BOOK NINE · THE AI ECONOMY MONETIZATION SERIES

Customer Success in the AI Economy

Adoption, expansion, retention, and the new role of customer success when the product runs autonomously

Stop asking whether they are using it. Start asking whether it is working.

Customer success in the AI era is not about product adoption — it is about outcome realisation. The product is doing things autonomously. CS is there to make sure those things are the right things, that they are working, and that the customer is capturing the full value of what the AI can deliver.

Framework F23: The AI Customer Health Architecture

PREFACE

The Question That Has Become Obsolete

'Are your users logging in?' — and why asking it for AI products is not just wrong, but actively misleading.

There is a fundamental question that has quietly become obsolete in AI customer success: "Are your users logging in?"

For twenty years of SaaS, this was the right question. Software that was not being used was not providing value. Usage — measured by logins, active users, features accessed, sessions completed — was the leading indicator of retention and the proxy for value

delivery. Customer success teams built their entire operating model around this question. QBRs were structured around feature adoption dashboards. Health scores were weighted heavily on monthly active users. Churn prediction models identified disengagement through declining login frequency.

Then the products started running themselves.

When Harvey AI reviews a customer's contracts overnight, the attorneys do not log in to operate the AI. When Intercom's AI resolves customer service tickets, the support team does not log in to activate each resolution. When Abridge generates clinical notes from physician conversations, the physicians are not sitting at dashboards managing the documentation process. The AI works. The humans receive the outputs. "Are your users logging in?" is not just the wrong question for these products — it is a question that has no meaningful answer. There is nothing for users to log into.

This is the transformation that customer success has not yet fully absorbed. The products have changed. The CS playbook has not kept pace. CS teams are still measuring logins for products where logins are irrelevant. They are still running QBRs structured around feature adoption for products where features run automatically. They are still predicting churn from engagement signals for products where engagement is indistinguishable from outcome delivery.

This book rebuilds the CS playbook from the ground up for AI-native products. It keeps what is still true — customer relationship management, renewal management, expansion conversation, advocacy creation — while replacing what is obsolete. The new foundation is not adoption. It is outcome realisation: is the AI delivering the specific business outcomes it was deployed to create, at the quality and reliability that justifies the commercial relationship?

This is a harder question than "are they logging in?" It requires measurement infrastructure that most CS teams have not built. It requires different conversations — with the customer's finance team, not just their operations team. It requires different health metrics — not engagement scores, but outcome delivery rates and value

realisation percentages. And it requires a different kind of CSM: not a product educator, but a value analyst and commercial advocate.

The companies that make this transition successfully — that rebuild their CS motion around outcome realisation rather than product adoption — will have customer relationships built on accountability and evidence. They will renew contracts because the evidence for renewal is overwhelming. They will expand accounts because the outcome data shows exactly where more AI deployment would create more value. And they will generate the kind of customer advocacy that no feature demonstration can create: customers who can say, with data, that the AI changed something important about how their business operates.

The companies that do not make this transition will continue to run QBRs structured around metrics that their customers have stopped caring about, and they will be confused when renewals become contentious despite strong product adoption numbers.

The product is running. The question now is whether CS is keeping up.

PART ONE

CS in the Age of Autonomous Products

The new mandate. The health architecture. The UAS.

CHAPTER ONE

The New Customer Success Mandate: From Adoption to Outcome

Why 'are they using it?' is the wrong question. The shift to 'is it working?' and 'are they growing?'

The customer success mandate in the AI era has shifted at its foundation. The old mandate was adoption-focused: ensure customers deploy the product, train users to use it effectively, monitor usage metrics to identify risk, and intervene when usage declines. The underlying logic was correct for SaaS: if users are using the software, they are getting value from it. If they stop using it, they are not getting value, which means they will eventually stop paying.

The AI era breaks this logic at every step. Users may not be using the software — the AI uses it on their behalf. Value is not proportional to usage — it is proportional to outcomes. Declining usage metrics may indicate that the AI is working well (the AI is handling things that previously required user intervention), not that value is declining. Increasing usage metrics may indicate that the AI is struggling (the AI is failing to resolve things autonomously, requiring more human intervention), not that value is growing.

The new mandate is outcome-focused: ensure the AI is delivering the specific business outcomes it was deployed to create, measure the economic value of those outcomes, identify opportunities to expand the AI's scope to create more value, and use the outcome evidence to support renewal and expansion conversations.

The specific shift in each CS operational area:

Onboarding: from user activation to value activation. The old onboarding success metric was "users have completed training and are using the product." The new success metric is "the AI has been deployed in a specific use case, is processing real work, and has delivered at least one verified outcome." The timeline for "activated" may be longer (deploying AI in a production workflow takes more setup than training users on a UI), but the definition is more precise and more commercially meaningful.

Health scoring: from engagement to outcome delivery. The old health score weighted user logins, feature adoption, and training completion. The new health score weights the AI's outcome delivery rate (what percentage of work items assigned to the AI are successfully completed), the outcome quality score (are the AI's outputs meeting the

quality standard the customer expects), and the value realisation rate (what percentage of the potential value the AI could create is actually being captured). These metrics require measurement infrastructure that is more complex to build than login tracking — but they are the metrics that actually predict retention and expansion.

Expansion identification: from usage patterns to outcome evidence. The old expansion conversation was: "You are using the product heavily in your sales team — have you considered deploying it in your marketing team?" The new expansion conversation is: "Your AI is resolving 72% of customer service inquiries automatically, saving your team 840 hours per month. Based on the inquiry mix we see in your data, we estimate that an expanded AI deployment in your technical support channel could save an additional 600 hours per month. Here is the analysis." The expansion is evidence-based, not pattern-based. The conversation is with the CFO, not just the operations team.

Renewal management: from relationship maintenance to value documentation. The old renewal process was: build the relationship, prepare a usage summary, have a conversation about the next term. The new renewal process is: prepare a value documentation package — the specific outcomes the AI delivered in the current term, the economic value of those outcomes, the comparison to the contract value, and the projection of value available in the next term. The renewal is not a relationship conversation. It is a value evidence conversation. The customer is not being asked to renew based on trust; they are being shown the evidence that makes renewal the obvious economic decision.

> ***"The SaaS CS playbook was built for a world where humans operated software. The AI CS playbook is built for a world where software operates autonomously on humans' behalf."***

CS Operational Shifts: SaaS vs AI-Native Products			
CS area	**SaaS model**	**AI-native model**	**Why it changed**
Success definition	Users trained and actively using product	AI deployed in production workflow delivering verified outcomes	The AI uses the product; humans receive outputs — usage metrics are irrelevant
Activation metric	Users completed onboarding and have logged in	AI has processed real work and delivered first verified outcome	Activation requires value delivery, not user training
Health score	Weighted on logins, feature adoption, active users	Weighted on outcome delivery rate, deployment breadth, expansion velocity	AI may have zero logins and be working perfectly; high logins may indicate AI is failing
Expansion trigger	Heavy usage in one team → offer to other teams	Outcome data shows AI underdeployed relative to workflow volume → evidence-based expansion brief	Expansion is evidence-based (what the data shows) not pattern-based (who is using it a lot)
Renewal conversation	Usage summary + relationship maintenance + feature roadmap	Value documentation: outcomes delivered, economic value, ROI vs contract cost, next-term opportunity	Customer's commercial leadership makes renewal decision on evidence, not relationship trust
Churn predictor	Declining login frequency, feature disengagement	Outcome quality plateau, scope contraction, stakeholder displacement	Disengaged humans don't churn AI products — underperforming AI churns AI products
CSM skills	Product knowledge, relationship management, user training	Data literacy, economic fluency, domain expertise, outcome analysis	CSM must translate AI performance data into financial business cases, not feature adoption stories

Chapter One — The Essentials

› The fundamental CS mandate shift: from adoption-focused (are they using it?) to outcome-focused (is it working?).

› Activation is redefined: not user training completed, but first verified meaningful outcome delivered.

› Health scoring is redesigned around outcome delivery, deployment breadth, and expansion

velocity — not logins.

› The renewal conversation is redesigned around value documentation and ROI evidence — not relationship maintenance.

› The CSM skills requirement expands: data literacy, economic fluency, and domain expertise are now required alongside product knowledge and relationship skills.

CHAPTER TWO

The AI Customer Health Architecture (Framework F23)

Five signals: usage depth · outcome realisation · expansion velocity · risk indicators · sentiment.

The AI Customer Health Architecture (Framework F23) is the operating framework for customer success in AI-native products. It identifies five distinct health signals that together provide a comprehensive view of whether a customer's AI deployment is healthy, growing, and commercially defensible.

The five health signals:

Signal 1 — Usage depth. Not whether users are logging in, but how deeply the AI is embedded in the customer's actual operational workflows. A shallow deployment is one where the AI handles a narrow, easily defined subset of the customer's work — the test cases, the simple examples, the scenarios where AI performance is most predictable. A deep deployment is one where the AI is integrated into the customer's primary operational workflows, handling the full range of complexity that those workflows involve. Usage depth is measured by the breadth of work item types the AI handles, the proportion of the customer's total workflow volume that the AI processes, and the degree to which the customer's operations would be disrupted if the AI were unavailable.

Signal 2 — Outcome realisation. The percentage of work items processed by the AI that result in verified, economically valuable outcomes. For a contract review AI, outcome realisation is the percentage of reviewed contracts for which the review was useful — the attorney acted on the AI's findings and did not need to significantly rework the analysis. For a customer service AI, outcome realisation is the percentage of tickets resolved by the AI without human escalation. For a clinical documentation AI, outcome realisation is the percentage of AI-generated notes that were accepted by the physician without significant modification. Outcome realisation is the core health signal: it directly measures whether the AI is delivering value on each interaction.

Signal 3 — Expansion velocity. The rate at which the customer is deploying the AI into new use cases, new teams, or new workflows. Positive expansion velocity — the customer is actively extending the AI's scope — is the strongest signal of health in an AI-native product. It indicates that the customer has validated the AI's value in initial deployments and is confident enough to extend it further. Negative expansion velocity — the deployment scope is contracting or stagnant — is a significant health risk signal even when outcome realisation metrics are positive, because it suggests the customer has reached the ceiling of what they are willing to deploy.

Signal 4 — Risk indicators. Specific signals that suggest the customer's AI deployment is at risk of underperforming, being reduced in scope, or being discontinued. Risk indicators include: outcome realisation rate declining over consecutive measurement periods (the AI is becoming less reliable), deployment scope contracting (the customer is pulling the AI back from workflows where it was previously deployed), billing disputes (the customer is questioning the commercial relationship), and stakeholder change (the executive champion for the AI deployment has left the organisation, leaving the deployment without internal advocacy).

Signal 5 — Sentiment. The qualitative signal from the customer's team about their experience with the AI deployment. Sentiment captures what the quantitative metrics miss: the frustration with an AI that is technically reliable but produces outputs in formats that require significant manual reformatting; the enthusiasm of a team that is

finding unexpected additional uses for the AI beyond the initial deployment scope; the anxiety of a compliance team that is unsure whether the AI's outputs meet their regulatory standards. Sentiment is assessed through regular structured conversations, customer satisfaction surveys, and informal signals from the CSM's relationship with the customer team.

The architecture's insight is that no single signal is sufficient. A customer with high outcome realisation but negative expansion velocity may be satisfied but capped — the AI is delivering well in its current scope, but the customer does not see additional deployment opportunities. A customer with positive expansion velocity but declining outcome realisation is growing their commitment while the AI's performance is degrading — a dangerous combination that requires immediate intervention. The five signals together provide the comprehensive picture that renewal and expansion decisions require.

HEALTH SIGNAL 1: USAGE DEPTH *How deeply is the AI embedded in the customer's actual operational workflows — not just whether users log in.*	
How measured	Work item type breadth (how many categories of work the AI handles), proportion of eligible workflow volume the AI processes, operational dependency score (would disruption of the AI significantly disrupt operations).
Healthy range	AI handles > 60% of eligible workflow volume; processes multiple work item types; high operational dependency rating.
Risk signal	AI scope narrowing over time; AI handling only 'safe' work types while complex work is manually handled; low operational dependency (AI could be removed without significant disruption).

HEALTH SIGNAL 2: OUTCOME REALISATION *The percentage of AI-processed work items that result in verified, economically valuable outcomes meeting the quality standard.*	
How measured	Outcome delivery rate (verified completions / total work items processed). Quality score from customer feedback or independent audit. Value realisation rate (documented value / contract value).
Healthy range	Outcome delivery rate > 85%; quality score trending stable or improving; value

	realisation rate > 2× (customer receiving at least 2× contract value in documented outcomes).
Risk signal	Outcome delivery rate declining over 3+ consecutive measurement periods; quality score declining; value realisation rate < 1× (customer receiving less value than the contract costs).

HEALTH SIGNAL 3: EXPANSION VELOCITY *The rate at which the customer is deploying the AI into new use cases, teams, or workflows — the leading indicator of future NRR.*	
How measured	Deployment scope change over trailing 90 days (new use cases added, new teams onboarded, new workflow categories enabled). Net expansion rate (expansions minus contractions).
Healthy range	Positive net expansion (deployment scope growing over trailing 90 days); at least one new use case or team deployment in the last 90 days.
Risk signal	Flat or declining deployment scope for 90+ days; customer explicitly contracting scope (routing work away from AI); no expansion conversations in the last 60 days.

HEALTH SIGNAL 4: RISK INDICATORS *Specific signals that suggest the AI deployment is at risk of underperforming, being reduced, or being discontinued.*	
How measured	Outcome quality plateau (performance flat for 60+ days when customer expects improvement); scope contraction; stakeholder displacement (key champion change); billing disputes; integration instability.
Healthy range	No risk indicators above threshold; stable stakeholder map; no billing disputes; integration uptime > 99.5%.
Risk signal	Any single risk indicator above threshold warrants immediate CSM investigation. Two or more risk indicators above threshold triggers CS Agent Huddle and formal intervention plan.

HEALTH SIGNAL 5: SENTIMENT *Qualitative signal from the customer's team about their experience with the AI deployment — capturing what metrics miss.*	
How measured	Regular structured conversations with operational team and commercial stakeholders; customer satisfaction pulse surveys; CSM qualitative assessment after each meaningful interaction.
Healthy range	Positive sentiment expressed by both operational team (the AI is making their work easier) and commercial stakeholders (the AI investment is worth it).

Risk signal	Expressed frustration with output quality or format; commercial stakeholder avoiding engagement; operational team creating workarounds rather than using AI outputs directly.

THE ARCHITECTURE INSIGHT

No single signal is sufficient — it is the combination that reveals true health

A customer with high outcome realisation but negative expansion velocity may be satisfied but capped — the AI is performing well in its current scope, but the customer does not see additional deployment opportunities. A customer with positive expansion velocity but declining outcome realisation is growing their commitment while the AI's performance is degrading — the most dangerous combination, requiring immediate intervention. The five signals together provide the comprehensive view that renewal and expansion decisions require. Weight the combination, not any single signal.

Chapter Two — The Essentials

› Five health signals: usage depth, outcome realisation, expansion velocity, risk indicators, sentiment — no single signal is sufficient alone.

› Usage depth measures workflow embedding, not user activity — an AI that runs autonomously with no user logins can have maximum usage depth.

› Outcome realisation is the core health signal: is the AI delivering value on each interaction? Trend matters more than absolute level.

› Expansion velocity is the leading indicator of NRR — positive expansion predicts NRR above 100%.

› The five signals together prevent the false positives and false negatives that any single metric produces for autonomous AI products.

CHAPTER THREE

The Usage Adoption Score (UAS) in Practice

Building, tracking, and acting on UAS. How UAS predicts NRR 90 days before the renewal.

The Usage Adoption Score (UAS) is the composite metric that combines multiple signals into a single number that predicts net revenue retention 90 days before the renewal. It is Framework F5 in this series, and it is the most commercially actionable metric in the AI customer success toolkit.

The UAS is not an engagement score. Traditional engagement scores measure how much users interact with the product — logins, feature clicks, session duration. The UAS measures how deeply the AI is embedded in value-creating work and how well it is performing in that work.

The UAS has four components weighted for AI-native products:

Component 1 (30% weight): Outcome delivery rate. The percentage of AI-processed work items that achieve the defined outcome successfully. For outcome-based billing products, this is the billing trigger metric. For subscription products, it is the primary evidence that the AI is delivering value for the subscription price.

Component 2 (25% weight): Deployment breadth. The percentage of the customer's total addressable workflow volume that the AI is currently processing. A deployment that handles 80% of the customer's eligible work volume is more deeply embedded — and more renewal-stable — than a deployment that handles 20% of eligible volume.

Component 3 (25% weight): Expansion trajectory. The rate at which deployment breadth is increasing over the trailing 90-day period. Positive trajectory (deployment breadth growing) is the leading indicator of NRR above 100%. Flat trajectory is a neutral signal requiring investigation. Declining trajectory is the strongest predictor of contraction or churn.

Component 4 (20% weight): Stakeholder engagement. The quality and frequency of meaningful commercial engagement between the vendor's CS team and the customer's decision-maker team — not just the operational team. Engagement with the CFO, the VP Operations, or the department head who owns the budget for the AI deployment is a strong health signal: the customer's commercial leadership is engaged with the AI's value, not just its operations.

UAS calculation: each component is scored on a scale of 0–100. The composite UAS is the weighted average of the four components. A UAS above 70 indicates a healthy deployment with high renewal probability and expansion potential. A UAS between 40 and 70 indicates a stable deployment with renewal risk if the score is declining or expansion opportunity if the score is improving. A UAS below 40 indicates a deployment at significant risk requiring immediate CS intervention.

The 90-day predictive horizon: the UAS predicts NRR outcomes 90 days in advance because the signals it captures — outcome delivery rate, deployment breadth, expansion trajectory, stakeholder engagement — are the underlying commercial health factors that will determine whether the renewal is strong, flat, or at risk. By the time the renewal conversation begins (typically 60–90 days before expiry), a UAS above 70 means the renewal is routine; a UAS below 40 means the renewal is at risk and the intervention must begin immediately.

Specific UAS benchmarks by AI product category:

For outcome-based AI services (contract review, clinical documentation, customer service AI): UAS above 70 requires outcome delivery rate above 85%, deployment breadth above 60% of eligible volume, positive expansion trajectory, and regular CFO/VP-level engagement. These thresholds reflect the higher commercial stakes of outcome-based products — the customer has a direct measure of value delivery that makes below-threshold performance commercially visible.

For productivity AI (coding assistants, writing AI, knowledge management): UAS above 70 requires team-level productivity improvement measurement (not just individual satisfaction), deployment across more than 70% of the eligible team, active use in high-value work (not just low-stakes tasks), and quarterly business review with department head-level sponsorship. The productivity AI benchmark reflects the challenge of measuring value in products where outcomes are diffuse across individual productivity improvements.

For decision support AI (analytics, forecasting, recommendation systems): UAS above 70 requires decision accuracy improvement above baseline, deployment in high-stakes decision workflows (not just exploratory use), evidence that AI recommendations are being acted upon, and evidence that the decisions made with AI support are producing better outcomes than baseline.

UAS Components — Weights and Calculation

Component	Weight	What it measures	Score 0–100	Calculation methodology
Outcome delivery rate	30%	Percentage of AI-processed work items achieving defined outcomes successfully	100 = delivery rate ≥ 95% trending stable/up · 70 = 80–94% stable · 40 = 65–79% or any declining trend · 0 = below 65% or significant decline	Rolling 90-day outcome delivery rate, adjusted for trend direction (stable, improving, or declining)
Deployment breadth	25%	Percentage of eligible workflow volume the AI is currently processing	100 = ≥ 80% of eligible volume · 70 = 50–79% · 40 = 25–49% · 0 = below 25%	(AI-processed work items / Total eligible work items) × 100 over trailing 90 days
Expansion trajectory	25%	Rate of change in deployment breadth over trailing 90 days	100 = ≥ 15% growth in 90 days · 70 = 5–14% growth · 40 = flat (±5%) · 0 = declining > 5%	(Current deployment breadth – 90-day-ago deployment breadth) / 90-day-ago deployment breadth × 100
Stakeholder engagement	20%	Quality and frequency of meaningful commercial engagement with	100 = regular CFO/VP-level engagement in last 60 days · 70	CSM assessment: seniority level of last

		decision-makers	= department head engagement · 40 = operational team only · 0 = no meaningful engagement in 90 days	meaningful engagement × recency score

UAS Interpretation and Action Guide

UAS score	Health status	Renewal prediction	Primary CS action	Commercial action
70–100	Healthy — AI deployment delivering well; customer capturing value	High confidence renewal; likely expansion	Proactive expansion conversation; reference customer programme	Prepare expansion brief; propose multi-year renewal at premium
55–70	Stable — deployment performing acceptably; some improvement opportunities	Renewal likely but not certain; flat NRR	Identify weakest component; targeted improvement conversation	Standard renewal preparation; modest expansion conversation
40–55	At-risk — health signals mixed; requires proactive attention	Renewal uncertain; contraction risk	CS Agent Huddle; root cause analysis; stabilisation plan	Renewal risk assessment; commercial options if stabilisation fails
25–40	Elevated risk — multiple negative signals; intervention required	Renewal at significant risk	Immediate intervention; executive reconnection; product investigation	Retention offer preparation; identify addressable commercial concerns
0–25	Critical — deployment in distress; urgent action required	High churn probability without rescue	CS emergency protocol; escalate to VP CS and account team; rapid remediation	Retention offer; executive relationship activation; last-resort commercial

				options

THE 90-DAY PREDICTION WINDOW

UAS predicts NRR outcomes 90 days in advance — giving CS teams a meaningful lead time for intervention

Analysis of AI product renewals shows that UAS assessed 90 days before renewal accurately predicts renewal outcomes: customers with UAS > 70 renew at 95%+ rates; customers with UAS 40–70 renew at 70–85% rates; customers with UAS < 40 renew at 45–60% rates. The 90-day window is not coincidental — it is the lead time between the health signals becoming visible and the renewal conversation beginning. A CS team that identifies a UAS below 40 with 90 days before renewal has time to implement the intervention plan and potentially rescue the renewal. A CS team that discovers the health issue 30 days before renewal has insufficient time for meaningful intervention.

Chapter Three — The Essentials

› UAS combines four components: outcome delivery rate (30%), deployment breadth (25%), expansion trajectory (25%), stakeholder engagement (20%).

› UAS > 70: healthy, high renewal confidence, prepare expansion conversation.

› UAS 40–70: stable but requiring attention; identify weakest component and intervene.

› UAS < 40: at-risk, immediate CS Agent Huddle, intervention plan required.

› The 90-day prediction window: UAS assessed 90 days before renewal predicts NRR with 85%+ accuracy — the lead time for effective intervention.

PART TWO

Health Scoring and Churn Prediction

New signals for autonomous products. Early warning. Intervention playbook.

CHAPTER FOUR

Health Scoring for AI Products: New Models for New Behaviours

When the product runs autonomously, traditional health scoring breaks. The new signals.

Health scoring for AI products requires a complete rethinking of what health means. Traditional SaaS health scoring models weight the signals that indicate a product is embedded in users' workflows: active users as a percentage of licensed seats, feature adoption breadth, login frequency, support ticket volume (low is good), and NPS score. These signals are reasonable proxies for value delivery in a product that humans actively operate.

For AI products, each of these signals either loses its meaning or actively misleads:

Active users as a percentage of licensed seats: In an AI product that operates autonomously, "active users" may mean the team members who occasionally review AI outputs — not the people whose work is being done by the AI. A customer who has deployed an AI to process their entire invoice management workflow might have only one "active user" who monitors exceptions, even though the AI is processing 5,000 invoices per month. Low active user count is not a health risk indicator — it is a success indicator.

Feature adoption breadth: AI products that run autonomously may have narrow feature adoption by design — the AI uses the features on the customer's behalf, and the features are not visible to the human users. A contract review AI that automatically applies the customer's negotiation playbook, flags specific clause types, and generates structured review documents is using multiple "features" internally without any human user activating them.

Login frequency: For AI products with autonomous operation, login frequency measures the rate at which humans intervene in or review AI activity — which declines as the AI performs better. Low login frequency is a positive health signal for a well-performing autonomous AI, not a negative one.

Support ticket volume: Low support volume can indicate either that the AI is performing well (no issues to report) or that the customer has disengaged (they no longer bother to report issues because they have mentally moved on from the product). Traditional health scoring cannot distinguish between these two very different situations.

The new health scoring framework for AI products replaces engagement proxies with value delivery metrics:

Outcome delivery rate trending: Is the percentage of successful AI completions improving, stable, or declining over the trailing 90 days? A stable or improving trend is healthy. A declining trend — even if the absolute level is still acceptable — is a warning signal. The trend is more important than the level because the trend reveals trajectory.

Work volume trends: Is the AI processing more or fewer work items per week than it was 90 days ago? Increasing volume indicates growing deployment. Stable volume indicates a mature deployment. Declining volume is the most actionable health risk signal — it means the customer is routing work away from the AI, which indicates dissatisfaction, loss of confidence, or organizational change.

Exception rate trends: What percentage of AI work items require human escalation or intervention? Rising exception rates indicate AI performance degradation or scope expansion beyond the AI's reliable capabilities. Falling exception rates indicate AI performance improvement and deeper customer confidence.

Economic value realised: The dollar value of outcomes delivered versus the contract value. A customer whose AI is delivering outcomes worth 5× the contract value is extraordinarily healthy. A customer whose AI is delivering outcomes worth 0.8× the contract value has a value realisation problem that will surface at renewal.

Stakeholder engagement quality: The seniority, frequency, and depth of meaningful engagement between the vendor CS team and the customer's commercial leadership. A customer whose CFO is engaged in the AI's value delivery is a customer whose renewal decision will be made on evidence. A customer where only the operations team is

engaged is a customer whose renewal decision may be made on budget review criteria that the operations team does not control.

Traditional Health Signals vs AI-Native Health Signals

Traditional SaaS signal	What it measures	Why it misleads for AI products	AI-native replacement
Monthly active users	Users logging into the product	Low MAU may mean the AI is working well (no need for human intervention) or the deployment has failed (humans have given up)	AI work volume processed: number of work items the AI handled in the month — regardless of human logins
Feature adoption breadth	Number of product features users have activated	AI activates features internally; users may never touch the interface at all	Output type coverage: number of distinct output types the AI produces in the month — breadth of AI activity
Login frequency	How often users access the product	Declining login frequency is a positive signal if the AI is resolving more autonomously; negative signal if users have disengaged	Exception rate: what percentage of AI work items required human escalation — declining is healthy, rising is risk
Support ticket volume (low)	Customers not reporting problems	Can mean no problems, or can mean customers have stopped trying to report problems — both look the same in the data	Outcome quality survey: structured monthly sample of AI outputs rated by the customer team — distinguishes 'no problems' from 'stopped reporting'
NPS score	Customer satisfaction with the product	NPS measures satisfaction with the interaction, not with the AI's performance — customers may be satisfied with the CS team but concerned about the AI	Value realisation score: (documented value delivered / contract cost) — the financial measure of whether the AI investment is paying off

Chapter Four — The Essentials

› Every traditional health signal either loses its meaning or actively misleads for autonomous AI products.

› Low logins: could be healthy (AI resolving autonomously) or critical (humans have given up) — without context, you cannot tell.

› Low support tickets: could be excellent performance or silent disengagement — the distinction requires proactive sentiment measurement.

› The new health signals are all output-side: work volume, exception rate, outcome quality, value realisation — measuring what the AI accomplishes, not how often humans click on it.

› Replace engagement proxies with value delivery metrics — the transition requires measurement infrastructure investment but produces signals that actually predict retention.

CHAPTER FIVE

AI-Powered Churn Prediction: Reading the Signals Before the Renewal

Agent-based churn models. Early warning indicators specific to AI product consumption.

Churn prediction for AI products requires early warning signals that differ fundamentally from SaaS churn models. Traditional SaaS churn prediction models identify patterns in historical data that precede customer departures: login decline, support ticket increases, feature disengagement, executive sponsor changes. These patterns are reliable because they reflect the underlying dynamic of SaaS churn — users who stop using the software stop paying for it.

AI product churn is different because the reasons for AI product churn are different. An AI product customer does not typically churn because they stop using the AI — they churn because the AI fails to deliver the outcomes they expected, because the AI's capability improvement has not kept pace with their evolving requirements, because a competitive alternative promises better outcomes, or because an internal change (reorganization, budget cut, leadership change) removes the business context that made the AI deployment valuable.

The early warning indicators specific to AI product churn:

Outcome quality plateau: When the AI's outcome delivery rate reaches a ceiling and stops improving — or worse, begins declining — the customer is experiencing the ceiling of the AI's current capability. If the customer's expectations continue to rise while the AI's performance is flat or declining, the gap will eventually trigger a commercial decision. The early warning is not the absolute performance level but the stagnation or decline.

Scope contraction: When the customer begins routing work items back to human processes that the AI was previously handling, the scope contraction is the clearest behavioral indicator of declining confidence in the AI. Scope contraction often precedes formal churn notification by 90–180 days — the customer is pulling back before they have made the formal decision to cancel.

Stakeholder displacement: When the internal champion for the AI deployment — the executive who sponsored the purchase, the VP who defended the budget, the director who drove the deployment — leaves the organization or changes roles, the AI deployment loses its internal advocacy. The successor does not have the same relationship with the product and may initiate a competitive review or apply greater scrutiny to the AI's value evidence.

Commercial engagement avoidance: When the customer's commercial leadership begins avoiding renewal conversations, declining QBR invitations, or responding slowly to commercial communications, the behavioral pattern suggests they are managing toward a non-renewal rather than engaging with the renewal decision collaboratively.

Integration instability: When the customer's integration with the AI begins experiencing frequent errors, data quality issues, or configuration problems that are not promptly resolved, the operational friction creates a churn risk that is separate from the AI's outcome quality. Integration instability signals that the AI deployment has not achieved the operational stability required for production reliance.

The at-risk identification model: each of these five signals is scored on a scale of 0–5 for severity. A customer scoring above 12 on the combined signals is at significant churn risk. A customer scoring above 8 is at elevated risk requiring proactive intervention. A customer scoring below 5 is healthy.

Specific AI company approaches to early churn detection:

Glean measures stakeholder displacement through its commercial engagement tracking: when the VP of Engineering who sponsored the Glean deployment changes roles or leaves, Glean's CS system automatically initiates an executive reconnection program — identifying the new stakeholder, scheduling a value review meeting, and preparing a customized value documentation package for the new decision-maker.

Intercom monitors scope contraction through its routing analytics: when the percentage of conversations being routed to AI assistance declines (customers manually routing inquiries to human agents rather than allowing AI to attempt resolution), the decline triggers a CS intervention workflow — the CSM investigates whether the routing change reflects a deliberate scope decision or a confidence issue with the AI's performance on specific inquiry types.

Harvey AI monitors outcome quality plateau through its attorney feedback system: when a law firm's attorneys begin systematically modifying Harvey's review outputs more substantially (taking longer to review, making more significant changes), the modification intensity signal triggers a CS review — the CSM examines whether the modifications reflect changing attorney preferences, a new document type that Harvey is not handling well, or a systemic quality issue.

Early Warning Indicators — AI Product Churn Prediction

Indicator	What it signals	Detection timing	Severity scoring	Intervention priority
Outcome quality plateau	AI performance has stopped improving; gap between customer	60–90 days before commercial impact	5 = significant plateau (>6 weeks flat or declining) · 3 = mild plateau (4–6	Score ≥ 4: immediate product/CS investigation

	expectations and AI capability is growing		weeks flat) · 1 = no plateau	and remediation plan
Scope contraction	Customer is routing work back to human processes; confidence in AI declining	30–120 days before churn notification	5 = >15% contraction in 30 days · 3 = 5–15% contraction · 1 = flat scope	Score ≥ 3: CSM escalation and intervention plan within 48 hours
Stakeholder displacement	Internal champion has left or changed roles; deployment lacks advocacy	Immediate — detectable from org chart monitoring	5 = champion has left; no identified successor · 3 = champion role change; relationship maintained · 1 = minor org change	Score = 5: executive reconnection programme initiated within 14 days
Commercial engagement avoidance	Customer avoiding renewal conversations; not responding to QBR invitations	60–120 days before renewal	5 = no response to 3+ commercial outreach attempts · 3 = selective engagement, declining QBRs · 1 = delayed but responsive	Score ≥ 3: escalate to CS manager; executive outreach initiated
Integration instability	Frequent errors or data quality issues in AI deployment; operational friction increasing	Real-time — detectable from integration monitoring	5 = >2% error rate for >7 days · 3 = elevated error rate, actively monitored · 1 = isolated incidents resolved quickly	Score ≥ 4: immediate engineering investigation; SLA credit if applicable

THE COMBINED RISK SCORE

Score each indicator 0–5; combined score > 12 = significant churn risk requiring CS Huddle

The five early warning indicators are not independent — they often appear in combination, and their combination is more predictive than any single indicator. Score each indicator on its severity (0–5) and sum the scores. Combined score 0–5: healthy. Combined score 6–8: monitor closely; investigate the specific driver. Combined score 9–12: elevated risk; proactive intervention required. Combined score 13+: significant churn risk; convene CS Agent Huddle immediately and prepare intervention plan.

The at-risk customer playbook is the operational guide for the CS intervention that prevents a deteriorating health situation from becoming a churn event. Effective interventions are early, specific, and evidence-based — not the generic "check-in call" that escalates only after the renewal conversation has already become adversarial.

The CS Agent Huddle for at-risk accounts (described in Book 2c as one of the seven huddle types) is the governance mechanism for managing high-stakes intervention decisions. When a customer's health indicators cross the at-risk threshold, the huddle assembles the relevant participants — CSM, CS manager, account executive, and relevant AI agents — to assess the situation and agree on the intervention approach.

The intervention playbook has three phases:

Phase 1 — Diagnosis (week 1–2 after risk threshold triggered): The CS Agent Huddle conducts a comprehensive review of the customer's health signals: outcome delivery rate trend, scope contraction evidence, stakeholder engagement quality, and any operational issues that may be contributing to the health deterioration. The diagnosis must identify the root cause of the health risk — is it AI performance, customer organization change, competitive threat, or commercial dissatisfaction — because each root cause requires a different intervention.

Phase 2 — Stabilisation (weeks 2–8): The intervention address the specific root cause. For AI performance issues: the product team investigates the specific failure modes, develops fixes or workarounds, and communicates a remediation plan with specific commitments. For organizational change: the CS team accelerates the executive reconnection program, prioritizes establishing the relationship with the new decision-maker, and presents a value documentation package that makes the AI's contribution visible to someone who has not experienced it personally. For competitive threat: the CSM prepares a competitive comparison analysis anchored on the customer's specific requirements, outcome data, and value evidence. For commercial dissatisfaction: the CSM prepares a renewal analysis showing value delivered versus contract cost with specific expansion recommendations.

Phase 3 — Rescue (if Phase 2 stabilisation fails): When the customer's commercial decision is already made and the formal churn notification is imminent, the rescue phase focuses on understanding the specific objections, addressing addressable concerns directly, and presenting a modified commercial structure that removes the commercial barriers while preserving the relationship. Rescue is expensive — it consumes significant senior time — and it is often unsuccessful if the underlying value delivery issue has not been resolved. The best time to intervene is in Phase 1, long before Phase 3 becomes necessary.

Specific intervention approaches that have demonstrated effectiveness:

Gong's executive reconnection program: when Gong identifies stakeholder displacement, the CS team automatically prepares a "value handoff package" — a document showing the departing executive's role in sponsoring the deployment, the specific outcomes the deployment has delivered, the current deployment scope, and the value evidence. This package is presented to the incoming decision-maker before they have formed an independent opinion of the deployment's value. The timing is critical: the executive reconnection must happen within 30 days of the stakeholder change, before the new decision-maker begins a competitive review.

ServiceNow's deployment health review: for customers whose deployment scope has contracted, ServiceNow conducts a structured deployment health review that examines each contracted workflow to identify which are performing well (and should remain AI-deployed), which are underperforming (and require AI improvement or scope adjustment), and which were never fully deployed (representing missed expansion opportunity). The health review converts a potentially adversarial scope contraction conversation into a collaborative optimization conversation.

Anthropic's outcome quality review: for enterprise customers experiencing AI performance plateau, Anthropic's CS team conducts a prompt engineering review — examining the customer's specific use cases, analyzing the patterns in outputs that are being modified or rejected, and implementing prompt optimizations that improve performance for the customer's specific domain. The prompt engineering review

demonstrates Anthropic's commitment to the customer's specific performance requirements and often produces measurable improvement within 30 days.

At-Risk Intervention Playbook — Three Phases				
Phase	**Timing**	**Activities**	**Responsible**	**Success indicator**
Phase 1: Diagnosis	Week 1–2 after risk threshold	CS Agent Huddle: review all five health signals, identify root cause (AI performance / org change / competitive threat / commercial dissatisfaction), agree intervention priority	CSM + CS Manager + AI risk agent	Root cause identified with confidence; intervention approach agreed by CS Agent Huddle
Phase 2: Stabilisation	Weeks 2–8	Root cause-specific intervention: AI performance → product investigation and fix commitment; org change → executive reconnection programme; competitive → evidence comparison; commercial → value documentation and renewal analysis	CSM leads; product/engineering/executive as required	UAS stabilisation (not declining); stakeholder engagement restored; specific issue addressed with evidence
Phase 3: Rescue	Only if Phase 2 fails	Understand specific objections; address addressable concerns; prepare modified commercial structure;	VP CS + Account Executive + Executive sponsor	At minimum: relationship maintained for future opportunity; ideally: modified

		executive relationship activation; retain relationship even if current term is lost		commercial structure agreed that retains revenue

Chapter Five — The Essentials

› Five early warning indicators: outcome quality plateau, scope contraction, stakeholder displacement, commercial engagement avoidance, integration instability.

› Combined risk score > 12: significant churn risk requiring immediate CS Agent Huddle.

› Stakeholder displacement requires the fastest response: executive reconnection programme must begin within 14 days of champion departure.

› Scope contraction precedes formal churn notification by 30–120 days — it is the most time-sensitive early warning signal.

› The three-phase intervention (diagnosis → stabilisation → rescue) is most effective when triggered at Phase 1 — rescue is expensive and often unsuccessful.

PART THREE

Expansion and Advocacy

How outcome evidence drives growth. The renewal as value documentation. The advocacy flywheel.

CHAPTER SIX

Expansion Motion Design for AI Products

How to build expansion into the product, the contract, and the CS model.

Expansion motion design for AI products is the commercial architecture of growth from within the installed base. It is the operational expression of the principle that AI products should grow revenue from existing customers automatically — through the elastic contract mechanisms described in Book 8 and through the deliberate expansion conversations that CS drives when the outcome evidence justifies them.

AI product expansion has four natural motions that the CS team must be equipped to execute:

Depth expansion: deploying the AI to handle more of the work within a workflow where it is already deployed. A contract review AI deployed for NDAs can expand to master service agreements, then to employment agreements, then to vendor contracts. Each expansion deepens the AI's role in the customer's workflow and generates more outcome evidence. Depth expansion conversations are straightforward: the CSM presents the outcome data from the current deployment and asks whether the customer would like the AI to handle the next document category.

Breadth expansion: deploying the AI in new departments or teams that have the same workflow needs as the initial deployment team. A customer service AI deployed in the customer support team can expand to the technical support team, then to the professional services team, then to the sales support team. Breadth expansion conversations require the CSM to present the outcome data from the initial deployment in terms that resonate with the new team's specific needs and workflows.

Tier expansion: moving the customer from a lower-quality tier to a higher-quality tier within the same use case. A customer using the Standard tier of a contract review AI (85% recall guarantee) can expand to the Professional tier (95% recall guarantee) when their legal team's confidence in the AI has grown and their use cases have evolved to higher-stakes agreements. Tier expansion conversations are value conversations: the CSM presents the outcome comparison between Standard and Professional tier performance and the economic value of the quality improvement for the customer's specific use case mix.

Cross-product expansion: deploying the vendor's AI products in adjacent use cases where the customer has equivalent workflow needs. An enterprise customer who deployed a contract review AI can expand to a contract management AI, a legal research AI, or a compliance monitoring AI. Cross-product expansion conversations require the CSM to connect the customer's demonstrated AI success in one area to the potential for equivalent success in adjacent areas.

The outcome data requirements for each expansion motion:

Depth expansion: the CS team needs outcome data showing the AI's performance on the current document/workflow types, and market data showing the AI's average performance on the proposed expansion document/workflow types. The conversation is: "Your AI is processing NDAs at 92% recall. Our average recall on MSAs for customers in your industry is 89%. Based on your MSA volume of approximately 150 per quarter, the AI would save your team approximately 60 hours per quarter."

Breadth expansion: the CS team needs the outcome data from the initial team's deployment in terms that translate to the expansion team's context. If the initial deployment is in customer support, the breadth expansion to technical support requires technical support-specific metrics: time to resolution for technical inquiries, escalation rate to engineering, customer satisfaction for technical support interactions.

Tier expansion: the CS team needs comparative outcome data between the current tier and the proposed tier, translated to the specific economic impact for this customer's use case mix. The conversation is not "Professional tier has better recall" — it is "For your specific mix of employment agreements and supplier contracts (which have higher legal stakes than your NDA volume), Professional tier recall improvement from 85% to 95% reduces your legal team's re-review requirement by approximately 8 hours per week."

Cross-product expansion: the CS team needs the outcome story from the initial product deployment combined with the outcome potential for the adjacent product. The connection must be made explicitly: "Your Harvey deployment for contract review has saved your legal team 12 hours per week on NDA review. Your legal team also spends

approximately 8 hours per week on regulatory compliance research — a workflow that Harvey's Legal Research product is designed for. Based on comparable deployments, the research AI reduces this time to approximately 2 hours per week."

Four Expansion Motions — Design and Data Requirements				
Motion	**Description**	**Required outcome data**	**Conversation anchor**	**Conversation level**
Depth expansion	Deploy AI to more work items within the same workflow category currently in use	Current category: outcome delivery rate, volume, attorney/operator time saved per item. Proposed category: benchmark delivery rate from comparable customers, volume estimate, projected time savings	'Your NDA recall is 93%. Our benchmark for MSA recall in your sector is 91%. At your MSA volume of 150/quarter, that is 90 additional attorney hours saved quarterly — worth $XX at your rates.'	Operational team + legal/department head
Breadth expansion	Deploy AI in new department or team with equivalent workflow needs	Current team: documented ROI, outcome metrics. Target team: workflow volume estimate, current cost model, comparable deployment benchmarks	'Your CS team's Fin resolution rate is 68%, saving 840 hours/month. Your technical support team has similar inquiry types. Our benchmark for technical support resolution is 62%. At their ticket volume, that is an additional 720 hours/month.'	New team operational lead + department head
Tier expansion	Upgrade from lower quality tier to higher quality tier for higher-stakes use cases	Current tier performance on the customer's actual use case mix. Incremental quality improvement of higher tier. Economic impact of quality	'Your Standard tier recall for NDAs is 87%. Professional tier averages 95% on NDAs. For your 200 monthly NDAs, the 8% improvement reduces attorney	Legal/operations lead + finance

		improvement for the specific use case mix.	re-review by approximately 6 hours per month — worth $XX. The tier upgrade is $XX/month.'	
Cross-product expansion	Deploy vendor's AI in adjacent use cases where the customer has equivalent workflow needs	Current product ROI and deployment success. Adjacent product category: benchmark outcome data, deployment timeline, integration requirements.	'Your Harvey contract review deployment is saving your legal team 3,400 hours/year at 9.2× ROI. Your compliance monitoring workflow currently consumes 200 attorney hours per month. Harvey's Regulatory Monitoring product averages 140 hours saved per month in comparable deployments. The projected ROI on a $180K contract is 4.7×.'	General Counsel + CFO

EXPANSION IS EVIDENCE-BASED — NOT RELATIONSHIP-BASED

The most effective expansion conversations lead with the customer's own outcome data, not with product features or relationship goodwill

The CSM who says 'we think you'd get a lot of value from our contract management product' is having a relationship conversation. The CSM who says 'your Harvey deployment has saved your legal team 3,400 hours this year. Your contract management process currently consumes approximately 200 paralegal hours per month on administrative tasks — creating contracts from templates, tracking signatures, monitoring deadlines. Our Contract Management AI averages 140 of those hours saved per month for comparable legal departments, at a projected ROI of 5.2× on a $180K annual contract. Would you like to see the data from three comparable deployments?' is having an evidence conversation. The second conversation does not require the CSM to be the customer's best friend. It requires the CSM to have done the analysis. Evidence-based expansion is both more persuasive and more scalable than relationship-based expansion.

Chapter Six — The Essentials
› Four expansion motions: depth (more work types), breadth (more teams), tier (higher quality), cross-product (adjacent use cases).
› Every expansion conversation must be anchored in the customer's own outcome data from the current deployment plus benchmark data for the proposed expansion.
› Depth expansion is the most accessible: start with the use case where AI performance is strongest and the expansion evidence is clearest.
› Cross-product expansion commands the largest commercial conversations: leads with the demonstrated ROI from the current product and projects the equivalent ROI for the adjacent product.
› The expansion signal agent identifies expansion-ready customers automatically — the CSM's job is to convert the signal into a prepared, evidence-based conversation.

CHAPTER SEVEN

The CS-Led Renewal: Using Health Data to Price and Negotiate

How CSMs use health data in renewal conversations. Expansion anchoring from outcome data.

The CS-led renewal is the process through which customer success transforms outcome data into commercial evidence for the renewal decision. In AI-native products, the renewal is not primarily a relationship decision — it is an economic decision. The customer's CFO and business unit leader are asking a specific question: did the AI investment pay off, and will it continue to pay off in the next contract term?

The CS team's job in the renewal process is to answer that question with evidence so specific and so compelling that the answer is obvious.

The renewal preparation package has five components:

Component 1 — Outcome summary. A quantified summary of the AI's outcomes delivered in the current contract term: the number of work items processed, the outcome delivery rate, the volume of successfully completed outcomes, and the categories of outcomes delivered. This summary should be expressed in terms that are meaningful to the customer's business — not in AI performance metrics, but in business activity metrics. "Your Harvey AI reviewed 847 contracts in the past 12 months, with a 93% first-pass acceptance rate" is less compelling than "Your Harvey AI reviewed 847 contracts, identifying 2,341 material issues across 623 agreements, in an average of 2.3 hours per contract rather than the 8 hours your attorneys previously required."

Component 2 — Economic value documentation. The dollar value of the outcomes delivered, calculated at the customer's specific economics. The economic value documentation translates outcome metrics into financial terms: hours saved × fully-loaded cost per attorney hour = labor value. Improved cycle time × contract volume × daily revenue cost of delayed contracts = cycle time value. Avoided legal disputes × average dispute resolution cost = risk reduction value. Each value component is calculated explicitly, sourced from the customer's own data where possible, and benchmarked against comparable deployments where the customer's data is incomplete.

Component 3 — ROI calculation. The explicit comparison between the total economic value documented in Component 2 and the contract value paid in the current term. The ROI statement should be specific and dramatic: "Your total documented value from the Harvey deployment in the past 12 months was $1.8M in labor savings and $400K in avoided dispute costs — $2.2M total. Your contract value was $240K. That is a 9.2× ROI." A 9.2× ROI statement is not a relationship argument for renewal. It is a financial argument. The CFO who sees a 9.2× ROI does not need to be sold on the renewal — they need to be sold on the expansion.

Component 4 — Next-term opportunity. The evidence-based projection of value available in the next contract term, including expansion opportunities that have been identified from the current deployment's data. This component converts the renewal

from a maintenance decision to a growth decision. "In the next term, deploying Harvey in your M&A due diligence workflow would add an estimated $800K in attorney time savings annually, based on your deal volume and our comparable M&A deployments. The expanded contract at $360K annually would yield a projected 8.3× ROI including the expansion."

Component 5 — Risk and concern acknowledgment. An honest acknowledgment of any performance issues, service failures, or capability limitations that occurred during the current term, with the remediation actions taken and the evidence that the issues have been resolved. Renewal packages that ignore problems lose credibility. Renewal packages that address problems directly and demonstrate resolution build the trust that makes the commercial argument credible.

Renewal Preparation Package — Five Components				
Component	**Content**	**Source**	**Compelling format**	**Audience**
1. Outcome summary	Work items processed, outcome delivery rate, categories of outcomes delivered — in business activity terms, not AI performance terms	Outcome tracking system; CS health monitoring agent	Table: work item type / volume / delivery rate / attorney-equivalent time per item / total hours saved. Narrative: what these numbers mean for the customer's operations.	Legal/operations team + department head
2. Economic value documentation	Dollar value of each outcome category at the customer's specific economics: hours saved × fully-loaded rate; cycle time improvement × contract volume	Customer-provided economic inputs + internal value calculation model	Financial table: value component / basis / calculation / dollar value. Total: $X in documented value.	CFO + General Counsel + VP Operations

	× daily revenue; risk reduction × average dispute cost			
3. ROI calculation	Contract value vs total documented economic value; explicit ROI statement	Economic value documentation ÷ contract value	Single headline: '$2.2M in documented value from a $240K contract = 9.2× ROI.' Supporting exhibit: component breakdown.	CFO + CEO/GC for strategic accounts
4. Next-term opportunity	Evidence-based projection of value available in the next term; expansion opportunities identified from outcome data; recommended contract structure	Expansion signal analysis from outcome data + benchmark data from comparable deployments	Section: 'What is available in the next term' with specific expansion opportunities, projected value, and recommended contract.	CFO + GC + department heads considering expansion
5. Risk and concern acknowledgment	Honest review of any performance issues, SLA misses, or service failures; remediation taken; evidence that issues are resolved	CS incident history + SLA performance reports	Brief, direct section: 'These issues occurred, these actions were taken, this is the evidence of resolution.' Nothing hidden.	All — trust-building component

Chapter Seven — The Essentials

› The renewal preparation package has five components: outcome summary, economic value documentation, ROI calculation, next-term opportunity, risk acknowledgment.

› The ROI statement is the most commercially powerful element: 9.2× ROI is not a relationship argument — it is a financial argument that the CFO can evaluate independently.

› The next-term opportunity converts the renewal from a maintenance decision to a growth decision: the customer is renewing and expanding, not just retaining.

› The risk acknowledgment builds trust — renewal packages that ignore problems lose credibility; packages that address them directly demonstrate accountability.

› The preparation package should be ready 60 days before renewal; the CFO conversation should be booked 45 days before renewal.

CHAPTER EIGHT

Turning Customers into Advocates: Reference, Case Study, and Co-Innovation

The advocacy flywheel for AI products. Outcome stories as sales assets.

Customer advocacy in the AI economy is the process of converting customers whose AI deployments have delivered documented, significant value into active promoters of the AI product to prospective buyers in their network.

The advocacy flywheel for AI products is particularly powerful because AI-delivered value is both compelling and credible. A customer who says "Harvey reviewed 847 contracts this year, saved our legal team 3,400 hours, and identified issues that prevented two significant disputes from reaching litigation" is a more persuasive advocate than any marketing content because their evidence is specific, their business context is real, and their experience is independently verifiable.

The four types of customer advocacy and how to cultivate each:

Reference conversations: direct peer conversations between the vendor's prospects and the vendor's satisfied customers, typically in the context of a prospect's evaluation. Reference conversations are the most persuasive form of advocacy because they are completely unmediated — the prospect asks the reference the specific questions they care about, and the reference answers from direct experience. AI product reference conversations are especially valuable when the reference customer can quantify their

outcomes: "We evaluated Harvey and three competitors. Harvey was not the cheapest. But after 12 months, our cost per contract review has gone from $800 to $175 and our turnaround time has gone from 21 days to 4 days. I can give you the data if you want it."

Case studies: written documentation of the customer's AI deployment story, from initial deployment through current value delivery, with specific metrics and economic outcomes. AI product case studies are differentiated by specificity — the case studies that generate the most prospect interest are those with specific numbers (outcomes delivered, time saved, cost reduced, revenue accelerated) rather than general statements about improved productivity.

Industry conference and community contributions: customers who speak at industry conferences, write about their AI deployment experiences in professional publications, or contribute to practitioner communities are creating advocacy that reaches large audiences and establishes the AI product as legitimate in the eyes of peers who have not yet adopted it. Encouraging and supporting these contributions — by helping customers prepare their presentations, by co-authoring thought leadership, by facilitating speaking opportunities — is one of the highest-leverage advocacy investments a CS team can make.

Co-innovation partnerships: customers who work with the vendor to develop new capabilities, test new features, or define new use cases are the most deeply invested advocates. Co-innovation partners have an ownership stake in the product's development and communicate that ownership when they discuss the product with peers. The co-innovation relationship is both the deepest form of advocacy and the deepest form of workflow lock — a customer who has contributed to the product's development is unlikely to replace it with a competitor.

The outcome evidence library: the CS team's most valuable advocacy tool is a structured library of customer outcome data, organized by industry, use case, and company size, that sales teams can use to construct value projections for prospects and that customers can use to benchmark their own deployments. This library is built incrementally — every

renewal preparation package contributes data to the library, which makes each successive value conversation more compelling.

Four Advocacy Types — Cultivation and Commercial Value				
Advocacy type	**What it requires**	**How to cultivate**	**Commercial value**	**Investment required**
Reference conversations	Customer willing to take peer calls; able to speak to specific outcomes	Identify customers with documented 5×+ ROI; ask during renewal when relationship is strongest; provide preparation support for reference calls	Highest conversion impact of any sales asset — peer conversations are trusted above all vendor-produced content	30 mins/reference to prepare; ongoing relationship maintenance
Written case studies	Customer willing to be named; specific outcome metrics; approval process complete	Propose during renewal or post-expansion when customer satisfaction is at peak; provide ghostwriting support; share drafts promptly for approval	Used across entire sales cycle; amplified through marketing; long shelf life — lasts until product significantly changes	4–8 hours to produce; 2–4 weeks for customer approval
Conference and community	Customer willing to speak publicly about AI deployment; interesting outcome story	Help customers submit to relevant conferences; co-author thought leadership; provide data support for their presentations	Reaches large audiences; establishes AI product as legitimate in the eyes of peers; more trusted than vendor marketing	Significant: conference preparation support, co-authoring, speaking coaching
Co-innovation partnership	Customer willing to develop new capabilities with vendor; has interesting edge case or expansion requirement	Invite customers with specific domain needs to join product advisory council; prioritize their requirements in roadmap; attribute contributions	Deepest advocacy (owner mindset) and deepest workflow lock simultaneously — co-innovation partner is almost never replaced by a competitor	High: product team time + CS time + executive relationship

		publicly		

Chapter Eight — The Essentials

› Reference conversations are the highest-conversion advocacy asset — peer credibility exceeds all vendor-produced content.

› The outcome evidence library is the CS team's most scalable advocacy tool: structured outcome data by industry, use case, and company size.

› Co-innovation partnerships create simultaneously the deepest advocacy (ownership mindset) and the deepest workflow lock (switching would abandon co-created capability).

› The best time to ask for advocacy is during renewal when satisfaction is highest and the ROI evidence is freshest.

› The advocacy flywheel: documented outcomes → reference conversations → case studies → conference talks → co-innovation → stronger evidence library → more compelling references.

PART FOUR

CS Operations in the AI Era

AI agents for CS. The CSM role. Metrics and the CS P&L.

CHAPTER NINE

AI Agents for Customer Success Operations

CS agents: health monitoring · QBR prep · risk alert · expansion signal · value documentation.

AI agents for customer success operations represent one of the highest-impact applications of AI in the commercial function — and one of the most underdeployed. The reasons for underdeployment are cultural: CS teams have historically been

relationship-oriented functions where personal connection and human judgment are valued highly, and there is resistance to the idea of AI agents playing a role in what should be human relationships.

This resistance is misapplied. AI agents in CS do not replace the relationship — they enable the relationship to be focused on what matters. When AI agents handle the data assembly (pulling together the outcome metrics, calculating the ROI, identifying expansion opportunities from usage patterns), the CSM can focus the conversation on what the data means for the customer's business and what the customer's strategic direction means for their AI deployment. The AI does the analytical work; the human does the relationship work.

Four specific AI agent types that generate the highest value in CS operations:

The health monitoring agent continuously monitors every customer's health signals against the five dimensions of the AI Customer Health Architecture. It runs continuously — not in a weekly manual review cycle — and alerts the CSM within hours of a significant health signal change. When a customer's outcome delivery rate drops by more than 5% in a week, or when a customer's work volume through the AI declines by more than 10% in a month, the health monitoring agent generates an alert with the specific signal data, the historical trend context, and a recommended intervention priority. The CSM receives a structured alert rather than discovering the issue weeks later during a scheduled review.

The QBR preparation agent assembles the complete data package for quarterly business reviews: pulling the outcome metrics for the quarter, calculating the ROI, identifying expansion opportunities from usage patterns, benchmarking the customer's performance against comparable deployments, and drafting the first version of the QBR narrative. A QBR preparation that previously consumed 4–6 hours of CSM time is reduced to a 30-minute review and customization of the agent's first draft. The CSM's time is spent on the strategic interpretation and the relationship conversation, not on the data assembly.

The risk alert agent specifically monitors the early warning indicators of churn risk described in the health scoring chapter: outcome quality plateau, scope contraction, stakeholder displacement, commercial engagement avoidance, and integration instability. When any of these signals reaches the at-risk threshold, the risk alert agent initiates the intervention workflow — notifying the CSM and CS manager, scheduling the CS Agent Huddle, and preparing the initial diagnosis package. The agent also queries comparable at-risk cases from the historical customer base to identify the intervention approach that has been most effective in similar situations.

The expansion signal agent specifically monitors customers for evidence that they are ready for expansion conversations: outcome delivery rate above threshold for more than 60 days, work volume growing toward the contracted capacity ceiling, stakeholder engagement quality improving (suggesting renewed executive interest), and absence of health risk signals. When the expansion signal agent identifies an expansion-ready customer, it generates a customized expansion brief for the CSM: the current deployment metrics, the expansion opportunity with projected value, and the recommended expansion conversation approach.

The value documentation agent assembles the renewal preparation package: pulling the contract history, calculating the cumulative outcomes delivered, translating outcomes to economic value using the customer-specific economic model, comparing the delivered value to the contract cost, and generating the initial draft of the renewal preparation document. The CSM's role in the renewal preparation process shifts from data assembly (which the agent handles) to strategic framing and relationship conversation (which the CSM handles best).

The Five CS Operations Agents — Reference

Agent	What it does	Trigger	Primary output	Human action required	Escalation criteria
Health monitoring agent	Continuously monitors all five health signals for every customer;	Continuous (hourly signal check);	Daily health briefing per CSM portfolio; immediate alert	Review alerts; investigate flagged changes;	UAS below 40 O two or more ris indicators abov threshold → C

	computes UAS; identifies signal changes	alerts on significant changes	on UAS drop > 10 points or any single signal reaching risk threshold	update intervention plans	Agent Huddle
QBR preparation agent	Assembles complete QBR data package: outcome metrics, ROI calculation, expansion opportunities, benchmark comparisons, first-draft narrative	Triggered 4 weeks before scheduled QBR	Complete QBR deck first draft with data, calculations, narrative, and recommended conversation agenda	CSM review and customization (30–60 mins); add relationship context and strategic framing; prepare CFO-level talking points	Significant dat gaps or unusua patterns tha require interpretation - CSM investigatio before QBR
Risk alert agent	Monitors five early warning churn indicators; calculates combined risk score; triggers intervention workflow when thresholds exceeded	Continuous (daily check); immediate alert when combined score crosses 9	Risk alert with: signal details, risk score, historical context, recommended intervention priority, comparable historical cases	Acknowledge alert within 24h; classify as valid or false positive; initiate appropriate intervention phase	Combined score 12 → automati CS Agent Huddl scheduling
Expansion signal agent	Identifies customers with positive expansion indicators: high UAS, growing outcome delivery, stakeholder engagement quality improving, approaching contracted capacity ceiling	Weekly analysis of portfolio; immediate alert when expansion-ready threshold crossed	Expansion brief: current deployment metrics, expansion opportunity with projected value, recommended conversation approach, comparable deployment benchmarks	Review expansion brief; customise for specific customer context; schedule expansion conversation	None - expansion signa agent produces recommendation CSM decide whether to pursu
Value documentation agent	Assembles renewal preparation package: contract history, outcome delivery summary, economic value calculation, ROI	Triggered 60 days before contract expiry	Complete renewal preparation package first draft	CSM review (60–90 mins); add strategic framing; prepare CFO talking points; check	Significant calculation questions o unusual custome circumstances - CSM/CS Manage review befor

	statement, next-term opportunity analysis, first draft of renewal narrative			customer-specific economic inputs for accuracy	finalisation

THE AGENT ELEVATION EFFECT

AI agents in CS do not replace the relationship — they enable the relationship to be focused on what matters

When health monitoring agents track 40 customers' UAS continuously rather than the CSM reviewing a spreadsheet weekly, the CSM is freed from data assembly to focus on the relationship and strategic conversations. When the QBR preparation agent produces the first draft of the QBR in two hours rather than the CSM spending six, the CSM is freed from slide-making to focus on the QBR conversation. The agent elevation effect is real: CSMs who deploy all five agents report managing 40–50% larger portfolios at higher quality than CSMs without agent support. The agents are multipliers, not replacements.

Chapter Nine — The Essentials

› Five CS operations agents: health monitoring (continuous UAS tracking), QBR prep (data assembly), risk alert (early warning), expansion signal (growth identification), value documentation (renewal package).

› The agent elevation effect: CSMs with full agent support manage 40–50% larger portfolios at higher quality.

› Health monitoring agent runs continuously — detects UAS changes in hours rather than in weekly manual reviews.

› QBR preparation agent reduces QBR data assembly from 4–6 hours to 30–60 minutes, allowing CSM to focus on the strategic conversation.

› Expansion signal agent converts passive portfolio management into an active growth identification system — no expansion opportunity falls through the cracks.

CHAPTER TEN

The CSM Role in the Agent Era: What Changes, What Stays

The skills that matter when the product runs autonomously. The human judgment layer.

The CSM role in the AI era requires a different profile from the SaaS CSM profile. The shift is from product educator to value analyst and commercial advocate — a shift that requires new skills, new tools, and a new understanding of what customer success means for autonomously operating products.

The SaaS CSM profile was centered on product knowledge and relationship management. The effective SaaS CSM knew the product deeply (could train users, answer feature questions, navigate the UI fluently), maintained strong personal relationships with the customer's operational team, was attentive to user engagement signals, and could navigate the renewal conversation as a relationship maintenance exercise.

These skills remain necessary but are no longer sufficient in the AI era. The AI CSM profile adds three new requirement clusters:

Data literacy and analytical capability. The AI CSM must be able to read and interpret outcome delivery data, calculate ROI from customer-specific economic inputs, identify patterns in usage and outcome data that indicate health or risk, and communicate analytical findings in business language to the customer's commercial leadership. This is not data science capability — it is business analytics capability: the ability to take numbers from a dashboard and turn them into a business narrative.

The CSM who says "your Harvey deployment has a 93% first-pass acceptance rate and an average review time of 2.3 hours" is doing data reporting. The CSM who says "your Harvey deployment is saving your legal team 3,400 hours this year — at your fully-loaded attorney cost of $165,000 per year, that is $557,000 in labor value from a $240,000 contract, and based on your M&A activity in the pipeline, the Harvey Due Diligence expansion would add another $320,000 in value" is doing business analytics. The second conversation is the one that gets CFO attention and drives expansion.

Economic fluency. The AI CSM must be able to have commercial conversations with CFOs and business unit leaders — conversations that are grounded in the language of

business outcomes and financial value rather than in the language of product features and usage metrics. This requires understanding how the customer's business makes money, what the economic stakes of the workflows the AI is processing are, and how to express the AI's contribution in terms that the CFO's decision framework recognizes.

Domain expertise. The AI CSM who serves legal AI customers needs to understand the economics of legal work — how law firms are structured, how corporate legal departments are evaluated, what an attorney's time is worth in different contexts, what the regulatory constraints on AI-generated legal work are, and what the senior partners care about when evaluating the AI's contribution. The AI CSM who serves healthcare customers needs to understand the clinical documentation requirements, the CMS billing implications of documentation quality, and the physician workflow in which the AI is deployed.

The transformation of the CSM's day-to-day work:

Old pattern: log in to the customer's account, review the usage dashboard, note which users are active and which are not, send a check-in email to the power user, prepare a feature adoption summary for the QBR.

New pattern: review the health monitoring agent's daily briefing, investigate any flagged signals, prepare expansion analysis for the customer identified as expansion-ready by the expansion signal agent, review the QBR draft prepared by the QBR agent and customize the narrative for the specific stakeholders attending, review the renewal preparation document drafted by the value documentation agent and prepare the CFO-level conversation points.

The time savings from AI agent support allow the CSM to serve more customers at a higher quality level — not by reducing relationship investment, but by ensuring that the relationship investment is focused on the highest-value activities rather than on data assembly that AI can do more quickly and more comprehensively.

CSM Skills: SaaS Era vs AI Era

Skill cluster	SaaS CSM requirement	AI CSM requirement	Development path
Product knowledge	Deep knowledge of all product features; ability to train users and answer feature questions	Understanding of AI capabilities and limitations; ability to explain AI performance metrics; knowledge of outcome definition and measurement	Reorient product training from 'how to use features' to 'how to interpret AI performance and outcomes'
Relationship management	Strong personal relationships with operational team; regular check-in cadence; user adoption advocacy	Relationships at multiple levels (operational, departmental, financial); ability to navigate organizational change; executive relationship development	Expand from single-level operational relationships to multi-level commercial relationships; add CFO-level communication skills
Data literacy	Ability to read basic product usage dashboards; identify users at risk of disengagement	Ability to interpret outcome delivery metrics; calculate ROI from customer economics; identify patterns in health signals; communicate analytical findings in business language	Analytics training: financial modeling, ROI calculation, statistical interpretation of AI performance trends
Domain expertise	General enterprise software knowledge; customer's industry at overview level	Deep domain knowledge specific to the AI product's application area: legal, healthcare, financial services, engineering, etc.	Industry-specific training; sector certification; customer industry conference participation
Commercial acumen	Renewal management and QBR delivery; basic expansion identification	Value documentation and ROI communication; CFO-level commercial conversations; expansion brief preparation; contract structure recommendation	Commercial training: value-based selling, financial conversation skills, renewal negotiation

Chapter Ten — The Essentials

› Three new skill clusters required for AI CSMs: data literacy and analytical capability, economic

fluency, and domain expertise.

› The old pattern (review usage dashboard, check on power users, prepare feature adoption summary) is replaced by the new pattern (review health agent briefing, analyze expansion opportunities, review renewal package draft, prepare CFO talking points).

› What stays: relationship management, customer advocacy, strategic conversation, human judgment on complex situations.

› What changes: data assembly, feature education, usage monitoring, basic check-in conversations — all partially or fully handled by AI agents.

› The transition requires active CSM development: the skills that made a great SaaS CSM are necessary but not sufficient for AI-native CS excellence.

CHAPTER ELEVEN

CS Metrics, Dashboards, and the CS P&L

NRR · GRR · Time-to-value · Health score distribution · CS cost per retained dollar.

Customer success metrics in the AI era require a complete reconstruction around the signals that predict renewal, expansion, and advocacy in AI-native products. The old metric set — MAU, feature adoption, NPS, support ticket volume — must be replaced or supplemented with outcome-based metrics that measure what actually matters.

The CS P&L: making CS economically accountable

The CS P&L measures the economic contribution of the CS function to business performance — not just the retention and expansion metrics that indicate what CS accomplished, but the economic value CS created through those accomplishments.

CS revenue contribution: the total ARR that CS-managed accounts represent, broken down by retained, expanded, and contracted. This is not just the renewal rate — it is the dollar value of what CS retained and grew versus what CS allowed to contract or churn.

CS cost per retained dollar: the total CS team cost (salaries, tools, benefits, overhead) divided by the total ARR retained in CS-managed accounts. This metric tells management whether the CS investment is economically justified relative to the retention it produces. A CS team that costs $2M annually and retains $40M in ARR has a cost per retained dollar of 5 cents — an excellent return. A CS team that costs $2M annually and retains $20M in ARR has a cost per retained dollar of 10 cents — still positive, but worth examining whether the CS model can be made more efficient.

Net Revenue Retention (NRR): the metric that most directly captures the commercial value of effective CS. NRR above 120% indicates that the CS team is not just retaining customers but growing revenue from them — a sign that the outcome evidence is compelling enough to drive expansion. NRR between 100% and 120% indicates healthy retention with moderate expansion. NRR below 100% indicates that CS is not fully offsetting customer contraction and churn — either the AI is not delivering sufficient value, or the CS model is not converting value delivery into commercial evidence.

Customer Health Score distribution: the percentage of the CS portfolio in each health tier (healthy above 70 UAS, at-risk between 40–70, critical below 40). The health score distribution tells management whether the CS team's proactive health monitoring is working — a healthy portfolio should have 70%+ of customers in the healthy tier, 25% in the at-risk tier receiving proactive attention, and less than 5% in the critical tier requiring urgent intervention.

Time to first value: the time from contract signing to the customer's first verified, meaningful outcome from the AI deployment. Time to first value is the most important metric for the onboarding phase — it predicts long-term retention better than any other early-stage signal. Customers who achieve first value within 30 days retain at significantly higher rates than those who take 90 days. The CS team's ability to accelerate time to first value — through deployment support, integration assistance, and use case configuration — is the most direct lever they have on long-term retention.

Expansion revenue per CSM: the total expansion ARR generated by each CSM's portfolio, reflecting both the quality of the expansion motion and the health of the

portfolio. A CSM with 20 customers at an average expansion rate of 15% is generating $3M in expansion ARR on a $20M portfolio — a strong result. A CSM with 20 customers and no expansion is indicating either portfolio health problems or an underdeveloped expansion motion.

AI CS Metrics Reference — Complete KPI Library

Metric	Definition	Formula	Target benchmark	Reporting cadence
Net Revenue Retention (NRR)	Total revenue from beginning-of-period customers at end of period, divided by beginning-of-period ARR × 100	(Beginning ARR + Expansion – Contraction – Churn) / Beginning ARR × 100	≥ 120% for healthy AI product CS; ≥ 100% for minimum acceptable performance	Monthly; trend is most important
Gross Revenue Retention (GRR)	Percentage of beginning ARR retained, excluding expansion	(Beginning ARR – Contraction – Churn) / Beginning ARR × 100	≥ 90%; below 85% indicates systemic value delivery or CS model problem	Monthly
Time to first value	Days from contract signing to first verified meaningful outcome	Contract signed date to first outcome milestone date; track by cohort	≤ 30 days: excellent. 30–60 days: acceptable. > 60 days: requires onboarding improvement	Per cohort; compare cohort-to-cohort for onboarding improvement tracking
Health score distribution	Percentage of CS portfolio in each UAS tier (healthy/at-risk/critical)	Count of customers in each UAS tier / total customers × 100	Healthy > 70%: above 70 UAS · At-risk manageable: 25–30% in 40–70 range · Critical < 5%: below 40 UAS	Weekly
Expansion revenue per CSM	Total expansion ARR generated by each CSM's portfolio in the period	Sum of expansion ARR from CSM's accounts / CSM count	$300K–$600K expansion ARR per CSM annually for high-performing AI	Quarterly

			CS teams	
CS cost per retained dollar	Total CS team cost divided by total ARR retained from CS-managed accounts	Total CS team cost (salary + tools + overhead) / Retained ARR	≤ 8 cents per retained dollar: efficient · ≤ 5 cents: high-performing · > 12 cents: CS model review required	Quarterly
UAS by portfolio tier	Average UAS across each CSM's portfolio, segmented by account tier	Average UAS for strategic / enterprise / mid-market / SMB accounts separately	Strategic: average UAS > 75 · Enterprise: > 70 · Mid-market: > 65	Monthly
Advocacy pipeline	Number of customers in each advocacy stage: reference-ready, case-study-in-progress, speaker/author, co-innovation partner	Count by stage in advocacy pipeline	1 reference-ready per 8 accounts; 1 case study per 12 accounts; 1 speaker per 25 accounts	Quarterly

Chapter Eleven — The Essentials

› The CS P&L has three primary metrics: CS revenue contribution, CS cost per retained dollar, and NRR.

› Time to first value is the most important leading indicator of long-term retention: customers achieving first value within 30 days retain at 95%+ rates at 12 months.

› Health score distribution is the operating health metric: > 70% of portfolio in healthy tier, < 5% in critical tier.

› CS cost per retained dollar benchmarks: ≤ 5 cents is high-performing; > 12 cents requires CS model review.

› Report CS metrics to leadership monthly with trend analysis — the CS team's commercial contribution is most visible when revenue metrics are connected to operational metrics.

CHAPTER TWELVE

Onboarding for AI Products: Getting to First Value Fast

The AI onboarding journey. What 'activated' means when the product is autonomous.

Onboarding for AI products has a specific success definition that differs from SaaS onboarding: the customer is "activated" when the AI has been deployed in a production workflow, is processing real work, and has delivered at least one verified, meaningful outcome. Not when users have completed training. Not when the integration is configured. When the AI has done something valuable.

This definition makes the onboarding process more demanding than SaaS onboarding because production deployment requires more than user training — it requires workflow integration, data connection, quality validation, and the operational confidence to trust the AI with real work.

The AI onboarding journey has five stages:

Stage 1 — Technical deployment (weeks 1–3): completing the technical setup required for the AI to begin processing work. For a contract review AI, this includes: connecting the AI to the customer's document management system, configuring the review playbook (the specific clause types and risk thresholds that the customer's legal team cares about), and running the AI on a test set of documents to validate that the configuration produces outputs in the expected format. The CS team's role is to manage the deployment project, coordinate between the vendor's technical team and the customer's IT team, and ensure that the deployment proceeds on schedule.

Stage 2 — Pilot operation (weeks 3–8): running the AI in a controlled setting where human review is applied to all AI outputs. The pilot operation validates the AI's performance on the customer's specific document types and workflow requirements before production deployment. The CS team's role is to collect the attorney or operational team's feedback on each AI output, identify patterns in the feedback, and work with the product team to address systematic issues before production deployment.

Stage 3 — First value verification (weeks 6–10): confirming that the AI has delivered its first verified, meaningful outcome. The first value milestone should be specific: "The AI reviewed 50 NDAs with a 90%+ first-pass acceptance rate, saving your legal team 40 hours." The CS team's role is to calculate and document this first value milestone, present it to the customer's decision-maker, and use it as the foundation for the customer's confidence in broader deployment.

Stage 4 — Production scale-up (weeks 8–16): expanding the AI's deployment from the pilot scope to the full production volume. This stage is where the most common onboarding failures occur — the pilot was designed for a controlled subset of the customer's work, and production deployment exposes the AI to the full complexity of the customer's actual work volume. The CS team must monitor the production deployment closely, catching performance issues before they become customer-visible quality problems.

Stage 5 — Expansion planning (weeks 12–20): using the production deployment data to identify the natural next deployment opportunities. When the production deployment is stable and performing well, the CS team prepares the first expansion analysis — using the outcome data from the initial deployment to project the value available from the next use case or department. Stage 5 converts a successful onboarding into the beginning of the expansion motion.

Time to first value is the metric that most accurately predicts long-term retention. Analysis of AI product deployments shows that customers who achieve a verified first value milestone within 30 days of contract signing retain at rates above 95% at 12 months. Customers who take 90 days to achieve first value retain at rates below 70% at 12 months. The 60-day difference in time to first value predicts a 25-point difference in 12-month retention. The CS team's most important operational contribution in the first 90 days is reducing time to first value.

AI Product Onboarding Journey — Five Stages					
Stage	**Weeks**	**What happens**	**CS role**	**Success**	**Risk if stage**

				milestone	fails
1. Technical deployment	1–3	System integration; configuration of AI parameters (playbooks, thresholds, document types); test dataset run	Deployment project management; coordinate vendor technical team and customer IT; troubleshoot integration issues	Integration complete; AI processes test set in expected format; configuration validated	Delayed deployment → delayed first value → increased early churn risk
2. Pilot operation	3–8	AI runs on controlled set of real work; human reviews all AI outputs; feedback collected systematically	Collect structured feedback on each AI output; identify patterns; coordinate with product on systematic issues before production deployment	Pilot performance meets quality threshold (≥ 85% first-pass acceptance rate on pilot set); systematic issues identified and addressed	Poor pilot performance → loss of confidence → production deployment delayed or abandoned
3. First value verification	6–10	Confirm first verified meaningful outcome; quantify the value; present to decision-maker	Calculate first value milestone; prepare brief value presentation; present to department head or economic buyer	First value documented: 'AI reviewed 50 NDAs with 92% first-pass acceptance, saving legal team 40 hours — $6,600 in attorney time.'	First value not verified or not presented → economic buyer has no evidence of value; renewal risk begins early
4. Production scale-up	8–16	Expand from pilot scope to full production volume; monitor closely for performance issues	Active monitoring of production deployment; rapid response to performance issues; weekly outcome review with customer operations team	Production outcome delivery rate ≥ 80% on full volume; exception rate stable or improving	Performance issues in production → scope contraction risk; recovery difficult once customer loses

					production confidence
5. Expansion planning	12–20	Use production data to identify natural next deployment opportunities; prepare first expansion brief	Prepare expansion analysis from production data; introduce expansion conversation in month 3–4 review meeting	Expansion brief prepared and presented; customer has explicitly considered the expansion opportunity	No expansion planning → first renewal is purely a maintenance decision; organic growth momentum lost

⚠ Time to First Value Is the Most Important Retention Predictor

Analysis of AI product deployments consistently shows that time to first value is the strongest predictor of 12-month retention. Customers achieving first value within 30 days retain at 95%+ rates. Customers taking 60 days retain at less than 75%. The 30-day difference represents a 20+ point retention gap that CS teams cannot easily recover from later in the customer relationship. The most important single investment in the onboarding process is accelerating the path from contract signing to first verified outcome. This means deploying the simplest, highest-confidence use case first — the one where AI performance is most predictable — rather than attempting the most complex use case immediately.

Chapter Twelve — The Essentials

› Five onboarding stages: technical deployment → pilot operation → first value verification → production scale-up → expansion planning.

› Activation in AI products means: AI has processed real work and delivered a verified, meaningful outcome — not user training completed.

› Time to first value ≤ 30 days: 95%+ 12-month retention. > 60 days: < 75% 12-month retention. The gap is the most important onboarding metric.

› Start with the simplest, highest-confidence use case — not the most ambitious one. First value must be achieved quickly; complexity comes after confidence is established.

› Stage 5 (expansion planning) must begin during onboarding — the first expansion conversation happens in month 3–4, not at renewal.

CHAPTER THIRTEEN

CS in Service as Software: Outcome Verification and SLA Management

When CS owns SLA delivery. Outcome verification. The CS-RevRec handoff.

When the CS team owns SLA delivery — when the contract commits the vendor to specific outcomes and the CS team is responsible for verifying and reporting on those outcomes — the CS function takes on a role that has no direct SaaS equivalent. It is simultaneously the operational quality assurance team, the commercial evidence gatherer, and the account manager.

This operational-commercial dual role is the defining characteristic of CS in Service as Software products. The CSM for a Harvey AI enterprise deployment is not just building the relationship and managing the renewal. They are verifying that the AI is delivering the contracted recall rate, calculating the quarterly SLA credit if it is not, documenting the remediation for any SLA misses, and building the renewal case from the evidence they have assembled in their quality assurance role.

The three operational responsibilities of the CS team in Service as Software:

SLA monitoring and reporting: tracking the AI's performance against each contracted SLA metric in real time and generating periodic performance reports for the customer. For a contract review AI with a 95% recall commitment, the CS team (with AI agent support) monitors the recall rate on a rolling 30-day basis, generates monthly performance reports showing the recall rate versus the committed standard, and calculates and initiates SLA credits whenever the performance falls below the threshold. The CSM delivers this performance report to the customer's legal team and explains any significant variances.

Outcome verification: confirming that the outcomes billed under outcome-based pricing were genuinely delivered. For an outcome-based contract review product, the CS team

verifies a sample of completed reviews by having an expert attorney independently assess whether the AI's review met the quality standard. The verification sample rate (typically 5–10% of completed reviews) is specified in the contract, as is the methodology for calculating the outcome delivery rate from the sample.

The CS-RevRec handoff: transferring verified outcome data to the revenue recognition team for recognition event recording. When the CS team's outcome verification confirms that an outcome was delivered, this confirmation triggers the billing event (for outcome-billed products) and the revenue recognition event (for all products where outcome delivery is the performance obligation satisfaction event). The CS team's outcome verification is not just a quality assurance activity — it is a financial accounting activity with direct revenue recognition implications.

Specific examples of CS teams managing SLA delivery in Service as Software:

Abridge's clinical documentation CS team monitors documentation quality continuously: sampling physician acceptance rates, tracking modification intensity, monitoring time-to-delivery against the committed SLA, and managing the SLA credit calculation for any delivery delays. The Abridge CSM's role includes presenting the monthly quality dashboard to the health system's Chief Medical Information Officer (CMIO) and explaining the quality metrics in clinical terms — not in AI performance terms.

Harvey AI's CS team manages the quarterly sample audit: selecting 20 contracts from the quarter's reviewed volume, arranging for independent attorney review of Harvey's analyses, comparing the attorney's findings to Harvey's findings, calculating the recall rate for the sample, and reporting the result to the law firm's managing partner. When the recall rate falls below the contracted standard, the CSM calculates and initiates the credit, presents the remediation plan, and schedules the next quarter's performance review.

ServiceNow's enterprise CS team manages workflow SLA delivery: monitoring the uptime, performance, and process completion rates for each contracted workflow,

generating the monthly SLA performance report, managing the credit process for any SLA misses, and building the annual value review from the cumulative workflow performance data.

CS Roles in Service as Software — Three Operational Responsibilities

Responsibility	What it involves	CS operational requirement	Commercial implication
SLA monitoring and reporting	Track AI performance against each contracted SLA metric in real time; generate periodic performance reports; calculate and initiate SLA credits when thresholds are missed	Real-time SLA monitoring dashboard; automated credit calculation; monthly performance report generation; CSM delivery of performance report to customer legal/operations leadership	SLA credits reduce revenue; SLA compliance evidence supports renewal; SLA performance history is the primary quality assurance signal
Outcome verification	Confirm that billed outcomes were genuinely delivered; conduct sample audits of AI output quality	Quarterly sample audit protocol: random selection, independent reviewer engagement, comparative analysis, recall/precision calculation, quality report generation	Verified outcomes trigger billing events and revenue recognition entries; unverified or disputed outcomes create billing disputes and revenue recognition uncertainty
CS-RevRec handoff	Transfer verified outcome data to revenue recognition team to trigger recognition events	Daily/weekly outcome verification data to RevRec team; clear protocol for which data constitutes a recognition trigger; documentation of verification methodology for audit defense	Outcome verification is a financial accounting activity: the CSM's quality assurance work directly generates revenue recognition events and audit evidence

Chapter Thirteen — The Essentials

› CS in Service as Software has three operational roles: SLA monitoring and reporting, outcome verification, and the CS-RevRec handoff.

› The CS-RevRec handoff converts the CSM's quality assurance work into revenue recognition events — a direct financial accounting responsibility.

› Sample audit protocols (typically 5–10% of completed work) must meet the evidentiary standard of the customer's industry: legal customers need attorney-verified audits; healthcare customers need clinician-verified audits.

› SLA credits are a revenue recognition event when applied — the CS team's credit calculation must align with the RevRec team's variable consideration methodology.

› The CSM for a Service as Software product is simultaneously account manager, quality assurance lead, and commercial evidence generator — a genuinely new role.

EXTENDED CASE STUDIES

Three CS Transformations

Intercom · Glean · Harvey AI — rebuilding CS for autonomous AI products.

CASE STUDY A

Intercom: Rebuilding CS for the Fin AI Era

How Intercom's CS team redesigned their entire operating model when Fin AI made usage metrics irrelevant.

Intercom's transition to AI-first customer service — and the corresponding evolution of its own customer success model — is the clearest current example of a company rebuilding its CS motion from the ground up for an autonomous AI product.

Intercom's CS challenge was specific: the company had built a SaaS CS model around helping customers configure chatbots, build help center articles, and train support

agents on the Intercom interface. When Intercom launched Fin AI — its large language model-powered resolution agent — the CS team suddenly had a completely different product to support. Fin AI does not need users to be trained. It does not need chatbot conversation flows to be configured. It answers customer questions autonomously, using the customer's existing help center content and support ticket history as its knowledge base.

The old CS questions ("have your support agents been trained on Intercom?", "how many conversation flows have you built?", "what is your agent adoption rate?") became irrelevant overnight. The new CS questions are: "What is Fin's resolution rate?", "Which inquiry categories is Fin handling well and which is it struggling with?", "What is the economic value of the tickets Fin is resolving versus what your human agents would have cost?", and "Are there inquiry categories you're currently routing to human agents that Fin could handle with some configuration changes?"

Intercom's CS motion rebuild:

The health metrics shifted entirely to outcome-based signals. The primary CS dashboard for Intercom enterprise customers now centers on: Fin resolution rate (percentage of conversations Fin handles to completion without human handover), CSAT score for Fin-handled conversations vs human-handled conversations, hours saved per month from Fin resolutions (calculated automatically as resolution count × average handle time for comparable human resolutions), and ticket deflection rate (conversations that would have been routed to human agents that Fin handled instead).

The QBR structure shifted from feature adoption to value documentation. Intercom's QBR template for Fin deployments is built around the economic value summary: total Fin resolutions, hours saved, dollar value saved (at the customer's average support agent cost), and comparison to the contract cost. A customer paying $2,000/month for Intercom with Fin who sees that Fin has saved them 200 support agent hours (valued at $6,000) in the quarter has a clear, undeniable case for the contract value.

The expansion motion shifted to resolution rate analysis. Intercom's CS team identifies expansion opportunities by analyzing which inquiry categories the customer is currently routing to human agents that have high Fin resolution potential — based on Intercom's aggregate data from similar customers. If a customer's technical support inquiries have a 35% Fin resolution rate while the Intercom average for technical support is 62%, there is a clear expansion opportunity: help the customer optimize their knowledge base for technical support content, and Fin's resolution rate should approach the platform average.

The commercial outcome: Intercom's enterprise NRR improved significantly after the Fin AI launch and the CS model rebuild. Customers who experienced strong Fin resolution rates (above 50%) renewed at near-100% rates and expanded their Intercom deployments. Customers with weak Fin performance (below 30%) required intensive CS intervention to diagnose and fix the knowledge base issues limiting Fin's resolution rate.

CASE STUDY B

Glean: Measurement-First CS for Diffuse Productivity Value

How Glean built the measurement infrastructure that converts invisible knowledge worker productivity improvements into specific, defensible ROI.

Glean's customer success model is built for a specific challenge: demonstrating the value of AI knowledge management to customers whose primary concern is information worker productivity — one of the hardest categories of AI value to measure because the value is diffuse across thousands of individual interactions rather than concentrated in discrete, countable outcomes.

The challenge: when Glean's AI reduces the time an analyst spends searching for information by 30%, that 30% is distributed across dozens of small interactions

throughout the day. There is no single "outcome" that a CSM can point to and say "that is the value Glean created." The value is the cumulative effect of hundreds of small efficiency improvements — which is real and significant, but hard to document.

Glean's measurement-first CS model:

Glean addresses the measurement challenge through a proprietary time-in-context methodology: an optional measurement tool that, with employee consent, tracks the time employees spend on specific information work activities (searching, reading, synthesizing, writing) before and after Glean deployment. The time-in-context measurement provides the before/after comparison that converts diffuse productivity improvements into a specific dollar value.

The measurement tool is deployed in the first 60 days of the CS engagement — before the customer has formed their own assessment of Glean's value. By establishing the baseline and collecting the post-deployment data systematically, Glean's CS team arrives at the QBR with objective measurement rather than self-reported satisfaction. The employee survey is one component; the time-in-context data is the primary evidence.

The QBR structure at Glean is built around the productivity value calculation: number of knowledge workers using Glean × hours saved per week × 52 weeks × fully-loaded hourly cost = annual productivity value. For a 500-person knowledge worker organization at $75/hour fully-loaded cost with 2 hours/week saved per worker: 500 × 2 × 52 × $75 = $3.9M in annual productivity value. Against a $300K annual Glean contract, the ROI is 13×.

The expansion motion at Glean is driven by department coverage analysis: which departments have the most knowledge workers, the highest information retrieval work burden, and the lowest Glean adoption? For enterprise customers, the expansion opportunity is almost always additional departments — Glean often starts in engineering or product, and the expansion to marketing, sales, finance, and legal provides 3–5× the initial deployment scope.

The commercial outcome of Glean's measurement-first CS model: enterprise customers who see the productivity value calculation in their first QBR renew at over 95% rates and typically expand the Glean deployment by 30–50% at the first renewal. The measurement data converts a fuzzy productivity story into a specific financial argument that CFOs can evaluate and approve.

CASE STUDY C

Harvey AI: The Service as Software CS Model

The most sophisticated application of CS-as-SLA-management currently operating at enterprise scale — quarterly sample audits, independent verification, attorney-grade evidence.

Harvey AI's customer success model for law firm and corporate legal department customers is the most sophisticated application of Service as Software CS currently operating at scale. The Harvey CSM role is a genuine hybrid of account management, quality assurance, and commercial evidence generation.

Harvey's CS challenge: law firms and corporate legal departments are among the most analytically demanding enterprise customers. They are accustomed to rigorous evidence standards from their own work. A Harvey CSM who presents a QBR with vague productivity claims will be challenged by attorneys who are professionally trained to scrutinize evidence. The CS model must produce evidence that meets the evidentiary standard of its customers.

Harvey's quality assurance program:

Every Harvey enterprise deployment is supported by a quarterly sample audit. A random sample of 20 contracts reviewed by Harvey in the quarter is selected, and an independent attorney (senior associate level) at a reference law firm conducts a fresh

review of each contract, without seeing Harvey's output. Harvey's findings are then compared to the independent review's findings to calculate the recall rate (percentage of material issues identified by both Harvey and the independent reviewer that Harvey caught) and precision rate (percentage of Harvey-identified issues that the independent reviewer considered material).

The sample audit generates a defensible quality report that Harvey's CSM presents to the law firm's managing partner or the corporation's general counsel. The report is not self-reported performance data — it is independently verified quality data, comparable to the evidentiary standard that attorneys apply in their own work. This evidentiary quality is why Harvey's enterprise customer retention is high: the managing partner is not being asked to trust Harvey's claims. They are being shown independently verified data.

Harvey's expansion motion is based on work type analysis: which legal work types are currently being handled by human attorneys that Harvey's demonstrated recall rate suggests could be partially automated? For a law firm that is using Harvey for NDA review (93% recall on NDAs), the expansion conversation is data-driven: "Harvey's average recall on employment agreements across our customer base is 91%. Your employment agreement volume is approximately 40 per month. Based on comparable deployments, deploying Harvey for employment agreements would save your team approximately 120 hours per month. Would you like to see a pilot on 10 employment agreements before the full deployment decision?"

The commercial outcome at Harvey: law firms and corporate legal departments that receive Harvey's quarterly quality reports — independently verified, specifically quantified, financially translated — renew at rates above 90% and expand their Harvey deployments at rates that generate NRR above 130%. The sample audit is not just a quality control mechanism. It is the commercial evidence engine that makes Harvey's renewal and expansion conversations compelling.

CLOSING

Stop Asking Whether They Are Using It

The product is running. The question is whether it is working — and whether your CS team can prove it.

Framework F23 — The AI Customer Health Architecture

Framework F23 — The AI Customer Health Architecture — provides the operating model for customer success in AI-native products.

The framework rests on two foundational reframings:

From adoption to outcome: the primary measure of customer health shifts from whether users are engaging with the product to whether the AI is delivering the business outcomes it was deployed to create. This reframing changes every CS operational metric, every QBR structure, every health score component, and every expansion conversation.

From reactive to proactive: traditional CS is largely reactive — it responds to declining usage, customer-reported issues, and renewal conversations. AI CS must be proactive — using real-time health signal monitoring, early warning indicators, and predictive UAS to identify opportunities and risks weeks or months before they become visible in renewal conversations.

The five health signals of the architecture: usage depth, outcome realisation, expansion velocity, risk indicators, and sentiment — provide the comprehensive view of customer health that renewal and expansion decisions require.

The UAS (Usage Adoption Score) translates the five health signals into a single predictive metric: UAS above 70 indicates a healthy deployment with high renewal probability; UAS below 40 indicates at-risk status requiring immediate intervention. The UAS predicts NRR outcomes 90 days in advance.

The four CS operational pivots: health scoring (from engagement to outcome delivery), expansion identification (from usage patterns to outcome evidence), renewal

management (from relationship maintenance to value documentation), and onboarding (from user activation to value activation).

Framework F23 — Health Signals and CS Actions Reference

Health signal	Healthy indicator	Risk indicator	CS action: healthy	CS action: at-risk
Usage depth	AI processes >60% of eligible workflow volume; high operational dependency	AI scope narrowing; <25% of eligible volume; operational team routing work away from AI	Proactive expansion conversation: which additional workflow categories could AI handle?	Scope contraction intervention: identify why work is being routed away from AI; address the specific confidence gap
Outcome realisation	Delivery rate >85%; quality score stable or improving; value realisation >2×	Delivery rate declining over 3+ periods; value realisation <1×	Include in renewal value documentation: the delivery rate is the primary evidence	Immediate investigation: is decline in delivery rate vs scope, quality, or customer expectations? Product and CS joint response
Expansion velocity	Positive net expansion: new use cases or teams being added in trailing 90 days	Flat or declining deployment scope for 90+ days; no new deployment conversations	Expansion signal agent: identify next natural expansion opportunity; prepare expansion brief	Scope stabilisation: address the reasons for expansion stagnation; understand the ceiling and whether it is removable
Risk indicators	No risk indicators above threshold; stable stakeholder map; integration uptime >99.5%	Combined risk score > 9; any single indicator at 5; stakeholder displacement	Monitor maintenance: quarterly check on stakeholder map; integration health review	CS Agent Huddle: diagnose root cause; intervention plan by risk category;

				executive reconnection if required
Sentiment	Positive sentiment from both operational team and commercial stakeholders; unsolicited positive feedback	Expressed frustration; commercial stakeholder avoiding engagement; operational workarounds being created	Reference programme: high-sentiment customers are reference and co-innovation candidates	Sentiment investigation: structured conversation to understand specific concerns; address operational friction immediately

The question that customer success must stop asking is "are they using it?" The product is using itself. The AI reviews the contracts, resolves the tickets, generates the notes, and processes the invoices — continuously, autonomously, at scale. Whether the users are logging in is irrelevant if the AI is working.

The question that customer success must start asking is "is it working?" — and then, more specifically: is the AI delivering the outcomes the customer deployed it to create? At what rate? At what quality? With what trajectory? What is that delivery worth to the customer's business? And what would more of it be worth?

These are harder questions to answer. They require measurement infrastructure that most CS teams have not built. They require analytical conversations that most CSMs have not been trained to have. They require a relationship with the customer's financial leadership that most CS teams have not developed.

But they are the questions that determine whether the customer succeeds with the AI, whether they renew the contract, whether they expand the deployment, and whether they become the advocates who bring the next customer.

The SaaS CS playbook was built for a world where humans operated software. The AI CS playbook is built for a world where software operates autonomously on humans'

behalf. The transition between those two worlds is the transformation that this book describes.

The customers who succeed with AI will be served by CS teams who understand what success means in the AI era: not adoption, but outcomes. Not engagement, but value. Not retention, but the compounding of a commercial relationship built on evidence that the AI investment is among the best commercial decisions the customer has made.

Stop asking whether they are using it. Start asking whether it is working. And then show them the data that proves it is.

> ***"Stop asking whether they are using it. Start asking whether it is working. And then show them the data that proves it is."***

The AI Economy Monetization Series concludes with Book Ten:

The Complete AI Economy Monetization Reference