Manuka

Beyond Pilot Purgatory

Getting AI Into Production

How to move from prototype to production-ready AI by closing technical, human and commercial gaps.

Gareth Martin

Chief Executive Officer

"Credibility is easily lost in this space," says Steve Janoo, Global Data, Analytics & AI Leader ex. Diageo and Dyson. "If you lose confidence quickly, it's going to be a challenge to get that back."

Most organisations can build a working AI model. The hard part (the part where value is won or lost) is everything that has to happen around it before that model can go live.

That gap between a promising pilot and a governed, measurable, trusted production system is where most AI programmes stall. Not because the technology falls short, but because the conditions for production were never put in place.

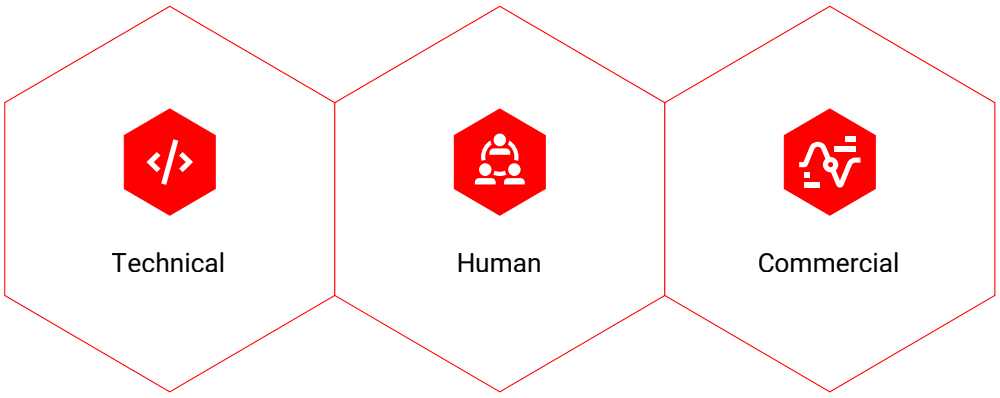
At a recent panel we hosted at the Databricks London office, leaders from Diageo, SEGA Europe and Databricks shared what it takes to close that gap. Four themes came through clearly:

- Scope the solution so everyone is clear about what you're building and why it matters.
- Understand the value to define a measurable benefit that justifies every step from pilot to production.
- Build in governance from the start so you can deploy to production without hitting barriers later.
- Bring sponsors and users with you to establish trust and momentum across the organisation from board level down.

Get these right and you create the conditions for production. Get them wrong, or skip them, and you risk joining the majority of organisations still stuck in pilot purgatory.

In this article we look at the three critical gaps you need to close to set these foundations.

The bridge to production-ready AI has three pillars:



- **Technical** – a clear evaluation framework, reliable CI/CD for LLM apps, strong observability, robust governance and reusable patterns that teams can build on.
- **Human** – a cultural shift that wins hearts and minds, gains leadership buy-in and includes early involvement of users so AI reflects real workflows and earns trust from day one.
- **Commercial** – tight scoping that underpins a direct, measurable link to a business KPI that makes the investment easy to justify.

Building a production-ready AI ecosystem

01 The Technical Gap: Lay the foundations of a production system

A demo can prove what's possible, but it's the production system that transforms it into a controlled value driver.

Teams often build AI applications without a proper production framework or the tools and platforms they need to scale them. That's why so many promising pilots do not make it past architectural or governance approval.

A production-ready AI system has four non-negotiables:

A clear evaluation design - What the system is meant to achieve, what it must avoid, how quality will be measured and how drift will be detected and corrected.

Evaluation is the foundation of trust. The LLM-as-a-Judge approach is a good example of how to achieve it at speed. It compresses validation cycles from weeks to hours, without removing human oversight. Faster evaluation means faster iteration, which means faster progress toward production.

> *"Think of evaluation first. Collect some examples of what a good answer could have been...and the other thing that we do a lot is go back to the data and understand what the data state is like. Do we need to add a little bit more metadata to help the model understand what we need to fetch? Do we need to clean it?"*
>
> **Maria Zervou (Databricks)**

Governance and observability baked in - You can't scale what you can't see. Production AI needs full traceability, guardrails and human oversight from day one, supported by tools such as Databricks' AI Gateway, MLflow Tracing and Unity Catalog.

Start with the lowest autonomy tier that works, prove you can monitor and control the system reliably, only then increase autonomy. That's how governance becomes an enabler of production, not a blocker to it.

Interoperability – AI works best when data, models and agents can communicate through open standards on a single interface.

This interoperability enables different AIs to use shared data and services, supports agent-to-agent coordination and avoids the brittle, proprietary models that create long-term lock-in. It also strengthens governance: a single interface can register, monitor and cost-control all AI usage across the organisation, giving leaders visibility of who's using what, when and why.

Reusable architecture – Every engagement should strengthen the asset library. For example, a multi-agent system built for one specific process can act as a foundation for others. Reuse accelerates delivery and reduces risk, and it means each new use case is easier than the last.

02 The Human Gap: Winning hearts and minds

Many data and AI teams are still fighting for credibility, navigating middle-management resistance and trying to balance boards that say they want AI while simultaneously blocking anything that feels risky.

Momentum only builds when people at every level understand the opportunity available right now, trust the approach and feel part of the journey. This requires a campaign to educate and excite everybody about AI, running alongside the tech transformation itself.

Leadership sponsorship from the top is essential, but when choosing where to pilot first, start with stakeholders who are genuinely receptive to AI and willing to iterate. Early credibility matters enormously. If your first high-profile use case underwhelms a sceptical executive, recovering confidence across the programme becomes much harder. Build momentum with advocates, prove value, and let that success pull senior leaders in rather than risking their trust before you've earned it.

Reinforcing momentum is then about shaping the product with the people who will use it. AI solutions aren't one-off projects; they're evolving products. If users only meet the system at handover, adoption is already at risk. Rather than being trained on a finished tool, users should shape products from the start so they are active contributors to a constantly improving system.

"Feedback and reinforcement is really important and it's always a missed thing. Consumers using our agents don't realise the importance of giving the agent feedback on how well it's done or how badly it's done. Collect it, encourage people to give it, and then lastly, and most importantly, do something with it."

Francis Hart (SEGA)

Many people's only reference point for AI is ChatGPT or Copilot. That makes it harder to understand AI embedded inside a workflow or process. Turning users into co-owners, not passive recipients, closes that gap and strengthens adoption.

This means three things:

Enable early and continuous user involvement	Implement human-focused change management	Move from a project to a product mindset
Not just in testing, but in designing the workflow, defining quality and understanding how the system behaves.	Help users understand new processes, how their role changes with AI, how they work with the system and how they can correct it if it drifts.	AI in production is not a finish line. The goal is not one live use case, it's a repeatable route to the next one. That means ongoing user input, evaluation and adjustment.

By instilling a sense of ownership, you enhance adoption and unlock the continuous improvement that ensures AI systems deliver long-term value.

03 The Commercial Gap: Keep your eye on the KPI

Broad, all-things-to-all-people, solutions are hard to design well and even harder to measure. A tightly scoped use case, linked to one KPI, gives you a clear baseline, a clear value path and a clear way to judge whether the system is doing what it was built to do.

There are also two very different types of AI investment. Organisations should absolutely modernise the workplace by rolling out general-purpose tools such as Claude or Copilot to create a broad productivity uplift. But productionising AI is a different category: these are specific, strategically aligned use cases with a tightly defined scope and measurable outcomes.

The strongest commercial cases we see start from the organisation's strategic objectives, not from the technology. Identify the business functions that genuinely need transformation, find the highest-value processes within them, and look for use cases that align to strategic initiatives already active and funded in the organisation. That's where the budget already sits, the executive sponsorship already exists and the appetite for change is highest. Aligning AI to these programmes turns a standalone investment case into an accelerant for transformation that's already underway.

That is what a credible commercial case for AI looks like. Tying a production system to a high-level goal that impacts financial performance focuses minds and unlocks investment. Without a single measurable outcome such as revenue, margin or cost reduction, it becomes impossible to justify the spend required to move from pilot to minimum viable product to production.

Governance matters here too. Every AI product must clear three approval routes: architecture, compliance and finance. A clear KPI unlocks all three. If the business case isn't obvious, finance won't fund it, architecture won't back it and compliance won't touch it. One weak link, and the pilot stalls before production.

"One of the core problems of why we are not getting pilots into production is because we're not focused on those real business problems and we are really looking at the value that solving those business problems is gonna have."

Steve Janoo (ex. Diageo)

Navigating your path to production

Close these three gaps and you create a credible route to production:

01 **Scope tightly and anchor to one KPI -** Identify a narrow purpose and a single measurable outcome. Tight scope makes the system easier to design, govern and evaluate, and the KPI provides the commercial anchor that unlocks investment.

02 **Win hearts and minds early -** Build belief in what AI can do now, not just in the future by educating leaders, engaging receptive stakeholders first and creating enough positive sentiment to carry the programme forward.

03 **Co-design evaluation with the business -** Define golden datasets, quality dimensions and acceptance criteria with SMEs. Build these into the product so users can see when performance changes and understand how the system behaves.

04 **Govern through a unified, interoperable gateway -** Model-specific approvals break as models evolve. A gateway approach unifies access, behaviour and logging, while open standards ensure agents, data and services can communicate across the organisation.

05 **Build reusable patterns -** Every production use case should make the next one easier. Reusable architecture accelerates delivery, reduces risk and creates a repeatable route from pilot to production.

This is how organisations escape pilot purgatory: by building the technical, human and commercial conditions that make production possible. **If you want to turn promising AI pilots into governed products with measurable business impact, we should talk.**

Manuka

Beyond Pilot Purgatory

Getting AI Into Production

How to move from prototype to production-ready AI by closing technical, human and commercial gaps.

We're Manuka. We build AI on Databricks. Let's build together.

- UK – Manchester – London – Liverpool
- linkedin.com/company/manukaai-emea
- ANZ – Sydney – Auckland
- linkedin.com/company/manukaaianz
- US – California – North Carolina – Texas
- linkedin.com/company/manukaai-us
- www.manuka-ai.com