# “Model Refinement” Module Transcript

## Chapter 1

### Intro, Topics Covered, & Learning Outcomes

Hey, my name is Hayley and I'm on the One AI team here at One Model. In previous modules, we've covered the foundations of machine learning and best practices for building models. You also explored how One AI recipes streamline the model-building process. In this module, we'll bring all that knowledge together to create a new hire failure model. This model is great for anyone trying to understand gaps in the hiring process, reducing turnover costs, and improving overall employee retention.

We will cover an overview of the new hire failure model, key considerations before beginning the model-building process, step by step instructions for building a new hire failure risk model in One AI with a recipe, and then we will close with a review of some of the important insights that can be drawn from this type of model.

After completing this module, you will understand and address important considerations before beginning the model-building process, ensuring effective and accurate predictions, be able to confidently create a new hire failure risk model that is relevant to your organization by using the One AI recipe, and you will use the model insights to pinpoint the factors that most significantly impact new hire failure, enabling you to make informed business decisions.

## Chapter 2

### Overview & Purpose

Section 2 - Overview and Purpose

The new hire success or failure model is designed to predict whether new hires, hired during a specified time period, will remain employed at your organization within a defined time frame after their hire date. By analyzing attributes, such as employee demographics, onboarding surveys, compensation, team attributes, and much more, this binary classification model identifies individuals at high risk of becoming new hire failures.

The model classifies new hires into two outcomes, new hire failures, meaning they are no longer with the organization after the specified time period, usually 6 months or one year, and new hire successes, meaning they are still with your organization.

Each instance in the model represents a hiring event typically defined by an external hire metric and identified by unique person IDs from the employee event table. This approach differs from models like the voluntary attrition risk model, where instances are the employees themselves rather than their hiring event, and data comes from the employee table rather than the employee event table.

The model creator chooses the model population, the prediction periods, how much data the model is trained on, and the input features used to train the model to make its predictions. Input features can include core attributes and individual and team generative attributes.

Additionally, model results will include the top drivers of new hire failures and new hire successes, which allows you to gain deeper insights into the underlying factors that influence whether new hires will remain employed with your organization. This enables organizations to identify potential high risk hires early, allowing for targeted interventions to improve retention and reduce turnover costs. This proactive approach enhances overall hiring strategies and workforce stability.

Some key use cases for the new hire failure model are improving retention rates. So identifying high risk hires early helps leaders gain insights into factors affecting retention and implement targeted interventions, such as additional training or mentorship programs to improve retention rates. By proactively reducing new hire failures, you also decrease turnover costs. According to the findings of the Allied Workforce Mobility Survey, it takes on average eight months for recently hired employees to achieve their peak productivity levels, so we wanna make sure good new hires hang around long after that.

For example, we helped a client use this model to find the optimal team size for their warehouse employees, which reduces new hire failure risk using team-based generative attributes and EDA. Previously, many new hires were on too large of teams to learn their roles effectively.

Additionally, enhancing hiring strategies. So some drivers of new hire failure are within your organization's control, while others are inherent to the new hires themselves. Analyze which of these attributes contribute to new hire success to refine recruitment strategies and select candidates more likely to succeed. For example, we helped a client find which candidate sources produce the most new hire successes again using EDA.

Also, optimizing onboarding programs. So by understanding the specific needs and risk factors of new hires, human resources can design more effective onboarding programs that foster engagement and productivity to ensure hires feel supported. For example, we helped a client use this model to identify the most important onboarding courses as well as find an optimal time period for onboarding for a specialty group of employees.

## Chapter 3

### Key Considerations Before Model Building

Section 3 - Key Considerations Before Beginning the Model-Building Process

Before creating a new hire failure model, there are some important factors that should be carefully considered to ensure its effectiveness. First, define your objective and goals for the model, such as improving retention rates, reducing turnover costs, costs, or enhancing hiring strategies.

Additionally, determine if the model aims to confirm hypotheses, conduct exploratory analysis to just learn more about this kind of data, or replicate a model previously built outside of One Model. This will help guide your approach.

Next, ensure you have accessible and high quality data to train the model. Verify that the necessary data sources are integrated and validated in One Model. Consider if there is sufficient historical data on hiring, onboarding, training records, compensation, and other interesting data the model may use to learn in order to make accurate predictions about new hire failure.

And finally, ensure you have the resources to build and maintain the model. One AI simplifies model building and visualization, but ongoing maintenance is essential. This includes rerunning and refining the model with new data, creating effective visualizations and storyboards, incorporating feedback, and monitoring performance.

Next, we'll move over to One Model for a demo on building a new hire failure model in One AI with a recipe.

## Chapter 4

### How to Build in One AI

Section 4 - How to Build a New Hire Failure Model in One AI

Let's get started. Click 'One AI' in the main ribbon menu. This will take you to where you can edit or create models. To add a new model, click 'Add Machine Learning Model' in the upper right hand corner.

First, give your model a unique name in the Display Name field. Click 'One AI Recipe' and then 'Configure One AI Recipe'. This will bring you to the One AI Query Builder screen. Screen. Under "What are you interested in predicting?", select 'New Hire Success / Failure' from the list of available recipes. Once selected, the problem statement is populated with the information we need to provide to build the model. One AI simplifies most settings by automatically selecting defaults or trying an intelligent subset of options, so we will primarily focus on data framing here.

In step one, we must select a metric to define new hire failure. This should be an event metric that captures hires terminating from the organization. A metric that includes all terminations is the most common and straightforward selection. So first, I'm going to select the 'Terminations' metric that includes all termination events to include voluntary, involuntary, and other. If desired, here you can filter the metric to exclude particular types of termination events, but I'm not going to do this for this model.

If you don't have an all encompassing terminations metric, you can build one by using the event occurrence column from the employee event table and filtering the event reasons to include everything related to terminations. Something like this works well.

Next, we must select the group of hires for which One AI will predict new hire failure risk. You can use a general external hires metric and filter down or a specific hires metric like hires in Texas or engineer hires to get to your group of interest. I'm just going to use a general External Hires metric and filter down in the last step of this step. Then choose a unique identifier for each instance in the model. Typically, person ID from the employee event table is the best choice for this model, but any unique ID from the employee event table will work. One AI will verify the IDs uniqueness for the model. Do not use an ID from the employee table or any table other than the employee event table because the unique ID must come from the same table that the population metric originates from.

Next, select your population date. The population date anchors the query with the predict frame and the train test dataset offset from it. Common choices are today, the end of last month, or a static date. I'm selecting 'Today', which means that this model will use history going back starting from today and will predict a year in the future starting from today as well.

In the final step for this step, use filters to refine your hires population. Instead of running the model on the entire organization, segment it by factors like department or job level because motivations and drivers may vary by group. For example, I'm selecting the 'Country' filter here to only include new hires from the United States in this model population.

Next, select how far in the future you want to predict from the population date. 6 months or one year is the most common choice because it's helpful to know who will be a new hire failure relatively soon. You can adjust this to match your organization's definition of new hire failure. I'm going to make it 6 months.

Then choose how much historical data to use for training the model. We have found that best performance results from enough training intervals to give us one to two years of historical data. So because I selected 6 months in the previous step, each interval equals 6 months. Therefore, I will enter a '2' here so that the model has one year of historical data to learn from.

Adding meaningful labels to the values you are predicting will simplify reading the result summary and model storyboards greatly. Null values and 0s indicate higher successes, while values greater than 0 indicate higher failures for this model. Click 'Load Target Metric Values' to set overrides. You can name the groups however you like, but just ensure there are only two unique labels total. Something like this works well.

Next, select core attributes, which are the input variables that form the dataset the model learns from and uses to make predictions. Aim for a balance, enough attributes for robust predictions, but not so many that the results become difficult to interpret.

If you don't intend to manually select each column into your model dataset, I recommend using a 'Balanced' scope for this model. If you use a narrow scope, you will be limited to the columns from the employee event table, which typically doesn't have a lot of predictive columns about the actual hire. The balanced scope will include the employee table, which is where a lot of that good predictive core data sits.

Then, go through the 'Included' columns and exclude anything you don't want One AI to use using the 'X' button. Then, go through each of the 'Excluded' tables and columns and bring in any additional features of interest using the checkmark button.

In the next step, select the generated attributes you want to use in your prediction using the check boxes to the left of the attribute name. You can create, edit, and delete generative attributes in this step. For more information, check out our module on generative attributes. I recommend working with your CS team to add a supervisor ID, manager ID, or however you define a team column to the employee event table if it's not already there. This will allow you to build team based generative attributes, which are particularly predictive for new hire failure failure models.

In the next step, click 'Generate Data Statistics' to verify that all of your previous selections are valid. This helps identify errors that cause the model to air while running. It also provides valuable data exploration information about the train test dataset, which can help you identify columns you want to exclude and just learn more about your model dataset. As you can see, we received the green light "Success" so we can proceed. If One AI finds a problem, it will display action needed in red with details about the errors that need to be corrected before moving on.

In the final step, you can download the train test dataset, predict dataset, or both that were generated by this recipe.

Once complete, review each step. Once it looks good, click the save icon in the upper righthand corner to return to the Machine Learning Model screen. Here, you can manually configure model settings. Refer to our Global Settings module series or Advanced Configuration module series if interested. At minimum, I recommend at least generating SHAP values if you intend to model this data on a storyboard.

Once complete, scroll to the bottom and click 'Create'. If you do not click create and navigate elsewhere in One Model, you will lose your work from the One AI Query Builder screen. You can now create new hire failure models in One AI. The next step is to run your model, evaluate and refine your model, and then deploy and share the insights.

## Chapter 5

## Insights from a New Hire Failure Model

Section 5 - What Insights Can be Drawn from a New Hire Failure Model?

New hire failure models can provide several insights. First and foremost, you will receive data on the model's predictions made at the instance level. These can and should be aggregated to larger groups for better accuracy.

Without creating any visualizations, you can view individual predictions by dataset ID in the Results Explorer. You can also use storyboards to analyze individual employees or groupings within the model population, examining which features drive behavior, how it compares to the mean, and an explanation for these drivers.

For example, this is filtered down to an individual new hire, but you could aggregate it up to a job level, a particular plant, or whatever other dimension you feel.

You can analyze correlation data from the EDA report to see how your input features interact. This can help you understand how your data is related to each other and uncover interesting patterns to provide even more insights on what makes a new hire a failure vs. a success.

You can also understand and visualize the factors influencing new hires to either become new hire failures or successes. If SHAP is enabled in your model and incorporated into storyboards, you can view the top drivers for both classes in order of importance.

For example, here we can see that being a quality hire and having a flex arrangement is very indicative of becoming a new hire success, while not being a quality hire and not having a flex arrangement is indicative to becoming a new hire failure.

You can also analyze new hire failure risk by groupings within your model population. Any dimension that can be used to filter the hires metric can break out your model population to identify high, medium, and low levels of new hire failure risk by grouping. For example, using the Job Level dimension, you can see that hires at the intermediate staff level have the highest population predicted to be new hire failures. This allows you to plan for targeted interventions.

Next, you can analyze individuals within your model population to get targeted insights about specific new hires. You can view the predicted risk of new hire failure, feature impacts, and feature value differences from the mean.

And finally, you can analyze by-name lists of folks at high risk of becoming new hire failures and view other information about these employees alongside this information, such as who their manager is, what org unit they belong to, their job title, and whatever else you want. These are just a few examples of insights that can be drawn from new hire failure models. Many more are available depending on your objectives and what questions you are using the model to try to answer.

## Chapter 6

## Conclusion & Thanks

By mastering the new hire failure model, you're equipped to proactively tackle new hire turnover, refine hiring processes, and enhance workforce stability. This model allows you to gain valuable insights to optimize onboarding and make informed decisions that improve retention and reduce costs. Continuous monitoring and refinement will ensure the model remains effective. Happy modeling!