Agents with Taste

Nov 20, 2026 · @David

The challenge

Anyone can generate a convincing website. Build the agent that isn't fooled.

Picture yourself buying software for a company. A team needs a new tool, there are a dozen candidates, and you have to decide quickly which product to buy and which company you can do business with. You can't audit every candidate, so you do what experienced buyers do: you look at the website, form an impression, and then check what matters.

That takes two skills. **Taste:** the copyright says 2017, the help center has three articles and two of them 404, the customer logos look borrowed. Or the site looks like 2009, and some of the biggest companies in Germany run on the product anyway. **Fact checking:** do the features exist, do the customers confirm, is anyone still building the product?

Give the same site to an AI agent and it reads the marketing copy back to you, with full confidence.

Build an agent with taste that also checks its facts: one that judges a software product and the company behind it like a seasoned buyer, quickly and without being fooled. We're Stackgini. We help companies choose software, so we read a lot of these websites, and we would love an agent that reads them better than we do.

Looks are only half of it

A screenshot tells you a lot about how a site looks. It tells you little about what is behind it, and the two often disagree:

- A polished site, generated last month: feature pages for features that don't exist, a help center nobody wrote, logos of customers who never heard of the vendor.
- A dated site, run by a vendor with decades of customers and a product that ships every month.
- A solid-looking product with a data breach in its past that you won't find on its blog.

Anything a vendor writes about itself is now free to produce. The signals that still mean something take time or need someone else: customers who name the vendor on their own sites, a domain older than last month, a team with a history, a press report the vendor didn't write.

An agent that believes everything fails. So does an agent that doubts everything.

What you build

An agent that takes the website of a software product and returns a verdict a buyer can act on: is the product real and alive, and would you trust the company behind it with your data and money? With the reasons, and a link behind each one.

How it gets there is up to you. Some places to start:

- **Looks.** Design age, the copyright year, dead links, a help center that goes nowhere, product screenshots that don't match the product.
- **Substance.** Claims that check out in the docs or a partner marketplace, customers who confirm, a company register entry, a team with a track record.
- **Risk.** Breaches, lawsuits, an acquisition that froze the product, what the founders did before.

Vision models, search APIs, crawlers, public registers, the Wayback Machine: use whatever works.

What you get

On Nov 23 you get labelled examples: real software products, each with our call and the reasons behind it. A small script lets you check your agent against our calls.

On the day, the other partners bring search, compute and n8n. Bring any model and any stack.

What counts as done

1. Hand it a product it has never seen, and it returns a verdict with reasons.
2. Every reason points to something a human can open and check: a page, a record, an archived snapshot.
3. On the live products it gets both sides right. It flags the hollow ones without trashing the solid ones.

How we pick the winner

Live cases decide, on the evening of Nov 27:

1. **19:00** Every team gets the same few products, new to everyone.
2. **19:45** Submissions close. Send us a file with your agent's verdicts on those products, next to the standard AIB submission.

We judge the files on two things, and the calls count most:

- **The calls.** We compare each verdict with ours, but our calls can be wrong too. Where your agent disagrees, we check its reasons. If they hold up, the call counts. If your agent found something we missed, you get a bonus.
- **The evidence.** Every reason points to a source that says what the agent claims.

The top three teams then run one more product live for us, picked on the spot, so we see that their agent made the file. We announce the winner the same evening.

Prize, judge and support

- **Prize:** each member of the winning team gets GTA 6 on the console of their choice and an IMAX coupon for Dune: Part Three. Then the whole team comes to dinner with us. Stackgini delivers the prize, and AIB adds a Builder Pass for each winner.
- **Judge:** David Mente, CTO and co-founder of Stackgini.
- **Support:** David and the Stackgini team, onsite from the kickoff until the 19:45 cutoff.

The fine print

- **Who can enter:** any team at the hackathon, solo or up to five, with at least one person onsite.
- **Your code stays yours.** With the winners' OK, we'd love to tell people how they did it.
- **Our calls are judgement calls** made for this challenge, not public ratings of these companies. Keep the data inside the event.
- **Let the agent do the work.** No researching the live products by hand.
- **Be a polite visitor.** Don't hammer vendor sites, and don't contact the vendors.