AMIT GANDHI: Hi. My name is Amit Gandhi. And I'm a graduate researcher at MIT. Welcome to the series on exploring fairness and machine learning for international development. In this module, we will cover the appropriate usage framework developed by the US Agency for International Development.

The Center for Digital Development at USAID has been studying the appropriate use of machine learning and developing country contexts. Among other activities, this research is involved engaging stakeholders, conducting case studies, and developing and publishing and appropriate use framework. The work done by the MIT CITE team builds on certain aspects of this report, which can be found in the linked materials.

In this section, we will be presenting some characteristics for the appropriate application of machine learning. Please keep these characteristics in mind as you think about the projects you were working with. They are intended to help practitioners think through the ethical and appropriate use of machine learning in international development.

The first criterion is relevance. Is the use of machine learning in this context solving an appropriate problem? As machine learning becomes more of a trend, we are seeing more and more organizations seeking to apply it to their work in an effort to distinguish themselves from their competitors or to increase their appeal to funders. Many of these organizations may try to implement prepackaged or off the shelf machine learning solutions without understanding if it is the right tool for the problem they are trying to solve.

For an example of relevance, let us consider the tracking of human assistance delivery. Organizations may want to create systems to monitor people in refugee camps and make sure they're only getting the food or other supplies that they're entitled to.

However, the major losses often happen further upstream, such as people diverting whole trucks full of supplies, not individuals taking two bags of rice instead of one. An AI solution to keep a closer eye on individual aid recipients would fail the

relevancy test because it is not addressing a major issue in the system.

The second criterion is representativeness. Is the data used to train the machine learning models appropriately selected? In order to evaluate representativeness, the organization should consider if the machine learning model uses data representative to the context in which it will be deployed and which strategies are important for ensuring models can be trained with appropriate data.

As an example, consider a startup medical diagnostics company that is trying to build a remote diagnostic tool for the West African population. High quality coded data sets from West Africa may not be available. So the startup uses a European data set to train their models. Some diagnoses may be accurate, but disease differences between Europe and West Africa may cause misdiagnoses for individuals, putting them at health risk.

Now consider if the startup used a data set based on East African patient data instead of European patient data. While this would probably provide better results, resulting diagnoses from this model would overlook diseases such as malaria and yellow fever, which tend to be more common in West Africa, and also result in improper diagnosis.

Finally, let's consider if the startup uses a data set from patients from the largest hospitals in West African countries. While this may seem like a good choice at first, this data set would probably be more representative of urban populations as compared to rural populations, also resulting in improper diagnosis. This example shows that there can be different scales of data representation and that coders need to be careful tests right questions about the context for each problem to design models appropriately.

The third criterion is value. Does the machine learning model produce predictions that are more accurate than alternative methods? Does it explain variation more completely than alternative models? Do the predicted values inform human decisions in a meaningful way?

For example, are they actionable? Telling farmers that they could improve their yields by moving to a different elevation is not useful to them. Are they timely? Having information but not enough time to act on it provides little to no value. Are

they delivered to the right people? You shouldn't build a system that provides information to frontline workers when decisions are made by their supervisors.

While machine learning is a powerful tool, it is not always the best approach to all problems. Organizations should have sufficient reason to think that machine learning would add value and make sure that they evaluate that assumption before scaling solutions.

The fourth criterion is explainability. How effectively is the use of machine learning communicated? It is important to ensure that the application is explained to end users in a way that is effectively communicating how outcomes were determined. Organizations seeking to apply machine learning outcomes without understanding the nuances of how models make decisions may use algorithm outputs inappropriately.

Let's look back at the earlier example of gender differentiated credit scoring in the previous module. An explainable solution could include information on why a specific individual was denied a loan and which factors they could change in order for them to increase their credit worthiness.

The fifth criterion is auditability. Can the models decision making processes be queried or monitored by external actors? Increasingly, organizations returning to black box machine learning solutions, whose inner workings can range from unintuitive to incomprehensible.

It is important that the outputs can be monitored externally to show that the model is fair, unbiased, and does not harm some users. This may require additional infrastructure, whether it is institutional or legal frameworks that require audits, provide auditors with secure access to data and algorithms, and require people to act on those findings.

The sixth criterion is equity. If used to guide decision making, has the machine learning model been tested to determine whether it is disproportionately harmful or beneficial to some individuals or groups more than others? Testing the results of algorithms against protected variables, such as gender, age, race, or skin color, is key to preventing the adoption of biased algorithms.

Does a specific algorithm fail for specific groups more often than it does for other groups of people? Does it misclassified different groups in different directions? Or did certain groups have different rates of false positives or false negatives?

It is also important to address the issue that accuracy and fairness are not necessarily correlated. Algorithms can be accurate technically but still inconsistent with the values that the organizations may want to promote when making decisions, such as who should be hired, who should receive medical care, or other similar decisions. Gaining an understanding of how these outcomes are derived and taking steps to mitigate them is an important piece in ensuring that unfair algorithms are not widely adopted and used.

The seventh criterion is accountability or responsibility. If used to guide decision making, are there mechanisms in place to ensure that someone will be responsible for responding to feedback and redress harms, if necessary? For example, an algorithm might be used to assist in diagnosing medical conditions. But the final diagnosis should be still provided by a trained medical professional.

When used by itself, the risk from false identifications from the algorithm can actually cause harm to individuals. However, consider if there is a shortage of trained medical professionals. Does the risk of misdiagnoses outweigh the risk of not treating people? These decisions are complicated, and there is not always a right answer.

It is important to keep accountability and responsibility in mind when designing systems. Remember, an accountable setup both make sure that there are systems in place to prevent harmful errors and make sure someone is responsible for correcting errors.

As a review, we talked about seven characteristics for the appropriate application of machine learning in this module.

The first is relevance. The second is representativeness. The third is value. The fourth is explainability. The fifth is auditability. The sixth is equity. And the seventh is accountability and responsibility.

Be sure to take this into consideration while you're implementing your own

solutions. Thank you for taking the time to take this course. We hope that you'll continue to watch the rest of the modules in this series.

[MUSIC PLAYING]