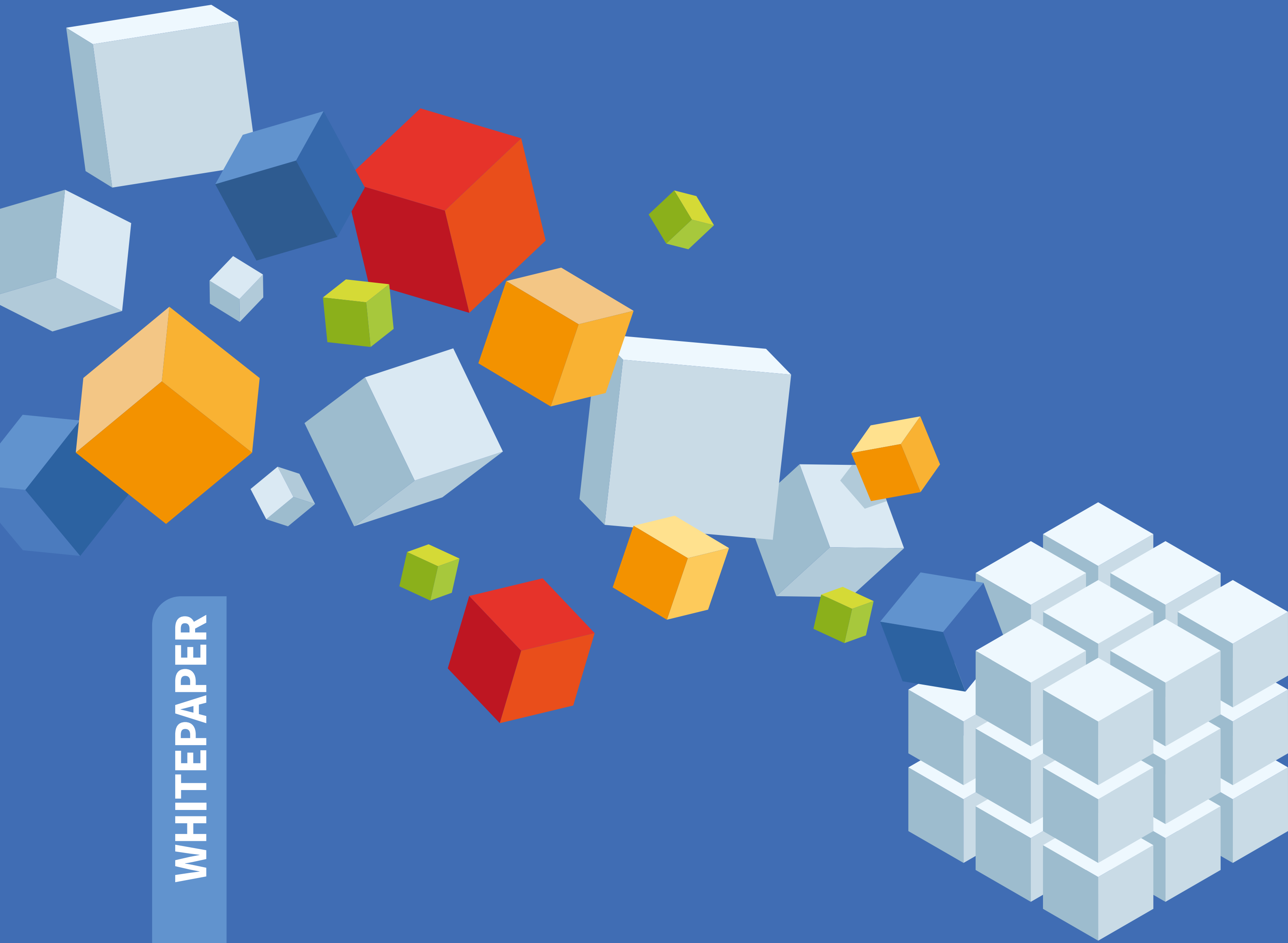
THE CONSOL PRIVACY PURGER: ADVANCED PII DETECTION AND CLEANSING



WHITEPAPER

DEFINITIONS

ConSol CM: A workflow-centric application platform developed by ConSol Software GmbH. This versatile platform not only underpins the ready-to-use solutions like CM/Ticketing, CM/Complaint, and CM/Helpdesk but also serves as a foundation to tailor individual business processes or service scenarios.

ConSol CM/AI Assist: An addon for ConSol CM that seamlessly integrates ChatGPT - or other Large Language Models - as a virtual assistant. Its prime function is to boost the efficiency of case handlers and amplify the automation capabilities of the platform.

PII or Personally Identifiable Information: Refers to any data that can be used to identify a specific individual, such as name, address, or social security number.

Privacy Purger: Developed by ConSol Software GmbH, this software component cleans input texts of PII, rendering them depersonalised to be safely shared with external parties like LLMs. While it's a notable feature of the ConSol CM/AI Assist subscription, it can also be availed as a standalone license. This whitepaper delves into its nuances.

INTEGRATING LLMS: BENEFITS AND CHALLENGES

Benefits of Integrating an LLM Into a Workflow System

Integrating Large Language Models (LLMs) like ChatGPT into service systems such as ConSol CM can really improve how things work. Our ConSol CM/AI Assist addon brings in ChatGPT to help case handlers work better and faster. This means they can deal with more cases in less time. Plus, with ChatGPT, the platform can do more things automatically. With AI Assist, ConSol CM can:

SUMMARIZE EMAILS

Pull out the main point from customer emails, so case handlers read less and can act faster.

CATEGORIZE EMAILS

Classify emails into relevant service categories. This automation ensures that emails are directed to the appropriate teams or even processed autonomously when suitable.

SPLIT CASES

If an incoming message touches on multiple issues, the system can intelligently split it into distinct cases, enhancing organizational efficiency.

SUGGEST RESPONSES

Generate contextually appropriate reply suggestions, allowing case handlers to focus on nuanced problem-solving rather than drafting routine responses.

SUMMARIZE CASES

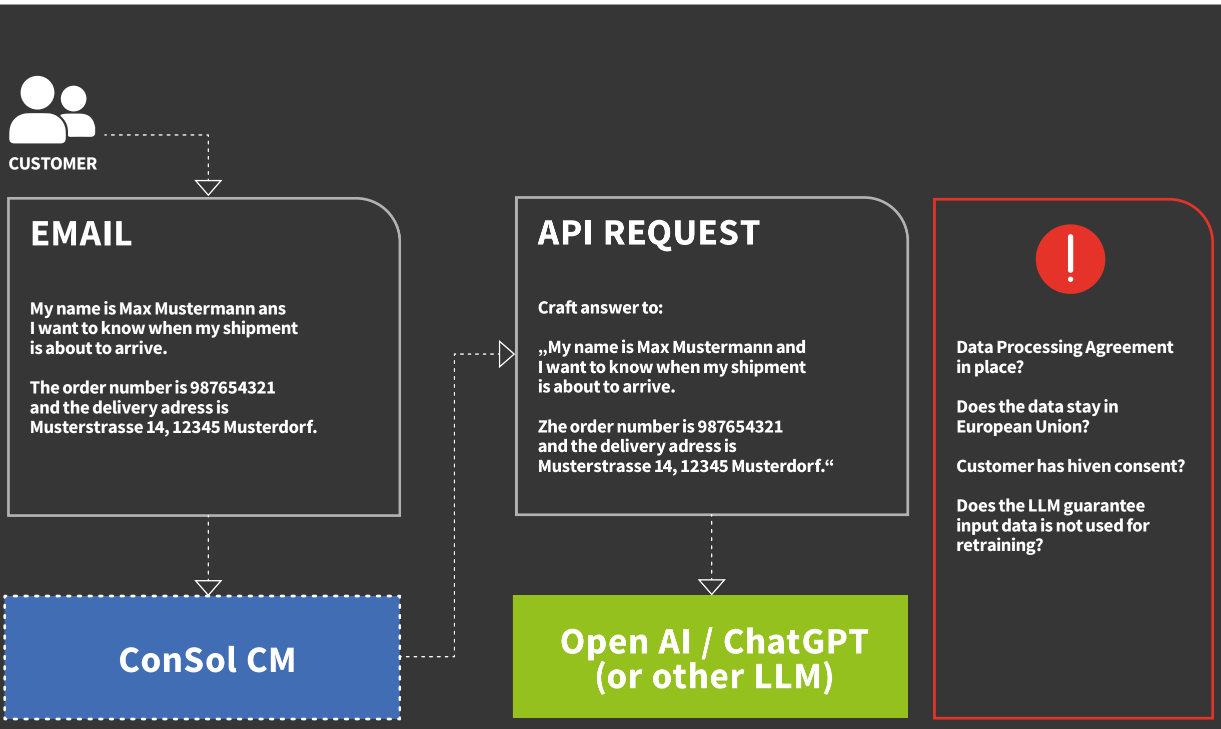
After the conclusion of a case, a summary is automatically crafted for easy future reference, eliminating the manual task of creating one.

However, while the integration of LLMs offers these transformative advantages, the concern for personal data privacy under the GDPR looms large. This is where our Privacy Purger steps in.

The Dilemma: Transmitting PII to LLMs

The advantages derived from the described use cases are evident. However, a significant hurdle looms: For the stated scenarios to work, ChatGPT requires access to the actual correspondence related to the cases. Whether these are emails, documents, or portal messages, the inherent risk is that they might contain Personally Identifiable Information (PII). When this PII is shared with ChatGPT, it presents potential GDPR compliance challenges. According to GDPR stipulations:

- A Data Processing Agreement (DPA, German: AVV) must be established between OpenAI and the entity supplying the data
- The data should remain exclusively within the boundaries of the European Union (EU)
- Explicit consent must be obtained from each individual customer whose data is shared, usually via acceptance of terms of service



CUSTOMER

EMAIL

My name is Max Mustermann ans
I want to know when my shipment
is about to arrive.

The order number is 987654321
and the delivery adress is
Musterstrasse 14, 12345 Musterdorf.

ConSol CM

API REQUEST

Craft answer to:

„My name is Max Mustermann and
I want to know when my shipment
is about to arrive.

Zhe order number is 987654321
and the delivery adress is
Musterstrasse 14, 12345 Musterdorf.“

Open AI / ChatGPT
(or other LLM)

Data Processing Agreement in place?

Does the data stay in European Union?

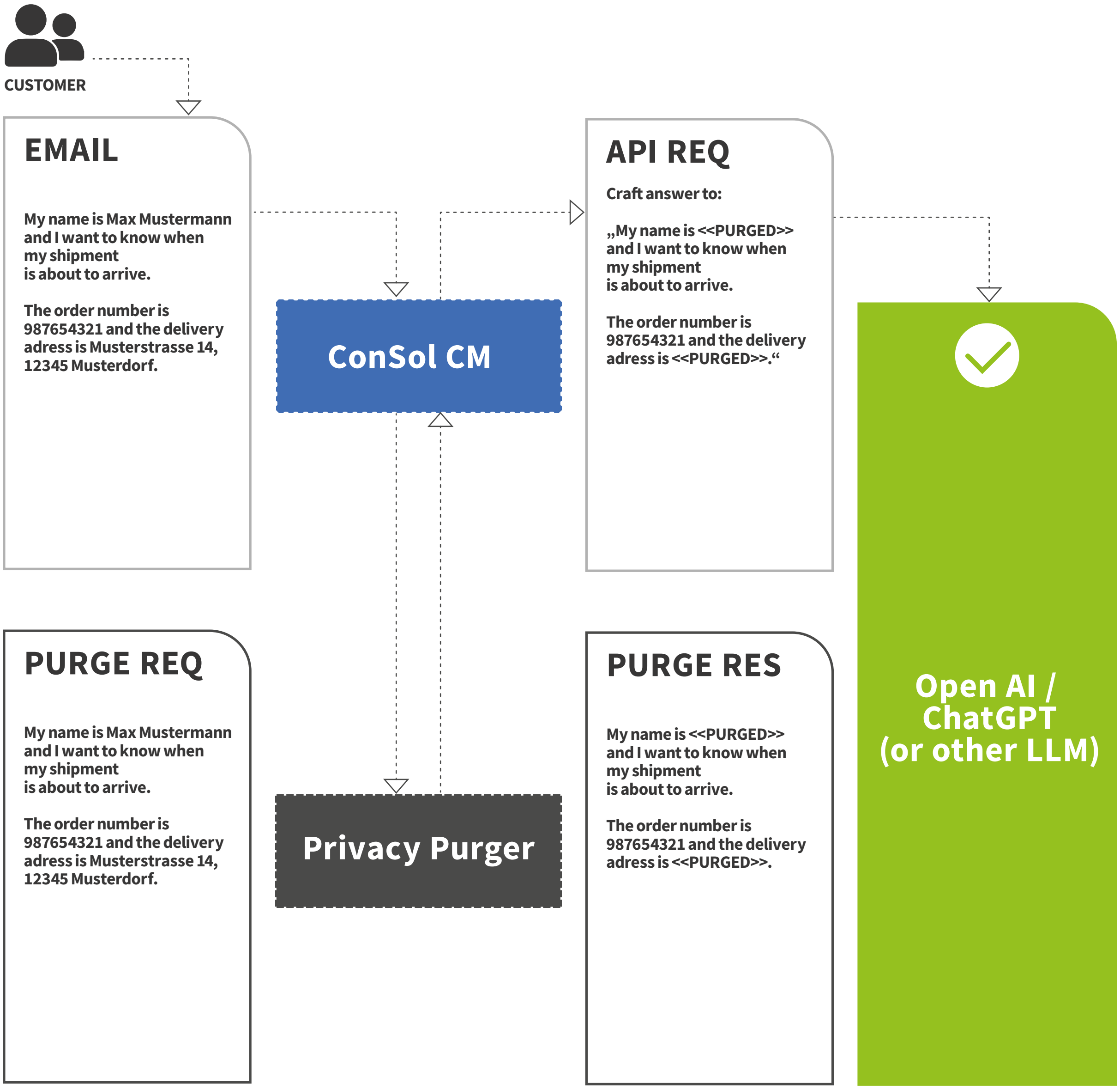
Customer has hiven consent?

Does the LLM guarantee input data is not used for retraining?

At the time this whitepaper was drafted, OpenAI does not facilitate the choice of data centers specifically within the European Union, rendering one of the GDPR requirements unfulfilled. Recently available enterprise versions of ChatGPT provide at least the possibility to sign a DPA with OpenAI. External ChatGPT providers, such as Microsoft Azure's offering, may offer more stringent guarantees, but come with the caveat of introducing another cloud entity to potentially sensitive PII data. Furthermore, a lingering concern with hosted LLMs like ChatGPT is data usage: only on enterprise plans there is concrete assurance that the input data isn't leveraged for refining the model. This raises the unsettling possibility of actual PII surfacing in subsequent model outputs, whether through direct recall or inadvertent generation.

The Resolution: Privacy Purging to Uphold Data Confidentiality

Given the challenges we mentioned, we created the Privacy Purger. This tool lets ConSol CM users get all the benefits of using LLMs, without the headache of legal and data concerns. Simply put, before any message goes to ChatGPT, the Privacy Purger removes any personal info from it. Even without this data, the main message and its purpose remain clear, so ChatGPT can still understand and work with it, ensuring both efficiency and privacy.



EMAIL

My name is Max Mustermann and I want to know when my shipment is about to arrive.

The order number is 987654321 and the delivery adress is Musterstrasse 14, 12345 Musterdorf.

API REQ

Craft answer to:

„My name is <<PURGED>> and I want to know when my shipment is about to arrive.

The order number is 987654321 and the delivery adress is <<PURGED>>.“

PURGE REQ

My name is Max Mustermann and I want to know when my shipment is about to arrive.

The order number is 987654321 and the delivery adress is Musterstrasse 14, 12345 Musterdorf.

PURGE RES

My name is <<PURGED>> and I want to know when my shipment is about to arrive.

The order number is 987654321 and the delivery adress is <<PURGED>>.

HOW WE PURGE DATA: METHODS AND TECHNIQUES

The Privacy Purger employs a two-fold approach to detect and eliminate PII data: the precision of Regular Expressions for easily discernible data types, complemented by the advanced capabilities of a Machine Learning technique known as Named Entity Recognition (NER) for more intricate data like names and addresses.

REGULAR EXPRESSIONS (REGEX)

Regular expressions (often abbreviated as "regex" or "regexp") are sequences of characters that define a search pattern, primarily used for string matching and manipulation. They provide a powerful way to identify and process strings in texts. RegEx patterns can range from simple matches, like finding specific words, to complex patterns that identify email addresses or URLs.

NAMED ENTITY RECOGNITION (NER)

Named Entity Recognition (NER) is a technique in computer language processing that helps identify and categorize specific names and terms in a text, like names of people, places, companies, or dates. Think of it as a tool that picks out important bits from a piece of writing, making it easier to understand and organize. It's often used to sort information in news articles, research papers, or any large set of written content.

Together, these tools give the Privacy Purger its ability to reliably find and remove any PII from different types of input. Here's a list of the data types we purge and the method we use:

INFORMATION TYPE	PURGE TECHNOLOGY
Names (including first names and last names)	NER
Street addresses	NER
Cities	NER
Zips	regex
Email addresses	regex
Birthdates	regex
"Postfach" addresses (German specific)	regex
IBAN	regex
Phone numbers	regex

FRAMEWORKS & SYSTEM DESIGN

Used Frameworks

Our Privacy Purger uses a mix of regular expressions and a Machine Learning technique called NER. Since the NER functionality is so crucial we tested several NER libraries to find the best fit for our needs.

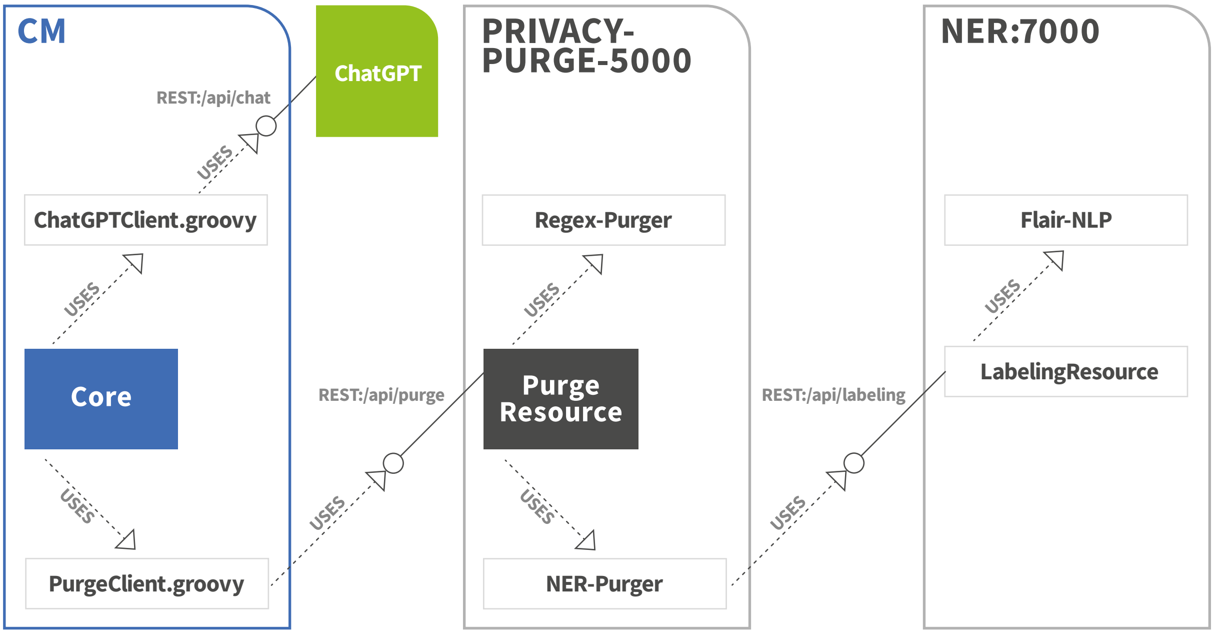
After careful evaluation, we picked the Flair NLP framework (https://www.informatik.hu-berlin.de/de/forschung/gebiete/ml/Flair/flair). It's developed by the Humboldt University Berlin and is a well-respected tool in the Python ML community. It's also open-source under the MIT license.

Additionally, we use Spring Boot from Java, as well as Flask and Waitress from Python.

System Overview

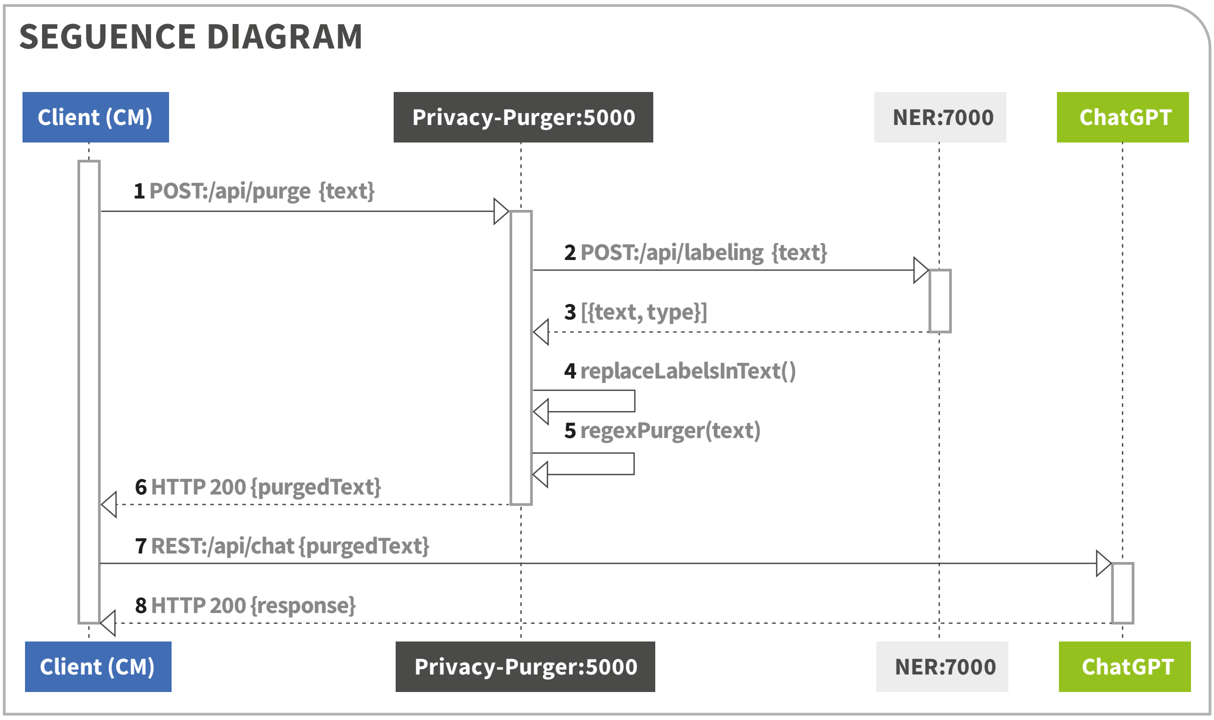
The Privacy Purger's architecture is modular, comprising two Microservices. Both offer a REST interface on different ports for easy communication. The main service is implemented in Java Spring Boot and provides the API entry point used by ConSol CM. Once called with an input text, the main service calls the NER service via REST, which is implemented in Python and performs the labelling of names and addresses.

This component diagram shows the interplay of the components and their communication paths.



The feedback from the NER service forms the basis for substituting names and addresses with generic placeholders. Following this, the engine invokes the regular expressions purger sequence, applying all predefined regular expressions to the remaining text.

The following sequence diagram illustrates the sequence of calls and operations for a purge call.



DEPLOYMENT & SYSTEM REQUIREMENTS

Deployment Options and Format

The Privacy Purger offers flexibility in deployment. If you use ConSol CM on premise, it's often best to also run the Privacy Purger in the same environment for top-notch data security. However, for those using the ConSol CM Cloud service, we'll provide a centrally managed Privacy Purger on our cloud cluster.

Our Privacy Purger is published as a set of two container images for deployment on a Docker or Podman runtime. So if you want to run it on premise, you will need one server or VM with **Podman** or **Docker** runtime installed.

If you're familiar with container platforms like Kubernetes or OpenShift, you'll be glad to know the Privacy Purger plays well with them. Just keep in mind, depending on your setup, some additional expertise and configurations might be required.

Hardware Requirements

Memory/RAM: At least 8 GB
CPU: No strict limitation, but strong processing power is recommended since the ML part will benefit from it. At the bare minimum, ensure you have 2 cores.

FREQUENTLY ASKED QUESTIONS

Can you guarantee full GDPR compliance with this approach?

Yes, our approach is considered fully compliant with the mandatory requirements of the GDPR for our application ConSol CM.

Does it work with English or other languages?

The Privacy Purger uses a NER model explicitely trained with German texts for best performance on German language. However, Flair provides also models specific to English language which could be easily switched. We might consider adding a configurability here in future.

Can we extend it to also purge additional data?

Yes, the regular expression engine of the Privacy Purger can be extended by pure configuration. If your use case requires cleansing of additional specific data we can extend this easily.

Can we run the Privacy Purger on premise?

Yes. All that is needed is a server with either Docker or Podman runtime.

Do you offer a hosted variant?

ConSol CM cloud customers can use an instance provided directly on our cloud. Other hosted options can be discussed on an individual basis.

Revision history

REVISON	DATE	AUTHOR	COMMENTS/NOTES
0.1	30.10.2023	Jan Zahalka (Product Manager ConSol CM)	Initial version
0.2	03.11.2023	Robert Niedermeier (Cyberlegis RA GmbH)	Legal Review of GDPR Compliance
1.0	02.02.2024	Jan Zahalka (Product Manager ConSol CM)	Small enhancement to reflect newest OpenAI enterprise offerings.

ConSol
Consulting & Solutions Software GmbH

St.-Cajetan-Straße 43
D-81669 München
Tel.: +49-89-45841-120
cm-vertrieb@consol.de
https://cm.consol.de

Für mehr Unternehmens-Infos,
Tech News und Einblicke in unsere Kultur:
Folgen Sie uns gerne auf LinkedIn.