Factored

A COMPREHENSIVE GUIDE TO SUCCESSFULLY DEPLOYING AI PROJECTS

ISSUE 02

What
are the
successful
building
blocks you
need to
**maximize
your AI
investment?**

Table of Contents

Introduction

The adoption of Artificial Intelligence technologies by companies both big and small is only increasing.

According to the International Data Corporation's (IDC) 2022 AI InfrastructureView[1] survey, 31% of companies say that they now have AI in production, while the majority are actively piloting AI technologies.

Why the increase in adoption? Because it's leading to increases in business profitability. McKinsey found that 27% of businesses claim that they can attribute at least 5% of their earnings before interest and taxes (EBIT) to Artificial Intelligence.

But while there are big gains to be made in this field, the problem is that deploying AI projects can be a challenge. In fact, over 80% of Artificial Intelligence projects never make it off the ground[2].

That's because, oftentimes, companies face significant challenges when it comes to infrastructure management, model monitoring and retraining, and versioning.

Infrastructure **Management**

Simply deploying an API isn't enough; there must be adequate infrastructure to support scaling up when needed, including if and when your data and/or the number of users increases—your single API won't be able to support this and the model will die.

Model Monitoring **& Retraining**

Data changes over time and models become obsolete. Therefore, you need proper observability in order to monitor data changes; additionally, model decay is required to know when and how to retrain models.

Versioning

If your company doesn't have the right systems in place, your data science teams won't be able to properly version and share their AI models, code, and datasets—and AI models are pointless if other people can't use them. You don't want to have to keep track of all the models and datasets created by teams so that they don't reinvent the wheel every time they receive a new business request.

"Without the right systems in place, your data science teams won't be able to properly version **and share their AI models, code, and datasets."**

So, how exactly can you deploy your business's AI initiatives effectively and efficiently and set yourself up for success?

To help you achieve exactly this, we pulled together the most common pitfalls to help your business avoid future mistakes.

We'll first take a look at what businesses are up against, and then review how your organization can better develop and share intricate models to predict, automate, and categorize processes for maximum efficiency.

A Lack of Technical Expertise is **Coming Back to Haunt Businesses**

In addition to organizations lacking the systems that allow their data science teams to properly share AI models, companies and technical leaders are facing a myriad of other challenges and pain points when it comes to deploying AI projects (we're guessing you can relate?).

One of these issues is that as machine learning models continue to exponentially increase in size, these large models often don't fit on commodity hardware; what's more, they can be slow and expensive to run or locked into proprietary APIs. This means data science teams can't dive into the code to evaluate or improve the Artificial Intelligence models.

"...without doing the work necessary in order to adapt AI models for one's business, data **isn't fine-tuned to fit the contexts in which it'll be used.**"

Furthermore, without doing the work necessary in order to adapt AI models for one's business, data isn't fine-tuned to fit the contexts in which it'll be used. This lack of accurate data was rated on par with a lack of internal expertise as setbacks to further implementing AI initiatives across organizations in a Dun & Bradstreet[3] survey.

It's important to remember that an off-the-shelf AI model won't understand the nuances of the enterprise environment. It simply won't be useful for decision-making until it has the required data, because it won't understand the context of the user intent.

So what do you need to do in order to solve these types of problems?

Getting Your Business **Back on Track**

There are some basic first steps you can take to get your Artificial Intelligence initiatives back on track.

First and foremost, focus on strategy. Start by looking at your business strategy, and then work backwards from there. Ask yourself: How does Artificial Intelligence fit into what we want to achieve as a company in the next month, six months, year, five years, and ten years?

Once you have that alignment, it's time to start focusing on the following...

Bringing **About Buy-In**

In order to get your project off the ground, once the project has been implemented, and throughout the iteration process, you're going to need to have stakeholders on your side.

In order to do this, make sure all stakeholders understand why you need to deploy Artificial Intelligence initiatives and—later on in the process—how to use what you've built, and use it well. After all, you'll need their continued buy-in to have a clear path forward to continue iterating and expanding your AI initiatives.

Business leaders will want data to confirm 1. The need for the project and 2. The project's success and its contribution to the bottom line. So you need to make sure it is used correctly and efficiently.

People also forget things, so don't be shy about scheduling quarterly reminders or check-ins with the teams using the new tool to ensure that everything is working as it should be and that they're using it in the most effective way.

Tagging **the Team**

For AI initiatives to work, you first need the right team in place.

Don't think that a "one-man-army engineer" can solve all of your problems. It's simply not possible, and—even if it was—it would be extremely inefficient. What your business needs is a team to successfully define, build, deploy, and manage your AI strategy. The ideal team would be made up of:

- **Data engineers** to help build the data pipelines that feed the models

- **Data scientists/Machine learning engineers** to build and deploy models according to the business requirements
- **Data analysts** to maintain the alignment between AI results and business objectives
- **DevOps/Cloud engineers** to manage deployments, infrastructure and security

It's important to understand that AI projects are a mixture of data and technology. Good teams understand that you have to start small, so don't expect to get the perfect model in a month. Remember that it's an iterative process: While you can get 80% done in 20% of the total time, the remaining 20% takes the other 80% of time.

Last but not least, don't underestimate the value of having good data. If your data is worthless, you won't be able to extract any value from it. After all, AI is not magic; AI is advanced pattern finding in data. Which brings us to our next point...

What your business needs is a team to **successfully define, build, deploy, and manage your AI strategy.**

Detecting **the Data**

In this day and age, it's not enough to simply own data. Just because data is in the data warehouse doesn't mean you can start creating AI models, setting up dashboards to help make key decisions, and delivering insightful business analysis.

For data to become useful to a business, one needs to first find the applicable data, then clean, order, format, partition, and organize it so they can have a solid foundation to work from and can effectively run these models on. That way, you can present it in a digestible way for all stakeholders.

Andrew Ng, our founder, calls this the "data-centric approach" to AI[4], which advocates that AI has reached a point where data is more important than models. In other words: If artificial intelligence is seen as a system with moving parts, it makes more sense to keep the models relatively fixed while focusing on quality data to fine-tune the models, rather than continuing to push for marginal improvements in the models.

However, expecting organizations to train their own custom AI models isn't realistic, which is why businesses need tools that can help them build their own models, engineer the data, and express their domain knowledge.

If you've ever felt the pain of an incomplete data catalog, stale data, wasted loads of computing power, or not being able to find relevant data, then you know that the road from gathering data to having tangible business outcomes isn't a straightforward one.

HERE'S HOW YOU CAN START MAKING YOUR DATA WORK FOR YOU:

Your data engineer[5] builds the systems to collect your business data into the data warehouse.

You iterate the data curation, as well as your analyses, dashboards, and machine learning models via data cataloging, data quality assessment, business logic validation, and business-aware transformations.

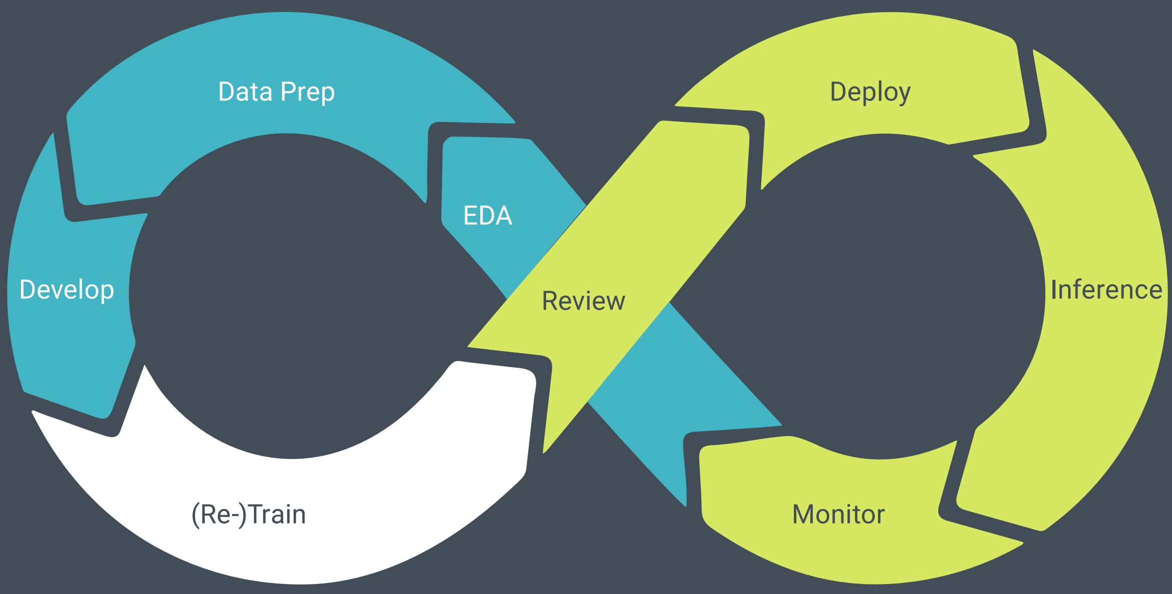
You are now able to test, transform, model, and catalog data so that it's ready for machine learning engineers (MLEs), data scientists, data analysts, and/or business stakeholders to consume.

If artificial intelligence is seen as a system with moving parts, it makes more sense to keep the models **relatively fixed while focusing on quality data to fine-tune the models...**

Molding the **Methodology**

The emerging discipline of MLOps, or machine learning operations, is a core function of Machine Learning engineering that's focused on streamlining the process of taking machine learning models to production, and then maintaining and monitoring them. MLOps is a collaborative function, often comprising data scientists, devops engineers, and IT.

MLOps Cycle



Via databricks[6]

By adopting an MLOps approach, data scientists and machine learning engineers can:

- Implement continuous integration and deployment (CI/CD) practices with proper monitoring, validation, and governance of ML models
- Collaborate and increase the pace of model development and production
- Address the disparate and siloed nature of AI development by establishing practices for collaboration between data scientists .

By simplifying AI management processes, MLops can help businesses:

- Automate the deployment of AI models into the core software systems of an organization
- Define and maintain data and features to enable multiple teams to access the data
- Improve the efficiency of teams developing machine learning models
- Version code, data, and models
- Ensure models are reproducible
- Track the different experiments performed
- Monitor models in production
- Ensure the best practices for deploying models are used
- Capture and utilize user feedback
- Retrain models in production

Along with implementing MLops, here are some additional tips to make your AI methodology work for everyone involved:

Instrument

Artificial Intelligence and Machine Learning systems can be affected by a lot of variables and—sometimes—bad predictions. Unfortunately, high CPU and RAM usage are not enough to successfully assess why a system is failing. Therefore, try to instrument your solution as much as you can to increase its level of observability.

Scale

This is a must for every project in production. You can have the best performing AI model out there, but if your infrastructure can't keep up with users' requests, it will be considered a bad implementation.

Automate

The nature of every AI project is constantly changing, so it will be a tedious process to manually deploy each one of those iterations. With this in mind, our guidance is to try to automate as much as you can in your AI development cycle. Make sure to also pay special attention to automating the deployment stage specifically.

Testing to Perfection

Never underestimate the importance of testing. Once you've ensured that what you're building is scalable, test it several times, and ask certain stakeholders to test it for you and provide feedback. Sometimes, the non-technical end-users might spot something your highly technical team might have missed.

For example, when Amazon[7] automated its recruiting process with AI in 2014, the company realized that the new recruiting system wasn't rating candidates fairly and showed bias against women, forcing them to ditch the algorithm altogether. Had they not continually been testing this, they never would have known!

OUR BEST PRACTICE ADVICE FROM FACTORED?

While best practices tend to be project or application specific in order to truly be relevant (a few such examples to elaborate: Explainability is crucial in medical applications, but irrelevant in many language-related applications. Or bias is crucial in dialog applications, but irrelevant when it comes to detecting manufacturing defects; and while fully-automated solutions are desirable for some businesses, others prefer to keep the human-in-the-loop and use AI as an assistant to improve the human's productivity.), here are some general best practices that apply to all projects:

Focus on maintaining high quality data, both for training and validating models: Guarantee that you have sufficient representation of all your classes and that the samples are properly labeled.

Standardize your development and production environments: Make it so that developers work in an environment as close to production as possible.

Define tests for your data: Make sure any transformations do what you expect them to do.

Keep track of all the artifacts related to AI: Data, models, code, preprocessing pipelines, tokenizers, post-processors, and version them appropriately.

Track your experiments: To make it easy to compare between strategies and understand what worked and what did not; this is very helpful to continue improving models.

Start small, receive feedback and iterate: This has the added benefit of giving the company incremental value, instead of having to wait six+ months to get some value.

Keep track of software and model metrics of your models in production: How many requests are you getting? What is the latency? What is the performance of the model? Measuring this will let you know when to scale up and when to re-train.

When working on projects for our clients, we always make sure to automate as much as possible so we don't have to manually deploy each iteration that crops up. We also always ensure our projects and models are built to be scalable, on infrastructure that we know will keep up with users' requests once deployed. And even then, we test, test, and test again!

Where to Go **From Here**

Don't jump in headfirst to building and deploying an AI project if you don't have the correct foundation in place. Your first step is to get experts on your side—people who've done this hundreds of times before—like those at Factored.

That's because, armed with the right experts, your company can get crucial AI initiatives off the ground and into implementation in no time.

At Factored, our team can bring value by managing the entire AI lifecycle: From developing and delivering AI models, to integrating those models efficiently and continuously with your business so that production can be scaled and you can see prompt results.

We know this because we work the entire lifecycle and know how the foundational setup affects future work, such as building pipelines to automate the AI implementation process and adjusting or retraining AI models that have degraded over time.

HERE'S HOW WE DO IT:

We identify the issues with the data and propose solutions to fix it.

We understand the data and the challenges of AI; once we understand the business, we can help define and scope projects and propose timelines and deliverables that add incremental value to the client.

We know the different regimes of data and can suggest how we should proceed: Is labeling data the best approach? Can we use the existing data? How much data do we need? Should we use existing models or create new ones? Do we think AutoML would work for this use case?

We understand the requirements for deploying solutions and know the techniques and technologies required to make them work: Do we need extremely low latency (fast predictions), or do we need high throughput (the ability to process many requests quickly)?

We understand that data changes and models change, and we help clients set up proper infrastructure for monitoring and retraining models.

Once we understand the business, we know the "tips and tricks" required to solve the problem successfully; this is something that automated ML solutions could never do (at least not in the short-to-medium term).

We love research, which means that with us, you'll get the best solutions to your problems. We don't stick with what we know; we understand the problem and research for the best solution to the problem.

Our experts have helped businesses understand the behavior of AI models so that they can better:

Build model sets to measure credit health of SMEs using only transactional data from bank accounts.

Predict the price movement of stocks and the probability of home equity line of credit default.

Build a natural language processing solution in order to use online customer reviews to determine the strengths and weaknesses of particular services.

Create a system connected to a dashboard where all stakeholders can extract insights about the efficiency of their services based on raw text reviews.

Are you ready to get started? Book a meeting with Factored **to speak with our AI and data experts today.**

+ BOOK A CALL

Source References

1. https://www.idc.com/getdoc.jsp?containerId=prUS48870422

2. https://venturebeat.com/2020/12/16/the-future-of-ai-deployments-reaching-production-is-bright-in-2021/

3. https://www.prnewswire.com/news-releases/artificial-intelligence-is-creating-jobs-dun--bradstreet-survey-finds-300774141.html

4. https://venturebeat.com/2022/03/21/andrew-ng-predicts-the-next-10-years-in-ai/

5. https://factored.ai/2021/08/12/the-responsibility-of-a-data-engineer/

6. https://databricks.com/glossary/mlops

7. https://www.techtarget.com/searchaws/news/252475880/How-Amazon-HR-influences-hiring-trends