Responsible AI in Consumer Enterprise

A framework to help organizations operationalize ethics, privacy, and security as they apply machine learning and artificial intelligence

integrate.ai

Executive Summary

Data has become such a business-critical asset that organizations from every sector are recharacterizing themselves as data companies.

Today, there's infinite opportunity for organizations to leverage and unlock the value inherent in their data repositories. Those companies that successfully deploy artificial intelligence (AI) to derive meaningful insights from their data will emerge as the leading innovators of tomorrow. But to achieve true success, organizations must implement the necessary guardrails for responsible data use. That's because the long-term sustainability of any enterprise is predicated on trust. For data companies, the respectful and ethical treatment of data has to be a core feature of any trust model.

The concept of data ethics is still in its formative stages and requires active, informed, and multi-stakeholder discussion. In the pages that follow, we provide a high-level framework designed to help facilitate a structured conversation about the ethical considerations and broader economic and social impacts of AI data initiatives. For a more in-depth examination of this topic, see the full **Responsible AI in Consumer Enterprise paper.**

Implementing Responsible AI

AI may be the biggest and most disruptive technology advance we see in our lifetimes.

And while the field has been around for a long time, over the past five years, it has shifted dramatically thanks to faster computation, smarter algorithms, and, most importantly, an exponential growth in data.

Machine learning — the subfield of AI where software systems learn from data and experience — is having the greatest impact on the enterprise. As Amazon CEO Jeff Bezos said in his 2017 letter to shareholders,

> “
>
> Over the past decades computers have broadly automated tasks that programmers could describe with clear rules and algorithms. Modern machine learning techniques now allow us to do the same for tasks where describing the precise rules is much harder.

Of course, machine learning does more than automate existing businesses processes. It changes how businesses form and strengthen relationships with customers. Using data and machine learning, businesses can turn every interaction into an opportunity to learn what people want and value. At a macro level, machine learning can optimize margins, directing budget dollars and human resources to those customers where outreach and engagement will generate the highest return.

But there are risks. AI requires enterprises to use customer data in new ways, making appropriate data usage one of their key responsibilities. That's because people feel shocked when they learn that their sensitive information was leaked. Suspicious when they sense businesses want to manipulate their behavior. Powerless when an automated system denies them a product without any explanation for why. Trust is not a constant. It's earned over years and can be lost in an instant.

Executives are ultimately responsible for striking the right balance between business risk (both legal or reputational) and opportunity. Leaders need a clear mental model for what AI can and cannot do and a means to effectively arbitrate between business and risk stakeholders to make the right decisions. In a world where major technology advances like AI challenge existing decision-making models, this is becoming increasingly difficult.

That's where our framework comes in. It presents the privacy, security, and ethics choices businesses face when using machine learning on consumer data. It breaks things down into the various small decisions teams need to make when building a machine learning system. It's an agile approach to ethics and risk management that aligns with agile software development practices.

Importantly, our framework is neither a regulatory compliance compendium nor an exhaustive list of risk management controls. Rather, it's a tool to help businesses applying AI to think about ethics and risk contextually. It provides insights for implementation teams and high-level questions for executive leadership.

How Machine Learning Systems Work

Before you start addressing the ethics and risks of machine learning, it helps if everyone has a common understanding of what machine learning systems do and how they work. This doesn't mean that everyone needs to become a machine learning scientist and grasp the nuances of different algorithms. They just need the grounded intuitions necessary to ask good questions.

Machine learning systems create useful mappings between inputs and outputs. These mappings, called models, are mathematical functions, equations of the form $y = mx + b$, where x is an input and y is an output (admittedly, however, the equations can be much more complicated!). You could use hand-written rules to define those mappings, but rules take a lot of time to write and usually don't handle a lot of cases.

With machine learning, computer programmers no longer write and update the mappings between inputs and outputs. Computers learn these mappings from data. So, in $y = mx + b$, the computer learns what value "m" and "b" should be after seeing lots of x's and y's. When the system is presented with new inputs it hasn't seen before, it uses the mappings it's learned to make a useful guess about the corresponding output. These mappings aren't certain, and they don't always generalize perfectly to new data.

Most machine learning applications boil down to making a prediction about the future (how likely is it that this individual will become a profitable customer?) or classifying data into useful categories (is this email spam? Is this cell phone stationary or in transit?). Emerging systems that do things like schedule meetings, make phone calls, or write emails on our behalf can deliver a range of possible outputs rather than just an output with a clear right or wrong answer (like correctly saying what object is in an image) or a strict binary decision (will this customer churn or not?).

Common machine learning applications in consumer enterprise

Recommendation Systems
Compare actions of consumers to infer similar taste or suggest affinity between consumers and products based on attributes and actions

Audience Segmentation
Separate consumers into groups that look like one another in a way that is relevant for marketing or product performance

Personalization
Modify the experience of a product, marketing message, or channel to best resonate with a consumer at a scale too large for human teams to execute

Chatbots
Help customers answer questions, resolve problems, or identify the right product mix to redirect human resources to higher-value interactions that require judgment

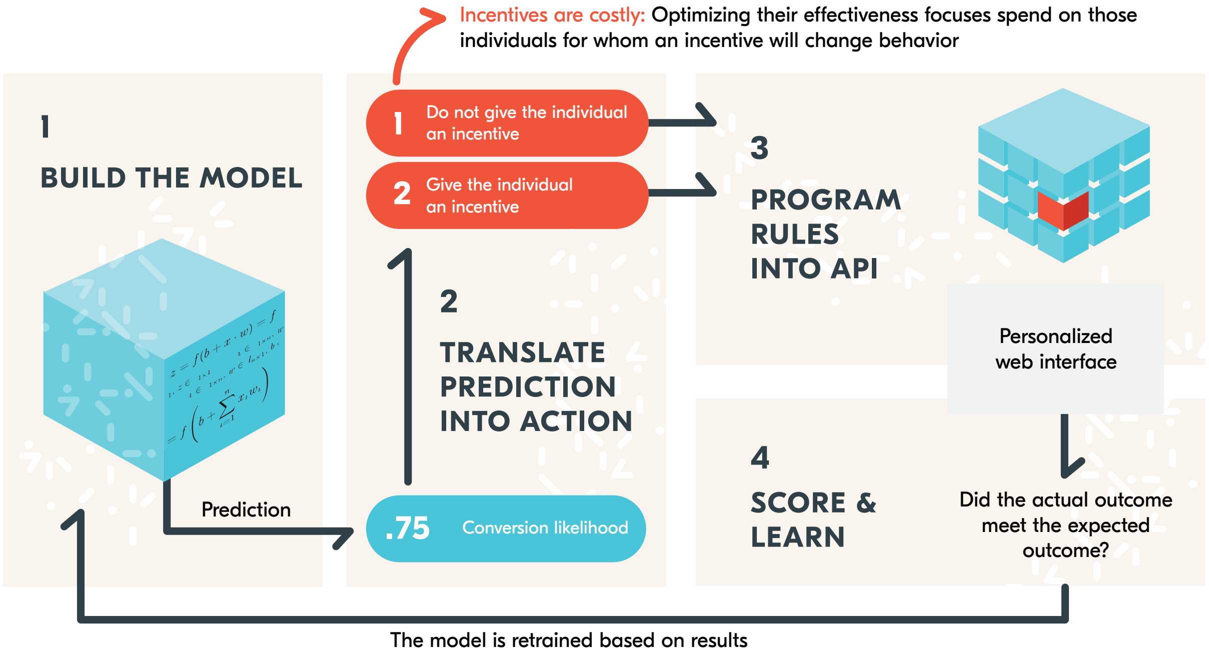
Risk Assessments
Modify offer and pricing on an insurance or banking product according to predicted risk or likelihood to default

Anomaly Detection
Identify a shift in customer behavior that could signal opportunity for upsell or risk of churn, or a shift in network or system behavior that could signal malicious activity

Anti-money Laundering and Compliance
Identify suspicious behavior or attributes and automate compliance reporting workflows using natural language generation

Data Products
Use algorithms to identify useful insights about consumer behavior that are packaged and sold to other businesses for targeted marketing

When machine learning is incorporated into a business process, businesses must design how to transform a model's output into an action. Feedback loops happen when businesses keep track of the difference between expected versus actual outcomes and use this difference to improve prediction accuracy over time.


Incentives are costly: Optimizing their effectiveness focuses spend on those individuals for whom an incentive will change behavior
1
BUILD THE MODEL
Prediction
1 Do not give the individual an incentive
2 Give the individual an incentive
2
TRANSLATE PREDICTION INTO ACTION
.75 Conversion likelihood
3
PROGRAM RULES INTO API
Personalized web interface
4
SCORE & LEARN
Did the actual outcome meet the expected outcome?
The model is retrained based on results

Ethics and risk questions arise across the machine learning system workflow. Say your system automates decisions on granting people housing loans. Does your historical data create a mapping that frequently denies loans to black people? Could a malicious attacker reverse engineer your model to access sensitive personal data? Can you explain why you denied someone a loan? If the score changes over time, can you reconstruct old mappings that have been updated and replaced? Our framework examines these questions, breaking them down according to the different tasks that go into building a machine learning system.

Guiding Principles

Our framework starts with some guiding principles. Effectively, they're the intuitions everyone in your business, including executive management, should internalize to inform risk-based thinking and ethical decisions.

Responsible AI Principles:

- In standard practice, machine learning assumes the future will look like the past. When the past is unfair or biased, machine learning will propagate these biases and enhance them through feedback loops. If you want the future to look different from the past, you need to design systems with that in mind. You can't just let the data guide you. Executive leadership should decide what future outcomes the business wants to achieve. These should include fairness, not just profit.
- The outcomes businesses want to optimize for are often hard to measure or occur far in the future (e.g., customer lifetime value). Businesses therefore resort to easier-to-measure proxies that stand in for their desired outcomes. Be clear about what these proxies do and don't optimize. You may learn they exacerbate bias or have downstream consequences that conflict with your values or goals.
- All of your customers are individuals. Representing them as data points necessarily transforms them from people into abstractions. When you deal with abstractions and groupings, you run the risk of treating humans unethically.
- Beware of correlations that mask sensitive data behind benign proxies. For example, postal code/ZIP code is often a proxy for ethnic background. If your machine learning system uses location to make decisions, you may end up treating various ethnic groups differently.
- Context is key for explainability and transparency. Systems that decide who gets a credit card or loan require more scrutiny than systems that personalize marketing offers. Business and risk teams should assess context and communicate required constraints to technology teams.
- Privacy is not just about personal data, notices, or consent forms, or a set of controls to minimize data use. It's about appropriate data flows that conform to social norms and expectations. Map these flows and ask if people would be surprised to learn how their data has been used.
- Accountability is a marathon, not a sprint. Once in production, machine learning systems often make errors on populations that are less well represented in training data. Develop a plan to catch and fix these errors. "Govern the optimizations. Patrol the results." [1]
- There's no silver bullet to responsible AI. It takes critical thinking and teamwork. Step outside the walls of your organization and ask communities and customers what matters to them.

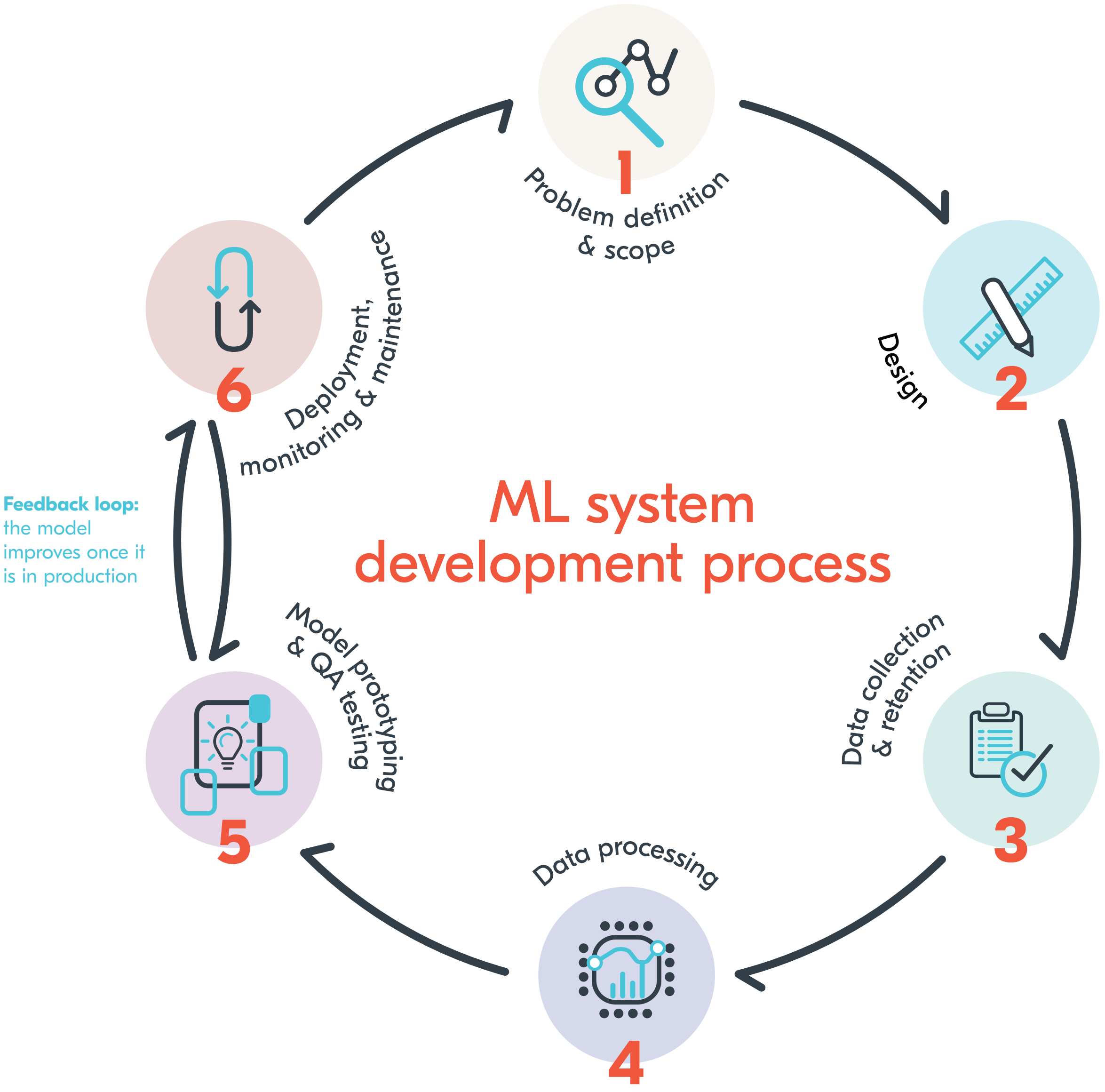
[1] Weinberger, David. "Optimization over Explanation: Maximizing the benefits of machine learning without sacrificing its intelligence." https://medium.com/berkman-klein-center/optimization-over-explanation-41ecb135763d

The Responsible AI Framework

Framework Summary

The responsible AI framework breaks down the steps used to build a machine learning system and highlights privacy, security, and ethics questions teams should consider at each step. Inspired by Privacy by Design, it's characterized by proactive rather than reactive measures to privacy and ethics, and embeds critical thinking and controls into the design and architecture of machine learning systems.

View this as an agile process with multiple iterations and decision points, not a waterfall process that plans everything in advance. You may discover you want to cut a project because you lack sufficient training data, require greater certainty to foster adoption, or have identified ethical concerns. Learn that quickly and free up resources to do something else. Remember that cross-functional teams should participate in most meetings (or at least have regular check-ins) throughout the process, in particular during the scoping phase.


1
Problem definition & scope
2
Design
3
Data collection & retention
4
Data processing
5
Model prototyping & QA testing
6
Deployment, monitoring & maintenance
ML system development process
Feedback loop: the model improves once it is in production

Step	Jobs to Be Done	Risk & Ethics Questions
1 Problem Definition & Scope	• Map current business process • Identify where machine learning system adds value or alters process • Define inputs, outputs, and what you are optimizing for • Measure baseline performance and expected lift	• How could your system negatively impact individuals? Who is most vulnerable and why? • How much error in predictions can your business accept for this use case? • Will you need to explain which input factors had the greatest influence on outputs? • Do you need personally identifiable information (PII) or can you provide group-level insights?
2 Design	• Analyze a user flow to understand how data is collected and where users hesitate on what to input • Decide whether this will be a fully-automated or human-in-the-loop system • Interview users and apply human-centric principles to understand their experience • Design how model outputs will translate into insight or action for internal users/external consumers	• How can you make data collection procedures transparent to consumers? • Will the formats you use to collect data alienate anyone? • How will you enable end users to control use of their data? • Should you make it clear to users when they engage with a system and not a human?
3 Data Collection & Retention	• Conduct a data census to identify what data you have and what data you need • Procure second- and third-party data sets • Align machine learning training needs with data retention schedule	• How will you manage the provenance of third-party data? • Who are the underrepresented minorities in your data set? • If a vendor processes your data, have you ensured it has appropriate security controls?
4 Data Processing	• Format and process the data to prepare it for machine learning algorithms • Pair subject matter experts with scientists to help understand data and features that matter for predictions	• Have you de-identified your data and taken measures to reduce the probability of re-identification? • Will socially sensitive features like gender or ethnic background influence outputs? • Are seemingly harmless features like location hiding proxies for socially sensitive features?
5 Model Prototyping & QA Testing	• Experiment with various algorithms to verify the problem can be solved and select the approach that performs best • Test model performance on reserved test data set to verify functionality beyond training set	• Does your use case require a more interpretable algorithm? • Should you be optimizing for a different outcome than accuracy to make your outcomes fairer? • Is it possible that a malicious actor has compromised training data and created misleading results?
6 Deployment, Monitoring & Maintenance	• Integrate model outputs into business process • Capture data on outcomes and provide feedback back to the system • Define model retraining frequency (batch or real-time) and how scientists evaluate future model changes • Monitor system for failures or bugs and update code regularly • Measure and report on results	• Can a malicious actor infer information about individuals from your system? • Are you able to identify anomalous activity on your system that might indicate a security breach? • Do you have a plan to monitor for poor performance on individuals or subgroups? • Do you have a plan to log and store historical predictions if a consumer requests access in the future? • Have you documented model retraining cycles and can you confirm that a subject's data has been removed from models?

A Word About Consumer Privacy

One of the main data privacy challenges you'll face relates to protecting consumer privacy beyond personally identifiable information (PII). Focusing narrowly on PII, including fields in databases like first and last names, social insurance numbers, or email addresses, isn't sufficient to guarantee privacy. You have to expand risk to protect the possibility of a breach even when a data set has been scrubbed of PII. The core ethics issues relate to deciding what types of inferred features or profiles your organization feels are appropriate and identifying tightly correlated features in data sets that can hide discriminatory treatment.

Identification risks grow when data is released publicly or third-party data is used to augment first-party data. That's because third-party data might fill in gaps, leading to an increased ability to reverse engineer an individual from a group. As such, enterprises should have consistent practices for sharing data with third parties. If two startups have two different views on people, each of which are private, but collaborate with one another, they'll have keys to unlock identity.

This is another area where the current best practice is to think critically and apply a risk-based approach. The Information and Privacy Commissioner of Ontario recommends taking a risk-based approach to de-identify data. The downside to de-identification is that it is not foolproof. There will be residual re-identification risk, which is why tolerance needs to be assessed and governed against.

An alternative technique that provides theoretical guarantees is differential privacy. Differential privacy modifies the data set in such a way that statistical features that matter for a model are preserved, but it's impossible to tell the difference between a distribution that does and does not contain an individual. Protections can be added at various points in the machine learning pipeline, with tradeoffs of model performance and privacy guarantees. As we saw above, the more questions you ask about an aggregate, the closer you get to an individual. Most differential privacy algorithms have a privacy budget, or number of queries they can support before privacy guarantees lessen. Product management leaders need to consider these tradeoffs during implementation.

At this time, differential privacy is in production in companies like Google, Facebook, Apple, and Uber, but has yet to become de facto best practice in startups or the enterprise. It is still relatively new and difficult to implement effectively. Other privacy techniques include one-way hash functions, which make a cryptographic mapping of input data that cannot be reversed, and masking, which removes variables or replaces them with pseudonymous or encrypted information.

DEGREES OF IDENTIFIABILITY

Information containing direct and indirect identifiers

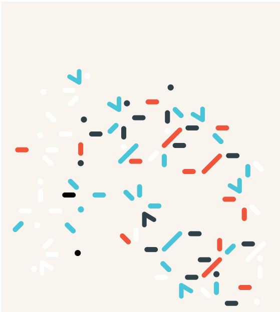
PSEUDONYMOUS DATA

Information from which direct identifiers have been eliminatied or transformed, but indirect identifiers remain intact

DE-IDENTIFIED DATA

Direct and known indirect indentifiers have been removed or manipulated to break the linkage to real world identities

ANONYMOUS DATA



Direct and indirect identifiers have been removed or manipulated together with mathematical and technical guarantees to prevent re-identification

Conclusion

AI offers immense potential to businesses and society. Our ability to process data at scale and use machine learning to learn from that data has shifted the balance between enterprises and consumers. Engagement used to be unidirectional and one size fits all. Thanks to machine learning, relationships between consumers and businesses are becoming bidirectional. The actions consumers take provide a window into who they are, what they want, and what they value. As in interpersonal relationships, businesses can listen to this feedback and use it to provide more relevant products, services, and experiences. The impact AI will have on society starts with the mindset we adopt to imagine its potential and the tasks we choose to apply it to.

We hope this framework will empower you to apply machine learning and innovate. We hope it will spark ideas and spur conversations between teams in your company or with new communities outside your company. AI is here to stay. We can use it for good. We simply have to ask the right questions and activate our ethical prerogative to express our values in the systems we build.

To learn more, download the full **Responsible AI in Consumer Enterprise paper.**

ABOUT INTEGRATE.AI

Integrate.ai helps consumer enterprises develop more meaningful relationships with customers using artificial intelligence.

Our cloud-based machine learning platform integrates directly into systems of engagement to focus marketing and customer service resources where they will have the greatest impact. Our trusted signals exchange enables enterprises to share insights and learnings about consumer behaviour without leaking competitive IP or compromising user privacy. We enable partnerships, collaboration, and a platform for businesses to innovate ethically as we shape our collective future.

integrate.ai

Toronto / San Francisco
info@integrate.ai
twitter LINKEDIN