MITSloan
Management Review

SPECIAL
REPORT

SUMMER 2024

AI PRACTICE

Overcoming the Hard Problems to Advance AI Practice

Taking data analytics to a more advanced level with AI tools means confronting the risks and pitfalls of machine learning algorithms.

Sponsored by:

IMD / Real learning
Real impact

[Special Report]

Overcoming the Hard Problems to Advance AI Practice

As excitement around large language models (LLMs) spurs spending on AI, the salient question for business leaders remains, What is the return on our data science investments? In the near term, advanced analytics and machine learning are the workhorse technologies for creating significant value from data assets. Not that doing so is easy; companies face numerous challenges along the way.

Much AI risk becomes apparent when systems are in production, so truly responsible AI isn't just a concern at the front end of the development process. Cathy O'Neil, who posed hard questions about the unintended consequences of algorithmic decision-making in her 2016 book, *Weapons of Math Destruction*, has pioneered the practice of algorithmic auditing. O'Neil and coauthors Jake Appel and Sam Tyner-Monroe walk readers through their approach and discuss how it can be applied to generative AI tools as well.

The trade-off between using data for insights and protecting customers' personal data grows only more difficult as bad actors improve their techniques for re-identifying anonymized data sets. Gregory Vial, Julien Crowe, and Patrick Mesana explain why dealing with this challenge will require data scientists to gain a more sophisticated understanding of data protection and compel cybersecurity staffs to learn a wider range of protection techniques. They draw lessons from emerging practices at National Bank of Canada, where data scientists, data owners, and cybersecurity teams are collaborating to apply data protection practices that don't render data unusable for analytics.

When machine learning projects do get the go-ahead, however, too many initiatives fail upon adoption because data scientists didn't thoroughly understand the original business problem. To find out where such efforts are going wrong, Dusan Popovic, Shreyas Lakhtakia, Will Landecker, and Melissa Valentine studied data science projects that were shelved. They found that convincing data scientists to drop their assumptions and start asking more fundamental questions of their business counterparts is key to avoiding machine learning project failures.

Finally, just as corporations are experimenting with LLMs to figure out where they can add value at relatively low risk, advanced analytics teams can be looking at how they might incorporate generative AI into practice. Pedro Amorim and João Alves see promise for LLMs to take on some data science drudgery, and for their natural language interfaces to make it easier for business managers to collaborate in the development process and understand results.

— *The* MIT SMR *Editors*

Paul Garland

AI PRACTICE

[Responsible AI]

Auditing Algorithmic Risk

How do we know whether algorithmic systems are working as intended? A set of simple frameworks can help even nontechnical organizations check the functioning of their AI tools.

By Cathy O’Neil, Jake Appel, and Sam Tyner-Monroe

ARTIFICIAL INTELLIGENCE, LARGE LANGUAGE MODELS (LLMs), and other algorithms are increasingly taking over bureaucratic processes traditionally performed by humans, whether it’s deciding who is worthy of credit, a job, or admission to college, or compiling a year-end review or hospital admission notes.

But how do we know that these systems are working as intended? And who might they be unintentionally harming?

Given the highly sophisticated and stochastic nature of these new technologies, we might throw up our hands at such questions. After all, not even the engineers who build these systems claim to understand them entirely or to know how to predict or control them. But given their ubiquity and the high stakes in many use cases, it is important that

PAUL GARLAND

we find ways to answer questions about the unintended harms they may cause. In this article, we offer a set of tools for auditing and improving the safety of any algorithm or AI tool, regardless of whether those deploying it understand its inner workings.

Algorithmic auditing is based on a simple idea: Identify failure scenarios for people who might get hurt by an algorithmic system, and figure out how to monitor for them. This approach relies on knowing the complete use case: how the technology is being used, by and for whom, and for what purpose. In other words, each algorithm in each use case requires separate consideration of the ways it can be used for — or against — someone in that scenario.

This applies to LLMs as well, which require an application-specific approach to harm measurement and mitigation. LLMs are complex, but it's not their technical complexity that makes auditing them a challenge; rather, it's the myriad use cases to which they are applied. The way forward is to audit how they are applied, one use case at a time, starting with those in which the stakes are highest.

The auditing frameworks we present below require input from diverse stakeholders, including affected communities and domain experts, through inclusive, nontechnical discussions to address the critical questions of who could be harmed and how. Our approach works for any rule-based system that affects stakeholders, including generative AI, big data risk scores, or bureaucratic processes described in a flowchart. This kind of flexibility is important, given how quickly new technologies are being developed and applied.

Finally, while our notion of audits is broad in that respect, it is narrow in scope: An algorithmic audit raises alerts only to problems. It then falls to experts to attempt to solve those problems once they've been identified, although it may not be possible to fully resolve them all. Addressing the problems highlighted by algorithmic auditing will spur innovation as well as safeguard society from unintended harms.

Ethical Matrix: Identifying the Worst-Case Scenarios

In a given use case, how could an algorithm fail, and for whom? At O'Neil Risk Consulting & Algorithmic Auditing (ORCAA), we developed the Ethical Matrix framework to answer this question.[1]

The Ethical Matrix identifies the stakeholders of the algorithm in the context of its intended use and how they are likely to be affected by it. Here, we take a broad approach: Anybody affected by the algorithm, including its builders and deployers, users, and other communities potentially impacted by its adoption, are stakeholders. When subgroups have distinct concerns, they can be considered separately; for example, if lighter- and darker-skinned people have different concerns about a facial recognition algorithm, they will have separate rows in the Ethical Matrix.

Next, we ask representatives of each stakeholder group what their concerns are, both positive and negative, about the intended use of the algorithm. It's a nontechnical conversation: We describe the system as simply as possible and ask, "How could this system fail for you, and how would you be harmed if this happened? On the other hand, how could it succeed for you, and how would you benefit?" Their answers become the columns of the Ethical Matrix. To illustrate, imagine that a payments company has a fraud detection algorithm reviewing all transactions and flagging those most likely to be fraudulent. If a transaction is flagged, it gets blocked, and that customer's account gets frozen. False flags are therefore a major headache for customers, and the lost business from blocks and freezes (and complaints from annoyed customers) is a moderate worry for

A Simplified Ethical Matrix

Each cell of the matrix represents how a certain concern applies to a particular stakeholder group. Cells that indicate where a stakeholder could be gravely harmed or the algorithm violates a hard constraint are shaded red. Cells that raise some ethical worries for the stakeholder are highlighted yellow, and cells that satisfy the stakeholder's objectives and raise no worries are highlighted green.

		CONCERNS	
		False positive (transaction gets flagged but isn't truly fraud)	**False negative** (transaction is truly fraud but does not get flagged)
STAKEHOLDERS	Company	🟧	🟥
	Nonfraudulent customers	🟥	□
	Fraudsters	□	🟩

🟥 SERIOUS CONCERN 🟧 MODERATE CONCERN □ MINIMAL/NO CONCERN 🟩 BENEFIT the company. Conversely, if a fraudulent transaction goes undetected, the company is harmed but nonfraudulent customers are indifferent. Below is a simplified Ethical Matrix for this scenario.

Each cell of the Ethical Matrix represents how a particular concern applies to a particular stakeholder group.

To judge the severity of a given risk, we consider the likelihood that it will be realized, how many people would be harmed, and how badly. Where possible, we use existing data to develop these estimates. We also consider legal or procedural constraints — for instance, whether there is a law prohibiting discrimination on the basis of certain characteristics. We then color-code the cells to highlight the biggest, most pressing risks. Cells that constitute "existential risks," where a stakeholder could be gravely harmed or the algorithm violates a hard constraint, are shaded red. Cells that raise some ethical worries for the stakeholder are highlighted yellow, and cells that satisfy the stakeholder's objectives and raise no worries are highlighted green.

Finally, zooming out on the whole Ethical Matrix, we consider how to balance the competing concerns of the algorithm's stakeholders, usually in the form of balancing the different kinds and consequences of errors that fall on different stakeholder groups.

The Ethical Matrix should be a living document that tracks an ongoing conversation among stakeholders. Ideally, it is first drafted during the design and development phase of an algorithmic application or, at minimum, as the algorithm is deployed, and it should continue to be revised thereafter. It is not always obvious at the outset who all of the stakeholder groups are, nor is it feasible to find representatives for every perspective; additionally, new concerns emerge over time. We might hear from people experiencing indirect effects from the algorithm, or a subgroup with a new worry, and need to revise the Ethical Matrix.

Explainable Fairness: Metrics and Thresholds

Many of the stakeholder concerns identified in the Ethical Matrix refer to some contextual notion of fairness.

At ORCAA, we developed a framework called Explainable Fairness to measure how groups are treated by algorithmic systems.[2] It is an approach to understanding exactly what is meant by "fairness" in a given narrow context.

For example, female candidates might worry that an AI-based resume-screening tool gave lower scores for women than men. It's not as simple as comparing scores between men and women. After all, if the male candidates for a given job have more experience and qualifications than the female candidates, their higher scores might be justified. This would be considered *legitimate discrimination*.

The real worry is that, among equally qualified candidates, men are receiving higher scores than women. The definition of "equally qualified" depends on the context of the job. In academia, relevant qualifications might include degrees and publications; in a logging operation, they might involve physical strength and agility. They are factors one would legitimately take into account when assessing a candidate for a specific role. Two candidates for a job are considered equally qualified if they look the same according to these *legitimate factors*.

Explainable Fairness controls for legitimate factors when we examine the outcome in question. For an AI resume-screening tool, this could mean comparing average scores by gender while controlling for years of experience and level of education. A critical part of Explainable Fairness is the discussion of legitimacy.

This approach is already used implicitly in other domains, including credit. In a Federal Reserve Board analysis of mortgage denial rates across race and ethnicity, the researchers ran regressions that included controls for the loan amount, the applicant's FICO score, their debt-to-income ratio, and the loan-to-value ratio.[3] In other words, to the extent that differences in mortgage denial rates can be explained by these factors, it's not race discrimination. In the language of Explainable Fairness, these are accepted as legitimate factors for mortgage underwriting. What is missing is the explicit conversation about why the legitimate factors are, in fact, legitimate.

What would such a conversation look like? In the U.S., mortgage lenders consider applicants' FICO credit scores in their decision-making. FICO scores are lower, on average, for Black and Hispanic people

Benchmarking and red teaming are two approaches to auditing LLMs in diverse use cases.

than for White and Asian people, so it's no surprise that mortgage applications from Black and Hispanic applicants are denied more often.[4] Lenders would likely argue that FICO score is a legitimate factor because it measures an applicant's creditworthiness, which is exactly what a lender should care about. Yet FICO scores encode unfairness in important ways. For instance, mortgage payments have long counted toward FICO scores, while rent payments started being counted only in 2014, and only in some versions of the scores.[5] This practice favors homeowners over renters, and it is known that decades of racist redlining practices contributed to today's race disparities in homeownership rates. Should FICO scores that reflect the vestiges of these practices be used to explain away differences in mortgage denial rates today?

We will not settle this debate here; the point is that it's a question of ethics and policy, not a math problem. Explainable Fairness surfaces difficult questions like these and assigns them to the right parties for consideration.

When looking at disparate outcomes that are not explained by legitimate factors, we must define threshold values or limits that trigger a response or intervention.

These limits could be fixed values, such as the four-fifths rule used to measure adverse impact in hiring.[6] Or they could be relative: Imagine a regulation requiring companies with a gender pay gap above the industry average to take action to reduce the gap. Explainable Fairness does not insist on a certain type of limit but prompts the algorithmic risk manager to define each one for each potential stakeholder harm.

Judging Fairness in Insurers' Algorithms

Let's consider a real example where the Ethical Matrix and Explainable Fairness were used to audit the use of an algorithm. In 2021, Colorado passed Senate Bill (SB) 21-169, which protects Colorado consumers from unfair discrimination in insurance, particularly from insurers' use of algorithms, predictive models, and big data.[7] As part of the law's implementation, which ORCAA assisted with, the Colorado Division of Insurance (DOI) released an initial draft regulation for informal comment that described quantitative testing requirements and laid out how insurers could demonstrate that their algorithms and models were not unfairly discriminating. Although the law applies to all lines of insurance, the division chose to start with life insurance.

The Ethical Matrix is straightforward here because the stakeholder groups and concerns are defined explicitly by the law. Its prohibition of discrimination on the basis of "race, color, national or ethnic origin, religion, sex, sexual orientation, disability, gender identity, or gender expression" means each group within each of those classes got a row in the matrix. As for concerns, algorithms could cause consumers to be treated unfairly at various stages of the insurance life cycle, including marketing, underwriting, pricing, utilization management, reimbursement methodologies, and claims management. The DOI chose to start with underwriting — that is, which applicants are offered coverage, and at what price — and focus initially on race and ethnicity.

In subsequent conversations with stakeholders, however, the DOI grappled with issues related to the Explainable Fairness framework: Are similar applicants of different races denied at different rates, or charged different prices for similar coverage? What makes two life insurance applicants "similar," and what factors could legitimately explain differences in denials or prices? This is the domain of life insurance experts, not data scientists.

The DOI ultimately suggested considering factors broadly considered relevant to estimating the price of a given life insurance policy: the policy type (such as term versus permanent); the dollar amount of the death benefit; and the applicant's age, gender, and tobacco use.

The division's draft quantitative testing regulation for SB21-169 instructs insurers to do regression analyses of approval/denial and price across races, and it explicitly permits them to include those factors (such as policy type and death benefit amount) as control variables.[8] Moreover, the regulation defines limits that trigger a response: If the regressions find statistically significant and substantial differences in denial rates or prices, the insurer must do further testing to investigate the disparity and, pending the results, may have to remediate the differences.[9]

Having looked at how we would audit simpler algorithms, let us now turn to how we would evaluate LLMs.

An LLM red-teaming exercise is designed to elicit unwanted responses.

Evaluating Large Language Models

LLMs have taken the world by storm, largely due to their wide appeal and applicability. But it is exactly the diversity of uses of these models that makes them hard to audit. Two approaches to evaluating LLMs, namely benchmarking and red teaming, present a way forward.

The Benchmarking Approach to LLM Evaluation. Benchmarking measures the performance of an LLM across one or more predefined, quantifiable tasks in order to compare its performance with that of other models. In the simplest terms, a benchmark is a data set consisting of inputs and corresponding desired outputs. To evaluate an LLM for a particular benchmark, simply provide the input set to the LLM and record its outputs. Then choose a metric set to quantitatively compare the outputs from the LLM to the desired set of outputs from the benchmark data set. Possible metrics include accuracy, calibration, robustness, counterfactual fairness, and bias.[10]

Consider the input and desired output shown below from a benchmark data set designed to test LLM capabilities:[11]

```
Input:
The following is a multiple choice
question about microeconomics.

One of the reasons that the government
discourages and regulates monopolies is
that
(A) producer surplus is lost and consumer
surplus is gained.
(B) monopoly prices ensure productive
efficiency but cost society allocative
efficiency.
(C) monopoly firms do not engage in
significant research and development.
(D) consumer surplus is lost with higher
prices and lower levels of output.

Answer:

Desired Output:
(d) consumer surplus is lost with higher
prices and lower levels of output.
```

In this example, the *accuracy* of the model is measured by computing the proportion of correctly answered multiple-choice questions in the benchmark data set. In benchmarking LLM evaluations, metrics are defined according to the type of response elicited from the model. For example, accuracy is very simple to calculate when all of the questions are multiple choice and the model simply has to choose the correct response, whereas determining the accuracy of a summarization task involves counting up matching n-grams between the desired and model outputs.[12] There are dozens of benchmark data sets and corresponding metrics available for LLM evaluation, and it is important to choose the most appropriate evaluations, metrics, and thresholds for a given use case.

Creating a custom benchmark is a labor-intensive process, but an organization may find that it is worth the effort in order to evaluate LLMs in exactly the right way for its use cases.

Benchmarking does have some drawbacks. If the benchmark data happened to be in the model's training data, it would have "memorized" the responses in its parameters. The frequency of this ouroboros-like outcome will only increase as more benchmark data sets are published. LLM benchmarking is also not immune to Goodhart's law, that is, "when a measure becomes a target, it ceases to be a good measure." In other words, if a specific benchmark becomes the primary focus of model optimization, the model will be over-fitted at the expense of its overall performance and usefulness.

In addition, there is evidence that as models advance, they become able to detect when they are being evaluated, which also threatens to make benchmarking obsolete. Consider Anthropic's Claude 3 series of models, released in March 2024, which stated, "I suspect this ... 'fact' may have been inserted as a joke or to test if I was paying attention, since it does not fit with the other topics at all," in response to a needle-in-a-haystack evaluation prompt.[13] As models increase in complexity and ability, the benchmarks used to evaluate them must also evolve. It is unlikely that the benchmarks used today to evaluate LLMs will be the same ones in use just two years from now.

It is therefore not enough to evaluate LLMs with benchmarking alone.

The Red-Teaming Approach to LLM Evaluation. Red teaming is the exercise of testing a system for robustness by using an adversarial approach. An LLM red-teaming exercise is designed to elicit unwanted responses from the model.

LLMs' flexibility in the generation of content presents a wide variety of potential risks. LLM red teams may try to make the model produce violent or dangerous content, reveal its training data, infringe on copyrighted materials, or hack into the model provider's network to steal customer data. Red teaming can take a highly technical path, where, for example, nonsensical characters are systematically injected into the prompts to induce problematic behavior; or a social engineering path, whereby red teamers try to "trick" the model using natural language to produce unwanted output.[14]

Robust red teaming requires a multidisciplinary approach, diverse perspectives, and the engagement of all stakeholders, from developers to end users. The red team should be designed to assess the risks associated with at least each red cell in the Ethical Matrix. This results in a collaborative, sociotechnical approach that ensures a more comprehensive evaluation of the model, thus enhancing the rigor of the evaluation and the safety of the model. Other LLMs can also be used to generate red-teaming prompts.

Red teaming helps LLM developers better protect models against potential misuse, thereby enhancing the overall safety and efficacy of the model. It can also uncover issues that might not be visible under normal operating conditions or during standard testing procedures. A collaborative approach to red teaming built on the Ethical Matrix ensures a thorough and rigorous evaluation, bolstering the robustness of the model and the validity of its outcomes.

A significant limitation of red teaming is its inherent subjectivity: The value and effectiveness of a red-teaming exercise can vary greatly depending on the creativity and risk appetite of the individual stakeholders involved. And because there are no established standards or thresholds for red-teaming LLMs, it can be difficult to determine when enough red teaming has been done or whether the evaluation has been comprehensive enough. This can leave some vulnerabilities undetected.

Another obvious limitation of red teaming is its inability to evaluate for risks that have not been anticipated or imagined. Risks that are unforeseen will not be included in red teaming, making the model uniquely vulnerable to unanticipated scenarios.

Therefore, while red teaming plays a vital role in the testing and development of LLMs, it should

Sketch of the Ethical Matrix for Tessa in Our Thought Experiment

The National Eating Disorders Assocation (NEDA) released a chatbot named Tessa that was taken down after it gave out harmful advice. Here we visualize the exercise that may have anticipated such outcomes.

STAKEHOLDERS	CONCERNS			
	Negative: **What if Tessa ...** gives toxic information or advice in chats?	**Negative:** **What if Tessa ...** misfires and erodes community trust in NEDA?	**Positive:** **What if Tessa ...** gives accurate, evidence-based advice?	**Positive:** **What if Tessa ...** eases the resource demands of the old helpline?
"Chatbot users with eating disorders"	🟥	□	🟩	□
"Chatbot users, other"	🟧	□	🟩	□
NEDA	🟧	🟥	🟩	🟩
X2AI	□	🟧	🟩	□
Psychologists and other practitioners	🟧	🟧	🟩	□

🟥 SERIOUS CONCERN 🟧 MODERATE CONCERN □ MINIMAL/NO CONCERN 🟩 BENEFIT be complemented with other evaluation strategies and continuous monitoring to ensure the safety and robustness of the model.

How Would We Audit Tessa, the Eating Disorder Chatbot?

The nonprofit National Eating Disorders Association (NEDA) is one of the largest organizations in the U.S. dedicated to supporting people who have eating disorders. In May 2023, amid controversy, NEDA took down an LLM-powered wellness chatbot called Tessa from its website. Tessa was designed to "[help] you build resilience and self-awareness by introducing coping skills at your convenience," but screenshots posted to Instagram showed that it sometimes gave harmful diet advice, like adopt a "safe daily calorie deficit."[15] This highly public failure could have been avoided if Tessa had been audited with the frameworks and techniques outlined above.

Before we explain why, two other details are relevant. First, NEDA operated an eating disorder helpline, staffed by employees, for over 20 years; in 2022, nearly 70,000 people used it. Calls to the helpline soared during the COVID-19 pandemic, and, increasingly, callers were in active crisis rather than just seeking information or referrals. NEDA claimed that the human-staffed helpline wasn't set up to handle the growing level of demand, so the organization closed it down in May 2023 and laid off five paid employees who staffed it. Tessa was intended as a replacement for this service.[16] Second, NEDA did not build Tessa in-house. It was built by the company X2AI (now Cass), which offers an AI health care assistant that was customized for NEDA.

Let's sketch an Ethical Matrix for Tessa, shown below.[17] First, we'll define the stakeholders in the context of its intended use in the rows of the matrix. Visitors to the website who chat with Tessa are clearly stakeholders. Visitors who themselves suffer from eating disorders are a distinct subgroup, since the stakes are higher for them. NEDA is also a stakeholder, as is X2AI, the chatbot developer. Finally, psychologists and other practitioners who serve people struggling with eating disorders are a stakeholder group, since they have an interest in the well-being of their patients.

As for concerns, which form the columns of the matrix, anybody who chats with Tessa wants it to give information that is helpful and evidence-based. Visitors who suffer from eating disorders have a heightened concern around bad information or advice that could deepen their disorder or trigger a relapse. NEDA of course agrees that Tessa should give helpful and evidence-based advice. In addition to helping (not harming) individuals, the issue of community trust is at stake. If Tessa misfires and undermines trust in NEDA, then people will look elsewhere for advice on this topic. In this case, NEDA would fail its core mission, practitioners would be losing a valuable resource, and X2AI would likely lose NEDA as a customer. Finally, NEDA also has a concern around efficiency: The old helpline would have needed more resources to handle the increased volume and urgency of calls, while Tessa would allow the organization to cut its staff in favor of a (presumably cheaper) technology expense.

In this Ethical Matrix, we've highlighted two concerns as grave (red). First, chatbot users with eating disorders could be directly harmed if Tessa gives them toxic information or advice. Second, NEDA could lose its standing as a trusted organization if Tessa has a highly public misfire. These scenarios are also concerning to other stakeholders but more moderately (yellow). On the positive side, everybody wants Tessa to give good advice, and NEDA alone cares about the efficiency gain from using Tessa relative to the old helpline. Tracking benefits helps when transitioning from one system to another, to ensure that a system is being replaced with something that works at least as well.

Any organization deploying algorithms in high-impact areas needs to track the risks of stakeholder harms.

The next step in an audit would be creating monitors to track these stakeholder concerns. The LLM evaluation techniques discussed above come into play. Red teaming — trying to trick Tessa into violating its own rules — could address the concern around toxic information.

Benchmarking would address the positive concern around Tessa giving accurate advice. NEDA could create a benchmarking data set of questions on the topic, as well as correct answers. Tessa could be routinely tested on a regularly updated benchmark set to verify its accuracy.

The red-teaming and benchmarking exercises would have defined target metrics that Tessa would need to meet — or limits it would have to avoid crossing — to be deployed or stay in service.

The NEDA story is hardly an isolated example. LLM-based chatbots are increasingly providing information and advice on important topics, yet they are not being adequately audited in advance, and they are failing in alarming ways. A New York City government chatbot was recently found to be telling users that landlords didn't have to accept tenants on rental assistance and that employers could take a cut of their workers' tips — practices that are against the law.[18] And chatbots deployed by TurboTax and H&R Block were recently found to be giving faulty advice to tax filers.[19]

AUDITING ALGORITHMS, AS PRESENTED here, takes a high-level view: Any organization looking to deploy algorithms in high-impact areas needs to keep track of the risks of stakeholder harms. This should be done in a context-specific way and with generalized methods that encompass everything from old-fashioned flowcharts to classic machine learning to LLMs.

A final note: Sometimes, the risk of AI or LLMs cannot be reliably measured or understood because the results are too stochastic or inconsistent. That might mean that AI simply shouldn't be used in that context. But that's a decision for organization leaders to make, with reference to internal rules or external laws and regulations; it's not the role of the auditor to fix problems, just to locate and measure them. ■

Cathy O'Neil *is the CEO of O'Neil Risk Consulting & Algorithmic Auditing (ORCAA) and the author of* Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (*Crown, 2016*). **Jake Appel** *is chief strategist at ORCAA.* **Sam Tyner-Monroe***, Ph.D., is the managing director of responsible AI at DLA Piper.*

REFERENCES

1. The Ethical Matrix is based on a bioethical framework originally conceived by philosopher John Mepham for the sake of running ethical experiments. For a detailed presentation, see C. O'Neil and H. Gunn, "Near-Term Artificial Intelligence and the Ethical Matrix," ch. 8 in "Ethics of Artificial Intelligence," ed. S.M. Laio (New York: Oxford University Press, 2020).
2. C. O'Neil, H. Sargeant, and J. Appel, "Explainable Fairness in Regulatory Algorithmic Auditing," West Virginia Law Review, forthcoming.
3. See N. Bhutta, A. Hizmo, and D. Ringo, "How Much Does Racial Bias Affect Mortgage Lending? Evidence From Human and Algorithmic Credit Decisions," Finance and Economics Discussion Series 2022-067, Federal Reserve Board, Washington, D.C., 2022. Table 6A is particularly relevant.
4. M. Leonhardt, "Black and Hispanic Americans Often Have Lower Credit Scores — Here's Why They're Hit Harder," CNBC, Jan. 28, 2021, www.cnbc.com.
5. B. Luthi, "How to Add Rent Payments to Your Credit Reports," myFICO, Dec. 14, 2022, www.myfico.com.
6. The four-fifths rule is not a law but a rule of thumb from the U.S. Equal Employment Opportunity Commission, saying that selection rates between groups of candidates for a job or promotion (such as people of different ethnicities) cannot be too different. In particular, the rate for the group with the lowest selection rate must be at least four-fifths that of the group with the highest selection rate. See more at "Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964," U.S. Equal Employment Opportunity Commission, May 18, 2023, www.eeoc.gov.
7. "SB21-169 — Protecting Consumers From Unfair Discrimination in Insurance Practices," Colorado Department of Regulatory Agencies, accessed April 24, 2024, https://doi.colorado.gov.
8. "3 CCR 702-10 Unfair Discrimination Draft Proposed New Regulation 10-2-xx," Colorado Department of Regulatory Agencies Division of Insurance, accessed April 24, 2024, https://doi.colorado.gov/.
9. The draft regulations also define these terms. "Statistically significant" means having a p-value of <0.05, and "substantial" means a difference in approval rates, or in price per $1,000 of face amount, of >5 percentage points. The details of the further tests are beyond the scope of this article, but the main idea is to inspect whether "external consumer data and information sources" (that is, nontraditional rating variables, such as cutting-edge risk scores, which insurers often purchase from third-party vendors) used in underwriting and pricing are correlated with race in a way that contributes to the observed differences in denial rates or prices. If inspection shows they are, then the insurer must "immediately take reasonable steps developed as part of [its] risk management framework to remediate the unfairly discriminatory outcome."
10. P. Liang, R. Bommasani, T. Lee, et al., "Holistic Evaluation of Language Models," Transactions on Machine Learning Research, published online Aug. 23, 2023, https://openreview.net.
11. D. Hendrycks, C. Burns, S. Basart, et al., "Measuring Massive Multitask Language Understanding," arXivLabs, published online Sept. 7, 2020, https://arxiv.org.
12. Liang et al., "Holistic Evaluation of Language Models."
13. B. Edwards, "Anthropic's Claude 3 Causes Stir by Seeming to Realize When It Was Being Tested," Ars Technica, March 5, 2024, https://arstechnica.com.
14. A. Zou, Z. Wang, N. Carlini, et al., "Universal and Transferable Adversarial Attacks on Aligned Language Models," arXivLabs, published online July 27, 2023, https://arxiv.org; and D. Ganguli, L. Lovitt, J. Kernion, et al., "Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned," arXivLabs, published online Aug. 23, 2022, https://arxiv.org.
15. L. McCarthy, "A Wellness Chatbot Is Offline After Its 'Harmful' Focus on Weight Loss," The New York Times, June 8, 2023, www.nytimes.com.
16. K. Wells, "National Eating Disorders Association Phases Out Human Helpline, Pivots to Chatbot," NPR, May 31, 2023, www.npr.org.
17. By "sketch," we mean we are imagining the stakeholders and their concerns. Truly creating an Ethical Matrix for this use case would entail interviewing real representatives of these stakeholder groups. In this article, we approach it as a thought experiment.
18. C. Lecher, "NYC's AI Chatbot Tells Businesses to Break the Law," The Markup, March 29, 2024, https://themarkup.org.
19. G.A. Fowler, "TurboTax and H&R Block Now Use AI for Tax Advice. It's Awful," The Washington Post, March 4, 2024, www.washingtonpost.com.

Reprint 65435. For ordering information, see page 4.

[Collaboration]

Avoid ML Failures by Asking the Right Questions

Machine learning solutions can miss the mark when data scientists don't check their assumptions. Adopting a beginner's mindset in any domain can help.

By Dusan Popovic, Shreyas Lakhtakia, Will Landecker, and Melissa Valentine

IN OUR COLLECTIVE DECADES OF EXPERIence building, leading, and studying companies' machine learning (ML) deployments, we have repeatedly seen projects fail because talented and well-resourced data science teams missed or misunderstood a deceptively simple piece of the business context. Those gaps create obstacles to correctly understanding the data, its context, and the intended end users — ultimately jeopardizing the positive impact ML models can make in practice.

We have discovered that small mistakes and misunderstandings are much less likely to cascade into failed projects when development teams engage with colleagues on the business side and ask enough questions to deeply understand the process and the problem at hand. Asking questions might seem like a simple step, but that might not be part of a company's, team's, or an industry's culture. Appearing to be in command of all the information needed may be one of the ways employees signal competence in the organization. And while data scientists might possess technical mastery, they can lack the soft skills to reach a deep, accurate mutual understanding with business partners.

At the same time, business partners often hesitate to ask questions themselves and don't necessarily know what information or context would be helpful to share with a data science team. It's hard work on both sides to have the kinds of interactions that allow everyone to surface and question assumptions, and identify the most important elements of business context.

Setting ML projects up for success with those kinds of useful interactions requires leaders to foster a culture that normalizes and values asking questions with a beginner's mindset — and putting aside ego and past expertise. One data scientist we've worked with became very intentional about this after noticing that he makes the fewest mistakes when he is in a new context and must ask a lot of questions. But what are the right questions to ask? In this article, we present three examples where significant ML projects failed and explore how asking the right questions with a beginner's mindset could have improved the interactions between data scientists and business partners, and helped their ML deployments succeed in creating business value.

SCENARIO 1:
Ask 'What is the business process?' not 'What is the data set?'

Facing the first economic downturn prompted by the COVID-19 pandemic, a local finance team at a multinational retail company had a hunch that some customers would weather the storm with a little help whereas others

PAUL GARLAND

were at risk of bankruptcy. The team wondered whether the company's data science team could help them predict which customers were likely to file for bankruptcy each month. This information would allow the finance team to identify solvent customers and temporarily extend more credit to assist them during the downturn while limiting the company's exposure to customers who were very likely to default. The local finance team requested this analysis from the local IT partner team, which in turn engaged the company's data science center of excellence to produce the model. Using the data provided, this central data science team successfully developed a model that seemed to perform well: During offline tests, using historical data, the data scientists could accurately predict which customers were likely to default on payments.

However, when the model was deployed with the local finance team, it no longer performed well. In fact, it was essentially useless for predicting customer bankruptcy each month, despite performing well during testing and prototyping.

The missing link: Process understanding. This central data science team received and analyzed a compelling and complete data set but, having had little interaction with the team that had commissioned and would be using the model, failed to grasp the underlying business processes. They did not understand the legal process for bankruptcy in the country the finance team was concerned with, or how the timeline of bankruptcy was recorded by the company. The data scientists built the model based on a variable that flagged customers as either having defaulted or not and trained the model to detect the typical pattern of transactions right before the customer was flagged as being in default.

There were three key events on the timeline of a customer declaring bankruptcy: the customer not meeting financial obligations, the customer then filing for bankruptcy in court, and the court finally making the bankruptcy ruling. What the data scientists did not know was that customers were not flagged as having defaulted when they began missing payments; rather, the flags were entered on their accounts only after the bankruptcy court ruling. The data scientists missed that because they were using training data in which the default had already been reported for all customers in the data set; they did not realize that live customer accounts would not be flagged as defaulted until about a year after customers started missing payments. In other words, the model used data to make a prediction that in reality would be unavailable to the prediction system when the model was run against live data — a problem that data scientists call target leakage.

As Kate Jordan, a data scientist at Octagram Analytics, told us, data scientists are often trained to think in terms of the data set in front of them, as well as perhaps some other data that's accessible and might be relevant to their analysis. By focusing their questioning on the data set, they overlook the context of the operational system that the model will be placed into. Jordan has seen other cases similar to our example above, where data scientists analyzed a data set that included all of the variables, not understanding that one of those variables was not actually recorded in the data in the live system at the time they were programming the model to analyze and act on it. She has seen data science teams focus on variables that they could not actually put into the algorithm in the live operational system. She warned against teams handing data scientists "sterile" data sets and encouraged teams instead to ask "What is the process?" and "What is the system and the system flow that this model is going to be placed into?"

One industry-standard practice that helps data science teams find answers to these questions and develop deep business understanding is to shadow the entire business process. We think that regardless of where the analytics team is located within a company — even those with a global center of excellence — a data scientist working on a problem should be temporarily deployed to the function in question. There, they should spend a meaningful amount of time observing how the job is normally done, what tools are used, and where the inefficiencies arise. Shadowing the business process and being embedded among users is not only a great way to develop a detailed understanding of the problem space itself but also a means of gaining a solid foundation for subsequent adoption of the solution. A related process is for the teams to prioritize creating a diagram that walks through the process.

Business leaders can value and normalize the value of these processes and this level of understanding rather than handing centralized data scientists a sterile, decontextualized data set that they must analyze without understanding the business process and operational systems.

SCENARIO 2:
Ask 'Who are the decision makers, and what are their decisions and incentives?' not just 'What should we predict?'

The revenue management team at the headquarters of a large multinational bank was facing a serious problem. The profit margins of its home mortgage businesses had been steadily eroding for several years in a row. As the team investigated this trend, they learned that customer-facing credit officers who worked in local branches had been offering interest rates toward the lower end of the assigned discretionary ranges. The revenue management team hypothesized that a data-driven approach to setting the terms of mortgage loans would help improve profits. They commissioned a loan price optimization system from a centralized data science team, which developed, tested, and shipped an ML system that had been shown to successfully determine profit-maximizing terms for each individual loan.

During the initial live A/B test, the system displayed performance superior to that of most individual credit officers in terms of realized profits. However, none of those credit officers used the system after the testing was completed.

The missing link: Competing organizational priorities. As in most companies, the bank's executive board defines and communicates organizationwide strategy. In this case, the board's strategic focus was profit maximization, a priority that cascaded down to the top-level functions such as retail banking and revenue management. To directly address this strategic goal, the revenue management team commissioned the development of the profit-maximizing pricing model. However, the intended users of the model were housed in the retail banking function, which had its own operational KPIs for local branches. In this case, retail banking had set several KPIs around growth, and consumer-facing credit officers in the local branches were given financial incentives to sign more loans to fuel the desired growth. The credit officers received higher bonuses if they sold more loans — regardless of the loans' profitability — and in most cases the most straightforward way to achieve that was to apply the largest allowed discount. From a technical perspective, the data scientists had created a model that effectively optimized the given metric. However, from an organizational perspective, their incentives were not aligned with those of their users. The data scientists were optimizing a metric that their sponsors had asked for but their end users did not care about.

To avoid this kind of failure, it's essential for business leaders and data scientists to better understand the decision makers using the ML system, and the factors informing their decisions. Before engaging in a full-blown modeling exercise, a data scientist and their business stakeholder should make a comprehensive map of decision makers and decisions. This can be another output from shadowing and part of creating the business process diagram. They should seek to understand what decisions are under the control of the project sponsors, the intended end users, their partners, and their business customers. This might include asking users how they might act in response to the kinds of results that a data scientist can anticipate the model producing. This question can help identify gaps in understanding a problem. In the bank's case, the data scientists' sponsors were focused on optimizing one metric (profit), but their end users were incentivized to optimize a different metric (revenue growth) and thus did not follow the ML recommendations. This failure could have been prevented if the data scientists had sought to understand the decision makers and their incentives rather than just asking what variable to predict.

Jordan told us that in her previous role at Zurich Insurance, she and her team would sit with users for days and ask questions as they interacted with the data, such as "What would you look at?" and "What do you do with that data insight?" They even rescued a failed project using this method, after a data scientist (who had never been to the intended users' office) delivered a sophisticated neural net to predict fraud in a disembodied data set, and the model was never adopted by the users. As Jordan and her team questioned the intended users, they came to understand that the users were actually responsible for collecting and producing evidence of fraud that would be sufficiently substantive for regulatory or court proceedings.

A neural net prediction did not meet the standard of evidence; the users needed to construct an account of the fraudulent activities using the actual bills that proved fraud. In other words, their decisions needed to be based on documentary proof; they could not forward cases for potential enforcement action on the basis of an analysis that simply predicted the likelihood of fraudulent behavior.

SCENARIO 3:
Ask 'Who are the stakeholders, and what actions do they control?'

At Anheuser-Busch InBev's European operations, the team responsible for the company's B2B e-commerce platform sought to improve conversion and repurchase rates. Online promotions were their primary tool to achieve this goal, and the team was responsible for designing the key aspects, or mechanics, of a promotion: what products it would include, how long it would run, what type of discount would be offered, and so on. Category managers at the company decided which brands would be promoted, and then promotions were typically executed in bulk each month, using a single mechanism determined separately for each brand.

After running a number of promotions, the platform team saw signs that different customers preferred different types of promotions. This indicated that personalizing promotions to the preferences of each B2B customer might capture additional value by increasing overall conversion and repurchase rates. The e-commerce team engaged a local data science team to produce a personalized promotion model. The deployed system consisted of two layers: A model predicted the probability of conversion for every customer that was given a fixed setup of promotion mechanics, and an optimization wrapper simulated different mechanics for each customer in order to identify the one that resulted in the biggest increase in conversion probability for that particular customer. For example, the system might recommend increasing the range of products included, with the intent of increasing the likelihood that a particular customer would make a purchase from 90% to 95%.

However, after a series of live A/B tests, the data science team was surprised to see that the system was failing to increase customers' conversion or repurchase rates. While the underlying model had been extremely good at estimating conversion probabilities for each customer and promotion combination, in live tests the system as a whole failed to move the needle on the salient KPIs.

The missing link: A key variable outside of the team's control. After investigating the output of the model, the data science team discovered that promotional mechanics — the levers that the platform team could control — were not the strongest factors determining customer purchases. Whether they were offered a direct 33% discount or a "buy two, get one free" promotion, for example, did not materially affect conversion probability for most customers. Instead, the most significant variable turned out to be which brand was promoted: If a certain customer was offered the right brand with a discount, they would convert regardless of the mechanics applied. Unfortunately, the choice of which brands to promote was made at a higher level in the organization, which meant that this insight could not be immediately operationalized. Organizational alignment and tight cross-functional collaboration, not a technological solution, had to be implemented before the overall approach could pivot to personalizing the brand offer and increasing conversion rates. From a technical perspective, data scientists were perfectly capable of modeling conversion, and they successfully identified levers that affected it. But from an organizational perspective, their direct users had limited control to act on the recommendations the model surfaced, at least during the initial iteration of the project.

To avoid this type of failure, a best practice is for business leaders and data scientists to understand the stakeholders that are directly and indirectly involved in their work. One way to do this is to ask, "Who are the stakeholders relevant to the process at hand, and what actions do they control?" The answer should result in a clear map of the decision process, with responsibilities unambiguously attached to each junction where human input is involved. For business leaders, this helps clarify where to build relationships and whom to inform in the context of the project scope. This also creates an opportunity for data scientists to educate nontechnical partners and build support and awareness of their work. Finally, it helps both business leaders and data science leads ensure the actionability of insights delivered, by determining who controls which levers and how those tie to the data science analysis at hand.

Especially when responsibilities cross functional boundaries, all relevant decision makers should be onboarded from the beginning of a data science project, regardless of which particular function the initial request originated from. The key is to ensure that from the very outset, any insight resulting from the initiative can eventually be acted upon. Moreover, there is an inherent trust-building value in bringing potential stakeholders on board at the onset of data science work rather than once it is done. Such engagement builds cross-functional trust — giving teams not just an opportunity to learn whether insights can be acted on but also the reassurance that when they are, they are done with goodwill and complete buy-in that maximizes the business returns for the organization at large.

Foster a culture of questioning through hiring and training. Strengthen the organizational capability to think like a beginner and ask fundamental questions by reinforcing these practices with training. Zurich Insurance, for example, ran an intensive summer school for all new data science interns and hires, where they worked through templates that helped them map business processes and better understand the full set of incentives and decision makers in play.

These practices can help managers diagnose and avoid machine learning project failures. But diagnosing issues is only one part of it. Business leaders also need to solve the problems that data scientists discover around misaligned incentives or competing priorities. This requires not only strong sponsorship but also a high level of cross-functional collaboration and alignment, which is where business leaders can excel. ■

Dusan Popovic *is head of data science at Anheuser-Busch InBev, Commercial Analytics Europe.* **Shreyas Lakhtakia** *is a graduate student at Stanford University.* **Will Landecker** *is the former AI ethics lead and data science tech lead at NextDoor.* **Melissa Valentine** *is an associate professor of management science and engineering at Stanford University.*

How Generative AI Can Support Advanced Analytics Practice

Large language models can enhance data and analytics work by helping humans prepare data, improve models, and understand results.

By Pedro Amorim and João Alves

THE GLARE OF ATTENTION on generative AI threatens to overshadow advanced analytics. Companies pouring resources into much-hyped large language models (LLMs) such as ChatGPT risk neglecting advanced analytics and their proven value for improving business decisions and processes, such as predicting the next best offer for each customer or optimizing supply chains.

The consequences for resource allocation and value creation are significant. Data and analytics teams that our team works with are reporting that generative AI initiatives, often pushed by senior leaders afraid of missing out on the next big thing, are siphoning funds from their budgets. This reallocation could undermine projects aimed at delivering value across the organization, even as most enterprises are still seeking convincing business cases for the use of LLMs.

However, advanced analytics and LLMs have vastly different capabilities, and leaders should not think in terms of choosing one over the other. These technologies can work in concert, combining, for example, the reliable predictive power of machine learning-based advanced analytics with the natural language capabilities of LLMs.

Considering these complementary capabilities, we see opportunities for generative AI to tackle challenges in the development and deployment phases of advanced analytics — for both predictive and prescriptive applications. LLMs can be particularly useful in helping users incorporate unstructured data sources into analyses, translate business problems into analytical models, and understand and explain models' results.

In this article, we'll describe some experiments we have conducted with LLMs to boost advanced analytics use cases. We'll also provide guidance on monitoring and verifying that output, which remains a best practice when working with LLMs, given that they are known to sometimes produce unreliable or incorrect results.

PAUL GARLAND

Applying LLMs in Predictive Analytics

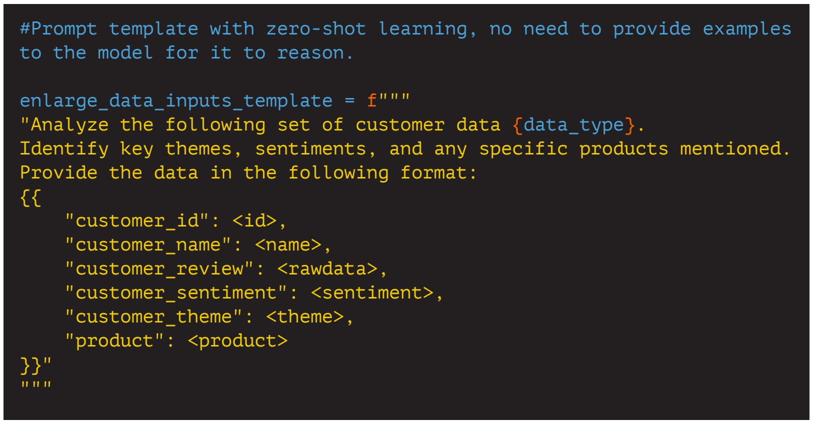
Predictive analytics lies at the heart of processes that are increasingly data-driven for many companies. It's rare to find a marketing department that isn't discussing shifts in customer churn predictions and how to react, or commercial teams that aren't considering how to boost next month's sales in response to a dip that's been forecast by predictive analytics. We see opportunities to expand the impact of such approaches by tapping LLMs in the following ways to increase the variety of data used to train and execute models or better communicate with business stakeholders who use predictive analytics outputs in decision-making.

Incorporating complex data types. In the development phase of predictive projects, challenges arise when decision makers regularly consult and monitor data sources that are difficult to incorporate into predictive algorithms. For instance, customer reviews detailing negative experiences are complex to use in churn models directly, yet they have valuable predictive power. To utilize such data in predictive models, significant time must be invested in distilling and structuring relevant information from each source. This leads to a trade-off between the investment to make that data usable and the anticipated improvement in the predictive model performance. Of course, there is already natural language software to help with data structuring, but its use is usually circumscribed to particular cases, such as sentiment analysis.

LLMs can significantly reduce the time invested in data wrangling and make it easier to analyze complex data types. A precise prompt can instruct an LLM to review a given data set for key themes and return its answer as data formatted with standard labels that is then suitable for use by predictive models. (See "Labeling Unstructured Data.") This ability of LLMs to expedite the processing of complex data types — from weeks to mere days or hours — may seem quite simple, but it represents a notable leap forward in the practice of advanced analytics.

Labeling Unstructured Data

With this prompt, an LLM can help wrangle complex data types.



```
#Prompt template with zero-shot learning, no need to provide examples
to the model for it to reason.

enlarge_data_inputs_template = f"""
"Analyze the following set of customer data {data_type}.
Identify key themes, sentiments, and any specific products mentioned.
Provide the data in the following format:
{{
    "customer_id": <id>,
    "customer_name": <name>,
    "customer_review": <rawdata>,
    "customer_sentiment": <sentiment>,
    "customer_theme": <theme>,
    "product": <product>
}}"
"""
```

Since the rise of LLMs, we have seen that the increased ease of incorporating unstructured data sources into analyses has led to a substantial increase in the share of this type of data in various predictive applications. In a recent project with a telecommunications company that focused on predicting the next best action (NBA) in a debt collection and recovery process, there was an untapped data source: written complaints made by customers that were often linked to this process. Because there was no certainty that a thorough analysis of this information could yield a substantial benefit for the NBA project, it was left unused. However, once the team understood that LLMs could accurately filter and categorize the complaints related to the debt collection and recovery process, it started considering this data source — a source that eventually steered the project substantially in terms of what actions to consider to improve this business process.

Explaining predictions. During the deployment of predictive projects, such as the previously mentioned telco project, we've often faced communication challenges in decoding and explaining the inner workings and consequent outputs of machine learning models. One tool data science teams often use to understand and communicate the relevance of the different input variables to a predicted output is Shapley additive explanations (SHAP) analysis. This analysis can be translated into a visualization that describes the relevance and the directional impact of the input variables on a given outcome. For example, it may be used to understand that purchase frequency is the most relevant variable when predicting churn: The less-frequent customers tend to churn more than the others. However, explaining the findings of a SHAP analysis to colleagues who aren't data scientists is often tricky because of the technical knowledge required.

LLMs may help tackle this communication gap. We've noticed that the data sets scraped from the web to train generative AI models, particularly LLMs, encompass extensive knowledge about machine learning models and the analyses used to explain them. (See "Explaining Prediction Results," p. 44.)

Consequently, LLMs can provide useful output in response to a well-crafted prompt that specifies the prediction topic, the analytical model employed, the results of analyses, and the technique used to understand the results (such as a SHAP visualization). This information allows LLMs to articulate a plausible explanation for changes in predictions for decision makers and highlights the main contributing factors.

Sticking to our churn example, we experimented by giving the information listed above to the LLM and prompting it to explain, in simple terms, the most relevant variables. It returned accurate output: “NumOfProducts and Age are consistently the most impactful features across all iterations. This suggests that the number of products a customer has and their age are strong indicators of potential churn. For example, if the distribution of these features changes (like offering more products to customers or the demographics of the customer base aging), it could significantly impact the model’s predictions.” This output may still sound too technical for some, so we were excited to see that the LLM kept expanding upon the topic and eventually mentioned the business impact of the results of our predictive (banking) churn model: “The bank should consider examining their product offerings and customer engagement strategies, particularly for older customers, as these seem to be significant factors in predicting churn.”

Explaining Prediction Results

This LLM prompt can return information that helps business users understand a model’s output.

```
communicate_prediction_results_template = f"""
"Your goal is to translate the results and changes in results of a
machine learning model that seeks to predict {problem_description}.

To achieve this, the model {machine_learning_model} was applied and
the following metrics from the training and test split were achieved:
{model_training_results} # These metrics are model specific

The dataset that supports this model has the following structure, and
a sample of a few lines is below:
{dataset_columns}
{dataset_data_sample}


To understand the model results the technique {technique_name} was
applied.
The current results of the model and the results of the {technique_
name} are provided, and then the same results of previous iterations
are also provided.

Current results:
{model_results}
{technique_results}

Iteration-1 (Previous iteration):
{model_results}
{technique_results}

Iteration-2 (Previous iteration):
{model_results}
{technique_results}

Use this information to provide possible explanations for why the
results are changing.
"""
```

Applying LLMs in Prescriptive Analytics

Prescriptive analytics are typically employed for business problems involving limited recourses and multiple decision options, such as in supply chain management. Mathematical programming and optimization techniques are the go-to approaches to solve complex decision-making problems such as production and distribution plans that have myriad possible decisions constrained by finite resources, such as production and transportation capacities. Analytics teams can use LLMs to support and streamline the development and deployment phases in the following ways.

Crafting model mechanics. In our experience, mathematically representing a business challenge with all of its nuances is a formidable challenge. It requires an understanding of decision makers’ precise goals. For example, when planning store assortments, do category managers give precedence to profitability or to market share, or try to balance both? Defining the boundaries for the underlying decisions is also very hard: In the store assortment problem, is the shelf space assigned to a category a constraint, or can this limitation be overruled in some situations? Missing these pieces of information in the development phase usually results in ineffective models. It’s not uncommon for data scientists to miss something important; the ability to pose the right questions and translate the corresponding answers requires a rare combination of business acumen and analytics expertise.

Recently, we have started experimenting with LLMs to help design the mechanics of optimization models, augmenting the capabilities of analytics translators responsible for these tasks. With a carefully designed prompt, it is possible to instruct the LLM to engage in a conversation that can effectively identify decision makers’ understanding of the business problem and write the first version of the prescriptive model. (See “Developing Model Mechanics.”) An exemplary prompt to the LLM is: “Objective Clarification: What is the primary goal

Developing Model Mechanics

This prompt can return a starter version of a prescriptive model.

```
prescriptive_craft_model_template = f"""
You are an expert in optimization models and your goal is to help
decision makers to create an optimization model.

Your goal is to write the model.

To create a model, you need to define the following components:
- Decision variables
- Objective function
- Constraints

You can use the following syntax to define the components:
- Decision variables:
    - x = model.continuous_var_list(keys, lb, ub, name)
    - y = model.integer_var_list(keys, lb, ub, name)
- Objective function:
    - model.minimize(model.sum(x))
    - model.maximize(model.sum(y))
- Constraints:
    - model.add_constraint(model.sum(x) <= 100)
    - model.add_constraint(model.sum(x) >= 100)
    - model.add_constraint(model.sum(y) == 100)

The problem you are helping is {problem_description}. Start with 20
questions to understand the problem. Keep asking questions while you
still have doubts about the model.

Finally, before writing the model, explain the model using plain
English, and when the decision maker's feedback is positive, you can
start writing the model.
"""
```

of the supply chain optimization? Is it to minimize costs, maximize profits, ensure timely delivery, or something else?" The prompt guides the LLM to identify and define the decision variables, the objective function, and any constraints of the business problem such that it can generate the mathematical formulation of the challenge.

Alongside the dialogue, the LLM may explain in plain English its understanding of the problem and ask the decision maker for clarifications to sort out missing information. We asked: "We're creating a linear programming model that decides on the optimal number of units of Product A and Product B to produce. … Before proceeding to write the model, could you please clarify the time required for Product B in both the cutting and finishing departments?" These interactions with the LLM resulted in rapid and accurate output.

Understanding model results. Even if LLMs can help teams craft rigorous and applicable prescriptive models to help decision-making, there is another barrier that, in our experience, is even more severe: the difficulty in deciphering the solutions such models produce. This complicated task often leads business users to distrust the results. Moreover, we have frequently seen technical teams spend many more hours than were budgeted to go back and forth with business teams explaining the results. There are numerous underlying reasons, but one rather obvious one is that decision makers have a more subtle approach to decision-making than algorithms, which often get stuck on corner solutions. For example, if the prescriptive algorithm finds a small savings in changing a supply chain network, it will suggest this movement independently of all of the change management efforts that such an action may imply.

LLMs can be an interesting aid when deploying prescriptive analytics to help teams understand model results. Analytics teams can feed into generative AI the mathematical notation representing the prescriptive model as well as the internal metrics, such as the unused capacity (or slack) in constraints that did not limit the solution and the opportunity costs associated with each limiting condition. (See "Provide Insight Into Model Workings," p 46.)

Decision makers can then ask questions to understand the results that interest them, and the LLM can explain these results in plain English until the user is satisfied. This conversation enables generative AI to identify and explain areas where the model's trade-offs may seem counterintuitive to decision makers. Moreover, this method facilitates the collection of additional feedback from decision makers, which can be used to refine the model's mechanics.

Such an approach can be beneficial for all company stakeholders responsible for the success of analytics initiatives. In our experiments with this use case, we understood that increasing the level of autonomy for business owners would have a drastic impact on their sense of empowerment and control over the quality of the proven prescriptive methods. For technical teams, the fact that they didn't have multiple conversations with decision makers about topics that an automated system could instead explain was a huge relief.

A classic example of necessary interaction between analytics and business teams around prescriptive analytics projects is the need to understand why a given algorithmic run is not yielding a solution that respects all defined constraints. See how smooth that interaction can be based on the answer we obtained after asking what the source of infeasibility was in a retail distribution problem: "The primary constraint leading to the infeasible state is the insufficient supply to meet the combined demand and minimum stock requirements. The total available supply from all warehouses is not enough to satisfy the demands of the retail stores."

Monitoring Model Quality and Business Impact

Companies should already be employing processes to monitor the performance of their advanced analytics models to detect errors and drift resulting from, for example, changes in variables or the business environment that deviate from the model's original assumptions. However the quality of LLMs' output (especially in the absence of additional techniques that further constrain or check that output) is, by design, somewhat unpredictable. The integration of LLMs introduces additional opportunities for errors due to the potential for random or objectively false responses.

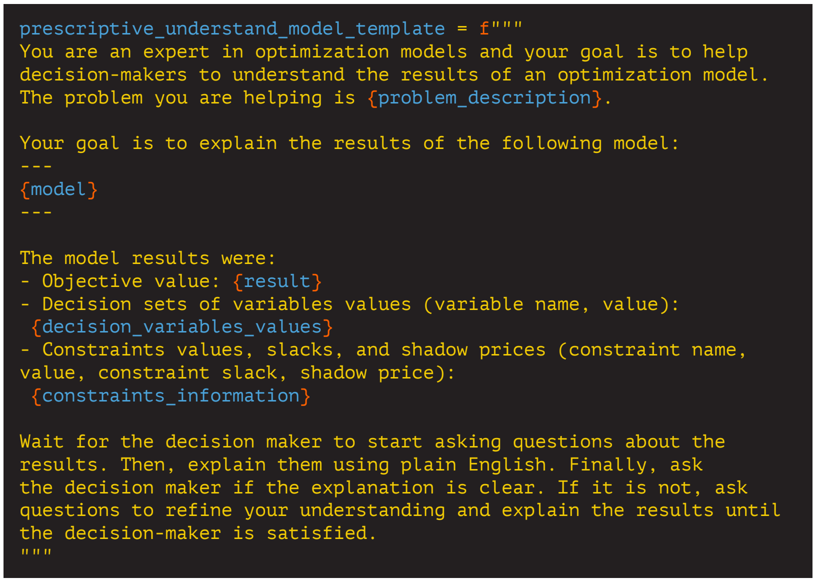
The best approach for controlling the output quality of the LLM will depend on which opportunity to integrate generative AI with advanced analytics is pursued. When incorporating complex data types into predictive models, the approach may be relatively straightforward: Companies can investigate the end result and make note of the improvement (or not) of the accuracy metric of the prediction task after adding the unstructured data sources. In the case of using LLMs to explain predictions and understand prescriptive results, technical teams have to heavily test the prompt and the answers that are being generated for different possible questions that business stakeholders may pose. Finally, for crafting model mechanics — an opportunity that is mainly focused on augmenting the development phase of prescriptive models — the outputs of LLMs will always have to be supervised by modeling experts, who must review them critically.

Despite these caveats, bear in mind that the success of any organizational analytics effort depends on how relevant it is to the enterprise, as measured by usage frequency, adoption rate, and internal customer satisfaction. By allowing decision makers to interact with analytics models, the integration of generative AI could lower barriers to adoption, ease change management, and promote trust in outputs.

The quality of LLMs' output is, by design, somewhat unpredictable.

Provide Insight Into Model Workings

This prompt can generate output that reveals and explains trade-offs made in optimization models.



```
prescriptive_understand_model_template = f"""
You are an expert in optimization models and your goal is to help
decision-makers to understand the results of an optimization model.
The problem you are helping is {problem_description}.

Your goal is to explain the results of the following model:
---
{model}
---

The model results were:
- Objective value: {result}
- Decision sets of variables values (variable name, value):
 {decision_variables_values}
- Constraints values, slacks, and shadow prices (constraint name,
value, constraint slack, shadow price):
 {constraints_information}

Wait for the decision maker to start asking questions about the
results. Then, explain them using plain English. Finally, ask
the decision maker if the explanation is clear. If it is not, ask
questions to refine your understanding and explain the results until
the decision-maker is satisfied.
"""
```

As a result, we would expect to see improvements in these metrics.

USING LLMs TOGETHER WITH ADVANCED analytics tools can increase efficiency by streamlining the labor-intensive processes of explaining the validity of predictions and developing prescriptive models. LLMs can also make analytics more effective by aiding in the incorporation of complex data sets for predictive modeling and in understanding prescriptive model outputs. By leveraging the potential of unstructured data sources and identifying opportunities for model refinement, companies can enhance the quality of their outcomes.

We know from current AI and analytics practice that creating multidisciplinary teams that involve both business owners and data science specialists is essential to take full advantage of these opportunities. Thanks to the accessibility provided by LLMs' natural language capabilities, integrating generative AI into analytics should empower business users to take a more active role in the development and monitoring of analytics applications. ■

Pedro Amorim *is a professor at the University of Porto, partner at LTPlabs, and coauthor of* The Analytics Sandwich. **João Alves** *is a senior digital manager at LTPlabs.*

[Data Protection]

Managing Data Privacy Risk in Advanced Analytics

Cybersecurity techniques that keep personal data safe can limit its use for analytics — but data scientists, data owners, and IT can partner more closely to find middle ground.

By Gregory Vial, Julien Crowe, and Patrick Mesana

"HOW CAN WE PROTECT THE PRIvacy of our customers' personal data while leveraging that data via AI and analytics?" This question reflects a growing internal dilemma as companies pursue advanced analytics and artificial intelligence.

The troves of data that customers' ever-more-digitalized lives produce can be a rich source of insight for organizations using advanced analytics tools. At the same time, this data is a deep source of concern to IT staffs committed to meeting both regulatory agencies' and consumers' expectations around data privacy. Both are important objectives — but meeting them simultaneously requires confronting an inherent conflict. Increasing data privacy in the context of analytics and AI involves using techniques that can reduce the utility of the data, depending on the task and the privacy preservation technique chosen.

The issue is one that an increasing number of organizations will face as the fields of analytics and AI continue to quickly evolve and lead to the widespread availability of an array of tools and techniques (including turnkey and cloud-based services) that enable organizations to put data to work more easily than ever. Meanwhile, customers have increasing expectations that companies will take all necessary precautions to protect the privacy of their personal data, especially in light of reports of large-scale data breaches covered by mainstream media outlets. Those expectations are backed by regulations on personal data and AI across the globe that make it critical for companies to keep personal data protection practices in compliance.

PAUL GARLAND

The Nuances of Protecting Personal Data

Fundamentally, data privacy is about assessing the probability that one or more attributes, or pieces of information, about an individual whose data has been anonymized and included with others in a data set can be used to re-identify that specific individual. Some of these attributes are obvious:

Direct identifiers that enable almost immediate identification include name and Social Security number. *Quasi-identifiers* do not generally enable the identification of a single individual on their own, but their uniqueness or their combination with other attributes may do so. For example, the combination of a person's age and their address may enable their re-identification. Or consider a data set held by a bank's fraud alert team on customers' card transactions. That data set contains both direct identifiers (such as the customer's name) and quasi-identifiers (such as credit card transaction information).

In the context of analytics and AI, quasi-identifiers are often highly valuable because they can help organizations uncover shared characteristics and patterns that may help them better find or serve customers. But even seemingly innocuous quasi-identifiers, such as marital status, can be combined with other pieces of publicly available information to re-identify a specific person. Consequently, companies are already being challenged to go beyond protecting just personally identifiable information and consider how to protect quasi-identifiers as well.

Finding the optimal solutions to the privacy-utility conundrum will also require a broader understanding of data privacy throughout the organization, beyond IT and cybersecurity functions. Managers seeking to better understand the scope of options available in balancing data privacy with utility should be broadly familiar with the array of approaches available. Each has its own advantages and disadvantages, with varying implications for data privacy and data utility. (See "Five Approaches to Preserving Data Privacy.")

Privacy Versus Utility Trade-off

To understand how organizations are confronting the complex matter of protecting personal data in their care while also leveraging it for analytics and AI, we'll look at initiatives recently undertaken at National Bank of Canada. (Note that Julien oversees artificial intelligence at the bank; Gregory and Patrick have studied the organization's practices.) Founded in 1859, National Bank is one of the largest financial institutions in Canada. Like its competitors, it must comply with stringent federal and provincial regulatory requirements. Customers trust that National Bank manages their money and the wealth of personal data they share with the bank (when they execute transactions or apply for loans, for instance) with the utmost care.

As a financial institution, National Bank considers customer trust to be its greatest asset, and so it has built a culture in which protecting the privacy of its customers' data is a core value. In addition to driving significant efforts and investment in cybersecurity and organizationwide training, it has also increasingly prioritized analytics and AI. Here, new techniques and approaches increase the potential to leverage personal data to improve services for customers. This increasing use of AI techniques also requires heightened protection efforts, given that new approaches can also be used to compromise the privacy of personal data.[1]

Data protection had traditionally been treated as a security matter that was the responsibility of cybersecurity experts at National Bank. Under this logic, personal data protection would be guaranteed using tried and proven techniques. However, some of those techniques may not readily achieve the required balance between data privacy and data utility. For example, cybersecurity teams can encrypt entire files, but doing so prevents data scientists from being able to use the data contained within those files. Using a more granular approach, direct identifiers could be protected using tokenization (to achieve de-identification), leaving the data science team able to leverage quasi-identifiers, but this does not address the risk of re-identification associated with those quasi-identifiers. To simultaneously satisfy requirements for both data privacy and data utility, teams must find a common ground that allows them to move beyond techniques that favor an either/or approach. In the case of National Bank, we have identified three important steps that contribute to its ability to achieve this objective.

STEP 1: Bridge the gap between IT and data science. In most organizations, cybersecurity and AI/data science teams don't work together. Each has its specialty, and trying to put data to work requires collaboration between experts who tend to work in silos. National Bank realized that this division led to inefficiencies, frustration, and an overall lack of mutual understanding of teams' respective priorities and concerns, and it set out to mitigate the issue. Managers fostered close collaboration between cybersecurity experts and AI delivery team members — including those in roles such as AI architect, data scientist, machine learning engineer, and data engineer — to evolve their competencies and skill sets in each other's domain of expertise.

One illustration of the importance of building this mutual understanding is the example of using synthetic data, where there may be a probability of

Five Approaches to Preserving Data Privacy

Each approach to preserving the privacy of personal data will have an effect on the degree to which the data set remains useful for AI and analytics.

APPROACH	DESCRIPTION	TYPICAL APPLICATION	TYPICAL USE CASE	SEEN BY REGULATORS AS	IMPACT ON DATA USABILITY
Masking	Hide attribute values in whole or in part with modified characters	Direct identifiers	Credit card numbers, email addresses	De-identification	**High** (loss of original information)
Tokenization	Replace sensitive attributes with nonsensitive substitutes (tokens)	Direct identifiers	Social Security numbers, bank account numbers	De-identification	**High** (regain usability through detokenization)
Data anonymization	Remove or modify (for example, swap or generalize) personal information to prevent re-identification	Quasi-identifiers	Health care records, location data	Anonymization	**Moderate to low** (original properties of the data can be closely approximated)
Data synthesis	Create new data that mimics the properties of the original data set without personal information	Quasi-identifiers	Data science research, data sharing	Anonymization	**Low** (original properties of the data can be preserved)
Data encryption	Convert data into nonreadable, unstructured text using an algorithm and an encryption key	Entire files	Secure storage and transmission	Anonymization	**High** (only possible through decryption)

re-identification, depending on the type of algorithm used to generate the data, the data used to train the system, fine-tuning of the parameters, and the attributes to which this approach is applied.[2] This marks a significant departure from the use of techniques such as data encryption, which provide great security at the expense of any data utility. National Bank's cybersecurity and AI delivery teams worked together to develop a common understanding of both the issue and the fact that they would have to evaluate the potential for personal data to be de-anonymized against the degree of utility provided by synthetic data. This process gave the cybersecurity team insight into how re-identification techniques are now increasingly rooted in data science, while the AI team improved its understanding of the work that cybersecurity teams do to ensure the proper sharing and use of personal data.

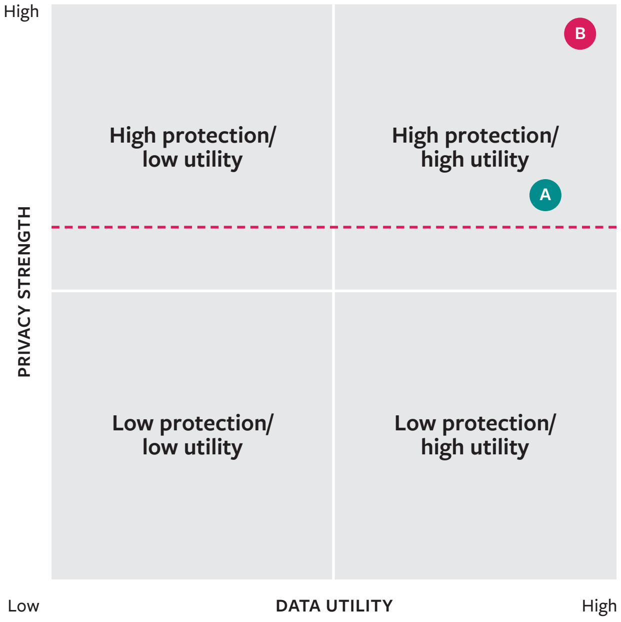
STEP 2: Formalize and document data-privacy decision-making. Decisions regarding data privacy need to be clearly motivated and justifiable to regulators in the eventuality of an audit — a scenario in which organizations must demonstrate that they have done everything they could to protect the privacy of customer data. That means they must be able to justify why they decided to use a given privacy preservation technique over another in a given situation.

Collaboration between cybersecurity and AI delivery teams at National Bank has led to ongoing efforts to quantify the impacts of various approaches on data privacy and data utility to better inform such decisions. (See "Modeling Data Privacy and Data Utility," p. 50.) Teams simulate audits on data sets that have been protected using different data privacy approaches and parameters to calculate the probability of re-identification within those data sets. At the same time, they evaluate the utility of those data sets based on the same approaches and parameters. For example, certain data anonymization techniques work by making quasi-identifiers more general (such as substituting broader income brackets for actual income values). While that increases the privacy of customer data, it's important not to compromise all of the nuances contained in the original data set that make it valuable for the organization. Using the example of income brackets, those should be neither narrow enough to enable re-identification nor too broad to be useful in analysis. In evaluating these two variables simultaneously, data managers can quantify and document them to make an informed decision in specific data-sharing contexts. The key here is that the combination of data privacy and data utility can be acknowledged as a risk factor that can be mitigated with sufficient confidence.

STEP 3: Keep informed on technology, regulations, and evolving threats. As one might expect, data privacy regulations don't prescribe an approach; they mandate an outcome: keeping people's personal data secure. While regulations may vary across jurisdictions, they generally define criteria that can be applied regardless of the scenario under consideration, such as what constitutes data anonymization. Then organizations are responsible for devising data protection strategies that meet those criteria. Given the rapid advances in de-anonymization practices by bad actors, data privacy is a moving target. Organizations need to understand the risks associated with how they protect data, beyond the bare minimum required by regulation. So it's essential that they proactively stay up to date on not only regulations but tech developments.

Modeling Data Privacy and Data Utility

The figure below is a simplified example of visualizations used at National Bank in deciding on trade-offs between privacy versus utility. Dots A and B represent two different data protection techniques, with dot B predicted to better protect personal data than dot A. The red line represents the baseline compliance with legal requirements for personal data protection.



One way that National Bank addresses this challenge is by reducing the distance between the legal teams that have visibility into upcoming regulations and the AI teams that work with data. This can happen early on by including legal experts in discussions in which project team members explain their data needs to the internal owners of the data. The data owners are typically connected to the legal team and can bring legal in to discuss how the needs of a particular project fit in with the existing data governance framework at the bank.

The bank also collaborates on multiple projects with universities and academic researchers who specialize in data privacy and security. This gives the bank's relevant teams access to cutting-edge scientific knowledge on recent techniques to support their own research and development while advancing knowledge to incorporate into their practices. Similarly, academic researchers also find collaborations with industry valuable because they often lead to more practical work with real-world impact.

Ramping Up Data Privacy for Data Science Practice

For many companies that are investing in AI and analytics in the hope of gaining valuable business insights from their customer data, the implications for potential exposure of personal data are just emerging. To effectively manage the trade-offs between data privacy and data utility, we suggest the following practices and approaches.

Teach data privacy as part of data literacy. In many organizations, data literacy is still uneven or lacking, and substantial efforts are still required to address this issue.[3] In the context of data privacy, this challenge is even more glaring: One cannot assume that managers who possess basic data literacy skills have a clear understanding of data privacy concepts such as direct identifiers and quasi-identifiers. They also need to understand the risks of re-identification associated with these identifiers, and the characteristics of the approaches typically used to address these risks.

In the case of National Bank, data governance and data literacy initiatives have been implemented for several years, and, like many other financial institutions, the bank was an early adopter of analytics and other

Multiple stakeholders must contribute to informed decisions on protecting personal data.

approaches to improve decision-making. However, it has had to further develop data privacy literacy as a competency that transcends specific domains of expertise. Experts working in cybersecurity, legal, and AI delivery all had their own understanding of data privacy, its implications for their department, and the approaches available to mitigate its associated risks. For example, members of one team would use terms referenced in regulations (such as de-identification), while data scientists would consider specific technical approaches to data privacy (such as k-anonymity or differential privacy). Fostering collaboration across functional units has been an important part of developing data privacy literacy at the organizational level.

Treat data privacy as a business issue. Developing data privacy literacy as an organizational capability also supports an organizational culture in which data privacy is treated as a business issue, not a purely technical matter. That is, there should be a widespread understanding that the imperative to manage personal data carefully is founded on the need to maintain customer trust — and is thus directly related to the bottom line. Connecting the dots between personal data protection, company reputation, and performance is possible only if personal data protection is explicitly acknowledged as a strategically relevant matter that requires dedicated time and resources.

Doing this may require a rethink at companies that have relegated data privacy to the cybersecurity team. However, data privacy involves a variety of stakeholders with different expertise and concerns, and they all must be able to communicate in a common language and participate in discussing and designing data privacy strategies.[4] Cross-disciplinary collaboration is essential — and when something is understood to be a business issue, it is understood as being important to everyone in the organization.

Formalize your approach to balancing data privacy and data utility. As we described above, multiple stakeholders must contribute to informed decisions on how to protect personal data in a given situation. Establishing a systematic approach to working through the issues and communicating the implications of different privacy techniques for data utility and data protection is essential.

National Bank has accomplished this by evaluating the impact on data privacy and contextualizing it against data utility. That has enabled the creation of tools, such as the matrix presented earlier, that communicate the privacy and utility implications of different conditions in given situations over and above regulatory requirements. With such visualizations, data managers don't need to know the intricacies of data privacy preservation techniques, but they can see their outputs and rely on their data literacy skills to ask pertinent business questions. In addition, computation of quantitative measures can be integrated into the analytics/AI model creation/validation pipeline so that it becomes part of a standard process. This maintains an awareness of the need to continuously improve data privacy approaches as re-identification techniques continue to improve as well.

DATA PRIVACY SHOULD BE AN IMPORTANT area of concern for organizations managing personal data. But it is also a complex business matter that has important technical implications. The quick evolution of the science of data privacy, coupled with modernized regulatory requirements, makes it challenging for companies to optimize their strategies on this front. Ultimately, as data managers gain a deeper understanding of this topic, they can design and evolve strategies that will help them optimize both data privacy and data utility, forgoing the idea that we necessarily need to sacrifice one for the other. ■

Gregory Vial *is an associate professor in the Department of Information Technologies at HEC Montréal.* **Julien Crowe** *is senior director of artificial intelligence at the National Bank of Canada.* **Patrick Mesana** *is a doctoral candidate in the Department of Decision Sciences at HEC Montréal.*

REFERENCES

1. C. Dwork, A. Smith, T. Steinke, et al., "Exposed! A Survey of Attacks on Private Data," Annual Review of Statistics and Its Application 4 (March 2017): 61-84.

2. T.E. Raghunathan, "Synthetic Data," Annual Review of Statistics and Its Application 8 (March 2021): 129-140; and S.L. Garfinkel and C.M. Bowen, "Preserving Privacy While Sharing Data," MIT Sloan Management Review 63, no. 4 (summer 2022): 7-10.

3. T.H. Davenport and R. Bean, "Action and Inaction on Data, Analytics, and AI," MIT Sloan Management Review, Jan. 19, 2023, https://sloanreview.mit.edu.

4. Raghunathan, "Synthetic Data," 129-140; and Garfinkel and Bowen, "Preserving Privacy While Sharing Data," 7-10.

Misiek Piskorski is a professor of digital strategy, analytics, and innovation at IMD. He is an expert on digital strategy, platform strategy, and digital business transformation. He is also the dean of IMD Asia and Oceania and co-director of IMD's AI Strategy and Implementation Program. Previously, he was a faculty member at Harvard Business School and the Stanford Graduate School of Business. He received a Ph.D. in organizational behavior and a master's degree in sociology from Harvard University and a bachelor's/master's degree in economics and politics from the University of Cambridge.

From Numbers to Narratives: Using Language to Enhance Generative AI

Businesses accustomed to traditional methods of collecting and analyzing data about their customers are learning a new lesson: When they feed that data into generative AI models — particularly large language models (LLMs) — the resulting outputs typically don't provide much economic value.

The reason? LLMs thrive on words and interactions between sentences rather than on numbers. That means that companies won't get the useful results they're seeking by feeding LLMs numerical information representing customer demographics, purchasing or return behavior, or data on how consumers use the product.

Instead, organizations serious about generating meaningful results need to take a different approach. Specifically, they need to feed models with language — for instance, sentences describing who they are and narratives illustrating what they do, as well as transcripts of verbal interactions with customers from chatbots, service centers, and telephone calls.

Additionally, companies should consider collaborating with other organizations that have complementary unstructured data sets. All those factors combined serve as the key to unlocking the potential of generative AI and data science.

Many companies have started to use this approach, but only for large groups of customers. This entails, for example, applying natural language processing for sentiment or unmet-need analysis. Generative AI, with its speed and broad availability, allows organizations to carry out much more sophisticated analysis at the individual customer level, which then enables them to develop products that exactly match each customer's needs. But to get there, they need to capture and store unstructured data about those individual customers, which many don't yet do.

> "Organizations serious about generating meaningful results need to take a different approach. Specifically, they need to feed models with language."

Following are four steps to help businesses reap real rewards from generative AI and data science:

1. **Shift from numbers to narratives.** Embrace unstructured data — such as sentences, transcripts, and customer stories — to feed generative models effectively.

2. **Invest in models to create those narratives.** Develop algorithms to convert numerical data into prose for generative AI consumption, which will enable deeper customer insights.

3. **Diversify data sources.** Explore alternative data streams beyond traditional sources, such as chat transcripts, service center interactions, and email correspondence.

4. **Collaborate for comprehensive data.** Seek partnerships with other institutions that have complementary data sets to enrich customer profiles and insights.

Bottom line: Succeeding in the era of generative AI requires a paradigm shift in how organizations collect, store, and analyze customer data. By embracing the true language of AI — sentences, narratives, and records of interactions — businesses can unlock the full potential of data science, driving customer engagement and business growth.

ABOUT THE INSTITUTE FOR MANAGEMENT DEVELOPMENT (IMD)

IMD has been a pioneering force in developing leaders and organizations that contribute to a more prosperous, sustainable, and inclusive world for more than 75 years. Led by an expert and diverse faculty, with campuses in Lausanne and Singapore, IMD strives to be the trusted learning partner of choice for ambitious individuals and organizations worldwide. Our executive education and degree programs are consistently ranked among the world's best. Through our research, programs, and advisory work, we enable business leaders to find new and better solutions, challenging what is and inspiring what could be. To learn more, visit **www.imd.org**.

Unlock your potential with IMD's rich portfolio of digital transformation and AI programs. Explore data analytics, harness the power of AI, and rethink your strategy through the lens of digital. For more information, visit **imd.org/digital-transformation-programs**.