♨️

IBM AI Ethics(mod6)

Module 6

What is privacy?

Meet the team

How important is privacy? This story looks at a large educational institution and how privacy must be considered when developing AI systems.

A large educational institution is seeking to expand its reach by offering online courses. While developing an AI model to recommend personalized learning curricula, the team runs into some issues around data privacy and AI privacy. This story covers how the team learns about privacy issues and how privacy can be protected.

Identify the issue

An educational institution seeks to expand its reach by launching an online global campus. After analyzing the market, the institution’s business development team proposes creating personalized learning experiences for each student. The team believes that the most efficient way to do this is to use an AI model to create personalized learning curricula. The institution’s data science team is brought in to create a proof-of-concept AI model using sample data collected from the institution’s existing learning portal data.

When the proof of concept is complete, the business development and data science teams gather to review the AI model.

During the presentation, a sample learning curriculum generated by the AI model changes the course of the meeting:

"Kamal, we were not informed that data from our learning portal would be used to create the AI model. Now I am concerned because I thought that data was private. We need to make sure that users’ privacy is respected. Can you tell us about the privacy safeguards this AI model has in place?”

Kamal replies, “May I take some time and come back to you with that answer? I’d like to confer with my team.”

Explain the issue

Kamal meets with the team to share the feedback he received about the proof of concept. Quickly, he realizes that he needs to learn more about the basic concepts of privacy so that appropriate safeguards can be determined and put into place. He asks Nia from the Data Privacy team and Adrian from the AI Development team to clarify some basic terms and concepts.

The issue of data privacy

Nia explains, “Individuals can be protective about information or data related to them. Plus, in some countries, people have a right to data privacy; although, it’s important to remember that the definition of privacy and the kinds of data it applies to vary from country to country. This map(opens in a new tab) shows you how privacy regulations vary around the world.

“To give you a basic idea about privacy, let’s consider two kinds of data. **Personal information (PI)** is any information relating to an identified or identifiable individual, like a name or postal code. **Sensitive personal information (SPI)** is information that, if compromised, could be misused to significantly harm or inconvenience an individual, like a bank account number or birth date. How would you categorize these types of information?”

Take a moment and see if you can guess what kind of information is shown on the front of the card.

Kamal asks, “Thanks, Nia. But how does this relate to AI?”

Adrian from the AI Development team says, “That’s a good question, Kamal. Because the machine learning models that drive AI often need to be trained using personal or sensitive information, it is critical that AI systems prioritize and safeguard privacy. If a model is trained using personal or sensitive information without any privacy controls applied, then it could be vulnerable to breaches or attacks.”

Kamal asks, “Breaches and attacks? Can you give us an example?”

Adrian explains, “Let’s consider one type of privacy attack: **membership inference attack**. In a membership inference attack, an attacker tries to determine whether a specific individual was part of the training data set.

Because the data of individuals included in the training data set is compromised, their privacy is violated. Therefore, when we develop an AI system or train a new model, our goal must be to preserve and protect individuals’ privacy as much as possible!”

Kamal asks, “But what can we do to protect privacy?”

The issue of data privacy

Adrian responds, “Well, there are many privacy controls that can be applied to fortify AI against potential breaches of personal or sensitive data. Two that occur during model training are **model anonymization** and **differential privacy**. One that occurs after model training is **data minimization**. Let’s take a look at each.”

> **Model anonymization**
>
> (during model training)
>
> The goal of model anonymization is to anonymize the training data with minimal accuracy loss. After all, if the model is trained on anonymous data, then the model itself is anonymous and there is little risk to any personal data used during training

> **Differential privacy**
>
> (during model training)
>
> In differential privacy, random noise is added during model training to reduce the impact of any single individual on the

> model’s outcomes and to give a guarantee that an individual in the training data set could not be identified.

> **Data minimization**
>
> (after model training)
>
> Data minimization means that only data that is needed is being collected. This control helps prevent privacy breaches by limiting the amount of personal data that is collected in the first place and by ensuring that collected data is only as granular as needed. For example, data minimization might mean that you collect only an individual’s zip code instead of their full address, or only their year of birth instead of their full birth date.

Adrian concludes, “By applying privacy controls like model anonymization and differential privacy during model training, and data minimization after model training, we can fortify our models against personal data breaches and safeguard individuals’ privacy!”

Kamal thanks Nia and Adrian for their help and makes plans with the management team to meet again to discuss the privacy controls they will implement in the AI system.

Address the issue

Now, the team knows the concepts of personal and sensitive information and understands that information that seems harmless may be able to be used to identify individuals. Securing personal data and the AI model is important for both the company and their users.

Reflection: Privacy in AI

Imagine that you are part of the team trying to deal with privacy issues. Think about the following questions. Take a few minutes to reflect and type your responses in the following text boxes. (Writing an answer is a good way to process your thoughts.

These answers are for your use only. You have the option to download your response and save it. It will not be saved in the text box when you move on in the course.)

For question 1:

How could making the model publicly available introduce a risk to privacy?

If an attacker has access to a model, they might be able to infer which individuals were included in the training data. That’s why applying AI privacy controls is so important — when training data has been anonymized or has had noise added to it, it is much more difficult for attackers to determine who was included in the training data during a membership inference attack.

For question 2:

If there is risk involved with using personal information, is it still worth using?

Yes, it is still worth using even if there is risk involved. Personal information can be used to train models in the appropriate circumstances, as long as privacy techniques are applied to the data to preserve the privacy of individuals whose data is included.

For question 3:

What is another example of data minimization?

Using an individual’s industry instead of company or job title, using an individual’s area code instead of telephone number, using yes/no questions instead of collecting specific details (for example, asking “Did you graduate from high school? Yes/No” instead of asking for the high school’s name or the individual’s graduation date).