

# **“Scheduling ML Models & Data Augmentations” Module Transcript**

## **Chapter 1**

### **Introductions & Learning Outcomes**

Hi, everyone. My name is Josh Lemoine and I'm the product manager for One AI here at One Model. In this recording, we are going to go over scheduling machine learning models and data augmentations in One AI.

After watching this video, you will understand the importance of scheduling models and augmentations for timely insights and consistent performance. You will know the step by step process for scheduling models and augmentations. You will gain confidence in managing and adjusting scheduled runs as needed, and you will recognize the importance in always validating scheduled model and augmentation results.

## **Chapter 2**

### **Benefits of Scheduling Models & Augmentations**

#### Section 2 - Benefits of Scheduling Models and Augmentations

Before we dive into the how-to, let's talk about why scheduling these processes is beneficial.

First and foremost, timeliness. Large machine learning models and data augmentations can take hours to run, so scheduling them to run regularly, say overnight, allows you to have fresh results ready to review on a specified cadence. It also takes the burden of remembering to rerun the models regularly off of the user and puts this responsibility back on One AI. You can schedule these processes to run at specific intervals that align with your business needs so you have the necessary data when you need to make informed decisions.

Additionally, scheduled models lend themselves to better consistency. Regularly scheduled runs ensure that your models are consistently updated with the latest data and are making predictions based on the most recent information available. This consistency helps maintain the model's relevance and accuracy over time.

Regularly running machine learning models also checks that the model is not overfit or underfit to your organization's data. We want to make sure models are well fit, meaning it strikes an appropriate balance between capturing the underlying patterns in the data without being overly complex or simplistic so it can generalize effectively to new unseen data. If the model is under or overfit, you will have the opportunity to fix it sooner rather than later.

## **Chapter 3**

### **How to Schedule Models & Augmentations**

Section 3 - How to Schedule Models and Augmentations.

Now on to how to schedule machine learning models and data augmentations in One AI. The process is nearly identical for machine learning models and data augmentations, except for one small difference that I will point out.

Please note that if the last run of a model or augmentation was unsuccessful, meaning it errored, was canceled, or was ignored, you will not be able to schedule it until a successful run has completed and the model is in a pending or deployed status.

You will see a message above the scheduler if this is the case.

First, navigate to the augmentations page by clicking 'Data' > 'Augmentations' from the main ribbon menu.

Then scroll to the machine learning model or data augmentation you wish to schedule and click the 'Schedule' button right next to the edit button.

Click on the enable checkbox to unlock the scheduling section.

Next, choose the time period level you want to schedule the model or augmentation to run on with the 'Run every' dropdown menu.

Your choices are 'Day', 'Week', or 'Month'. Monthly is most common. Then, if you selected 'Week' or 'Month' for the time period level, you must pick the days or dates you want the model to run on. You must select at least one, but can select multiple if desired.

If using weekly, there is no default day. If using monthly, it defaults to the 31st, which not all months have. We recommend switching to the 1st or another date that meets your needs and exists in every month. Next, you have the option to set the time you would

like to schedule the model to run at. This defaults to 00:00 UTC, but you can change the hour and minute with the 'At time' dropdowns, and you can change the time zone with the 'Time Zone' dropdown.

## **Chapter 4**

### **Auto Deploying Data Augmentations**

If you are scheduling a data augmentation, you will also see an 'Auto Deploy' checkbox. Checking this box means that the results will be automatically deployed to the front end of One Model for use in Explore and Storyboards once the data augmentation runs and data processing has been run and successfully completed. If it errors, the results will not deploy, and you will have to address the issue and rerun.

Data augmentations are not predictive, so performance is not a factor. Therefore, auto deploying the results does not hold the risks that auto deploying a machine learning model has. However, results should still be validated.

The auto deploy checkbox does not appear when scheduling a machine learning model because the task of deploying machine learning models results requires human intervention since performance of models can change drastically over time. Once you have reviewed the EDA and results summary reports, you can manually deploy the results if you are happy with them. Finally, once you are happy with your schedule configuration, you will click the 'Save' button.

If you are scheduling a commute time data augmentation, when you click save, you will need to read the warning and click 'Agree' and 'Confirm' if you are comfortable with your organization's zip code combinations being sent to the MapQuest API. This is the same warning message that appears every time you create or run a commute time augmentation. If scheduling was configured correctly, you will see information populate under next scheduled run which states the date and time of the 'Next Scheduled Run' and how many days there are until the scheduled date.

## **Chapter 5**

### **Conclusion & Thanks**

Thank you for joining me to learn about scheduling machine learning models and data augmentations. Happy scheduling!