The 10 Principles of Applied Artificial Intelligence

HOW TO IMPLEMENT AI IN YOUR SOFTWARE SOLUTION

Georgian

Introduction

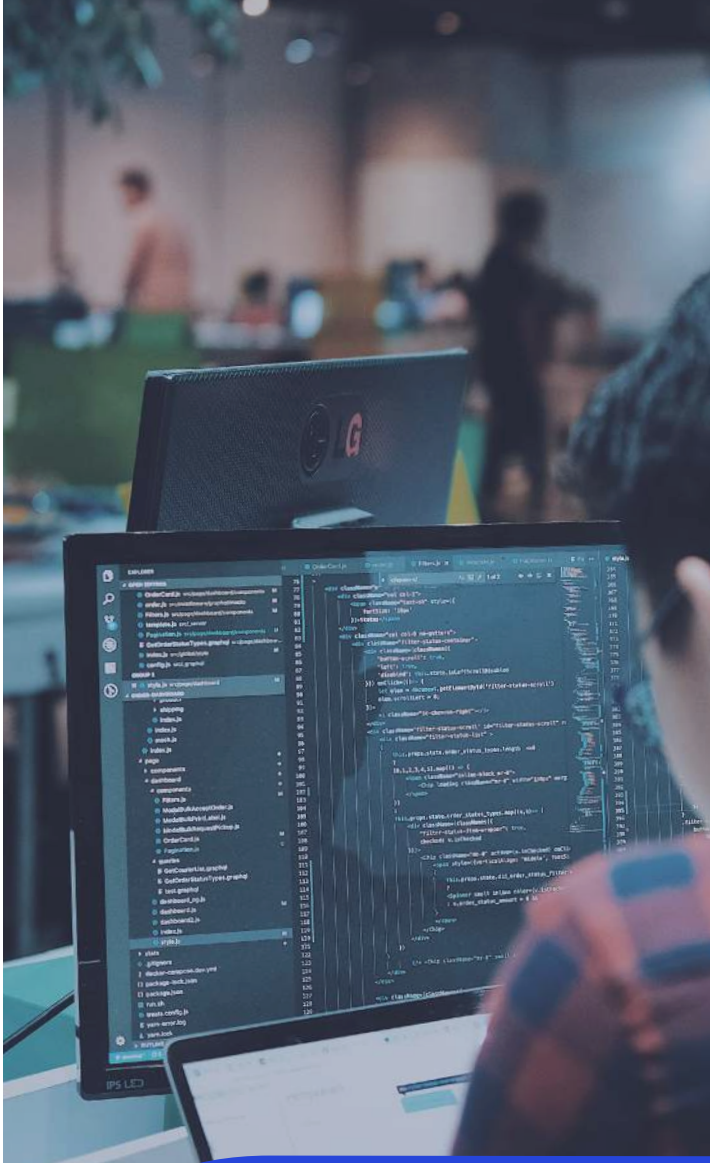
Artificial intelligence (AI) is rapidly moving out of the laboratory and into business and consumer applications. As a result, we're seeing a fundamental shift in how software is built and what it's capable of doing.

Of course, the type of AI we're talking about isn't the artificial general intelligence of science fiction, where a robot or software program can do whatever a person can and more. Instead, we're referring to artificial narrow intelligence. In other words, the use of AI in very specific contexts. It's this type of AI that's already being used to power some of today's most successful technology businesses, including Google, Facebook, Amazon, Netflix, Uber and Airbnb.

AI is a rich and complex topic, and one that has evolved as a result of years of computer science research. Today, we have reached the point where AI can be applied within most businesses, provided that they have the right talent, and that they're focused on applying it to the class of problems that AI is best suited to.

In recent years we've witnessed major progress in the field of AI. This is primarily thanks to advances in machine learning, one of its best-known and most widely talked about subfields. With machine learning, software algorithms learn new things from data without having to be specifically programmed to do so. In other words, machine learning can learn from and make predictions on data. The result is software that can tackle tasks that would otherwise be too complex to program manually.

This paper isn't intended to help you understand the latest advancements in the field of AI. Rather, our goal is to help you recognize the problems to which artificial intelligence can be applied. We also want to help you develop a strategy for delivering capability and customer value that would otherwise be impractical, if not impossible, without AI. At its core, this is what we mean by the term applied artificial intelligence (applied AI).



Today, we have reached the point where AI can be applied within most businesses, provided that they have the right talent, and that they're focused on applying it to the class of problems that AI is best suited to.

From Applied Analytics to Applied Artificial Intelligence

For those readers familiar with our principles of applied analytics, or who have implemented applied analytics solutions, you'll notice some similarities in the structure presented here. That's because applied analytics can be an example of AI models applied to the delivery of insights into a business process. However, applied AI extends the applied analytics framework with:

- An increased focus on end-to-end process automation
- Use of advanced machine learning techniques
- Better articulated performance objectives and model accuracy
- Appropriate use of human judgment and automated predictions
- Continuous learning and adaptation through feedback loops
- Approaches to limit the impact of model error and bias

A notable departure from the applied analytics framework is that we talk about integration first and data last. By integration, we mean how an AI-enabled process will be consumed by other business processes or people. There are many options and techniques possible for integrating AI into a business. The selection of modeling techniques may differ depending on what integrations are required. In addition, the data required to support the models depends on the selected modeling techniques, so we address that topic later.

While the applied analytics principles support businesses that aspire to derive greater insights from their data to drive business processes, the applied AI principles are designed for businesses that aspire to total process automation.

How to Use This White Paper

This white paper introduces the principles of applied AI, along with a framework and maturity model for applying AI to your organization. These principles bring structure to an important, and at times highly complex, topic, and can be a useful reference point when it comes time to develop and execute your AI strategy.

A Framework for Applied Artificial Intelligence

The principles themselves fit into five related groupings as shown in the framework below. The purpose of grouping them in this way is to separate the principles into areas that can be discussed both individually and as part of the larger framework.

START WITH THE PROCESSES

AI has the potential to affect every business process and job role either through augmentation or automation. Business value can be impacted by injecting insights into existing processes, optimizing processes, or automating entire or parts of business processes. Beyond that, AI provides opportunities to redefine or create entirely new processes that weren't possible before, even with human effort.

To identify the places with the greatest potential for AI in your business, start with the processes your business controls or should control, and where improved outcomes or efficiency would have the highest benefit to your and/or your customers' businesses. Understand which parts of the process involve human judgment, and which of those parts should be optimized, augmented, automated or reinvented.

INTEGRATION

Integrate artificial intelligence-enabled processes into your business via APIs for software and intuitive interfaces for people. Integrate human judgment as required to improve predictions. Leverage your customer base, in-house domain experts or crowdsourcing engines to supplement model predictions. Enable continuous learning by observing, measuring and improving outcomes as actions are delivered within a business process.

MODELS

The first step in model selection is to determine the desired level of performance from a business perspective. Select techniques to build machine learning models that achieve desired outcomes and to fit budget, objectives and available training data for each of the most valuable predictions and integration points. A lack of planning and achievement of objectives for error and bias will increase business risk and reputational harm. Business differentiation will come from the optimal combination of model integration, the modeling techniques themselves, and the data used to train and optimize the models.

DATA

The type and volume of data required for applied AI depends on the machine learning technique(s) selected. This data may be that which you already collect and have access to (albeit perhaps with additional labeling), or it may need to be acquired from your own sensors or from third parties. Develop plans and infrastructure to capture and cleanse relevant data to support your AI models, and adjust frequency and breadth as necessary.

GOVERNANCE

Monitor performance and fairness, and build customer trust. Ensure sufficient understanding of models and data to satisfy expectations of repeatability and interpretability. Adopt a market leadership position on ethical, privacy and legal concerns related to use of artificial intelligence.

Maturity Levels

LEVEL ONE

CAPABILITY AND INGREDIENTS

Your organization has intrinsic company characteristics, assets or strategic intent that give you an advantage for executing on trust. You have a basic understanding of priorities and opportunities, and there are few or no specialist skills within your organization.

LEVEL TWO

READINESS AND PROCESS ADVANTAGE

You have formalized company standards, practices, procedures and skills that give you an advantage for execution. You have partially documented your approach and priorities, and acquired some specialized skills within your organization.

LEVEL THREE

EXECUTION ADVANTAGE

You're exploiting advantages with mature processes that capture outcomes. You're measuring trust and regularly checking your approach with customers to drive continuous improvement. Last, but not least, you have complete documentation of your approach, including roadmaps for implementation, and a complete set of appropriate skills to support the implementation of trust within your organization.

START WITH THE PROCESSES

PRINCIPLE 1

Understand all processes

Catalog the processes your software solution enables, any adjacent customer processes and relevant third-party processes.

LEVEL ONE

Opportunities identified in an ad-hoc manner, primarily within your company's solution.

LEVEL TWO

Key business processes documented and prioritized.

LEVEL THREE

All business processes prioritized across wider ecosystem; roadmap developed.

Applied AI represents an opportunity to intelligently automate or augment a wide range of business processes, including those your business doesn't control today. To best direct your efforts, focus on understanding the ecosystem of business processes in which you operate. To do so, catalog all of the relevant ones.

Start by documenting the business processes that your existing software solution enables (e.g., the management of payments to suppliers). Next, consider areas of your customers' business where they have processes adjacent to your area of expertise that are delivered using manual effort or custom software solutions (e.g., manual processing of an incoming invoice in an otherwise automated payments workflow). Finally, take note of adjacent business processes delivered by third parties, including partners and competitors (e.g., customer support software for managing inquiries from suppliers awaiting payment).

Don't limit your thinking to improving existing processes. Instead, also consider how processes might be extended further if new types of predictions and automations were possible. Think, too, about opportunities to create new processes that aren't currently possible with human effort and intelligence alone. For example, commercial loan approval has traditionally been a manual process with loan officers using simple models and limited data to make their decisions. New machine learning approaches now enable much richer decision-making that augments human judgment with models that can include dozens of variables while also incorporating historic data.

Document how predictions and decisions are currently made for each process. Are they made by people? Automated using business rules? Pay attention to potential information gaps within the ecosystem that limit current predictions or decision-making. Finally, identify and document any key performance indicators (KPIs) used to measure a particular business process.

Key Takeaways

1. Map all existing (and potential) business processes.
2. Consider processes managed by you, your customers or third parties.
3. Document the decisions and actions for each.
4. Consider key control points and any existing automation.

PRINCIPLE 2

Prioritize the most valuable processes

Identify and prioritize which business processes are best suited to be automated using applied AI.

LEVEL ONE

Prioritization of opportunities for automating processes started, but incomplete.

LEVEL TWO

Some high-value opportunities for optimizing existing insights and automation have been prioritized and are part of the business solution.

LEVEL THREE

Most opportunities to optimize existing automation and augmentation have been mapped out and some are in production. Some high-value opportunities for automation have been identified.

There are a variety of ways that artificial intelligence could be applied to your business. These range from simply improving the quality or efficiency of existing insights to augmenting human decision-making. Other common applications include fully automating business processes that previously could only be undertaken by people, as well as processes that weren't possible before even with human insights and decision-making.

Not every opportunity is equally valuable. Some uses of artificial intelligence in your business will create outsized value compared to others. It's therefore important to understand value, cost, risk and timelines to help identify and prioritize the best opportunities to focus on.

We recommend scoring each potential process using the following attributes. With each, consider the relative opportunity to create value, what the cost might be and what the risk of implementation is:

- **Understanding.** How well must the process be understood? Is the process well enough defined that a model could be trained to deliver human-like outcomes with sufficient quality?
- **Automation.** Does the opportunity represent a means to improve the quality of insights and decision-making in a process (augmentation), or is it likely to enable full intelligent automation of the process?
- **Scale.** Would automating this process remove constraints in the system, allowing more work to be done without an increase in human effort?
- **Quality.** Is the required quality of the process outcomes understood? Are there specific legal or other governance requirements on the quality of outcomes? Would AI-driven automation meet, or even exceed, the quality of any current human or automated decisions if they exist?
- **Data.** Is there data being captured or could it be captured to allow the training of models with sufficiently high quality? Is there sufficient labeled data to support machine learning or a cost-effective means to create labels for the data?

- **Control.** Are you in control of a sufficient part of the process and do you have integrations into all decision endpoints so that you can have a high level of impact through automated actions?

The output of the exercise should be a list of candidate processes for automation and augmentation that have been prioritized (based on relative value, cost and risk) for further investigation.

Key Takeaways

1. Look for opportunities within your business to apply AI ranging from improved insights through to full automation.
2. Start with the desired outcomes that would have the greatest impact within your customers' processes.
3. Recognize where in those processes you're short on instrumentation or integrations.
4. Score potential opportunities across various dimensions to identify your top priorities.

INTEGRATION

PRINCIPLE 3

Design frictionless integrations

Understand what a proposed solution needs from other systems, what triggers its execution, interdependencies with other processes, and how the predictions it makes will cause actions to be taken.

LEVEL ONE

Easiest integrations handled first. Some friction exists and simplification is possible. More integrations possible to increase AI utility.

LEVEL TWO

Roadmap exists for major integrations. Care being taken to simplify use and reduce friction for a wide set of users. Implementation underway.

LEVEL THREE

AI capabilities that deliver high levels of automation are easy to consume across the business. All processes have necessary integrations. Have found best way to present AI results to users. New integrations added systematically as needed.

As with all design processes, once the general objectives for the solution are understood, a good place to start is by considering the external aspects of the solution. That includes how it looks from the outside and how it interacts with the environment in which it operates. There will be plenty of time to deal with the details of data sources, machine learning models and quality. Initially, get as clear as possible on what the solution needs from other systems, what triggers its execution, and how the predictions it makes will cause actions to be taken. These are the integrations to be designed.

By frictionless, we mean that the objective for every integration is to be as automated, natural and complete as possible. Some questions that you may want to answer as part of the design to achieve the most frictionless interface include:

- **Is the AI solution fully embedded within automated processes?** If so, the AI predictions might be able to drive follow-on actions with no intervention or friction whatsoever.
- **Does the optimal AI solution need to deliver a recommendation to a person?** Then choose the most natural persona to simulate the expert, and the most natural interface the person would use.
- **Is the AI meant to engage a person in a conversation?** Think about the how, when and where of that conversation to ensure the most natural medium, language and timing to reduce friction. In some cases, the AI might provide feedback via visuals or other UX elements such as buttons.
- **Are the steps that follow the AI prediction within your company's control?** If not, try to reduce AI friction through partnerships to extend the solution's reach.
- **Is the AI solution one that might be valuable to new users, systems or even third parties over time?** Consider building interfaces through APIs to increase future flexibility and reduce friction.

At this stage, you can ignore the training process, the sources of data and the complexity of the models. They are all important, but the highest-leverage design decision for the AI solution is how it integrates into existing processes in a way that makes it most natural to use.

Key Takeaways

1. Take an outside-in approach.
2. Identify interdependencies with other processes.
3. Aim for solutions to be as automated, natural and complete as possible.
4. The more integrated your applied AI solution is, the greater the value.

Company Example

Lemonade applies AI to disrupt the insurance industry. The software provides a frictionless user experience via conversational interfaces and natural language processing. There are rich back-end integrations to process claims using machine learning and to generate payments automatically.

See more at:

Shai Winiger, "The Secret Behind Lemonade's Instant Insurance."

INTEGRATION

PRINCIPLE 4

Integrate human judgment as required

Consider where people may need to play a role in your solution.

LEVEL ONE

Little thought given to which processes need human judgment and how it will be used to continuously improve model predictions and actions.

LEVEL TWO

Some thought given to which processes need humans in the loop. Some use of people to take actions in cases of predictions with low confidence. Some feedback incorporated into learning process and not all done in real time.

LEVEL THREE

People integrated into select processes to add judgment to predictions to optimize outcomes and meet regulatory and user requirements. Higher levels of human judgment are iteratively replaced with more accurate predictions to increase levels of automation.

While in many cases the goal may be to complete process automation using applied AI, there are a number of reasons why that may not be possible. These include insufficient model accuracy, government regulations, and negative market or customer sentiment toward full automation. There may also be particular attributes of a business process that currently make full automation too costly or impossible.

Human involvement may include providing a level of oversight for automated predictions as well as providing feedback to improve performance. For example, when an AI solution is bootstrapped with little data, human decisions may accelerate model training to meet performance quality objectives. That human input can be captured either implicitly through product interaction data, including historic data provided in bulk to initially train the model, as well as through newly generated user interaction data. The system may also generate specific requests to multiple human trainers to classify a new piece of data and then incorporate that feedback into the model. Incorporation of this human interaction data is an important way of optimizing the system faster.

In-house experts such as financial analysts, compliance officers, nurses, attorneys and sales agents can also be an important source of judgment and training data. Having such domain expertise can be a competitive advantage. That's particularly true when the impact of decisions is high, and the availability of experts in the market who can reliably make those decisions is low.

Additionally, people can serve as gatekeepers in situations where predictions may introduce disproportionate risk or when actions are irreversible. Regulation often plays a role in these domains and using AI to augment user decisions may be the only option.

For some processes, users may expect to be in control, thereby constraining the level of automation. Gaining the confidence of your users over time should allow you to increase levels of automation without human triggers.

Finally, there may be business processes in which the AI model provides a prediction, but where people must be involved to complete the action because the end-to-end process isn't within your business's control. For example, where the action resulting from an automated prediction in your business process is to be performed by a business partner.

Key Takeaways

1. Understand the roles people play in your solution.
2. People may be sources of data, trainers, reviewers or users of predictions.
3. Having humans in the loop may be a user, market or regulatory requirement.
4. Even when full automation is possible, human oversight may be required.

Company Example

Stitch Fix, an online provider of custom garments, makes recommendations and ships clothing to customers, without the clothes being seen by the customer ahead of time.

The company applies machine learning to select clothing items based on inputs provided by the customer (images, answers to questions, etc.). Fashion experts then make the final selection of a subset of the items. The machine learning algorithm is used to crunch large amounts of input data, while the human experts are able to apply knowledge of social norms, cognition and improvisation when selecting the items that are eventually shipped. Feedback from users is also incorporated into the AI learning process.

See more at:

Simone Ahuja, "What Stitch Fix Figured Out About Mass Customization," Harvard Business Review, May 26, 2015.

MODELS

PRINCIPLE 5

Understand performance objectives

Analyze performance expectations in terms of errors, robustness and bias.

LEVEL ONE

Incomplete and generic model performance objectives employed.

LEVEL TWO

Model performance objectives and the business impact of model predictions are somewhat understood.

LEVEL THREE

Rich user-based definition of model performance, including error tolerance, robustness and bias for every prediction. Expected performance levels are continuously validated with customers.

Given that achieving perfect performance is rarely possible, and even more difficult to prove, it's essential to understand (and meet or exceed) the performance expectations for your market and application in terms of:

- **Errors.** False predictions originating from incorrect, incomplete or noisy training data, or from programming errors in the model itself.
- **Robustness.** Also known as model stability, robustness refers to how much the model is impacted by small changes in input.
- **Bias.** The unintended differential treatment of, or impact on, a specific group on the basis of common characteristics that are injected into an AI model through training data.

Performance expectations will most likely be industry- and even use case-specific. For example, in medical applications, there's very little tolerance for errors such as false negatives that cause a condition to be missed. By contrast, marketing applications that recommend products to consumers are much more easily forgiven for mistakes.

It's also important to consider the level of consistency with which the model must perform to meet expectations. Perfect performance is often not be required, but there will be some expectations of how well the model will perform with different inputs. Be aware, too, that some industries will have set rules, and penalties, to ensure that bias doesn't enter into important decisions. For example, in financial services, loan decisions must not take certain factors such as race or gender into consideration.

Key Takeaways

1. Understand what performance levels are required to add business value.
2. Model performance includes error rates, consistency and bias.
3. Determine the cost of improving model accuracy vs. the impact of false predictions.
4. Avoid models where the impact of false predictions is greater than the business can tolerate.

Company Example

PatternEx, a start-up applying machine learning to security operations, is an example of a company that understands and measures against model performance goals. In security operations, errors can be expensive. The typical high cost of intrusions makes minimizing false negatives a key concern. At the same time, false positives must be kept low, due to the costs of unnecessary analyst investigations. The company has optimized its human-in-the-loop AI system on both goals: the system can detect significantly more events than the baseline (85 percent of attacks versus less than 8 percent) with significantly fewer false positives (5 times fewer).

See more at:

"AI^2: An AI-Driven Predictive Cybersecurity Platform," MITCSAIL, April 19, 2016.

MODELS

PRINCIPLE 6

Start with proven modeling techniques

Aim for the simplest model that can do the largest part of the job.

LEVEL ONE

Limited understanding of all possible machine learning techniques. Most simple techniques available are generally used. Acceptable fit to data and performance objectives.

LEVEL TWO

Some process decomposition to achieve simplification as required as well as use of richer techniques. Experimentation used to optimize selection. Ongoing review of model performance and investigation of better alternatives. AI plan exists.

LEVEL THREE

Expert understanding of the current ML landscape driving selection of the best possible techniques for the widest set of problems. Active in the research community, advancing the state of the art, and consistently testing results to optimize models. Rich AI plan.

The process of technique selection is not prescriptive itself. In the fast-changing world of artificial intelligence in general, and machine learning specifically, what couldn't be done yesterday can be done today. And, what's now possible may be faster and more accurate in the near future.

Generally, use proven methods. The resulting models are typically easier to implement, maintain and explain to customers. Using a proven method doesn't necessarily sacrifice the richness or sophistication of the approach. Advanced machine learning methods are continuously being proven to work for different problems. Moreover, these methods are rapidly supported by commercial tools, open-source implementations, and educational resources that help reduce the complexity and cost of adoption.

The required degree of explainability of model predictions varies by problem and should be a consideration in technique selection. Users adopting a new AI solution often want to understand how the output of the model was generated. Regulation is also driving model explainability requirements. Models vary in the degree of explainability, and research is continuously improving the explainability of even the most advanced techniques. Understand your explainability requirements, and keep current on research work to ensure they can be met with selected techniques.

In some cases, business problems can be simplified by breaking them down into smaller issues. Different modeling techniques can be matched to each problem to get the best overall model and highest accuracy predictions. Multiple modeling techniques and resulting models can also be used for a single portion of the problem. This is referred to as ensemble learning.

To support optimal technique selection for all business solutions, hire experienced and curious machine learning scientists and engineers, and enable them to keep current on, and even advance, the state of the art. The machine learning scientist's job also includes choosing and tuning relevant model parameters and testing trained models. Some of the machine learning scientist workflows may be automated over time with the maturing of AI techniques and frameworks.

Each trained model should be tested using carefully crafted test sets tailored for different business objectives. These will be used to ensure the

documented performance objectives are met and to compare performance across different models for the same problem, or to support model tuning to improve the performance of a selected model. Different business problems require different levels of performance. Moreover, testing has a high cost and model tuning provides diminishing returns. As a result, only do sufficient testing to satisfy the objectives. Managing model quality must continue throughout the use of the models, since model results may change over time.

Document how each business process will be decomposed for modeling, the modeling techniques that will be used for each portion, and the data and skills required for training and testing the models. The plan should include future considerations of model and data improvement, the opportunity to increase automation and get to a best AI solution, and some guidance as to the improvement cycle timeline. Consider the cadence at which models will be updated.

Your leadership team should review, understand and champion those plans and ask questions about training, testing, meeting performance objectives, interpretability and the degree of automation. While AI models are complicated and constantly changing, the approach to their implementation and use should make business sense to all members of your company's leadership team.

Key Takeaways

1. AI and machine learning techniques are changing fast.
2. Aim for the simplest model that can do the largest part of the job.
3. Consider whether explainability of the model is a requirement.
4. Simplify processes using decomposition as needed to fit simpler models.
5. Hire experienced machine learning scientists for optimal outcomes.
6. Test for and compare model performance.
7. Have a plan for evaluating new, more sophisticated models.

Company Example

The payments company Stripe has had success using proven models such as random forests for a wide range of machine learning challenges. This is despite having constantly experimented with other models for fraud detection such as deep learning.

Airbnb considers key trade-offs such as model predictive capability vs. explainability when deciding on the suitability of one approach vs. another. Techniques such as deep learning, for example, are less easily explained than a simpler linear model. However, for certain applications (e.g., machine translation), the performance of more complex approaches such as deep learning have been shown to be significantly better while explainability is also less important.

See more at:

Michael Manapat, "A Primer on Machine Learning for Fraud Detection," Stripe.
Robert Chang, "Using Machine Learning to Predict Value of Homes on Airbnb," July 17, 2017, Medium.
Gideon Lewis-Kraus, "The Great A.I. Awakening," The New York Times, December 14, 2016.

DATA

PRINCIPLE 7

Capture relevant data to support models

Develop any necessary plans and infrastructure to capture relevant data to support your machine learning models.

LEVEL ONE

Historical data captured for most processes, including most inputs, actions and outcomes, and some capture of context. Little to no understanding of how relevant the data is. Little to no infrastructure in place to match frequency.

LEVEL TWO

Most relevant data inputs, context, decisions, actions and outcomes captured for some processes at the right frequency.

LEVEL THREE

Full capture of relevant data inputs, context, decisions, actions and outcomes for all processes at the required frequency to match techniques and performance objectives. Appropriate use of data generation and simulation to amplify model training.

Depending on the type of problem you're incorporating AI into, and the machine learning techniques used, you'll need appropriate data to train and optimize the machine learning models. The quality of data is as important as your choice of machine learning techniques. You may already collect and have access to the data needed, although that data may require additional labeling. You might also need to acquire data from third parties. Given the performance objectives, the data characteristics will also inform the techniques you can apply.

Develop plans and infrastructure, as necessary, to capture relevant data to support your machine learning models. You may need to increase the frequency with which you collect data and instrument processes with new sensors to capture input, context, decision, action and outcome data end-to-end and in real time. Design the data-capturing pipeline to minimize data loss and, when using real-time data, use mechanisms to monitor data quality in real time.

The first step in solving any problem using ML techniques is to transform raw data to best represent the real-world entity. Identifying the best set of features for each task to maximize the utility of the model is called feature engineering and can have a large impact on model performance. One of the key advantages of deep neural networks for different applications is that these models can be trained through an end-to-end process to learn both input representation and parameters. Thus, task and domain-specific feature engineering may not always be required.

Finally, you can import subject matter expertise or leverage pre-trained models as starting points, or even derive data using data generation and simulation techniques, if necessary. These approaches can complement real data and can help address cold start issues, particularly in cases where capturing and leveraging sufficient amounts of in-process real data is a slow process.

Key Takeaways

1. Capture comprehensive and diverse data from business processes.
2. Such relevant data will be used to train your models.
3. Include input, context, action and outcome data from end-to-end processes in real time.
4. Derive general and edge case data using data generation and simulation techniques if necessary.

Company Example

Blue River Technology applies AI to treat agriculture crops and optimize output. The company uses high-resolution plant images captured in real time via cameras installed on the equipment to understand plant characteristics and direct the application of herbicide to weeds and in the right amounts. The technology stack includes a second set of cameras that capture images and verify that the right actions are being taken. Finally, drones are used to collect data over field plots and understand the effects of the actions and close the feedback loop.

See more at:

Blue River Technology

GOVERNANCE

PRINCIPLE 8

Manage quality for error and bias

Understand that bias may come from human inputs that you are using to train or evaluate your models. Implement processes for monitoring for bias and other errors.

LEVEL ONE

Some awareness of the sources of bias and error, but no comprehensive approach to prevent and detect them.

LEVEL TWO

Processes implemented to prevent the introduction of bias and errors and understand the remedies that can be used when they're detected.

LEVEL THREE

Effective execution to achieve and maintain the highest possible model utility while preventing the introduction of bias. Monitoring for errors throughout automated machine learning systems.

Organizations must proactively recognize the potential to introduce bias into an AI system and design processes to avoid it. They must then automate the monitoring of system behavior at each stage of the AI process, and correct error and bias on detection. Most organizations will understand the impact of errors on their business, but many don't yet consider the impact of bias.

While machine learning is often portrayed as an objective, fair and data-driven approach to decision-making, that can only be true if models are free of bias. Bias does not include the implementation of intended business decisions that, for example, may limit a company's business activity to a certain geography, or certain market segment. However, there is risk when a business has experience in one portion of a market, and then uses its customer and transaction data in the context of a broader market. This exposes a blind spot, and can introduce bias into an AI model.

Bias and errors are related but different. However, the processes to avoid, detect and correct for either are similar. Often the problem in a model outcome will need to be investigated before it's clear if the source of the problem was an error or an underlying bias originating from the input data. But, as with software development in general, errors are more difficult and expensive to find and repair the later they are found. Frequent review and early defect removal are possible and recommended in the development of AI solutions, and should result in better, more consistent and easier-to-interpret results.

Many types of input can be used to train machine learning models, including transactional data, which represents the real-time capture of business process execution. If the execution of the processes themselves have bias (hiring, promoting, lending, etc.), the model will likely be trained to reflect the bias, if no remedial steps are taken. In addition, some data may be labeled by employees, customers or a third party. If the rules they use to create the labels have bias, or the samples they are labeling are not fully representative of what the models may see at run-time, then the models will again reflect the bias.

In some cases, human judgment will be added to a AI process to monitor and tune the results of models. People's reactions to AI-based predictions can also be biased based on their assumptions of higher or lower accuracy than reality, depending on the person. Further, human judgment may have the same underlying business process bias as the organization itself, making it difficult to detect a model error.

To minimize the risk of introducing bias, you first need to broadly consider potential biases in your organization's processes, training data and human participation in training. Increase the diversity of data and participants, and use redundancy to reduce the potential impact of bias. Second, build assessments into your AI processes to detect incomplete or erroneous input data that might lead to bias and error. Finally, review model outcomes over time to ensure they continue to meet your business practice standards as model behavior can shift with new data.

Key Takeaways

1. Machine learning is only objective and fair if it's free of bias.
2. Implement processes for monitoring for bias and other errors.
3. Bias may come from existing business processes and their data.
4. Human input into training or evaluating AI predictions may be biased itself.

Company Example

An example of machine learning bias in the real world can be seen at LinkedIn, where high-paying jobs weren't being displayed as frequently to women as men. For its part, Google, also no stranger to issues with bias in its algorithms, has developed a tool that can be used to detect bias in machine learning systems.

See more at:

Hope Reese, "Bias in Machine Learning and How to Stop It," TechRepublic, November 18, 2016.

Molly Mulshine, "A Major Flaw in Google's Algorithm Allegedly Tagged Two Black People's Faces with the Word 'Gorillas'," Business Insider, July 2, 2015.

Mike Wheatley, "Google Researchers Develop a Test for Machine Learning Bias," Silicon Angle, December 22, 2016.

GOVERNANCE

PRINCIPLE 9

Build fault tolerance as part of model integration

Recognize that models will have errors, behave unexpectedly and sometimes fail. Try to increase system reliability through fault tolerance.

LEVEL ONE

No specific fault-tolerance techniques implemented.

LEVEL TWO

Some fault tolerance is implemented to prevent unwanted impacts from the most critical errors.

LEVEL THREE

Fault tolerance is built into the system from the start. Uses a rich set of techniques to allow earliest possible detection and minimize the impact of errors in automated processes.

All systems built by people have errors that can result in unexpected behavior when they're put into production. Non-AI systems have the benefit of providing generally consistent execution based on system inputs. That is, testing a non-AI system with a fixed set of inputs can largely eliminate errors prior to promotion to a production environment. A notable exception are errors that originate through timing — either sequences of events, time-of-day, or multi-process race conditions. These are a difficult kind of error to test for and eliminate, and most closely resemble those that can affect a learning AI system.

For AI-based systems, the behavior of the system can change over time as the model learns from new data, including human interventions. Even if you expand your training and testing data to include disaster scenarios, eliminating all problems in advance of the system's use in a production environment is unlikely. In addition, increased automation of an AI system may allow an error to be introduced and go undetected for some time, affecting the quality of the actions, and reinforcing and potentially increasing the degree of error over time. To prevent this, all traditional fault detection techniques for downstream processes must be in place when actions are initiated by an AI system.

Fault tolerance must be designed and built into AI-based systems. Techniques such as rules-based circuit breakers, and human and AI-based monitoring and audits, check-and-balance redundancy, mandatory human review, recognizing data shifts, kill switches and least-harm actions are all viable and must be used selectively to provide results that sufficiently reduce the consequences of error.

Key Takeaways

1. Models will have errors, behave unexpectedly and sometimes fail.
2. Learning systems can shift over time with new data.
3. Automation can amplify the impact of errors.
4. Traditional approaches to error detection and elimination are inadequate.
5. Increase system dependency through fault tolerance.
6. Consider rules-based circuit breakers, human and AI-based monitoring and audits, check-and-balance redundancy, mandatory human review, kill switches and least-harm actions.

Company Example

Google is researching approaches to create a “kill switch” for artificial intelligence that would enable humans to shut down an algorithm that is not behaving optimally. This is particularly important in situations where reinforcement learning is being used and there is a need for a human operator to stop the algorithm if it starts taking actions that are harmful to either itself or the environment in which it is operating.

See more at:

“Google Developing Kill Switch for AI,” BBC News, June 8, 2016.

Laurent Orseau and Stuart Armstrong, “Safely Interruptible Agents,” Intelligence.org.

RELIABILITY

PRINCIPLE 10

Defend your legal and ethical stance

Consider external factors such as legislation and customer sentiment, and be open and transparent.

LEVEL ONE

Meeting legal requirements on information rights and privacy. Little understanding of the implications of automation/AI around social bias, repeatability, interpretability and future of work.

LEVEL TWO

Adopting a market leadership position on information rights and privacy. Some understanding of market expectations on AI and social bias, repeatability, interpretability and future of work.

LEVEL THREE

Adopting a market leadership position on all ethical and legal issues. Arguments for business process automation are supported with facts.

AI and the data associated with it will continually raise ethical and legal issues. The ethical and legal investments you make will be relative to the domain space and market expectations, the problem, and the data. Develop a strong understanding of the overall benefits of AI in your problem domain. Benefits can include achieving massive efficiency gains and freeing up people to work on more engaging, high-value activities, reducing risks by bringing consistency into a process, or helping people solve the otherwise unsolvable. Frame ethical and legal concerns in the context of the overall benefits and support your arguments with facts.

First, you will have to accommodate external factors, such as legislation, that introduce restrictions on the information you gather and store. An example of such legislation is the General Data Protection Regulation (GDPR) in the European Union. Your customers may also restrict you from using data via information rights agreements. Have a decision framework for whether you need some specific data points and how the respective data is being used to support your AI strategy and create value.

In addition, have policies in place to trim data that is no longer useful both on the time and attribute dimensions. Finally, incorporate privacy-preserving techniques that provide guarantees, such as differential privacy, where possible. This should enable you to proactively address legislation requirements, be transparent with your customers, give them insights into what data you're recording and how they benefit from it, and gain their trust in the process.

Second, regulation can introduce restrictions on your AI-driven decisions and actions, to ensure fairness, but also reduce the risk introduced by automation in certain domains. Consider the consequences of an individual error when automating decision-making and have an explanation for why your AI-driven solution is better than the state of the art. If you start with augmenting experts because of regulatory restrictions or user requirements, how you explain your solution performance relative to the best alternative can allow more automation in the process, by influencing your customers and even regulation in the long run.

Third, acknowledge that increased automation will shift workloads and impact jobs, as most new technology waves prior to AI have. In this case, however, the effects may be more pronounced as AI solutions directly replace some human activities, as AI solutions evolve from augmentation to automation. Understand the impact and the issues surrounding these transitions.

Finally, it's important that you show transparency and openness and adopt a market leadership position about ethical and legal issues. With the use of AI, there is a responsibility to understand it and explain it to customers and educate the market. Talk at conferences and publish articles on privacy, fairness, error, interpretability of AI decisions and the overall changes automation brings to your market.

Key Takeaways

1. Understand the overall benefits of AI to your market.
2. Consider external factors such as legislation and customer sentiment.
3. Regulation may restrict automation levels to reduce risk and consequence.
4. Don't gloss over the replacement of human effort with automation.
5. Be open and transparent.

Company Example

Tesla has had to defend the performance of its "autopilot" AI-based driving capabilities following a fatal accident in 2016. The company made its case by focusing on the statistical improvement in safety with its autopilot, when compared to regular driving. Third-party data is also emerging that further validates the company's position.

See more at:

Will Oremus, "A Tesla Driver Died in a Crash While His Car Was on Autopilot," Slate, June 30, 2016.

"A Tragic Loss," Tesla.com, June 30, 2016.

Danielle Muoio, "Tesla's Autopilot Has Slashed Crash Rates for Its Cars by 40%," Business Insider, January 20, 2017.

Applied Artificial Intelligence Maturity Model

PRINCIPLE	LEVEL 1	LEVEL 2	LEVEL 3
1. Understand all processes	Opportunities identified in an ad-hoc manner, primarily within your company's solution.	Key business processes documented and prioritized.	All business processes prioritized across wider ecosystem; roadmap developed.
2. Prioritize the most valuable processes	Prioritization of opportunities for automating processes started, but incomplete.	Some high-value opportunities for optimizing existing insights and automation have been prioritized and are part of the business solution.	Most opportunities to optimize existing automation and augmentation have been mapped out and some are in production. Some high-value opportunities for automation have been identified.
3. Design frictionless integrations	Easiest integrations handled first. Some friction exists and simplification is possible. More integrations possible to increase AI utility.	Roadmap exists for major integrations. Care being taken to simplify use and reduce friction for a wide set of users. Implementation underway.	AI capabilities that deliver high levels of automation are easy to consume across the business. All processes have necessary integrations. Have found best way to present AI results to users. New integrations added systematically as needed.
4. Integrate human judgment as required	Little thought given to which processes need human judgment and how it will be used to continuously improve model predictions and actions.	Some thought given to which processes need humans in the loop. Some use of people to take actions in cases of predictions with low confidence. Some feedback incorporated into learning process and not all done in real time.	People integrated into select processes to add judgment to predictions to optimize outcomes and meet regulatory and user requirements. Higher levels of human judgment are iteratively replaced with more accurate predictions to increase levels of automation.

PRINCIPLE	LEVEL 1	LEVEL 2	LEVEL 3
5. Understand performance objectives	Incomplete and generic model performance objectives employed.	Model performance objectives and the business impact of model predictions are somewhat understood.	Rich user-based definition of model performance, including error tolerance, robustness and bias for every prediction. Expected performance levels are continuously validated with customers.
6. Start with proven modeling techniques	Limited understanding of all possible machine learning techniques. Most simple techniques available are generally used. Acceptable fit to data and performance objectives.	Some process decomposition to achieve simplification as required as well as use of richer techniques. Experimentation used to optimize selection. Ongoing review of model performance and investigation of better alternatives. AI plan exists.	Expert understanding of the current ML landscape driving selection of the best possible techniques for the widest set of problems. Active in the research community, advancing the state of the art, and consistently testing results to optimize models. Rich AI plan.
7. Capture relevant data to support models	Historical data captured for most processes, including most inputs, actions and outcomes, and some capture of context. Little to no understanding of how relevant the data is. Little to no infrastructure in place to match frequency.	Most relevant data inputs, context, decisions, actions and outcomes captured for some processes at the right frequency.	Full capture of relevant data inputs, context, decisions, actions and outcomes for all processes at the required frequency to match techniques and performance objectives. Appropriate use of data generation and simulation to amplify model training.
8. Manage quality for error and bias	Some awareness of the sources of bias and error, but no comprehensive approach to prevent and detect them.	Processes implemented to prevent the introduction of bias and errors and understand the remedies that can be used when they're detected.	Effective execution to achieve and maintain the highest possible model utility while preventing the introduction of bias. Monitoring for errors throughout automated machine learning systems.

PRINCIPLE	LEVEL 1	LEVEL 2	LEVEL 3
9. Build fault tolerance as part of model integration	No specific fault-tolerance techniques implemented.	Some fault tolerance is implemented to prevent unwanted impacts from the most critical errors.	Fault tolerance is built into the system from the start. Uses a rich set of techniques to allow earliest possible detection and minimize the impact of errors in automated processes.
10. Defend your legal and ethical stance	Meeting legal requirements on information rights and privacy. Little understanding of the implications of automation/ AI around social bias, repeatability, interpretability and future of work.	Adopting a market leadership position on information rights and privacy. Some understanding of market expectations on AI and social bias, repeatability, interpretability and future of work.	Adopting a market leadership position on all ethical and legal issues. Arguments for business process automation are supported with facts.

Georgian

At Georgian, we're building a platform to provide a better experience of growth capital to CEOs and their teams.

Georgian's platform is designed to identify and accelerate the best growth-stage software companies, taking an intelligent, data-driven approach to solving the key challenges CEOs face as they grow their businesses.

We invest in high growth companies across North America that harness the power of data in a trustworthy way. Based in Toronto, Georgian's team brings together software entrepreneurs, machine learning experts, experienced operators and investment professionals.

info@georgian.io
georgian.io