

Responding to the top 10 causes of stalled AI/ML projects ¹

Yogi Schulz, BComm

Organizations observe too many AI/ML projects stalling and eventually being cancelled. This article discusses the ten most common causes of this unfortunate situation and what project managers can do to correct the problems. Project managers and teams can evaluate their AI/ML project work in light of these ten causes, address them in their project plan and take corrective action to avoid stalling their AI/ML projects. The ten most common causes are:

1. Deteriorating business case
2. Underestimating model training
3. Lacking data quality
4. Addressing data integration
5. Managing data volumes
6. Incorporating iterative development
7. Responding to data shift
8. Underspecifying the model
9. Validating results
10. Complicating the algorithm

Deteriorating business case

Organizations approve an AI/ML project based on an appealing business case. As the project proceeds, some events, such as the following, can undermine the business case:

- Discovery of additional complexity in the solution leading to a significant increase in the cost-to-complete forecast and the annual operating cost estimate.
- Recognition that the solution requires more data the organization does not own and must pay for.
- Changes in customer expectations or preferences.
- Actions by competitors require a more elaborate response.

The best practice response for project teams is to update the business case and determine if continuing the project is still appealing. Allowing projects that are unlikely to produce a net benefit to drag on wastes the organization's resources.

¹ How to cite this article: Schulz, Y. (2023). Responding to the top 10 causes of stalled AI/ML projects, *PM World Journal*, Vol. XII, Issue X, October.

Underestimating model training

Organizations often underestimate the work that goes into training AI/ML models. AI/ML project teams tend to underestimate the following:

- Data scientists' efforts required to train models.
- Business expertise and effort needed to collaborate with the data scientists.
- Software development effort even when significant use of open-source libraries is planned.
- Data volume and variety required.

The best practice response is for project teams to arbitrarily double the model training estimate during project planning. If this increase materially undermines the project's business case, the project manager should recommend cancelling the project.

Lacking data quality

Too frequently, data quality issues in internal datastores stall AI/ML project progress. The common issues are:

- Inaccuracy – The available data is not accurate enough to produce reliable results from the model.
- Insufficient richness – The available data often lacks values and rows required to train the model effectively.

Improving internal data quality requires the organization to recognize the value of its data and improve its data stewardship processes. Improving data richness to achieve high-quality training data requires the project team to create synthetic data. These corrective actions add schedule and cost to the project.

Addressing data integration

Too frequently, these data integration issues stall AI/ML project progress:

- Data accessibility complexity – The required data is scattered across too many internal and external datastores housed in multiple data centers and clouds to support reliable real-time access.
- Inability to integrate data – Differences in data models, values and formats across datastores are too complex to integrate the data cost-effectively.

Addressing complex data integration issues requires the project to abandon the preferred approach of accessing data in its original location. Instead, the project team will be forced to extract, transform and load the data into a data warehouse. Adding the creation and operation of a data warehouse to the project scope adds schedule and cost.

Managing data volumes

As AI/ML models become larger by feeding them more data and increasing the number of parameters, they become better and better. However, increasing the data volume will stall the AI/ML project's progress by increasing the:

- Cost and elapsed time for each model test run.
- Effort to evaluate the results of each test run.

Managing growing data volumes requires the following:

- Consumption of more computing resources.
- Implementation of a graph DBMS.
- More attention to optimizing the algorithm.

Incorporating iterative development

Developing AI/ML applications is an iterative or agile development process, even though it's not widely recognized. It is difficult to predict precisely:

- How many algorithm-tuning iterations will be required to achieve sufficient reliable functionality.
- Which data or parameters you will need to complete the model until you learn from the algorithm training iterations.

Incorporate iterative development into the project plan by explicitly including at least five iterations. Plan for more if your team is new to the problem space of the project.

Responding to data shift

Data shift is the term used when the difference between the training and real-world data is too much for the AI/ML model. The result is unreliable model recommendations and false positives and negatives.

Respond to data shift issues by creating a richer training dataset and tweaking the algorithm.

Underspecifying the model

Underspecification of an AI/ML model produces multiple results with high confidence. It occurs when model versions differ in small, arbitrary ways, such as in:

- The random values assigned to the neural network nodes before training starts.
- The way training data is selected or represented.
- The number of training runs conducted until the model is deemed ready for production use.

These minor, often random, differences are typically overlooked because they don't affect model results during training. However, these subtle differences can lead to enormous variations in model performance in the real world.

Underspecification is addressed by conducting additional tests using different datasets focusing on specific features. This testing work can materially increase the cost of AI/ML model development to reduce the risk of nasty surprises during the production use of the model.

Validating results

If your AI/ML model is not delivering the intended results, it's most likely because:

- The training data was incomplete or skewed, creating unreliable model results.
- The algorithm selected is not appropriate for the problem space.

These issues are best addressed by carefully validating results after every test run and not deferring validation until later in the project. In most cases, the corrective action will be to curate additional training data.

Complicating the algorithm

Recent advances in AI/ML algorithms have come chiefly from minor tweaks and significant additional computing power rather than breakthrough innovations. Examples of limited algorithm advances, where the last breakthrough is about a decade old, include:

- Information retrieval algorithms – are used heavily in search engines.
- Pruning algorithms – make neural networks more efficient by trimming unneeded connections.
- Loss functions – specify the algorithm's objective mathematically.
- Generative adversarial networks (GANs) – pair neural networks in a create-and-critique cycle.
- Recommendation algorithms – apply neural networks to product and entertainment recommendations.

Addressing AI/ML algorithm issues requires the AI/ML project team to review if the complexity that's been added to the algorithm during the project has, in fact, produced accuracy and performance benefits. Simplifying the algorithm used to generate the AI/ML model will improve performance with no reduction in accuracy.

When project managers insist that their teams evaluate their AI/ML project scope for risks related to these ten causes and take corrective action, they avoid stalling or cancelling their AI/ML projects.

This article is derived from the following article: [Top 10 causes of stalled AI/ML projects.](#)

About the Author



Yogi Schulz

Calgary, Alberta, Canada



Yogi Schulz has over 40 years of Information Technology experience in various industries. Yogi works extensively in the petroleum industry to select and implement financial, production revenue accounting, land & contracts and geotechnical systems. He manages projects that arise from changes in business requirements, from the need to leverage technology opportunities and from mergers. His specialties include IT strategy, web strategy and systems project management.

Mr. Schulz regularly speaks to industry groups and writes a regular column for [IT World Canada and for Engineering.com](#). He has written for Microsoft.com and the Calgary Herald. His writing focuses on project management and IT developments of interest to management. Mr. Schulz served as a member of the Board of Directors of the PPDM Association for twenty years until 2015. Learn more at <https://www.corvelle.com/>. He can be contacted at yogischulz@corvelle.com

His new book, co-authored by Jocelyn Schulz Lapointe, is "[A Project Sponsor's Warp-Speed Guide: Improving Project Performance.](#)"