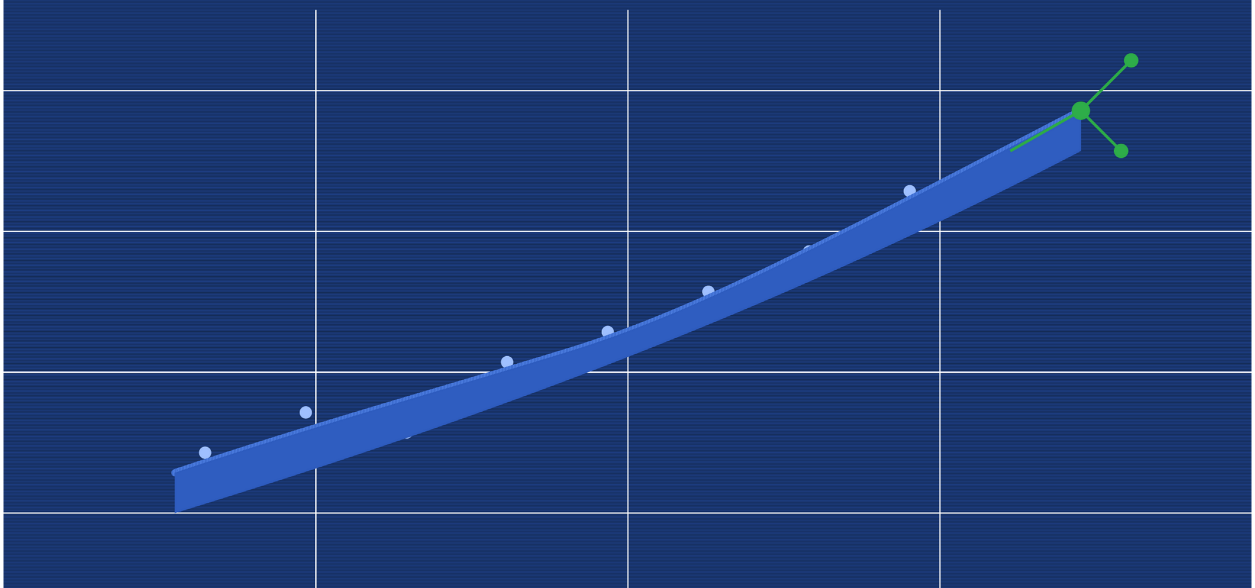
ccplanning.net
workforce planning, done well



A CCPLANNING.NET WHITE PAPER

AI and the Workforce Planner

An honest assessment of what AI actually does —
and does not do — for contact centre planning.

June 2026 · ccplanning.net

Contents

Right-click and choose "Update Field" to populate page numbers in Word.

Executive summary

Artificial intelligence is the most discussed and least understood topic in workforce planning today. The vendor pitch promises transformation; the experienced planner quietly suspects oversell. Both are partly right, and the gap between them is where most AI projects in contact centre planning quietly fail.

This paper takes a deliberately honest position. AI adds real, measurable value to specific parts of the planning function — and adds little or nothing, sometimes negative value, in others. The single most useful skill a planner can develop on this topic is the ability to tell the two apart. That skill is worth more than any individual tool.

> **The thesis in one paragraph**
>
> AI delivers less than vendors claim and more than sceptics expect. The genuine gains are concentrated in narrow, specific places: combining structured drivers into a forecast, reconciling forecasts across a hierarchy, detecting anomalies faster than a human can, drafting and summarising the analytical layer, and producing competent forecasts for queues that previously had none. The larger gains in any planning function still come from getting the foundations right — clean data, honest measurement, defensible method, and a credible operating rhythm — not from a more sophisticated algorithm. A planning team that adopts AI before it has built those foundations simply automates its existing problems faster.

The paper that follows sets out, function by function, where AI helps a planning team and where it does not; how to adopt it in an order that builds rather than burns credibility; what the maintenance and governance burden actually is; and how to have the leadership conversation that keeps expectations honest. It is written for the working planner, the planning manager, and the operations leader who has to decide whether the latest AI proposal is worth the money.

1. A topic that has lost its measure

There is no shortage of vendors claiming AI will transform contact centre forecasting and scheduling, and no shortage of planners quietly suspecting they are being oversold. The discussion has lost its measure. On one side, conference keynotes and pitch decks present AI as a step change that will make traditional planning obsolete. On the other, the planner who has lived through three "transformational" technology cycles has learned to discount the claims to near zero.

Neither posture is useful. The keynote optimism leads to expensive projects that under-deliver and damage the planning team's credibility when they do. The reflexive scepticism leads to genuine, capturable gains being left on the table while competitors quietly bank them. The planner who can hold both truths at once — that the technology is real and that most of the marketing about it is not — is the one who gets value from it.

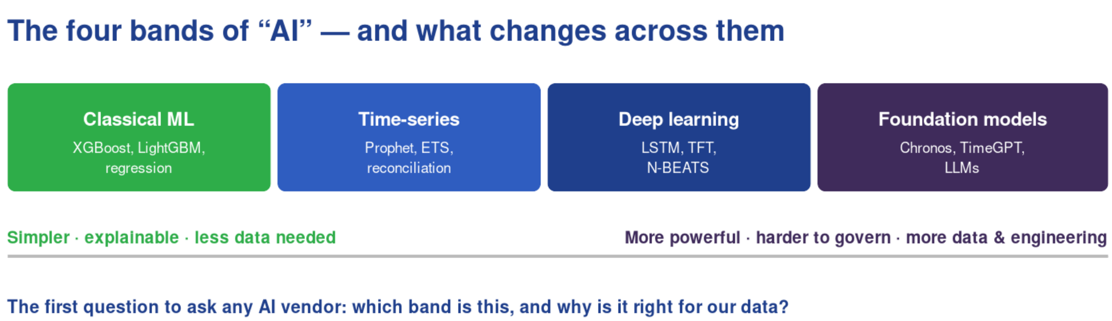
The reason the topic has lost its measure is partly that "AI" has stopped meaning anything specific. It is used to describe a regularised regression, a gradient-boosted tree, a deep neural network, a pre-trained foundation model, and a large language model summarising a report — technologies with wildly different data requirements, accuracy profiles, costs, and risks. Talking about "AI for planning" as a single thing confuses the conversation before it starts. The first discipline is precision about what is actually being proposed.

What we mean by AI

For a planning audience it helps to think of four bands on a spectrum, each with a different profile.

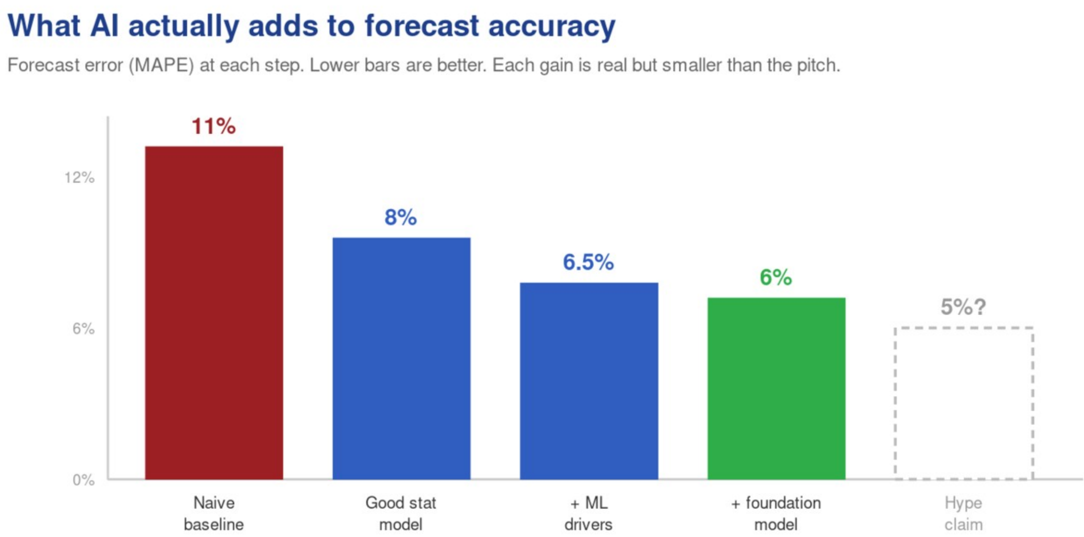
- **Classical machine learning** — gradient boosting (XGBoost, LightGBM), random forests, regularised regression. These take a feature table (volume, day-of-week, holiday flags, campaigns, weather) and learn the weights that best predict the target. Well understood, fast to train, explainable.
- **Time-series-specific methods** — Prophet, ETS state-space models, hierarchical reconciliation. These combine statistical rigour with some learned components and are purpose-built for the shape of demand data.
- **Deep learning** — LSTMs, temporal fusion transformers, N-BEATS, PatchTST. Powerful, data-hungry, and substantially harder to explain and govern.
- **Foundation models** — Chronos, TimeGPT, Lag-Llama for time series, and the general large language models (the GPT and Claude families) for text. Pre-trained on enormous corpora, capable of competent output on unseen problems with little setup.

When a vendor says "AI," ask which band. The honest answer determines everything that follows — how much data you need, how explainable the result will be, how much engineering the thing will require to stay alive, and whether your audit and finance functions will accept its output.



2. AI in forecasting: the accuracy reality

Forecasting is where AI is most heavily marketed to planning teams and where the honest accuracy picture is best understood. The headline most planners need is simple: the accuracy lift from AI is real, but it is smaller than the pitch decks suggest, and it is conditional on having a clean foundation underneath it.



The honest accuracy lift from AI on a typical forecast. Real gains exist, but each step is smaller than vendor decks imply, and the largest single improvement is usually getting the data clean — which is not an AI step at all.

A useful way to picture it: a naive last-year-same-week baseline might run at around 11% MAPE on a stable queue. A properly tuned classical model — Holt-Winters or Prophet with sensible event overrides — brings that to perhaps 8%. Adding a gradient-boosted model that combines the baseline with structured drivers might reach 6.5%. A foundation model on top might shave it to 6%. The vendor slide claiming 5% or better is usually comparing against a straw-man baseline, not against a well-run classical forecast. Each real step is genuine; none is the revolution the marketing implies.

Where AI genuinely helps

Five wins are real and worth pursuing in roughly this order of reliability.

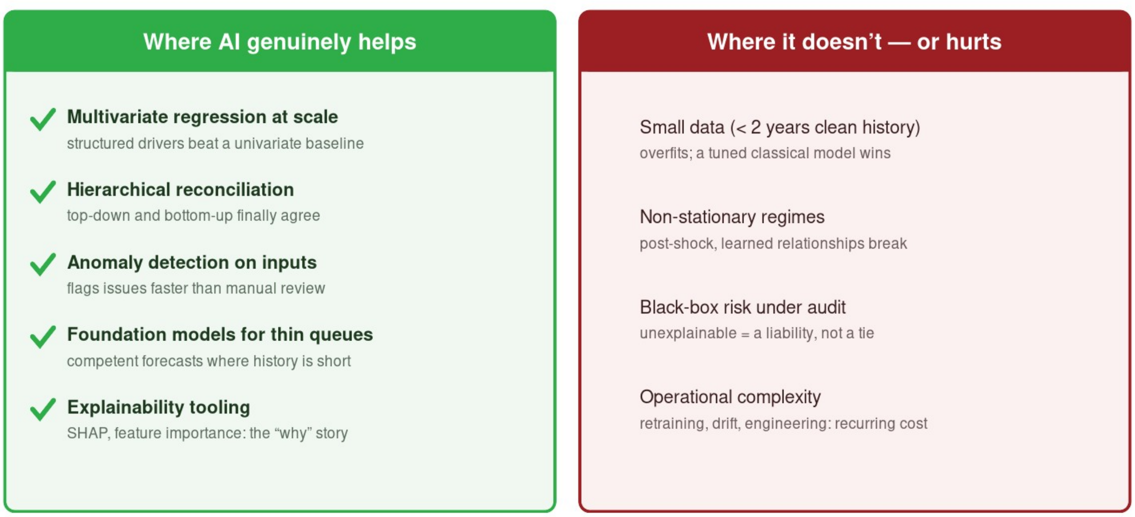
- **Multivariate regression at scale.** A planner with a good driver dataset — volume, day-of-week, holidays, marketing sends, product launches, weather, even competitor-outage signals — can use a gradient-boosted model to combine them into a forecast that consistently beats a univariate baseline. The gain is a few points of MAPE on stable queues, and substantially more on queues that respond strongly to drivers a time-series-only model cannot see.
- **Hierarchical reconciliation.** Operations with many sub-queues face the perennial problem that the bottom-up sum does not equal the top-down number. Modern reconciliation methods (the MinT family being the most cited) resolve this with machine-learned weights, usually beating both pure top-down and pure bottom-up.
- **Anomaly detection on inputs.** Models that flag unusual arrival patterns, AHT shifts, or interval-level outliers surface issues faster than manual inspection, freeing the planner for the judgement-heavy parts of the job.

- **Foundation models for thin-history queues.** For small or newly launched queues that lack the history to fit a bespoke model, pre-trained time-series models can produce a competent forecast out of the box with no model-selection effort — a genuine step forward.
- **Explainability tooling.** SHAP values, partial-dependence plots, and feature-importance reports turn a model into a story about why the forecast moved — which is exactly the conversation finance and operations want to have.

Where AI does not help — or actively hurts

Several conditions flip the comparison the other way. **Small data** is the most common: queues with less than two years of clean history are usually better served by a well-tuned classical model, because the training process consumes degrees of freedom a small dataset does not have and the result overfits. **Non-stationary regimes** are dangerous: after a pandemic, a restructure, a launch, or a regulatory change, the relationships the model learned no longer hold, and a simple model with a planner's manual override is more robust than a sophisticated one nobody can override cleanly.

Black-box risk matters more than vendors admit. In a regulated industry, under internal audit, or in any forecast that must be defended at senior level, a model nobody on the team can explain is a liability — the same accuracy from a less explainable model is a worse outcome, not an equivalent one. And **operational complexity** is the recurring cost: a pipeline that retrains, watches its own drift, and reconciles its hierarchy is far more infrastructure than a spreadsheet plus exponential smoothing. Teams that adopt the technology without resourcing the engineering find the model degrades silently and the analyst spends more time keeping it alive than using it.



> *An ML model that beats a naive last-year comparison is not the same as one that beats a properly tuned Holt-Winters with event overrides. Honest benchmarking is the difference between a real decision and an expensive one.*

3. AI in scheduling and capacity planning

Scheduling is, mathematically, an optimisation problem: fit the right number of the right people across intervals to meet a service target at the lowest sustainable cost, subject to a thicket of constraints — contracts, skills, fairness, statutory rest, individual preferences. This is fertile ground for algorithms, and the better WFM platforms have used optimisation engines for years. The question is what the current wave of AI adds on top.

The genuine contribution is in two areas. The first is **scenario speed**. A planner who wants to test "what happens to cost and service if I move the part-time layer two hours later, or add a Saturday shift, or change the shrinkage assumption from 30% to 36%" can now get a credible answer in seconds rather than rebuilding a spreadsheet for an afternoon. Faster iteration changes the quality of the decision, because the planner can explore the trade-off space instead of defending the first feasible answer they found.

The second is **preference-aware optimisation**. Modern engines can balance the operation’s coverage needs against thousands of individual agent preferences far more effectively than a manual roster, which has measurable retention benefits. A schedule that respects when people actually want to work, within the bounds the operation can sustain, is a quietly powerful lever — and one of the few places where the technology improves both cost and the human experience at the same time.

> **Where scheduling AI disappoints**
> Optimisation is only as good as its inputs. An engine fed an optimistic shrinkage figure produces a beautifully optimised schedule that is structurally short of cover — the maths is flawless and the answer is wrong. The most common scheduling failure is not a weak algorithm; it is a strong algorithm applied to a shrinkage assumption that was never measured honestly, or a demand curve that aged out two seasons ago. Fix the assumptions before you buy the optimiser.

In capacity planning the same logic holds. The twelve-month plan is driven by five inputs — volume forecast, AHT trend, realised shrinkage, attrition by cohort, and ramp-up curve. AI can sharpen the volume forecast and detect attrition patterns, but the plan fails far more often because shrinkage was modelled at 30% when the reality was 36%, or ramp-up was assumed at six weeks when it was nearer ten. No model rescues a plan built on inputs nobody measured. The leverage is in honest inputs, and honest inputs are a discipline, not a technology.

4. AI in real-time management

Real-time management is the discipline of responding to the day as it actually unfolds, and it is dominated by a single under-appreciated truth: most short-term variance is noise that reverts to the mean, and the most valuable real-time skill is the discipline to do nothing in response to it. This has a direct and slightly counter-intuitive implication for AI.

The genuine AI win in real-time is **anomaly detection that distinguishes signal from noise**. A model that has learned the normal distribution of interval-level variance can tell the analyst, with reasonable confidence, whether a 14% spike at 10:30 is within the band of natural fluctuation or a genuine event worth acting on. Used well, this makes the team calmer and more accurate — it reduces the over-reaction that is the most expensive real-time failure mode, where an analyst authorises overtime and cancels training in response to a blip that has reverted by the time the actions take effect.

Used badly, AI does the opposite. A dashboard that surfaces every deviation as an alert, dressed up as "AI-powered monitoring," trains the floor to chase noise and erodes the very discipline that makes real-time credible. The technology should raise the threshold for action, not lower it. The right question to ask a real-time AI vendor is not "how many things can it alert on" but "how well does it tell me when to do nothing."

> *The right real-time AI raises the threshold for action. A tool that surfaces every deviation as an alert is not intelligence — it is a faster way to chase noise.*

There is also a real contribution in **intraday re-forecasting** — updating the rest-of-day projection as actuals arrive — and in suggesting, from a known menu of levers, which response best fits the situation. But these are decision-support tools. The accountable human still decides, because the cost of a confident wrong automated action on a live floor is high and immediate.

5. AI in quality, MI, and the analytical layer

This is where the newest and, for many planning teams, most immediately useful AI lives — and where the honest assessment is most nuanced, because the gains are real but they are not the gains the slide deck promises.

Quality assurance

AI-led QA can score a far larger share of contacts than a human team ever could, apply a consistent rubric without evaluator drift, and surface coverage that random sampling misses. That is genuinely valuable. What it does poorly is judge the things that need judgement — tone in a genuinely difficult call, the unscripted moment where an agent did exactly the right human thing, the edge case the rubric never anticipated. AI QA is strong on the measurable and weak on the meaningful.

The combination that works is neither pure-AI nor pure-human. AI scores at volume and flags the contacts worth a human look; humans calibrate the model, handle the judgement calls, and own the coaching conversation. Crucially, the coaching loop — the part that actually changes behaviour — remains human, because an agent will not change for a score from a machine the way they will for a respected team leader who watched the call. AI makes QA scalable; it does not make QA matter. People still do that.

Management information and reporting

Generative AI is genuinely good at the analytical layer's grunt work: drafting the narrative around a set of numbers, summarising a week of activity into a readable update, turning a dense pack into a one-page summary for an executive, and translating a planner's technical finding into the language a different audience needs. A planner who spends Friday afternoon writing the same commentary they wrote last Friday can reasonably hand the first draft to a model and spend the reclaimed time on analysis.

But there is a hard limit, and it is the limit that matters most for MI. Generative AI will confidently produce plausible narrative around numbers it does not understand, including causal claims the data does not support. It will write "service dipped because of higher volume" when the real cause was an absence spike, because the first explanation is more common in its training and reads well. The planner who lets a model write the causal story without checking it is outsourcing the one part of the job — understanding why — that gives the planning function its credibility. Use it for the draft; never use it for the diagnosis.

> **The MI rule for gen-AI**
> Let the model draft the words. Never let it decide the cause. The value of an MI pack is not the prose around the numbers — it is the correct identification of what is driving them, and that remains the planner's job. A beautifully written pack with a wrong causal story is worse than a plain one with a right one, because people act on it.

Speech and interaction analytics

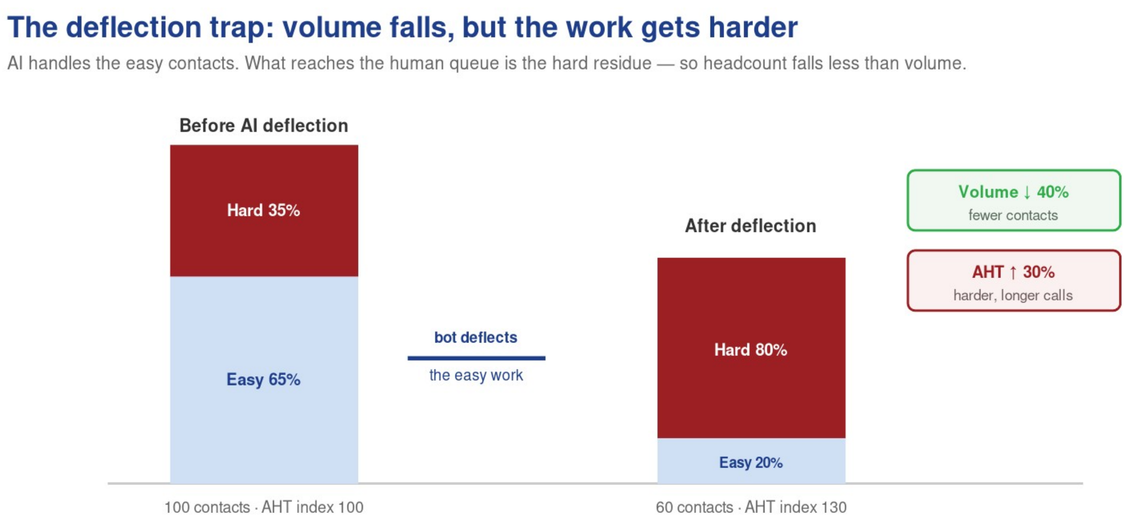
For planners, the under-used contribution of speech analytics is demand decomposition: understanding *why* customers are contacting, at scale, from the content of the contacts themselves. A forecast built on call reason — rather than on undifferentiated volume — is a better forecast, because it can anticipate the effect of an upstream change (a billing run, a website outage, a product recall) that raw volume only reveals after the fact. This is one of the clearest cases where AI gives the planner information they genuinely could not get any other way.

6. AI, self-service, and the changing shape of demand

The AI development that will affect planners most over the next three years is not a forecasting model at all. It is the customer-facing AI — the chatbots, the voice assistants, the generative self-service — that increasingly resolves contacts before they ever reach a human agent. This deflection changes the workload planners are planning for, and the change is more subtle and more dangerous than the headline "AI will reduce volume" suggests.

The headline is true as far as it goes: capable self-service reduces the count of contacts reaching the human queue. But it does not reduce them uniformly. AI handles the simple, repetitive, high-volume contacts best — balance checks, password resets, opening hours, order status. What it deflects, therefore, is precisely the easy work. What remains for the human agent is the residue: the complex, the emotional, the multi-issue, the regulated, the cases where the customer has already tried the bot and arrives frustrated. Volume falls and average complexity rises at the same time.

> *AI deflection does not shrink the queue evenly. It removes the easy contacts and leaves the hard ones — so volume falls while average handle time and complexity rise. A planner who models only the volume drop will under-staff a harder operation.*



This has direct and frequently mishandled consequences for the planning model. AHT rises as the simple contacts leave the mix, so a forecast that applies the old AHT to the new lower volume will under-staff. Occupancy tolerance should arguably fall, because sustained handling of harder, more emotionally demanding contacts burns agents out faster than a mix that included easy breathers. Shrinkage assumptions may need revisiting, because a harder contact mix changes the coaching, wellbeing, and off-phone support the operation must build in. And forecast volatility increases, because the residual human volume is a smaller, less stable number more sensitive to how well the bot performed that day.

The deflection trap

The finance case for self-service AI is almost always built on the volume reduction alone — "the bot handles 40% of contacts, so we need 40% fewer agents." The real reduction in required headcount is smaller, because the remaining contacts are longer and harder, and because someone must now plan for, monitor, and improve the bot itself. A planner who signs up to the naive headcount saving will spend the following year explaining a service miss that was baked in on day one.

There is also a new planning object to account for: the bot's own performance is now a driver of human demand. A degraded or poorly tuned assistant pushes more contacts — and more annoyed contacts — into the human queue, exactly the way a website outage does. The mature planning function treats containment rate as a forecastable, monitorable driver, not a fixed assumption, and builds the relationship with the team that owns the self-service layer so that changes there are flagged before they hit the floor. This is one of the clearest examples of AI creating work for planners even as it removes work from agents.

7. The planner’s role: augmented, not replaced

The anxious question under every AI conversation in a planning team is whether the technology replaces the planner. The honest answer is that it changes the job substantially and eliminates it almost nowhere — and the planners who thrive are the ones who understand which of their tasks are now cheap and which are now more valuable than ever.

AI makes the **mechanical** parts of planning cheap: producing a baseline forecast, generating a feasible schedule, scoring a contact, drafting a report. It makes the **judgement** parts more valuable, not less, because when the baseline is automated the differentiator becomes the overlay — the event knowledge, the operational context, the relationship with operations leadership, the ability to defend a number under pressure and own it when it is wrong. The credibility of a planning function was never the maths; the maths was the easy part. AI automates the easy part, which raises the premium on everything else.

> *AI automates the easy part of planning. The hard part — judgement, context, credibility, the relationship around the number — is exactly the part it cannot do, and exactly the part that was always the real job.*

Augmented, not replaced: where the value moves

The planner — judgement layer
Event knowledge · operational context · the relationship with operations · defends the number · owns the miss
More valuable as the layer below is automated

The model — mechanical layer
Baseline forecast · feasible roster · scored contact · first-draft report — now cheap and fast

AI automates the easy part of planning. The hard part — judgement, context, credibility — is exactly what it cannot do.

Practically, this points the modern planner toward a different skill mix. Less time hand-building forecasts and rosters; more time on data quality, on translating analysis for different audiences, on the operating rhythm that makes the function trusted, and on governing the models that now produce the first draft of the work. The planner becomes the judgement-and-communication layer on top of the machine. That is a more senior, more durable, and frankly more interesting job than the one the machine is taking over.

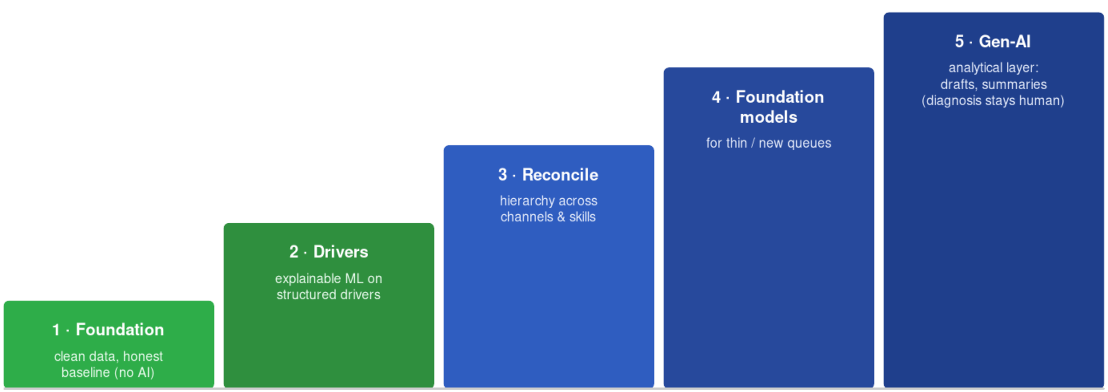
8. A practical adoption roadmap

The path that consistently works treats AI as an addition to a clean foundation, never a replacement for one. The sequence matters, because adopting in the wrong order is how teams burn money and credibility at the same time.

1. **Fix the foundation first.** Clean, granular history; channels separated; shrinkage and AHT measured honestly; a defensible classical baseline; and a measurement framework that tracks accuracy across multiple horizons. Without these four, layering AI on top mostly automates the existing problems faster. This step has no AI in it, and it is the highest-return step in the whole programme.
2. **Add structured drivers with an explainable model.** The highest-return first AI step is usually a gradient-boosted regression combining the time-series baseline with structured drivers. It is well understood, explainable, fast to train, and easy to govern. If it adds two or three points of MAPE in honest benchmarking, the case for going further is strong. If it does not, the case is weak and the project should stop there — cleanly, without embarrassment.
3. **Reconcile across the hierarchy.** Where the data justifies it, add hierarchical reconciliation across channels and skills so the bottom-up and top-down numbers finally agree.
4. **Explore foundation models for thin queues.** For small, new, or unstable queues, test pre-trained time-series models where bespoke training is hard. Treat this as exploration, not commitment.
5. **Layer generative AI onto the analytical workflow.** Use it for first-draft commentary, summarisation, and audience translation — with the diagnosis always checked by a human.

A staged adoption path that builds credibility, not burns it

Each step earns the next. The highest-return step has no AI in it at all.



If step 2 doesn't beat a tuned baseline in honest benchmarking, stop there — cleanly.

Two principles hold throughout. The planner is augmented, not replaced — the model produces a baseline; the human overlays judgement and signs off. And explainability is not optional — every published output should be defensible by a person who can articulate the drivers, the assumptions, and the known limitations. A team that cannot explain its own forecast does not own it.

9. Governance, drift, and the maintenance burden

The part of AI adoption that vendor decks skip is what happens after the impressive demo. Machine-learning outputs decay. A model trained on last year's data gradually misaligns with this year's reality, and the misalignment creeps in long before it becomes obvious. Any team adopting AI must plan for ongoing monitoring — accuracy tracked weekly, automated alerts when drift exceeds a threshold, scheduled retraining, and clear ownership of the decision to retrain or roll back.

This is engineering work, and a planning team that takes it on without engineering support typically ends up running an unmaintained model that nobody fully trusts within twelve months. The cost is real, recurring, and almost always under-stated in the business case. The honest version of the business case includes the maintenance line.

Governance follows the same logic. A policy document covering model lineage, training-data provenance, validation methodology, drift monitoring, and override authority is not glamorous, but it is what lets the model be trusted by audit, finance, and senior management. Without it, the first time the model is wrong in a way the planner cannot quickly explain, confidence collapses — and the team retreats to spreadsheets, usually with more bitterness than they had before they started.

> **The maintenance question to ask before you buy**
> Who owns this model in eighteen months, when the analyst who built it has moved on and the demand pattern has shifted twice? If the answer is "nobody has the time," the project will fail — not because the technology is bad, but because an unmaintained model is worse than a maintained spreadsheet. Resourcing the upkeep is part of the decision, not an afterthought.

10. The leadership conversation

Most planning leaders will at some point have to respond to an AI proposal from above — a board that read an article, a vendor who got to the executive first, a peer operation claiming a transformation. The planning function's credibility depends on handling that conversation honestly rather than either dismissing it or capitulating to it.

The strongest position is the measured one. Yes, AI adds value here, here, and here — and the honest gain is this size, conditional on this foundation, with this maintenance cost. No, it does not make the planning team obsolete, and here is why the judgement layer matters more as the mechanical layer is automated. That answer earns more trust than either uncritical enthusiasm or reflexive resistance, because it demonstrates the one thing leadership most wants from its planning function: a clear-eyed grasp of what is real.

Vendor due diligence: five questions

6. **Which band of AI is this, specifically?** Classical ML, deep learning, or foundation model — and why is that the right choice for our data?
7. **What is the accuracy gain against a properly tuned classical baseline,** not against a naive last-year comparison?
8. **How explainable is the output,** and will our audit and finance functions accept it?
9. **What does the model require to stay accurate** — retraining cadence, drift monitoring, engineering ownership — and who is resourcing that?
10. **What happens when it is wrong?** Can a human override it cleanly, and who is accountable for the published number?

A vendor who answers these five well is worth listening to. A vendor who deflects them is selling the slide deck, not the system.

Conclusion: the honest opportunity

AI in workforce planning is real, useful, and overhyped, all at once. The practical opportunity is meaningful but specific: better use of structured drivers in the forecast, cleaner reconciliation across hierarchies, faster and calmer anomaly detection in real time, scalable quality assurance with a human judgement layer, and a generative-AI assist on the analytical grunt work. Captured well, it frees planners from the mechanical work and lets them spend their time where their value actually is.

Captured carelessly — model before data, hype before benchmarking, automation in place of judgement — it produces a more sophisticated version of the existing problems plus a new layer of opacity, at higher cost and with the planning team's credibility spent. The difference between the two outcomes is not the quality of the technology. It is the quality of the planning discipline the technology is layered onto.

> *The planners who get the most from AI are the ones who use it to augment a discipline they already do well — not as a substitute for one they have not yet built.*

That is the honest assessment. The opportunity is genuine, the foundations come first, and the planner — augmented, not replaced — remains the most important part of the system. The teams that internalise that will quietly outperform both the believers and the sceptics, because they will be the only ones measuring the right thing.

About ccplanning.net

ccplanning.net is an opinionated, practitioner-focused resource for contact centre workforce planning — forecasting, scheduling, real-time management, capacity planning, MI, and the leadership of the planning function. It publishes free articles, browser-based planning calculators (Erlang C, Erlang A, shrinkage, cost of attrition, and more), and a fortnightly newsletter for working planners.

If this paper was useful, the newsletter is where the thinking continues — one article, one resource, and one honest observation every other Thursday. Subscribe at ccplanning.net.

Read next on ccplanning.net

Using AI for contact centre forecasting · Generative AI for planners · AI vs human QA · The Excel forecasting paradox · A beginner's guide to forecasting — the foundation any AI work should sit on top of.