How to build risk management into AI development

Prasanthi Desiraju, Vittorio Pepe, and Will Shin

MSDS 485 Data Governance, Ethics & Law

May 15, 2021

Data-driven decision-making has become an inherent part of many Organizations with the advancements in Artificial Intelligence (AI) and Machine Learning (ML). ML helps automate decisions and provides meaningful insights enabling decision-makers to take informed actions and justify their decisions. However, data isn't infallible and is susceptible to biases, incoherence, and other such blind spots if not curated and might lead business in the wrong direction. Thus the introduction of AI capabilities in your company might expose you to new risks that need to be addressed and in this document, we will introduce a framework to identify and manage them.

There are four main factors that make managing risk associated with AI a challenging endeavor:

1. AI risks embrace many aspects as model, compliance, operational, legal, reputational, and regulatory. Moreover, it introduces new risk factors as systematic bias. Even in companies with a dedicated risk management team and are used to dealing with some forms of the previous risks, AI will introduce a different angle to them that needs to be addressed. These teams will be asked to manage risks across a broader spectrum of business functions, often straining the team resources and moving in areas of business on which the team has no previous expertise.
2. The availability of third-party 'out of the box' and vendor-provided AI solutions contributes to a fragmented and widespread usage of AI capabilities that are often not centrally governed, making the task of managing the risk associated with them more complicated.
3. AI risk management is a new field, and introducing it in an existing organization provides challenges on its own. You need to define the level of risk acceptable in this area, how it

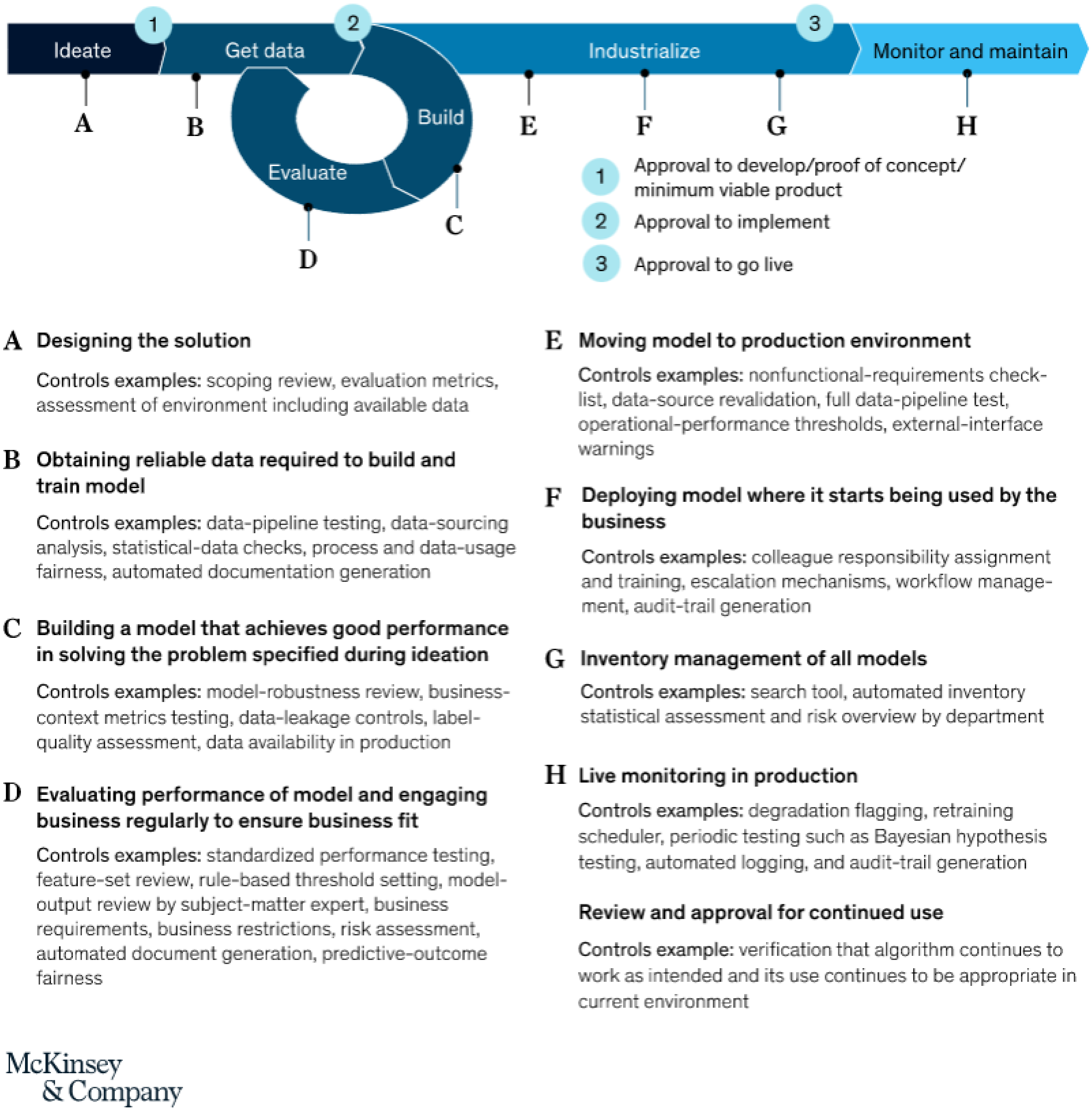
will be managed (locally or globally), and how it is related to other risks such as data privacy, cybers security, data ethics, and others.

4. AI technology is usually based on the analysis of historical data that can change rapidly, triggering model retraining or adjustment. This process could be quick, and a traditional post-development risk review would delay the implementation and nullify the update effort. The data used could become' old' during the review process. AI is a very dynamic field and needs an equally dynamic way of addressing the risks.

Many of the risks posed by AI are already present in other areas of companies that use predictive analytical models. As mentioned, the usual strategy in dealing with these issues is to establish a review phase after development. This kind of strategy is based on assumptions as the model is mostly developed in-house, and it is not updated quickly, with this review typically happening every year. In the AI landscape, those assumptions are not always true. The AI software can be embedded in third-party vendor software or provided as SaaS, significantly reducing the visibility of the entire process. And models could be updated as frequently as every week.

To address these peculiar issues, we propose a framework introduced by McKinsey & Co, de-risking AI by design. The risk review phase is embedded in the life cycle of the software (development, implementation, and production), establishing reviews and metrics since the beginning and consistent across the company.

The framework can be summarized in the below image:



A Designing the solution

Controls examples: scoping review, evaluation metrics, assessment of environment including available data

B Obtaining reliable data required to build and train model

Controls examples: data-pipeline testing, data-sourcing analysis, statistical-data checks, process and data-usage fairness, automated documentation generation

C Building a model that achieves good performance in solving the problem specified during ideation

Controls examples: model-robustness review, business-context metrics testing, data-leakage controls, label-quality assessment, data availability in production

D Evaluating performance of model and engaging business regularly to ensure business fit

Controls examples: standardized performance testing, feature-set review, rule-based threshold setting, model-output review by subject-matter expert, business requirements, business restrictions, risk assessment, automated document generation, predictive-outcome fairness

E Moving model to production environment

Controls examples: nonfunctional-requirements check-list, data-source revalidation, full data-pipeline test, operational-performance thresholds, external-interface warnings

F Deploying model where it starts being used by the business

Controls examples: colleague responsibility assignment and training, escalation mechanisms, workflow management, audit-trail generation

G Inventory management of all models

Controls examples: search tool, automated inventory statistical assessment and risk overview by department

H Live monitoring in production

Controls examples: degradation flagging, retraining scheduler, periodic testing such as Bayesian hypothesis testing, automated logging, and audit-trail generation

Review and approval for continued use

Controls example: verification that algorithm continues to work as intended and its use continues to be appropriate in current environment

McKinsey
& Company

Fig 1: De Risking AI by Design - Risk Management through algorithmic model’s life cycle

While the above process builds a strong framework by embedding AI at every stage of model development, having an additional checklist to be monitored at every phase will ensure the majority of risks have already been accounted for and mitigated. The following suggests a brief overview of the checklist to be followed to avoid any unintended consequences of AI throughout the lifecycle and have the system in control.

Step 1 : Ideate and Get Started

1. Articulate the key question in the form of a hypothesis, framing a null hypothesis that can be proven or disproven.
2. Identify the area of business which will have a major impact on.
3. Identify the roles and responsibilities and assign people to achieve the targeted measurable goals.
4. Establish data governance and develop right capabilities
5. Documentation of the processes and approvals

Step 2 : Design And Data

1. Identify the data and data sources you need.
2. Questions Around Data - Is the data reliable, Is the data accessible, Is the data real-time, Is the data available and at what cost ?
3. Is it worth collecting all the data - What are costs, Data Minimization, Legal implications, Data privacy
4. Set budget and get necessary approvals to collect the unavailable data.
5. Define processes and protocols for collection of data.

6. Documentation of the processes and approvals
7. Are there any potential risks or cybersecurity issues while extracting data from a third party system ?

Step 3: Data Analysis

1. Feature engineering and EDA on the preliminary data.
2. Is the data a true representation of the Problem Statement ? Evaluate bias-detection techniques and fair-representation techniques
3. Guidance on convening a diverse team , creation of bias-risk metrics, possible impact of bias on the decision and brand reputation.

Step 4: Extract Insights

1. Develop an understandable and explainable model as opposed to a black box algorithmic model.
2. Model different scenarios based on business use-case.
3. Extract analytics and test the hypothesis

Step 5: Communicate

1. Articulate and present the findings via graphs, visualizations etc
2. Understand the impact of proposed changes and make recommendations/decisions.
3. Create a step-by-step execution plan.
4. Communicate with key stakeholders
5. Document the results.

Step 6: Monitoring and Maintenance

1. Define Performance monitoring requirements
2. System Scalability as data load increases
3. Scope for reinforcement learning

This risk strategy presentation serves as a general framework and guide to minimize and mitigate risks associated with AI. Each organization should apply this framework as best as can meet its goal based on its mission and resources, while remaining within the boundaries of all applicable laws and regulations. And regardless of whether an AI meets legal standards or no legal rules prohibit or exist, corporate or communal ethical standards and expectations should also be considered, both within the organization and outside with the public.

REFERENCES

1. Baquero, Juan Aristi, Roger Burkhardt, Arvind Govindarajan, and Thomas Wallace. 2020. "Derisking AI by Design: How to Build Risk Management into AI Development." *McKinsey & Company*. McKinsey & Company. August 13. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development.
2. IC-Data-Driven-Decision-Making-Checklist-Template "https://www.smartsheet.com/sites/default/files/IC-Data-Driven-Decision-Making-Checklist-Template-10545_PDF.pdf"