

ABBYY



ABBYY Vantage 2.3.2

Release Notes

Table of Contents

ABBYY Vantage 2.3.2 release notes 3

 About the release 3

 Main features 3

 Available Document skills 3

 Classification skill 5

 Handwriting 5

 Improvements in OCR technologies 5

 Mobile Input 6

 Adminstrating Vantage 6

 ABBYY Vantage Advanced Designer 7

 Bug fixes 9

 Limitations and known issues 12

ABBYY Vantage 2.3.2 release notes

About the release

This is a minor Vantage release that introduces new features and bug fixes.

The [Main features](#) section describes updates to the product's functionality.

The [Bug fixes](#) section contains a detailed list of issues that have been fixed in this release.

This release has some known limitations and issues, which are described in the [Limitations and known issues](#) section.

Main features

Available Document skills

ABBYY Vantage includes multiple Document skills in the Skill Catalog in order to handle different types of documents from various industries: customer onboarding, accounts payable, transportation, insurance, etc.

For more information about ABBYY Vantage skills, see:

- <https://vantage-eu.abbyy.com/help/english/SkillCatalog/?topic=abbyyskills> if you are in Western Europe
- <https://vantage-us.abbyy.com/help/english/SkillCatalog/?topic=abbyyskills> if you are in North America
- <https://vantage-au.abbyy.com/help/English/SkillDesigner/?topic=abbyyskills> if you are in Australia

Field regions in the Identity Documents (ID) skill during Manual Review

In the Identity Documents (ID) skill, Manual Review will display the specific regions where extracted fields from identity documents are located.

Updates to existing built-in skills

▼ Invoices

- Invoices JP: extraction of official company IDs is now supported.
- Invoices UA: new skill for processing Ukrainian invoices.

▼ Purchase Orders

Improved field extraction quality after training the neural network on more documents.

▼ Brokerage Statement

Improved field extraction quality after training the neural network on more documents.

▼ Receipts

Extraction of line items is now supported. Extracted columns: Description, Quantity, Price, SKU, Amount.

▼ Tax Forms

- Processing of handwritten text is enabled for forms:
 - 1095-C
 - 1095-B
- Address parsing is enabled for forms:
 - 1099 Composite
 - W2-G
 - 1099-OID
 - 1099-G
 - 1099-MISC
 - 1099-NEC
 - 1099-INT
 - 1099-B
 - 1099-R
 - 1099-DIV
 - 1098

New built-in skills

The following new skills have been added in version 2.3.2:

▼ Loan Origination skills

- Initial Escrow Account Disclosure Statement
- Desktop Underwriting
- Appraisal 1025
- Appraisal 1073
- Deed of Trust

▼ Tax Forms

- 1095-A, Health Insurance Marketplace Statement
- W-2G, Certain Gambling Winnings
- 1099-OID, Original Issue Discount

Classification skill

Online Learning for a Classification skill with multi-document transactions

In version 2.3.2, when a Manual Review Operator verifies classification and extraction from several documents of a transaction, Vantage will use this information to improve both Classification skills and Document skills.

Handwriting


Vantage 2.3.2 supports the following four languages for ICR:

- English
- German
- French
- Spanish

The Handwritten option is enabled by default for Document skills and disabled by default for the OCR skill.

Users can select several recognition languages and enable the Handwritten option. The system will automatically select the best approach to process each document.

Additionally, starting from version 2.3.1, multiline handwritten fields are supported.

 **Note:** For all languages, the Handwritten text recognition option works best if only the target language is selected in the list of recognition languages for the skill.

In version 2.3.2, the handwriting recognition quality for French and Spanish forms, agreements, and some other document types has been improved.

Improved ICR recognition on fields with combinations of letters and digits for English

The latest version demonstrates improved recognition accuracy for handwritten text areas such as ID numbers, serial numbers, and other alphanumeric combinations in English, surpassing the performance of previous versions.

Improvements in OCR technologies

Further improvements in the quality of Japanese OCR

The quality of Japanese OCR in Vantage has been further improved for condensed and expanded characters, as well as half-width katakana, with fewer OCR recognition errors caused by diacritics.

Rotate incorrectly oriented pages in digital-born PDFs

Automatic rotation preserves the text layer of PDF documents oriented incorrectly.

In previous versions, OCR was used to recreate the text layer. In version 2.3.2, incorrectly oriented pages are automatically rotated and the original text layer is preserved.

Mobile Input

For detailed instructions on working with Vantage Mobile Input, see


- <https://vantage-eu.abbyy.com/help/english/QuickStart/?topic=mobileupload> if you are in Western Europe
- <https://vantage-us.abbyy.com/help/english/QuickStart/?topic=mobileupload> if you are in North America
- <https://vantage-au.abbyy.com/help/English/QuickStart/?topic=mobileupload> if you are in Australia

Special interface for capturing passports

This Vantage release introduces a new special user interface for capturing passports and improves user experience when working with business processes that require the processing of ID documents.

When enabled, the special interface for passports provides the user with hints regarding how the document should be placed in order to achieve the best quality of the resulting image.

To enable the special user interface for capturing passports, a new URL parameter should be added to the mobile input link that is generated by the Vantage API.

 **Note:** The capture of passports is not currently automated. Users should manually capture the document image once the document has been centered on the screen.

Documentation on Vantage Mobile Input internet communications

Vantage Mobile Input transfers captured documents from the user's mobile device to Vantage servers for further processing.

The Vantage documentation describes in detail the endpoints and protocols used for the data transfer, allowing the customer's security team to assess Vantage Mobile Input communications.

Administering Vantage

Server-to-server authentication scheme now supported

A new authentication flow called Client Credentials, which is suitable for server-to-server communications, is now supported in the Vantage API. This flow is most useful when there is an application that allows end-users to process documents using Vantage without having accounts in Vantage.

The new flow can be enabled for any tenant's API client. When enabling the flow, the tenant administrator can assign roles that will be granted to the application that connects to Vantage using this flow. When connecting to Vantage, the application will get an access token with the roles specified in the client's API settings.

This flow is recommended for situations where an external identity provider is enabled for the tenant, allowing for authentication in Vantage even when the Authorization Code flow would not be suitable.

For more information, see:

- https://vantage-eu.abbyy.com/help/English/AdminGuide/?topic=admin_managingrestapiclients_createclient and https://vantage-eu.abbyy.com/help/English/DevGuide/?topic=developer_apiauthentication_clientcredentialsflow if you are in Western Europe
- https://vantage-us.abbyy.com/help/English/AdminGuide/?topic=admin_managingrestapiclients_createclient and https://vantage-us.abbyy.com/help/English/DevGuide/?topic=developer_apiauthentication_clientcredentialsflow if you are in North America
- https://vantage-au.abbyy.com/help/English/AdminGuide/?topic=admin_managingrestapiclients_createclient and https://vantage-au.abbyy.com/help/English/DevGuide/?topic=developer_apiauthentication_clientcredentialsflow if you are in Australia

ABBYY Vantage Advanced Designer

For detailed instructions on working with Vantage Advanced Designer, see the ABBYY Vantage Advanced Designer section in online Help:

- https://vantage-eu.abbyy.com/help/English/AdvancedDesigner/?topic=ad_introduction if you are in Western Europe
- https://vantage-us.abbyy.com/help/English/AdvancedDesigner/?topic=ad_introduction if you are in North America
- https://vantage-au.abbyy.com/help/English/AdvancedDesigner/?topic=ad_introduction if you are in Australia

Added ability to search for two related companies

A Classify By Company activity uses Data Catalogs available on the Vantage server to which you are connected. In some cases, you may want to search a document for a pair of related companies (for example, a supplier and a purchaser of goods). To do this, you can set up a simultaneous search for both the issuer and the receiver of a document by adding a Data Catalog of type Document Issuer Companies and one of type Document Receiver Companies (optional).

If there is a likelihood that your document issuer companies work with any of the document receivers, you don't need to fill in the Company Correlation ID column in your Document Issuer Companies catalog for the search to take into account all possible company pairings. However, sometimes business units may have their own supplier databases. Any supplier that works with several business units will have multiple entries with different IDs. In this case, you will need to find the exact entry for the supplier that corresponds to the business unit. Fill in the Company Correlation ID column in the Document Issuer Companies catalog. The search will then look for pairs of companies and require that the correlation ID for the document issuer matches the document receiver ID. If some correlation IDs are missing, pairs with matching correlation IDs will be prioritized.

To configure the Classify By Company activity, you need to select two directories of the following types: Document Issuer Companies and Document Receiver Companies. Next, configure the mapping of the fields and columns of both directories. If the Correlation ID column in your issuer's catalog is empty, the company will be paired randomly. If the Correlation ID column in your catalog is not empty, then the only pairs that will be searched for are those where the Correlation ID in the issuer's catalog is identical to the BUID in the recipient's catalog. If no such pairs are found, an entry with an empty Correlation ID can be selected as the issuer.

For more information, see:

- https://vantage-eu.abbyy.com/help/English/AdvancedDesigner/?topic=ad_bycompanyactivity if you are in Western Europe
- https://vantage-us.abbyy.com/help/English/AdvancedDesigner/?topic=ad_bycompanyactivity if you are in North America
- https://vantage-au.abbyy.com/help/English/AdvancedDesigner/?topic=ad_bycompanyactivity if you are in Australia

Testing Deep Learning results after each epoch

Changes have been made to the display of statistics for the training of the Deep Learning activity in Advanced Designer. Previously, internal training metrics were displayed during the training process. Now, however, a validation test is conducted after each training epoch, after which statistics are displayed. All metrics in this output are the same as those in the final report generated after activity testing.

The test performed while training is in progress is carried out on 20% of the documents, while the final test is carried out on all of the documents. For this reason, the resulting numbers of the two tests may differ. The more documents are used for training, the smaller that difference should be.

For more information, see:

- https://vantage-eu.abbyy.com/help/English/AdvancedDesigner/?topic=ad_deeplearningsemistructured_progress if you are in Western Europe
- https://vantage-us.abbyy.com/help/English/AdvancedDesigner/?topic=ad_deeplearningsemistructured_progress if you are in North America
- https://vantage-au.abbyy.com/help/English/AdvancedDesigner/?topic=ad_deeplearningsemistructured_progress if you are in Australia

NLP Language support

Vantage 2.3.2 now supports new languages for Advanced Designer NLP and NER activities.

The NLP Deep Learning and Named Entities (NER) activities now also support Dutch.

Starting from version 2.3.2, Vantage offers full NLP support* for:

- English
- French
- Spanish

- German
- Russian
- Japanese (preview)
- Portuguese
- Italian
- Dutch

*"Full NLP support" means support for Segmentation, Named Entities (NER), and NLP Deep Learning activities.

Bug fixes

In Vantage 2.3.2, the following bugs have been fixed:

Zendesk Ticket #	Description
402844	In the "Select Skill" dialog box, skills were not sorted alphabetically.
539757	The extraction quality of Line Items was improved for the Purchase Orders skill.
536970	Pages were not rotated correctly.
534679, 543675	A "400 Bad Request. Request Header or Cookie Large" error occurred when trying to authenticate via Azure Active Directory.
545378	Dots were not recognized in decimal values.
549694	Extraction slowed down when a document skill was trained on a form with a lot of text.
551557	Trying to log in to Vantage in Incognito mode or after clearing the browser cache resulted in the cursor spinning indefinitely.
554301	Extraction training on a document set containing 10000 invoices failed with the "Could not set vantage model, because could not find training step" error.
554276	Some Process Skills were not shown in Skill Monitor after being published.
557400	A "Get Results: Header 'Content-Length' is missing or invalid." error occurred while using UiPath connector.
473168	Training Invoice EU in full text recognition caused horizontal fields to incorrectly be displayed as vertical fields.
518006	Extra spaces between numbers were detected on some documents.
537524	The recognition quality of line items was low when processing Chinese invoices.

Zendesk Ticket #	Description
541801	Training for new columns succeeded in Advanced Designer but failed in Vantage.
541619	Original file names (containing symbols) was not preserved on import.
545267	The paragraph separator symbol was detected when performing full text recognition, causing problems with Catalogue lookup.
543855	The slash symbol (/) was not recognized correctly in full text recognition.
47978	Unknown symbols in the recognition results for some documents.
551718	Handwritten numbers were not recognized correctly.
551864	A loss of quality occurred when exporting to PDF/A using OCR skill.
552624	Extraction field training failed for customer's skill because of a time-out.
552624	Training failed for a specific Document Skill.
553411	A copy of Invoice CN skill (v13) stopped capturing fields after it was opened in Advanced Designer and published again.
554507	A "Not enough memory" error occurred when restoring a skill from a local folder in Advanced Designer.
555341	An empty screen appeared in the Document skill editor for a specific skill.
555197	Having a document with a content restriction in the training document set caused the training of Deep Learning for NLP activity to enter an endless loop.
549863	Data Catalog lookup rules were not populating data.
555943	An "Internal program error "PerspectiveFixerConverter.cpp, 162" error occurred during training.
374940, 413881, 443290, 513475	An option to resize columns width was added to Skill Catalog, Skill Monitor, and Document Repository.
559471	Records for some steps were not exported via Vantage Reporting API.
529674	Language settings were not displayed correctly in Invoice EU skill.
523175, 534488	Digital-born PDFs with an incorrect text layer could not be rotated.
539699	Document Skills in the Extract activity list were not sorted in alphabetical order.
536318, 552151, 535416	The ResultClassName and ClassName tags were missing from the output JSON file.
547209	An invalid workflow could be published in Process skill editor.

Zendesk Ticket #	Description
548144	An error occurred in some cases after selecting the data catalog record manually on Manual Review.
550030	FlexiLayout Studio projects with the Object collection element were imported incorrectly to Advanced Designer.
551540	The "Due Date" field was recognized incorrectly when recognizing a JP invoice document.
553209	Max error percentage property was imported incorrectly from FlexiLayout projects to Advanced Designer if its value was set to 0, causing an error and elements to not match.
560730	A "RunTime error" occurred when trying to view rule errors on Manual Review.
547763	Skill versions that had active transactions could be automatically deleted if the "Keep 3 latest and versions currently in use" option was selected in the skill versions settings.
552661	Added the following error message: "Cannot import skills that already exist in the tenant.".
518518	ID Skill transaction failed when trying to process non-ID documents.
550786, 552997	User documentation inaccuracies were found.
534853	Some transactions in Skill Monitor displayed an incorrect status.
558004	It was not possible to create or delete fields in some skills.
546724	Japanese characters were not recognized in line items.
560275	Export skill operations failed to create an archive and were stuck in the Active state.
558307	Verification results were discarded for non-table fields with regions after applying the continue line items option.
553890	Incorrect skill description in the invoice guide on Marketplace.
555197	Uploaded password-protected document remained in the set.
565531	Classification Skill returned low-confidence results after upgrading to v. 2.3.2.
565600	If a copy of a built-in skill was retrained using certain document samples, the quality of the processing of other documents deteriorated considerably.
564520	Redundant underscores were added to recognized text when processing documents via OCR skill.

Zendesk Ticket #	Description
564817	Chinese characters typed in any field were not saved during Manual review.
565811	The "Column settings" dialog box did not open when choosing the "Column settings" menu item.
565655, 565838, 565910, 566706, 567499	When using the "Continue table from this row" option, the values in the line items column were displayed in an incorrect format.
567193	Filed values were extracted incorrectly for some documents.
564506	Redundant characters appeared in recognized text after performing OCR.
564807	Vantage automatically logged out users if IP restriction was enabled, even if their IPs were not black-listed.
473286	Ability to select all/deselect all barcode types has been added in the "Barcode Settings" dialog box.
565323, 564663	Some values were not extracted from line items when processing documents using the Document skill.
565776, 568843, 565196, 548497, 566393, 566778, 567758, 565779	Some numbers were not recognized on certain documents.
568260	If, after opening the table with line items for the first document, the operator opened another document in the batch, the values of one column could be moved to another column. This resulted in the values being in the wrong column in the export file.
569367	If a whole table was removed and then recreated during Manual Review clicking the "+" button inserted the second row above the first .
569615, 569615	A wrong Vantage version was displayed in the About dialog box.
568224	Skill export could fail if there was a large number of documents in some document set and multiple document sets were queried at once.

Limitations and known issues

In this release, there are the following limitations and known issues that will be addressed in future releases:

- General
 - The maximum number of files that can be added to a transaction is 1,000.

- If more than one file is added to OCR, Document, or Classification skill, they will be merged and processed as a single document.
- Each imported file should not exceed 2 GB.
- This release has only been tested with Google Chrome and Safari. Some issues may be encountered when using Firefox, Opera, and other browsers. These browsers will be fully supported in future releases.
- Automatic skill selection for the files uploaded into **Documents** currently only works for the following languages: English, French, German, and Spanish.
- Process skill
 - If a Process skill contains multiple **Classify/Extract** activities, online machine learning will be applied to the last activity.
 - **Review**, **Input/Output**, **Assemble**, and **For each document** activities cannot be wrapped in a **For each document** activity. Wrapping some of them in a **For each document** activity will be supported in future releases.
 - Each ZIP archive to be imported by ABBYY Vantage from a shared folder should not exceed 1.5 GB. Each file inside a ZIP archive should not exceed 100 MB. A ZIP archive must contain files in supported formats only, otherwise the import of the ZIP archive will fail.
 - Custom activity supports only synchronous Vantage API calls. A synchronous call implies that processing in the external system takes only a few seconds and that processing results are sent in the third-party system response almost instantly. Asynchronous methods will be implemented in future releases.
 - Process skills support no more than 50 Document skills in one **Extract** activity.
 - When importing from an e-mail, only first-level attachments are processed as transaction documents. If there is an attached e-mail file, its attachments will not be imported as files of the transaction.
 - When importing from an e-mail, the "FilePath" document registration parameter will not be specified, while the "FileName" document registration parameter will be specified.
 - Image fields cannot be exported to CSV. If there is an image field in the Document skill, the value of such field will always be empty in the export results.
 - During JPEG image processing with the Manual Review step, the image orientation may change if the exif value is not set to "topleft".
 - A "Missing Parent Extracted Object" JavaScript error occurred when adding an instance of a nested repeating group in Custom Activity.
- OCR skill and OCR activity in the Process skill
 - Selecting too many languages (more than 20) may slow down performance and impair the OCR skill.

- The maximum recommended number of pages to be processed with an OCR skill is 3,000 pages per file.
- In OCR skills, recognition of handwritten text is carried out only if the output format is set to either JSON/XML (text only) or TXT (either in combination with JSON/XML (text only)) or alone). To turn on handwritten text recognition, enable the **Handwritten** option in the **Text Appearance** section.
- Document and Classification skills
 - ABBYY Vantage can work with training sets of up to 10,000 documents.
 - Classification and Document skills support files containing up to 500 pages. Bigger documents may cause RAM issues and long processing times.
 - Classification skill is available in the following languages: Armenian, Azeri (Latin), Bashkir, Bulgarian, Catalan, Chinese Simplified, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Indonesian, Italian, Japanese, Kazakh, Korean, Latvian, Lithuanian, Norwegian (Bokmål), Norwegian (Nynorsk), Polish, Portuguese, Portuguese (Brazilian), Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, Tatar, Turkish, and Ukrainian.
 - When new documents are added into a trained Document or Classification skill, the **Results** tab will show outdated information until the new documents are labeled and the skill is retrained.
 - When comparing training results to reference in the Document skill, the **Difference** tab will not show the error type and reference value for repeating fields.
- Document Splitter skill
 - Reference labeling and statistics are not yet supported for the Document Splitter skill.
 - It is not possible to revert to a previous version of the Document Splitter skill.
- Online Learning
 - Online Learning does not support the following types of fields:
 - Image fields
 - Repeating groups containing image fields, nested repeating groups, groups inside repeating groups or tables
 - Generally, the results of Online Learning for checkmark fields are of a sufficiently high quality. However, sometimes checkmarks may not be detected or trained properly because of either the geometric form of the checkmark field (e.g. round or square) or the way the user chose to select the checkmark field (e.g. if the user crossed or underlined the field instead of placing a checkmark inside). Detection and training of such checkmarks will be improved in future releases.
 - For fields with handwritten text, ABBYY Vantage forms larger regions than those for fields with printed text. This may result in capturing and recognizing redundant adjacent text (e.g. field names).
 - The maximum number of documents is 10,000 for the training set and 1,000 for the testing set.

- Deep Learning in Advanced Designer
 - Deep Learning works for text fields only.
 - One Deep Learning activity can be trained for 50 fields, including repeating groups and fields, or for one table of up to 32 columns.
- Data catalogs
 - A data catalog cannot be deleted or changed if it is used in any skills, including any versions of a skill. To delete a data catalog that you do not need anymore, you must stop using it in the skill (i.e. by deleting or modifying the corresponding rule) and delete all the previous versions of the corresponding skill. Skill versions can be deleted in skill settings. If you do not want to delete old skill versions but want to change the data catalog, simply create a new data catalog.
 - If catalogs are uploaded to Advanced Designer for the Classify by Company activity, stricter requirements apply to the CSV file. It is not possible to upload a catalog to Advanced Designer if the CSV file does not meet the following criteria:
 - The number of data entries in each row must be equal to the number of columns.
 - A character that is used as a separator cannot be used inside the data (i.e., if your address data entries contain commas, the comma character cannot be used as a separator).
To overcome these limitations, modify your CSV file to meet the above criteria or upload your CSV file via Vantage.
- Skill Monitor
 - Transactions that have been created via the API and have not been started yet are not displayed on the **Skill Monitor–Transactions** page.
 - Only the latest 10,000 errors can be exported to a CSV file from the Error Log. You can use filters and search for specific lines in the Error Log to reduce the number of errors to be exported.
- Advanced Designer
 - If there is a large number of fields present in the **Manage Fields** dialog of the main menu, they extend beyond the visible part of the dialog area. When attempting to drag a field down the list while simultaneously scrolling, an error 901 occurs.
To overcome this limitation, avoid scrolling while dragging.
- Built-in skills
 - It is not possible to enable additional languages in copies of a Default Skill. Default Skills are intended to process documents in specific languages. Enabling additional languages in a Default Skill will not let you extract fields from documents in such languages. You can disable any of the supported languages, and we recommend disabling any languages that will not be used in your documents.
- Mobile Input
 - The capture of passports is not currently automated. Users should manually capture the document image once the document has been centered on the screen.

- Scanning Station

- When installing the Scanning Station without administrative permissions, it will be installed with the English UI, even if a different UI language was selected during setup.

To overcome this limitation, use one of the following workarounds: (a) secure administrative permissions prior to installing the Scanning Station; (b) reinstall the Scanning Station; or (c) modify the following registry key after installing:

[HKEY_CURRENT_USER\SOFTWARE\ABBYY\ScanStationVantage\5.0\Fine Objects]

"InterfaceLanguage"="X" where

X = 64 for Chinese (Simplified)

X = 3 for French

X = 2 for German

X = 6 for Italian

X = 26 for Japanese

X = 63 for Korean

X = 27 for Portuguese (Brazil)

X = 5 for Spanish

ABBYY Vantage © 2023 ABBYY Development, Inc.

ABBYY, ABBYY Vantage, Vantage are either registered trademarks or trademarks of ABBYY Development Inc. and/or its affiliates in the USA or other countries. These designations can also be logos, product or company names (or part of any of the above) of ABBYY Development Inc. and/or its affiliates and may not be used without consent of their respective owners.

Information in this document is subject to change without notice and does not bear any commitment on the part of ABBYY.

The software described in this document is supplied under a license agreement. The software may only be used or copied in strict accordance with the terms of the agreement. It is a breach of the United States copyright law and international laws to copy the software onto any medium unless specifically allowed in the license agreement or nondisclosure agreements.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or other, for any purpose, without the express written permission of ABBYY.