1. A debate, with peer review & a vote inside it	No	No	No	Open source	Small, maintenance in progress
AiZolo	Free / \$9.90/mo / \$99.90 a year, their own annual price	None	None found	No	No	No	None	Small, built for search traffic
OneScales / MultiLLM.pro	Not verifiable	Not verifiable	Not verifiable	No	No	No	Not verifiable	Thin

* GTM partner to help decide.



Prices & features are each company's own claim about itself, re-read off their live pricing & marketing pages on 14 September 2026 (first read 7 August 2026), not independently verified. Billing period is shown where the company publishes one; Talkory.ai does not publish one for its paid tier. **The yearly line under each price is there so the column can be compared at a glance. Every 1 of the other 7 that publishes a period bills monthly**, checked live the same day, so theirs is the listed monthly price × 12 & is a ceiling: most software discounts an annual plan. AiZolo is the 1 exception, publishing its own annual price. **Ours is our real yearly plan: 10 months paid, 2 months free.** The loops column counts what each company advertises on those pages, so read "None found" as "not advertised", never as proof there is none. Debate/cross-exam modes are now common across this group, so we never pitch ours as the first to try it, only as the one that goes further. Full sourcing: the competition appendix.

Not 1 of them keeps a record of who turned out to be right.



The direct competition (pay experts by the hour to grade & correct AI): Mercor · Surge AI · Handshake AI · Micro1 · Outlier (Scale AI) · Alignerr (Labelbox) · Mindrift (Toloka) · DataAnnotation · Turing · Invisible · Pareto.AI · Prolific · Uber AI Solutions · xAI hiring direct · Appen/CrowdGen · Snorkel. Rates run \$15 to \$200/hr, Mercor at the top.

The old guard who already pay experts hourly: GLG · AlphaSights · Third Bridge · Guidepoint · AlphaSense/Tegus · Dialectica, plus about 10 smaller ones. A \$3B market, ~150 firms.

Closest to what we actually do: Doximity PeerCheck has 10,000+ doctors reviewing AI answers. Then JustAnswer, Wyzant, Chegg.

& not 1 lets the expert keep what they earned or turns specialized hobbyists or professionals (not looking for a job) into opt-in instant consultants. They are all work-for-hire for the labs, paid by the lab rather than by the people asking, so the record stays with the lab.

Sources



Hourly rates come from recruiting blogs & worker reports, not from the companies themselves, so treat them as rough. Company sizes come from press reporting. Gathered 16 August 2026.

60 features built & running. 7 left to build.

17



Answers & accuracy



The search engine & the 3-AI comparison
Auto-routing to the best AI for your topic
Certainizer™ settles the disagreement
Hand-picking your own 3 AIs
Fact Check
AI Debates, free for everyone
Search by photo
A commentator on each answer, in the tone you choose
3 suggested follow-up questions
1 thumbs up or down grades all 3 AIs, & typed corrections are kept
The Right/Wrong slider
Share an answer & get the credit
Live accuracy & safety scores for each AI on each topic
Catching fake corrections & groups voting together
Life Coaches: advice compared across several AIs
Safe-to-Act Check, before an AI agent buys, books or sends
Get Measured: free grading for anyone who runs an AI

- Search by photo
- A commentator in the tone you choose
- The Right/Wrong slider

15



Your notebook & staff



The notebook
Adding your own files & links
Finding an old question by meaning
Job Scout, 3 months while hunting
The notebook auto-organizer
Importing your notes from Notion & Evernote
Your own AI assistant can read & add to your notebook
Your knowledge as a 3D picture you can spin
Each staff member keeps its own to-do list
Agenda reminders with a notification bell
Search history page
CV quick-fill from Word, Pages, photos & more
Welcome email
The advisor on every page that answers questions about the product
A feedback box on every page that reaches a real person

- The auto-organizer
- Imports from Notion & Evernote
- Your knowledge as a 3D picture

9



Community & experts



A group forum with our AI built in, used by a real group
A map of the group's knowledge
Find an expert by describing what you need
Expert profiles & a short personal address
Track-record points & achievements
Invite-a-friend links & referral badges
A weekly email sending open disputes to experts
Trending in your topics
Emails to followers & topic subscribers

- A live group forum with our AI inside
- Expert profiles & a short personal address
- Track-record points

11



Companies



Company plan with seats & an admin page
Online setup questions & pilot sign-up
Shared team history: choose My team or Only me
Team space: already asked, open questions, threads, department folders, translation & who to ask
The company's own Job Scout
Sign in with Microsoft
Slack & Microsoft Teams apps
Company browser extension
A shared monthly allowance of the top AIs
A different AI level for each department & a monthly AI savings report
Security & data sheet

- A company plan with an admin page
- A shared team space
- Slack & Teams apps

3



Sites & sign-in



1 shared set of parts (sign-in, payments, invites, email) runs across our 9 live sites, so each new site reuses them instead of starting over.

1 sign-in across all our sites, by Google, email link or Microsoft

9 live web addresses

A summary page AIs can read, on both business sites

- 9 live sites
- 1 sign-in across all of them
- Google, email link or Microsoft

5



Plans & payments



7-day free Plus trial, no card

Free plan: 5 searches, with a countdown to the next 5

Checkout for every plan, weekly, monthly or yearly

A billing page people manage themselves

Founding pass: Plus free

- A 7-day free Plus trial
- Checkout for every plan
- A billing page people manage themselves

Still building

1. Being found & hired by companies



Free: companies pay experts through us, we take no cut.

2. Watch my own documents



Insurance, mortgage, tariff, handbook. Tests said it is worth **\$15/mo.**

3. Called it first



Timestamped, public proof. Tests said it is worth **\$10/mo.**

4. Deep Check



one careful answer on something that matters, from real people. **Price & plan are both open.** It is the 1 thing here that costs a person's time rather than a machine's, so what it is worth & what a week of it can promise are a GTM conversation, not a number we set alone.

5. What AI tells people about you or your company



Companies can already pay to find out IF an AI mentions them. **Profound, Otterly, Peec, Scrunch & Rankscale** sell it, & **Semrush & Ahrefs** give a version away free. Every one of them tracks the same 3 things: are you mentioned, are you cited, is the tone kind. **None of them check whether the thing the AI said about you is TRUE.** We will run the claim through our own panel, show which AI said what, & flag what is not correct like we do.

Phase 2, after launch.

Looking for a GTM Partner

THE OFFER

Tell us what your terms would be to take over the promotion & running of the websites. Our current idea is: take 100% of quick ramp-up net revenues



Net is after low AI costs, card fees & contributors' share. Contributors are paid in parallel, not after your 3x: when a company hires one, it pays them through our Stripe checkout, straight to the expert, & we take no cut.

until you've recouped 3x your time & costs + equity & ongoing profit sharing from AI lab revenue & all other revenue. Propose the deal that makes sense for your cost structure, current GTM staff, & resources.

Revenue progression: Companies & Consumer first, then anonymized human corrected data revenue from AI labs.

What flows to the GTM/Ops partner



Built from real people & real hours, never a slice of a market. The floor is what subscriptions alone produce. The second line is the AI-lab line, worked out on its own.

Reaching... (paying members)	Subscriptions		Share 50%?*		Share 25%?*	
	Year 1	Every year after	Year 1	Every year after	Year 1	Every year after
1,000 + 5 companies	~\$41K	~\$77K	~\$250K	~\$490K	~\$370K	~\$740K
10,000 + 20 companies	~\$476K	~\$880K	~\$2.5M	~\$4.9M	~\$3.7M	~\$7.4M
100,000 + 100 companies	~\$4.3M	~\$8.0M	~\$25M	~\$49M	~\$37M	~\$74M
1,000,000 + 500 companies	~\$42M	~\$78M	~\$250M	~\$490M	~\$370M	~\$740M

* The AI-lab line, after members get their share. How much to share with members is your decision as the partner.

Mature content filters are your decision too: whether they ever launch, & whether they start on or off.

The AI-lab line is upside, not the plan. It is never counted inside the subscription figures, & nothing here depends on it.



Subscriptions: 100% of net revenue goes to you until you have recouped 3x your cost basis, then equity on top. Not a share of it, all of it. The year-1 column is a ramp to that size, not a full year at it.

The year-1 columns apply a 50% ramp discount, a rounder & more conservative number than the ~54% the subscription figures already use, so this errs low.

How the AI-lab line is built, & the second way it was checked, are set out in full below.

No advertising money is assumed anywhere in this table. Every sign-up in it arrives through outreach, not through buying attention. That is the half we are asking a partner to own.

The margins, plan by plan

Somebody else’s numbers, not ours. A crypto exchange with 2,500 engineers cut its AI bill by roughly half by sending each job to whichever model could do it. Usage went *up* while spend went *down*. **Our cost argument, at scale, from a company with no reason to help us.** The Information, 4 August 2026.

	\$5/mo Certainizer™	\$9/mo Plus	\$19/mo Pro																
How much of every \$1 we keep																			
After paying for the AI & the card fees. How much conversation we remember is a dial, not a fixed cost. Today we remember your last 3 questions, so a follow up makes sense without you repeating yourself. Remember more & people can hold a longer back & forth, the way they do in their AI app. Remember less & we keep more of every dollar. A partner sets it wherever they want it. <table><tr><th>Memory (before card fees)</th><th>\$5/mo</th><th>\$9/mo</th><th>\$19/mo</th></tr><tr><td>1 question (estimate)</td><td>93%</td><td>94%</td><td>97%</td></tr><tr><td>3 (today)</td><td>89%</td><td>92%</td><td>96%</td></tr><tr><td>6 questions (estimate)</td><td>83%</td><td>87%</td><td>94%</td></tr></table> The figures on this row are after the AI & card fees, so the whole table reads on 1 basis. Before card fees they are ~89%, ~92% & ~96%. ~80% ~85% ~92%	Memory (before card fees)	\$5/mo	\$9/mo	\$19/mo	1 question (estimate)	93%	94%	97%	3 (today)	89%	92%	96%	6 questions (estimate)	83%	87%	94%			
Memory (before card fees)	\$5/mo	\$9/mo	\$19/mo																
1 question (estimate)	93%	94%	97%																
3 (today)	89%	92%	96%																
6 questions (estimate)	83%	87%	94%																
Profit from 1 subscriber, 1 month																			
real money, not a percentage. After the AI & card fees, same basis as every row below. \$4.02 \$7.69 \$17.40																			
What 1 subscriber is worth over a year																			
Lifetime value, or LTV. Total profit we would expect across 12 months, illustrative. After the AI & card fees, worked out row by row in the forecast page at the back. \$48.24 \$92.28 \$208.80																			
	\$5/mo Certainizer™	\$9/mo Plus	\$19/mo Pro																

What we get back for every \$1 spent winning a subscriber i A year's profit ÷ the cost to win them. The trade writes it LTV : CAC. CAC is customer acquisition cost, what winning 1 customer costs us. LTV is lifetime value, what that customer is worth to us once they are here. Every row uses profit after the AI & card fees: \$48, \$92 & \$209 a year. Investors look for at least \$3 back for every \$1 spent winning a customer. Pro gets there at every cost shown, Plus up to \$30 a customer & Certainizer up to \$10. We have not spent anything on advertising yet, so there is no real cost to divide by. Setting that number is a partner's first job, & we would rather agree it with you than invent it.			
at \$0.17 i The AI cost of the free trials it takes to win 1 subscriber, if outreach & the viral loops carry it with no advertising. What the AI calls cost us to win 1 paying customer, assuming 1 in 10 free-trial users converts. It is not \$0: we do pay for the free trials. There are no card fees in this figure because a free trial never touches a card, & the card fees on the ones who do pay are already taken out of the profit every row above divides. It is the smallest piece of what winning somebody costs, not the whole of it.	~284 : 1	~543 : 1	~1,228 : 1
at \$10 to win them	~5 : 1	~9 : 1	~21 : 1
at \$30 to win them	~1.6 : 1	~3 : 1	~7 : 1
at \$60 to win them	~0.8 : 1	~1.5 : 1	~3.5 : 1
How fast that \$0.17 comes back i The finance phrase is CAC payback. This row answers the \$0.17 line only. At \$10 to win somebody it is under 3 months on every plan, at \$30 about 2 to 8 depending on the plan, at \$60 about 3 to 15. Investors want that money back within 12 months. Every plan does it at \$10 & \$30 a customer, & Plus & Pro do it at \$60 too.	~1 day	<1 day	<1 day

We are guessing at how fast users arrive & stick. We are hoping our GTM partner will have better guesses & the experience to know which of our strategies are good, which need fixing, & better strategies to share.

3 revenue streams are EXCLUDED from every figure above.

All 3 are built into the product. **Not a cent of any of them appears anywhere above, so every number here is a floor, not a forecast.** Deep Check pay-per-answer · bounty commission · companies buying verified answers. **The AI-lab corrections corpus is now forecast on its own, in section 6 of the forecast page. The company & AI-lab lines are the destination. The consumer subscriptions above are the bridge.**

The AI-lab line is the biggest of them all.

What the AI-lab line is worth

Where those come from: items times a price per item. A quick 👍 or 👎 sells for little, a written correction for much more, & both add up.



The middle case, & what it assumes: each paying member averages 5 👍 or 👎 a day out of about 15 searches, sold at **30¢** each, the middle of the 10¢ to 50¢ a lab pays for a quick rating, plus **a written correction every 2 to 3 days**, sold at **\$3** each, the published middle for an expert correction (the range is \$1.50 to \$8+). Some do none & some a lot; this is the average. That is **~\$985 a year per paying member**, if every rating & correction sells. Low & high cases are on the forecast page.

Shown in full. Members are paid 50% of it when it earns money.

The market we are selling into is ~\$2-7 Billion/yr today, growing about 30% a year, reaching **\$16.4B by 2030** (Research & Markets) & **\$25.0B by 2032** (Global Market Insights). 1 firm forecasts \$44.68B by 2035; **we used the low end on purpose**. Set against those: at 10,000 paying members our line is ~0.1% to 0.5% of today's market, at 100,000 it is ~1.4% to 4.9%, & at 1,000,000 it is ~3.9% to 6.0% of the 2030 to 2032 forecast.

Checked a second way, from the other end. Scale alone was making about \$2B a year, roughly 45% of it goes to the people doing the work, split across Scale's ~240,000 contractors. That is **~\$3,750 a year each**, against our **~\$985** a year per paying member.

We did not use share of market. On those forecasts 1% would have been ~\$210M to ~\$260M. **Not a cent of it is in any number above.**

Separately, & this is our own estimate rather than a published figure: companies will spend somewhere around **\$10B to \$15B a year of their own staff time** checking & fixing what AI gave them. That is not a market we sell into. **It is a bill we take off their desk.**

Our users can produce more of this data than the world currently buys, which solves the problem of AI being out of control.



The people side costs us \$0 up front. The people who check answers are paid out of what that check earns, split 50/50* (* what we would promise users if the GTM team agrees). A share of the money, not a wage. Paying them by the hour instead is the 1 line that would sink a margin like the one above.

Every figure in this money block is our own estimate, built bottom-up, & is reproduced from the line-by-line forecast where each one is worked out in full. Built & running; opens to the public at launch.

What's in each plan

Unlimited on anything that only thinks. Metered on anything that looks something up, posts, or runs on a timer.

Accuracy climbs with the plan: ~60% asking 3 AIs yourself, unaided → ~82% on Free & Certainizer \$5/mo → ~95% on Plus \$9/mo → ~95% on Pro \$19/mo.



Deliberate: those 2 tiers compete on features & staff, not accuracy. 3 well-chosen AIs already gets you within 2 points of Claude Opus 5 alone. **~60%:** on hard & fast-changing

questions, asking on your own.

Feature	\$5/mo	\$9/mo	\$19/mo	What tests said it's worth stand-alone
What changes from plan to plan				
Als that can answer i A better set chosen for your topic. On Plus any of 100 can take your question, on Pro any of 169, & the best 3 run, weighted to whichever 1 has the strongest track record on it. You can let the topic routing pick, or hand-pick 3 yourself on Plus & Pro, which is built & live today.	80	100	169	Plus is 85% cheaper · Pro saves ~\$7,000/yr, vs paying each AI i A year against a year, both sides. The yearly plan charges 10 months & gives 2 free, so it is \$90 a year on Plus & \$190 on Pro. Buying 3 mid-tier AI subscriptions separately runs about \$720 a year, so Plus costs over 85% less. The 3 top-tier ones run about \$7,200 a year, so Pro saves about \$7,010.
The top Als: Fable, Opus, o3 & GPT-5.6 Sol i 150 calls a month on Opus & o3, shared, & 10 a month each on Fable & GPT-5.6 Sol. Each can be raised for \$4/mo more.	—	—	✓	
Als from outside the USA i Plus: Kimi, Qwen Max, DeepSeek & GLM-5.2 (China). Pro adds: DeepSeek R1 (China) & Mistral (France).	—	4	6	
Follow-up & out-of-date answers i The 1 thing all 3 professional testers picked unprompted as the feature they would keep. It is metered either way, never unlimited, because every watched question is a bill that repeats forever.	5 facts	20 facts	50 facts	\$40-50/mo as a build check
Find an old question by meaning	—	30 days	all	
Search by photo	—	✓	✓	
Fact Check	✓	more checks	unlimited	
What email needs you today	✓	+ catch-up on threads	+ drafts replies	
Staff i Certainizer: Fact Check, Out-of-date answers, Self-contradiction checker, What email needs you today, Drafting social posts, First Draft, Life Coaches, Note Tidier, Warranty Manager, Refund Chaser, Question Monitor & Job Scout. Plus adds: Meeting Prepper, Scheduling Desk, Product Hunter, Company Reviewer & Form Filler. Pro adds: Visibility Fixer, Notebook Auto-Organizer, Translator & Out-of-date things you wrote.	starter set	+5 more	+4 more	
Notebook Auto-Organizer: files on your PC & cloud storage	—	—	150 entries a month	
The same on every plan				

Auto-routing to the best AI for your topic				
<i>i</i> Free on every plan, including the free one, & that is deliberate. Your question goes to the AIs that score best on its subject, & every graded answer makes that routing sharper. What you get free is the right AI on every search. The scoreboard underneath stays ours.	✓	✓	✓	Free on every plan
Certainizer™ settles the disagreement	✓	✓	✓	
History kept while you're subscribed	✓	✓	✓	
Job Scout, 3 months, while hunting · in Certainizer	✓	✓	✓	\$15/mo hunting, \$0 the week they sign

People independently priced a single high-stakes answer between \$5 & \$20. No price shown for Fact Check: unanimous that nobody would pay for it, & that charging for it would change how they read the company.

Who buys companies like this

- **Stripe agreed to pay \$7.5B to see where AI spending goes, but hasn't closed as of this writing. We tell them who got it right. That has to be worth more.**



Reported by Bloomberg on 16 August 2026 ("more than \$7 billion") & at \$7.5B by the New York Times & CNBC on 19 August. OpenRouter had raised at \$1.3B a few months earlier.

The deal is agreed, not closed, & we have no involvement in it. What those records hold is which model answered, what it cost & how fast. OpenRouter grades AIs on fixed test questions. **Nothing grades the real answers its customers got.** OpenRouter's own ranking page says the ranking measures "adoption, not quality".

- **Lovable is worth \$13.3B & is openly shopping for companies to buy.**



Lovable raised **\$400M at a \$13.3B valuation** in August 2026. Its own chief executive asked publicly, in March 2026, for startups & teams to bring acquisitions to its head of M&A. **So this is a buyer with money, looking now, rather than a price anybody paid for a company like ours.**

- **Reddit & LinkedIn hold the argument & the credentials. Neither holds the answer. (potential acquirer)**



Reddit has the arguments & the votes, which record what people **preferred**, never what was right. **LinkedIn** has the credentials people claim, never a record of whether their answers held up. **Both already hold the audience we would bring an answer layer to.**

⚠ **LinkedIn is owned by Microsoft**, so these are 2 doors into 1 buyer, not 2 buyers. Said plainly here because a reader who knows that & is not told will read the list as padded.

- **Microsoft says the moat is the learning loop, not the model. We could be a natural independent fit for them because they can't appear independent.**
(potential acquirer)



From Satya Nadella's essay *A Frontier Without an Ecosystem Is Not Stable*: "*The real opportunity is not in picking the best model but instead in building a learning loop on top of models where human capital & token capital compound.*" **Microsoft has not bought anything here & has not bid for anything of ours.** The point is the strategy: it does not own one of the leading models, so its whole plan depends on the loop being the moat rather than the model.

Funded

Founded & funded by a former VC, not raising money & tiny burn (mostly patent attorney)

Running

Built & tested today, not a plan

~80-92%

Of every \$1 we keep, after the AI & card fees

Patent Pending

Reviewed by a top 10 patent firm

Founders & team



Bob Haya, Founder

Retired venture capitalist

Earlier in his career, Bob had 2 exits & co-invested alongside firms like Andreessen Horowitz, Kleiner Perkins, & Khosla Ventures, etc. See:

createlabs.com/portfolio/portfolio.html

Bob still collaborates with his billionaire VC partner & informal advisor Cyrus Hodes, both personal friends with contacts at the highest levels of today's top AI companies & the White House.

[Click on ABOUT US here](#)

The longer bio reallysolved.com/gtm/bio

Open to any ideas, including adjusting the business model, or maximizing the underlying engines for entirely different applications.

Already built several acquisition funnels with persona variations (more verticals & personas already spec'd). Patent pending, reviewed by a top-10 patent firm.

"I want another source of ground truth in the world, and I want to contribute to the gig economy."

"I'm easy to work with, Open to changing Branding...just blow it up!"

"Twice I walked into a government office with the wrong forms & procedures. Both times I had asked the best AI there is. Both times it answered confidently, & both times I was embarrassed & sent home. AI is helpful but useless if I can't tell which parts are correct & there's no way to fix it for the next person."



Mathew Svensson

Co-founder, B2B

Mathew spearheads the Hands-On B2B consultative services of ReallySolved & is focused on the workflows between departments & how to apply AI most efficiently. MSc in Technology Entrepreneurship, Technical University of Denmark. Over a decade helping companies build, streamline, & scale their operations. Recently got the highest possible grade on his AI related thesis, 12 (A+), and was told that all his sub thesis work would have been normally a thesis each on their own.

"I design intelligent organizations. Most companies think they have a technology problem. Usually it's a flow problem: information, decisions & work get stuck between teams, and adding AI on top doesn't unstick them. So I ask what kind of organization you want to become, and what's stopping you. I address flow first &

tools only where needed. I connect specialists so their work creates value. When I leave, your organization understands itself better, executes better, and can keep evolving without me."



Cyrus Hodes

Informal advisor

Founder of AI Safety Connect, which convenes governments, frontier AI labs & investors alongside the UN General Assembly & the international AI summits. Co-founder of The Future Society at the Harvard Kennedy School. Leads the Safety & Assurance of Generative AI project at the Global Partnership on AI, sits on the OECD expert group on AI futures, & runs the GRASP global risk mapping at the Mohammed Bin Rashid School of Government in Dubai. Venture partner at Lionheart Ventures. Formerly advisor to the UAE's AI minister.

Where the people come from

People do the work for money, status & a reputation they can use.

STEP 1

Consumers

Free multi-AI search. An answer costs us almost nothing. * **BUYER 1, IN SMALL AMOUNTS**

STEP 2

Experts



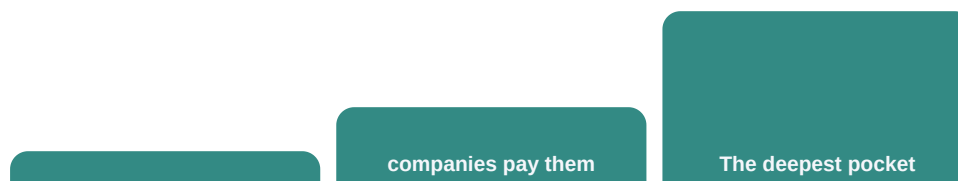
Expert-supply channels: direct recruiting (Reddit, StackOverflow, GitHub), professional-association partnerships, corporate HR outreach, content-led (LinkedIn first), or a paid referral program. Which first, & how much budget?

A few searchers know a subject better than the AIs do. They show it in public. †

STEP 3

AI labs

They buy the questions the AIs got wrong, plus the checked answer. **BUYER 2, THE BUSINESS**



* An answer costs us **85¢ per 10,000**. 4 tiers: free, \$5/mo, \$9/mo, \$19/mo.

† **We line the buyers up first, then pay.** Companies are phase 1, run by hand: ask 20 what they would pay before a line of code. If nobody says yes, we saved a year. We will ask companies what they want & design the bounties program around that before we engage experts on the platform & outreach to other experts. **That is bounties. Everywhere else both sides arrive together:** the games bring the people in while the outreach brings the companies, & neither waits for the other.

EARN MONEY · 1 OF 4

EARN MONEY · 2 OF 4

Get hired directly



Build a strong reputation & real companies pay you for your expertise - no résumé, no cold outreach.

Answer for a stated price



companies offer a price for work: **"\$40 to answer this. About 10 minutes."** We pay 50% when your 👍/👎 earns money. **Written corrections that hold up are paid, & lab sales are split 50/50* with members.**

* What we would promise users if the GTM team agrees.

EARN MONEY · 3 OF 4

Win disputes



Catch an AI's mistake, get featured for it, get noticed by companies watching who's actually right.

EARN MONEY · 4 OF 4

Weekly contests



Find the best AI mistake of the week, winners get real money or platform credits.

GROW FOR FREE

Invite a friend



You both get free access to better AI engines. If they become a paying customer, your reward doubles.

GROW FOR FREE

Creator badges



We find creators, verify their expertise with AI, & give them a badge worth posting, their followers join us.

HABIT & STATUS

Streaks & leaderboards



Keep a daily search streak alive, climb the public rankings - the same habit loops as Duolingo or Snapchat.

HABIT & STATUS

Dispute Hunter challenge



A weekly goal, "find 10 AI mistakes this week", turns the core job into a game people want to win.

EXTRA LOOPS DISCOVERED ALONG THE WAY

Innovative group features not available on any platform today



We are building a layer that sits inside communities somebody else already runs. It asks 3 AI models every new question, posts an answer only when they agree, & stays silent & calls a human when they do not. For fault-finding questions it returns every possible cause rather than one answer, & later asks the person which one it actually was. The whole idea depends on people answering that question, & we measured how often they do it today: between 1% & 7%. We Believe with our new features we can make groups more engaging, productive, & pay Members for their human corrected data as a byproduct.

THE QUIET ONE · RUNS FIRST

Your notebook



What you already know, saved as you search. It tells us **which companies would want your expertise**, so it can earn for you before you answer a single question.

We provide you a staff that uses your notebook to provide all kinds of services.

Working now: Job Scout, the self-contradiction checker, Meeting Prepper & Fact Check.

Offered now: Out-of-date answers, your inbox, your calendar, your warranties & a first draft instead of a blank page. **All 20 are written out in full at the back of this deck. You choose which ones run, & nothing is ever sent anywhere until you press send.**



1 Track record - moves on everything they do, including a 👍 or a 👎. It is standing, never cash. It is what gets them offered the paid work

2 Correction Earnings - moves only when a solution or a correction is accepted. Shown in \$

A bot that farms the first one gets nothing it can spend.

Our first real test.



We already have an existing enthusiast community, ~2,155 members & ~99 posts a week in its Facebook group. We & its admins are adding features to its new forum on our Oracle server, then bringing those members over. Normally we don't host the group we just add to it but this is a good test to see if we can improve their group, charge \$29 a month (our hosted price for groups up to 5,000), & get the human correction data, the real value to us.

We only post when the answers are confident, so the handful of people who really know things only see the hard ones. **Nobody else does our last step.** When a thread goes quiet, we gently follow up with people who generated the questions or proposed answers depending on the semantic meaning of the thread. We read 285 real fault-finding threads across 3 communities to measure it. Not because people are unhelpful. Because nobody asks them. It pays for itself, & the corpus pays more.

We are considering turning the brightest forum members into instant consultants having their advisor send out their Job Scout. It is also how people find us. Members watch it work in their own forum every day.

A group with an active Facebook membership now has its own forum, & its members come over next. **The wider plan is to give group owners better tools at a price they can afford**, by partnering with the companies that host communities like theirs.



Built & running; opens to the public at launch. Every number in this deck is either our own testing or somebody else's published research, labelled as one or the other. **What a partner is looking at is a machine that runs, waiting for the people to point it at.**

 **Why it compounds:** about 3 searches in 5 turn up a disagreement, & every disagreement is free raw material for the next answer. **On our 91-question everyday bank, 55 did.** 9 loops feed each one back to the top of the flywheel.



Answer for a stated price · Get hired directly · Win disputes · Weekly contests · Invite a friend · Creator badges · Streaks & leaderboards · Dispute Hunter challenge · Your notebook

The cheapest door in: Fact Check

Paste what another AI told you. We say where it's shaky. **No account, no email, & never a signup in front of the result.**

- **It works on people who have never heard of us.** That is what makes it distribution, not a feature.
- In people & persona testing, every one said they'd tell a friend about it, & every one also said they would not pay for it.
- **The email ask, if any, comes after the answer is already on screen.**
- It is also the cheapest proof we have that we are not just a wrapper around one AI: it exists specifically to catch the others.

17 / 22

Contributor, sign up, first search

Solver, first paid job · \$500-\$2K/yr

Senior Solver, a track record · \$5K-\$15K/yr

Master Solver, a strong year · \$50K-\$200K+/yr

Solver Laureate, a career capstone

The ~500 Topic & Safety Matrix. The board below shows 8 topics out of ~500 researched or tested variables, because 1 AI is better at stem cell science & another at robotics, & calling them both “science” hides that. The same record runs a 2nd time for safety, across 245 safety topics, & a 3rd time for the 7 questions with no right answer & 4th time to check out of date answers (all included in the ~500).

The leaders keep changing. We watched it move this month: the model we tested ourselves against was retired 6 days later, & a second one was discontinued the same week. Yesterday's best is not today's best.

So we keep the list, subject by subject. From launch, which AIs are best at which topic is rewritten by real corrections on our own platform, as answers come back & people correct them. Here is the board.

ENGINE	TIER	CRYPTO	BIOHACK	CURRENT EVENTS	INVESTING	CYBERSEC	MEDICAL	SCIENCE	LEGAL										
Claude Fable										Pro	78	85	70	88	90	92	95	91	
Claude Opus										Pro	63	82	68	86	88	90	93	89	
OpenAI o3										Pro	82	84	72	90	94	93	96	90	
GLM (non-USA)										Plus	70	72	78	75	76	78	82	79	
GPT-5.6 Sol										Pro	80	83	82	88	91	91	94	88	
DeepSeek R1 (non-USA)										Pro	80	77	60	88	88	81	94	79	
Claude Sonnet										Plus	75	80	68	82	86	86	90	85	
Claude Haiku										Plus	62	65	60	68	70	72	76	72	
GPT-4o										Plus	72	75	80	80	82	84	86	82	
Gemini Pro										Tested	44	78	90	0*	84	88	92	84	
Kimi (non-USA)										Plus	66	68	78	74	72	75	80	80	
Qwen Max (non-USA)										Plus	72	74	80	78	80	82	88	81	
DeepSeek (non-USA)										Plus	76	75	65	82	84	80	91	78	
GPT-OSS										Free	50	52	45	55	58	56	62	54	
Gemini Flash Lite										Free	48	60	82	45	62	68	72	60	
Llama 3.1 8B										Tested	54	55	48	58	62	60	68	58	
Grok										Pro	75	74	92	80	80	78	86	76	
Mistral										Pro	70	72	65	76	82	78	85	84	
Command R+										Tested	60	62	75	78	68	70	72	82	

SEE 200 OF THE 203 ACCURACY TOPICS ▲

FORMAL & ABSTRACT SCIENCES

Logic & Reasoning

1. Formal Logic
2. Informal Fallacies
3. Mathematical Logic
4. Modal Logic
5. Proof Theory

Pure Mathematics

6. Number Theory
7. Abstract Algebra
8. Geometry
9. Topology
10. Mathematical Analysis

Applied Mathematics

11. Calculus
12. Linear Algebra
13. Differential Equations
14. Numerical Analysis
15. Game Theory

Statistics & Probability

16. Descriptive Statistics
17. Inferential Statistics
18. Bayesian Statistics
19. Stochastic Processes
20. Probability Theory

● 91-100 expert ● 76-90 strong ● 51-75 competent ● 21-50 weak ● 0-20 unreliable

*These board scores are our best current read of the landscape (public tests, July 2026), not our own lab measurement yet. The 169 AIs we tested ourselves are a separate count, on page 1. "Tested" = scored by us, not on a plan. *Gemini's investing score is a real result, not a data-entry error: it's built to refuse investment questions.*

The safety half: 245 safety topics, 22 categories reallysolved.com/gtm/safety

We are not the only ones **saying this.**

Bill Gates published a nearly 6,000-word essay on August 26th 2026 warning that artificial intelligence is advancing far faster than governments and society are prepared to handle.

BILL GATES · AUG 2026

“AI is confident, not always correct.”

ETHAN MOLLICK, WHARTON

“Base model capability is commoditizing on schedule. **Trust isn't.**”

DR SETH DOBRIN, SILICON SANDS NEWS · AUG 2026



His argument across 35,000 words is that the AI itself keeps getting cheaper & more interchangeable, while being able to PROVE an answer is right stays scarce. That is the business we are in.

“Chat AIs are trained to be agreeable, and **agreeable advice is worthless the day you're about to make a real mistake.**”

HYPERAUTOMATION AI REPORT · AUG 2026



They shipped a prompt that makes 1 AI argue with itself, & attacked it 8 ways to check it would not cave. Ours is 3 AIs that disagree on their own, so nobody has to remember to ask for the argument.

Fractl's survey found that people are not walking away from AI. **They are using it more & believing it less.**

FRACTL / SEARCH ENGINE LAND · JUN 2026



70% say they use AI tools for search more than they did last year, & only **3%** use them less. Over the same 12 months the share who find AI search **more** helpful than a normal search fell from **82% to 54%**, & the share who find it **less** helpful went from **3% to 17%**. Survey of 1,008 US consumers & 150 marketers, Q2 2026. **People are not leaving. They are staying & trusting it less, & that gap is the business we are in.**

Dr Seth Dobrin has said that the AI that answered you last month is not the one answering today, & nothing tells you it changed.

DR SETH DOBRIN, SILICON SANDS NEWS · AUG 2026



His words: *"The model you approved is not the model you're running."* Security review, evaluation results & risk sign-off all attach to a version that no longer exists in your environment. **Watching what you have already been told, & flagging it the moment something newer disagrees, is the feature every professional tester picked unprompted as the one they would keep.**

→ *We watch for that.*

CustomerThink has said that the people who built it cannot say why it answered that.

CUSTOMERTHINK · AUG 2026



The article's point: nobody can see what a customer asked an AI or what it told them, because the systems are closed & the conversations are private. **Even the developers cannot fully account for their own systems' outputs.** Paraphrased, not quoted. **So the only place left to check is the answer, which is the one thing anybody can hold up & compare.**

→ *Somebody outside has to check the answer itself.*

Dr Seth Dobrin has said that as AI starts **doing** things instead of just saying them, an unchecked answer stops being a nuisance **& becomes a liability.**

DR SETH DOBRIN, SILICON SANDS NEWS · AUG 2026



His words: *"A wrong answer is an information-quality problem. A wrong tool call with live credentials is a security incident."* **The cost of not checking is about to rise sharply, & it rises for the buyer, not for the AI company.**

VentureBeat has reported that a new \$6B AI company graded its own accuracy against OpenAI, Anthropic & Google, **on a test it built itself.**

VENTUREBEAT · AUG 2026



Hark raised **\$700M at a \$6B valuation** in May 2026, with Nvidia, AMD, Intel, Qualcomm & Salesforce among its backers. At its first product launch it published a score of **97.7** for itself, against **92.8** for OpenAI's best, **84.1** for Claude & **69** for Gemini. VentureBeat's own caveats: Hark supplied the comparisons, 2 of the 3 tests ran *"inside Hark's own harness, with pass rates computed by Hark's internal LLM judge"*, the rival models were set to their slowest setting, & *"no third party has posted"* results for the newest rivals, so the claim *"cannot currently be checked"*. **The scores a buyer sees were written by the company being scored.**

American Express has said it will now cover the bill when an AI agent buys the wrong thing.

AMERICAN EXPRESS · APR 2026



On 14 April 2026 Amex launched a kit for AI agents that shop for you, with **Amex Agent Purchase Protection**, which it calls “a first-of-its-kind commitment to protect registered AI agent purchases”. It is conditional: the agent has to be registered, you have to have authorized it, & the purchase intent has to reach Amex. **A card company putting its own money behind an AI’s mistakes is the moment being right starts having a price on it.**

→ *Somebody still has to decide what “wrong” was.*

The industry says picking **the right model for each job** is how AI stays affordable.

NVIDIA, ANDREESSEN HOROWITZ, Y COMBINATOR, MICROSOFT, META & OTHERS · JUL 2026



From the open letter *Open Weights & American AI Leadership*, 24 July 2026, also signed by Hugging Face, Mistral, Perplexity & IBM: “Open weights let every organization **match the right model to the right job at the right cost**, reserving frontier-scale capability for genuine frontier problems. That discipline is what will make AI economically sustainable.”

→ *That is our engine.*

→ *We match each question to the right AI at the right price, & cheap independent AIs check the expensive ones.*

“The real opportunity is not in picking the best model but instead in **building a learning loop on top of models where human capital and token capital compound.**”

SATYA NADELLA, MICROSOFT · JUN 2026



From his essay *A Frontier Without an Ecosystem Is Not Stable*.

The biggest company in business software is describing a need for our business.

Microsoft does not own one of the leading models, so its whole plan depends on the loop being the moat rather than the model.

→ *The winner is not the best model. **It is the loop of people correcting it.***

Ben Lorica has said that companies are learning to turn their own data into specialist AI they own.

BEN LORICA, GRADIENT FLOW · JUL 2026



His point: companies are learning to turn their own data into specialist AI they own, so the big labs “start competing against their own customers’ data.” When training a specialist model gets cheap, the scarce input stops being compute & becomes verified human correction with a source attached. That is the one thing money cannot shortcut, & it is exactly what our flywheel manufactures.

→ *When everyone can train their own AI, **verified human correction becomes the scarce part.***

The Financial Times & Bloomberg have reported that the market is now pricing what **frontier AI costs to build**.

FINANCIAL TIMES / BLOOMBERG · JUL 2026



“Cost of insuring against default by AI hyperscalers **hits record levels**.” Oracle’s 5-year default insurance is at its highest since 2008, & S&P cut it to 1 notch above junk, on data-center spending. We buy intelligence by the answer, from whoever is best & cheapest that day.

→ *We never take that cost on.*

Why we get better every day, & why a copycat can't catch up.

Published test, February 2024: the same AI got 93% right on facts that don't move, & 26% on facts that do. 125 questions in each group.



Google researchers built a set of everyday questions & split them by whether the answer changes. The writers were told to make each 1 **"appear natural, plausible for a real person to type into a search engine"**, so this is not a set of trick questions. GPT-4 answered with no web search. **93% right when the answer stays put. 26% right when it moves.** Same AI, same day.

The same team ran it a year earlier & got **69% & 14.4%**, so the gap is not a 1-off.

Source: *FreshLLMs: Refreshing Large Language Models with Search Engine Augmentation*, Vu et al., Findings of the Association for Computational Linguistics, ACL 2024.

Knowing which AI is right **cuts our own bill, starting day 1**. If somebody pays us for that knowledge later, **treat it as a bonus & count it as zero**.

PAYS OFF FROM DAY 1

1. We stop using the AI that got worse



Als quietly get worse at things. When one does, we send that kind of question somewhere else. **Everyone else keeps using the bad one, because nobody told them.** & the AI that is right is often the cheap one, so being right costs us less, not more.

PAYS OFF FROM DAY 1

2. We tell people when their answer goes out of date



You save an answer today. In 6 months the rule behind it changes. **We are already watching, so we just tell you.** People pay every month for that.

WE NEED 1 TO TRY IT

3. Companies that put an AI inside their own product



Their AI got worse last Tuesday & nobody told them. **There is no alarm bell for that.** It makes sense, but nobody has paid us yet.

BEST IDEA, WORST ODDS

4. The AI companies themselves



They would want to know where their own AI is weak. Trouble is they can build it themselves. **We would take the meeting, but do not put it in the model.**

SLOW, FEW SALES, BIG CONTRACTS

5. Reports for big companies choosing an AI



"Which AI should our lawyers use?" People already sell reports like that. **Ours falls out of what we are collecting anyway.**

BIGGEST, & THE LONGEST WAIT

6. Insurance, one day



An insurance table for AI mistakes, the kind an insurer calls an actuarial table. Which AI got someone in trouble, how often, & on what. You can only build that by writing it down for years, **which is why nobody can build it in a hurry.**

Read it this way: the first 2 pay for themselves. The other 4 are what those 2 make possible.

Specialist AIs too, not just the big ones.



Some sell their answers to any company today. Others need a partner deal, & a few only work for their own paying customers. **Each one has to earn its place:** a specialist that doesn't beat the big AIs on its own subject costs money & adds nothing.

We've mapped 200+ AIs built for 1 field: medicine, law, finance, science, math & dozens of languages. **Every one that makes sense will be in the lineup by the time the sites launch,** tested against the big AIs on its own subject. Actuarial work has no specialist AI yet, **so that is the one we intend to own.**

Funds are buying small firms & handing the work to AI. **We check that work before a client sees it.**

Accountants, bookkeepers, tax preparers, IT help desks & property managers. Owners are retiring, funds are buying, & their plan already has a step where the AI's work gets checked. **We are built to be that step.**

STEP 1 · THEIR AI

Collects the paperwork

Emails, invoices, uploads & forms, sorted & filed.

STEP 2 · THEIR AI

Writes the first draft

The tax return, the ledger, the monthly report.

STEP 3 · US

Several AIs check the draft

AIs from different companies read it. Where they agree, it moves on.

Where they disagree, a person sees exactly which line & why.

STEP 4 · THEIR PERSON

A person signs off

Nothing reaches a client without it. Their time goes to the lines we flagged, not to rereading everything.

Every fix a person makes becomes a rule.

The correction is saved in the firm's notebook, so the same mistake gets caught next time, in every firm the fund owns. **Their checking gets better every week. The data stay with the company & we keep an anonymized record if we can sell the anonymized human corrected data.** It is the same human correction data the rest of this deck is built to collect.

Nothing new to build: the several-AI check, the expert crowd & the notebook already exist. This is 1 more buyer for the same machine.

Every number, every test, every rival.

CAP TABLE: Uncomplicated. Bob currently owns 100%.

TAM = everyone frustrated with wrong AI answers



TAM = everyone who is frustrated with AI answers being wrong. People who are curious, have hobbies & professions & don't mind taking a little time to 👍/👎 or even type in a short correction when they're frustrated with an answer, & if they can earn some money on the side, even if it's not much, it's more than they're getting today sharing their knowledge & sometimes arguing on social networks, & not helping make AI as accurate as possible at the same time, & helping create another source of trustable information.

The part with published forecasts is ~\$2-7 Billion a year.

Below are 6 more pages. They are not short summaries where we picked the good bits. In 2 of them we say, out loud, that we got a number wrong earlier & then fixed it, which most companies would quietly delete instead.

The machine, drawn as 1 gearbox live, tap any gear

The safety & online-harm matrix scored the same way as accuracy

Who else is doing this the full map, plus Microsoft & the 11 small rivals

The forecast, line by line every row, every assumption

GTM notes & the stack for the GTM engineer

Where every number came from including the ones we refused to use

Let's build the growth machine together →