CrowdWave

February 2026 | Confidential

Executive Summary

Situation: Market research takes 4-6 weeks and $25K+ per study. By the time you have answers, the market has moved.

Complication: Competitors running 20 concept tests per quarter will outlearn teams running 2-3. Speed is the new competitive moat.

Resolution: CrowdWave's calibrated simulation delivers 95% directional accuracy in minutes. Validated against Pew, Gallup, and AARP — averaging 2-point error. Test 10x more ideas, kill losers instantly, validate only winners.

Your research budget buys 3 studies per quarter — competitors are testing 20

Traditional research economics make experimentation impossible

Constraint	Traditional	Impact
Cost per study	$25,000+	Budget supports 2-3 tests/quarter
Time to insight	4-6 weeks	Decision window closes before data
Iteration capacity	1-2 rounds max	No room to explore alternatives

The math: At $25K/test, a $75K quarterly budget buys 3 tests. Competitors using simulation run 50 for near-zero marginal cost.

Winner: The team that learns faster.

We predicted real consumer behavior within 2 points — blindly, across 27 tests

Blind validation against authoritative sources

Prediction	CrowdWave	Actual	Error
Adults 50+ smartphone ownership	89%	90%	**1 pt**
Americans identifying as Independent	44%	45%	**1 pt**
Trust in scientists	77%	77%	**0 pts**
"Very concerned" about AI	50%	48%	**2 pts**
Manufacturing industry NPS	64	65	**1 pt**

Sources: Pew (N=5,000+), Gallup (N=13,000+), AARP (N=3,838)

27 tests. Mean error: 1.9 points. 100% within 5 points.

Raw AI predictions are wrong by 9 points — calibration cuts error by 79%

The calibration difference

Metric	Raw LLM	Calibrated	Improvement
Mean absolute error	9.1 pts	1.9 pts	**79% reduction**
Predictions within 2 pts	7%	81%	11x better
Predictions within 5 pts	30%	100%	Complete

Why raw AI fails: Training data averages across contexts. No demographic calibration. Missing behavioral adjustments. Can't handle polarization.

What we built: 8 bias patterns with corrections, 20+ domain calibrations, 5M+ human responses as ground truth.

Accuracy is predictable: trust hits ±2 pts, purchase intent needs validation

Match the tool to the question

Accuracy Zone	Error	Question Types	Use
HIGH	±2-3 pts	Trust, awareness, party ID, demographics	Decisions
MEDIUM	±4-5 pts	Satisfaction, NPS, concern levels	Direction
LOW	±8-15 pts	Purchase intent, price, polarized topics	Validate

Examples:

- "Which 3 of 10 concepts resonate?" → **Simulation alone** (80%+ top-3 match)
- "How much would customers pay?" → **Validate** (intent overstates 3-5x)
- "How do voters feel about immigration?" → **Segment by party or miss by 50 pts**

LLMs underestimate seniors by 25% — we found the correction factors

Documented bias: Adults 60+ and technology

Prediction	Raw LLM	Calibrated	Actual	Fix
Smartphone ownership (50+)	72%	89%	90%	×1.25
Daily internet use (60+)	60%	82%	83%	×1.35
Video streaming (70+)	40%	65%	64%	×1.60

Why: LLM training over-represents stereotypes. Reality: 90% of adults 50+ own smartphones.

Fix: Validated against AARP Tech Trends (N=3,838). Built correction multipliers by age.

This pattern repeats across 8 documented biases — each with validated corrections.

Political topics require segmentation — the "average American" doesn't exist

Pew Research, February 2025 (N=5,086)

Issue	Republican	Democrat	Gap	"Average"
Immigration concern	75%	25%	**50 pts**	48%
Climate concern	25%	70%	**45 pts**	45%
Gun violence concern	35%	70%	**35 pts**	52%

The trap: Report "48% concerned about immigration" and you've described no one. Republicans: 75%. Democrats: 25%. The average is fiction.

Our rule: CrowdWave flags polarized topics and enforces segmentation. No misleading averages.

Stated purchase intent overstates reality by 3-5x

The intent-action gap (meta-analysis, 50+ studies)

Survey Response	Stated	Actual	Gap
"Definitely will buy"	80-90%	25-35%	**3x**
"Probably will buy"	50-60%	10-20%	**4x**
"Might consider"	30-40%	3-8%	**5x**

Implications:

- Never report raw intent as conversion
- Apply factors: Definitely $\rightarrow$ ×0.30, Probably $\rightarrow$ ×0.15, Might $\rightarrow$ ×0.05
- Pricing decisions: Always validate with behavioral data

CrowdWave applies corrections automatically on intent questions.

C-suite needs role-specific calibration — CHROs are 75% more worried about AI than CEOs

Conference Board C-Suite Survey 2026 (N=1,732)

Concern	CEO	CFO	CHRO	CMO
Cyberattacks	+30%	+40%	**+60%**	-10%
AI disruption	-10%	+5%	**+40%**	+10%
Economic uncertainty	+35%	+50%	+50%	+25%

Calibration vs. generic "executive" baseline

Insight: "Executive concern" predictions miss role variation by 40+ points. CHROs worry about AI. CMOs don't worry about cyber.

Application: Specify the role. Generic predictions waste accuracy.

Industry NPS varies 35 points — LLMs assume everyone is at 40

Survicate NPS Benchmark 2025 (N=5.4M responses)

Industry	Actual NPS	LLM Prediction	Error
Manufacturing	65	40	**-25 pts**
Healthcare	61	40	**-21 pts**
Retail	55	40	**-15 pts**
Fintech	46	40	**-6 pts**
Software	30	40	**+10 pts**

Problem: LLMs anchor on "NPS ≈ 40" regardless of industry. Reality: 30 to 65+.

Fix: Industry-specific baselines. Manufacturing starts at 65, not 40.

Simulation changes research economics: 10x velocity, 1/100th cost

Metric	Traditional	CrowdWave	Multiple
Time to insight	4-6 weeks	Minutes	**1000x**
Cost per concept	$25,000	~$0	**Near-infinite**
Tests per quarter	2-3	20-50+	**10-20x**
Iterations	1-2 rounds	Unlimited	**Continuous**

The compounding effect:

- Week 1: Simulate 20 concepts, kill 15 losers
- Week 2: Iterate on 5 survivors
- Week 3: Validate top 2 with real respondents ($50K)
- Week 4: Launch with confidence

Traditional: Test 2 concepts in 6 weeks. Hope you picked right.

Match simulation confidence to decision stakes

	LOW STAKES	HIGH STAKES
HIGH ACCURACY	✅ Simulation only	✅ Simulation + validation
MEDIUM ACCURACY	✅ Directional use	⚠️ Validate before spend
LOW ACCURACY	⚠️ Directional only	❌ Always validate

Thresholds:

- Under $100K → Simulation sufficient (high/medium accuracy)

- $100K - $1M → Simulate first, validate finalists

- Over $1M → Simulation screens, humans decide

Principle: Simulation accelerates decisions. Doesn't replace judgment.

What we built: 20+ domains, 8 bias corrections, 5M+ human responses

Component	Coverage
Validated domains	20+ (trust, tech, NPS, executive, consumer, travel, healthcare)
Bias corrections	8 patterns with validated fixes
Human data	5M+ responses from Tier 1 sources
Calibration factors	100+ domain-specific multipliers
Test cases	27 blind predictions, published validation

Source tiers:

- **Tier 1:** Pew, Gallup, AARP (probability samples, N>1K, peer review)
- **Tier 2:** McKinsey, Deloitte, Conference Board (large N, established)
- **Tier 3:** YouGov, Harris (online panels, directional)

Three actions to capture the speed advantage

1. Integrate simulation into every research project

Simulate first. Screen concepts, kill losers. Then decide what needs validation.

2. Set decision thresholds

- Screening/ranking → Simulation only
- Major campaigns → Simulation + validate finalists
- Pricing/conversion → Always validate

3. Track and compound accuracy

Log predictions vs. outcomes. Feed misses back. Calibration improves continuously.

The question isn't whether to use simulation — it's how much ground you'll lose to competitors who start first.

Appendix

Appendix A: Validation Detail (27 tests)

Domain	Tests	Mean Error	Range
Trust/institutions	5	1.8 pts	0-3 pts
Technology adoption	6	2.1 pts	1-4 pts
Political identity	4	1.2 pts	0-2 pts
NPS by industry	5	2.4 pts	1-4 pts
Consumer concerns	4	1.6 pts	1-3 pts
Executive attitudes	3	2.8 pts	2-4 pts

Sources: Pew Research, Gallup, AARP Tech Trends, Survicate, Conference Board, Edelman

Appendix B: Demographic Calibration

Segment	Tech Adoption	Emotional Intensity	Price Sensitivity
Adults 50-69	×1.30	—	—
Adults 70-79	×1.40	—	—
Adults 80+	×1.50	—	—
Women 60+	×1.35	×1.30	×0.85
High-income ($150K+)	+0.3	—	×0.60
Parents (child context)	—	+0.6	×0.80

Appendix C: Bias Corrections

Bias	Direction	Correction	Source
Senior tech	Under	×1.30-1.65	AARP 2025
AI concern	Over	×0.90	Pew/YouGov
Status quo	Under	+15-20 pts	Behavioral
Intent gap	Over	×0.30-0.55	Meta-analysis
Emotion	Under	×1.20-1.30	Study
Life satisfaction	Over	-3 to -5 pts	Gallup
Partisan avg	Wrong	Segment	Pew
Open-end polish	Over	20% low-qual	Benchmark

CrowdWave