Data Literacy for Responsible AI

About the authors

Ted Kwartler – VP, Trusted AI, DataRobot

Ted Kwartler is Vice President, Trusted AI at DataRobot. In that role, he sets product strategy for explainable and ethical uses of data technology. Through his current position and his previous employment in a major insurance company and the IT industry, he brings unique insights and experience with data ethics and transparency. Ted also teaches at the Harvard University Extension School and serves as an advisor to the Bureau of Economic Affairs' Advisory Committee on Data for Evidence Building.

Haniyeh Mahmoudian, PhD – Global AI Ethicist, DataRobot

Haniyeh is a Global AI Ethicist at DataRobot's Trusted AI team. Her research focuses on bias, privacy, robustness and stability, and ethics in AI and Machine Learning. She has a demonstrated history of implementing ML and AI in a variety of industries and initiated the incorporation of bias and fairness features into DataRobot products. She is a thought leader in the area of AI bias and ethical AI. Haniyeh holds a PhD in Astronomy and Astrophysics from the Rheinische Friedrich-Wilhelms-Universität Bonn.

Sarah Khatry – Managing Director, AI Ethics, DataRobot

Sarah Khatry is a managing director on the Trusted AI team at DataRobot. Her work focuses on the ethical use of AI, particularly the creation of tools, frameworks, and approaches to support responsible but pragmatic AI stewardship, and the advancement of thought leadership and education on AI ethics.

Adel Nehme – Data Science Evangelist, DataCamp

Adel is a Data Science educator, speaker, and Evangelist at DataCamp where he has released various courses and live training on data analysis, machine learning, and data engineering. He is passionate about spreading data skills and data literacy throughout organizations and the intersection of technology and society.

Introduction

Over the past decade, Artificial Intelligence (AI) has rocketed out of academia and into the public consciousness, fueled by an increase in the accessibility of computational resources and an explosion of data. Algorithms proposed and envisioned back in the 1950s-80s, have become realized, and served as building blocks for ever-accelerating innovation. Predictive analytics is making our businesses smarter and more efficient. Machine vision is bringing AI into direct interaction with our physical world. Natural language processing (NLP) is changing how we interact with computer systems, on the web, on our phones, and in our homes, with new smart speakers.

Artificial intelligence (AI) is maturing as a technology, and with that comes the expectation that it should be used responsibly and ethically. For example, a recent Capgemini report states that 70% of customers expect organizations to provide AI interactions and products that are transparent and fair.[1] Moreover, increasing regulatory frameworks for AI governance mean that organizations will be more and more liable for providing transparency to consumers around AI-driven decisions.[2]

AI is no longer in a safe sandbox of pure research and hypotheticals but integral to our day-to-day lives, in arenas both minor and critical. In high-sensitivity applications, like finance or healthcare, the risks are most apparent, and the standards accordingly high. But even in seemingly more everyday areas, like advertising, AI can have unfortunate repercussions, as the scale and speed of impact enabled by AI can not only reproduce but also amplify to a systematic level historical issues like bias and discrimination. This means every AI use case that in some way touches human lives requires careful forethought, consideration, and strategic planning.

Algorithmic bias has been uncovered in high-profile examples of both supervised and unsupervised machine learning. Machine learning is a subfield of AI in which statistical algorithms discover and teach themselves patterns in a dataset. In supervised machine learning, the goal is to predict certain outcomes from those patterns.

In 2016, ProPublica investigated the COMPAS algorithm: a proprietary algorithm used to inform sentencing. COMPAS aimed to predict the likelihood of an individual to reoffend. ProPublica uncovered that Black defendants were far more likely to receive a false positive recommendation, that they were at a high risk of reoffending when in reality they would not, while white defendants were more likely to receive a false negative. False positives for Black defendants could unfairly lead to more severe sentencing and diminished opportunities for parole. In contrast, false negatives would enable white defendants who were going to commit a crime again a more lenient sentence[3].

[1] "AI and the ethical conundrum" Report, Accessed April 14, 2021.

[2] "Using Artificial Intelligence and Algorithms", Accessed April 14, 2021.

[3] "How We Analyzed the COMPAS Recidivism Algorithm" Accessed April 2, 2021.

In a supervised example, algorithmic bias may be traced back to historical bias and discrimination, informing the outcomes used to train a model. But bias is also becoming apparent as an issue in unsupervised examples, where no human labeling can be attributed as the cause. Instead, the makeup and distribution of the underlying data itself reflect societal dynamics and stereotypes that may be damaging. For example, GPT-3 was unveiled by OpenAI in 2020, representing the third generation of OpenAI's implementation of a new type of NLP algorithm called a Language Transformer. Language Transformers essentially operate as highly sophisticated autocomplete software for a variety of language tasks, requiring a massive amount of unlabeled text to learn from. GPT-3 was trained on gigabytes of text data mined from the internet, resulting in a model with over 175 billion parameters.

This extreme complexity seems to have paid off. GPT-3 can generate incredibly convincing responses to few-sentence prompts in a variety of styles, from examples of news bulletins you can imagine encountering the next time you open the paper, to the openings of noir novels[4]. However, this isn't all that GPT-3 has learned. Researchers from Stanford and McMaster universities discovered that even when given as short a prompt as "Two Muslims walked into a…", GPT-3 with high frequency would generate a response associating "Muslims" with violence, at far higher rates than for any other religious denomination[5]. In both of these cases, the undesirable behavior of the AI system can be attributed to us: the text data that GPT-3 learned from, and the historical outcomes that trained COMPAS.

AI is an opportunity to do better, to deprogram our society of bias, and directly encode the ethics and values we'd like to see reflected in AI-driven processes. The path there requires a nuanced understanding of algorithmic bias, how it is produced, and how it can be mitigated, and also a more comprehensive framework for accountability and governance of AI systems in general, encompassing all risks including bias. Every use case is unique, and context is pivotal to extrapolating how an AI will interact with a process and impact different groups of people. It is never just math.

Trust provides us with a powerful paradigm to approach the pragmatic application of AI ethics. Trust is not an intrinsic quality. Instead, it describes a relationship: the human-AI interaction. Trust must be earned and means different things to different people.

AI requires a multistakeholder approach, recognizing that there is a wide range of personas an AI system touches or impacts. That includes everyone from the team in charge of creating the AI—product managers, data scientists, data engineers—to the team in charge of operationalizing the model—machine learning engineers, IT, software developers—and finally who uses and who is impacted by the decisions made—internal and external consumers, including indirect stakeholders.

Depending on the nature of the use case, risk, compliance, and security personnel—or even the general public—may also have a stake in the final AI product. This range of stakeholders will have differing degrees of technical literacy. For each, trust will require something different, and they will be seeking unique trust signals to establish and verify the system is working in a way that protects their interests.

[4] "OpenAI's new language generator GPT-3 is shockingly good—and completely mindless" Accessed April 2, 2021.

5 Abid, Farooqi, and Zhou. "Persistent Anti-Muslim Bias in Large Language Models". arXiv Preprint, 2021.

Not all stakeholders in AI need the level of understanding of a data scientist, but a baseline data literacy is essential to being able to grasp the meaning and impact of an AI system. At this point in time, with the saturation of AI through the systems that surround us, we are all AI stakeholders, and overall deeper fluency in data is key to making sure those systems are worthy of our trust.

In this white paper, the following key elements of a trustworthy dynamic between humans and AI will be explored:

1. Algorithmic bias, how to define and identify bias in AI, what are the sources of bias, and how to mitigate it.
2. Accountability and governance frameworks, to develop a comprehensive understanding of the risks in an AI use case, and to put into action guardrails to monitor and reduce those risks.
3. The cultivation of data literacy, to make available to stakeholders of all technical levels a shared understanding of an AI and its data, for more informed decision-making.

Algorithmic bias

When it comes to algorithmic bias, we have all seen by now the many headline stories of machine learning and AI gone wrong: a hiring algorithm that disproportionately favored men[6], online advertising that seemed to reproduce the phenomenon of redlining on a digital map[7], or even a widely-used healthcare algorithm that exhibited significant racial bias.[8]

What is less publicly known and understood is how to define algorithmic bias. This is in fact a bit of a trick question, as there are dozens upon dozens of different mathematical metrics that can be used to define bias. In general, algorithmic bias refers to the observation that an algorithm is treating groups in your data differently. Those groups, when we are concerned about societal bias, are identified by what are referred to as protected or sensitive characteristics—like race, gender, age, pregnancy or veteran status. What is 'fair' is informed most prominently by the context of a use case. When identifying what bias metric is most appropriate, it is essential to interrogate the use case and determine if there is a particular group most vulnerable to harmful impacts.

While there are somewhere beyond 70 different metrics proposed by research and academia, a handful can cover the majority of algorithmic bias and fairness concerns; the rest will be useful in only hyper-specialized cases. All of these metrics can be broken up into two main categories: fairness by representation and fairness by error.

[6] "Amazon scraps secret AI recruiting tool that showed bias ...." Accessed April 6, 2020.
[7] "Facebook charged with housing discrimination in targeted ads ...." Accessed April 6, 2020.
[8] "Dissecting racial bias in an algorithm used to manage the ...." Accessed April 6, 2020.

Fairness by representation or error

Fairness by representation focuses directly on what outcomes the model predicts to evaluate if there are different likelihoods of receiving the more favorable outcome by each group. In fairness by error, the quality of model performance and accuracy is compared across groups; are some groups disproportionately affected by certain kinds of error?

To illustrate the difference between these types of metrics, let us first focus on a fairness metric called statistical parity. This test, as a test of representation, calculates the likelihood of a group being predicted to have a favorable outcome and compares each group to the group with the highest likelihood. A common area to use this metric is in hiring to ensure diversity and equal representation in the workforce. In this example, say men have a 50% likelihood of being called back to an interview after submitting a CV and cover letter, which is evaluated by an AI system. If women have a lesser chance, this may indicate that gendered bias is creeping in.

In another example, consider a use case in which AI is used to identify patients that need to be enrolled in a treatment program. The ground truth of what percent of individuals from different groups need this treatment may be different due to biological factors. Thus, instead of fairness by representation, fairness by error will be most appropriate. If the greatest potential harm is denial of access to this treatment to someone who needs it, we will want to compare the false negative rates of different groups and ensure that no group is being denied necessary treatment at a disproportionate rate. A test exists referred to as false negative parity, and that will guide our assessment of harmful bias in this use case.

Where bias comes from

Now that we have discussed the two major types of fairness metrics and have given examples of the contextual considerations that inform the selection of a metric, it is essential to understand the source of bias. Below are some of the ways bias can manifest from data.

- **Skewed dataset:** Lack of representation in the data can affect an AI's ability to learn from diverse sets of examples, which can result in biased model performance.

- **Tainted examples:** Unreliable labels or historical bias in the data have a direct impact on AI's discriminatory behavior.

- **Limited features:** Feature collection for certain groups may not be informative or reliable, which can occur under bad data collection practices. Similar to a skewed dataset, this will impair an AI's ability to predict accurately for those groups.

- **Sample size:** Small datasets limit the ability of the AI's effective learning process and can result in bias.

- **Proxy features:** Features can indirectly leak information about the protected attributes, even in cases when that protected feature has been removed. Zipcodes, sports activities, and university attended can be used by the model to indirectly infer race or gender; if the examples are then tainted by historical bias, even without direct access to that protected feature, the model will learn the pattern of discrimination from proxies.

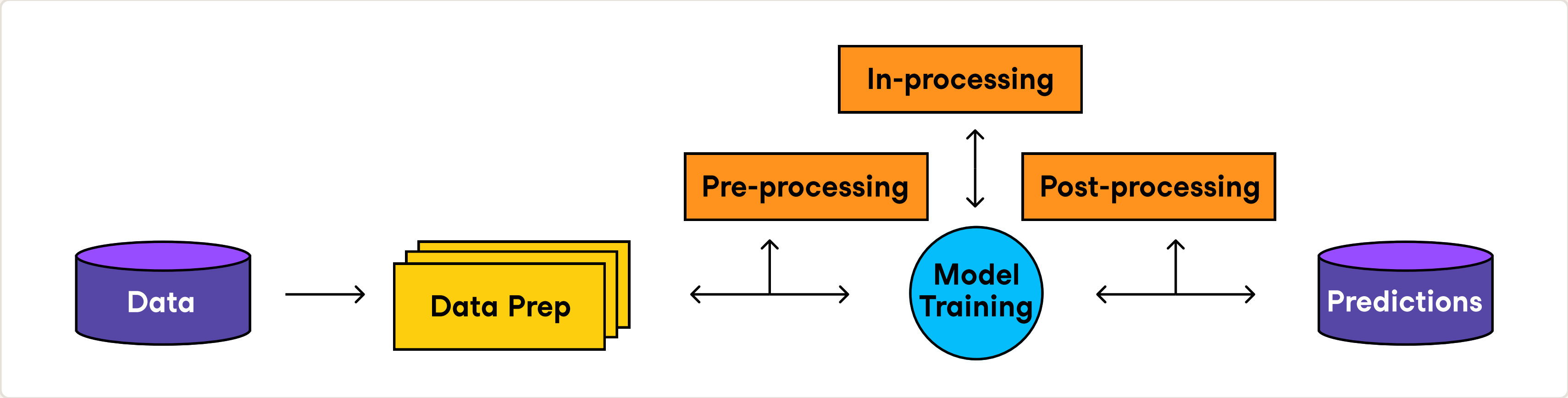
If a metric reveals bias, the first step will be an investigation of the data. One method is to compare the distribution of different features across all the protected groups in the dataset, and note any observed disparities as potential contributors to bias in predicted outcomes.

To identify the proxy features, an effective strategy is the replacement of the predictive target in the data with the desired protected attribute. Building a model predicting the protected attribute will surface the presence of proxies in the dataset. For a comprehensive look at fairness by error, we can also assess the raw accuracy values for each group.

That investigation may offer paths to mitigating the bias observed, through the elimination of proxies, or, if the data were a sample, to a curated training dataset with improved representation of groups across outcomes and features.

Mitigation techniques

However, there are also other algorithmic methods of mitigation available. Mitigation strategies can occur at different points of the machine learning pipeline: pre-processing, in-processing, and post-processing.



Pre-processing techniques aim at reducing bias in the original data before you train your model. Here are some examples:

- Data Preprocessing Techniques for Classification without Discrimination[9]
- Optimized Pre-Processing for Discrimination Prevention[10]
- Certifying and Removing Disparate Impact[11]

These techniques can range from simple sampling the data to have more balanced and curated data or giving weights to rows representing unprivileged groups9 to more complex data transformation methods to reduce the correlation between features, target and protected attributes, and predictability of protected or sensitive attributes[10]-[11].

[9] Kamiran, Faisal and Calders, Toon. Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1):1–33, 2012

[10] Calmon, F., Wei, D., Vinzamuri, B., Ramamurthy, K. N., and Varshney, K. R. Optimized pre-processing for discrimination prevention. In Advances in Neural Information Processing Systems 30, 2017.

[11] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian. Certifying and removing disparate impact. In KDD, 2015.

In-processing methods tackle bias during model training.

- [Classification with Fairness Constraints: A Meta-Algorithm with Provable Guarantees][12]
- [Fairness-Aware Classifier with Prejudice Remover Regularizer][13]

In-processing techniques force models to learn less biased patterns. Classification with fairness constraints[12], enforces fairness constraints in the optimization process, and fairness aware classifiers[13] use regularization as the means to reduce bias.

Post-processing algorithms reduce bias by modifying model predictions.

- On Fairness and Calibration[14] and Equality of Opportunity in Supervised Learning[15] change model predictions for privileged and unprivileged groups to reduce bias in model predictions based on specific fairness metrics.
- Decision Theory for Discrimination-aware Classification[16] assumes that most discrimination occurs close to the decision boundary, in which the mode is uncertain about the prediction, and thus exploits the low confidence region of a classifier for discrimination reduction.

Presume you've used one of these techniques—or attempted a variety of them—and now have a cohort of models to evaluate. The last step in the modeling process is a critical one: choosing the final model to deploy and use. This will require a multidimensional analysis of overall performance and accuracy, and the potential impact of predictions on the different groups in your data. Data science practitioners are very used to weighing the trade-off between model performance and speed and scalability. Now, with rising awareness of algorithmic fairness, we can consider the trade-off as well between a model's accuracy and its level of discrimination.

No model is likely going to be perfect, in either respect. Thinking through correcting bias by representation, you may be forcing the model to predict outcomes that did not happen historically, due to discrimination; that alone will be a source of a perceived loss in accuracy. Ultimately, this decision is best guided by a recognition of what tolerance is permissible on each axis of evaluation, accuracy, and bias. This requires a multistakeholder discussion and is best pursued with full transparency into the potential risk and impact of any modeling choice.

In the next section, we will discuss governance best practices that lay the foundation for navigating these complex decisions.

[12] L. Elisa Celis, Lingxiao Huang, Vijay Keswani, and Nisheeth K. Vishnoi. Classification with fairness constraints: A meta-algorithm with provable guarantees. In Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, 2019.

[13] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma. Fairness-aware classifier with prejudice remover regularizer. Machine Learning and Knowledge Discovery in Databases, pages 35—50, 2012.

[14] Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., and Weinberger, K. Q. (2017). On fairness and calibration. In Advances in Neural Information Processing Systems, pages 5680–5689.

[15] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In Advances in Neural Information Processing Systems, 2016.

[16] Kamiran, F., Karim, A., Zhang, X. 2012. Decision theory for discrimination-aware classification. In Proceedings of the IEEE International Conference on Data Mining (ICDM 2012), Zaki M. J., Siebes A., Yu J. X., Goethals B., Webb G. I. & Wu X. (eds). IEEE Computer Society, 924–929

Governance & Accountability

From a strategic standpoint, AI systems represent a novel and valuable proposition, but also can increase organizational risk. Organizational risks can be measured directly as impacts to revenue, but there is also the reputational risk to the brand incurred by an algorithm that is perceived as discriminatory and harmful to vulnerable groups to consider.

However, AI is not starting from scratch. Risk management as a field has been studied and iterated upon since World War II. Technologies like aviation or nuclear power would not exist without robust frameworks to evaluate and mitigate risks. The difference for AI is fueled by the ability of the technology to rapidly have impacts of great scope, and unique dimensions of its implementation. Following in the footsteps of other innovative technologies, governance utilizing risk management is needed comprehensively in AI.

The first step is to classify the type of AI decision associated with a specific use case. Next, a well-defined impact assessment can assist in the identification of stakeholders and the potential for risk for each of them due to an AI use case. Guided by the impact assessment, the potentially harmful AI behaviors can be understood, tested for, and sometimes mitigated. In conclusion, rigorous AI behavioral tests mitigating risk for stakeholders represents proper AI governance.

At DataRobot, that first step of AI decision type classification is enabled through a framework. AI decisions can be understood as existing on a spectrum of risk ranging from low, only small monetary losses, to high, with possibly significant monetary or life-impacting loss. The following table illustrates the DataRobot AI Risk Framework. An AI decision of low risk does not need as robust governance as one of medium or high risk. A high-risk decision should likely never be made in an automated fashion by an AI system alone; instead, it should be an augmented decision shared between human and machine intelligence. For example, this would be the optimal procedure for integrating an AI system into a medical diagnosis. It is not the AI system that makes the final decision, but only a recommendation, with as much transparency and interpretability functionality as possible, to a physician who is evaluating the patient as a whole.

	Type I - Low	Type II - Medium	Type III - High
Risk Size	Loss <$100K	$100k < Loss < $1M, or Injury to Human Livelihood	Loss > $1M, or Death
Example	Probability of an ad-click	Probability of Mortgage Default	Medical Imaging Machine Vision
Human Role in Governance	Construction, Maintenance & Monitoring	Type I & Risk Assessment & Mitigation	Type I, II & Final Augmented Decision Outcome “Human over the loop”

DataRobot’s AI Risk Framework. Thresholds need to be adjusted according to organizational definitions.

Next, on the basis of the recognition of decision type, an organization should apply stakeholder theory to define all parties to the system, and conduct an impact assessment. Although currently not a regulated mandate, creating an impact assessment is a worthwhile proactive planning exercise when developing AI systems. Consider manufacturing. Currently, when a factory is built, environmental impact assessments weigh the benefit of the facility—such as job creation and increased economic output—against the factory's impact on the surrounding wildlife, watershed, and population. This informs whether the project should even be pursued from the outset, and what risks need to be intentionally mitigated throughout the project lifecycle. Similarly, an AI impact assessment helps justify the technology's use and identifies stakeholder risks.

Once the organization has a handle on the potential loss associated with an AI system, and all stakeholders and associated risks are defined in an impact assessment, each hazard needs to be understood by its severity and likelihood in a risk matrix. Practical unit tests can then be applied with rigor, corresponding to the loss size identified in the first framework.

AI governance best practice is based on both model and organizational behaviors requiring:

1. **Comprehensive and transparent checklists for internal AI stakeholders**

   An open, transparent implementation process encourages diversity of thought.

2. **Automated and user-guided testing**

   Defined testing procedures ensure consistent, repeatable, and, most importantly, auditable data science methods, implementations, and impacts. It is important to avoid bespoke product releases; models should be monitored and updated at a regular cadence.

3. **Business rule testing**

   Often AI products do not fit into existing operational processes. Models should not only adhere to data science and IT system requirements but also satisfy business standard operating procedures. An AI should not operate outside any procedural bounds you'd expect an employee to adhere to.

4. **Detailed test reports**

   Due to the interdisciplinary nature of AI products, testing results should be shared broadly and at a regular cadence so that system stability and ethical ends are achieved.

5. **Directly responsible individuals/Group assertions**

   - Data Science Sign-Offs—Do the AI creators assert the model's accuracy, validity, and proper model training steps were undertaken?
   - Legal, model risk management, and IT team Sign-Off—Do all AI facilitators declare the system is robust, fault-tolerant, and adheres to regulatory restrictions?
   - Business SME Sign-Off—Does the business owner accept the risks of AI in the operation or procedure, and can they quantify the value?

This is the foundation of proper AI governance. Ultimately, accountability rests with us: the human creators, facilitators, and operators of AI. Arming all responsible parties with the needed understanding and motivation to govern an AI system is also integral to handling anything that may go wrong in the future.

Responsible AI relies on data literacy

Scaling responsible and trusted AI requires organizations to develop a baseline data literacy to ensure a common understanding of how AI projects are scoped, deployed, and governed, alongside their impact and projected risk.

AI governance efforts necessitate multi-stakeholder engagement, the adoption of a common framework, and the sign-off of various personas stemming from different backgrounds, roles, and levels of technical proficiency. As such, a common understanding of AI and data concepts is needed for these efforts to scale. Moreover, a baseline data literacy is desired for individuals interacting with AI systems. This is especially relevant for high-risk AI use-cases that would require human intervention in the final decision output.

Data literacy is a key component of developing responsible AI, as it ensures diverse perspectives are baked into AI governance efforts, and that AI systems in production achieve better and consistent outcomes[17].

How data literacy promotes trust

Data literacy can be defined as the ability to critically understand data science and AI applications, distinguish between various data roles, communicate insights from data, and derive data-driven decisions. Expanding on this definition, data literacy allows for non-technical stakeholders to become conversational with data and AI experts, and to understand the limitations of AI systems. More importantly, it promotes a two-way conversation between subject matter experts and AI experts that allows non-technical stakeholders to inject their domain expertise into the problem set-up, scoping, and implementation of AI projects. This empowers subject matter experts to guide the development and governance of AI systems that maximize value for end-users, and minimize potential harm perpetuated by these systems.

Cultivating data literacy enables the adoption of a common data language throughout the organization that promotes mutual understanding between diverse sets of stakeholders. Moreover, data literacy empowers sectoral regulators to provide guidance around the implementation of AI systems in a particular industry[18].

[17] "Responsible AI: A Framework for Building Trust in Your AI Solutions", Accessed April 15, 2021

[18] "On Artificial Intelligence - A European approach to excellence and trust", Accessed April 15, 2021

While data literacy includes basic data visualization skills, the ability to make data-informed decisions, and the ability to communicate and reason with data, in the context of scaling responsible AI, the most important building blocks of data literacy can be described as the following:

- **The data science and machine learning workflow:** Understanding the data science and machine learning workflow and the steps needed to create predictions out of raw data ensures that stakeholders understand how AI projects are implemented. This simplifies the creation of an open and transparent implementation checklist that is agreed and iterated upon by the organization.

- **The distinction between various data roles:** Understanding the distinctions between various data roles (e.g, data engineers, data scientists, machine learning engineers, etc...) and their contribution towards building an AI system facilitates collaboration between technical and non-technical experts and ensures mutual understanding of accountability for different elements of an AI project.

- **The flow of data through an organization:** Understanding how data flows through an organization from raw data extraction, to transformation, to loading into an AI system to consume guarantees organizations are aligned on potential bias risks with data collection. Moreover, this promotes a mutual understanding of the risks associated with data degradation[19], and the harmful impacts it may have on machine learning systems in production.

- **The distinction between various types of AI systems:** Grasping the distinction between various types of AI technologies will empower stakeholders to be part of evaluating which type of model is suitable for deployment. A great starting point would be understanding the difference between rule-based AI systems, machine learning, and deep learning. More importantly, grasping the distinction in their level of explainability will be paramount for organizations to understand as they scale and operationalize AI systems. This will empower subject matter experts to have an accurate understanding of the risks associated with a particular AI system in production.

- **Evaluation metrics for machine learning models:** Depending on the use-case and the AI system in place, there could be varying evaluation metrics that optimize for different outcomes (e.g., accuracy, precision, recall, etc...). Understanding what these metrics optimize for and how they intersect with the definitions of fairness outlined earlier will equip stakeholders with the language to better qualify risk associated with an AI system.

As every organization becomes a data and AI organization, raising data literacy will not only ensure organizations have the skills to remain competitive but will allow diverse sets of stakeholders to take ownership and accountability of their AI governance charter. This means subject matter experts will constructively contribute to comprehensive implementation checklists, define business rules for AI systems in production, critically evaluate testing reports for AI systems, assess the risks associated with an AI system, and more.

[19] "How can you Prevent ML Models from Degrading?", Accessed April 16, 2021

Upskilling is the only way to close the data literacy skills gap

As organizations grapple with a shortage in data talent and skills, training stakeholders of different backgrounds is the only way to close the data literacy skills gap. For example, Bosch launched an AI training program for 20,000 employees aimed at both software engineers and managers[20], for them to be able to understand the use-cases of AI and how to leverage it responsibly. The European Union recently unveiled their European Skills Agenda[21], allocating €1.1 trillion for upskilling and reskilling initiatives, key amongst them is AI skills for regulators.

In summary, as AI and data are rapidly evolving fields, continuous learning will be paramount for stakeholders to understand what is possible and responsible to achieve with AI. Ultimately, data literacy is the currency that enables the flow of sharp and critical discourse around the deployment of AI systems, scalable AI governance efforts, and increased trust in AI.

Close the data literacy skills gap with
DataCamp's data literacy fundamentals skill track

[20] "Bosh Launches AI Training Program for Developers and Managers", Accessed April 16, 2021

[21] "European Skills Agenda For Sustainable Competitiveness, Social Fairness, and Resilience", Accessed April 16, 2021