



adaptTo()

EUROPE'S LEADING AEM DEVELOPER CONFERENCE
28th – 30th SEPTEMBER 2026

Measuring the Unmeasurable: Pre/Post
Experimentation for Brand Visibility in LLMs
Nitin Gupta, Adobe

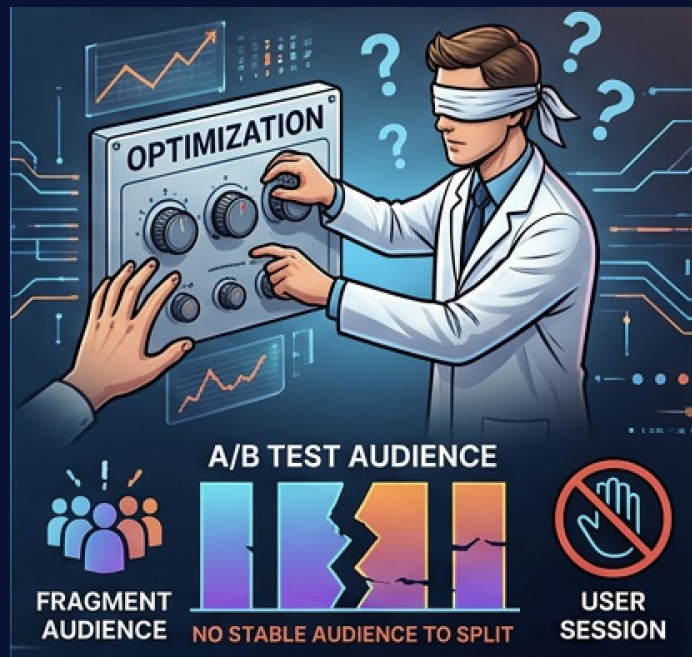
Agenda

- The measurement problem
- Under the hood
- Proving causation: the evidence ladder
- Product experience
- Where this is headed

01 — The Problem

Old measurement playbooks don't work here

- Buyers now ask ChatGPT, not just Google — citations are invisible without measurement
- No stable audience to split for A/B — LLM traffic is crawlers and one-shot chats
- No cookies. No logs. Optimizing blind is an act of faith.

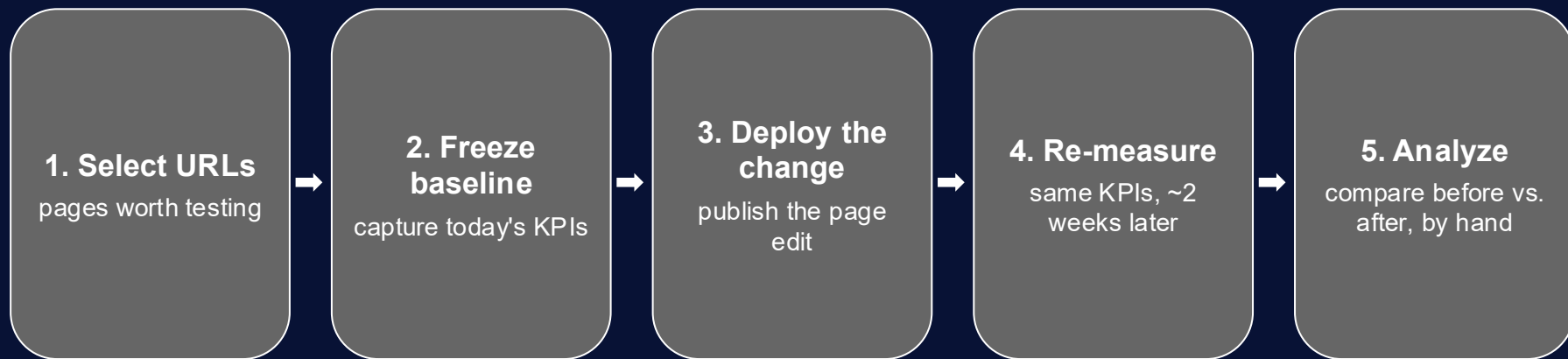


The real enemy: non-determinism & ambient drift

- Same prompt, same model, different answer
- Naive before/after can't isolate the signal
- **Ambient drift:** models, market, index — all moving

Doing this by hand: the manual workflow

- No tooling — a person does every step below, one page at a time



02 — Under the Hood

Where this lives: the Adobe Brand Visibility Impact Measurement Engine

Opportunities

Brand: Frescopa Coffee

Prioritized optimization opportunities based on player gaps, trending topics, and performance data

Opportunities Opportunity Workspace

Strategy Status: All Opportunity Status: All

Total Strategies

1

Completed, not yet tracked

0

Completed, tracked

0

Strategies

+ Create Strategy

Recover Content Visibility Completed, tracked View Performance

Created: Jun 17, 2026 | URLs: View 12 URLs | Platforms: ChatGPT (Free), Perplexity, ChatGPT (Paid) | Prompts: View 12 prompts | Goal: Citations | Completed: Jul 3, 2026

Key Outcomes

(measured over 14 days)

Citation Rate

25% → 43%

+72% improvement

Optimization Process

Baseline collected Jun 17, 2026

Deployed to edge Jun 18, 2026

Measuring impact Jun 18 – Jul 3, 2026

Experiment complete Jul 3, 2026

The unit of experimentation: one suggestion

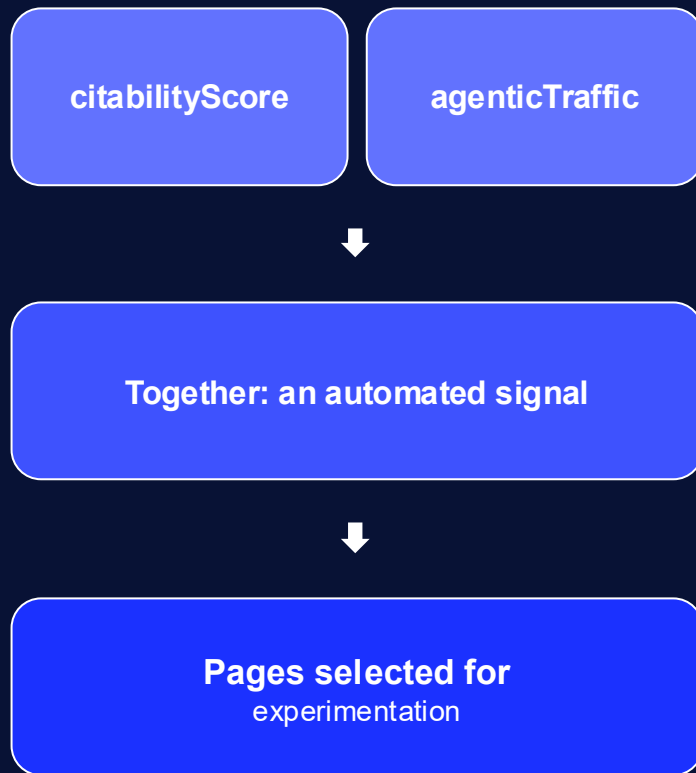
- Prompts are scoped to that page and optimization
- **60% prompt coverage** gates the experiment start
- Missing prompts? Auto-generated for you

	URL	Agentic Traffic (4 Weeks) ±1	Content Visibility	Content Gain Ratio ⓘ	Actions	Details
☐	All Domain URLs	—	—	—		
✓ ☒	https://frescopa.coffee/coffee	1,614,855	13%	7.9	 Preview	 Details
✓ ☐	https://frescopa.coffee/jp/coffee	12,507	11%	9.1	 Preview	 Details
✓ ☐	https://frescopa.coffee/en/coffee	18,973	13%	7.8	 Preview	 Details
✓ ☒	https://frescopa.coffee/blog	760,281	23%	4.4	 Preview	 Details
✓ ☒	https://frescopa.coffee/tea	1,619,515	17%	6.0	 Preview	 Details
✓ ☐	https://frescopa.coffee/jp/tea	17,286	16%	6.4	 Preview	 Details
✓ ☐	https://frescopa.coffee/en/tea	17,936	17%	5.9	 Preview	 Details
✓ ☐	https://frescopa.coffee/jp/machines	24,157	21%	4.7	 Preview	 Details

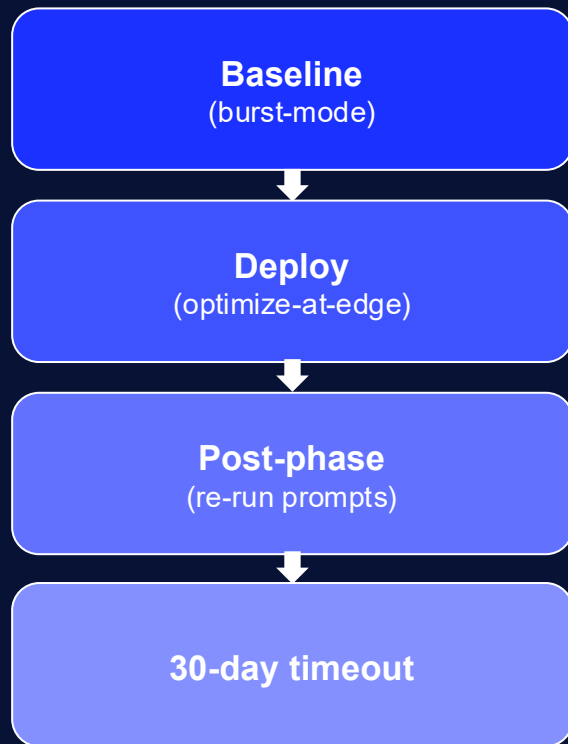
3 suggestions selected  Mark as Fixed  Ignore Suggestions  Deploy optimizations

Finding the right pages: automated scoring

- **citabilityScore**: content gain from optimization
- **agenticTraffic**: existing AI-crawler traffic
- Together: an automated signal for what's worth fixing

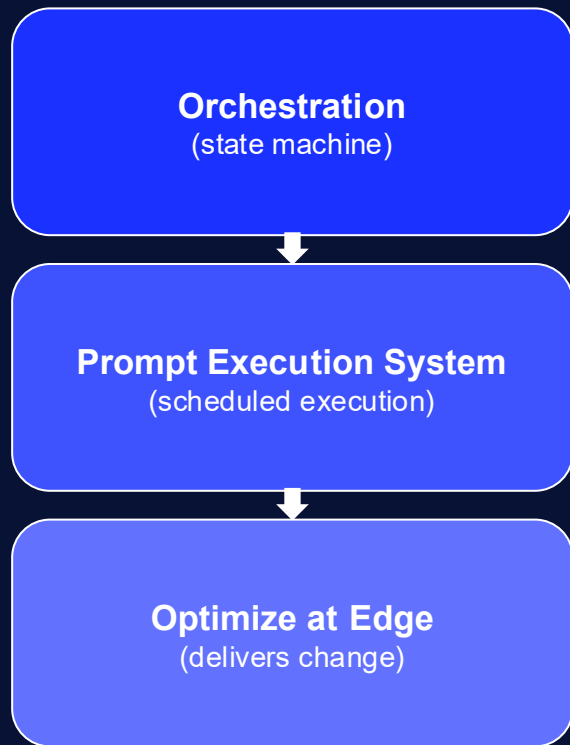


Baseline, deploy, measure — the real lifecycle



- Burst-mode baseline first
- Deploy via Optimize at Edge
- Post-phase re-runs the prompts
- 30-day timeout, end to end

Three systems, one workflow



- Orchestration: the experiment state machine
- Prompt Execution System: scheduled execution, citation data
- **Optimize at Edge**: delivers the change

Where's the Control group?

- Prompts aren't scoped only to the target URL — any citation lands in the same pre/post data
- No separately-scheduled control group needed
- Direct answer to the ambient drift problem

What real pilot timing taught the design

- ChatGPT: live-fetch, visible in ~40–45 min
- Pre/post windows are sized around this, not guessed
- CDN Configuration bugs are real
- Resiliency is super-critical

03 — Proving Causation: The Evidence Ladder

Raising the evidence bar: the three-tier ladder

TIER 1

What moved?

Raw citation lift
New URLs cited

TIER 2

Did it beat the market?

Share of voice

TIER 3

Can we claim impact?

DiD* = brand's lift vs.
competitors,
bootstrap CI** across runs

Decision layer

• SCALE

CI** excludes zero → a real effect, scale it

• ITERATE

CI crosses zero → can't rule out "no effect"

*DiD – Difference-in-Differences **CI – Confidence Interval

DiD and Confidence Interval, explained simply

DiD — Difference-in-Differences

Not "did citations go up?" but "did they go up more than the market's own drift, over the same window?"

DiD = your lift – the market's lift

Dev analogy: it's your A/B test control group — except the "control" is competitors moving through the same time window, since LLM traffic can't be split.

CI — Confidence Interval

We don't get one number back — we get a range of plausible true values.

95% CI [3.5, 8.3] — excludes zero

Dev analogy: a flaky test re-run 1,000 times (bootstrap)
— CI is the range the result landed in 95% of the time.

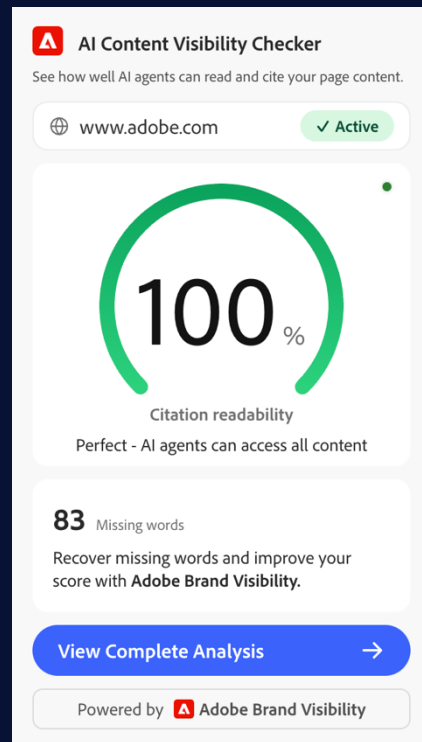
Bottom line: DiD isolates your effect from market noise. CI tells you whether that effect is real, or could just be zero.

04 — Product Experience

Live Demo: The Full Experimentation Flow

Case in point: content invisible to AI crawlers

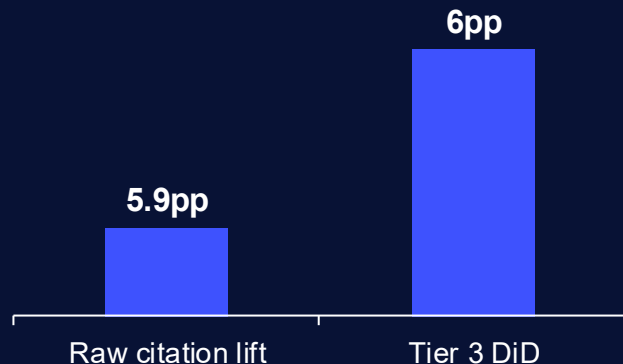
- Direct-fetch HTML vs. scraped HTML, compared
- A 1.2× content-gain ratio flags the page
- **Fix: Optimize at Edge** — pre-rendered, AI-only



A pilot: clearing the causation bar

- 131 prompts, 30 runs, 2 platforms, 52 competitors
- Treated-URL citation lift: **+5.9 pp***
- Tier 3 DiD^{**}: **+6.0 pp** (+13.2% relative)
- 95% bootstrap CI^{***} **[3.5, 8.3]** — **excludes zero**

Treated-URL lift vs. DiD



**pp – percentage points*

***DiD – Difference-in-Differences*

****CI – Confidence Interval*

05 — Where This Is Headed

From today's workspace to measuring any change, anywhere

- Opportunity Workspace already holds manual strategies Adobe didn't automate (Brand Claims, Earned Content)
- Today: organize and track — no automated measurement for these
- Same engine, pointed at a manual or offsite strategy, still measures it

Key takeaways

- You don't need cookies — you need shared prompts
- Lift alone isn't proof — demand a CI excluding zero
- The container for "any change" exists — automation's next

Thank you — Questions?

Scan for links

Chrome Extension



Demo

