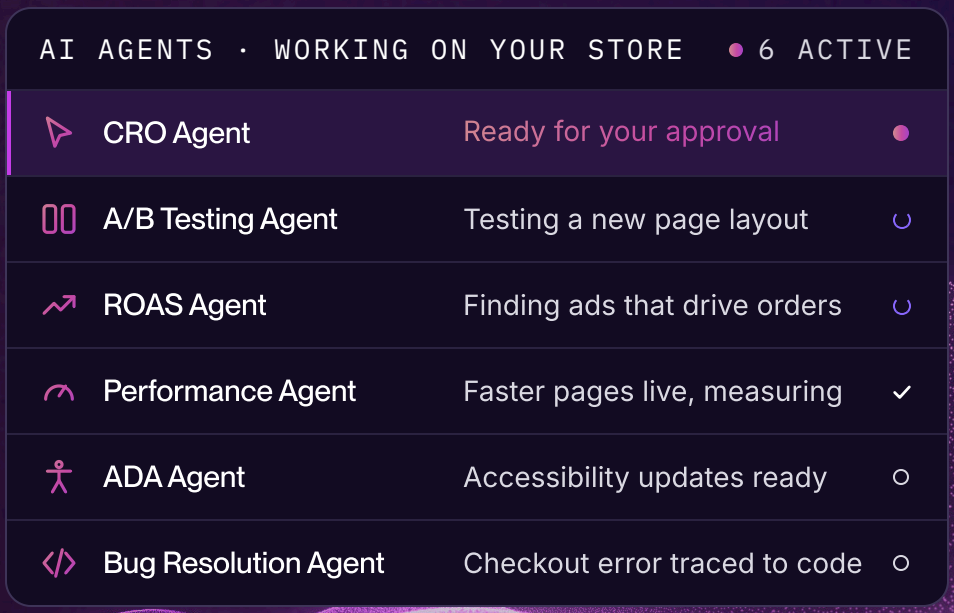
Noibu

WHITE PAPER

The Self-Improving Website

A Retailer's Guide to AI Agents



BUILT ON NOIBU DATA FROM 2,400+ ONLINE STORES



From left: Filip Slatinac, Robert Boukine and Kailin Noivo, Noibu's co-founders

FOREWORD

A note from our founders

For nine years, Noibu ended at a dashboard. We showed retailers what was broken, slow or confusing on their websites. What happened next was somebody else's job.

Last year, that changed. Our customers started pulling our data into AI tools and asking for changes, not reports. Then they started asking whether those changes had worked.

That shift is what this paper is about. A retail website has always improved at the speed of the people working on it. For the first time, it doesn't have to.

A website can now find what to change, draft the change, wait for a person to approve it, ship it, and measure what happened. Then it does it again, a little sharper each time. We call that a self-improving website.

We wrote this for retail leaders deciding how their digital storefronts should run in 2027. It explains what a self-improving website is, what it takes to run one safely, and how to tell whether yours is improving. We hope it's useful whichever tools you choose.

Kailin Noivo, Robert Boukine and Filip Slatinac
Co-founders, Noibu

Short on time? Page 4 has the short version.
Want to see the loop on your own store? Start a Noibu free trial →

INSIDE THIS PAPER

Contents

A practical guide for retail leaders, built on Noibu data from more than 2,400 online stores.

01	The short version	04
02	The constraint	05
03	What it costs	06
04	What a self-improving website is	07
05	The loop, step by step	09
06	What it takes to run the loop	10
07	The metric that matters	11
08	What it looks like in practice	12
09	Readiness check	14
10	Where to start	15
11	Meet Noibu's six agents	16

01

THE SHORT VERSION

Your website improves at the speed of your team. It doesn't have to.

THE LEAK

48%

of shoppers who start checkout don't finish

THE WAIT

71%

of revenue-costing issues are still live 90+ days after being flagged

THE FIX

1

human decision per change. Agents do the rest, and nothing ships until a person approves it

01 **The constraint is capacity, not insight.**

Most retailers already know where their sites lose shoppers. The changes wait in sprint queues, dev backlogs and code freezes.

02 **AI agents can now do the work in between.**

They find what to change, draft the change, ship it once a person approves, and measure the result.

03 **The gap opens one measured improvement at a time.**

Start with one a week on one goal, then add more as your team gets comfortable. Track measured improvements shipped per week.

THE CONSTRAINT

Capacity, not insight

02

Every retail website carries a backlog of changes it should make. Some are problems: a checkout step mobile shoppers abandon, a slow product page, an error firing on add to cart. Others are opportunities: a better filter, stronger product copy, a test worth running.

Knowing about these changes is rarely the hard part. Analytics, session replay, performance monitoring and testing tools make the signals easy to see. Turning a signal into a shipped change is where stores stall.



of senior marketing leaders say the gap between insight and action at their organization is significant or substantial.[1]

Why changes stall

Sprint calendars	Site changes compete with roadmap work for engineering time, and most lose.
Release batching	Changes wait to ship together. One multi-brand retailer planned a first release after its replatform that bundled nine months of changes.
Code freezes	Peak season locks the site for weeks, exactly when traffic is highest.
Test capacity	Many conversion programs, in-house or outsourced, test a handful of ideas a month on a few priority pages.
Fragmented tools	Each tool sees one slice: traffic, shopper behavior or technical health. Someone has to piece them together before anyone can decide.

WHAT IT COSTS

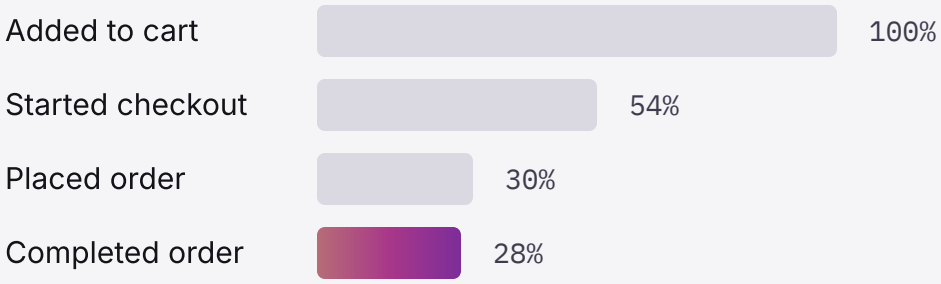
The cost of waiting compounds

03

Every day a known problem stays live, it keeps costing conversions. Every day a good idea waits, its upside waits with it.

WHAT THE DATA SHOWS

Most carts never become orders



72%

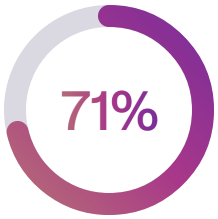
of carts didn't become completed orders

Share of carts reaching each step · 2,400+ stores, 1.8B sessions, Jul–Sep 2026

48%



of shoppers who start checkout don't complete their order.



of revenue-costing issues are still live 90+ days after being flagged. Nearly 1 in 4 have been waiting more than a year.

> The answer isn't more insight. Retailers already have plenty. It's the capacity to act on what they know.

THE IDEA

What a self-improving website is

04

A self-improving website runs a continuous cycle: it finds something worth changing, drafts the change, ships it once a person approves, and measures the result.

That result becomes the starting point for the next change. Run every week, the store improves on a schedule rather than whenever someone has time.

The idea comes from AI research, where it's called recursive self-improvement. In 1965, the mathematician I. J. Good described a machine able to design a better version of itself, which would design a better one still.[2]

His condition is the part most people skip: the machine had to stay under human control.

On a retail website, that control has a concrete form: the approval step. Nothing reaches the live store until someone who knows the business says yes, no, or "not like that." Everything else in the cycle can be automated, but that decision stays human.

1965

I. J. Good imagined a machine that could design a better version of itself, on one condition: it had to stay under human control.

WHAT IT ISN'T

It isn't autopilot.

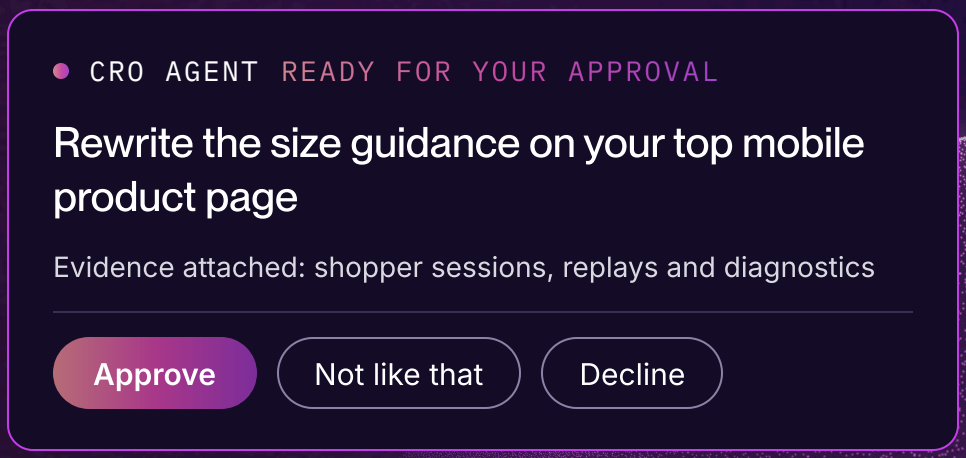
No change ships on its own, and every change arrives with its evidence: the sessions, replays and diagnostics behind it.

It isn't another dashboard.

A dashboard ends at an insight someone still has to act on. A self-improving website ends at a shipped, measured change.

THE RULE

One rule runs through all of it: nothing reaches a live store until a human approves it.



THE LOOP

05

Six steps, one human decision

Every cycle runs the same six steps. Only one has to stay human: the approval.

THE LOOP · RUNS EVERY WEEK			WHO DOES IT
01	Find	Spot what's worth changing in the store's own shopper data.	AGENT
02	Build	Draft the change: a code fix, a page rewrite, an A/B test or a shift in ad spend, ready for review.	AGENT
03	Approve	A person reviews the evidence, then says yes, no or "not like that."	YOU
04	Ship	The change goes live and is logged as a release.	AGENT
05	Measure	Check the result against a baseline or control.	AGENT
06	Compound	The result becomes the starting point for the next cycle.	AGENT

↻ THEN BACK TO STEP 01, A LITTLE SHARPER EACH TIME

In Noibu, six agents each run this loop for a different job: bug resolution, performance, A/B testing, accessibility, CRO and ROAS. Meet the agents on page 16 →

WHAT IT TAKES

What it takes to run the loop

06

Any AI model can point to the drop-off in a conversion report. A self-improving website depends on everything after the suggestion.

01

A safe way to ship the change

Agents can draft a fix or a test, but something has to get it in front of shoppers safely. That means a release path that can roll back fast if anything breaks. For tests, it also means keeping each visitor in the same version, without the original page flashing first. Done badly, it slows the whole site: in one setup, hiding the page until a test loaded more than doubled load time, from 2.7 to 6.0 seconds.[3]

02

A log of every change

Without a change log, a store has activity but no loop. If conversion dips on the 15th, the team needs to know whether the theme update on the 14th or the app installed on the 13th caused it. Every deploy, theme publish, app install and merged pull request lands as a dated release with a before-and-after comparison.

03

Shopper data the store controls

Top-line analytics show that conversion dropped, but not why. The why lives in sessions, clicks, journeys, page speed, errors and checkout events, held in one place, unsampled, with bots filtered out. A “lift” that only shows up in unknown-source traffic is not a lift.

Platform reality

Shopify, Salesforce Commerce Cloud, Adobe Commerce, BigCommerce and headless builds each handle releases, test delivery and checkout visibility differently. Ask how any approach works on your platform specifically.

THE METRIC

Measured improvements per week

07

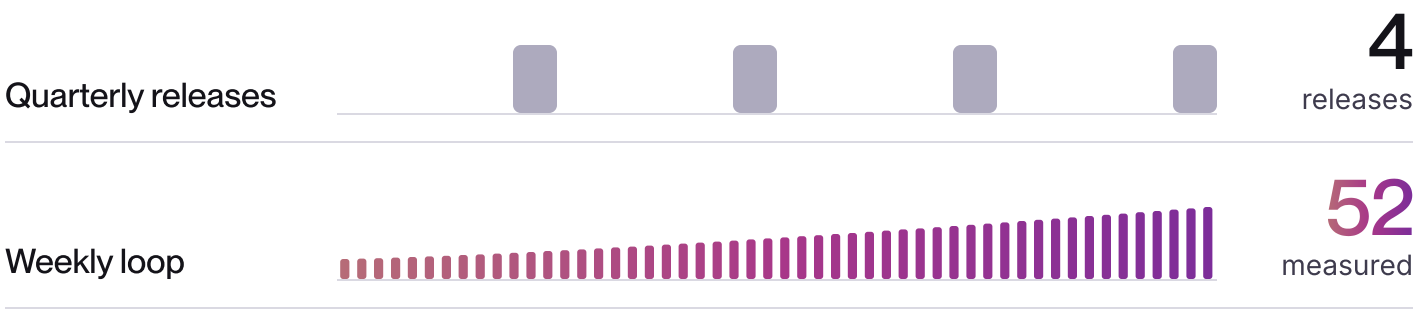
Track how many measured improvements your store ships each week.

DEFINITION

A measured improvement is a change that went live, was logged as a release, and had its effect checked against a baseline or control a week later.

- ✓ **Counts:** a change that shipped, was logged, and was measured against a baseline.
- ○ **Doesn't count yet:** ideas, tickets, tests started and open pull requests. Count them as inputs.

ONE YEAR, TWO CADENCES



Illustrative. Why weekly: a week gives each change enough traffic to check against a baseline, and it's short enough for improvements to stack up. A quarterly release ships many changes at once, so none can be measured cleanly.

> This number shows whether your store improves on a schedule or only when someone has time. Once the loop is running, it stops tracking headcount.

IN PRACTICE

What it looks like in practice

08

Retailers are already putting AI to work on their sites, and it happened fast.

1% → 100%

of Noibu's conversations with customers mentioned AI, January to September 2026[4]

~320

data lookups a month by the average store using Noibu inside AI tools like Claude or ChatGPT[4]

1 in 4

of those lookups ran on their own: an agent started them, not a person typing a question[4]

RIBBLE CUSTOMER STORY

One of the UK's largest online bike retailers

Ribble's site now flags its own problems, ranked by cost[5]

£35k–£100k

estimated annual sales protected by fixes shipped in 11 weeks

33

fixes shipped for issues Noibu surfaced

9, not 7,600

issues returned when the team asks what's broken, ranked by what's costing orders

"Before the MCP, Noibu's value was capped by how often a human opened it. Now it isn't capped at all."

Matthew Lawson, Chief Digital Officer, Ribble Cycles

Read the story →

IN PRACTICE

More from the field

08

> “Noibu’s data paired with best-of-breed AI has basically unlocked what it would feel like to hire a team of ten people.”

CARS N CARDS .com Donny Araujo, Founder, CarsnCards[9]

RENEGADE GAME STUDIOS

6

accessibility fixes merged in the first week

Across six BigCommerce stores, Noibu’s ADA agent drafts each change as a pull request, and a person reviews it before it merges.[6]

Read the story →

Totally Bamboo

+7%

conversion over the next 30 days

One person runs direct-to-consumer across 600+ products. A collection-page filter test lifted conversion 24%.[7]

Read the story →

CREATIVE BAG

6+

production fixes shipped by the COO, with no developer on staff

The COO uses Noibu’s AI plugin to find what’s costing the store revenue and ship production changes himself.[8]

Read the story →

READINESS CHECK

Before you start

09

Settle who approves what

Senior teams will ask these first. Agree the answers before the first cycle.

01
Who can approve a change, and does that differ for copy, design and code?

02
Which changes need a second reviewer?

03
How is a change rolled back, and how fast?

04
Where is the record of what shipped, who approved it and what it did?

SELF-ASSESSMENT

Answer these for your own store. If you answer “I don’t know” to more than two, start by logging every site change with a date (see page 10).

1. How many changes shipped to your website last week, and how many did you measure?
2. Is every site change, including theme publishes and app installs, logged with a date?
3. How long does a known, revenue-impacting issue typically stay live on your site?
4. Can you run an A/B test without slowing the page down?
5. Does one source hold your behavioral, technical and commercial shopper data, with bots filtered out?
6. Who approves a site change, and how long does approval usually take?
7. Did you check last week’s changes before starting this week’s?

GET STARTED

Where to start: a 90-day path

10

Start with one storefront and one goal, and aim for one measured improvement a week by day 60.

DAYS 1–30

Instrument and log

- Log every site change as a dated release, including deploys, theme publishes and app installs.
- Bring behavioral, technical and commercial shopper data into one place, with bots filtered out.
- Set baselines for conversion, page speed and error rates by page type.

DAYS 31–60

Run the loop on one goal

- Pick one outcome, such as mobile checkout completion.
- Agree who approves changes and how quickly.
- Ship and measure one improvement a week, checking each before starting the next.

DAYS 61–90

Raise the cadence

- Add a second goal, such as product page speed or conversion.
- Review measured improvements weekly as a team.
- Report the weekly count to leadership alongside conversion rate.

The gap won't open in a single leap. It opens one measured improvement at a time.

Want to run the loop on your own store? **Start a Noibu free trial →**

HOW NOIBU RUNS THE LOOP

Meet Noibu's six agents

11

Every agent runs the same loop: it finds the change, builds it, waits for your approval, ships it and measures the result. Teams work with the agents in the Noibu console or from the AI tools they already use.

Bug Resolution Agent

Traces checkout-blocking errors to the code and opens a pull request.

See how it works →

Performance Agent

Finds the slowest page groups and stages a change for review.

See how it works →

A/B Testing Agent

Builds and runs experiments, measures the lift and reports the winner to ship.

See how it works →

ADA Agent

Audits top pages against WCAG, stages changes and re-audits after launch.

See how it works →

CRO Agent

Rewrites underperforming product pages to convert better, ready to review and publish.

See how it works →

ROAS Agent

Scores paid traffic on orders, not clicks, and shows where to shift spend.

See how it works →

See what they would change on your store.

Your agents run on your own store until two changes are live.

Start a free trial

SOURCES

Sources and methodology

NOIBU NETWORK DATA

All figures are aggregated and anonymized, and never attributable to a single store. Funnel figures: 2,400+ stores and 1.8 billion sessions, July 1 to September 30, 2026. Issue age uses an October 19, 2026 snapshot of 105,934 issues with a revenue-loss estimate that fired in the prior seven days across 1,135 stores, measured from when Noibu first detected each issue.

1 Gain Theory, "Unfinished Business: Closing the Insight-to-Action Gap", 2026. Insight-to-Action Survey of senior marketing leaders, Q1 2026.

2 I. J. Good, "Speculations Concerning the First Ultraintelligent Machine," *Advances in Computers*, vol. 6, 1965.

3 DebugBear, "Anti-Flicker Snippets From A/B Testing Tools And Page Speed", updated December 7, 2025.

4 Noibu, "Recursive Self-Improving Websites Are Here", September 15, 2026.

5 Noibu, "How Ribble Cycles uses Noibu to make its site tell them when it's broken", October 22, 2026.

6 Noibu, "How Renegade Game Studios shipped six accessibility fixes in a week with Noibu's AI agents", October 20, 2026.

7 Noibu, "How Totally Bamboo's One-Person Ecommerce Team Runs Like a Dev Department", July 22, 2026.

8 Noibu, "How Creative Bag's COO ships his own production fixes with the Noibu AI Plugin", June 25, 2026.

9 Noibu, "Carsncards.com: A Solo Operator Running a $1M+ Business with Noibu's AI Agent Stack", May 13, 2026.

ABOUT NOIBU

Noibu is AI for ecommerce operations, the always-on optimization system that keeps online stores running at their best. A team of supervised AI agents finds what's broken, slow, or leaking revenue, does the work, and hands results back for human approval. Trusted by retailers like Free People, Guess, Liquid I.V., and Ariat.

© 2026 Noibu Technologies Inc. All rights reserved. **noibu.com**

Noibu

AI FOR ECOMMERCE OPERATIONS

See what six AI agents can do for your store.

Noibu is AI for ecommerce operations, the always-on optimization system that keeps online stores running at their best. A team of supervised AI agents finds what's broken, slow, or leaking revenue, does the work, and hands results back for human approval.

Start a free trial

See how Noibu's AI agents work →

Your agents run on your own store until two approved changes are live.

TRUSTED BY FREE PEOPLE GUESS LIQUID I.V. ARIAT

NOIBU.COM