

ayfie

Q&A

**EFFECTIVE WAYS TO
RETRIEVE PERSONALLY
IDENTIFIABLE
INFORMATION (PII) FOR
COMPANIES**

GDPR Webinar from July 24, 2018

What is the difference between structured and unstructured data?

Structured data is usually coming from a database (columns like first name, last name, etc.). Unstructured data is more of a natural language type of documents with sentences (Word Document, PDF, PowerPoint, etc.) and not data that is structured in a format like Excel, CSV or so. We bring structure to unstructured documents. ayfie finds PII in unstructured data.

Where does ayfie Supervisor get the data?

It could be any data source that exists within your company. We have over 70 different types of connectors that we sell alongside our search platform, which is the foundation of the GDPR reporting tool. While we have specific connectors that connect to a particular system like Sharepoint, Exchange or other enterprise level systems. We also provide more general purpose connectors. For instance, connecting to relational database type systems, which indexes the data in and makes it searchable.

How exactly do you identify PII?

People create about 80% of data in an unstructured manner, and there is a lot of PII in that data. For instance, a Support Agent makes notes about a customer call with personally identifying information, which with traditional systems would not be captured - our technology reads through the content and uses a mixture of large proprietary dictionaries and context respecting mechanisms to disambiguate certain phrases and terms. This way we can provide a higher probability for a specific data points, e.g., a phone number or a credit card number.

So are you using regular expressions?

It's a combination of different mechanisms. Some information can be identified through regular expressions (e.g., credit card numbers, that have a specific format), but in many cases, it goes beyond the possibilities of regular expressions, due to semantic and linguistic contexts. That's where our automatic entity extractors, which are more advanced, come into play.

Is there a demo version that my company can try?

Please go to <https://www.ayfie.com/solutions/knowledge-discovery/free-trial/> to try ayfie Locator (the company-wide search platform) for 30 days. Send us an email if you want to try ayfie Supervisor (the GDPR solution), that sits on top of ayfie Locator.

What makes ayfie Supervisor better than other solutions?

At ayfie, we focus on the very difficult part of the technology which is the unstructured data. To give companies the most control over their unstructured data, we leverage OCR detection and analyze about 300 different file formats across multiple data sources.

ayfie Supervisor has been built on more than 30 years of linguistic analysis experience.

Is there an API to be able to pull data related to data subject when requests for data come?

ayfie products come with a full set of RESTful APIs to control every aspect of platform programmatically.

If the documents contain multiple data subject, does PII information attach to different data subject?

If we understood this correctly, the question is about processing multiple PII datasets (subjects) in one document. This is something that ayfie Supervisor can do very well. Every fraction of PII in a document gets analyzed and extracted. E.g., if a document contains hundreds of different name entities, all of them would be subject of a report.

Does the system convert protected pdf documents?

The ayfie converter respects the security guidelines of every document. If a PDF file is password protected, the platform will not convert the document, like the author most likely intended.

If the supervisor can view sensitive PII data of the employees and customers - isn't that a GDPR violation in itself?

ayfie Supervisor is equipped with a restricting security mechanism to make sure only eligible people get access to the generated GDPR reports. The intention of a GDPR report requires software solutions to manage data in one place, to report on it. The regulation dictates to hold copies of created reports and thus expects companies to gather sensitive PII in one place.

You used the term “practically read” what does that mean? How does your company process data differently from others claiming to be able to help with GDPR?

By “reading” we mean, that we look for specific terms and phrases in sentence structures, to extract semantic entities. This extraction of such entities under consideration of the context can be considered reading.

For the second part of your question, please refer to the USPs above.

How long does it take to install Supervisor and begin using?

ayfie Supervisor is designed to easily work out of the box, after a short installation. Depending on the amount of content you're indexing, you can be up and be operating within the same day.

Is the solution scalable?

ayfie solutions are designed to scale up to many millions of documents. It allows being set up in a distributed manner, in order to spread the database and search index across multiple hardware nodes. This way it's entirely geared for the most demanding environments.

Which languages does the frontend support?

Bulgarian, Czech, Danish, German, Spanish, French, Italian, Dutch, Norwegian, Polish, Portuguese, English, Russian, Swedish and Ukrainian.

Contact

For more information please do not hesitate to contact us at the addresses given below:

Heather Murphy · heather.murphy@ayfie.com

Senior Director of Inbound Marketing

www.ayfie.com