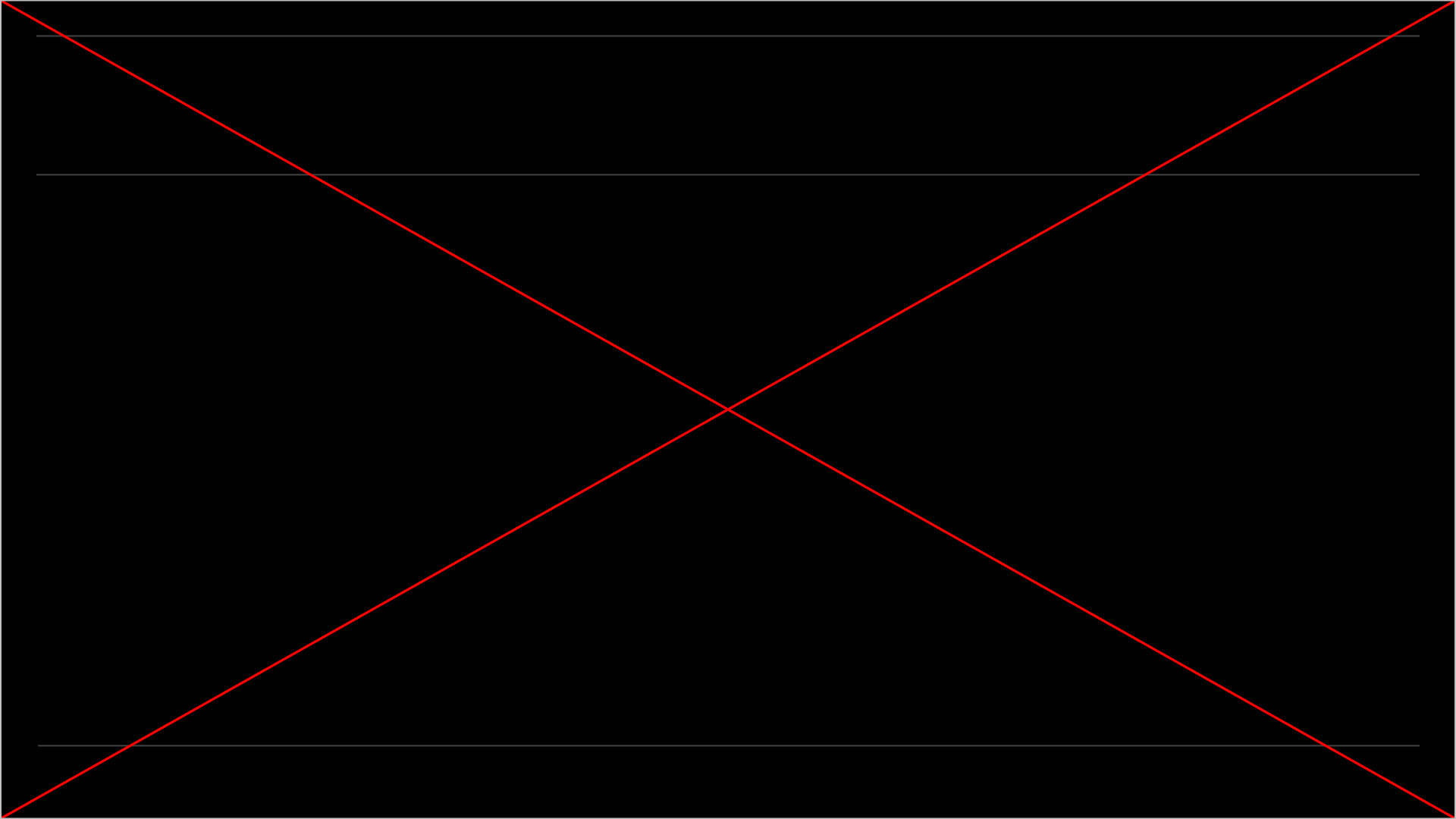


Deceptive Pattern Detector

Jane Lai, Chris Rezny, Bianka Paul, Eduardo Gonzalez, & Saleh Alhumaid

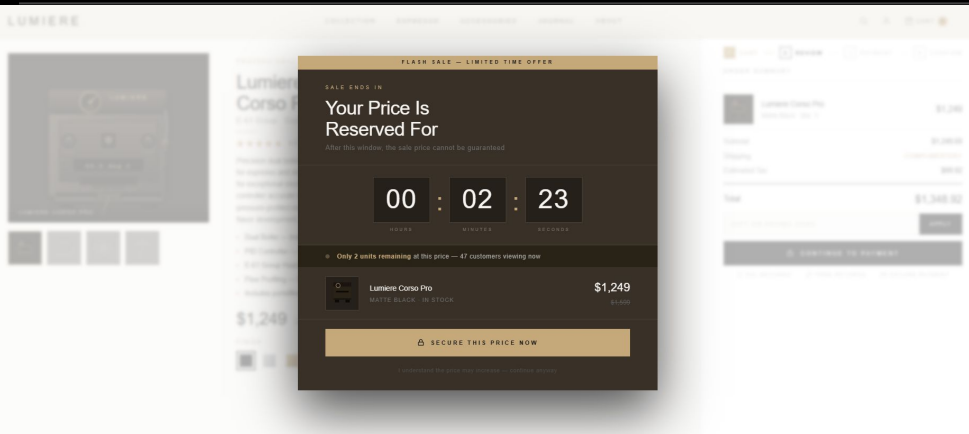




Was that relatable?



That Wasn't an Accident





That Wasn't an Accident

FLASH SALE — LIMITED TIME OFFER

SALE ENDS IN

Your Price Is Reserved For

After this window, the sale price cannot be guaranteed

00

:

02

:

23

HOURS MINUTES SECONDS

Only 2 units remaining at this price — 47 customers viewing now

 Lumiere Corso Pro

\$1,249

WHITE, BLACK, IN STOCK

SECURE THIS PRICE NOW

Unsubscribe the price once inventory is depleted.

COMPLETE YOUR ORDER

Recommended Additions

THE FOLLOWING HAVE BEEN SELECTED FOR YOU — UNCHECK TO REMOVE

- ☒ **Lumiere 3-Year Craft Protection** +\$89.00
Full parts and labor coverage including accidental damage, heating element failure, and boiler service.
- ☒ **White-Glove Delivery and Setup** +\$49.00
In-home delivery, placement, and initial calibration by a certified technician.
- ☒ **Lumiere Atelier Membership** Free Value \$24.99/mo
Priority support, member pricing, and quarterly specialty coffee sourcing reports.
Total period: approximately for 30 days from date of delivery. Upon conclusion of trial, the payment method provided at checkout will be automatically charged \$24.99 per month on a recurring basis. Membership may be cancelled via account settings or by emailing member_services@lumiere.com. Customers who opt-in to the next scheduled billing date. Cancellation does not entitle the member to a pro-rated refund of any active billing period.

ADDED TO ORDER +\$138.00

CONTINUE TO PAYMENT →

[REMOVE ALL ADD-ONS](#)



That Wasn't an Accident

LUMIERE

FLASH SALE — LIMITED TIME OFFER

SALE ENDS IN

Your Price Is Reserved For

After this window, the sale price cannot be guaranteed

00

:

02

:

23

HOURS MINUTES SECONDS

Only 2 units remaining at this price — 47 customers viewing now

Lumiere Corso Pro

\$1,249

WHITE BLACK IN STOCK

SECURE THIS PRICE NOW

Unsubscribes for your data marketing & Lumiere account

LUMIERE

COMPLETE YOUR ORDER

Recommended Additions

THE FOLLOWING HAVE BEEN SELECTED FOR YOU — UNCHECK TO REMOVE

- ☒ **Lumiere 3-Year Craft Protection** **+\$89.00**
Full parts and labor coverage including accidental damage, heating element failure, and boiler service.
- ☒ **White-Glove Delivery and Setup** **+\$49.00**
In-home delivery, placement, and initial calibration by a certified technician.
- ☒ **Lumiere Atelier Membership** **Free**
Priority support, member pricing, and quarterly specialty coffee sourcing reports. Value \$24.95/mo

Trial period: complimentary for 30 days from date of delivery. Upon conclusion of trial, the payment method provided at checkout will be automatically charged \$24.95 per month on a recurring basis. Membership may be cancelled at no-cost at any time by pausing or canceling your membership in your Lumiere account or by the next scheduled billing date. Cancellation does not entitle the member to a pro-rated refund of any portion of the trial period.

ADDED TO ORDER **+\$138.00**

CONTINUE TO PAYMENT →

REMOVE ALL ADD-ONS

LUMIERE

BEFORE YOU GO

Are you certain you want to leave these savings behind?

Your exclusive sale price and selections are held for the next 8 minutes. Once released, we cannot guarantee availability at this rate.

WHAT YOU WOULD BE FORFEITING	
Flash Sale Discount	-\$350.00
Craft Protection Plan	-\$89.00
White-Glove Setup	-\$49.00
Atelier Membership Trial	Complimentary
Total walkaway value	\$488.00

✓ KEEP MY SAVINGS AND COMPLETE PURCHASE

No thank you — I prefer to pay full \$1,550 and forfeit \$488 in savings



Designed Decisions, Not Bad UX

MAISONEDIT.COM

We value your privacy

We use cookies to personalise content, measure performance, and improve your experience. You can choose which categories you allow, or accept all to continue reading. See our [Cookie Policy](#) and [Privacy Notice](#)

Strictly Necessary

Essential for the site to function. Cannot be disabled.

ALWAYS
ON

Analytics & Performance

Help us understand how visitors interact with our content.



Personalisation

Tailor content and recommendations to your interests.



Advertising & Targeting

Allow our partners to show you relevant ads across the web.



ACCEPT ALL

[Manage Settings](#)

By clicking "Accept All" you consent to our use of all cookie categories. You may change your preferences at any time via [Cookie Settings](#).

AERIS.

BOOK

MY TRIPS

CHECK-IN

FLIGHT STATUS

AERIS MILES

HELP

MANAGE BOOKING

SIGN IN

JFK

NEW YORK · JOHN F. KENNEDY



5H 18M · NON-STOP

LAX

LOS ANGELES · INTERNATIONAL

DEPARTS

07:15 AM

Terminal 4 · Gate B22

ARRIVES

10:33 AM

Terminal 2 · Gate 42A

DATE

Apr 15, 2026

Tuesday

FLIGHT

MR 2847

Economy Select

14B

YOUR SEAT

Window · Zone 4 · Standard Legroom

✓ SEARCH

✓ OPTIONS

1 REVIEW

2 PAYMENT

BOOKING SUMMARY

JFK → LAX

Flight MR 2847 · Apr 15, 2026

1 Passenger · Economy Select

Base Fare — Economy Select

1 adult · Carry-on included

\$149.00

Seat 14B — Window, Zone 4

\$28.00

✓ Added

Checked Bag — 23 kg (1 bag)

\$39.00

✓ Added

Priority Boarding

\$19.00

✓ Added

Travel Protection — Basic

\$24.99

✓ Added

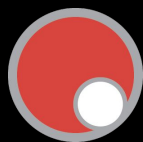
Taxes & Airport Fees

\$29.40

TOTAL \$449

\$289.39

CONFIRM & PAY — \$289.39



The Real Impact

Privacy

Forma

FEATURES

TEAMS

PRICING

DOCS

FORMA FOR TEAMS

Work better, together.

Forma brings your team's projects, docs, and workflows into one organized space. Built for clarity, designed for speed.

- ✓ Unlimited projects and workspaces
- ✓ Real-time collaboration and live presence
- ✓ Integrates with Slack, GitHub, Figma and 90+ apps
- ✓ Enterprise-grade security and access controls

Trusted by 300,000+ teams at Stripe, Shopify, and Notion

LOG IN

TRY FREE

Create your account

Free forever. No credit card required.

FIRST NAME

Alex

LAST NAME

Morgan

WORK EMAIL

alex@company.com

PASSWORD

Min. 8 characters

Use a mix of letters, numbers, and symbols

COMMUNICATION & DATA PREFERENCES

- ☒ Keep me updated on new features and product news
Occasional emails about releases, tips, and platform improvements
- ☒ Share my usage data to help improve Forma
Anonymized activity data used to enhance product experience
- ☒ Personalize my experience based on usage patterns
Tailored recommendations, onboarding flows, and feature highlights
- ☒ Allow Forma's integration partners to contact me
Relevant offers from tools in the Forma partner network

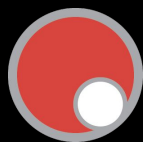
☐ I agree to Forma's [Terms of Service](#) and [Privacy Policy](#)

CREATE FREE ACCOUNT

OR

 Continue with Google

 Continue with Microsoft



The Real Impact

Autonomy

Strata

FEATURES

TEAMS

PRICING

ENTERPRISE

SIGN IN

GET STARTED

Simple, transparent pricing

Start free. Scale when you're ready. No hidden fees.

MONTHLY

ANNUAL

SAVE 20%

STARTER

Free

For individuals and personal projects

\$0

/mo

Free forever

- ✓ Up to 3 active projects
- ✓ 5 team members
- ✓ Basic task management
- ✓ 5 GB storage
- ✓ Community support

GET STARTED

No credit card required

MOST POPULAR

PROFESSIONAL

Pro

For growing teams that need more power

\$29

/mo per seat

Billed annually (\$36 if billed monthly)

- ✓ Everything in Starter
- ✓ Unlimited projects and members
- ✓ Advanced automations and workflows
- ✓ Analytics and reporting dashboard
- ✓ 100 GB storage
- ✓ Priority email and chat support
- ✓ 90+ integrations including Slack and GitHub
- ✓ Custom fields, views, and templates

START FREE TRIAL

14-day trial, cancel anytime

BUSINESS

Enterprise

For large organizations with complex needs

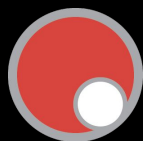
Custom

Volume pricing available

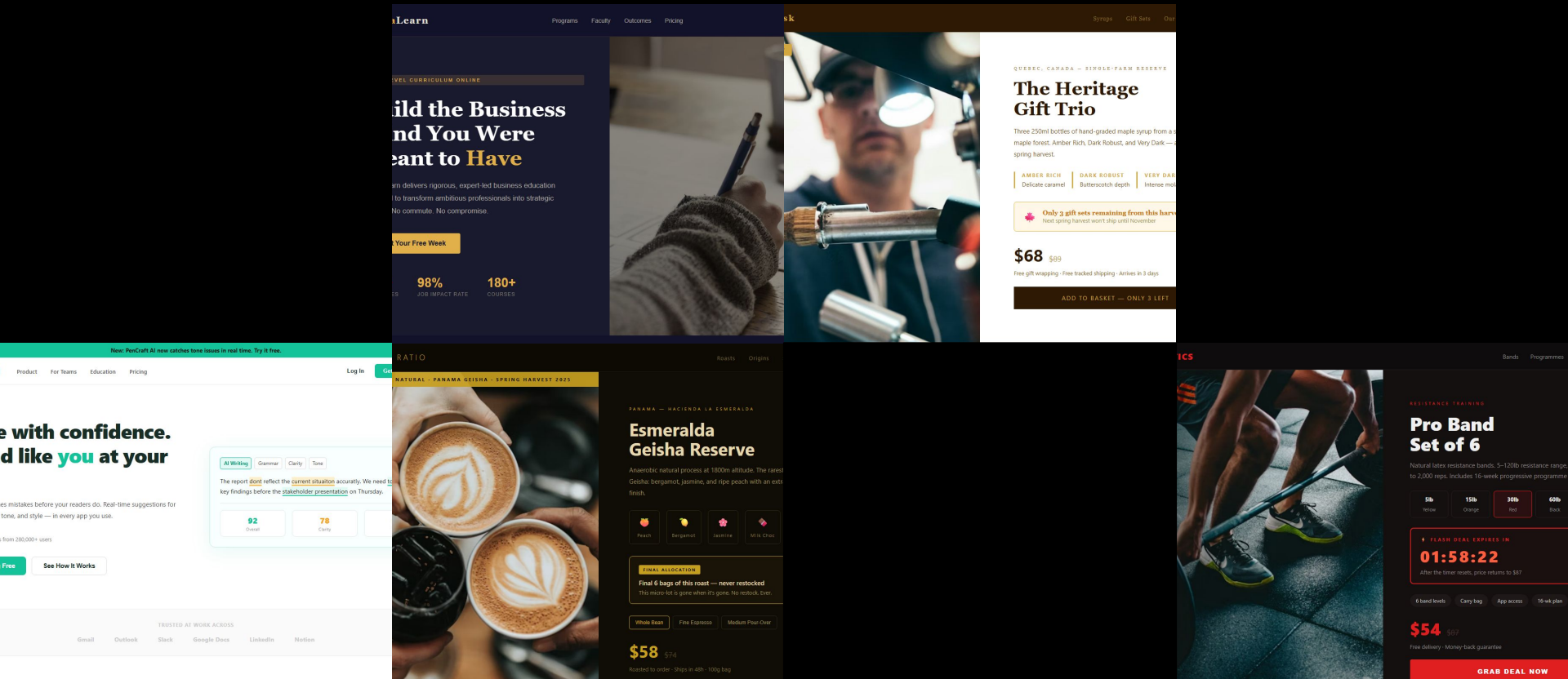
- ✓ Everything in Professional
- ✓ Single sign-on (SAML SSO)
- ✓ Advanced security and audit logs
- ✓ Dedicated customer success manager
- ✓ Custom SLA and uptime guarantees

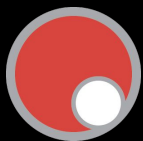
CONTACT SALES

Talk to our team first

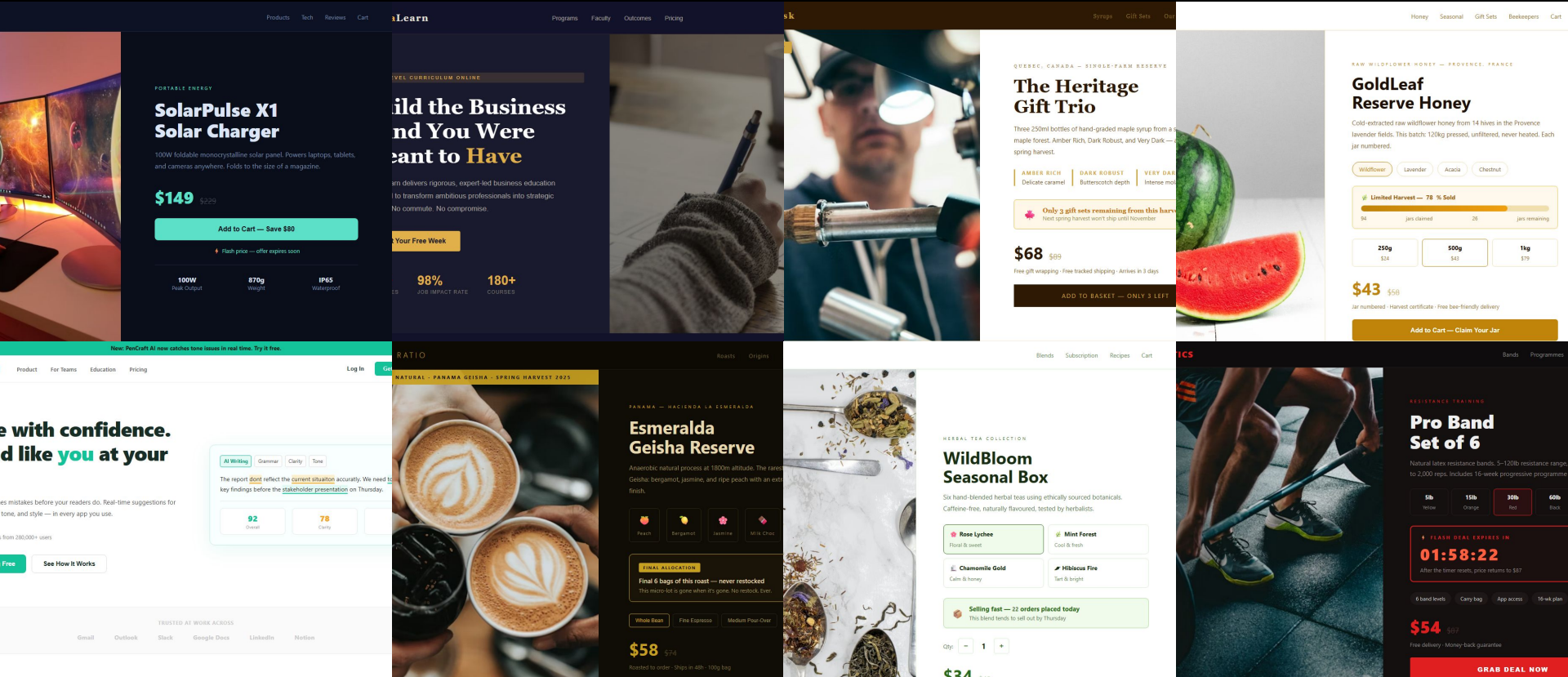


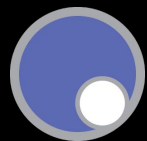
Why Nothing Stops This Today





Why Nothing Stops This Today





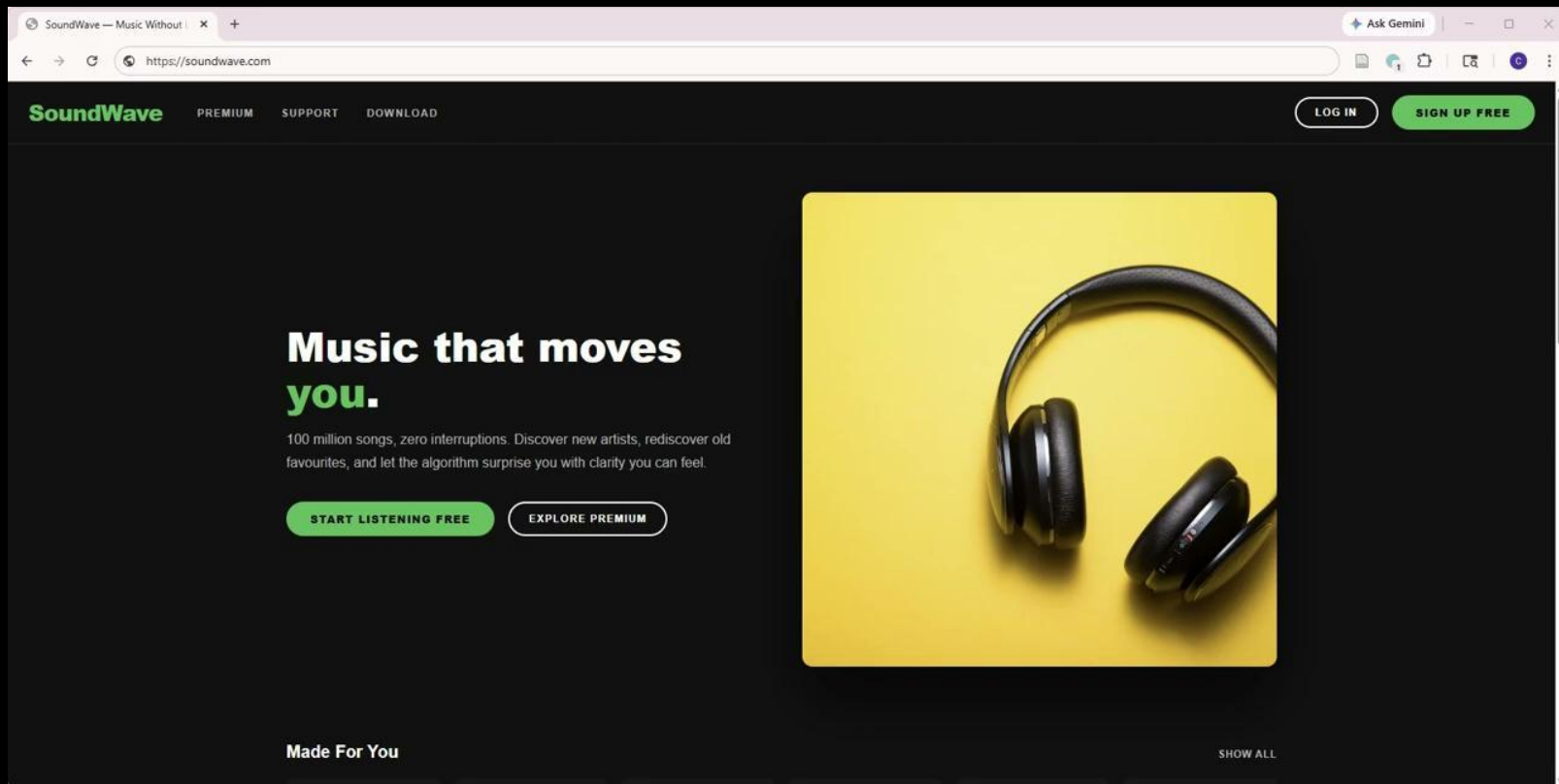
Deceptive Pattern Detector

- Browser Extension 

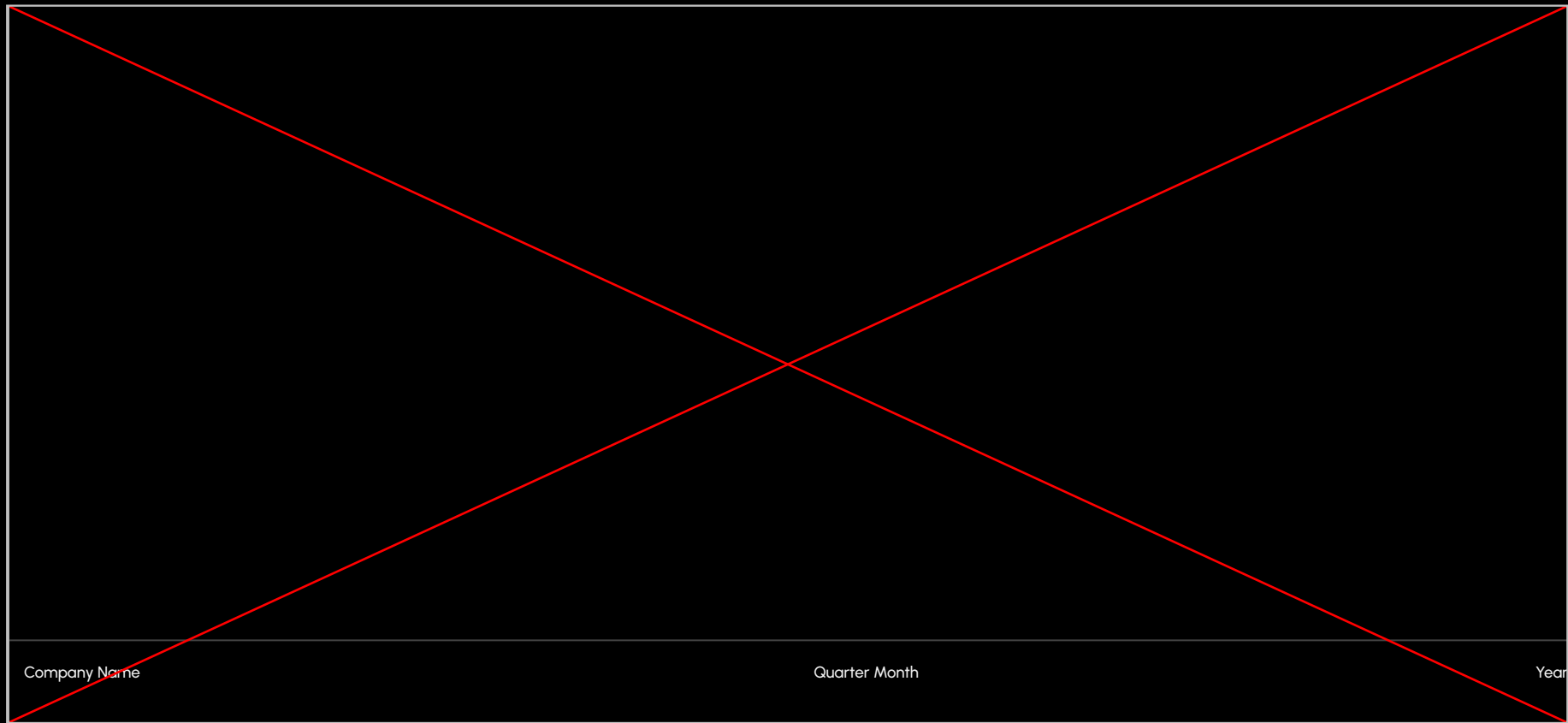
- pattern-specific ML models 

- gets better over time 

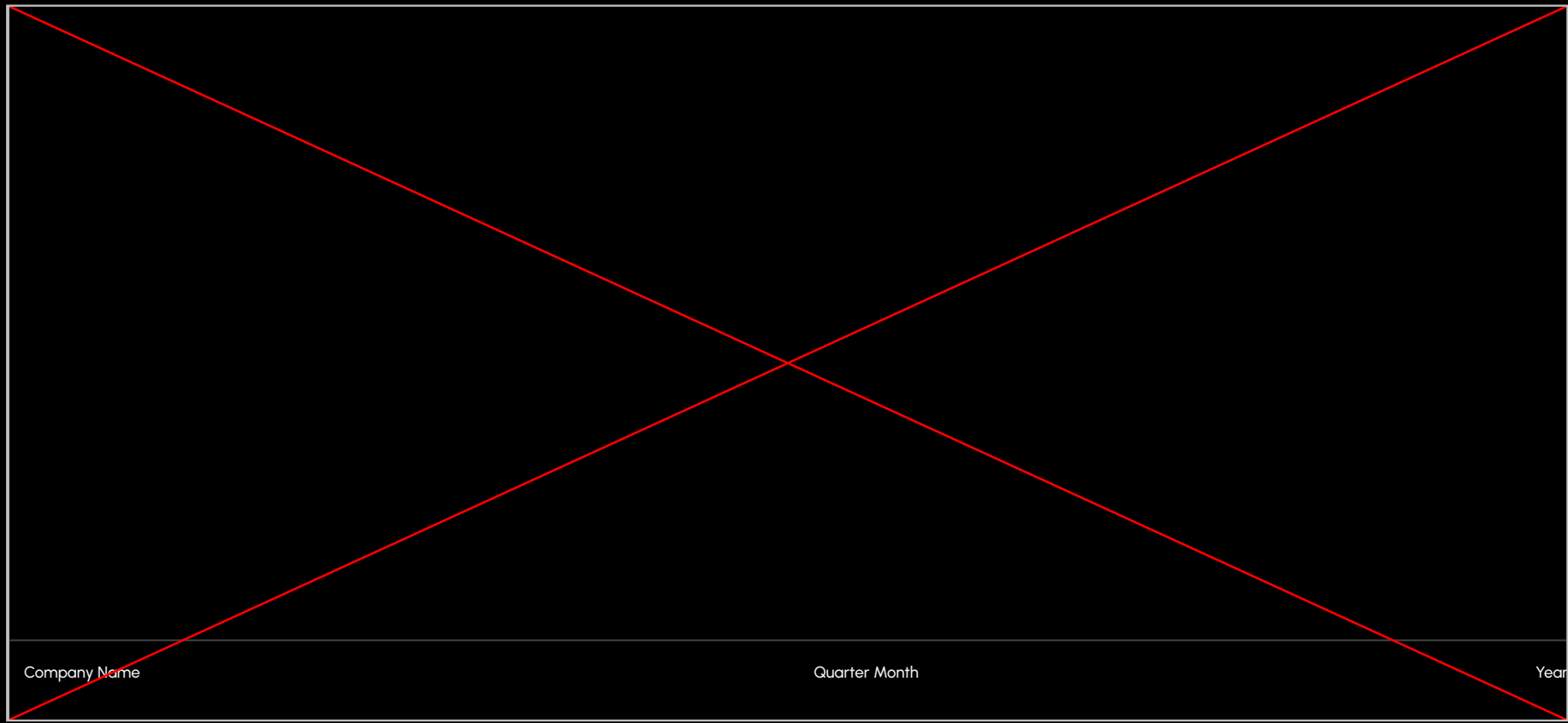
DEMO - EXTENSION

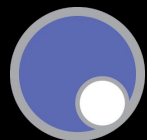


DEMO - GATEKEEPER DASHBOARD

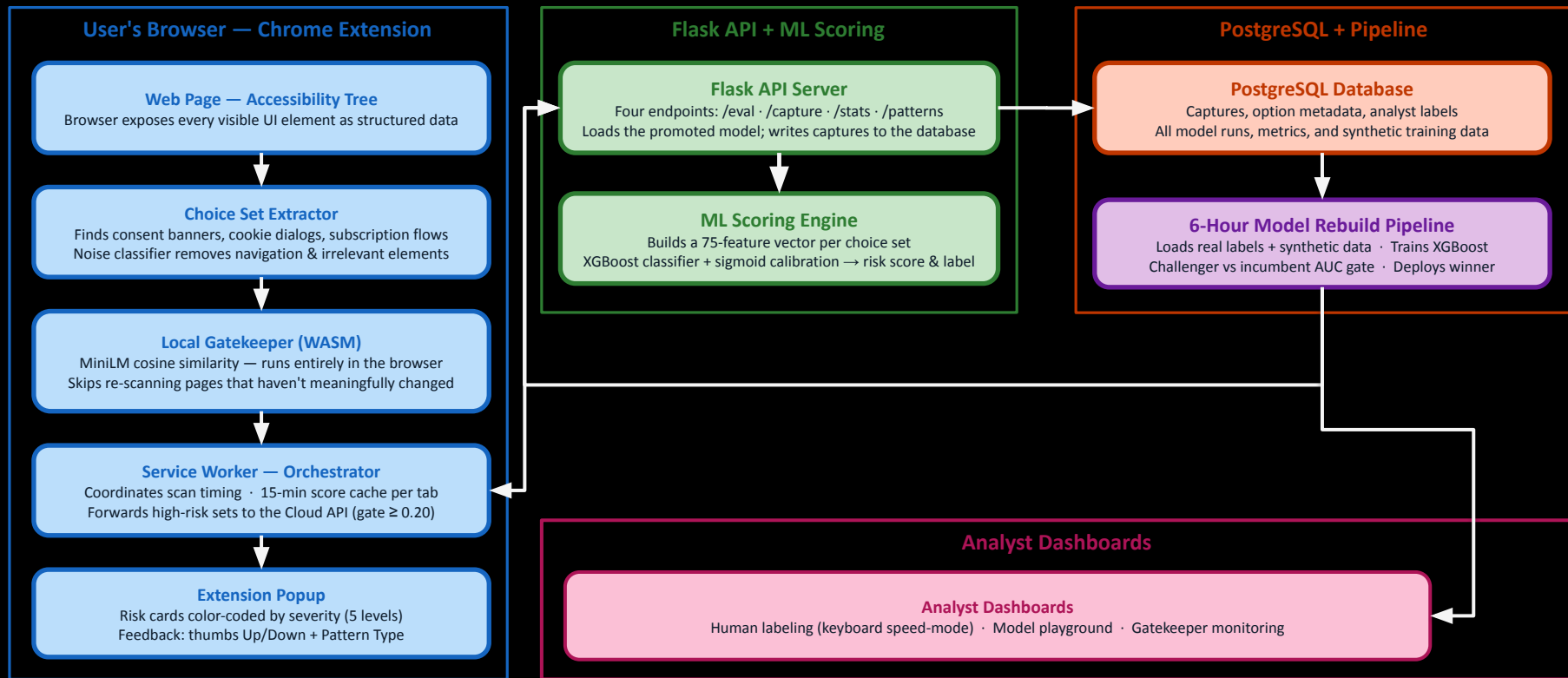


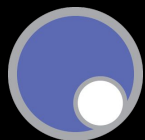
DEMO - MODEL DASHBOARD





How the System Works

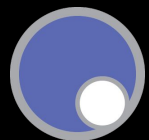




What is a "Choice Set"?

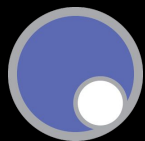
A decision UI (cookie banner, modal, subscription prompt)

- Extracted from live DOM via accessibility tree
 - Includes:
 - text
 - options
 - visual styling (CSS)
 - structural context
-



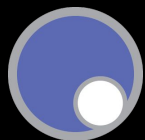
Why This is Hard

- No standard structure across websites
 - Patterns are rare and subtle
 - Heavy class imbalance
 - Visual tricks not always visible in text
-



Two-Stage Detection Pipeline

- Stage 1: Fast Client Side (Local) filter
- Stage 2: Deep pattern classification (API Endpoint)



Stage 1: Gatekeeper

- Runs entirely in the browser (ONNX/WASM)
 - Blended suspicion score
 - MiniLM semantic similarity (40%)
 - Keyword matching (20%)
 - CSS prominence gap (20%)
 - Structure signals (10%)
 - Urgency/scarcity patterns (10%)
 - Only choice sets scoring ≥ 0.20 are forwarded to the cloud API
-

Gatekeeper Model



Gatekeeper Dashboard

LOCAL EMBEDDING GATE — PASS / BLOCK ANALYTICS

threshold 0.20

flag 0.40

Loaded

TOTAL SCANS

2706

page loads recorded

TOTAL EXTRACTED

850456

choice sets found

NOISE FILTERED

843990

before gate

GATE CANDIDATES

6466

scored by embedding

GATE FILTERED

808

score < 0.20

GATE PASSED

5658

sent to server

SERVER HITS

942

pattern confirmed

AVG PASS RATE

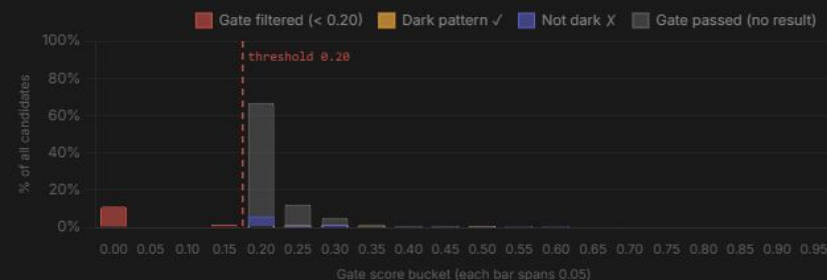
86%

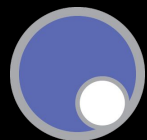
passed / candidates

PASS VS. BLOCK — BY DAY



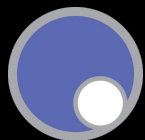
SCORE DISTRIBUTION — ALL CANDIDATES





Stage 2: Pattern Models

- One XGBoost binary classifier per pattern type (7 models)
 - Trained on structured feature vectors
 - Text, CSS, AX-tree
 - Class imbalance
 - Handled via `scale_pos_weight`
 - Rare positives up-weighted proportionally
 - Sigmoid calibration for true probabilities before thresholding
-



Stage 2: What the Models Looks At

~75 Engineered features looking at:

- Text
 - Urgency phrases , Scarcity phrases , Social proof phrases
 - Visual (CSS)
 - Per-button opacity, Hidden elements, Font size range, underlines, position:absolute, high z-index
 - Structure (AX-tree)
 - Dialog-like, Role counts, Intent counts
-

Stage 2: Model Performance & Limits



Why it Works



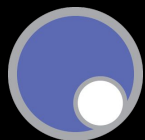
How We Validate



Where it Works Best

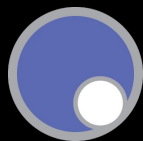


Where it Struggles



Stage 2: How Models Are Built

- Interactive training environment (per-pattern models)
 - Feature selection driven by real data
 - Configurable sampling & reproducibility
 - Compare, evaluate, and promote models
 - Auto-promotion when performance improves
-



Model Configuration



Model Playground
XGBOOST BINARY CLASSIFIER - TRAINING ENVIRONMENT

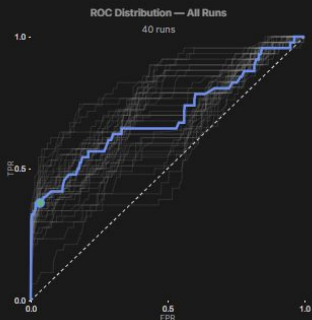
split v3

features ?

Done

CONFIGURATION

Pattern



Select a pattern row. The chosen saved run is only for inspection; training uses the inputs below unless you load params from a run.

PATTERN	LABELLED	SAVED RUNS	PROMOTED
☒ Asymmetric Choice	46	run 164 • 2026-03-31 • Best F1=0.476 • Best	☒
☐ Confirm Shaming	46	(none)	☐
☐ Misdirection	4	(none)	☐
☐ Privacy Zuckering	2	(none)	☐
☐ Scarcity	3	(none)	☐
☐ Social Proof	2	(none)	☐
☐ Trick Question	2	(none)	☐
☐ Urgency	7	(none)	☐

Features

☒ show distribution plots ☐ dark pattern ☐ not dark pattern

Select All Select None

Loader

Sampling + split controls

XGBoost

Model hyperparameters

RESULTS

METRICS

PR AUC 0.442 -0.036	ROC AUC 0.695 -0.046	BEST F1 0.476 -0.245	BEST THR 0.468 -0.245
LOGLOSS 0.232 +0.012	BRIER 0.053 -0.000	TEST POS 46 -1	TEST NEG 562 +1

OVERALL VS PROMOTED

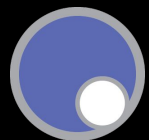
PR AUC	0.442	0.476
ROC AUC	0.695	0.741
Log loss	0.232	0.219
Brier	0.053	0.054
Best F1	0.476	0.476
Best F1 threshold	0.468	
Best F1 precision	0.882	0.938
Best F1 recall	0.326	0.319
Best Iteration	161	
Scale pos weight	7.969	
Test positives	46	47
Test negatives	562	561

DATASET

THRESHOLD 0.10

THRESHOLD 0.25

THRESHOLD 0.50



What the User Sees

- Non-intrusive overlay
- Risk score (5 levels)
- Plain-language explanation
- Pattern type identified (e.g., asymmetric choice)

asymmetric choice

One option is prominently styled while the other is de-emphasized or hidden.

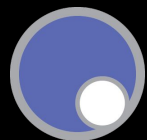


05

✗ Close

✓ CLAIM 3 MONTHS FREE

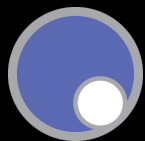
✗ No thanks, I'll just smile and nod when traveling abroad.



Feedback Loop

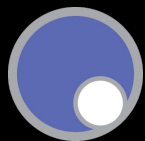


- Detection: pattern flagged in real time
- Understanding: user reviews explanation
- Improvement: optional feedback updates future models
- Models retrained every 6 hours and auto-promoted if better



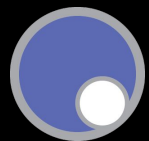
What Makes This Different

- Real-time detection (not post-analysis)
 - Interprets behavior (text + visual + structure)
 - Pattern-specific models (not one-size-fits-all)
 - Continuously improves through real-world use
-



Hard Problems We Solved

- Turning messy DOM into structured ML input
- Learning from rare events (class imbalance)
- Running ML in real time inside a browser



Idris Elba - The Real Voice Over

