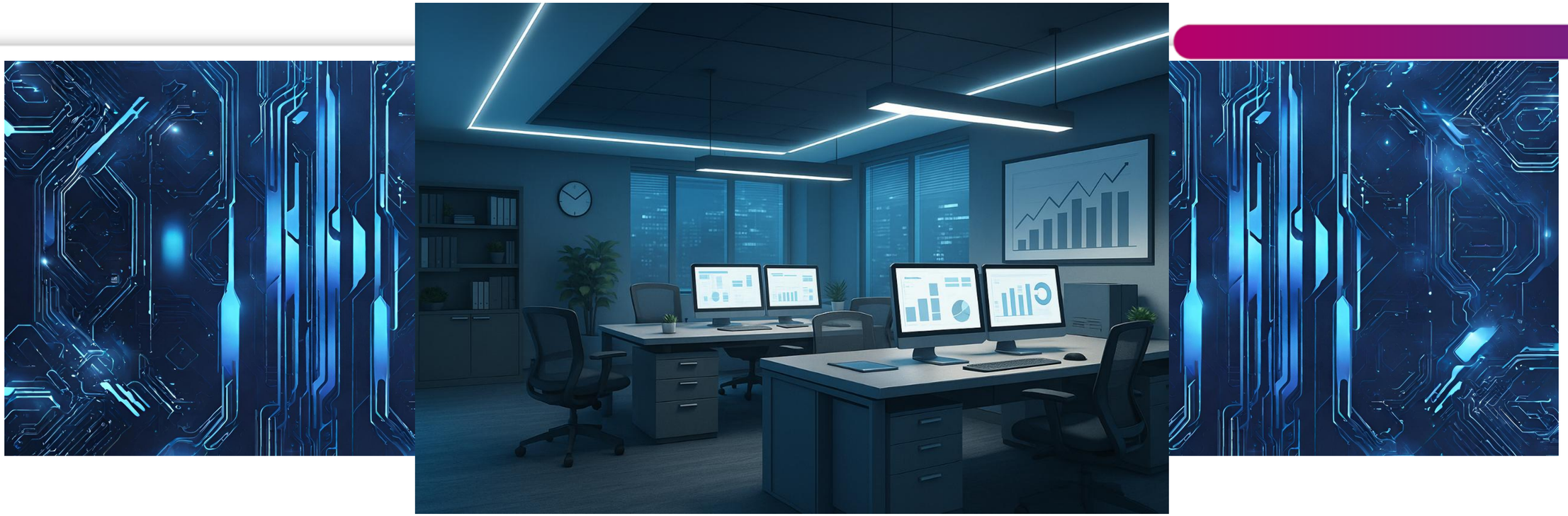
AFP FP&A FORUM

From Now to Next:
Roadmap to Your AI Future

Mariya Guttoh, FPAC
Director of FP&A and Treasury
PayJoy

AI Trend in Finance



If we don’t understand it, we ***won’t trust it****!*
An if we don’t trust it,
we won’t use it*!*

Why AI Fades in Finance?

AI doesn’t fail loudly. It fades when trust is not structurally built.

- ✓ AI Initiatives rarely fail loudly
- ✓ The model runs, but no one relies on it
- ✓ Trust erodes slowly
- ✓ AI fails not because it’s wrong, but because the environment isn’t ready
- ✓ In finance, AI success lives or dies on trust

Understanding → Trust → Usage → Scale

Myths vs Reality

AI is doing its job. Finance must do its job also.

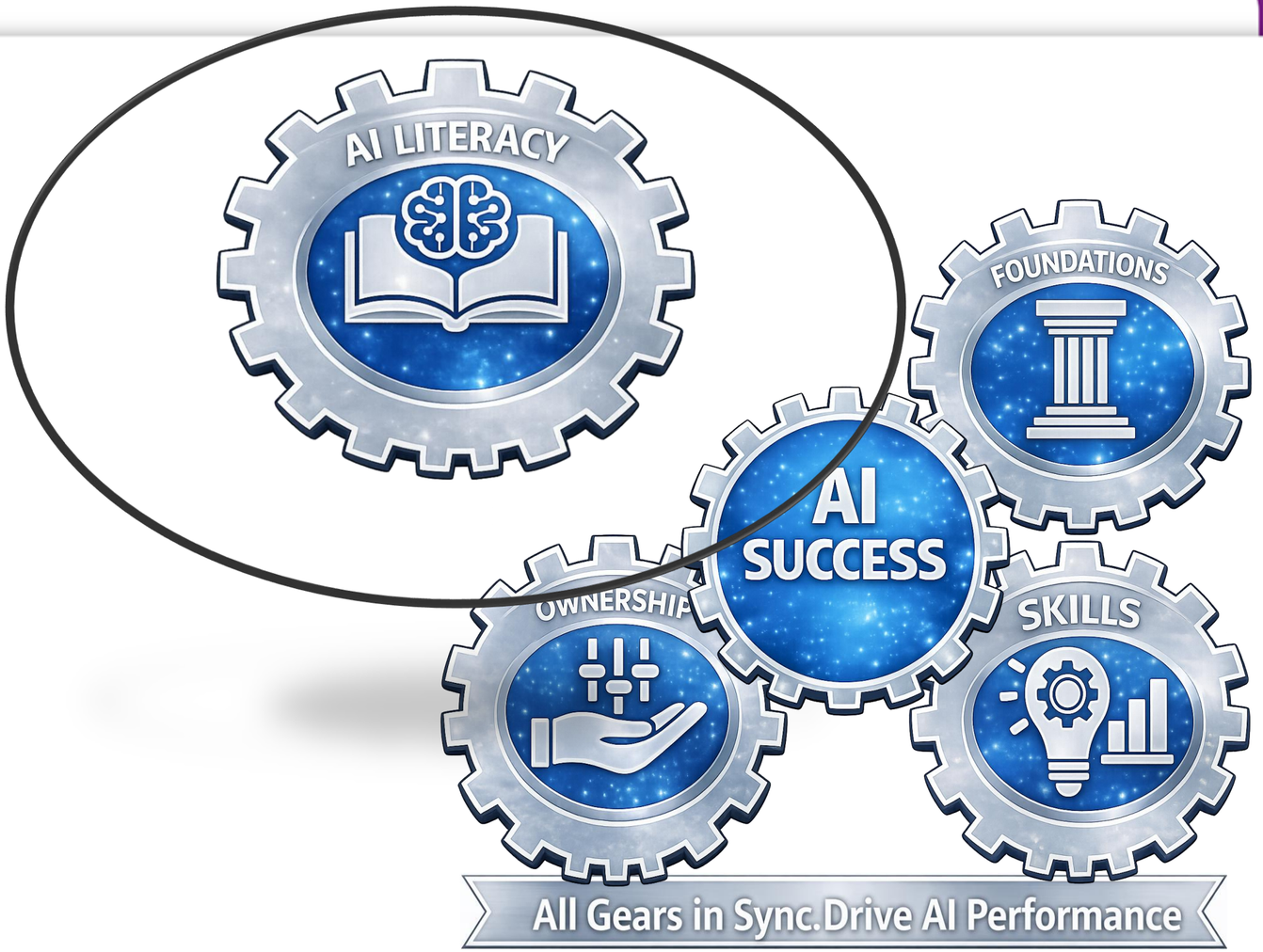
Myths	Reality
The tool isn't smart enough	Logic was never aligned
Our data is too messy	AI can handle mess, not ambiguity
This is a data science problem	It's a finance leadership problem
The model failed	It scaled our unclear assumptions
AI is a black box	It's only a black box when no one owns the inputs

AI is doing its job! It's us who have not fully done ours!

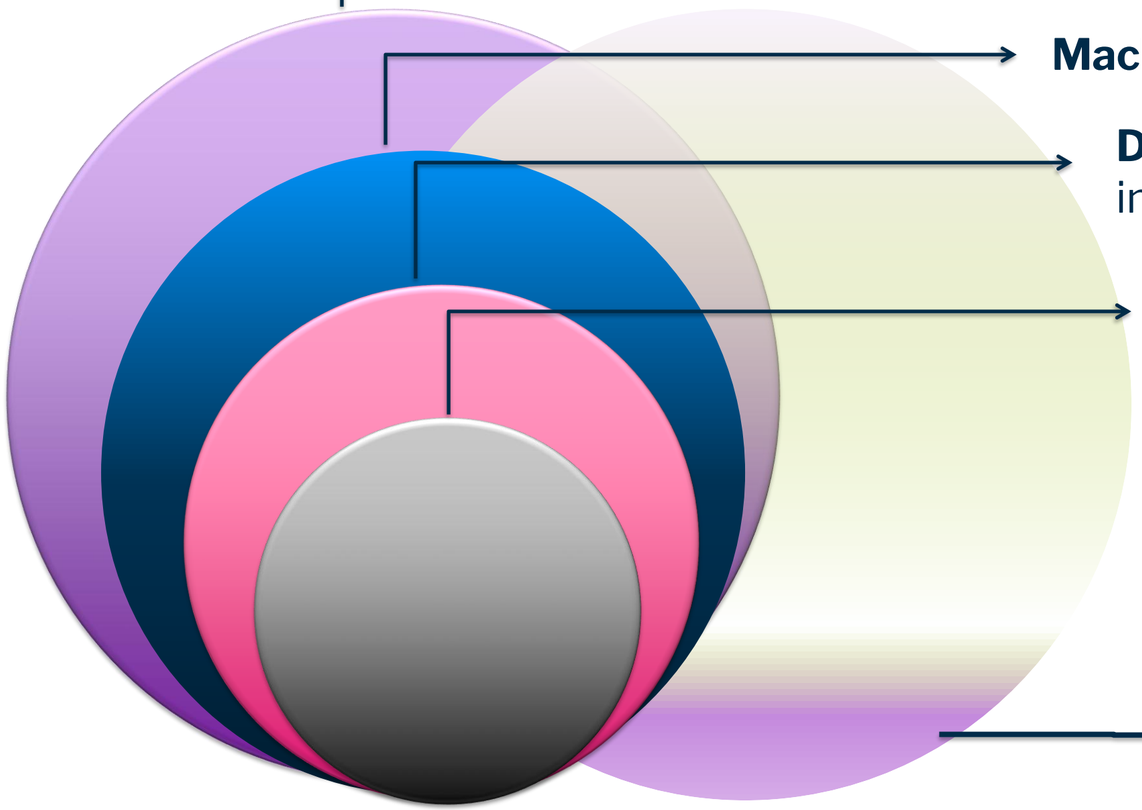
What AI Needs to Work?

AFP FP&A FORUM

AI Literacy



What is AI, Really?



Artificial Intelligence Machine that mimics human intelligence & behavior

Machine Learning Machine that learns from the data automatically

Deep Learning Subset of ML, learning happens via algorithms inspired by neural networks in human brain

Generative AI Using AI to create new content from existing ones, using advanced algorithms

Human Intervention

Human Judgment Converts model output into accountable decisions

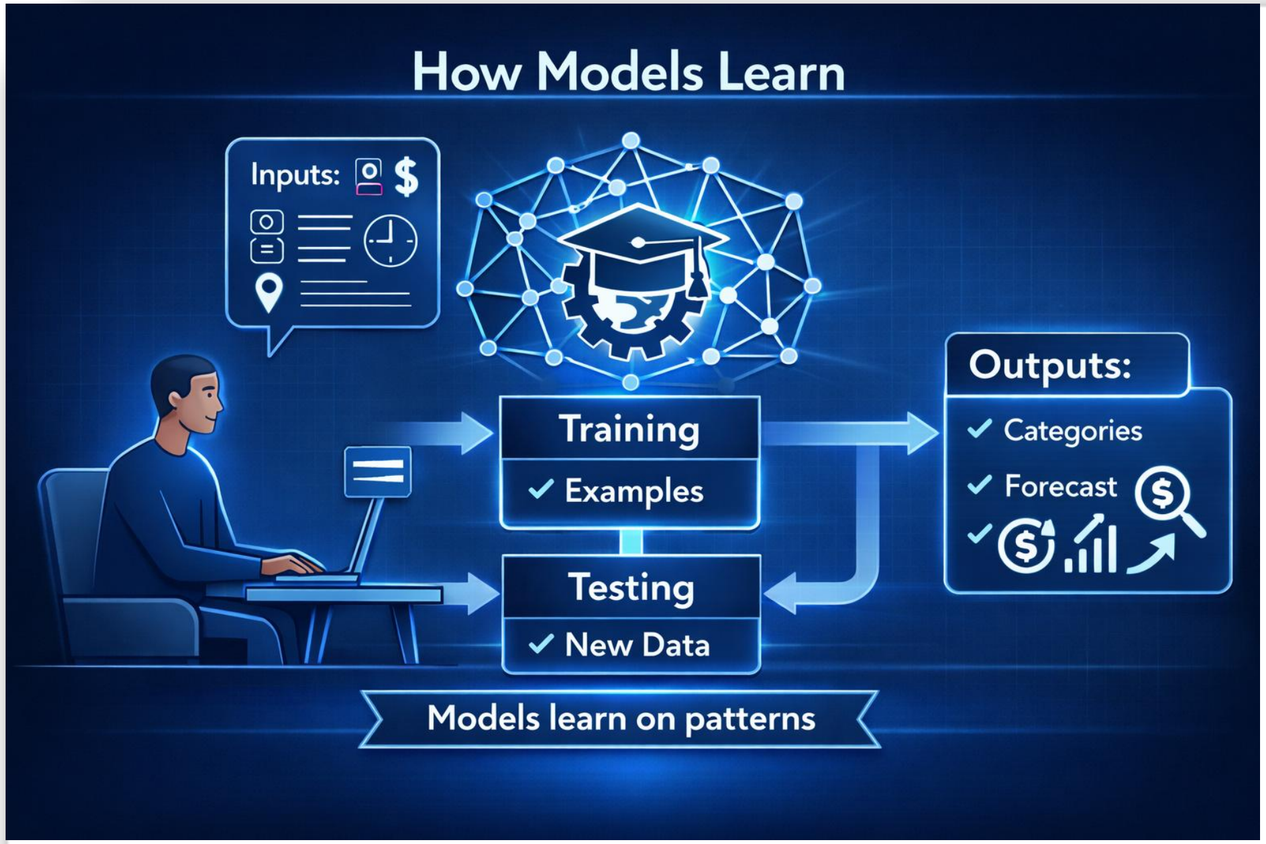
AI generates capability. People provide accountability!

Models that Power Finance

Category	Model	Example
Forecasting	Linear Regression	Forecasting revenue growth
Forecasting	Long-Short Term Memory (LSTM)	Collections over time
Classification	Random Forest	Expense type tagging
Segmentation	K-Means Clustering	Segmenting vendors
Interpretation	Transformers	Variance explanations

Different problems require different models!

How Models Learn?



- ✓ Training: historical inputs + labeled outputs
- ✓ Testing: checks if the model generalized
- ✓ The trainer’s judgment matters
- ✓ Iteration is essential

AI learns exactly what it’s shown!

Prompt Engineering for Open AI: CSI + FBI

CSI	FBI
Context	Format
Specifics	Behavior
Intent	Instructions

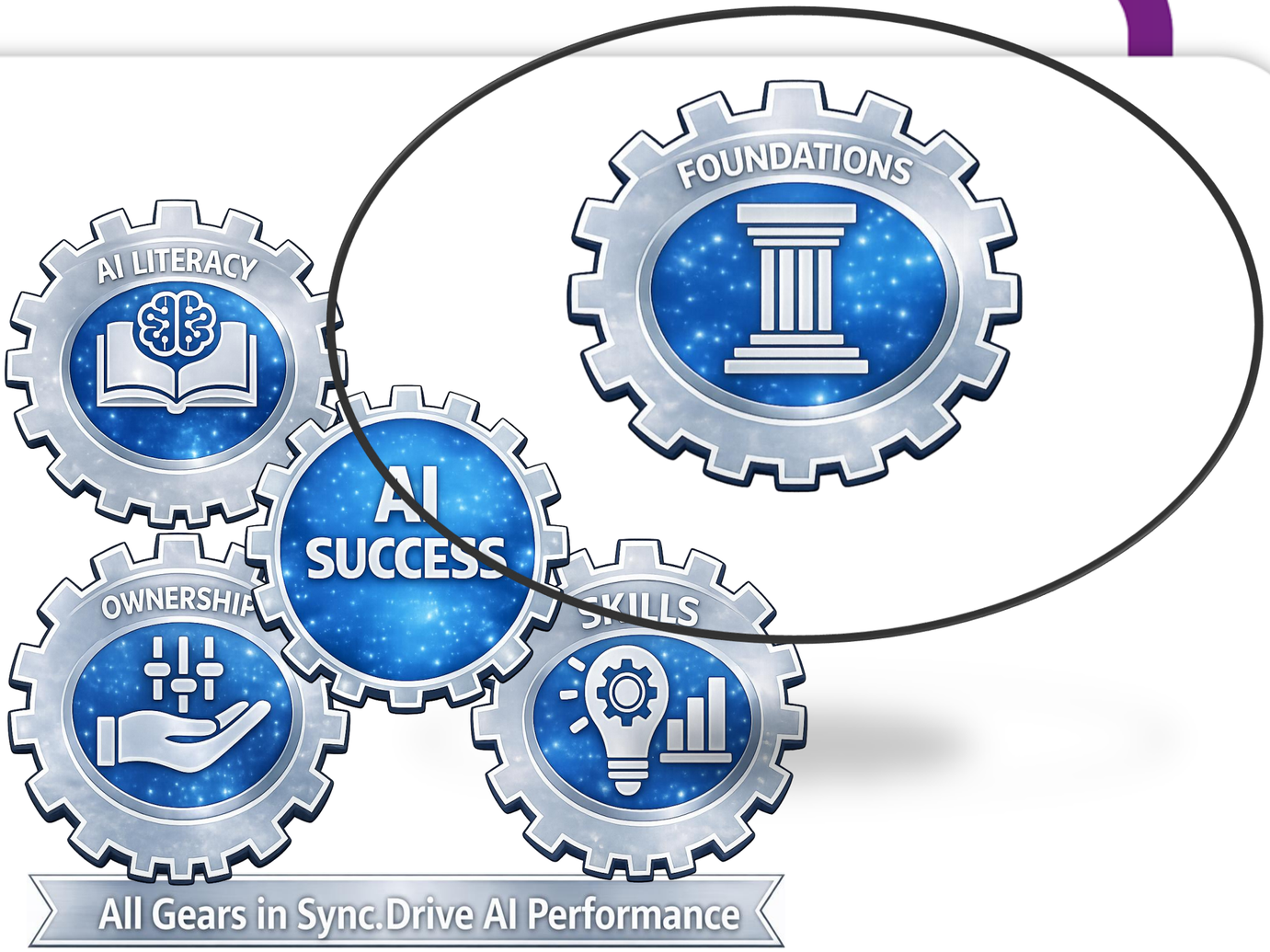
✓ **Weak prompt**: Can you write a quick summary of the Q2 financials? Be sure to mention bookings, churn, and margin

✓ **Same prompt using CSI + FBI**:

- I am an FP&A analyst preparing the CFO quarterly board update.
- The company operates in the SaaS space with a multi-product portfolio. Bookings are down 8% QoQ in Europe, but churn has improved across Small and mid-size business segments. Gross margin is under pressure due to rising cloud infrastructure costs.
- We need a concise narrative summary that explains the variance between forecast and actuals, highlights key risk areas, and surfaces at least two potential follow-up questions from the board.
- Use a bullet-point format with no more than 5 bullets.
- Keep the tone analytical and appropriate for an executive audience. Avoid jargon, technical model references, and speculation.
- If there are gaps in the logic or data, highlight them explicitly rather than smoothing over them.

AFP FP&A FORUM

Strong Foundations



Data

- ✓ AI now handles messy formats
- ✓ It can’t handle vague definitions
- ✓ Misalignment = mislearning
- ✓ Data Dictionary = knowledge
- ✓ Governance = operational clarity
- ✓ Consistence before sophistication

AI fails because data is ambiguous!

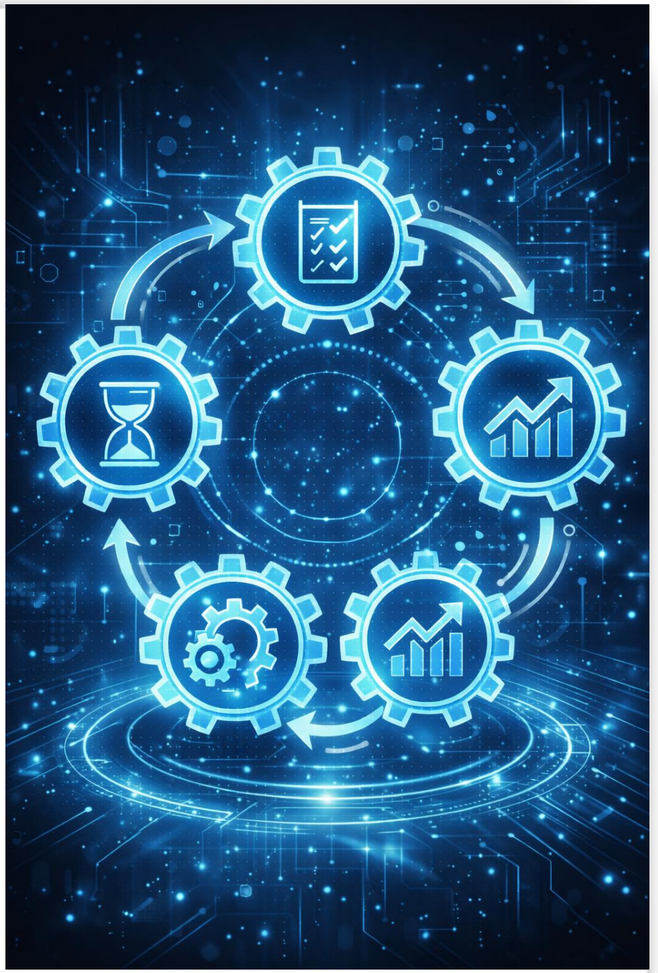
Systems

- ✓ Disconnected systems = blind spots
- ✓ AI fills gaps with outdated patterns
- ✓ Integration supports traceability and trust

If systems are siloed, AI doesn't forecast the business, just part of it!

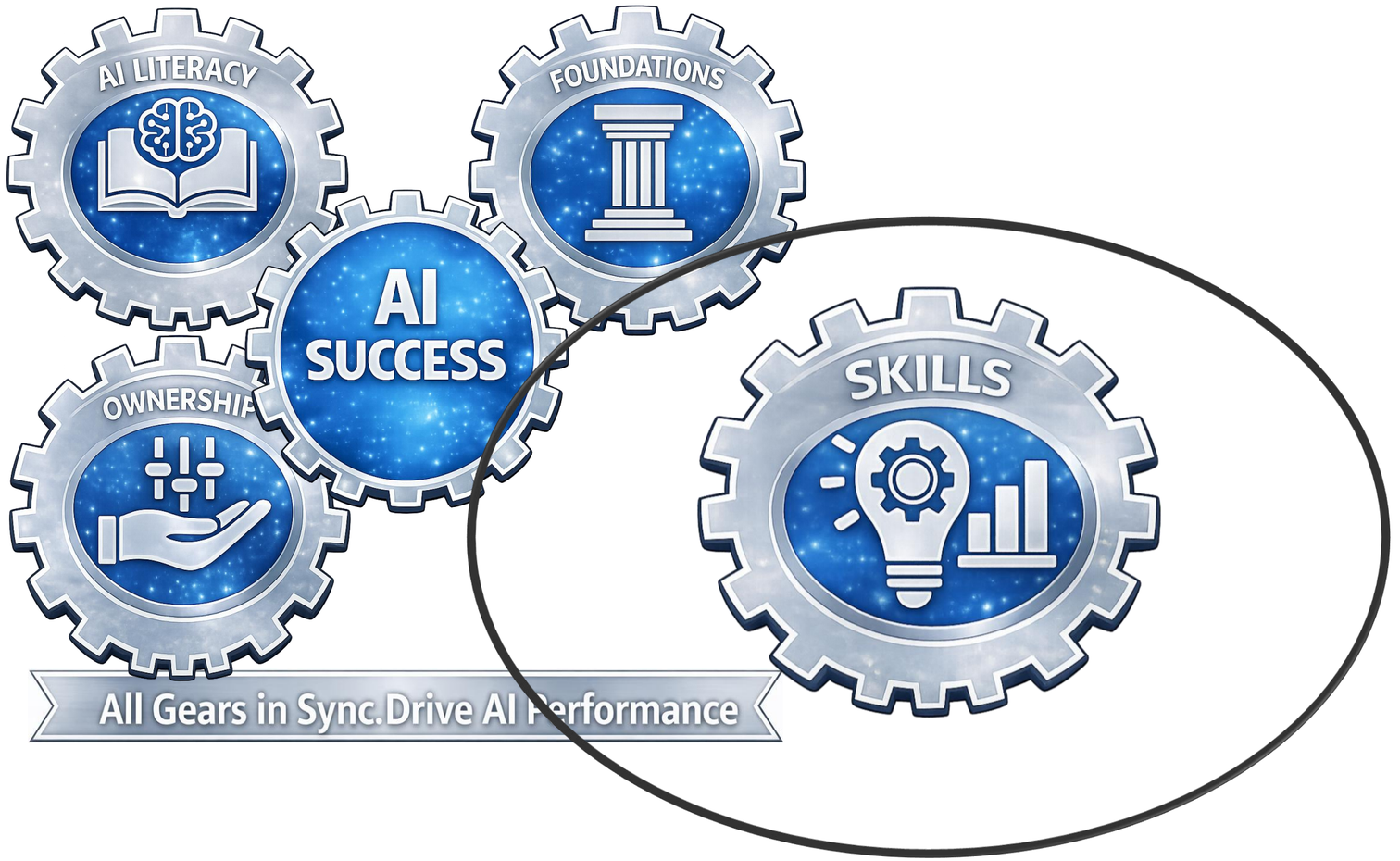
Processes

- ✓ AI needs consistency, not perfection
- ✓ Constantly shifting drivers = noise
- ✓ Predictable process = learnable pattern
- ✓ AI learns from patterns
  - No pattern = No reliable learning



Stability enables scale!

Skills


AI LITERACY
FOUNDATIONS
AI SUCCESS
OWNERSHIP
SKILLS
All Gears in Sync.Drive AI Performance

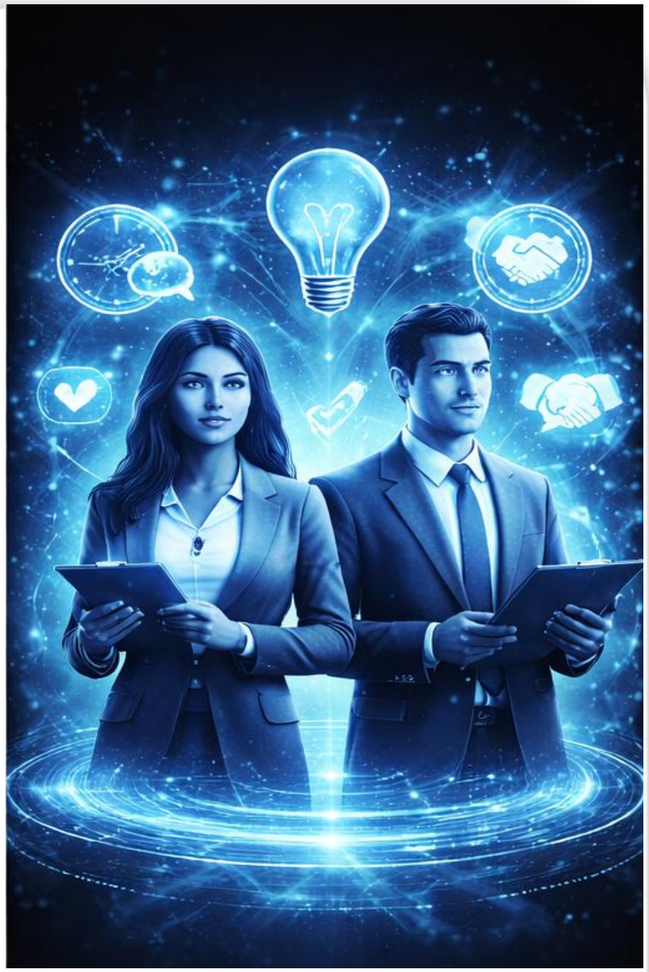
Hard Skills

- ✓ Basic Statistics
- ✓ Predictive analytics
- ✓ Data Fluency
- ✓ Automation
- ✓ Cybersecurity

If we can’t interpret it, we shouldn’t scale it!

Soft Skills

- ✓ Storytelling
- ✓ Cross-functional Collaboration
- ✓ Business Acumen
- ✓ Adaptability
- ✓ Active Listening
- ✓ Ethical Thinking



Models calculate. Finance decides what matters!

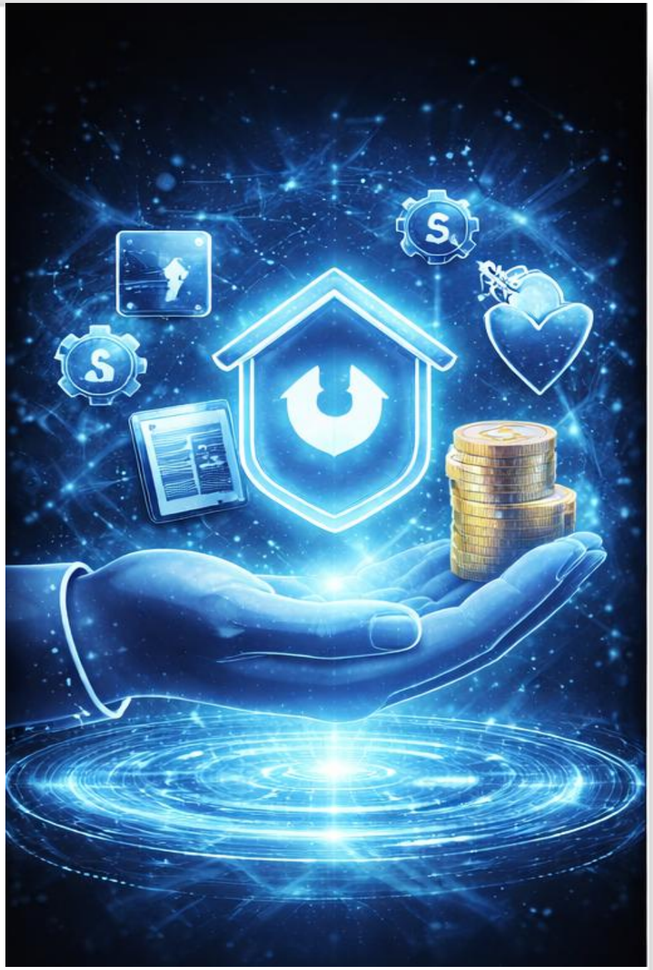
FP&A FORUM

Ownership Shift

The Ownership Shift

The goal is not control, the goal is no surprises

- ✓ Logic is moving into workflows, assumptions become defaults
- ✓ Decisions will be made anyway, finance will still be asked to explain them
- ✓ Finance needs to show up earlier to co-shape the logic



From Explaining Numbers to Designing Logic!

Thank You!

Appendix

Instructions

- ✓ Step 1: Score Individually
  - Be honest, as this is a diagnostic, not a performance review
- ✓ Step 2: Compare Within the Team
  - Discuss differences in scoring across roles:
    - Where do perceptions diverge?
    - Where is confidence assumed but not shared?
    - Where is clarity individual, not institutional?
  - Misalignment in scores is often more revealing than the score itself.
- ✓ Step 3: Identify Your Weakest Lever
  - AI readiness is systemic. Your lowest scoring category (Literacy, Data, Systems, Process, Skills, Ownership) is your constraint.
  - Strength in one area does not compensate for weakness in another.
- ✓ Step 4: Elevate One Lever at a Time
  - Use the “To Elevate” actions in the results table:
    - Pick 2–3 practical actions
    - Assign ownership and set a 90-day checkpoint
  - Do not try to fix everything at once.
  - AI readiness improves through structural progress, not enthusiasm.

Readiness determines whether AI delivers leverage or exposes gaps!

AI Literacy Score Card

Dimension	1	2	3	4	5	Your Score
What AI Is	AI viewed as magic, threat, or black box	General awareness of AI terms but unclear distinctions	Understands AI categories and basic differences	Has a clear mental model of AI vs ML vs GenAI	Can clearly explain AI concepts and implications to stakeholders	
Model Purpose	No clarity on use cases	Vague understanding of where AI might help	Knows common finance AI use cases	Consistently matches models to appropriate finance problems	Intentionally designs AI use cases aligned to decision needs	
How Models Learn	No understanding	Knows models use historical data	Understands training and testing conceptually	Understands inputs, outputs, validation, and bias risks	Can anticipate learning limitations and model failure modes	
Model Limitations	Unable to challenge outputs	Aware models can be wrong but unsure why	Knows errors exist but struggles to diagnose	Can challenge outputs and identify likely causes	Can mitigate root causes and adjust logic or data	
Explainability	Cannot explain results	Repeats tool or vendor explanations	Explains results at a high level	Explains logic, assumptions, and tradeoffs	Defends AI logic with confidence in decision settings	
Total Score						

AI Literacy Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**AI is understood and actively evaluated**	Overconfidence without governance discipline	• Standardize a shared vocabulary across FP&A (forecasting vs classification vs generative use cases). • Formalize model review checkpoints (What problem is it solving? What assumptions does it embed? What category does it fall into?). • Embed structured prompting standards (e.g., CSI + FBI) into recurring workflows like commentary and board prep. • Develop internal AI champions who can coach others and challenge outputs confidently.
11–17	**AI is used, but literacy is uneven**	Reliance on a few “AI translators”	• Run short internal education sessions on model categories (regression vs classification vs language). • Train FP&A to distinguish between explainable vs probabilistic outputs. • Introduce structured prompting templates to reduce vague AI usage. • Create a simple checklist for reviewing AI outputs before they are shared with leadership.
5–10	**AI feels like a black box**	Blind rejection or blind trust	• Start with foundational workshops on: What is AI? What is ML? What is generative AI? Teach basic concepts like correlation vs causation, model training vs testing, and overfitting in plain finance language. • Limit AI usage to controlled pilot cases with guided interpretation. • Pair FP&A with data teams during early exposure to build shared understanding.

Data Foundations Scorecard

Dimension	1	2	3	4	5	Your Score
Metric definitions	No shared definitions. Same metric means different things depending on report	Informal definitions vary by team and region	Definitions exist but are not enforced	One definition used consistently and documented	Definitions embedded in systems and reporting	
Calculation logic	Different logic used everywhere	Mostly aligned, many exceptions handled manually	Standard logic exists, but overrides are common	Fully consistent logic, clearly documented	Logic from centrally governed, version-controlled, and change-managed	
Data literacy	No shared data dictionary or understanding	Basic awareness, partially documented, relies on tribal knowledge	Mixed understanding with multiple documents and interpretations	Strong understanding of data meaning and limits within FP&A	Data literacy embedded across the org, not dependent on individuals	
Data governance	No ownership or accountability	Ownership unclear or informal	Owners exist but controls are weak	Clear owners with defined controls	Governance actively maintained, reviewed, and audited	
Total Score						

Data Foundations Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**Definitions are aligned and governed**	Subtle drift over time. As the business evolves, definitions quietly change without formal review, creating future misalignment.	• Formalize a metric ownership registry (clear owner per KPI). • Implement change-control for metric definition updates. Introduce periodic definition audits (e.g., quarterly logic reviews). • Embed definition documentation directly into reporting tools to reduce interpretation gaps.
11–17	**Definitions exist, but inconsistencies remain**	AI scales hidden definition differences across teams or regions, creating conflicting outputs and debates in meetings.	• Identify top 10 KPIs used in forecasting and reporting and align definitions across all systems. Reconcile gross vs net, accrual vs cash, management vs accounting views. • Establish a single authoritative source per key metric. Create a shared data glossary accessible to FP&A and stakeholders.
5–10	**Metrics are ambiguous or locally defined**	AI automates inconsistency. Meetings focus on debating numbers rather than discussing decisions. Trust weakens quickly.	• Map where core KPIs are calculated and how they differ across teams. • Document current definitions before attempting automation. • Pause AI expansion in areas with unclear definitions. • Assign explicit metric ownership within finance before scaling models further.

Systems Integration Scorecard

Dimension	1	2	3	4	5	Your Score
Core system integration	Systems fully siloed and disconnected	Manual exports between systems	Partial integrations with gaps	End-to-end integration for core workflows	Seamless, scalable architecture supporting all finance needs	
Upstream visibility (CRM, Ops)	Finance sees only results	Limited upstream signals available	Key drivers visible, but incomplete	Full driver visibility across functions	Real-time or near-real-time operational signals	
Data flow timing	Lagged data, manual refreshes	Periodic automation with delays	Mostly automated with some timing gaps	Timely and reliable data flows	Designed for decision speed, not just reporting	
Explainability	Numbers hard to trace	Traceable, but only with effort	Mostly traceable across systems	Clearly traceable from driver to output	Fully transparent driver-to-output logic	
Total Score						

Systems Integration Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**Systems are connected and financially traceable**	Silent drift between operational and financial systems as new tools or products are added. Integration becomes outdated without review.	• Establish periodic reconciliation between operational and financial drivers (CRM ↔ ERP ↔ HR). • Assign clear ownership for cross-system logic alignment. • Build driver-to-outcome traceability dashboards (pipeline → revenue → margin). • Review integrations when new products, regions, or channels are launched.
11–17	**Partial integration with manual bridges**	AI analyzes fragmented views of the business. Forecasts reflect incomplete signals, leading to inconsistent executive conversations.	• Identify top decision-critical data flows (e.g., pipeline → forecast, headcount → expense). • Reduce manual reconciliations between systems. Align timing logic across systems (close calendars, accrual timing, operational cutoffs). • Prioritize integrations based on explainability needs, not just IT convenience.
5–10	**Fragmented systems and siloed reporting**	AI produces confident outputs based on partial business visibility. Finance cannot clearly trace drivers to results, weakening trust.	• Map where core financial drivers originate and where they are stored. Identify critical gaps where operational signals do not reach finance datasets. • Create a cross-functional alignment session (Finance + IT + Operations) focused on driver visibility. • Avoid scaling AI use cases until key driver integrations are addressed.

Process Consistency Scorecard

Dimension	1	2	3	4	5	Your Score
Forecast methodology	Method changes every cycle with no continuity	Frequent changes driven by short-term needs	Mostly stable, occasional changes	Stable with clear, documented rules	Designed intentionally for learning and reuse	
Adjustments	Ad hoc, undocumented adjustments	Semi-documented, inconsistent use	Documented but inconsistent	Structured, controlled, and reviewable	Governed, intentional, and minimal	
Timing & cadence	Timing varies by team and cycle. Data arrives inconsistently	Target schedules exist but are loosely followed	Most inputs arrive on a consistent cadence	Timing is standardized and enforced across cycles	Timing is designed to support early, decision-ready signals	
Repeatability Across Cycles	Every cycle is built differently	Some structure repeats, many exceptions	Core structure repeats with noise	Same structure reused across cycles	Structure intentionally designed, reused, and refined	
Total Score						

Process Consistency Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**Processes are repeatable and decision logic is stable**	Subtle process drift over time. Informal adjustments begin creeping back in without documentation, weakening learning consistency.	• Formalize change control for forecast drivers and key assumptions. • Document when and why drivers change (not just that they changed). • Establish clear override rules: when human judgment is allowed and how it is recorded. • Periodically review whether process changes are intentional or reactive.
11–17	**Core processes exist but vary by team, region, or cycle**	AI learns unstable patterns. Forecast volatility increases. Trust becomes dependent on individual explanations rather than system logic.	• Identify where forecast logic changes quarter to quarter. • Standardize driver definitions across teams and regions. • Reduce ad-hoc overrides by clarifying when exceptions are appropriate. • Align timing assumptions across close, forecast, and reporting cycles.
5–10	**Processes rely heavily on manual adjustments and informal judgment**	AI cannot anchor to stable patterns. Outputs feel inconsistent, leading to manual rework and eventual tool abandonment.	• Map where manual adjustments are routinely applied. Separate structural logic from one-off judgment. • Define baseline forecasting methodology before layering exceptions. • Avoid scaling AI use cases until core logic is stabilized long enough for learning to occur.

Hard Skills Readiness Scorecard

Skill	1	2	3	4	5	Your Score
Basic Statistics	No understanding	Remembers concept	Applies averages & trends	Interprets distributions & variances	Uses statistics to judge model reliability	
Predictive analytics	Output accepted blindly	Directional understanding	Interprets results	Confidently challenges outputs	Shapes how predictive outputs are used in decisions	
Data fluency	Fully dependent on others	Basic comfort	Works with data tools	Comfortable working with data tools (PQ, SQL, Python, or equivalent)	Uses data strategically to explore and resolve issues	
Automation & workflows	Manual processes	Basic automation	Some workflow design	Understands scalable automation (RPA, APIs, workflows)	Designs reusable, scalable finance workflows	
AI & ML literacy	AI feels opaque	Knows buzzwords	Understands basics	Can assess model quality	Anticipates model risks	
Governance	Ignored	Aware but passive	Basic controls	Active stewardship	Ethical data leadership	
Total Score						

Hard Skills Readiness Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**AI outputs are rigorously evaluated and confidently challenged**	Overconfidence. The team understands outputs but may not consistently pressure-test assumptions or monitor model drift over time.	• Formalize model review checkpoints (performance, assumptions, drift). • Create a standard “AI output challenge” checklist (data, drivers, assumptions, stability). • Document known model limitations and communicate them clearly to stakeholders. • Train additional team members to avoid concentration of expertise.
11–17	**Basic literacy exists, but capability is uneven**	Dependence on a few knowledgeable individuals. Outputs are accepted or questioned inconsistently. Confidence fluctuates with complexity.	• Strengthen statistical literacy (mean vs median, correlation vs causation, variance interpretation). • Ensure the team understands what each model is predicting and what it is not. • Increase data fluency (SQL, Power Query, basic validation techniques). • Build structured training around interpreting predictive outputs and confidence levels.
5–10	**AI outputs are treated as black boxes or dismissed**	Blind trust or blanket skepticism. Decisions rely on intuition or manual overrides instead of structured evaluation.	• Establish foundational statistics training across FP&A. • Clarify the difference between forecasting, classification, segmentation, and language models. • Introduce basic predictive analytics literacy (inputs, outputs, assumptions). • Pair finance professionals with data teams for structured knowledge transfer.

Soft Skills Readiness Scorecard

Skill	1	2	3	4	5	Your Score
Storytelling	Reports numbers	Explains variances	Links drivers	Answers “so what”	Frames insights that consistently influence decisions	
Cross-functional collaboration	Finance-only	Limited engagement	Regular collaboration	Communicates effectively across functions	Trusted strategic partner across functions	
Business acumen	Variance-focused	Some context	Outcome-aware	Strong operational insight	Anticipates business impact and trade-offs	
Adaptability	Resistant to change	Learns slowly	Learns when required	Learns proactively	Proactively adjusts approach as tools, data, and context change	
Ethical & critical thinking	Accepts outputs	Questions occasionally	Evaluates logic	Challenges recommendations	Sets and enforces ethical boundaries	
Communication & Active listening	One-way	Reactive	Clear explanations	Active listener	Shapes dialogue and alignment	
Total Score						

Soft Skills Readiness Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**Finance leads decisions confidently in an AI-enabled environment**	Strategic blind spots. Strong analytical capability but occasional disconnect from stakeholder priorities or long-term implications.	• Formalize structured storytelling around AI outputs (clear narrative, risk framing, implications). • Strengthen cross-functional AI governance forums (Finance + Ops + Tech). • Embed ethical review into major AI-driven decisions. • Develop succession depth so leadership capability isn't concentrated.
11–17	**Finance interprets AI outputs, but influence varies**	Insights may be technically correct but fail to drive decisions. Communication or stakeholder alignment is inconsistent.	• Strengthen executive storytelling: always connect output to decision and impact. • Build structured collaboration routines with operations and data teams. • Train teams to translate probabilities and uncertainty into business language. • Encourage proactive challenge instead of reactive explanation.
5–10	**Finance reacts to AI outputs rather than shaping decisions**	Over-reliance on technical output. Decisions drift back to intuition or hierarchy when tension arises.	• Develop core storytelling capability (problem → insight → implication → recommendation). Increase cross-functional exposure and business immersion. • Train teams in critical thinking and scenario framing. • Reinforce active listening — clarify what stakeholders truly need before presenting analysis.

Ownership & Accountability Readiness Scorecard

Dimension	1	2	3	4	5	Your Score
Metric ownership	No clear owner	Informal ownership	Named owners, weak authority	Clear owners with authority	Ownership embedded in operating model	
Logic ownership (assumptions, rules)	Logic undocumented	Logic implicit	Logic documented	Logic enforced	Logic governed and version-controlled	
Accountability for outputs	Finance explains only	Shared accountability unclear	Finance accountable after the fact	Finance accountable end-to-end	Finance designs and owns outcomes	
Upstream involvement	Finance downstream only	Limited early involvement	Involved selectively	Involved in most decisions	Finance shapes decisions upstream	
Cross-functional decision rights	Undefined	Informal	Case-by-case	Clear and respected	Explicit and embedded	
Escalation & challenge	Avoided	Ad hoc	Possible but slow	Structured	Expected and institutionalized	
Total Score						

Ownership & Accountability Readiness Score Card Results

Score Range	Current Stage	Primary Risk at This Level	To Elevate to the Next Level, Your Team Must…
18–25	**Finance owns decision logic upstream**	Overconfidence or governance fatigue. Logic ownership exists, but review cadence and cross-functional reinforcement may weaken over time.	• Formalize AI logic ownership charters (who defines metrics, assumptions, override rules). • Implement structured review cycles for model assumptions and embedded logic. • Create clear documentation of “where judgment lives” vs automation. • Ensure finance participates in model design discussions before build, not after deployment.
11–17	**Finance partially owns logic but reacts in some areas**	Inconsistent accountability. Finance explains outputs but does not consistently define assumptions before models run.	• Define metric ownership explicitly (no shared ambiguity). • Document forecast logic and decision rules before automation. • Establish clear override policies (who can adjust and why). • Shift AI conversations from “Is the number right?” to “Is the logic right?”
5–10	**Finance primarily explains results downstream**	Decision logic is embedded elsewhere (IT, Data Science, Operations) without finance governance. High risk of scaled misalignment.	• Insert finance earlier into model design and data definition conversations. • Clarify ownership of key KPIs and their calculation logic. • Define what “correct” means before automation. • Create a single accountable owner for each AI-driven metric.

Readiness Pattern Interpretation

What Your Score Combinations Mean?

Pattern Combination	What It Indicates	What Happens in Practice	Leadership Priority
High Literacy + Weak Foundations	Smart people, unstable system	Teams understand AI concepts but struggle to trust outputs due to inconsistent data, systems, or processes	Stabilize definitions, integration, and repeatability before scaling AI
Strong Foundations + Weak Ownership	AI runs, no one stands behind it	Models operate technically well, but accountability for logic is unclear	Clarify metric definitions, assumptions, and decision authority
Strong Skills + Weak Governance/Security	Fast, risky AI	Teams build and automate quickly but expose the organization to control or compliance risk	Define access rules, approval boundaries, and data security standards
High Ownership + Low Literacy	Confident decisions, fragile evaluation	Finance owns logic but cannot properly evaluate model behavior	Strengthen conceptual AI understanding across FP&A
Strong Systems + Weak Processes	Connected but inconsistent	Data flows end-to-end, but process changes disrupt learning stability	Standardize forecasting cadence and driver governance
High Literacy + Weak Skills	Theoretical understanding, limited execution	Teams understand concepts but cannot challenge or operationalize outputs effectively	Invest in statistical fluency, data capability, and workflow design
Strong Processes + Weak Systems	Stable logic, fragmented visibility	Repeatable workflows exist, but missing system signals create blind spots	Improve system integration aligned to decision explainability
All High but Misaligned Across Roles	Perception gap	Leadership believes readiness is high, but frontline teams disagree	Recommended to facilitate alignment session before scaling further