North Star Intel

EDITION 2.2 · NOVEMBER 16, 2026
NSI-D2E-2.2-EN

THE STANDARD — UNITED STATES

TEN MOVES · ONE RECORD · FOUR VERDICTS · ONE-HOUR AUDIT

DECISION TO EVIDENCE

The standard for decisions you can prove. Ten moves past decision intelligence, and a one-hour audit that shows where you stand.

NORTH STAR DECISION LOOP — SIX STAGES · EACH ONE LEAVES A RECEIPT

SIGNAL → SCORE → ASSIGN → EXECUTE → VERIFY → LEARN

We read the decision intelligence and operational intelligence market, the research behind it and our own system. Most of it sells faster decisions. Little of it proves a decision worked. This edition sets the standard for proving it.

INSIDE · PAGE 19 IS THE AUDIT

02 How to use this
03 Four kinds of intelligence
04 Decision to evidence
05 Why it matters now
06 The standard, version 1.0
07 Ten moves, 1 to 5
08 Ten moves, 6 to 10
09 The Decision Record
10 Grading evidence
11 What verified means
12 Prove it with what you have
13 Before AI decides
14 Scoreboard + six levels
15 Seven decision leaks
16 Four findings that survive
17 Put a number on it
18 Worked example
19 The scored audit → /100
20 The horizon
21 Seven days + review
22 Sources and grades
23 About + terms
24 The standard on one page



PAGES 6 TO 14 ARE THE STANDARD. PAGE 19 IS THE TEST.

Read the ten moves on pages 7 and 8. Score yourself on page 19. Then pick one move to make this week.

© 2026 North Star Intelligence LLC. All rights reserved. Educational guide, not legal or financial advice. Terms on page 23.

02 · START HERE

You are not short on data. You are short on decisions that land.

Leadership teams that run several locations make more decisions in a month than they can list. A few get written down. Fewer get one owner and one date. Almost none get checked against a number written before anyone acted. That is where money and credibility leak, one location at a time.

This standard runs the other way. Find the decisions that never landed. Prove which ones paid. Repeat what worked.

01 FIND

Score the last 20 decisions your leadership team made, from records, not memory.

02 PROVE

Write a baseline, a comparison and a check date before acting. Then check on the date.

03 REPEAT

Keep score of your calls, so each decision starts smarter than the last one.

THREE PARTS

FOUNDATION
Pages 3 to 5
What decision intelligence is, the difference evidence makes, and why it matters now.

THE STANDARD
Pages 6 to 14
Principles, ten moves, the record, grading, verification, proof, the AI line and the scoreboard.

THE AUDIT
Pages 15 to 21
Leaks, research, the calculator, a worked example, the one-hour audit and your first seven days.

THREE IDEAS TO REMEMBER

- **A decision is a record, not a meeting.**
  A question, an owner, a date and a number. If it is not written, it was not decided.
- **Evidence is graded, not collected.**
  One reconciled number outweighs a thousand rows from a report nobody trusts.
- **Judge the process first, then the outcome.**
  Good decisions can lose and bad ones can win. Only the record shows which one happened.

WHO IT IS FOR

Multi-location operators, franchise systems and the leadership teams that run them. The standard works from one location up. Test and hold, on page 12, needs two or more.

WHAT TO HAVE ON THE TABLE

- ✓ Meeting notes from the last 90 days
- ✓ Email and chat threads where choices were made
- ✓ Job, sales or point-of-sale exports
- ✓ Your calendar, for the dates that were promised

We will tell you when the answer is not software.

Many guides in this market are published by companies that sell decision software, and they tend to conclude that you should buy it. On page 18 you will find a business whose first fix was a 15-minute Monday meeting. If the honest answer for you is a spreadsheet and a standing review, you should be able to reach it with this standard.

03 · FOUR KINDS OF INTELLIGENCE, PLAINLY

Four kinds of intelligence. Each answers one question.

The terms sound new. The problem is old. For eighty years, armies, intelligence services and companies have tried to turn information into choices they could defend. Each generation of tools answered one more question.

BUSINESS INTELLIGENCE

What happened?

Reports and dashboards on the past.

ENDS AT THE CHART

OPERATIONAL INTELLIGENCE

What is happening right now?

Watches live events and steers work in progress.

ENDS AT THE ALERT

DECISION INTELLIGENCE

What should we decide?

Models, scores and automates choices.

ENDS AT THE RECOMMENDATION

DECISION TO EVIDENCE

Did it happen, and did it work?

Writes each decision first, then proves the outcome.

ENDS AT A VERIFIED RESULT

WHEN	SHIFT	WHAT CHANGED
1940s	**Operations research**	Teams used math to plan convoys, routes and supply under pressure.
1951	**A word problem**	An estimate called an attack a “serious possibility.” Its own authors meant odds from 20 to 80 percent.
1990s	**Business intelligence**	Dashboards showed what happened. They rarely said what to do next.
2000s	**Operational intelligence**	Systems began watching events as they happened and raising alerts.
2007	**Analytic standards**	U.S. intelligence analysts got written rules for sources and uncertainty.
2011 to 2015	**Scored judgment**	Forecasting tournaments showed judgment can be measured and trained.
2020s	**Decision intelligence**	Analysts named a discipline for engineering decisions and feeding outcomes back.
Now	**Agentic AI**	Systems recommend and act. The record of who decided becomes the asset.

DECISION INTELLIGENCE

The discipline of designing how a decision gets made, who owns it, and how its outcome is checked and fed back into the next one.

NOT THIS

Not a dashboard.
A dashboard shows what happened. A record shows what you chose and whether it worked.

Not a forecast.
A prediction nobody checks later is an opinion with a chart.

Not an AI that decides for you.
A system can rank options. A person answers for the choice.

Not more data.
Most teams already hold the numbers. What they lack is a baseline and a date.

04 · DECISION TO EVIDENCE, THE DIFFERENCE

Decision intelligence improves the decision. Decision to evidence proves it.

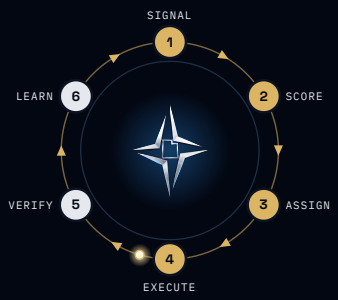
As analysts define it, decision intelligence includes feedback on outcomes. In practice, much of the software sold under the name centers on modeling and automating decisions at scale. Decision to evidence makes the feedback the product.



Signals in. Owned decisions out. Each step leaves a receipt. Artwork from northstarintel.ai.

	DECISION INTELLIGENCE, AS COMMONLY PRACTICED	DECISION TO EVIDENCE
The question	What should we decide, and how fast?	Did the decision happen, and did it work?
The unit	Models, data and decision logic	One Decision Record per decision
Where it ends	At the recommendation or the automated action	At a verified outcome on a set date
Success means	Faster, more consistent decisions at scale	A result measured against a baseline written first
What remains	Dashboards, scores and logs	Four receipts: interaction, decision, action, outcome
Who answers	Often the model, the system or the team	One named owner, checked by someone else
Best fit	Decisions made thousands of times a day	The decisions a leadership team makes each week

WHAT EACH STAGE LEAVES BEHIND



●● Gold: the stage left its receipt. Silver: not yet.

1 · SIGNAL
The interaction: source, time, consent

2 · SCORE
Priority and confidence, as numbers

3 · ASSIGN
One owner, one date, one next action

4 · EXECUTE
Proof the action happened

5 · VERIFY
The outcome against its baseline

6 · LEARN
The verdict, the lesson, the calibration

WHAT EVIDENCE LOOKS LIKE

RECEIPT READING — ILLUSTRATIVE

AS OF 2026-11-08

DECISION	Add a second dispatcher before storm season JUDGMENT
OWNER	Dispatch lead
DEADLINE	2026-11-21
PRIORITY	88 / 100
CONFIDENCE	●●○ Moderate confidence. Two seasons of dispatch logs agree. The hiring timeline is untested.
EVIDENCE	Dispatch log export, phone system FACT
SOURCE GRADE	B2 · Usually reliable · Probably true
VERDICT	Inconclusive until the indicator is checked against a record.
SEAL FINGERPRINT	sha256:3f9c…a1e7

Sample decision, not a client's record.

05 · WHY IT MATTERS NOW

Deciding is much of the job. Few teams can prove their decisions worked.

Two large bodies of research point the same way. Leaders spend much of their time deciding, most say that time is poorly used, and companies that decide and deliver well tend to perform better. Neither measures whether a single decision worked. That is the gap this standard closes.

20%
of executives said their organizations excel at decision making
McKINSEY GLOBAL SURVEY · 2019 · B2

61%
said most of their decision-making time is used ineffectively
McKINSEY GLOBAL SURVEY · 2019 · B2

37%
of their working time, on average, went to making decisions
McKINSEY GLOBAL SURVEY · 2019 · B2

1,000+
companies in a ten-year program linking decision effectiveness to business performance
BAIN AND COMPANY · 2013 · B2

WHAT THE ANALYSTS SAY

“ungoverned decision-making”

What Gartner says raises legal, operational and reputational risk as AI agents make more decisions. Press release, June 16, 2026.

In March 2025 Gartner said three in four global enterprises would apply decision intelligence by 2026. Logging decisions is the first step. A log written after the fact proves little. This standard is the rest: write it first, verify it, keep the proof.

READ RESEARCH THE WAY YOU READ RECORDS

Bain reported a link between decision effectiveness and business performance that held at a minimum 95 percent confidence level. An article republished on Bain’s own site called it a “95 percent correlation.” Those are different claims. A confidence level says the link is unlikely to be chance. It does not say how strong the link is. We cite the first and refuse the second.

WHAT IT MEANS FOR AN OPERATOR

The bottleneck is follow-through.

Most teams have the numbers. What they lack is an owner, a baseline and a date.

Records beat recollection.

A dated record settles arguments that memory restarts every quarter.

Proof travels.

A verified outcome can be shown to a franchisor, a lender or a buyer. A story cannot.

06 · THE DECISION TO EVIDENCE STANDARD · VERSION 1.0

A decision is finished when the evidence says so.

This standard sets what a decision record must contain, when an outcome may be called verified, and how the proof is kept. It applies to any decision that commits money, people or time, at one location or five hundred. Must means required. Should means expected, unless you write down why not.

SEVEN PRINCIPLES

1	**Write it before you act.**	The question, owner, measure, baseline and check date must exist before the work starts.
2	**One owner, separate roles.**	One person must answer for each decision. Approval, action and review should be named, and the reviewer must not be the owner.
3	**Grade every claim.**	Each claim should state how it is known, and each source should carry a grade.
4	**Keep the chain.**	Interaction, decision, action and outcome should each leave a dated record.
5	**Verify against a baseline.**	Nothing may be called verified until it clears all seven bars on page 11, including a baseline written first and two independent sources.
6	**Record every verdict.**	Gain, no gain, inconclusive and disputed must all be kept.
7	**People decide what matters.**	A system may suggest. A named person must approve anything touching pay, hiring, safety, the law or one specific person.

TEN DEFINITIONS

TERM	MEANING
Decision	A choice that commits money, people or time.
Decision Record	One decision, written down: question, options, owner, date, evidence, measure, baseline and result.
Owner	The one person who answers for the decision. A person, not a team.
Baseline	The value of the measure before acting.
Comparison window	The period the result will be measured over, set in advance.
Claim state	How a claim is known: observed, inferred, predicted, attested by a person, or verified.
Receipt	A dated record that proves one step happened.
Verdict	One of four: verified gain, verified with no gain, inconclusive or disputed.
Seal	A digital signature showing a receipt has not changed since it was verified.
Calibration	Whether your 70 percent calls come true about 70 percent of the time.

The Evidence Audit on page 19 is the conformance test for this standard. A score of 20 or more out of 24 meets it at Level 3, as defined on page 14. The audit does not test every bar on page 11 or principle 7.

07 · TEN MOVES FORWARD · 1 TO 5

Ten moves from deciding faster to proving it worked.

Each move names where decision and operational intelligence usually stop, what this standard requires instead, where the move already lives in North Star, and a first step you can take on Monday without buying anything.

1

FROM ~~Logged after the fact~~ → TO **Written before the act**

WHY
A record written afterward can be bent to fit the result. Science calls the fix pre-registration: state the measure, baseline and check date first.

IN NORTH STAR TODAY
An outcome cannot be recorded without a metric, a result, a baseline and a comparison window.

START MONDAY
Write the measure, the baseline and the check date before your next decision ships.

2

FROM ~~The team decided~~ → TO **Four named roles**

WHY
Shared ownership is no ownership, and nobody should grade their own result. Name the owner, the approver, the person who acts and the reviewer.

IN NORTH STAR TODAY
Approval needs a named approver and a policy, and action an authorized actor. The independent reviewer accepts the assignment in a separate reviewer portal, and a seal needs the owner's sign-off.

START MONDAY
Add approver and reviewer columns to your decision register.

3

FROM ~~Confident prose~~ → TO **Graded claims**

WHY
Words are not odds and sources are not equal. Draw on four kinds of evidence: research, your own records, experienced people and the people affected.

IN NORTH STAR TODAY
Every claim carries a state: observed, inferred, predicted, attested by a person, or verified. Sources are graded A to F and 1 to 6.

START MONDAY
Label each line of your next decision memo with its claim state and its grade.

4

FROM ~~Events on a dashboard~~ → TO **An evidence chain**

WHY
Operational intelligence shows events as they happen. It rarely ties the event to the decision, the action and the result.

IN NORTH STAR TODAY
Four linked receipts: the interaction with its time, channel, consent and retention, the decision with its approver and baseline, the authorized action, and the reviewed outcome.

START MONDAY
For one open decision, find one record for each of the four links.

5

FROM ~~Marked done~~ → TO **Verified**

WHY
A closed ticket says the work stopped. It does not say the result arrived.

IN NORTH STAR TODAY
Verified requires two distinct sources, a baseline, a window, a metric, a named reviewer, a review date and alternatives on file. We check distinct, not independent.

START MONDAY
Do not call anything verified on one source.

08 · TEN MOVES FORWARD · 6 TO 10

The second five: comparison, honesty and the human line.

Moves 1 to 5 make a decision checkable. Moves 6 to 10 make the check fair, the record trustworthy to outsiders, and the line between people and systems explicit.

6

FROM ~~Before and after~~ → TO **A counterfactual**

WHY
Markets move while you act, so before and after credits you with the tide. Locations that did not change show what would have happened anyway.

IN NORTH STAR TODAY
The outcome record asks for the comparison window and anything else that changed in it.

START MONDAY
Hold back at least as many locations as you change, or stagger the rollout so every wave has a comparison.

7

FROM ~~Wins only~~ → TO **Every verdict**

WHY
When only wins get reported, every win looks suspect. Researchers call it the file-drawer problem.

IN NORTH STAR TODAY
Four verdicts: verified gain, verified with no gain, inconclusive and disputed. A review starts at inconclusive.

START MONDAY
Report your misses in the same meeting as your wins.

8

FROM ~~Trust us~~ → TO **Check it yourself**

WHY
When any AI can write a convincing memo in seconds, a memo proves nothing. A signed record that cannot be quietly edited does.

IN NORTH STAR TODAY
Receipts are signed with Ed25519 and chained by fingerprint. Reviewer files get a SHA-256 fingerprint on upload and can't be overwritten. Anyone can check a receipt at northstarintel.ai/verify, and consenting owners publish at northstarintel.ai/receipts.

START MONDAY
Store each verified result with its date, sources and reviewer where nobody can edit it later.

9

FROM ~~Opinions about judgment~~ → TO **A calibration record**

WHY
Short probability training and keeping score improved forecasters in randomized trials. Judgment nobody scores never improves.

IN NORTH STAR TODAY
Not built yet, so this move is manual for now.

START MONDAY
Start the calibration log on page 12: chance before, result after, Brier score each month.

10

FROM ~~AI that acts~~ → TO **AI inside a written boundary**

WHY
Rules on automated decisions are arriving. The EU's AI Act will require deployers of high-risk systems to keep their logs for at least six months. A boundary written first costs less than one written after a complaint.

IN NORTH STAR TODAY
Human approval is on by default. AI suggests ratings and wording only, the score comes from the published formula, and any decision can be escalated to a person.

START MONDAY
Write the list of decisions that always go to a named person.

09 · THE DECISION RECORD + RECEIPTS

Write the decision before you automate it.

If you cannot write the decision down without software, software will not save it. These are the fields North Star keeps for every decision, stage by stage. Use the same fields on paper until you need more.

EVERY RECORD CARRIES

Question	Owner	Due date	Priority	Status
What we are choosing, and by when	The one person who answers for it	A date, not someday	Opportunity, risk of waiting and effort, each 1 to 3, scored 0 to 100	Open, approved, in progress, completed or escalated

THEN, STAGE BY STAGE · AND THE RULE THAT GATES EACH ONE

01
Interaction
- What happened
- When
- Channel
- Consent
- Disclosure
- Retention

GATE
Needs the event, time, channel, permission and retention.

02
Decision
- Options, including doing nothing
- Graded evidence
- Approver and policy
- Expected result
- Baseline

GATE
No approval without a named approver and a policy.

03
Action
- Next action
- Authorized actor
- Authorization
- Target system
- Human approval, on by default
- Action time

GATE
No action without a named actor and an authorization.

04
Outcome
- Metric
- Result
- Comparison window
- Reviewer and review date
- Alternative explanations
- Verdict

GATE
No outcome without metric, result, baseline and window.

> Human approval is on by default for a reason. Before you start, write down what always goes to a named person: pay, hiring and firing, a refund above your limit, a safety issue, a legal question, and any decision about one specific employee or customer. Every time.

EVERY DECISION SHOULD LEAVE FOUR RECEIPTS · THE EVIDENCE CHAIN

01 **Interaction** — Source · Time · Consent ▶
02 **Decision** — Owner · Options · Approval ▶
03 **Action** — Authority · Action · Record ▶
04 **Outcome** — Baseline · Evidence · Review

THREE FREE CHECKS

Two minutes
Six questions, one per loop stage
northstarintel.ai/decision-gap-score

Five minutes
One decision against the record standard, before anyone acts
northstarintel.ai/decision-audit

One hour
The last 20 decisions, scored out of 24
Page 19 of this standard

10 · GRADING EVIDENCE · THE TRADECRAFT, BORROWED

How an intelligence analyst writes a judgment.

Intelligence services spent decades on one problem: how to state a judgment so the reader knows how far to trust it. The U.S. standard is public. It is Intelligence Community Directive 203. Borrowing its method implies no government affiliation. It simply works on a Monday meeting too.

SAY THE ODDS · SEVEN WORDS, SEVEN RANGES

IF YOU SAY	YOU MEAN
Almost no chance	1 to 5%
Very unlikely	5 to 20%
Unlikely	20 to 45%
Roughly even chance	45 to 55%
Likely	55 to 80%
Very likely	80 to 95%
Almost certainly	95 to 99%

The scale stops at 99 percent. The directive also tells analysts not to put a confidence level and a likelihood word in the same sentence.

GRADE EVERY SOURCE · TWO JUDGMENTS, NOT ONE

SOURCE RELIABILITY	
A	Completely reliable
B	Usually reliable
C	Fairly reliable
D	Not usually reliable
E	Unreliable
F	Cannot be judged

INFORMATION CREDIBILITY	
1	Confirmed
2	Probably true
3	Possibly true
4	Doubtful
5	Improbable
6	Cannot be judged

The two grades are independent. A trusted manager can still pass on a rumor. That is a B4. This grid is the NATO scale. It is not part of ICD 203.

DRAW ON FOUR KINDS OF EVIDENCE

Research
Published studies, graded like any other source

Your records
Jobs, calls, sales, schedules, books

Experienced people
Managers and crews, asked separately

People affected
Customers, staff and partners

HOW IT READS IN A BUSINESS

A1
Sales export, reconciled to the bank deposit

B3
A manager's memory of last month

C4
A vendor's own case study

F6
An anonymous online comment

FIVE PROMPTS THAT MAKE AI A BETTER ANALYST

01	**Separate**	Label every line of your answer with its claim state: observed, inferred, predicted, attested or verified. Use only the records I gave you.
02	**Grade**	Grade each source A to F and each claim 1 to 6. Give one line of reason for each grade.
03	**Odds**	State the chance this works as a number from 1 to 99. Say what would move it ten points.
04	**Rival**	Make the strongest case for the option we are leaning against. What evidence would make it win?
05	**Premortem**	It is six months from now and this decision failed. Write the three most likely reasons.

COPY + PASTE STARTER

I lead a company with [number] locations. Here are the records behind one open decision: [paste]. Use only these records and do not invent missing facts. Return: (1) the decision as one question with a real choice, (2) each source with a grade, (3) what the records prove and what they only suggest, (4) the chance it works, as a number, with the reason, (5) what to measure, its baseline and a check date, (6) anything that needs a person to decide.

Do not paste customer personal data or employee records into an AI service until you understand its privacy and retention terms.

11 · VERIFICATION, VERDICTS AND RECEIPTS

Verified means something. Here is exactly what.

Most systems close a decision when someone marks it done. This standard closes it when the evidence clears seven bars, someone other than the owner checks it, and the verdict is recorded, whatever it is.

A RESULT COUNTS AS VERIFIED ONLY WHEN

1. The baseline was written before acting
2. The result window was set in advance
3. The metric is named
4. Two independent sources agree
5. Known alternative explanations are written down
6. A reviewer who is not the owner checks it
7. The owner signs off

FOUR VERDICTS, ALL KEPT

Verified gain
The measure moved the right way, beyond what the comparison explains.

Verified, no gain
Checked properly, and it did not work. Still a result.

Inconclusive
Not enough evidence either way. Every review starts here.

Disputed
The evidence conflicts, or someone with standing challenges it.

WHAT A SEALED RECEIPT CARRIES · OPEN FORMAT, VERSION 1



Seal intact: the signature and the chain check out.

IDENTITY	Receipt ID, signing key and issue date	CHAIN	The fingerprint of the receipt before it
VERDICT	One of the four, nothing else	MEASUREMENT	Metric, baseline and result, each with its window
EVIDENCE	Fingerprints of at least two sources	CONTEXT	The system involved and other changes in the window
PEOPLE	Independent reviewer, owner sign-off, owner's role	PUBLICATION	None, numbers only, anonymized or named

HOW ANYONE CAN CHECK IT, WITHOUT TRUSTING NORTH STAR

Put the fields in a fixed order and take the SHA-256 fingerprint. Check the Ed25519 signature against North Star's published key, then match the fingerprint of the receipt before it. In 2025 the W3C made Ed25519 signatures part of its Recommendation for verifiable credentials. A North Star receipt uses the same signature method. It is not a W3C credential. Try it at northstarintel.ai/verify. The format is at northstarintel.ai/receipt-spec. The keys are at northstarintel.ai/api/public/receipt-keys.

A seal proves the record has not changed. It does not prove the action caused the outcome. The comparison does. That is why move 6 exists.

THE QUESTION TO ASK ANY DECISION SOFTWARE VENDOR

"Show me one decision your system carried from question to verified outcome: the owner, the date, the graded evidence, the baseline, the verdict, and a receipt I can check without trusting you." A vendor who can do that in one sitting is worth your time. One who opens a dashboard is showing you a picture of the problem.

12 · PROVE IT WITH WHAT YOU ALREADY HAVE

Your locations are your evidence.

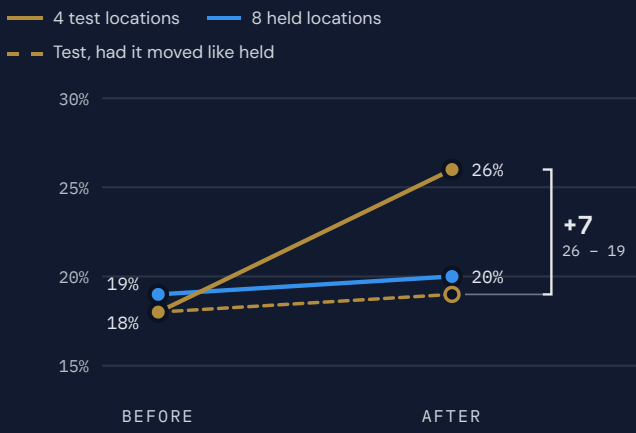
A single-site business can only compare itself with last year. An operator with several locations can compare sites in the same weeks, under the same weather, prices and economy. Multi-unit operators run 56.5 percent of U.S. franchised locations (FRANdata, 2025). Three tools turn that structure into proof.

TOOL 1 · TEST AND HOLD

Change some locations. Hold the rest, or stagger the rollout so every wave has a comparison. Compare the change in each group, not the ending numbers.

REBOOK RATE · ILLUSTRATIVE	4 TEST	8 HELD
8 weeks before the membership	18%	19%
8 weeks after	26%	20%
Change	+8 points	+1 point

Effect of the membership: +7 points. Without the held locations, the team would have claimed 8, and one of them would have arrived anyway.



- ✓ Pick the test locations before you see any results
- ✓ Hold at least as many locations as you test
- ✓ Write the measure, the window and the check date first
- ✓ Name anything else that changed in the window

TOOL 2 · THE NOISE CHECK

Give the same three real cases to five managers, separately: a refund request, a staffing call and a price exception. Compare the answers.

Refund for the same late-arrival complaint · five managers · ILLUSTRATIVE



Answers: $0, $50, $75, $100, $150

```
Spread = (highest − lowest) ÷ average
= 150 ÷ 75 = 200%
```

If two managers land further apart than you would accept from two vendor quotes, the problem is not one location. It is a missing rule.

TOOL 3 · THE CALIBRATION LOG

Before each decision, write the chance it works. On the check date, write 1 if it did and 0 if it did not. Score each call as (chance − result)², then average the scores. That is a Brier score. Lower is better.



The score to beat is 0.25, what you get by saying 50 percent every time. Twenty scored calls can tell you more about a manager's judgment than an interview.

FRANCHISE SYSTEMS ALREADY KNOW THIS STANDARD

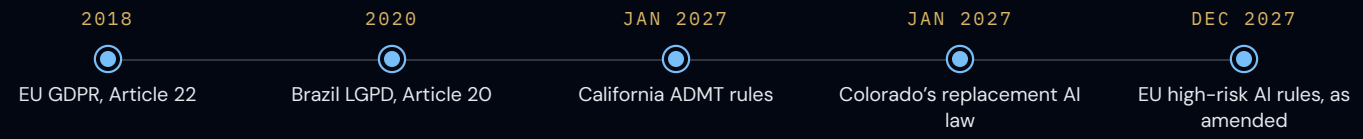
Under the FTC Franchise Rule, a franchisor that makes an earnings claim to a prospective franchisee must have a reasonable basis and written substantiation for it at the time the claim is made. Hold your internal decisions to the same bar the law sets for earnings claims.

13 · BEFORE AI MAKES OR SHAPES A DECISION

This is the page most decision guides leave out.

It describes general principles and is not legal advice. Rules vary by place and change often. Confirm your obligations with qualified counsel before a system makes or shapes decisions about people.

WHEN THE RULES APPLY · CHECK EACH ONE AGAINST ITS PRIMARY SOURCE



1 Decisions about people carry rights

The EU's GDPR (Article 22) and Brazil's LGPD (Article 20) give people rights around decisions made about them solely by automated processing. Since 2019, the LGPD no longer requires that a person do the review. California's rules on automated decision-making technology require notice, and in many cases a way to opt out, for significant decisions from January 1, 2027. Hiring, scheduling, pay and discipline are where operators meet these rules first.

2 The rules move in both directions

In May 2026, Colorado replaced its first AI law with a narrower, notice-based version weeks before the original was due to start. The new law takes effect January 1, 2027. In July 2026 the EU moved its high-risk AI rules for hiring and HR systems to December 2, 2027. Treat every rule claim you read, including these, as a pointer to a primary source.

3 Records are the common thread

Wherever you operate, the defensible posture looks the same. Know which decisions a system makes or shapes. Keep the input, the output and the person who approved it. Give people a way to ask for a review, and make sure a person actually does it. Once its high-risk rules apply, the EU's AI Act requires deployers of those systems to keep the logs they generate for at least six months.

FIVE QUESTIONS BEFORE AI TOUCHES A DECISION

CHECK	THE QUESTION	WHERE IT LIVES
Boundary	Which decisions may a system make alone, and which need a named person?	Decision Record, action stage
Record	Are the input, the output and the approver written down for every decision?	Decision and action receipts
Explanation	Could you tell the affected person why, in plain words?	Graded evidence and confidence
Review	Can someone ask for a human review, and does a person actually do it?	A named reviewer, not the owner
Drift	Who checks, on a set date, whether the system's decisions still hold?	Outcome stage: check date and verdict

Question 2 is the one that fails silently. A system recommends, a manager clicks approve, and nothing records who saw what. The cheapest protection is a written record of every approval. In North Star, AI can suggest a rating and wording. A person confirms every rating, and the saved score always comes from the published formula. Nothing is approved or verified by an AI.

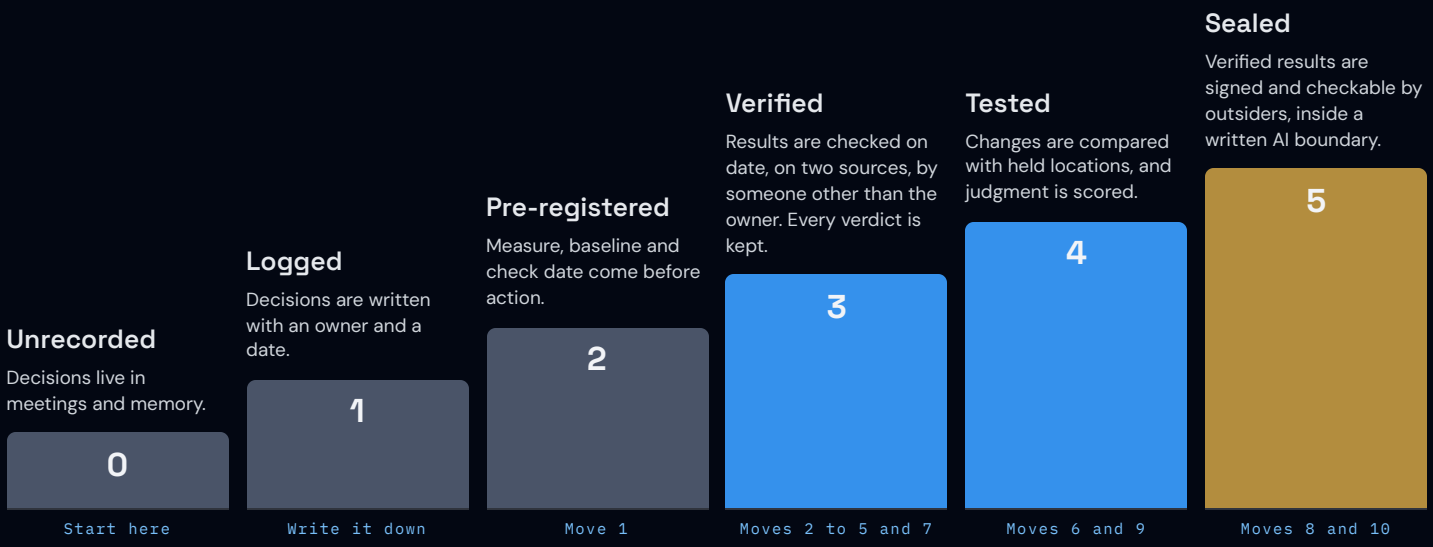
14 · THE SCOREBOARD AND THE SIX LEVELS

Eight numbers that show whether your decisions land.

Revenue tells you how the business did. These tell you how the decisions did. Each one comes straight from your Decision Records and matches a line of the audit on page 19.

#	METRIC	HOW TO COUNT IT	AUDIT LINE
1	**Ownership**	Decisions with one named person as owner, divided by all decisions	5
2	**Pre-registration**	Decisions with measure, baseline and window written before acting, divided by all decisions	7
3	**Decision latency**	Days from the first signal to a named owner and date	6
4	**Proof of action**	Actions with a record that they happened, divided by approved actions	8
5	**On-date checks**	Results checked on the date set, divided by results due	9
6	**Verification**	Outcomes verified on two sources, divided by outcomes checked	10
7	**Honest verdicts**	Share of no gain, inconclusive and disputed among all verdicts. Zero is a warning, not a win.	11
8	**Calibration**	Average Brier score of your chance estimates. Lower is better.	3

SIX LEVELS · WHERE YOU ARE, AND THE MOVES THAT GET YOU THERE



Audit bands: 0 to 11 is Level 0 or 1. 12 to 19 is Level 1 or 2. 20 to 24 is Level 3. Levels 4 and 5 add the practices on pages 11 to 13.

15 · THE SEVEN DECISION LEAKS

Where decisions escape between the meeting and the result.

A decision leak is a choice your team made, or meant to make, that never produced a result anyone checked. The seven tells below live in your notes, inboxes and job systems. Read your records, not a benchmark.

#	LEAK	THE TELL IN YOUR RECORDS
1	**Unwritten decision**	"We agreed" in a meeting, and no record says what, who or by when
2	**Owned by a team**	The owner field says ops, the managers or everyone
3	**Open-ended date**	Next quarter, soon, after the busy season
4	**Loudest-location evidence**	One manager's story became the rule for every location
5	**No baseline**	The result was judged after the change, against memory
6	**All-at-once rollout**	Every location changed the same week, so nothing is left to compare
7	**Closed, not verified**	The status says done, no result is attached and no lesson is written

You may recognize all seven. That is not the point. The point is which one costs you the most, and you cannot know that until you price it on page 17.

WHERE EACH LEAK SITS IN THE LOOP, AND WHICH AUDIT LINES CATCH IT

SIGNAL	SCORE	ASSIGN	EXECUTE	VERIFY	LEARN
LEAKS 1	LEAKS 4	LEAKS 2 3	LEAKS 5 6	LEAKS 7	LEAKS 7
AUDIT LINES 1 · 2	AUDIT LINES 3 · 4	AUDIT LINES 5 · 6	AUDIT LINES 7 · 8	AUDIT LINES 9 · 10	AUDIT LINES 11 · 12

WHAT WE REFUSE TO CITE

One of the most repeated numbers in this market says data-driven companies are 23 times more likely to win customers. We traced it to a 2013 consulting survey of about 400 managers on customer analytics. Retellings changed the subject, the source and one of the numbers. Market sizing is no better. Eight research firms put the 2025 decision intelligence market anywhere from $12.8 billion to $21.2 billion. Neither figure tells you whether your last decision worked. Your records can.

16 · FOUR FINDINGS THAT SURVIVE

We graded the research the way we ask you to grade your records.

Each finding carries a source grade from page 10 and a plain note on its limits. We use them for direction, not for decimals.

PROCESS BEATS ANALYSIS

Lovallo and Sibony · 2010

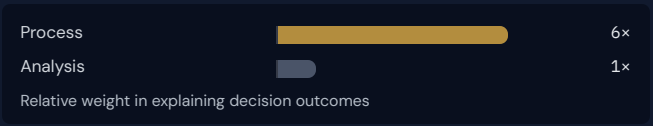
B2
GRADE

METHOD

Managers reported on 1,048 major decisions made over five years. A regression controlled for industry, geography and company size.

FINDING

How the decision was made mattered six times more than how much analysis went into it.



LIMIT

Self-reported by managers. Published by a consulting firm, not peer reviewed.

WHAT WE TAKE

Fix how decisions get made before you buy more analysis.

JUDGMENT IS NOISY

Kahneman, Sibony and Sunstein · 2021

B2
GRADE

METHOD

Forty-eight underwriters at one insurer priced the same realistic cases. Executives first predicted how far apart two underwriters would land.

FINDING

Executives expected about 10 percent. The typical gap between two underwriters was 55 percent of the average premium.



LIMIT

One company and one audit. The same pattern appears across many professions.

WHAT WE TAKE

Before you blame a location, check whether your managers would make the same call on the same case.

WORDS ARE NOT ODDS

Kent · 1964 · Mauboussin and Mauboussin · 2018

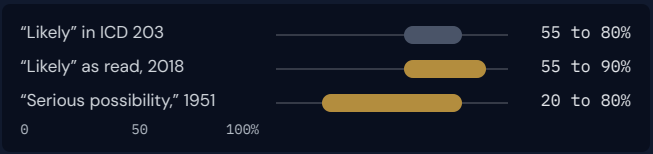
B1
GRADE

METHOD

Kent asked the authors of a 1951 estimate what “serious possibility” meant. In 2018, 1,700 people put numbers on common probability words.

FINDING

Kent’s colleagues meant anything from 20 to 80 percent. In 2018, the middle 90 percent of answers for “likely” ran from 55 to 90 percent.



LIMIT

These studies measure how words are read, not how decisions turn out.

WHAT WE TAKE

Write confidence as a number.

JUDGMENT CAN BE TRAINED

Mellers and colleagues · 2014 · Chang and colleagues · 2016

A2
GRADE

METHOD

Randomized experiments inside a four-year forecasting tournament sponsored by the U.S. intelligence community.

FINDING

Probability training, teamwork and tracking each improved accuracy. Training that took under an hour improved Brier scores by 6 to 11 percent over a control group.



LIMIT

World-events questions answered by volunteers, not business decisions made by executives.

WHAT WE TAKE

Keep score. A calibration log is the cheapest training you will ever buy.

HOW TO READ A GRADE

The letter grades the source and the number grades the information, on the scale from page 10. A2 means a completely reliable source and information that is probably true. These grades are our judgment, and page 22 gives the reason for each one.

17 · PUT A NUMBER ON IT

A decision without a number is an opinion.

A decision with a number and a date is a commitment. Two numbers settle most arguments in a leadership meeting: what each week of waiting costs, and where the decision ranks against everything else that is open.

THE THREE-MINUTE VERSION · FROM MEMORY

What is this decision worth each month if it works, after what it costs to run? $ _____

Multiply that by 12, then divide by 52. Each week of waiting costs about $ _____

How likely is it to work, as a number? % _____

THE FULL VERSION · RUN IT TWICE, REPORT THE RANGE

INPUT	CONSERVATIVE	REALISTIC
A **Monthly gain if it works**	$ _____	$ _____
B **Monthly cost to run it**	$ _____	$ _____
C **Chance it works**	% _____	% _____
D **Hours to set it up**	hrs _____	hrs _____

Net monthly value
`A − B`

Cost of each week you wait, if it works
`(A − B) × 12 ÷ 52`

Value per setup hour, risk-adjusted
`(A − B) × C ÷ D`

THEN RANK IT · THE PRIORITY RULE PUBLISHED ON NORTHSTARINTEL.AI

```
Priority = ((2 × Opportunity) + Risk of waiting − Effort) ÷ 8 × 100
```

Rate each from 1 (low) to 3 (high). Opportunity 3, risk of waiting 3 and effort 2 give ((6 + 3 − 2) ÷ 8) × 100 = 87.5, shown as 88.

IF YOU ARE GUESSING, START HERE · THEN REPLACE THE GUESSES WITHIN A MONTH

INPUT	STARTING POINT
C · chance it works	50% until you have a base rate from your own locations
A · monthly gain	The smallest number you would defend to your finance lead
D · hours to set up	Count every manager's hours, not only the vendor's

These starting points are conventions, not research, and we are telling you so. Fifty percent is the honest default for a decision you have never checked before. If the weekly cost of waiting is smaller than the cost of the meeting about it, park the decision and write down why. That is a legitimate answer. If it is several times larger, you have found this month's priority.

Deciding which services to keep or reprice? Which Jobs Pay, our free tool, works out margin per paid hour from your own job export, inside your browser. It is at northstarintel.ai/which-jobs-pay.

18 · WHAT THIS LOOKS LIKE IN A REAL LEADERSHIP TEAM

Illustrative. Every number is invented and every assumption is stated.

A home services operator with twelve locations believes its leadership team is decisive. The CEO pulls the last 20 leadership decisions from meeting notes and email instead of trusting that belief.

20	14	4	1
decisions found in the records	named one owner	had a baseline before acting	was checked on its date

AUDIT RESULT

5 / 24 · 21 / 100

RUNNING ON MEMORY · LEVEL 1

STAGE PROFILE · OUT OF 4

SIGNAL	■■□□	2/4
SCORE	□□□□	0/4
ASSIGN	■■□□	2/4
EXECUTE	■□□□	1/4
VERIFY	□□□□	0/4
LEARN	□□□□	0/4

THE WIN THAT WAS NEVER PROVEN

The biggest decision of the year was an 8 percent price increase at all twelve locations on April 1. Second-quarter revenue rose 6 percent, and leadership called it a win.

SECOND QUARTER, YEAR OVER YEAR	LAST YEAR · NO PRICE CHANGE	THIS YEAR · PRICES UP 8%
Revenue	+7%	+6%
Completed jobs	+2%	−2%

Revenue had already grown 7 percent the year before, with no price change. This year completed jobs fell 2 percent where they had grown 2 percent. Did the increase add revenue, or trade jobs for price while the market did the rest? Nobody wrote a baseline, held a comparison location or set a check date. The only honest verdict is inconclusive.

THE ONE-PAGE REGISTER · WHAT MONDAY LOOKS LIKE NOW · EXCERPT

DECISION	OWNER	NEXT ACTION DUE	MEASURE · BASELINE	CHECK
Test a maintenance membership at 4 of 12 locations	Operations manager	Week 1	Rebook rate · 18%	Week 9
Add Saturday dispatch at 3 locations	District manager	Week 2	Saturday jobs · 0	Week 6
Reprice the five lowest-margin job types	Finance lead	Week 1	Margin per paid hour · $48	Week 13

The fix was not software.

Every number the team needed already sat in its job system and its books. What was missing was a written baseline, a comparison and a date. The first fix was a 15-minute Monday review of a one-page register: open decisions, owners, dates and the number each one should move. Software came second, when the register passed 40 open decisions across twelve locations.

The next big decision, a maintenance membership, was tested at four locations with eight held back. Page 12 shows how that one was proven.

ILLUSTRATIVE Illustrative composite. Not a client, not a result and not a benchmark.

19 · THE ONE-HOUR EVIDENCE AUDIT · THE CONFORMANCE TEST

Twelve checks. Score out of 24, then out of 100.

Pull the last 20 decisions your leadership team made, at any location. Last, not best. A decision is any choice that committed money, people or time. Use records, not memory: notes, email, chat, tickets. If you cannot find twenty on record, you have found the first leak. Count each missing record as a no on every line and keep going.

2 = true for 18 or more of the 20. 1 = true for 6 to 17. 0 = true for 5 or fewer. Index /100 = score × 100 ÷ 24.

#	STAGE	CHECK	2	1	0
1	SIGNAL	Was the decision written as one question with a real choice?	○	○	○
2		Is the evidence behind it on file, with each source named?	○	○	○
3	SCORE	Was the chance of success written as a number before acting?	○	○	○
4		Was it ranked against the other open decisions by a written rule?	○	○	○
5	ASSIGN	Does it name one owner, a person and not a team?	○	○	○
6		Does it have a due date and a written next action?	○	○	○
7	EXECUTE	Were the measure and its baseline written before anyone acted?	○	○	○
8		Is there a record that the action happened, not only that it was approved?	○	○	○
9	VERIFY	Was the result checked on the date that was set?	○	○	○
10		Do two independent records support the verdict?	○	○	○
11	LEARN	Was a verdict recorded: gain, no gain, inconclusive or disputed?	○	○	○
12		Was the lesson written where the next decision will see it?	○	○	○

YOUR SCORE

_____ / 24 _____ / 100

WEAKEST STAGE _____

STAGE PROFILE · EACH STAGE OUT OF 4

SIGNAL	__ / 4	SCORE	__ / 4
ASSIGN	__ / 4	EXECUTE	__ / 4
VERIFY	__ / 4	LEARN	__ / 4

READING THE NUMBER · WORKING BANDS, NOT INDUSTRY BENCHMARKS

20 to 24 · 83 to 100

Defensible

LEVEL 3

Decisions carry owners, numbers and receipts. Ready to automate intake and the Monday review.

12 to 19 · 50 to 79

Decided, not proven

LEVEL 1 TO 2

Work happens. Proof does not. Fix lines 7, 9 and 10: baseline, check date, second record.

0 to 11 · 0 to 46

Running on memory

LEVEL 0 TO 1

Do not buy software yet. Get one decision a week written with an owner, a date and a number.

20 · WHAT STILL MATTERS AS AI DECIDES MORE

No predictions of our own. Four of theirs.

A leading technology analyst firm published all four between June 2025 and June 2026. All four can be true. The useful question is which parts of deciding get easier, and which get harder.

50% of business decisions augmented or automated by AI agents by 2027

15%+ of day-to-day work decisions made autonomously through agentic AI by 2028, up from zero in 2024

40%+ of agentic AI projects canceled by the end of 2027

5× more trust in explicitly modeled decisions than in ungoverned ones by 2029, and 80 percent faster

PREDICTIONS, NOT MEASUREMENTS · SOURCES ON PAGE 22

RUNG	WHAT THE SYSTEM DOES	WHAT YOU DO
1. Inform	Summarizes, flags and drafts	You decide everything
2. Recommend	Ranks options with reasons	You choose and own the choice
3. Decide inside limits	Acts within a written boundary	You audit a sample every week
4. Act and report	Executes and leaves receipts	You handle the exceptions
5. Answer for it	Nothing. This rung is not for systems	You answer for every rung below

A team can sit on rung 1 and believe it is on rung 3. Every rung needs the Decision Record on page 9, never less. A system that acts without supervision needs a sharper human boundary than one that only drafts.

THREE THINGS GET SCARCER

Records

When any AI can write a persuasive memo in seconds, memos prove nothing. A dated record of what was decided, by whom and on what evidence keeps its value.

Calibration

When everyone rents the same models, the edge is knowing how often your own judgment is right. That number only comes from a log.

Accountability

A system can rank options. It cannot be answerable. The name on the decision is the product.

THE FUTURE-PROOF TEST · USE IT ON MONDAY

1
If AI got ten times better overnight, would your decision process get stronger or disappear?

2
If AI stopped improving tomorrow, was the process still worth building?

3
When a decision fails, who answers, and can they show what was known at the time?

A process that survives all three is an investment. One that needs the technology to keep improving is a bet. If capability stalls tomorrow, every requirement in this standard still holds.

21 · FROM GUT CALLS TO GRADED CALLS IN ONE WEEK

Seven days. Three numbers your leadership team may never have written down.

DAY	MOVE	DO THIS
1	**Pull the records**	The last 20 leadership decisions, from notes, email and chat. Last, not best.
2	**Score them**	The twelve checks on page 19. Your number out of 24 and out of 100.
3	**Find the weak stage**	The stage profile. Start where the score is lowest.
4	**Write one record**	A Decision Record for the open decision costing you most.
5	**Price it**	Page 17, twice: conservative and realistic. Report the range.
6	**Run a noise check**	Three cases, five managers, answers written separately.
7	**Hold the review**	Fifteen minutes. The register, the owners, the dates, the numbers.

By Sunday you have three numbers: your decision score, your weakest stage score and the weekly cost of waiting on your costliest decision. Then make one move from pages 7 and 8 each month, and aim to climb one level each quarter.

YOUR NEXT STEP

You do not need another dashboard. You need decisions that land.

Request a decision review. A decision review is 20 minutes on one decision that keeps stalling. You leave with one owner, one date and a way to check the result against your own numbers.

BRING THREE THINGS

- ✓ Your audit score from page 19
- ✓ The records behind your costliest open decision
- ✓ The person who will own it

No records yet? Spend two weeks writing every decision down, then come back.

See what a finished report looks like: northstarintel.ai/decision-to-evidence-report

northstarintel.ai/contact · (435) 265-3483 · admin@northstarintel.ai

REQUEST

WHY THIS STANDARD READS DIFFERENTLY

North Star Intel is not a dashboard company. The method is evidence first: write the decision, grade what it rests on, name one owner and one date, then check the result against a baseline written before anyone acted. That is why page 18 ends with a Monday meeting, page 16 grades the research we cite, and move 9 says plainly what North Star has not built yet. A guide that only ever concludes you should buy the author's software is a brochure with footnotes.

22 · SOURCES AND GRADES

Every number in this standard, and where it came from.

Facts were checked against the sources below in November 2026. Grades use the scale on page 10 and are our judgment. Where a source sits behind a paywall, we relied on what its authors published in the open.

GRADED FINDINGS · PAGES 5 AND 16

B2 Lovallo, D. and Sibony, O. (2010). The case for behavioral strategy. McKinsey Quarterly.
B: established authors and a published method. 2: manager-reported and not independently replicated, though consistent with research on debiasing.

B2 Kahneman, D., Sibony, O. and Sunstein, C. R. (2021). Noise: A Flaw in Human Judgment. Little, Brown Spark. Also Kahneman, D., Rosenfield, A. M., Gandhi, L. and Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. Harvard Business Review.
B: established researchers and a described method. 2: one audit, consistent with noise found across many professions.

B1 Kent, S. (1964). Words of estimative probability. Studies in Intelligence, released 1993. Mauboussin, A. and Mauboussin, M. J. (2018). If you say something is "likely," how likely do people think it is? Harvard Business Review.
B: a first-hand account and a large published survey. 1: two independent observations, decades apart, agree.

A2 Mellers, B. and colleagues (2014). Psychological strategies for winning a geopolitical forecasting tournament. Psychological Science 25(5). Chang, W. and colleagues (2016). Developing expert political judgment. Judgment and Decision Making.
A: peer reviewed and randomized. 2: the follow-up came from the same research group, not an independent replication.

B2 McKinsey and Company (April 2019). Decision making in the age of urgency. McKinsey Global Survey of more than 1,200 executives.
B: a large published survey. 2: self-reported, by a firm that sells decision advice.

B2 Blenko, M. W., Mankins, M. C. and Rogers, P. (2013). Decision Insights: The five steps to better decisions. Bain and Company. Also Decide and Deliver (2010).
B: established authors and a long research program. 2: survey-based and reported at a confidence level, not as an effect size.

STANDARDS, METHODS AND FACTS

- Office of the Director of National Intelligence. Intelligence Community Directive 203, Analytic Standards. First issued 2007, reissued 2015, amended June 2023. Source reliability and information credibility scale (A to F, 1 to 6) as used in NATO intelligence doctrine, often called the Admiralty Code.
- Barends, E., Rousseau, D. M. and Briner, R. B. (2014). Evidence-Based Management: The Basic Principles. Center for Evidence-Based Management. Nosek, B. A. and colleagues (2018). The preregistration revolution. PNAS 115(11). Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin 86(3).
- Klein, G. (2007). Performing a project premortem. Harvard Business Review. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review 78(1). Hussey, M. A. and Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 28(2).
- W3C (May 15, 2025). Verifiable Credentials Data Model 2.0 and Data Integrity EdDSA Cryptosuites, W3C Recommendations. North Star open receipt format, version 1, northstarintel.ai/receipt-spec.
- U.S. Federal Trade Commission. Franchise Rule, 16 CFR 436.9(c) and Item 19, 16 CFR 436.5(s). International Franchise Association and FRANdata (2026). Franchising Economic Outlook. FRANdata (2025): 42,571 multi-unit operators run 243,278 franchised units, 56.5 percent of the total.
- Gartner press releases at gartner.com/en/newsroom: "Gartner Data & Analytics Summit 2025 Orlando: Day 1 Highlights," March 3, 2025. "Gartner Announces the Top Data & Analytics Predictions," June 17, 2025. "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," June 25, 2025. "Gartner Identifies the Top Trends for Data and Analytics," June 16, 2026.

RULES POINTED TO ON PAGES 8 AND 13

- EU General Data Protection Regulation, Article 22. Brazil, Law 13.709/2018 (LGPD), Article 20, as amended by Law 13.853/2019. California Privacy Protection Agency, regulations on automated decision-making technology, approved September 2025, compliance from January 1, 2027.
- Colorado SB 26-189, signed May 14, 2026, effective January 1, 2027, replacing SB 24-205. Regulation (EU) 2026/1744, the Digital Omnibus on AI, in force July 27, 2026. Regulation (EU) 2024/1689, the AI Act, Articles 12 and 26(6).

WHAT WE REFUSE TO CITE, TRACED · PAGES 5 AND 15

- McKinsey and Company, Five facts: How customer analytics boosts corporate performance (DataMatics 2013 survey of more than 400 managers). The original compares intensive and light users of customer analytics: 23 times more likely to outperform on new-customer acquisition and 9 times on loyalty.
- 2025 market-size estimates for decision intelligence: Global Market Insights $12.8B, 360iResearch $14.2B, The Business Research Company $17.5B, Bizwit Research $17.6B, SNS Insider $18.1B, Cervicorn Consulting $18.8B, Fortune Business Insights $19.4B, Straits Research $21.2B.
- A "95 percent correlation" between decision effectiveness and financial results, a retelling of a link Bain reported at a minimum 95 percent confidence level.

23 · ABOUT THIS STANDARD AND TERMS OF USE

How this standard is versioned, cited and used.

A standard is only useful if people can tell which version they are reading and what it promises. This page says both.

VERSION AND EDITION

Standard	Decision to Evidence Standard, version 1.0
Edition	2.2, November 16, 2026. Replaces Editions 2.1 and 2.0, and Edition 1.0, published as The Evidence Audit.
New in 2.2	Adds the whole standard on one page, as page 24. The EU date is now law, and forecasts and citations are updated. Four broad claims are narrowed, and two web addresses are added. Level labels match the ten moves, and section numbers match page numbers. The twelve checks are unchanged.
Earlier editions	2.1 matched the current release and made the contents clickable. 2.0 added the standard and its ten moves.
Next review	Within twelve months, or sooner when a rule or a cited study changes

HOW TO CITE IT

North Star Intel (2026). Decision to Evidence: The Standard, Edition 2.2. North Star Intelligence LLC. northstarintel.ai/decision-to-evidence-standard

WHAT NORTH STAR HAS BUILT, AND WHAT IT HAS NOT

Every “In North Star today” line on pages 7 and 8 describes how the current release works. Calibration is not part of it, and move 9 says so. No client record appears in this standard, and no example in it is a client.

TERMS OF USE

Educational only. This standard is not legal, financial, tax, investment, employment or compliance advice. Talk to qualified professionals before you act on it.

No warranty. Provided as is and current as of November 16, 2026. Laws, studies and facts change, and we may update this standard without notice.

Limitation of liability. To the fullest extent the law allows, North Star Intelligence LLC is not liable for any loss arising from the use of this standard or from decisions made with it.

Illustrative examples. Examples are composites, not clients, results, benchmarks, earnings claims or guarantees. Starting points on page 17 are conventions, not research.

Third parties. Studies, standards, agencies and firms are cited for reference only. Their mention implies no affiliation, sponsorship or endorsement, and their names and marks belong to their owners. North Star Intel has no government affiliation.

License. You may read, print and use this standard inside your own organization and apply it to your own decisions. Without written permission, you may not resell, republish, translate, rebrand or redistribute it, in whole or in part, or use it to train AI models. Short quotations with credit to North Star Intel and a link to northstarintel.ai are welcome. This file blocks copying, so retype them.

Trademarks. North Star Intel™ and Decision to Evidence™ are trademarks of North Star Intelligence LLC.

Edition 2.2 · November 16, 2026 · Document NSI-D2E-2.2-EN · © 2026 North Star Intelligence LLC, a Utah limited liability company founded in North Logan, Utah. All rights reserved. northstarintel.ai · admin@northstarintel.ai · (435) 265-3483

24 · THE STANDARD ON ONE PAGE

The whole standard on one page.

Print it and keep it where decisions get made. Each block names the page that explains it.

THREE IDEAS TO REMEMBER · PAGE 2

A decision is a record, not a meeting.

Evidence is graded, not collected.

Judge the process first, then the outcome.

TEN MOVES · PAGES 7 AND 8

1. ~~Logged after the fact~~ → **Written before the act**
2. ~~The team decided~~ → **Four named roles**
3. ~~Confident prose~~ → **Graded claims**
4. ~~Events on a dashboard~~ → **An evidence chain**
5. ~~Marked done~~ → **Verified**
6. ~~Before and after~~ → **A counterfactual**
7. ~~Wins only~~ → **Every verdict**
8. ~~Trust us~~ → **Check it yourself**
9. ~~Opinions about judgment~~ → **A calibration record**
10. ~~AI that acts~~ → **AI inside a written boundary**

VERIFIED ONLY WHEN · PAGE 11

1. The baseline was written before acting
2. The result window was set in advance
3. The metric is named
4. Two independent sources agree
5. Known alternative explanations are written down
6. A reviewer who is not the owner checks it
7. The owner signs off

EVERY RECORD CARRIES · PAGE 9

01 Question
02 Owner
03 Due date
04 Priority
05 Status

FOUR VERDICTS · PAGE 11

- Verified gain
- Verified, no gain
- Inconclusive
- Disputed

SIX LEVELS · PAGE 14

0 Unrecorded
1 Logged
2 Pre-registered
3 Verified
4 Tested
5 Sealed

THE ONE-HOUR EVIDENCE AUDIT · PAGE 19

Twelve checks. Score out of 24, then out of 100.

SIGNAL > SCORE > ASSIGN > EXECUTE > VERIFY > LEARN

20 to 24 · 83 to 100
Defensible
LEVEL 3

12 to 19 · 50 to 79
Decided, not proven
LEVEL 1 TO 2

0 to 11 · 0 to 46
Running on memory
LEVEL 0 TO 1

YOUR FIRST SEVEN DAYS · PAGE 21

1. Pull the records
2. Score them
3. Find the weak stage
4. Write one record
5. Price it
6. Run a noise check
7. Hold the review

ALWAYS A NAMED PERSON · PAGE 9

> Before you start, write down what always goes to a named person: pay, hiring and firing, a refund above your limit, a safety issue, a legal question, and any decision about one specific employee or customer. Every time.

Then make one move from pages 7 and 8 each month, and aim to climb one level each quarter.