THE GRAYSON COMPANY | POINT OF VIEW

Always On

AI Agents and the Future of Retail Measurement

A point of view for retail and consumer brand executives and their investors

The Grayson Company | November 2026



EXECUTIVE SUMMARY

Measurement must become a continuous system

$200.4B

Global retail media ad spend forecast for 2026 (WARC)

81%

Share of US off-site retail media going to Amazon and Walmart in 2027 (eMarketer)

+393%

Growth in AI-referred traffic to US retailers, Q1 2026 year over year (Adobe)

1

Continuous, not periodic

A standing system senses, diagnoses, tests and reports every week.

2

Incrementality is the currency

Seller-reported, attributed ROAS becomes a secondary metric.

3

Governance is the hard part

Decision rights and thresholds decide whether agents add value or noise.

4

Measurement drives valuation

Proving return on commercial spend is now a diligence item.

01

The case for change

Why today's measurement model no longer fits the market



Four forces have opened a gap between spend and proof



The cadence mismatch

Trackers report quarterly and mix models refresh yearly, while prices, rankings and auctions change daily.



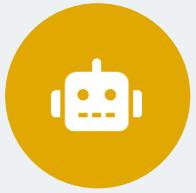
Fragmentation

A dozen or more retail media networks, each with its own attribution window and definitions.



Self-grading

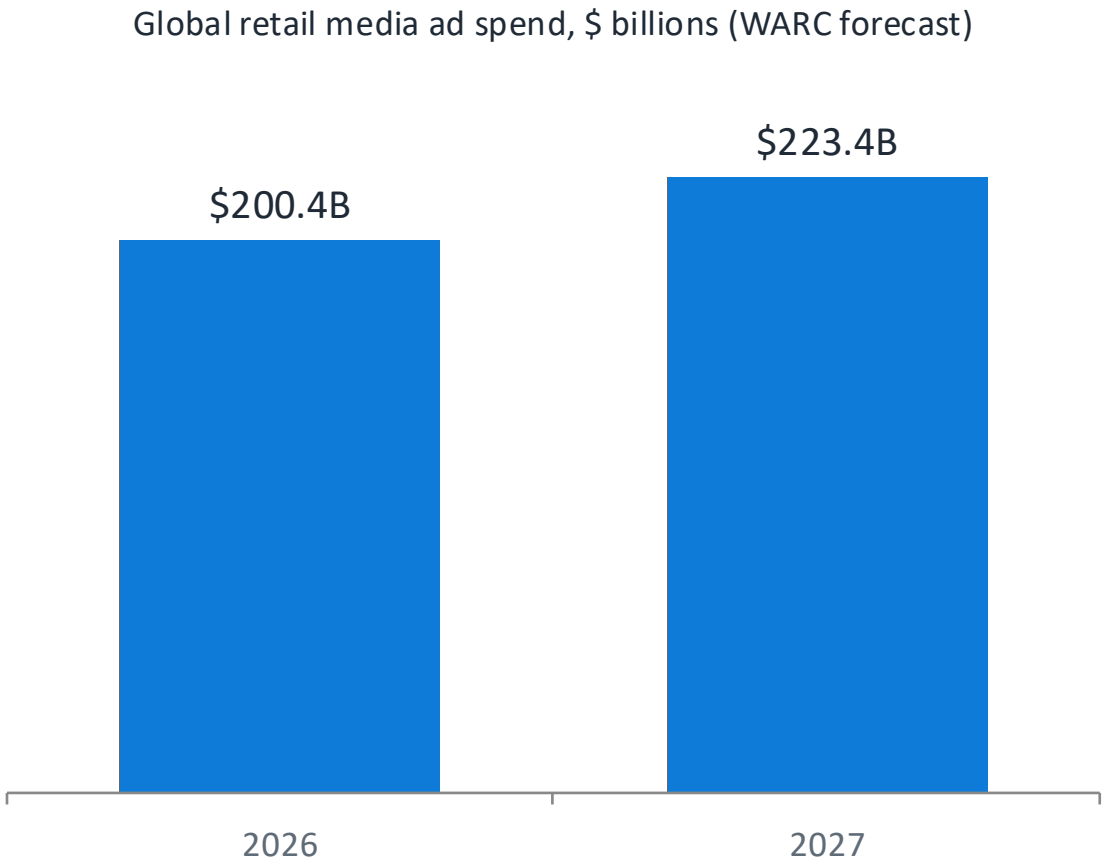
Most reported returns are attributed, not incremental, and certified by the seller of the media.

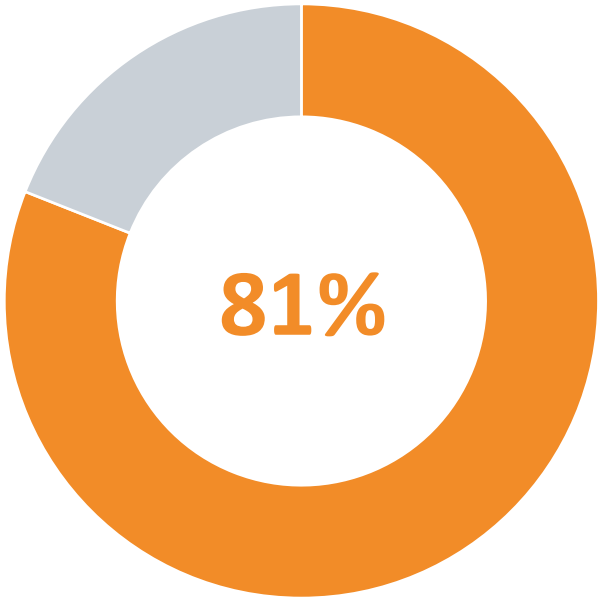


The invisible shelf

AI assistants shape choices before shoppers reach a retailer. No standard metric sees it.

Retail media is a core budget line, concentrated in two networks





US off-site retail media, 2027
Amazon and Walmart (orange) versus all other networks (gray). Source: eMarketer.

Brands increasingly spend in closed ecosystems that report on themselves, so independent proof of return matters more each year.

AGENTIC COMMERCE

AI assistants are becoming a shelf that brands cannot see

+393%

Year-over-year growth in AI-referred traffic to US retail sites, Q1 2026 (Adobe Analytics)

US agentic commerce by 2030, Morgan Stanley scenarios



$0 **$190B base case** **$385B bull case**

Roughly 10% to 20% of US e-commerce

Discovery moves upstream

Shoppers research and compare in assistants before reaching a retailer.

Attribution goes dark

Standard brand and share metrics do not record an assistant's recommendation.

02

What agents do, and how to govern them

Five agent roles and three client frameworks



FIVE AGENT ROLES

Always-on agents play five distinct roles



Sentinel

Watches every signal and flags material change

Overnight 30% drop in add-to-cart for a hero SKU



Diagnostician

Breaks a change into its causes

Lost Buy Box, competitor price cut and a content error



Experimenter

Runs controlled tests on a rolling basis

Geo and audience holdouts on every network



Auditor

Checks reported results against evidence

Network ROAS versus holdout lift



Narrator

Turns findings into decisions by audience

CFO sees incrementality; category lead sees shelf

Fast, visible wins **Durable financial value** **Executive usability**

Every decision sits on an explicit rung of authority

Authority is earned with evidence

Decisions move up a rung only after their recommendations are checked against outcomes. Start with nearly everything on rungs 1 and 2.

1

Observe

Monitors and alerts

Human interprets and acts

Anomaly alerts on sales, shelf and price

2

Diagnose

Explains why something changed

Human decides

Root cause of a decline at one retailer

3

Recommend

Proposes an action with evidence

Human approves

Reallocation across networks

4

Act with veto

Acts unless a human objects

Human monitors

Budget shifts within one network

5

Autonomous

Acts within limits, reports after

Human sets limits

Bids within 15%; pause ads on out-of-stocks

Three rules keep the ladder safe and credible

01

Reversibility sets the ceiling

Only quick, low-cost, reversible actions belong on rungs 4 and 5. Consumer-facing price, contracts and brand positioning stay at rung 3 or below.

02

Promotion is earned

A decision moves up only after its recommendations have been checked against outcomes, ideally with holdouts.

03

Every rung has an owner

A named executive owns the limits on each rung and reviews the agent's performance there.

Where to start: put nearly every decision on rungs 1 and 2, then make the path upward deliberate and visible to the board.

Decisions are only as good as the layers beneath them

Decide Narrator | Autonomy Ladder	Budget, assortment, content and price decisions, each on an explicit rung
Prove Experimenter | Auditor	Continuous holdouts, incremental return by network, monthly mix model
Sense and diagnose Sentinel | Diagnostician	Material changes flagged within a day, causes ranked by impact
Unify Data foundation	Common definitions, reconciled feeds, clean-room access
Signals Inputs	POS, retail media, digital shelf, price, reviews, inventory, AI answers

Value flows up

The practical test

Can two networks' return figures be compared directly today?

If not, start with the unify layer. Most programs invest at the top while the foundation stays broken.

Score six dimensions from 1 to 4 before scaling autonomy

Dimension	1 Fragmented	2 Developing	3 Established	4 Leading
Data access	Network dashboards; manual exports	Some API feeds; partial retailer data	Most feeds in a common store	Clean-room and first-party unified daily
Common definitions	Each network's metrics as given	Agreed for top networks	One dictionary across channels	Enforced by the agent automatically
Incrementality	Attributed ROAS only	Occasional lift studies	Holdouts on major networks	Continuous tests feed the mix model
AI-shelf visibility	Not measured	Ad hoc spot checks	Regular Share of Agent tracking	Share of Agent tied to fixes and sales
Decision rights	Unclear who acts	Owners for some metrics	Autonomy Ladder documented	Ladder reviewed quarterly
Talent and model	Analysts build reports	Analysts investigate on request	Analysts supervise agents	Measurement run as a product

Below 12: fix foundations first

12 to 17: build and prove

18 or more: move up the ladder

03

Worked examples

Retail media and agentic commerce. Figures are illustrative and hypothetical.



EXAMPLE 1 | RETAIL MEDIA | ILLUSTRATIVE

An agent turns reported return into proven return

$1.5B
premium snack brand

$80M
annual retail media

8
retail media networks

4x to 9x
reported ROAS

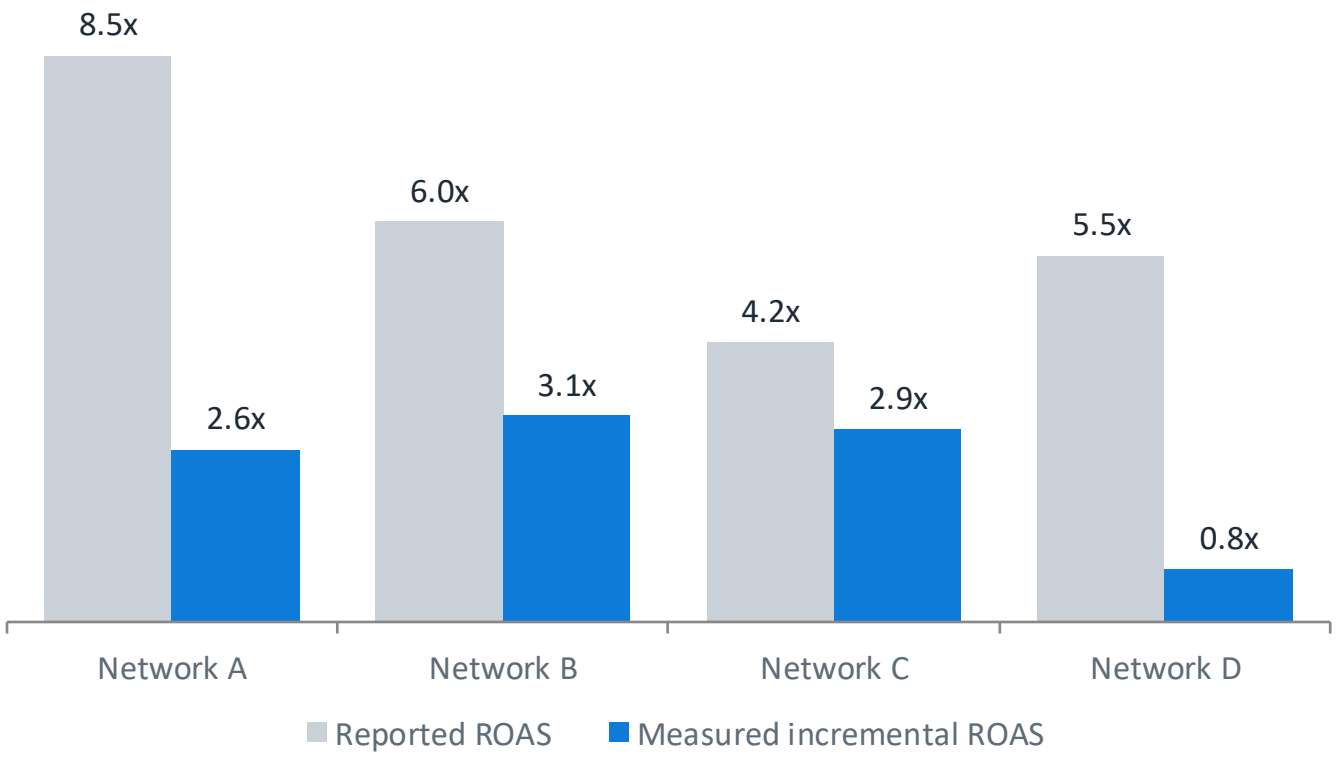
The CFO cannot reconcile these returns with flat category share. The CMO cannot say which dollars work.

1	**Normalize**	Restate every network on common definitions
2	**Hold out**	Keep a rotating 5% to 10% holdout per network
3	**Measure lift**	Incremental ROAS by network, tactic and SKU
4	**Audit**	Publish each network's reporting gap
5	**Recommend**	Monthly reallocation (rung 3); bids (rung 5)
6	**Learn**	Recalibrate the mix model monthly

EXAMPLE 1 | RETAIL MEDIA | ILLUSTRATIVE

Only 15% to 69% of reported return is incremental

Reported versus measured incremental ROAS after one quarter



$14M

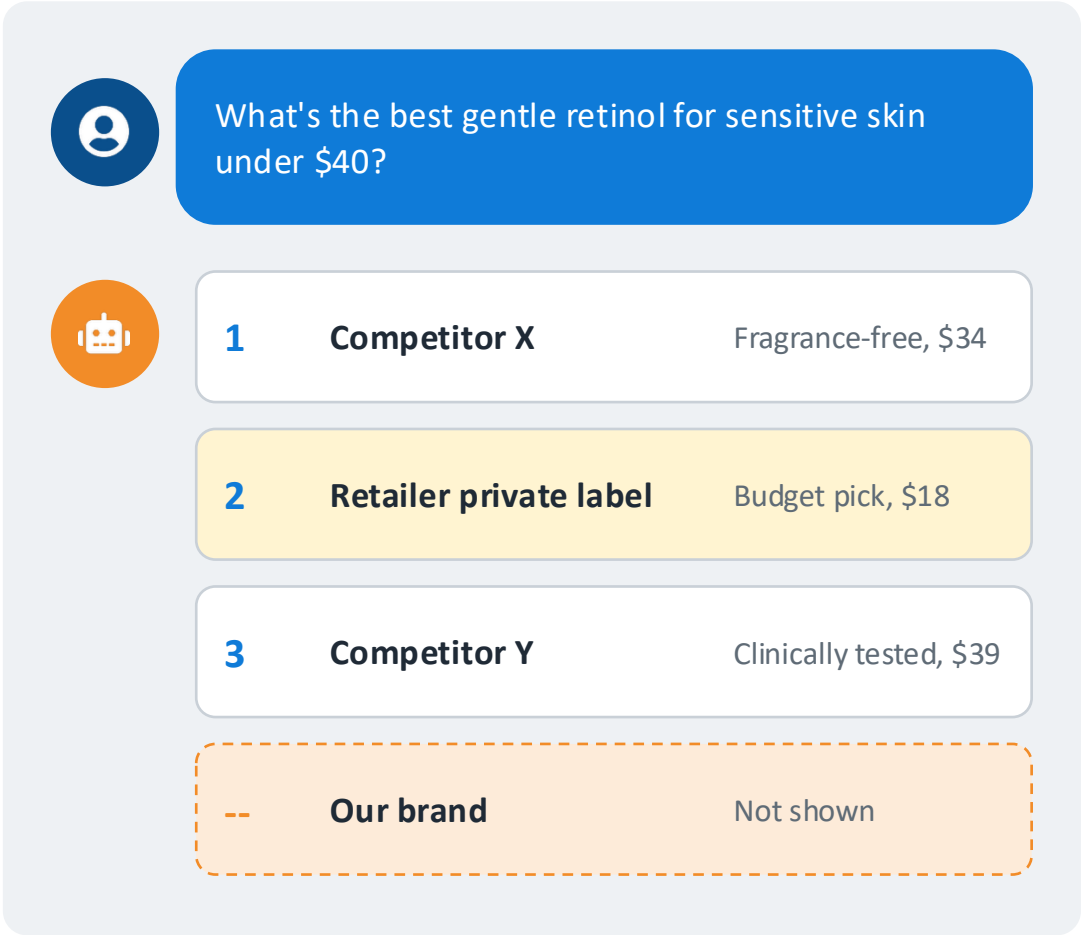
moved to stronger networks and tactics at the same total spend

Actions by network

- **A** Cut branded search; keep new items
- **B** Hold; shift toward off-site video
- **C** Increase budget 20%
- **D** Cut to the trade-terms minimum

EXAMPLE 2 | AGENTIC COMMERCE | ILLUSTRATIVE

The AI shelf has three to five slots and no paid placement

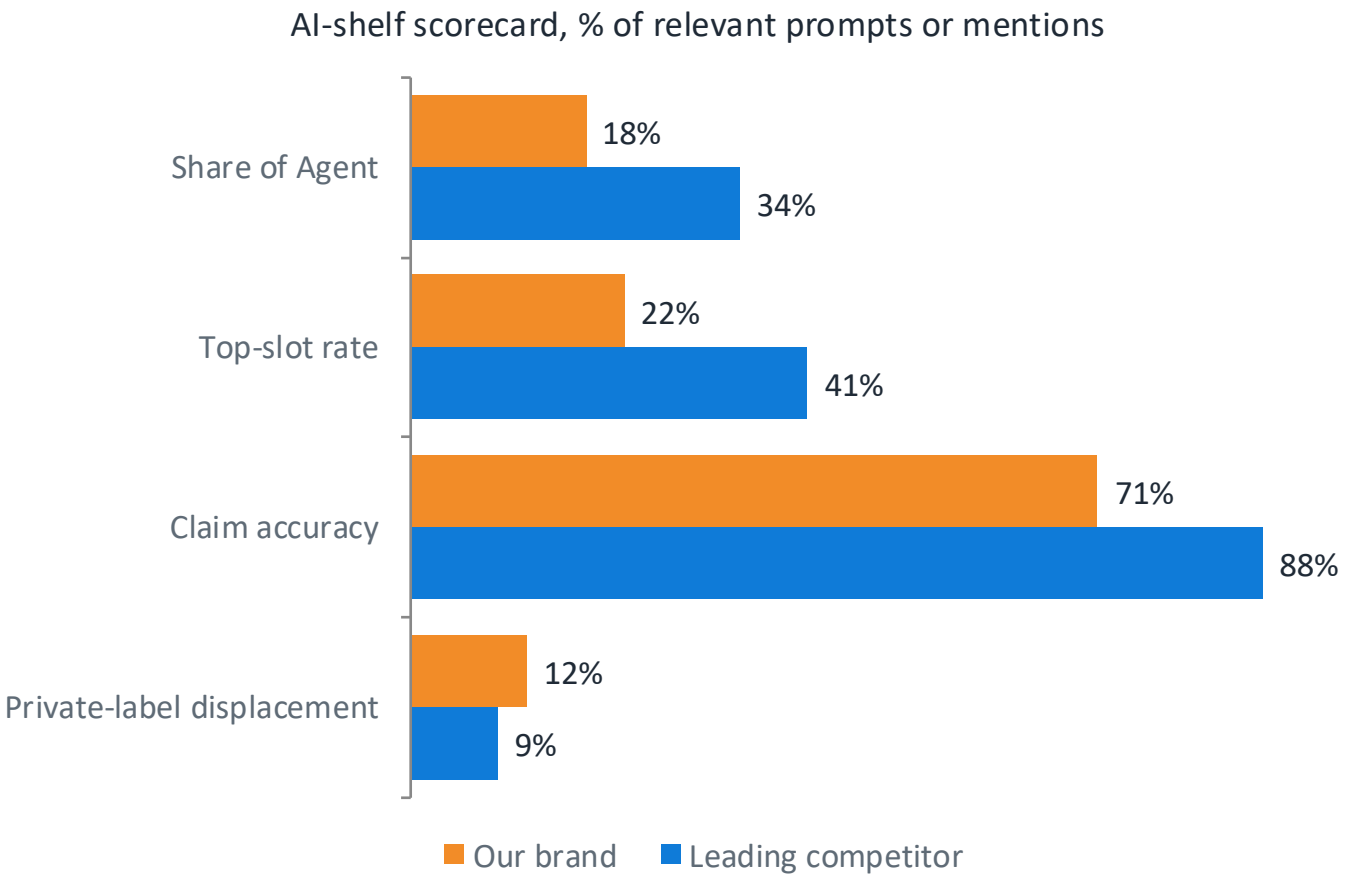


Share of Agent: how often the brand is recommended across a large panel of realistic shopping prompts

1	**Build the panel**	Thousands of prompts by need, price and persona
2	**Probe assistants**	General and retailer assistants, repeated
3	**Score answers**	Presence, rank, rivals, claims, price, retailer
4	**Check accuracy**	Claims and prices against product data
5	**Diagnose**	Thin data, missing attributes, weak reviews
6	**Close the loop**	Fix content, then measure the response

EXAMPLE 2 | AGENTIC COMMERCE | ILLUSTRATIVE

The brand trails on every AI-shelf metric that matters



3.1% vs 4.4%

Conversion on AI-referred traffic, brand versus competitor

The diagnosis is not media

- Assistants describe an old formulation
- The fragrance-free claim is missing
- A retailer feed lacks key attributes
- Fixes are cheap compared with media

Caution: read Share of Agent as a trend across a large, stable panel, never as a single figure.

04

Governance, investors and the roadmap

Managing the risks, reading the value, and getting started



The main risk is succeeding at the wrong objective, faster

Risk	How it shows up	Safeguard
False discoveries	Daily watching of hundreds of metrics produces noise	Thresholds set in advance; alerts ranked by impact
Optimizing the proxy	Spend drifts to the best attributed return	Incremental return governs; holdouts never switch off
Alert fatigue	Many flags, no owner; trust erodes	Every alert type has an owner and response time
Data dependency	Retailer terms block the cross-network view	Data access written into joint business plans
Model drift	Weekly refreshes quietly change conclusions	Versioned models; the agent explains changes
Unchecked autonomy	Actions compound before anyone notices	Spend caps, reversibility rules, a kill switch
Agents measuring agents	AI-shelf metrics shift with model updates	Large stable panels; model versions logged



CFO and CMO own it together

The agent increasingly decides where commercial capital goes. It warrants the oversight of any capital allocation process, not delegation to IT or a vendor.

IMPLICATIONS FOR INVESTORS

Measurement capability is a read on the quality of earnings

Diligence questions

1. What share of spend is measured as incremental, not attributed?
2. What happens to sales if the largest network is cut 20%?
3. Which trade terms require media, and is it tracked apart?
4. What is the brand's Share of Agent in its top need states?
5. Is product data a managed asset with an owner?
6. Who can act on findings, and how fast does budget move?

Where value is created

Margin

Shift spend to high-incrementality uses

Growth

Early AI-shelf position compounds

Leverage

Evidence strengthens retailer talks

Exit story

Proven returns support valuation

Portfolio angle: one capability across holdings pools cost and learning.

ROADMAP

Prove it on one network in 90 days, then earn autonomy

Prove the concept

Days 0 to 90

- Score readiness
- Unify top three networks
- Holdout on largest network
- Share of Agent baseline
- Draft the Autonomy Ladder

Scale what works

Months 4 to 6

- Holdouts on all networks
- First budget reallocation
- Monthly mix model refresh
- Product-data fixes live

Gate 1: Incremental return proven on one network

Extend and delegate

Months 7 to 12

- Chosen decisions to rung 4
- AI shelf tied to sales
- Evidence in JBP talks
- Board review of the Ladder

Gate 2: First reallocation beats the prior plan

NEXT STEPS

Three moves executives can make this quarter



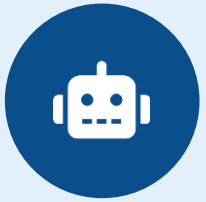
Run the Readiness Scorecard

With marketing, finance, sales and e-commerce in the same room.



Commission an independent holdout

On the largest retail media network, separate from the network's own reporting.



Baseline Share of Agent

Across the brand's top five need states, on a stable prompt panel.

The Grayson Company helps retailers and consumer brands design measurement governance, decision rights and the retailer data relationship.