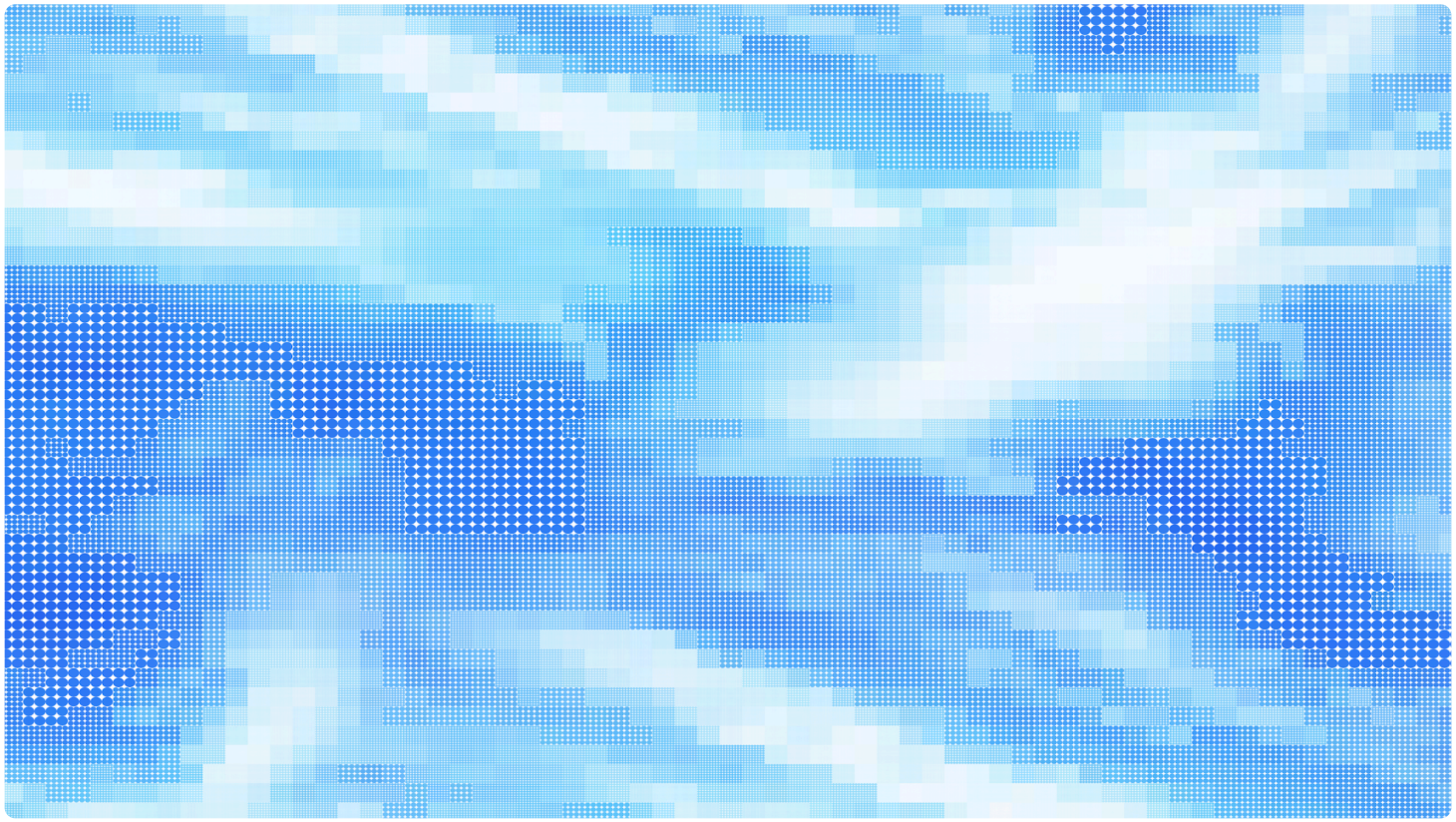
OpenAI

From experiments to deployments

A practical path to scaling AI



Foreword

Most organizations are experimenting with AI. Some are stuck in pilots, others are already weaving it into daily operations and customer products. The difference is approach, not ambition.

Traditional software playbooks were built for slower, linear cycles. AI moves faster, requires teams to learn as they build, and draws on innovation from every corner of the business. Progress now depends on finding repeatable ways to experiment, learn, and scale faster than the technology evolves.

At OpenAI, we partner with companies across industries to turn experimentation into execution. Together, we've seen what accelerates progress, and what stalls it. This guide distills those lessons at a high level, and shares where you can go for deeper guidance.

A new playbook for AI deployments

For years, companies have focused on validating whether software was fit for purpose. The approach was simple: start small, test a specific use case, and scale once results are proven. This worked when technology evolved slowly and served a single department at a time.

AI moves differently. Its capabilities evolve in weeks, not quarters, and its impact reaches every part of the organization. Success depends less on a single tool's performance and more on how quickly teams can learn, adapt, and apply AI to solve the problems in front of them.

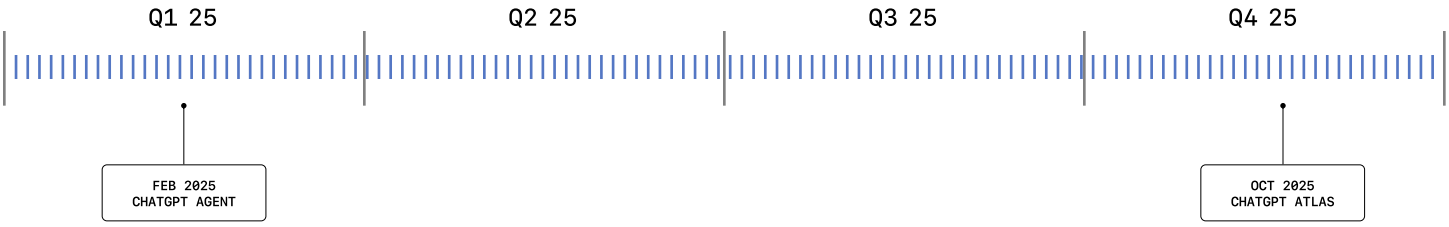
These shifts demand a new operating rhythm that balances speed with structure and evolves as fast as the technology itself.

Why AI requires a new playbook

Speed defines success

Product cycles that once ran quarterly now update weekly. Long validation phases mean teams ship what's already obsolete. Organizations that stay current build in shorter loops, remove friction, and decentralize literacy and learning.

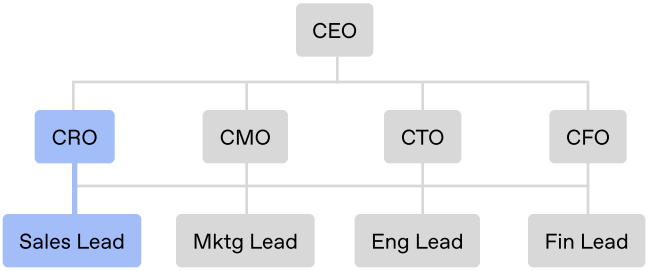
There has been a new feature every ~3 days across ChatGPT & API so far this year



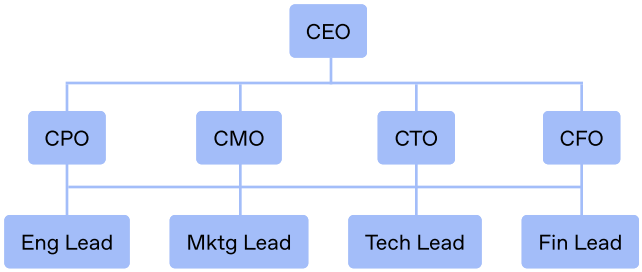
Innovation can come from any team

Traditional tools lived inside departments; AI spans all teams. A marketing analyst who automates reporting can find use cases that scale across the whole company. The best outcomes come when innovation is distributed, celebrated, and shared.

Traditional solutions



AI solutions



ROI compounds over time

AI's value grows with every use. Early time savings expand into organization-wide efficiency, creating the foundation for new revenue and cost savings through new homegrown products. As data access and skills mature, each new build becomes faster and more valuable than the last. Developing literacy and scaling use cases will have exponential impact on ROI.

AI requires new skills, not just new tools

AI is changing how work happens, not just the tools we use. The fastest path to impact is a workforce fluent in AI, with people who know when and how to apply it. That fluency takes practice: teams need room to test, iterate, and make AI part of everyday workflows, or new products that they build.

AI systems don’t follow fixed rules

Unlike traditional software, AI systems generate outputs based on patterns, data, and context rather than predefined logic. Their success changes with new inputs, so quality and reliability require continuous evaluation, feedback, and iteration.

Together these shifts require a new playbook, built around iteration, intentionally removing friction, and continuous learning.

From strategy to practice

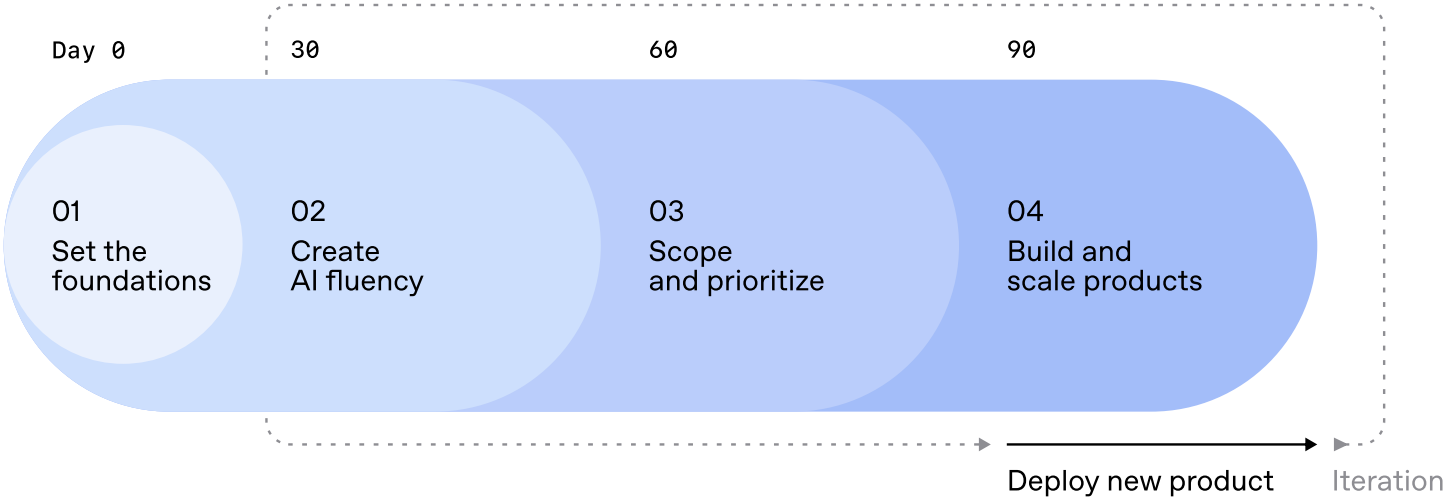
Many companies are already seeing measurable gains from tools like ChatGPT. Productivity improvements appear almost instantly and can quickly scale across teams and organizations.

The challenge is what comes next; turning those early use cases into durable products and workflows that deliver sustained value at scale.

Moving from experimentation to deployment takes more than intent or good ideas. It requires coordinated progress across data access, governance, literacy, and iterative measurement. Together, these capabilities create the structure and momentum required to turn experimentation into new internal and external products.

The organizations that reach production consistently focus on four connected phases that form a repeatable system for scaling AI:

01	**Set the foundations**	Establish executive alignment, governance, and data access.
02	**Create AI fluency**	Build literacy, champion networks, and share learnings across teams.
03	**Scope and prioritize**	Capture and prioritize ideas through a repeatable intake process focused on business impact.
04	**Build and scale products**	Combine orchestration, measurement, and feedback loops to deliver safely and efficiently.



Teams can begin at any phase based on maturity, but progress depends on revisiting and strengthening each as capabilities evolve. These phases are a continuous cycle that reinforces itself as new use cases and learnings emerge to improve the way teams work.

Phase 01
Set the foundations

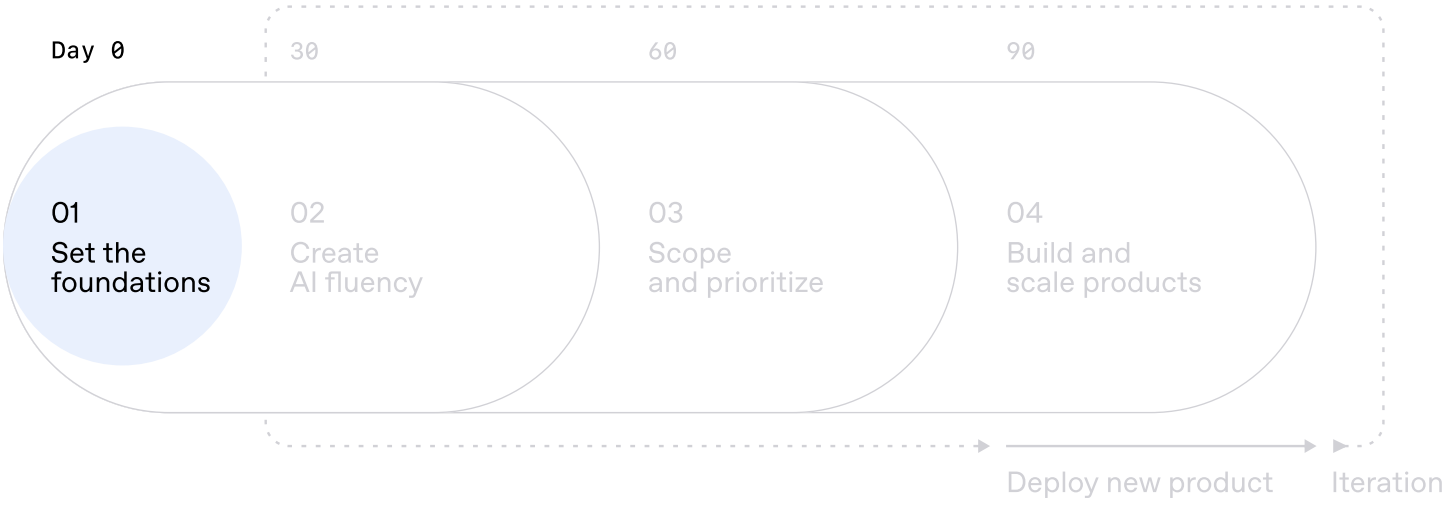
Objective

Establish the structures that let experimentation scale safely and mature over time; executive alignment, governance, data access, and clear goals.

Organizations that successfully embed AI into their businesses begin with sponsorship from leadership, trusted data, and governance that balances speed with risk. Teams that skip this groundwork often move fast at first but stall when gaps appear.

These foundations strengthen with every deployment. Each use case improves governance, data access, and measurement, building momentum that makes future work faster and safer.

Read our **Staying Ahead in the Age of AI guide** for more on how you can drive change as a leader.



How to get started

Step 1 **Assess your maturity**	Evaluate where your organization stands across core enablers; data access, governance, literacy, technical capacity, and use case development. Map the friction points that slow experimentation and prioritize where to invest first.

Start small and strengthen as you scale

Any organization can begin today. Your first step may be a GPT, simple agent, or internal automation that unlocks learning and value. Put lightweight guardrails, data boundaries, a clear owner, and simple ROI measurement in place. As you progress, expand access to data, refine governance, and improve ROI tracking so each build makes the next one faster and more valuable.

Step 2 **Bring executives into AI early**	Leaders who use AI make better, faster decisions. Run short, hands-on sessions where executives use ChatGPT or internal tools to tackle their own tasks and connect insights to business priorities. Visible participation builds confidence and sets the tone for responsible experimentation.
Step 3 **Strengthen access to data**	Reliable data and tools underpin every AI initiative. Start with low-sensitivity datasets to move quickly while improving quality and governance in parallel. Clear classification rules show what can be used safely. As you scope new opportunities use potential ROI as a way to prioritize new data access.
Step 4 **Design governance for motion**	Create a cross-functional Center of Excellence that includes business, IT, compliance, and an executive sponsor to remove roadblocks. Start simple—clarify principles, intake flows, decision rights and escalation paths. Effective governance reduces uncertainty and gives teams the confidence to innovate safely.
Step 5 **Set clear goals and incentives**	Connect experimentation to business outcomes. Early metrics can track time saved or pilots launched; later ones should measure reuse, ROI, and speed to deployment. Align incentives so business and technical teams share credit for impact. As maturity grows, shift from activity-based to outcome-based measures.

How to get started, cont.

With executive involvement, governance, and data access in place, teams move faster and face fewer surprises. Each deployment strengthens the next, creating a foundation for fluency, experimentation, and scaled AI products. The next phase turns this structure into daily habits that accelerate adoption across the organization.

In practice

Chime refined its data inputs to build trustworthy AI.
Early pilots underperformed until data quality improved. Calibrating a custom GPT on Chime's best-performing blog content produced more accurate, on-brand outputs.

Figma created a "compliance fast path" for AI experimentation.
Data guardrails allowed teams to test new tools quickly and responsibly. Lowering the cost of trying revealed real value and sparked innovation across the organization.

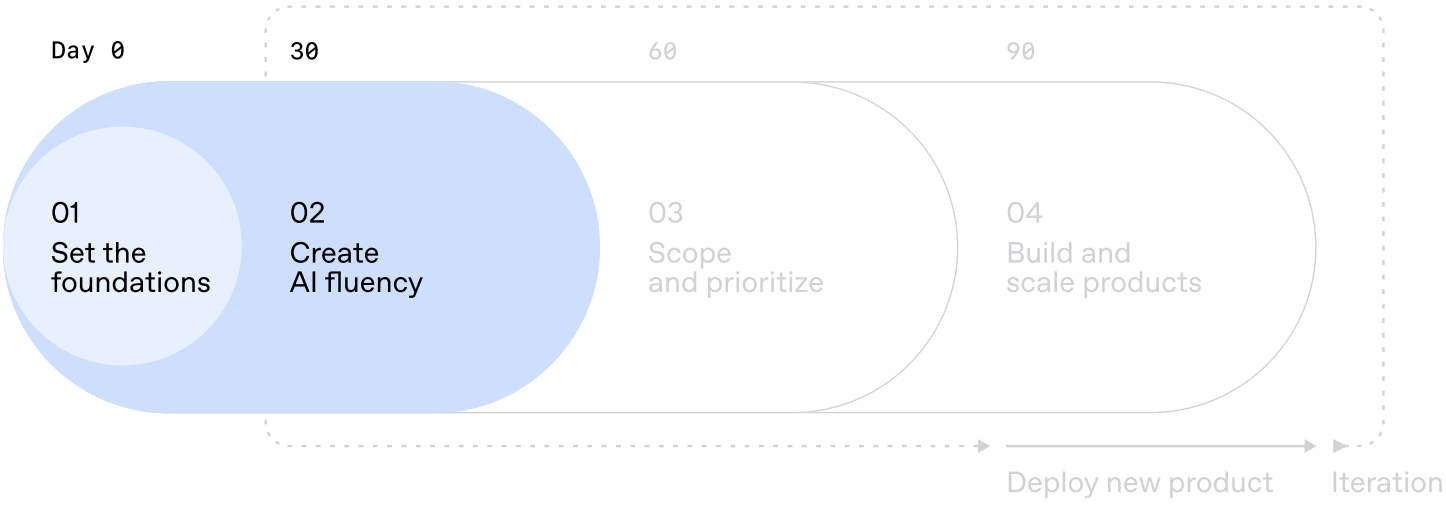
Phase 02
Create AI fluency

Objective

Develop the skills, confidence, and culture that make AI part of everyday work rather than a separate initiative.

Many organizations roll out AI tools before building skills, and adoption and experimentation stalls. The companies progressing fastest treat AI as a discipline that must be learned, reinforced, and rewarded.

This phase builds fluency, deepens practical use, and connects early adopters so knowledge spreads naturally. As AI is horizontal, education is the ignition that sparks use-case creation across every team. Over time, this builds a bench of subject-matter experts who understand both their domains and AI—making them invaluable partners for evaluating, testing, and refining new products. The expertise developed here compounds in later phases, accelerating use-case development, quality assurance, and innovation at scale.



How to get started

Step 1 **Scale learning, then tailor by role**	Start broad, then specialize. Offer short sessions on prompting and everyday use cases before adapting by function. Marketing might focus on campaign ideation, finance on forecasting, and engineering on pair programming. Relevant, role-based learning turns curiosity into practical, high-value use.

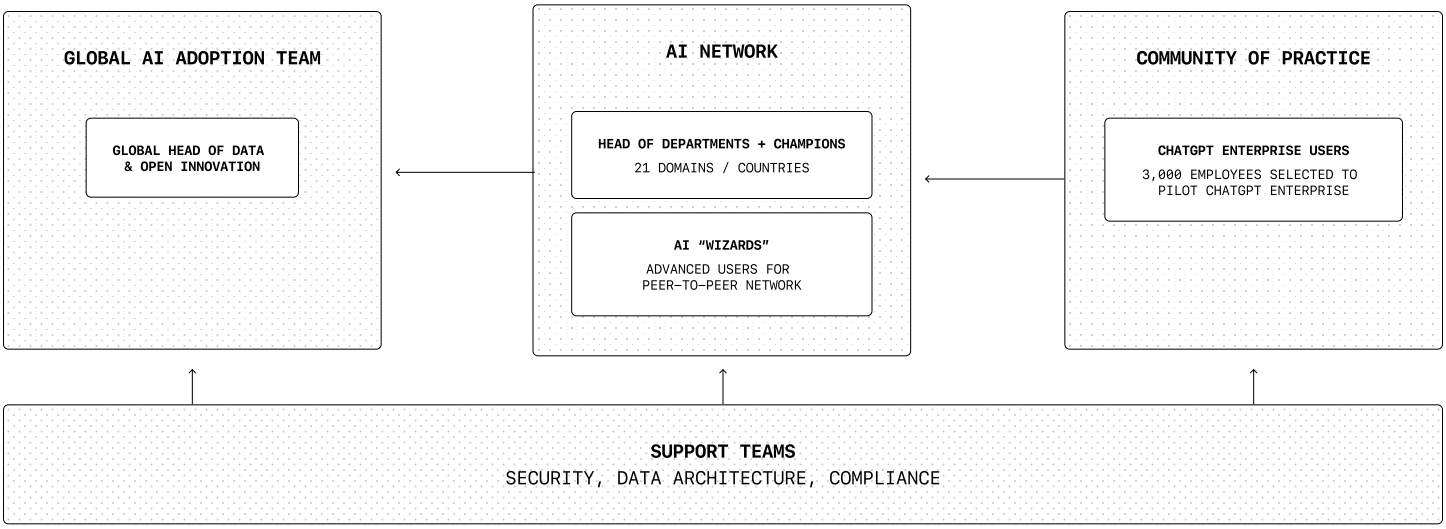
Engineering as engine for change

Engineering teams play a critical role in scaling AI. When engineers master AI-powered tools and agents, they ship faster, improve quality, and become a force multiplier for every new AI product or tool that you decide to build

Step 2 **Create rituals that sustain learning**	Fluency comes from repetition, not instruction. Create consistent spaces for teams to test ideas, share outcomes, and learn from peers. Weekly showcases, short hackathons, or “use case of the week” posts make experimentation visible and normalize progress through participation.
Step 3 **Build champion networks and Subject Matter Experts (SMEs)**	Formalize a champion network of early adopters who mentor peers, document learnings, and connect back to the Center of Excellence. Over time, this distributed system becomes a living learning network that scales faster and sustains curiosity beyond formal training.
Step 4 **Recognize and reward experimentation**	Make success visible. Highlight teams that create value through AI and connect results to professional growth. When curiosity feels rewarded, participation spreads. With fluency and community established, organizations move faster. Shared understanding removes bottlenecks, SMEs elevate quality, and distributed learning turns experimentation into a steady stream of new ideas. The next phase captures these ideas and turns them into a shared, prioritized backlog for development.

In practice

BBVA scaled adoption through a distributed champion network.
The bank expanded from 3,000 to 11,000 ChatGPT licenses by building a champion network across departments and regions. Champions ran hackathons, and shared learnings—surfacing 2,900 GPTs and driving weekly AI use among 83% of employees in 5 months.



The San Antonio Spurs built fluency through hands-on experimentation.
Through internal hackathons and peer-led sessions, teams identified high-impact GPTs and shared use cases across departments. Within months, AI fluency rose from 14% to 85%, and staff collectively saved over 1,800 hours each month.

Wayfair tailored AI training to every function.
Specialized programs equipped technical teams with advanced tools and non-technical teams with practical skills that improved daily productivity.

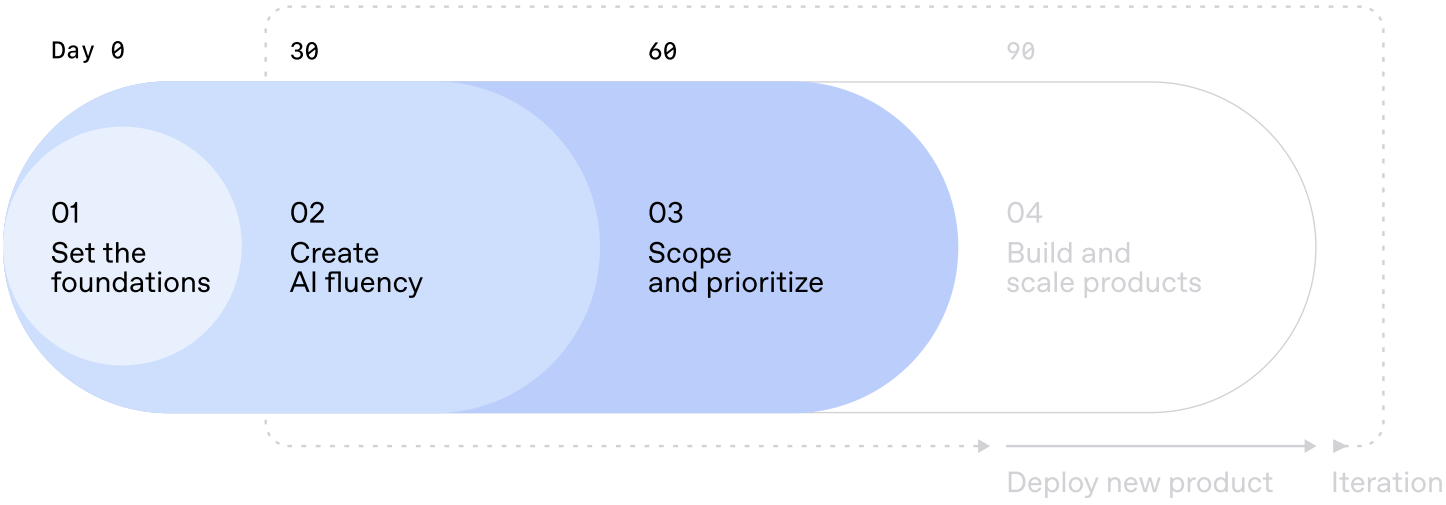
Phase 03
Scope and prioritize

Objective

Create a clear, repeatable system for capturing, evaluating, and prioritizing opportunities across the organization.

Teams are now experimenting and surfacing strong ideas, but many remain isolated. The organizations that move fastest use a shared, transparent process to collect ideas, evaluate them by value and feasibility, and direct resources to where they deliver the most impact.

A strong backlog becomes the organization's compass for AI development. It helps leaders adapt to change, pivot quickly toward higher-value opportunities, and focus resources where results compound. This formal scoping process also rallies the right data, governance, and technical foundations before work begins, ensuring teams build on solid ground and move faster when it's time to execute.



How to get started

Step 1 **Create open channels for idea intake**	Encourage anyone to propose use cases through a simple, visible process. Submissions should capture the problem, who it helps, and the potential value. Intake can flow through champions, individual contributors, or an internal AI hub that models what “good” looks like. Open intake broadens participation and signals that innovation is everyone’s responsibility, not the privilege of a few teams.
Step 2 **Host discovery sessions that turn ideas into prototypes**	Run short hackathons or ideation sessions that bring SMEs, engineers, and designers together to refine ideas. These sessions act as both filters and accelerators—the strongest advance to proof of concept; others feed insights back into the backlog to guide future work.

How to get started, cont.

Step 3
Scope and prioritize systematically

Review ideas regularly using a simple rubric—impact, effort, risk, and reuse potential. This helps identify quick wins while planning for higher-value, deeper integrations.

Score ideas collaboratively with both domain and technical experts to balance feasibility and value. Over time, this creates a shared view of priorities and a backlog that makes every subsequent build faster.

When reviewing ideas, ask:

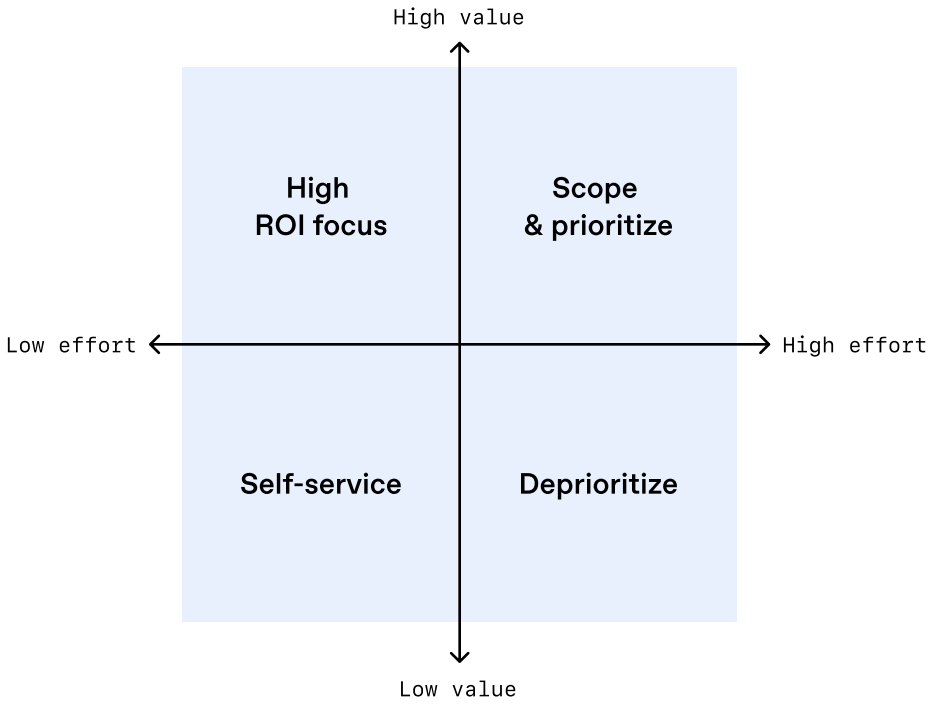
What's the lift to scale and implement it?

What's the quantifiable impact—team, cost, or revenue? Is it strategic?

Is data accessible, and capable of being integrated, is there data for evals?

Are risks manageable?

Does the use case align with LLM strengths?



Step 4
Design for reuse from the start

As you prioritize, look for recurring patterns—code, orchestration flows, or data assets that can support multiple use cases. Designing with reuse in mind compounds speed, lowers cost, and creates a technical memory that turns each project into a launchpad for the next.

Example

A customer-service workflow that retrieves CRM context can later support IT, HR, or retail assistants. Each reused pattern cuts delivery time and boosts ROI across future builds.

In practice

Promega encouraged creativity but also put resources behind high value use cases.
By giving teams the freedom to explore creative use cases, tracking how people used ChatGPT across the enterprise, and running regular workshops and share-outs, Promega surfaced high-impact ideas early and focused resources where they delivered the most value.

With a transparent backlog and consistent scoring, teams move from collecting ideas to delivering value. Shared prioritization and reusable patterns turn experimentation into a reliable pipeline of validated use cases—ready for build.

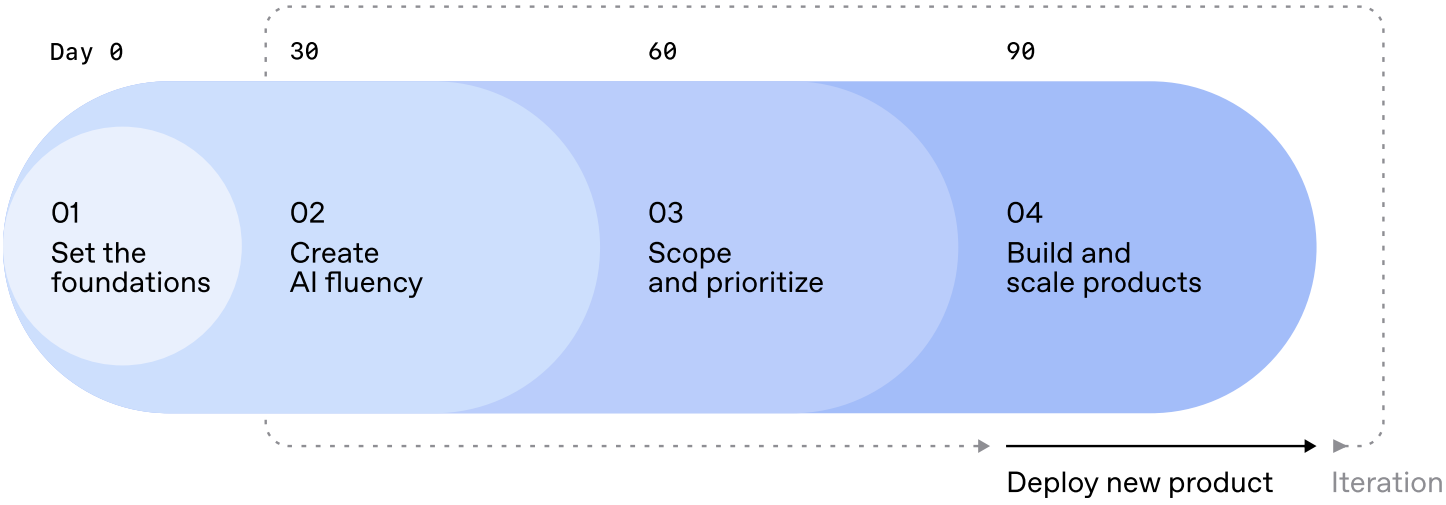
Phase 04
Build and scale products

Objective:

Develop a consistent, reliable method for turning new ideas and use cases into internal and external products.

Building with AI is uniquely powerful because AI systems can learn and adapt rather than relying on fixed logic. AI products improve through repeated iterations of the product itself: each new version is assessed on how it responds to real data, context, and whether it is reliable and cost-effective. As teams run evaluations, integrate new information, and adjust system prompts or workflows, these refinements strengthen the final product.

This phase introduces the structure that supports that iterative rhythm by bringing the right expertise together, creating a clear cadence of iterating, measuring, and improving, and pre-empting where progress may slow. These ways of working make development more predictable and clarify what will scale versus what may be a wasted cycle.



How to get started

Step 1
Build the right teams

The fastest progress comes from small, focused teams with the right mix of expertise. Pair engineers who understand AI tools with subject-matter experts who define success, data leads who ensure access to the right information, and an executive sponsor to remove blockers. Treat this as a standing cross-functional team with shared goals, accountability, and regular touchpoints.

Your engineering engine

An AI-ready engineering team determines how quickly you can move from concept to production. The more engineers who can build, test, run evals, and write prompts, the faster your pipeline grows. When engineering lags, innovation slows.

Step 2
Unblock the path

Most slowdowns stem from access and approvals. By Phase 4, governance and data foundations should be in place, but new use cases often introduce fresh dependencies. Establish direct channels between build teams and IT, Legal, and your AI Center of Excellence to resolve issues early.

Governance decisions should follow the principles set in Phase 1 so teams can move quickly while staying within agreed boundaries.

Step 3
Build incrementally and measure as you go

Building with AI is iterative. Unlike traditional software, generative systems improve through continuous tuning, evaluation, and user feedback. Each build tests assumptions, strengthens data quality, and defines what's ready to scale.

Successful companies break work into smaller capability steps instead of jumping from proof of concept to full production. They use gated checkpoints (typically MVP, Pilot, and Phased Production rollouts) to confirm that builds are useful, performant, and cost-effective before expanding scope.

These checkpoints function on pre-defined OKRs, usually on tool adoption and performance. The latter can be defined both on a system-level as well as on building blocks, i.e., the individual agents or even the components of an agent.

Taking the example of a Q&A agent, that retrieves data, reasons on the questions, formulates an answer and checks against guardrails, this could look like the process in the following table.

How to get started, cont.

Stage	Key goal	Measurement	Decision made
Retrieval	Does the agent reliably find the right information?	Evals on traces/logs, token costs	☐ Continue ☐ Refine ☐ Stop
Summarization and grounding	Does it synthesize clear, consistent, useful, and cited answers? Has the agent followed the right steps and accessed the right data to provide its answer?	Evals on traces/logs + SME review, token costs	☐ Continue ☐ Refine ☐ Stop
Guardrails	Does it stay within approved data, tone, and safety guidelines?	Evals on traces/logs + safety review, token costs	☐ Continue ☐ Refine ☐ Stop

Once proven in a narrow scope, the same checks repeat at pilot and production scale across broader data sources and safeguards.

Evaluating readiness at each step

Evals are run at every stage and before each gate. Their frequency and discipline often determine whether a system moves beyond pilot to a trusted, production-ready solution.

Teams test with real examples of everyday tasks and edge cases, comparing outputs to what a capable teammate would produce. SMEs define what "good" looks like in context and catch issues automated checks may miss.

Instrumentation begins early with basic logs and simple checks, then evolves into automated evaluations for consistency, latency, and cost. When results fall short, teams refine prompts, data, or workflow logic and rerun tests to confirm improvement. Over time, this steady practice builds trust and ensures quality grows alongside capability.

Capture and Reuse What Works

As teams build, they keep their learnings, prompts, connectors, code, evaluation data, and guardrails are captured and then—share them with other teams, and use them again. Reuse reduces lift, improves consistency, and helps new projects reach production faster.

Eventually this creates a shared foundation shaped by real deployments, helping teams move quickly without sacrificing quality and giving governance a stronger footing anchored in proven patterns.

OpenAI Blueprints and Cookbooks

We recently built a set of reusable blueprints and supporting code for the use cases we see most often. They offer a practical starting point so teams can accelerate future builds with patterns proven in the field.

We also consistently share new techniques, patterns, and build guides in our Cookbooks.

A note on scaling simple use cases

Many companies see their first wins with ChatGPT itself before developing complex systems. Simple use cases can deliver quick value when scaled responsibly. Apply the same discipline: validate with SMEs, ensure access to the right data and tools, survey users for feedback, and increase complexity gradually.

Keeping a steady rhythm

Throughout development and after launch, teams maintain a rhythm: build in short loops, evaluate performance on real work, and refine based on results. The same signals determine progress at each stage of deployment.

Focus	What teams look for
Quality and relevance	Outputs are accurate, grounded, and useful, validated by SMEs and evaluations
Responsiveness and reliability	Systems respond quickly and consistently, with logs and traces to surface issues early
Efficiency	Token use, orchestration, and model choice support sustainable scaling
Business value	Time saved, productivity gains, or new capabilities are visible and measurable

This steady cadence keeps progress predictable and ensures each cycle strengthens the next. Teams that pair disciplined evaluation with reuse learn faster, ship with confidence, and scale what works with greater impact. Every deployment reinforces data, governance, tooling, and skills—making the next build easier and more reliable.

In practice

Uber built confidence through continuous measurement.
The team tracked customer experience, resolution speed, automation rates, and productivity gains, comparing AI-augmented workflows with traditional ones through controlled experiments. Metrics like engagement and incremental bookings revealed clear business impact.

Intercom scaled safely through structured evaluation.
Every model powering Fin Voice and Fin Tasks went through offline tests and live A/B trials to validate instruction following, tool accuracy, and coherence before release.

OpenAI refined an inbound sales assistant through human feedback.
Sales reps reviewed every draft response, turning corrections into training data. Accuracy rose from 60 to 98 percent in weeks as the system learned the habits of top performers.

Lowe's accelerated delivery by pairing associates with engineers.
Engineers and store teams co-designed Mylow Companion, an AI assistant now in 1,700 stores, ensuring it fit real workflows from day one.

Booking.com moved from prototype to production in 10 weeks.
By combining OpenAI models with structured and unstructured data, the team built a Trip Planner that handles complex requests and deep personalization.

A new path to production

The path to production is through building the systems, skills, and confidence that turn experimentation into execution and execution into lasting capability. Each phase of this guide strengthens the next, creating a cycle that compounds value with every build.

Organizations that succeed focus on continuous progress. They invest in strong foundations, expand literacy and confidence across teams, and build shared infrastructure that accelerates new experiments. Experimentation evolves into learning, learning into new opportunities, and progress becomes more repeatable.

At OpenAI, we partner with organizations to build these capabilities together. When these foundations are in place, innovation stops being a series of pilots and becomes a source of sustained growth.