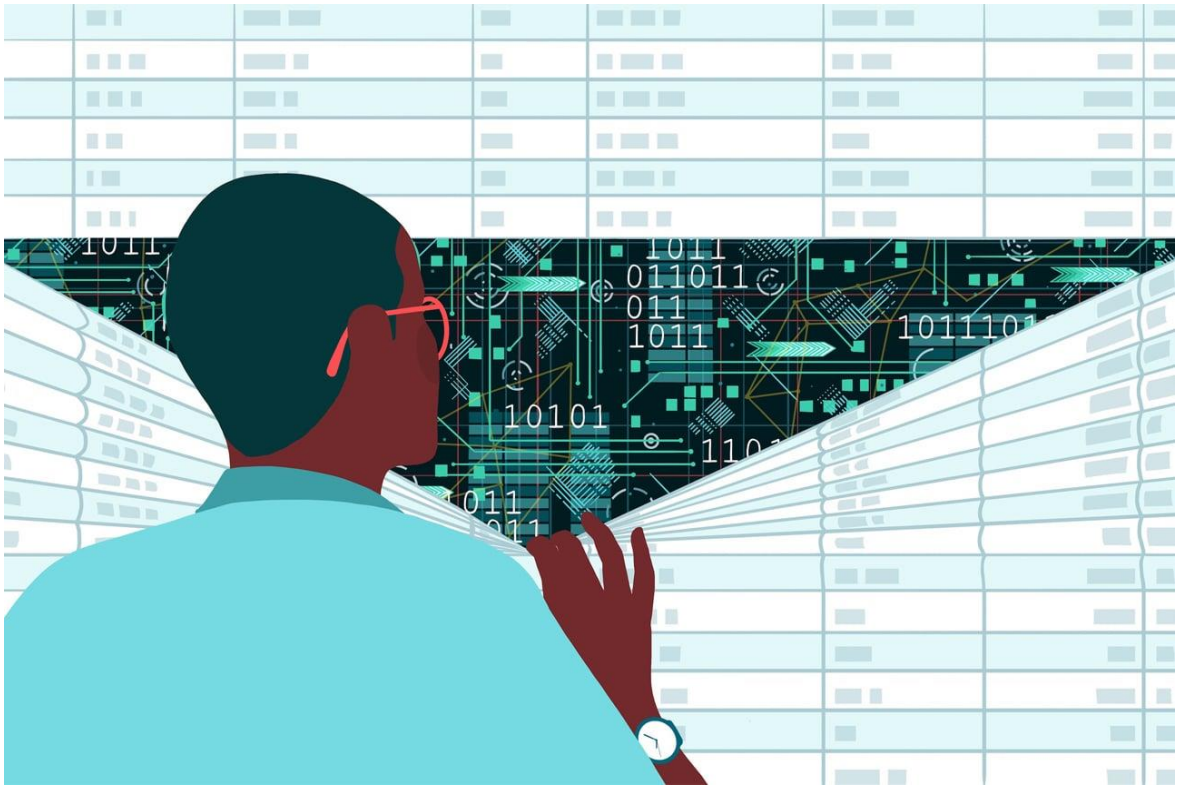
What Managers Should Ask About AI Models and Data Sets

Business leaders must make the right calls on whether and how broadly to deploy AI models. Ask these tough questions.

Roger W. Hoerl and Thomas C. Redman December 19, 2023Reading Time: 12 min



The power of AI and the machine learning models on which it is based continue to reshape the rules of business. However, too many AI projects are failing — often after deployment, which is especially costly and embarrassing. Just ask Amazon about its facial recognition fiascos, or Microsoft about its blunders with its Tay chatbot. Too often, data scientists write off such failures as individual anomalies without looking for patterns that could help prevent future failures. Today's senior business managers have the power — and the responsibility — to prevent post-deployment failures. But to do so, they must understand more about the data sets and data models in order to both ask the right questions of AI model developers and evaluate the answers.

Maybe you're thinking, "But aren't data scientists highly trained?" The vast majority of training for today's data scientists focuses on the mechanics of machine learning, not its limitations. This leaves data scientists ill-equipped to prevent or properly diagnose AI model failures. AI developers must gauge a model's ability to work into the future and beyond the limits of its training data sets — a concept they call *generalizability*. Today this concept is poorly defined and lacks rigor.

A saying in analytics asserts that model developers and artists share the same bad habit of falling in love with their models. Data, on the other hand, doesn't get the attention it requires. For example, it's all too easy for AI model developers to settle for readily available data sets rather than seeking ones more fit for the problem at hand.

Senior business managers, lacking advanced degrees in technical disciplines, are even less equipped to spot trouble related to AI models and data sets. Yet it's these business leaders who ultimately decide whether and how broadly to deploy AI models. Our goal for this article is to help managers do that better, using:

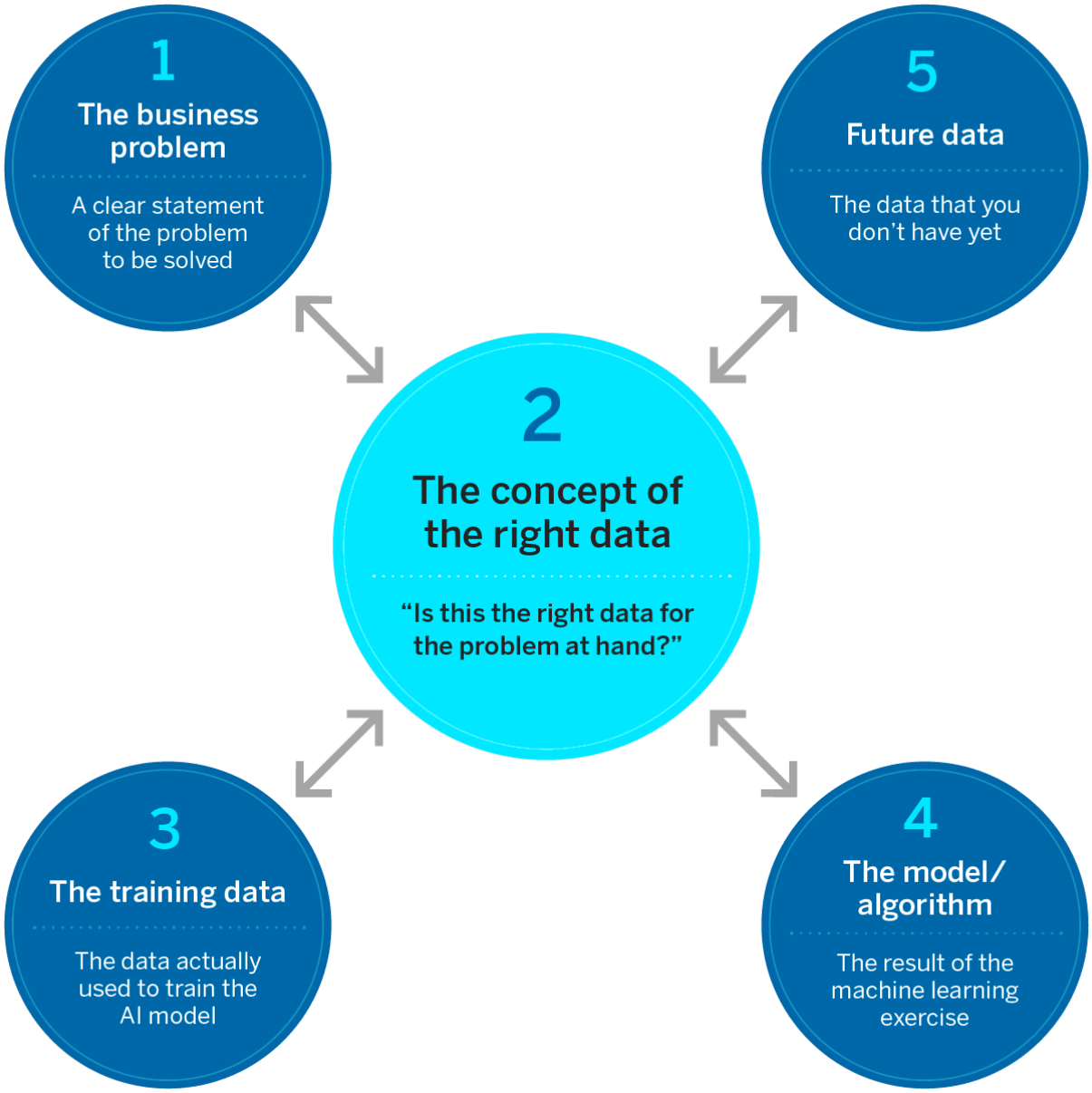
- A framework that delivers needed context. In particular, we'll introduce the concept of "the right data." Mismatches between the right data and the data actually employed in an AI project can be risky.
- A set of six questions to ask their organization's AI model developers before and during modeling work and deployment.
- Guidance on how to assess AI model developers' answers to those six questions.

How to Identify the Right Data: A Framework

An AI project's success or failure depends on the data set it uses. To help teams get to the right data, we offer a five-element framework depicted below. Let's break it down into its components.

The Right Data Framework

Is that AI project destined to succeed? This framework will help you have more informed discussions with AI model developers. After you've confirmed the business problem to be solved, identifying the right data is central, as shown here. Without the right data set, your AI project will fail. But you'll also need to evaluate the related pieces: the training data, the AI model, and the data that will be applied to the model in the future.



1. The problem and the population of interest. All good data science requires a clear statement of the problem to be solved. As obvious as this might seem, we find that AI model developers often haven't given this enough thought or that team members have different understandings of the problem they are trying to solve. For example, what is our ultimate business objective relative to a new automated credit-scoring model? Is it to save time? Is it to replace underwriters — or simply to advise them? Is the goal to make fewer lending errors, or reduce bias? The answer might involve some combination of those objectives. One more detail: Must the logic be explainable to people rejected for credit, or will black-box models

suffice? The answers to these business-problem questions will lead to different solutions.

2. The concept of the right data. A key contribution to the fundamentals of data quality has been the concept of fitness for use — whether a given set of data is appropriate, or fit, for a given decision, operation, or analysis. Depending on the problem, there can be many and varied aspects of fitness for use, but two — "Is the data right?" and "Is this the right data?" — are always important.

Here, we'll focus on the question "Is this the right data?" because it is critical to evaluating generalizability and preventing project failure.

To answer this question, managers must focus on six criteria (sometimes also called dimensions, features, properties, requirements, or aspects):

- Relevancy/completeness: The data should have predictive power. In our credit-scoring example, attributes such as age, late payment history, and income might contribute. Ideally, all such attributes are included (that is, the set of attributes is complete) and all misleading, extraneous, or illegal attributes have been excluded.
- Comprehensiveness/adequate representation: The two major issues are "Does the data adequately cover the population of interest?" and "Is there enough of it to adequately train the model?" Importantly, privacy or other concerns might dictate that certain data must be excluded.
- Freedom from bias: Many types of biases can be hidden in data, and this dimension demands their elimination. This is a special concern in our credit-scoring example and anytime the problem of interest involves human beings.
- Timeliness: The essential issue is "How new must the data be?" For some problems, older data might contain biases that are difficult to remove. And in some applications, (future) data is no longer relevant mere seconds after having been created.
- Clear definition: All terms, including units of measurement, should be clearly defined.
- Appropriate exclusions: In the discussions of relevancy and comprehensiveness above, we noted that some data should be excluded, given legal, regulatory, ethical, and intellectual property considerations. For example, using ZIP codes can be a surrogate for race in credit decisions, and organizations must avoid violating laws that stipulate how personally

identifiable information can be used. There is growing concern that AI models trained on public sources might violate intellectual property rights. Managers, or their companies' legal teams, should spell out the requirements as fully as possible.

3. The training data. This refers to the data actually used to train the model — regardless of whether it is, in fact, the right data or is some more easily obtained substitute.

4. The model/algorithm. This is the result of the machine learning exercise. Once trained, the model can be updated in the future using new data. That's commonly referred to as "future data."

5. Future data. This refers to data that you don't have yet but will apply to the AI model in the future.

Consideration of the right data, the training data, and the future data is the best defense against embarrassing AI project failures.

As depicted in the figure, the *concept of the right data* is central to everything. We call it the *concept* of the right data because it is more often about the criteria one *hopes* the data will satisfy rather than an actual data set. Model developers should first clarify the problem and population of interest. Next, they should define the criteria that the data used to train the model should meet to address that problem. Third, they should compare the training data they can actually obtain with these criteria. Then they should make similar comparisons with future data. Increasingly large gaps or mismatches signal trouble.

As we noted earlier, it's easy for developers to become enamored with the models they've built, and the easily obtained data and the data best suited for the business problem might be quite different. To head off trouble, business leaders must guide the team through a stone-cold sober consideration of the right data, the training data, and the future data. This is the best defense against overenthusiasm, outright model hubris, and embarrassing AI project failures.

Ask Six Penetrating Questions

With this background, we recommend that managers ask a sequence of pointed questions at three key stages, based on the Right Data Framework. Questioning should begin at the time you're defining the problem and continue through deployment. By the way, if you want assistance with the technical vocabulary of

data science, seek out someone who can translate for you in meetings with model developers. Some companies are even establishing specific roles and groups to bridge this gap between data science teams and business leaders.

Questions to Ask During Problem Definition

Developers should begin by asking the following two questions:

1. Assuming this project is successful, how and where do you anticipate the models that you develop will be utilized?

What to look for in the answers: This question is intended to determine the model developers' grasp of the real problem that the business is trying to solve; what's in and what's out of scope, with respect to the population of interest; and how long into the future the developers intend the model to be applied.

Furthermore, this question sets the stage for the next two. We advise managers to be extremely demanding with this query. Too many data science efforts doom themselves from the very start by failing to nail the problem statement.

2. How will you acquire training data that meets the right data criteria?

What to look for in the answers: This question might be the most critical. At this point, model developers are anticipating what data they can acquire. Make sure that the developers have sorted out the right data criteria (using the six considerations noted above, starting with relevancy). Next, examine whether the developers have a credible plan to obtain data that meets those criteria. If their answers come up short at this stage, send them back to the drawing board.

Questions to Ask During AI Model Development

Once developers have completed the problem definition stage, it's time to build the AI model. Focus on the following questions:

3. What steps have you taken to understand the full history, subtlety, advantages, and limitations of the training data acquired? How does it compare with the right data criteria?

What to look for in the answers: Here, you want to verify that the model developers *actually* obtained the data that they anticipated acquiring in Question 2. Press them to work criterion by criterion, listing gaps in the training data vis-à-

vis the right data, evaluating the severity of gaps, and explaining their plans to close important gaps.

Importantly, outside of textbooks, there is no such thing as a perfect data set, so expect gaps. Be very suspicious if model developers report that there are no gaps.

4. How will you check that future data meets the right data criteria?

What to look for in the answers: Probe model developers at this time, when they've just finished working with the training data, to make sure they've thought about the sources of future data. If not, ask them to give this careful consideration. (Question 5 follows up on this.)

Please note that validating a model by "holding back" some training data is *not* an acceptable solution. On Kaggle.com and other data science competition platforms, training data is assumed to be of acceptable, even pristine, quality. Model developers compete to build the best prediction based on a "holdout" set of data, taken from the same original data set as the training data. Therefore, in all the important ways, the holdout data looks just like the training data. In real-life situations, this is not what happens. For example, in Amazon's facial recognition system, the training data came from the local geographical area, while the algorithm was to be applied more broadly. This led to "poor calibration of the algorithm," in Amazon's own words.

Questions to Ask Before and During Deployment

Why ask these questions before *and* during deployment? You will want to ask them multiple times because the team will learn new things during deployment.

5. How will you ensure that future data meets your expectations? What controls for both data and models are in place to ensure both successful deployment and model accuracy with future data?

What to look for in the answers: This builds on Question 4's discussion of future data. Here, you're ensuring that developers have a system in place to evaluate future data before it's used by the model or used to update the model.

That's important because while companies hope for the best with AI model deployment, they are well advised to prepare for the worst. Managers should ensure that model developers have a control plan that will prevent, or at least provide early warning of, changes in future data or poor model performance. For

example, how will a gradual decrease in accuracy of the model over time be detected?

While companies hope for the best with AI model deployment, they are well advised to prepare for the worst.

Finally, probe developers' plans for updating their models as future data becomes available.

6. What are the top three ways you could envision your models failing in deployment? What steps have you taken to mitigate them?

What to look for in the answers: Engineers learned long ago that technical systems often fail despite their best efforts. They, therefore, developed failure mode and effects analysis, or FMEA, to help anticipate potential failures before they happen and put contingency plans in place to avoid or at least detect them.

Unfortunately, many data scientists have yet to embrace this method. Insist that model developers put in the equivalent work. Force them to think broadly over a range of potential failures related to technology, people, data quality, changes in the environment, and other issues.

Hard but Vital Conversations

We are well aware that many data scientists and AI model developers will not like answering these questions. But given the high failure rate of data science projects, asking, "How will you prevent yours from failing?" is simply good management.

What's more, as one Google research team noted, "Everyone wants to do the model work, not the data work." Business leaders don't have this luxury. Demanding an emphasis on the right data, not only for building models but also for validating and utilizing them in the future, is perhaps the single most important thing managers can do to increase the success rate for machine learning and AI projects.

ABOUT THE AUTHORS
Roger W. Hoerl is the Brate-Peschel Professor of Statistics at Union College in Schenectady, New York, and coauthor with Ronald D. Snee of *Leading Holistic Improvement With Lean Six Sigma 2.0*, 2nd ed. (Pearson FT Press, 2018). Thomas C. Redman is president of Data Quality Solutions and author of *People and Data: Uniting to Transform Your Organization* (KoganPage, 2023).