

AI Summit Webinar: Artificial Intelligence 101

Frequently Asked Questions

Q: I want to extract PDF and import extracted pdf data into SAP Business one Outgoing Payment. How can I do this? please help me?

A: You can do this using our Document Understanding capabilities. There are courses on <https://academy.uipath.com/> to help get you started.

Q: How can you retrain a model through the Action Center?

A: Data validated via Action Center can be added to a training dataset. The dataset is used to train/retrain the ML model as needed.

Q: When do you plan to support Cyrillic , e.g. Bulgarian language in the ML functionality for labeling and text extraction?

A: Processing Cyrillic is a functionality related to OCR engine capabilities. Our Document Understanding product can work with several OCR engines that are able to process it. Just make sure the OCR engine you're using has this capability and you're all set.

Q: With the OCR it works somehow. But if we want to use ML for invoice processing, e.g. with a ""digitized"" pdf, Cyrillic is not working at all!

A: Did you retrain the ML model using your own sample documents? Our out-of-the-box invoices model was trained on English invoices, so it is expected that it will not work on Cyrillic.

Q: Hello everyone, how accurate is the RPA in reading and validating documents, where documents from several vendors have different models?

A: Accuracy heavily depends on the documents themselves and the scan quality (if they are scanned). With proper training on a relevant dataset, we can achieve very good results.

Q: Do you have any courses for learning detailed AI Center capabilities and related AI products from UiPath? I know we have a workshop coming up tomorrow, but we need deep dive in that area. Are there any technical pre-requisites for AI courses?

A: If you go to academy.uipath.com, search AI Center (or AI Fabric - the old product name), you will find a course.

Q: How will the models from the UiPath Partners be available to use? On AI Center - like UiPath models? Will be there a platform like the marketplace where users can share the models?

A: As long as the models can be exported as python code, they can be deployed via AI Center. Our product and engineering teams are working on marketplace for users to share the models.

Q: How long (time needed) and or how many samples that needed for machine learning models to ensure us for gaining better understanding docs? 70-80%?

A: This is heavily dependent on your use case and your data. Recommendation is to test against the out of the box model (if applicable) and that will be a good base performance (there will be a difference between a model that is performing well on 50% of the predictions compared to 80%). Best is to always train in iterations – start with a smaller sample, see the performance, if you're not happy, retrain on a bigger sample, see the difference and you'll start to gain some experience and know what to expect depending on the use case.

Q: Are those evaluation metrics included in the AI Center? In other words, will that be automatically calculated for us?

A: Yes, depending on the model you will see the different KPIs for the Out of the Box Models. If you're building your own model, you will be the one writing the code to do (show) this in AI Center.

Q: I don't understand the comment that this doesn't require a data scientist. This is way beyond what even an experienced automation developer could do on their own. Comments?

A: With AI Center, you have the possibility to bring RPA developers and data scientists together, which was a huge gap before that. With out-of-the-box models UiPath offers, you don't need data science experience in the sense that you don't need to write any line of code to build a model. All of the code and meta data are already there. With a few clicks, you will be able to train and evaluate the model. With our OOTB models, you need to know the basic concept, which is what we went through today. You might also need to know how to prepare data for model training and retraining.

Q: For instance, regarding understanding document (Purchase Order or Invoices / Billings). How many samples or how many tests needed, so that I can say to end user, you can rely to this DU / ML.

A: We answered this question live, please check the recording on the minute 49:20 – 51:14.

Q: Is there a limit to the training data that a model can be trained on?

A: No, it only depends on the compute power. it might take 2-5-10 days to finish training though.

Q: Can you adjust the model parameters for OOB models provided by UiPath or 3rd party vendors?

A: User customizable parameters are supported today. And we will have a much nicer experience for this later this year that will auto-populate user-defined parameters in the UI.

Q: Why do you train from base model & not incremental?

A: It doesn't always make sense to do incremental learning, though we can absolutely support it today if a model is built to do incremental learning.

Q: Does UiPath AI center assist in cleansing unstructured data, like twitter feeds, etc.?

A: It depends on the model. On the pages of our out-of-the-box models, you will see what input is required. For example, if you are using the English Text Classification model, as dataset, you need to put all the text without any cleaning. And after the pipeline is finished, you will also see clean data.

Q: Do you have software/hardware requirements for AI Center infrastructure for an on-premise environment?

A: For on-prem, here are the software and hardware requirements for AI Center:
<https://docs.uipath.com/ai-fabric/docs/ai-fabric-singlenode-requirements>

Q: Quick question regarding the methods we saw at the beginning of the webinar: how do you make a relevant choice when choosing between document understanding or language analysis when you need to analyze and extract data from a document which structure can change but concerns a same topic, as a resume for example?

A: So this is the part where the business analysis comes in, based on what is required, you would use one or another, or maybe even both but it really depends on the use case. Maybe you want to process emails and based on the contents of that email, you want to do some language analysis on some sentiment or classification analysis. In this case, you will need sentiment analysis model and the classification models. If you want to analyze invoices, you will need to use an invoice model. More on this in the recording, see from minute 58:10 to 59:35.

Q: When retraining on the base model, how do you account for different versions and how the 'training' may be adversely affected in the live session to limit the amount of re-work and errors?

A: A I would recommend to take the Academy course for AI Center, which will provide better understanding of these concepts. There are 2 separate things: ML Packages – where the pretrained model is stored within different versions and the ML Skills – which is the model up and running ready to give predictions.

Q: When we create new version of existing model, will it get previous trainings from previous version into the new version (by default) or new version require brand new training?

A: We answered this question live, please check the recording minutes 45:30-47:28

Q: How is the retraining implemented (with the validation station and without it)?

A: Retraining works the same with or without Validation Station, the only difference is how you get the additional data for retraining: it can come from Business Users when they're correcting the data in Validation Station (or Action Center) or if the dataset is labelled separately – directly in Data Manager for Document Understanding models or in another tool or format for other models. Regarding retraining, the same principles apply as for training. Check more details here: <https://docs.uipath.com/ai-fabric/v0/docs/uipath-document-understanding#training-and-retraining-pipelines>

Q: I'm very interested in TPOT models. Do you have a place where we can see training models explained? So far, I haven't found how to design the CSV of the dataset (where do I set the expected output, how many attributes are accepted, etc.)

A: Please check out our documentation pages for the TPOT models: <https://docs.uipath.com/ai-fabric/docs/tpot-automl-classification>
<https://docs.uipath.com/ai-fabric/v0/docs/tpot-automl-regression>

Q: How do you measure confidence threshold?

A: Each model makes its own confidence calculations.

Q: How optimal are the times at the time of re-training to ML models?

A: In the beginning you might need to retrain more often and in time to become less and less often, but this is per case basis. Check how your model is performing, how much additional data you've gathered, and you will decide when retraining is needed.

Q: Is there any recommendation on the minimum number of samples that we need for a given number of available data set count?

A: Our recommendation is to always retrain with more data on the base spot. This is due to the deep learning architectures that are data-hungry: the more data you have, the better they will be able to learn and to find patterns, and to predict on top of them. See the recording for more details: 42:41 – 44:02

Q: Models that have been trained on a dev environment, can I use that/import into Prod environment or do I have to train the model in the Prod environment, again? Is there a way to extract/export the models for future use (and different customers)?

A: You can do it on-prem. For cloud, this functionality will be available soon. For now, you can export the data set and retrain the model in the new environment.

Q: What's the link for the immersion lab? It won't click to it?

A: <https://www.uipath.com/automation-immersion-lab>

Q: For supplier invoice ML training, do you recommend one training model per supplier or one training model for a group (similar invoice format) of suppliers? We have about 50 suppliers who send in invoices daily.

A: If all invoices are in the same language, we suggest using a single model for everything. Just make sure that the dataset is diverse and contains sufficient samples of all formats/suppliers.

Q: Can we think of a use case where the bot works as an operator to submit job statement production job through a web-based platform? And can the bot help in error recovery and reprinting?

A: Yes, robots and humans work well together, the use case you have in mind is possible.

Q: If I wanted to train a bot to reply to incoming emails, would I be able to train a model to answer multiple questions in one response?

A: It really depends on the questions and answers, this might fall into expert systems (like Alexa, Siri, etc.) which are quite complex to build from scratch. Have you considered chatbots?

Q: Is it possible to create a setup where the orchestrator is on-premise and the cloud AI Center is on cloud platform.

A: Yes. AI Center is connected to Orchestrator, sitting on the same SQL Server. But you can have mix deployments as well. You will need to use this by opening ML skills as endpoints. So you can have all deployment options you want.

Q: Hi, if I purchase license for "UiPath Document Understanding", does it include OOTB model as well that be applied and retrained?

A: Yes, UiPath Document Understanding license includes all the OOTB models in the Document Understanding category. If you want to use other ML models, you will need AI Center licenses.

Q: Are there any UiPath H2O use cases?

A: H2O is an AutoML platform – you dump your data, and it will automatically create models on top of this data. You can pick the one that you like best and use it within AI Center. There are many, many use cases that can be tackled with these technologies.

Q: Is there any plan to support Chinese in Language model of OOB?

A: For Document Understanding, the languages we cover include Latin, Cyrillic, Greek. We have a model in public preview for Japanese and other Asian languages.