

# **End User License Agreement (EU- LA)**

*When renting the Konfuzio Server Software*

**Helm & Nagel GmbH**

Oct 19, 2024

## **Table of Contents**

<b>1</b>	<b>Scope</b>	<b>3</b>
<b>2</b>	<b>Subject matter</b>	<b>4</b>
<b>3</b>	<b>Granted rights</b>	<b>4</b>
<b>4</b>	<b>Property rights of third parties</b>	<b>5</b>
<b>5</b>	<b>Obligations of the customer to cooperate</b>	<b>5</b>
<b>6</b>	<b>Liability</b>	<b>6</b>
<b>7</b>	<b>Duty of care and audit right</b>	<b>7</b>
<b>8</b>	<b>Source code protection</b>	<b>8</b>
<b>9</b>	<b>Warranty</b>	<b>8</b>
<b>10</b>	<b>Intellectual Property Rights</b>	<b>9</b>
<b>11</b>	<b>References</b>	<b>10</b>
<b>12</b>	<b>Other</b>	<b>10</b>
<b>13</b>	<b>Anlagen</b>	<b>11</b>
13.1	System Requirements . . . . .	11
13.2	Service descriptions . . . . .	12

## 1 Scope

---

### END USER LICENSE AGREEMENT

between

the **Helm & Nagel GmbH** | |  
Rosenweg 5

35614 Aßlar

- hereinafter referred to as *manufacturer* -

and

the **customer**

- hereinafter referred to as the *Customer* - -

hereinafter referred to individually as a *party* and collectively as the *parties*.

The *Customer* results from the main contract, see § 1 para. 3.

## 1 Scope

1. The following terms and conditions govern the transfer of the Software as well as the rights to operate and use the Software by the Customer, who is not a consumer within the meaning of § 13 BGB (German Civil Code) and not an institution within the meaning of § 1 para. 1b KWG (German Banking Act) with its registered office in the European Union, including Switzerland and the United Kingdom, in case of direct purchase from Helm & Nagel GmbH, hereinafter referred to as the Manufacturer. In this case, this EULA contract supplements the lease relationship concluded between the Manufacturer and the Customer.
2. In addition, the following terms and conditions shall govern the transfer of the Software to the Customer by a third party (partner of the Manufacturer), none of which is a consumer within the meaning of Section 13 of the German Civil Code (BGB) and none of which is an institution within the meaning of Section 1 (1b) of the German Banking Act (KWG) with its registered office in the European Union, including Switzerland and the United Kingdom, and the rights to operate and use the Software. In this case, this EULA Agreement shall apply in addition to the agreement concluded between the third party and the Customer.
3. The lease concluded between the Manufacturer and the Customer or the contract concluded between the Third Party and the Customer shall hereinafter be referred to as the Main Contract.
4. In case the Software is leased by a third party, the Manufacturer shall not incur any obligation towards the Customer under the contract concluded with the third party.
5. Any object code or source code contained in the Software or Software Container and/or

otherwise transmitted by the otherwise provided by Manufacturer to Customer shall be considered part of the Software, and shall be subject to this subject to this Agreement and the applicable terms and conditions.

## 2 Subject matter

1. The subject matter of the Agreement is, on the one hand, the establishment of the rights of operation and use between the Manufacturer and the Customer of the Software, as well as all related data, media, printed materials and other electronic documentation and data, in the remainder of this document referred to as Software for short.
2. This agreement further includes all updates and enhancements to the Software provided to the Customer during the term of the lease. By installing, downloading, accessing, running or otherwise using the Software, Customer agrees to this Agreement.
3. Manufacturer may modify this Agreement by notifying Customer in writing of the modifications. Manufacturer agrees to notify Customer expressly and in writing of any changes to the terms and conditions whenever such changes are made. The Customer has a 14-day right of objection in this regard. The objection must be declared in writing. If the customer fails to object, the amendment shall become effective upon expiry of the one-month period. If the manufacturer fails to notify the customer, the change shall not become effective for the customer.
4. In the event of withdrawal or termination of the main contract, further use of the software is excluded. The customer must cease operation and use and delete the software from all storage media. In addition, the customer is entitled to claim back the paid purchase price pro rata temporis.
5. The manufacturer publishes the current version of these EULA contract terms on its website: [https://help.konfuzio.com/EULA\\_DE.pdf](https://help.konfuzio.com/EULA_DE.pdf).
6. The maintenance and support of the software is not part of this contract.

## 3 Granted rights

1. The customer receives the non-transferable, non-exclusive right to install and administer the software on one server, one cluster or several servers for a limited period of time during the contract period.
2. The customer receives the rights to use the software for the temporary transfer of the software, and to use it for processing documents. Documents are files in the sense of the [technical documentation](#). The use of this software may only be made through the interface of the software and documented interfaces. Any other use, any misuse of the software is not allowed. In case of misuse or other non-legal use of the software, which does not arise in causal connection with the direct use of the software, the customer is liable.
3. The customer is entitled to connect the provided software with other computer programs via defined REST interfaces, as well as to develop own processes.
4. The rights of use of the customer according to the point 3.2 exist also for the customer in the respective time in the sense of §§ 15 ff. AktG (German Stock Corporation Act) and other companies in order to provide services to their customers and/or group companies, hereinafter "organizations entitled to use". Entitled to use are companies provided that the customer informs the manufacturer in writing with the data of the organizations (company and address). Therefore, the Customer will inform the Manufacturer in writing about these Eligible Organizations and will guarantee that they will also comply with all contrac-

## 4 Property rights of third parties

---

tual obligations of the Customer towards the Manufacturer, and will conclude appropriate agreements with the Eligible Organizations in favor of the Manufacturer.

5. By using own servers, the customer is responsible for the availability and functionality. Since this system environment is operated autonomously by the customer, the manufacturer assumes no liability for this.
6. The customer has no claim to the surrender of the source code of the software.
7. The customer may not rent, sell, lend, sublicense, publicly reproduce or make available the software in any other way. publicly reproduce or make it accessible or make it available to third parties, be it against payment or free of charge.
8. The customer is not granted any rights to product or brand names of the manufacturer.
9. The software is protected by copyright laws and international copyright treaties and treaties and laws. The customer is aware that the exclusive rights remain with the manufacturer and that none of these rights (with the exception of the temporary transfer of the rights of operation and use according to points 3.1, 3.2, and 3.3) are transferred to him with this contract.
10. The Manufacturer is entitled to terminate the Contract without notice in case of violation of these provisions and to deprive the Customer of all rights to the Software. Other rights continue to exist. In this case the customer has to stop the use of the software immediately and completely, delete all copies of the software installed on its systems and, if applicable, delete the backup copy and the backup copy and the documentation or hand them over to the manufacturer.
11. All aforementioned rights of use and exploitation are granted to the customer under the condition precedent that he fulfills his obligations from the main contract and in particular has paid the remuneration.

## 4 Property rights of third parties

1. In the event of a claim against Customer alleging that the Software or any part thereof infringes any patent, trademark, copyright or trade secret of any third party, Manufacturer shall indemnify and hold Customer harmless if.
  1. Customer promptly notifies Manufacturer in writing of the claim;
  2. Manufacturer has sole control of the defense of the action and freedom to negotiate for indemnification or other settlement; and
  3. the customer does not take any action that impairs the defense of the action by the manufacturer.

## 5 Obligations of the customer to cooperate

1. The following obligations to cooperate are primary obligations of the customer and not to be classified solely as secondary obligations or obligations.
2. The customer shall name a person responsible for the system and a deputy to the manufacturer, who shall provide the information required for the execution of the contract and to make or arrange for decisions himself/herself. The system manager and his deputy are the contact persons of the manufacturer in all questions concerning the implementation of the contract. execution of the contract. The contact data (telephone and e-mail address) of the persons will be provided to the manufacturer by the customer. Customer at

the disposal.

3. It is incumbent on the customer to carry out proper data backups and to keep the software and hardware environment not software and hardware environment not covered by the service description. The customer shall protect the hardware and software against unauthorized access by employees or other third parties, viruses, Trojans and other malware. malware.
4. The customer is solely responsible for the contents and data processed with the software. The customer hereby undertakes to use the application only in accordance with the contract and within the scope of the applicable legal provisions and and not to violate any rights of third parties during use. The customer shall inform the manufacturer immediately, if possible in writing, if possible, of: (i) the misuse or the suspicion of misuse of the contractually agreed upon performance; (ii) a risk or suspicion of a risk to compliance with data protection or data security that occurs in the course of the provision of the contracted service; (iii) a risk or suspicion of a risk for the suspicion of a risk to the service provided by the Manufacturer, e.g. through loss of access data or Hacker Attack.
5. Customer shall take reasonable precautions in the event that all or part of the Software does not work properly (e.g. by daily data backup, fault diagnosis, regular check of the data processing results). data processing results).
6. Before commissioning or updating the software, the customer is required to test all functions of the contractual software under the customer's hardware and software environment and to check the documentation provided.
7. The customer is obligated to immediately report any errors occurring during the use of the contractual software to the manufacturer as a `ticket the manufacturer immediately as a 'ticket <<https://konfuzio.com/support/>>' and to indicate and describe how the defect manifests itself in each case, what its effects are and under which effects are and under which circumstances it occurs. In particular, this includes providing the manufacturer with defect manufacturer upon request in text form and to provide other data and protocols suitable for analyzing the of the defect.
8. For optimal use of the application, the customer will apply the browser Google Chrome or Microsoft Edge in the current version. In addition, the use of cookies must be allowed in the settings in the browser used. If these technical requirements are not met by the customer, there may be restrictions on the usability of the application. application may be restricted.
9. The customer bears disadvantages from a violation of these obligations. The manufacturer is not responsible for these restrictions.

## 6 Liability

The manufacturer and the customer agree that software cannot be produced free of defects according to the state of the art and accordingly agree on the following limitation of liability:

1. The manufacturer is liable in accordance with the statutory provisions for damages arising from injury to life, body or health, as well as for other damages, which are based on an intentional or grossly negligent breach of duty as well as fraudulent intent. In addition, Manufacturer shall be liable for contracts in accordance with the statutory provisions for damages covered by liability under mandatory statutory provisions, such as, for example, in the event of the guarantees, fraudulent concealment of a defect or under the Product Liability Act. Guarantees by by the manufacturer shall only be made in writing and, in case of doubt, shall only be interpreted as such if they are designated as a "right to warranty".

## 7 Duty of care and audit right

---

2. The limitation of liability applies in the same way in favor of the organs, other representatives, employees and vicarious agents of the manufacturer.
3. The manufacturer is not liable for loss of data - except for intentional or grossly negligent actions.
4. For other culpable injury of substantial contract obligations the manufacturer is responsible, equal from which legal grounds, only on the merits. The legal right of withdrawal remains unaffected, however, the manufacturer is only liable for the otherwise only in the amount of the typically foreseeable damage.
5. The strict liability of the lessor according to § 536 a para. 1 BGB (German Civil Code) for defects of the contractual at the time of the conclusion of the contract is expressly excluded.
6. Any further liability for damages than provided for in § 9 is excluded, regardless of the legal nature of the claim asserted. of the asserted claim - is excluded. This applies in particular to claims for damages arising from culpa in contrahendo, other breaches of duty contract, due to other breaches of duty or due to tortious claims for compensation for property damage in accordance with according to § 823 BGB.
7. As far as the liability for damages towards the manufacturer is excluded or limited, this also applies with regard to the personal liability for damages. the personal liability for damages of the employees, representatives and vicarious agents of the manufacturer. vicarious agents of the manufacturer.
8. As far as the manufacturer is liable according to the aforementioned regulations, the liability is limited to the amount of cover of the liability insurance of the manufacturer. Any further liability on the part of the manufacturer is excluded. The insurance coverage can be viewed at [exali](#) can be viewed.
9. The manufacturer is not liable in particular for the recovery of data, if these are not reproducible with justifiable reproducible with reasonable effort. The customer is responsible for data backup. The manufacturer is not liable for damages, as far as the customer could have prevented their occurrence by reasonable measures - in particular program and data backup. could have prevented.
10. Unless otherwise stipulated above, liability is excluded.

## 7 Duty of care and audit right

1. The customer is obligated to take appropriate measures to protect the software and the documentation from access by unauthorized third parties. unauthorized third parties, in particular to keep all copies of the software in a protected place. The employees of the customer are to be emphatically instructed to comply with these contractual terms and copyright law. In particular, the customer shall request his employees not to make any unauthorized make unauthorized copies of the software or the documentation.
2. If an employee of the customer violates the copyright of the manufacturer, the customer is obligated to cooperate to the best of his ability in the clarification of the copyright violation. to clarify the copyright infringement, in particular to inform the manufacturer immediately after becoming aware of the corresponding of the infringing acts in question.
3. The customer shall enable the manufacturer, at the latter's request, to check the intended use of the software, in particular to determine the purpose of the software, in particular whether the customer uses the program qualitatively and quantitatively within the scope acquired scope. To this end, the customer shall provide the manufacturer with information, grant access to relevant documents and relevant documents and records as well as to enable an examination of the hardware and software environment used. Provided that this

happens only once in a calendar year, within the usual office hours and after previous announcement at least two weeks in advance.

## 8 Source code protection

1. All rights to the source code of the Software, including the right to modify it, remain with the Manufacturer. The producer has the right to encrypt the source code of the software.
2. As far as the software is not provided with a copy protection, the customer is allowed to make one single backup copy for backup purposes only. The customer has to mark the created backup copy with the note "Backup copy" as well as a copyright notice of the manufacturer visibly on the created backup copy. A copyright notice copyright notice and registration numbers included in the software may not be removed. It is the Software or the written material in whole or in part in its original or modified form or in any other in its original or modified form or in a form mixed with other software or included in software, or to otherwise reproduce.
3. The customer is not entitled to make changes or interventions in the source code of the software, either by himself or by third source code of the software, not even to eliminate possible program errors. The latter does not apply if the manufacturer the elimination of program errors has been finally rejected in writing. If the customer in such a case or by third parties in such a case, any further warranty by the manufacturer shall expire. the manufacturer.
4. The customer is not entitled, outside the legal regulations, to modify the software or a part of it, to translate it, to change the modify, translate, reverse engineer, debug, disassemble, decompile, or otherwise modify the software or any part thereof. disassemble, decompile, or in any other way to explore or reproduce the source code, structure, sequences or Explore or reverse engineer the source code, structure, sequences or design of the Software or any part thereof.
5. The contractual software is delivered in executable form, i.e. in object code. As far as it is technically the software without the source code, the customer is not permitted to use the delivered software in any way software beyond the exceptions defined in this contract, or to process it in any other way.
6. The software is protected by copyright laws and international copyright treaties (§§69a et seq. UrhG). Copyright law is subject matter of this contract. References to copyrights or to other rights, which are on or in the software, may neither be changed, removed nor otherwise made unrecognizable. made unrecognizable. The customer shall be liable for all damages due to copyright infringements, which the manufacturer from a violation of these contractual provisions by the customer.

## 9 Warranty

1. The rights of the customer in the event of material defects and defects of title (hereinafter referred to as "Defects") of the Contract Software shall be governed by the statutory provisions shall apply, unless otherwise stipulated in the following.
2. The manufacturer warrants that the software for the intended use complies with its performance description and is not The manufacturer guarantees that the software corresponds to its performance description for the intended use and is not afflicted with defects that impair the suitability of the software for the intended use more than insignificantly. Insignificant deviations from the performance description shall not be deemed to be defects.

## 10 Intellectual Property Rights

---

3. The manufacturer warrants that its owed services are free of third party rights. Should this is not the case, the manufacturer will, at his discretion, provide the customer with the required the necessary rights of use or modify his performance in such a way that the rights of third parties are no longer affected. are no longer impaired.
4. Warranty claims of the customer shall become statute-barred within a period of one year from receipt of the notice of defect.
5. The warranty for material defects does not apply to defects that are due to the fact that the software is used in a hardware and software environment that is software environment that does not meet the requirements specified in the documentation and the release notes, or that is the documentation and the release notes, for use contrary to the terms of the contract, or for changes and modifications made to the software without being entitled to do so by law, by this agreement or on the basis of a prior written written consent of the Seller.
6. The Software is designed for a variety of uses and cannot cover every conceivable use in detail. every conceivable application in every detail. The manufacturer shall ensure, in accordance with criteria customary in the industry and within the reasonable framework, that the software complies with the performance description.
7. The customer has informed himself about the essential functional features of the software and bears the risk as to whether it his wishes and needs; he has to be informed about questions of doubt by employees of the manufacturer or by of the manufacturer or by competent third parties before conclusion of the contract.
8. If the manufacturer publishes a non-binding statement according to which the operation of the software is also under other system requirements, the manufacturer does not guarantee that the software will always run properly under the changed software will always function properly under the changed system requirements.
9. The customer is responsible for regular backup and maintenance of his individual data. The manufacturer points out that a data backup is mandatory, especially in case of warranty.
10. The manufacturer is entitled at his own discretion to remedy defects either by removal, by delivery of defect-free software or a new release of the software. software or a new release of the software and to reject an unreasonable type of remedy chosen by the customer. reject.
11. The customer agrees that the use of the software is not intended in connection with a high risk. Examples of use scenarios include but are not limited to air travel, space travel, fire fighting, Police operations, power plant operations, or power generation, transportation management systems, military operations, Rescue operations, medical operations. Customer agrees not to use or make available the Software in any place that or provide the Software in any place where the use of the Software could cause damage to property or injury to persons. persons.

## 10 Intellectual Property Rights

1. All intellectual property rights in and to the Software, including all images, animations, text of all printed documentation, and all copies of the Software are the property of the developer.
2. All intellectual property rights in and to the data and content transmitted to the customer, prospective customer, reseller, or third party through the use of the Software (for example, through image or file transfer) remain the property of the respective owners of the content. This Agreement does not grant you any rights to such data (except for the temporary

transfer of the rights of use under Sections 3.1, 3.2 and 3.3).

## 11 References

1. The manufacturer obtains the right to provide the customer and organizations entitled to use the data in the context of providing references publicly also by using the logo.
2. The manufacturer is aware of the reputation attached to the names of customers and organizations entitled to use them and undertakes not to use the aforementioned rights to the detriment of customers and authorized organizations.

## 12 Other

1. Verbal side agreements are not made. Amendments to these agreements must be made in writing, this also applies to also for the modification of this written form requirement.
2. Conflicting general terms and conditions of the customer shall not become part of the contract, even without express objection and even in the event of performance, unless this has been expressly regulated otherwise in the individual case.
3. If a provision is wholly or partially invalid, the manufacturer and the customer undertake to replace it wholly or partially with such a effective provision that comes as close as possible to the economic purpose of the wholly or partially invalid provision. The validity of the rest of the provision or of the other provisions shall not be affected thereby.
4. The customer undertakes, during this contractual relationship and for a period of two years after the the termination of the cooperation, to refrain from any enticement of employees of the manufacturer, who are active in the context of projects of the software himself or by third parties. In the event of a breach of this obligation, the Customer shall pay to the Manufacturer a contractual penalty of contractual penalty in the amount of € 250,000. The manufacturer remains entitled to claim damages. The contractual penalty shall be credited against this amount.
5. This contractual relationship is subject to the law of the Federal Republic of Germany.
6. Place of performance and jurisdiction for all disputes arising from this contract is Wetzlar in Germany.
7. The headings contained in this contract are for convenience only and are not relevant to the interpretation of the provisions of this contract. provisions of this contract.
8. Customer shall not be permitted to modify this Agreement or the rights and obligations hereunder, whether by operation of law or otherwise, without the prior written consent of Contractor, which consent shall not be unreasonably withheld. which consent shall not be unreasonably withheld.
9. Failure of the Contractor to exercise any right existing under this Contract shall not be deemed a Be construed as a waiver of such right.
10. If these contractual conditions are in a language other than German, these are translations. In the event of any inconsistency between these Terms of Use in German and the Terms of Use in any other language, the provisions in German shall prevail over the versions in the other language and shall supersede them in the event of any conflict.

# 13 Anlagen

## 13.1 System Requirements

### 13.1.1 Operating requirements & system environments for installation on site

The system environment includes three types of VMs. The Konfuzio Server software is run on the master VM. The speed of processing the tasks in the Redis task queue can be extended by including additional worker VM(s) in addition to the master VM, see also Performance under load. If text recognition (OCR) is required, at least one OCR VM must be operated per Worker VM.

#### Design of the Master VM

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM
- We recommend Redhat Linux as the operating system for the VM.
- All VMs require the AVX2 CPU command extension.
- PostgreSQL version 10 or newer is used as database (current stable version is recommended).
- Redis version 5 or newer is used as the task queue (current stable version is recommended).
- Each VM should be connected within the network with at least 1 Gbit/s
- Network storage for files with at least 1 TB storage space
- Internet connection is not required.

Technical instructions for installing Konfuzio Server software can be found [here](#).

#### Design of the Worker VM

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM.
- We recommend Redhat Linux as the operating system for the VM.
- All VMs need the AVX2 CPU command extension.
- Each VM should be connected within the network with at least 1 Gbit/s
- Read and write access to the network storage of the master VM
- Internet connection is not required

#### Design of the OCR VM (optional)

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM
- We recommend Redhat Linux as the VM's operating system.
- All VMs require the AVX2 CPU command extension.
- Each VM should be connected within the network with at least 1 Gbit/s
- Read and write access to the network storage of the master VM
- The use of [Tesseract 4.1.1](#) does not require an internet connection

- Using [On-Prem Container](#) requires an internet connection approximately every 100 minutes to report the number of pages processed to Microsoft. However, no other data is transferred during this process. More details can be found in the [Documentation](#).

### Performance under load

A system environments with one Master VM and one Worker VM process 3,000 pages per hour. A system environments with one Master VM and two Worker VMs process 6,000 pages per hour. The data describes the state when using Tesseract 4.1.1 <https://github.com/tesseract-ocr/tesseract/releases/tag/4.1.1> and provide for no training of the no training of the AI is performed at load time.

### Development / test system

The following is the design of development or staging servers to enable a development/test system independent of operation in production.

- 1 VM for databases, data storage and Konfuzio server (each for development and test) + all tasks of the development system. Tasks refer to tasks from the “task queue” such as preprocessing, classification, extraction and training.
- 1 VM for OCR (development and test)

## 13.2 Service descriptions

### 13.2.1 Preamble to the standard software

With the Konfuzio software, information from documents is automated, simplified and retrievable at any time in a structured manner. Documents from various business processes can be integrated and consolidated. The software serves as a platform and offers different components for the simple and fast processing of individual documents of any kind as well as for the individual structuring of documents. The software serves as a platform and offers different components for the simple and fast processing of individual documents of any kind as well as for the individual structuring of the information contained therein. It is standard software that is designed for use by a large number of customers. The software must therefore be individually adapted and processed by the client in each case. The overall responsibility for the introduction of the software lies with the customer. The manufacturer of the software mentioned below is Helm & Nagel GmbH.

The Konfuzio software currently consists of three modules.

### 13.2.2 Konfuzio Server

The service-oriented architecture of Konfuzio provides an AI web service for processing documents. The results of the document processing are provided via multi-client REST API services in JSON format. The currently documented functions of the API are available at <https://app.konfuzio.com/api/>. The application differentiates users according to roles and offers the possibility to configure Create, Read, Update and Delete (CRUD) permissions. In simple terms, Konfuzio processes documents in three steps:

#### Text recognition in scans and images through OCR

When loading documents, documents are loaded into Konfuzio via the REST API. Depending on the incoming quality of the documents, technical correction procedures are used for damaged files and then OCR for full-page text recognition. The OCR engine used can be freely selected by the customer. By default, the Open Source OCR Engine

Tesseract 4.1.1 is installed as standard. With use of the Tesseract OCR does not incur any further costs for the customer. Other OCR engines are purchased separately. The manufacturer provides the customer with connectors in order to use the OCR engine separately for each project. A unique ID is generated for each document. Supported input formats, see [Documentation](#) are saved as archivable PDF documents (PDF/A) including embedded text layer. The originals of the uploaded file and the PDF/A generated from it can be retrieved via REST-API.

In the version of the Konfuzio Server hosted on [app.konfuzio.com](http://app.konfuzio.com), Helm & Nagel GmbH uses the Azure Read API 3.2. If the customer wants to ensure the same results of the text recognition in his own installation the customer is recommended to use the OCR of Azure as an [On-Prem as an `On-Prem Container](#) or as a [REST API](#). to be purchased. If the customer so wishes, Helm & Nagel shall provide this OCR engine and charge for it separately. The price for use shall be based on Microsoft's prices. These can be obtained for the REST API from [Microsoft](#) and for the on-prem [Container from Microsoft](#) can be viewed. The price without quantity discount applies.

### **Categorising and later extracting the individual pieces of information**

Categorising and extracting documents: Each incoming document is assigned to a class by a supervised learning model. The exact procedure for classification does not require any manual rules and goes far beyond the keyword-, phrase-, layout- or graphic-based classification. The classification gives per class per incoming document a confidence value. Classes are configured and trained by users. For each class information can be extracted for each class, if desired. The extraction of individual information in the context of tables, unstructured free text or by the layout of the document is only possible by defining training documents. For this, users adapt the AI through training documents. For each training run and category automatically trained on the basis of the training documents, stored and then used for inference. During inference, a confidence value is output for each piece of information. The recognition of the individual pieces of information is made possible by the use of labels, see below. If annotated in the training documents the context of individual pieces of information is learned by the AI. Thus, a single piece of information, e.g. first name, occur several times in the document. For example, one of the two recognised first names can be assigned to the recipient and one to the sender. The recognition of the context of the individual piece of information is made possible by the use of label sets, see next page. The manufacturer evaluates the latest AI research on an ongoing basis and includes further AI models in the product range. models into the application after positive test results.

### **Use of the data via REST API**

After loading and extracting the documents, the contents of the documents are made available as structured data as REST API in the JSON format. The data can be retrieved with the document ID. In addition, it is possible to webhook for each document, which actively sends the structured data to a previously defined service after the defined service after the processing process. Feedback on AI results from classification, extraction and context recognition can be given by authorised users through the web-based SmartView. This SmartView provides direct access and synchronous display of the recognised information via document ID. Through feedback, the quality of the AI is continuously improved. In addition, new classes, individual pieces of information or contexts can be trained to the AI in this way.

The software Konfuzio is designed to classify any type of document and to display information in the professional context on the basis of the document type. This generic applicability of the system is made possible by three essential elements of the software.

**Category:** Each incoming document is assigned to a document type, a so-called category, assigned. A document has a document type. If extraction is desired, a category can be supplemented with label sets. can be added to a category.

**Label set:** A label set is a bundle of labels. A label set can be used to recognise tables or to extract individual information in the subject context. extract individual pieces of information in the subject context, e.g. that of the sender of a letter.

**Label:** Labels define the individual pieces of information to be extracted from a document. For each label, an auto typing, e.g. conversion into machine-readable date formats.

With these three modules, users build up a comprehensive data set that can be used to apply AI with supervised learning. methods in both classification and extraction through sample documents. Data for both initial and ongoing training are applied via the web browser-based Konfuzio SmartView per point and click in the document by users. in the document by users. Once information has been saved, it can be accessed via an individual URL in the SmartView. In addition, each individual piece of information can be accessed via a unique URL directly in the SmartView. directly in the SmartView.

The application offers extensive logging. From import to export, the technical processing steps are logging of the technical processing steps takes place for each document. This view can be viewed by users authorised as superusers. authorised as superusers. The logging at module level, e.g. the classification of a document, is accessible through the tasks in the Redis messaging system.

In addition, these three modules make it possible to build up a well-ordered database and use it for both technical and content reporting purposes. technical as well as content-related reporting purposes. Konfuzio offers standard reports and allows for the individual reports: prefabricated reports can be downloaded directly from the application for each trained AI model and project as a CSV. directly from the application. Individual reports can be created using the free Konfuzio Python SDK or MS Excel Power Query.

The application is multi-client capable. An AI model can be used in different projects. Users with separately configurable roles can be invited to a project. One API endpoint is available per project. An AI model can be made available to different authenticated groups of users after training or retraining. users after training or retraining.

In addition to Konfuzio's internal reporting options, Konfuzio is operated in a Kubernetes environment. This allows a comprehensive control of the technical operation. The continuous export of reporting-relevant data enables end-to-end reporting.

### **Operating requirements & system environments for installation on the customer's servers (on-prem / (private) cloud)**

The following is a system design with 3 Konfuzio instances (DEV, TEST, PROD) in order to implement a development/test system independent of the operation in production.

The system environment comprises three types of VMs. The Konfuzio server software is run on the master VM. The speed of processing the tasks in the Redis task queue can be increased by integrating additional worker VM(s) in addition to the master VM, see also Performance under load. If text recognition (OCR) is required, at least one OCR VM must be operated per Worker VM.

#### *Design of the Master VM*

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM
- We recommend Redhat Linux as the VM's operating system.
- All VMs require the AVX2 CPU command extension.
- PostgreSQL version 9.5 or newer is used as the database (version 11.11 is recommended).
- Redis version 4 or newer is used as the task queue (version 4.0.9 is recommended).
- Each VM should be connected to the network with at least 1 Gbit/s.
- Network storage for files with at least 1 TB storage space

- Internet connection is not required.

The technical instructions for installing the Konfuzio Server software can be found [here](#).

### *Design of the Worker VM*

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM
- We recommend Redhat Linux as the VM's operating system.
- All VMs need the AVX2 CPU command extension.
- Each VM should be connected within the network with at least 1 Gbit/s
- Read and write access to the network storage of the master VM
- Internet connection is not required

### *Design of the OCR VM (optional)*

- Resources: 8 vCPU (min. 2.6 GHz) and 64 GB RAM
- We recommend Redhat Linux as the VM's operating system.
- All VMs require the AVX2 CPU command extension.
- Each VM should be connected within the network with at least 1 Gbit/s
- Read and write access to the network storage of the master VM
- The use of [Tesseract 4.1.1](#) does not require an internet connection
- Use of [On-Prem Container](#) requires an internet connection approximately every 100 minutes to report the number of pages processed to Microsoft. However, no further data is transferred during this process. Further details can be found in the [Documentation](#).

### *Performance under load*

A system environment with one Master VM and one Worker VM processes 3,000 pages per hour. A system environment with one Master VM and two Worker VMs processes 6,000 pages per hour. The specifications describe the state when using [Tesseract 4.1.1](#) `<https://github.com/tesseract-ocr/tesseract/releases/tag/4.1.1>` and assume that no training of the AI is carried out at load time.

### *Development / test system:*

- 1 VM for databases, data storage and web server (each for development and test) + all tasks of the development system. Tasks refer to tasks from the "task queue" such as preprocessing, classification, extraction and training.
- 1 VM for OCR (development and test)

## 13.2.3 Konfuzio Trainer

The service description can be found in the technical documentation at [https://dev.konfuzio.com/training/training\\_documentation.html](https://dev.konfuzio.com/training/training_documentation.html).

## 13.2.4 Confucius Python SDK

The Konfuzio Python SDK provides a free and MIT-licensed extensible Python API that allows data scientists and developers to access and interact with the Konfuzio server. developers to access and interact with the functionality of the Konfuzio server. The Konfuzio Python SDK works independently of the chosen hosting concept of the Konfuzio Servers.

A common use case is the complete download of all data available on the Konfuzio Server to which the user has access. the user has access to. This kind of download enables a complete and self-sufficient data backup or transfer of the transfer of data to the customer's server. With

a good internet connection, thus a download speed of more than 200 MBit/s, the following storage space requirements and duration for the download can be expected:

- The text of the document requires a storage space of approx. 0.05 MB per page at a download speed of 26,000 pages per hour
- An additional 1 MB per page is required if the optical properties, so-called bounding boxes, are to be saved down to the individual letter. individual letter are to be saved. The download speed is approx. 16,000 pages per hour.
- A further 0.125 MB per page is required if the archivable OCR version of the PDF is also to be saved. These files can be downloaded at approx. 48,000 pages per hour.
- A further 0.15 MB per page is required if each page is to be saved as an image. This download is possible at a speed of 16,000 pages per hour.

Technical instructions on how to use the Python SDK can be found at [dev.konfuzio.com](https://dev.konfuzio.com).

The service description can be found in the technical documentation at [https://dev.konfuzio.com/sdk/configuration\\_reference.html](https://dev.konfuzio.com/sdk/configuration_reference.html). If the Konfuzio Python SDK is not included in the is not included in the maintenance contract, the MIT licence applies. This can be on [GitHub](#).

The MIT licence offers a decisive advantage for commercial use. obligation to integrate the corresponding copyright notice together with the licence text into one's own application. own application. In this respect, the implementation or modification of MIT-licensed software is also not unconditional; this However, this single hurdle is low from a legal and practical point of view. Unlike many open source licences with copyleft clauses, proprietary applications can be developed or distributed with the Konfuzio Python SDK. Here It is up to the user under which conditions the proprietary application is published.