fractal

Why Generative AI PoCs struggle to scale to production

White Paper

Contents

Executive summary

Generative AI has the potential to transform various industries by automating tasks, generating innovative solutions, and predicting outcomes. However, many proof-of-concepts (PoCs) struggle to transition from experimentation to production. Organizations facing these challenges often encounter setbacks, but these setbacks can lead to growth and innovation if approached strategically.

Understanding and planning for common obstacles can significantly enhance the chances of successful scaling. This whitepaper identifies typical issues such as misaligned objectives, technological readiness gaps, adoption difficulties, and infrastructure challenges. It also provides practical strategies for overcoming these barriers. It emphasizes the importance of aligning PoCs with business goals, proactively addressing technical and human factors, and promoting a culture that learns from failures.

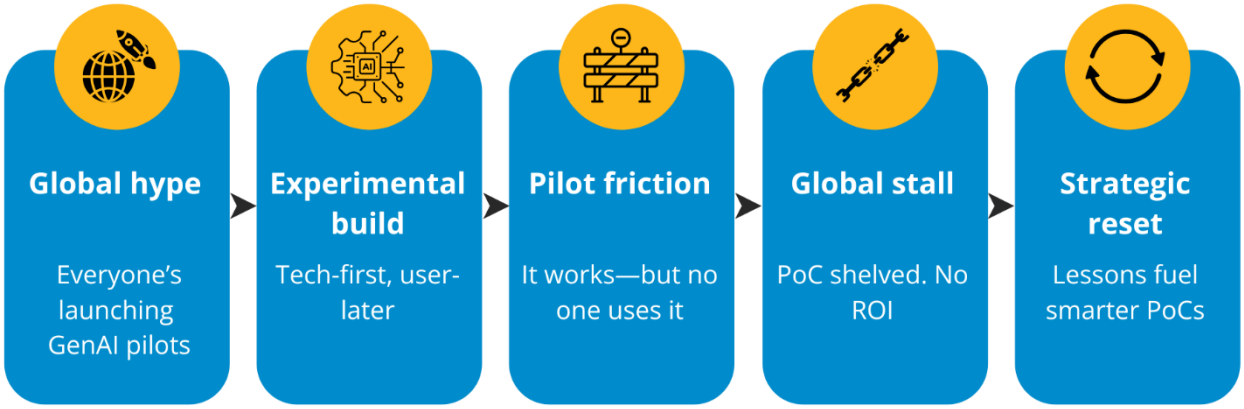
By adopting a structured approach, organizations can improve their ROI and lead the way for impactful deployment of generative AI solutions.

The reality of GenAI PoCs

The belief that all generative AI PoCs must succeed to be valuable is misleading. Most GenAI PoCs are likely to fail, and this should be seen positively. Each failure provides a cost-effective learning opportunity.

Investing in GenAI experiments means accepting that failures help prioritize promising ideas and improve strategies. Every failed PoC provides important insights into what works and what doesn't, forming the bedrock for future successes. This approach highlights the importance of focusing on ideas with real potential while minimizing time and resources spent on less promising projects.

GenAI project life cycle



This visual shows the typical path most GenAI PoCs take across industries. They start with excitement, but stall before reaching production. The stages highlight common reasons why these projects fail and where companies often need to rethink their approach.

Some common issues with failed PoCs include:

- **Overemphasis on technical aspects**: Solutions that focus heavily on technical details often ignore practical usability and impact.
- **Lack of user-centric design**: Designs that fail to consider end-user needs and experiences typically lead to poor adoption.
- **Unrealistic projections**: Setting overly optimistic expectations for outcomes and timelines can result in disappointment when these are not met.

- **Integration challenges**: Solutions that do not integrate well with existing systems or workflows can cause disruptions.
- **Slow adaptation**: Projects that do not quickly adapt to changes or feedback often stagnate.
- **Low return on investment**: Investments that do not yield significant returns fail to justify the resources spent.

Poor implementation of PoCs highlights several pitfalls. Users may not receive the expected results, whether in images or text, leading to inadequate results. Tools that do not fit well into existing processes are less likely to be used, resulting in poor usability. Additionally, users often cannot customize or fine-tune the results to meet their needs, highlighting limited control.

Issues of trust are also common; users may not rely on the tool's output due to poor explainability or auditability. Finally, solutions developed in isolation without engaging users often fail to gain widespread adoption, illustrating insufficient buy-in.

Understanding these common challenges is key for turning initial experiments into successful GenAI solutions that can effectively scale to production.

Key challenges in scaling GenAI PoCs

Here, we explore the specific challenges in scaling successful GenAI PoCs to production. We will also highlight the importance of addressing these hurdles to transform initial experiments into impactful business solutions.

PoCs doomed from inception

Many GenAI PoCs face inherent challenges from the start:

- **Misaligned success metrics:** Overemphasizing technical KPIs without considering business impact leads to solutions that do not address real-world complexities. For instance, focusing solely on performance metrics like accuracy, while ignoring user experience and overall business goals, results in misaligned priorities.
- **Solving idealized problems:** PoCs often tackle issues that are too theoretical and disconnected from practical applications. This results in solutions that may work in controlled environments but fail under real-world conditions.

- **Lack of end-user engagement:** Ignoring critical user requirements such as speed, explainability, and process integration leads to poor adoption. Successful implementations require early and continuous involvement of end-users to ensure the solution meets their needs.
- **Overestimation of GenAI capabilities:** Stakeholder misconceptions about AI's current limitations create unrealistic expectations, setting up projects for disappointment when actual capabilities fall short.

Rapidly evolving tech stacks

Technological advancements can both drive progress and hinder decision-making:

- **Vertical integration of generic solutions:** Integrating tools like RAG and chatbots into existing platforms without customizing them or setting up entirely new platforms splitting tooling can create inefficiencies and disjointed user experiences.
- **Decision paralysis:** The rapid pace of tooling and model advancements can overwhelm decision-makers, leading to delays in implementation. Constantly evolving technologies make it difficult to commit to a particular solution.

Post-PoC cost-benefit analysis

Effective financial assessment is often overlooked in the initial stages of PoCs:

- **Delayed financial viability assessments:** Postponing the evaluation of financial metrics and breakeven calculations can lead to overlooking key success indicators, resulting in missed opportunities for scaling.
- **High costs of agentic solutions:** Despite the declining "cost of intelligence," expensive solutions may become feasible as technologies evolve. It's important to balance the immediate costs with potential future benefits.

Technological readiness gaps

Some PoCs face challenges due to the current state of technology:

- **Niche use cases:** Specific applications requiring specialized data or high accuracy levels can struggle with existing models. These PoCs should be kept in the backlog for reevaluation as newer, more capable models become available.

- **Backlog for reevaluation:** Superior models with multi-modality and better reasoning are emerging monthly, offering opportunities to revisit and potentially succeed with previously failed PoCs.

The "Curse of the successful PoC"

Scaling a successful PoC presents several unique challenges including:

- **Technical scaling barriers**
    - **Foundational infrastructure gaps:** Ecosystem integration and hyperscalar capacity are crucial for scaling but often overlooked during initial PoC development.
    - **Skill mismatches:** The skills required to scale a PoC differ from those needed to build it. Bridging this gap is essential for successful implementation.
- **Hyperscalar capacity limitations:** Quota restrictions or high incremental capacity costs can hinder the scaling process, though this is likely to improve rapidly.
- **Risk and compliance concerns:** Security, regulatory hurdles, and adversarial usage must be addressed. PoCs may lack comprehensive risk management strategies, causing delays or failures in scaling. Implementing a secure & robust MCP can mitigate these risks by ensuring that security measures are consistently applied throughout the development and deployment phases.

Human adoption hurdles

- **End-user unfamiliarity:** Most end-users are still largely unexposed to GenAI, requiring time and training to become comfortable with new technologies.
- **Localization challenges:** Translation, cultural adaptation, and bias mitigation are critical for global organizations but often overlooked in PoCs. Market-specific data with unique features and biases need to be addressed.
- **Process reengineering costs:** Integrating GenAI into existing workflows requires modifying tools and processes, which can be time-consuming and expensive. These costs may not be considered in the initial cost-benefit analysis.
- **Leadership inertia:** Slow decision-making and resistance to change from leadership can impede the adoption of successful PoCs.

Other adoption challenges include:

- Resistance to adopting tools perceived as immature or disruptive to existing processes.
- High switching costs with no clear incentives for adoption.
- Lack of widespread use or endorsement can hinder acceptance.
- Concerns about the reliability of the tool's output, fears of job replacement, and doubts about enhancing creativity.
- Tools that are not exciting, visually appealing, or simple to use fail to capture user interest.

Strategies for successful scaling

Building a framework that supports the scaling of GenAI PoCs from experimentation to full production involves several key strategies:

Embrace failure as a learning tool

Failure should be embraced as an opportunity for growth and learning. Each failed PoC provides valuable insights that can enhance future projects.

- **Document lessons from failed PoCs:**
  - Create a knowledge base from documented failures to identify common pitfalls and guide future projects.
  - Conduct root cause analysis to understand primary reasons for failure. These insights ensure that future PoCs benefit from past experiences, streamlining the scaling process.
- **Promote a culture of experimentation:**
  - Encourage a culture that values experimentation and iterative improvement where failure is accepted as part of the innovation process.
  - Accepting that failure is part of the process supports a more resilient approach to scaling AI solutions.

Avoid doomed PoCs from the start

Early alignment of PoCs with business objectives and stakeholder expectations can prevent many common pitfalls.

- **Align PoCs with business objectives:**
  - Collaborate closely with business leaders and end-users during the PoC definition phase to ensure goals align with business objectives.
  - Define success metrics based on business impact, rather than technical performance, to ensure practicality and relevance.
- **Conduct upfront cost-benefit analysis:**
  - Engage in thorough cost-benefit analysis and user-journey reimagination workshops before initiating PoCs.
  - Identify critical requirements and ensure they are met.
  - Conduct reasonable upfront cost-benefit analysis to determine the worthiness of the PoC and identify break-even points.
- **Educate stakeholders on realistic capabilities:**
  - Provide access to tools like chat, code, and image generation to familiarize stakeholders with GenAI capabilities.
  - Conduct workshops and training sessions to set realistic expectations and educate stakeholders about AI's current limitations.

Focus on specialized use cases

Targeting unique data, workflows, or ecosystem needs rather than duplicating generic solutions can maximize the impact of Generative AI.

- **Avoid duplicating vertically integrated solutions:**
  - Focus on specialized use cases that showcase the uniqueness of data, workflows, or ecosystem needs and leverage Generative AI's strengths.
  - If none are unique, adopt existing generic solutions.
- **Stay in the learning game:**
  - Regularly review your current and state-of-the-art tech stack for capabilities that address niche and specialized needs.

- Accept that certain PoCs serve to learn the tech stack evolution, not necessarily for immediate value generation and revive past experiments as AI capabilities improve.

Build scalable foundations early

Investing in technical infrastructure and skill development early in the process lays the groundwork for successful scaling.

- **Invest in technical infrastructure:**
  - Proactively build the necessary technical infrastructure for scaling, including ecosystem integration and hyperscaler capacity.
  - Plan for localization, translation, and cultural adaptation early in the project to avoid scaling barriers.
- **Develop relevant skills:**
  - Address skill gaps by providing training and hiring specialists in scaling technologies.
  - Ensure the team is equipped with the necessary skills to scale PoCs effectively.

Drive adoption through change management

Successful scaling requires more than just technical readiness; it needs the active involvement and acceptance of end-users and leadership.

- **Co-create solutions with end-users:**
  - Collaborate with users on problem-solving to build familiarity and trust with the solution. Involve them early to ensure the solution meets their needs.
  - Provide comprehensive training and support to help users acclimatize to new technology.
- **Modify tooling and processes:**
  - Build or modify tools and processes to match reengineered solutions, helping with seamless integration.
  - Secure leadership support to drive organizational change, build a strong leadership coalition to promote adoption of successful PoCs.

Incorporating these strategies helps organizations address the challenges of scaling Generative AI PoCs. This ensures that projects can transition from experimental stages to impactful production-level deployments, thereby maximizing ROI and driving continuous improvement.

Future Outlook

Generative AI continues to evolve, with falling costs and improved model capabilities such as multimodality and reasoning. This ongoing advancement opens opportunities to revisit shelved PoCs, transforming them into viable solutions.

As the ecosystem matures, early collaboration will foster shared innovation and broader adoption. Organizations must stay adaptable and regularly revisit past experiments to leverage the latest advancements and remain competitive.

Organizations should keep up with the latest technological improvements and update PoCs to take advantage of new capabilities. Working with partners and sharing insights can increase innovation and improve adoption rates. Embracing ongoing education and adjusting strategies as AI evolves are also important.

Generative AI holds the promise to drive meaningful growth and create impactful solutions. Organizations that adjust and integrate these advancements will secure a lasting competitive edge.

Conclusion

Scaling generative AI PoCs can be challenging due to issues like misaligned objectives, changing technology, adoption difficulties, and infrastructure limitations.

Organizations can address these challenges by viewing failure as an opportunity to learn and aligning PoCs with business objectives. Additionally, focusing on specialized use cases, building strong foundations early, and promoting effective change management can help overcome these obstacles.

A clear, strategic approach is crucial to improving ROI and realizing the potential of generative AI. As technology progresses and practices evolve, scaling generative AI solutions will become more viable, leading to significant innovation and growth.

Organizations that remain flexible and proactive will be better equipped to benefit from these advancements, securing a competitive advantage in a constantly evolving environment.

fractal.ai

info@fractal.ai @fractalai linkedin.com/company/fractal-analytics