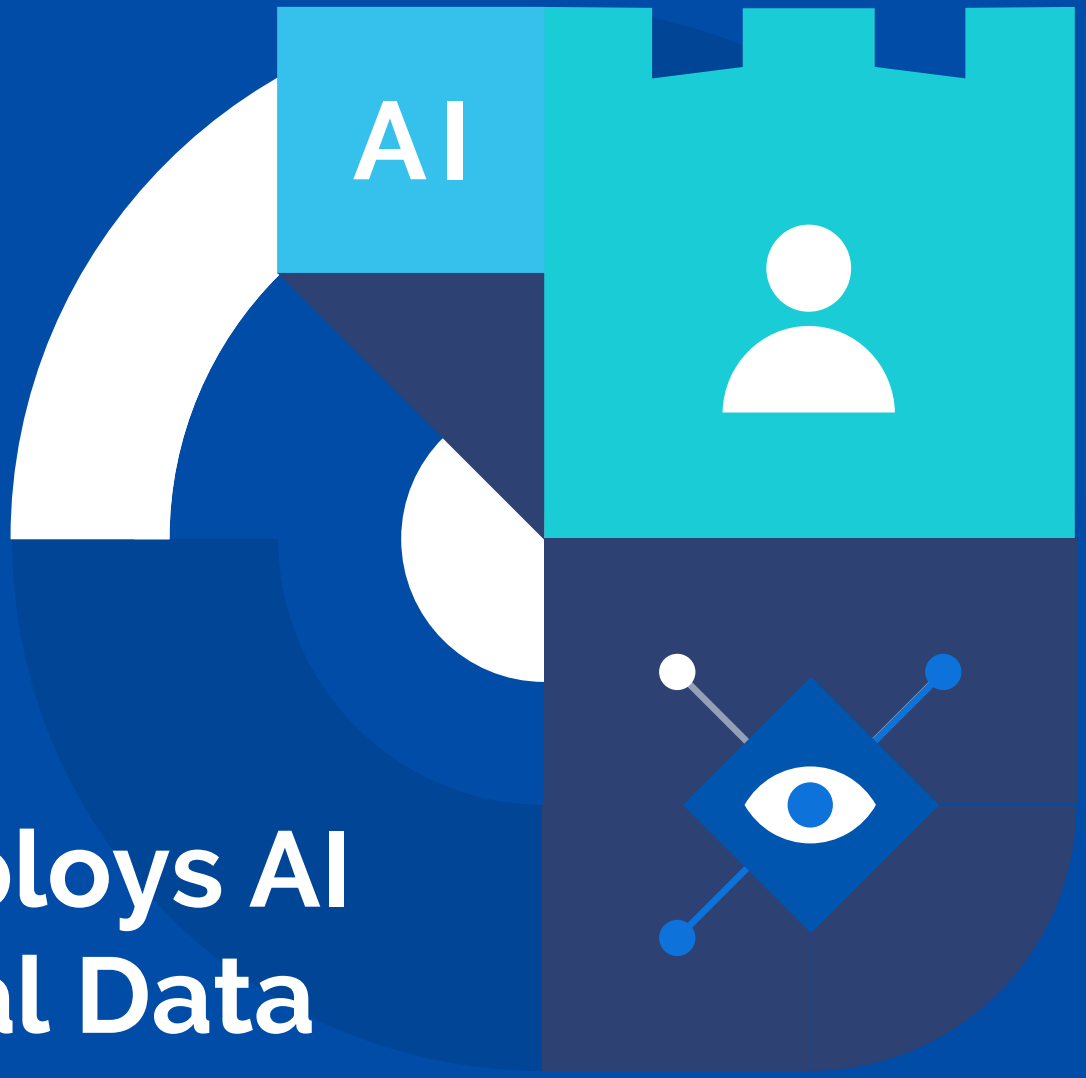
Dathena Privacy



AI Primer:

How Dathena Employs AI to Protect Personal Data

dathena

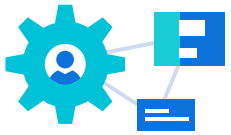
Dathena Privacy

Dathena leverages **Artificial Intelligence (AI)** technologies across its entire solution portfolio. In particular, Dathena employs proprietary AI technologies to address the critical problem of data privacy management and accelerate data privacy compliance.

Read this document to learn about the general and Dathena-proprietary AI technologies used in the development of Dathena Privacy.

Dathena Privacy

Dathena Privacy is a pioneering, AI-powered solution that is transforming how organizations protect personal data. Authorized individuals have a complete view of organizational data, its locations, and the access rights to every file and folder in the company's file system, allowing users to detect documents at risk based on the sensitive information they contain. With **Dathena Privacy** you can:



Link personal data to the subject of processing to automatically respond to compliance regulation requests

Inventory all personal information, whether structured, unstructured, on-premise, or in the cloud

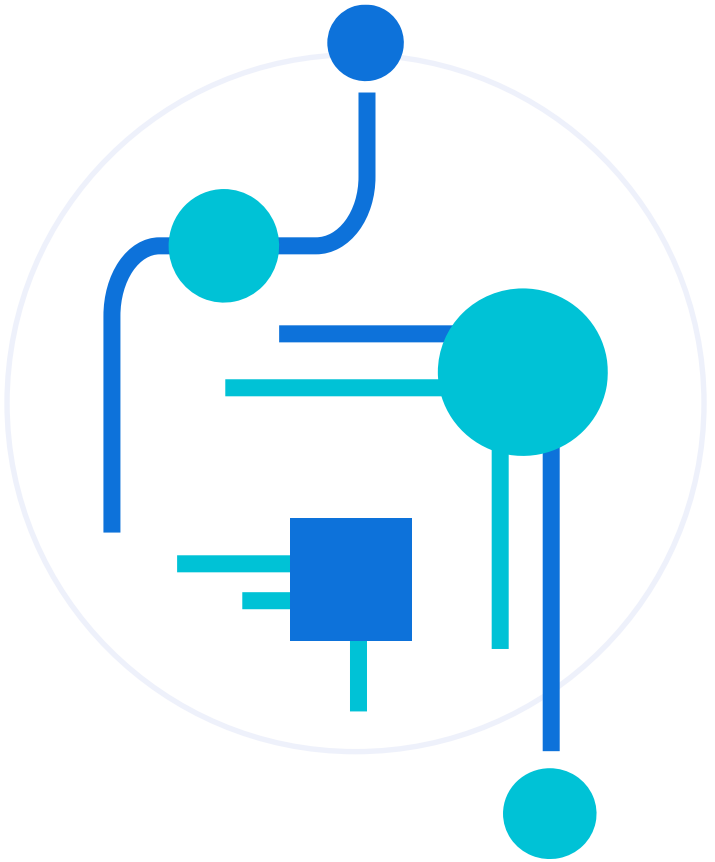
Define the purpose of personal data processing

Identify privacy risk to comply with General Data Protection Regulation (GDPR) or California Consumer Privacy Act (CCPA) and suggest remediation action

Provide a Record of Processing Activity (RoPA)

Get recommendations on how to optimize your storage system

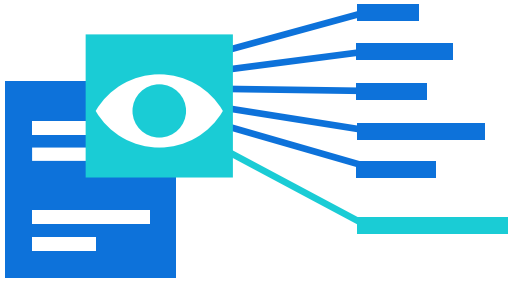
01 Data Engineering



- Automatically analyzes textual and numerical data from different types of documents and databases
- Automatically detects the language from the document content, folders, file names, and database tables
- Cleans, tokenizes, and lemmatizes the text [1].
- Analyzes document metadata (date of creation/modification, size, owner, etc.), clean, and match, etc.

1. Lemmatize: sort words by grouping inflected or variant forms of the same word.

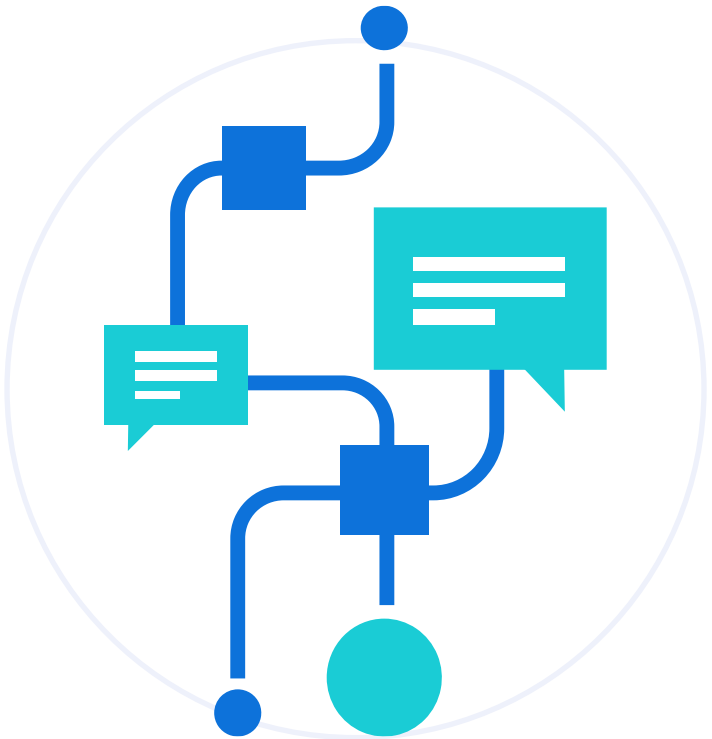
Selected AI Technologies Used in Dathena Privacy



Optical Character Recognition (OCR) converts a scanned image into normal raw text, which goes through data engineering steps to clean and convert it into "prepared data."

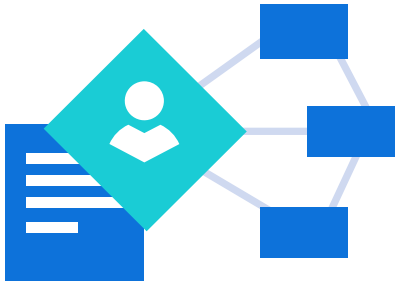
Feature Engineering extracts meaningful features from prepared contextual data and metadata and transforms them into a numerical vector used in downstream tasks. The system then deletes all source information and only keeps numerical representations to assure the highest level of security.

02 Natural Language Processing



Dathena uniquely leverages Natural Language Processing (NLP) methods to transform, aggregate, and generate textual data into meaningful information using language-specific techniques.

Selected AI Technologies Used in Dathena Privacy

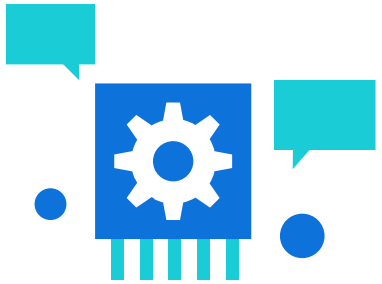


Named Entity Recognition (NER) automatically scans documents and databases and identifies and extracts personal data. The NER module incorporates Probabilistic Context-Based Modelling (PCBM) and Deep Learning (DL) state-of-the-art techniques to identify personal data that cannot be extracted using context alone, but by leveraging some specific linguistic features including metadata.

Dathena Proprietary

Probabilistic Context-Based Modeling uses context to determine if the data is personal data and what the probability is that the data is personal data. It also provides explainability to the extracted information.

Selected AI Technologies Used in Dathena Privacy



Language Modeling is a task aimed at training an unsupervised AI model to understand human language. Dathena uses specific language models to focus on specific words and their relationships in a mathematical way. When the text has been transformed into numerical values, any NLP task can be performed on the original text itself.

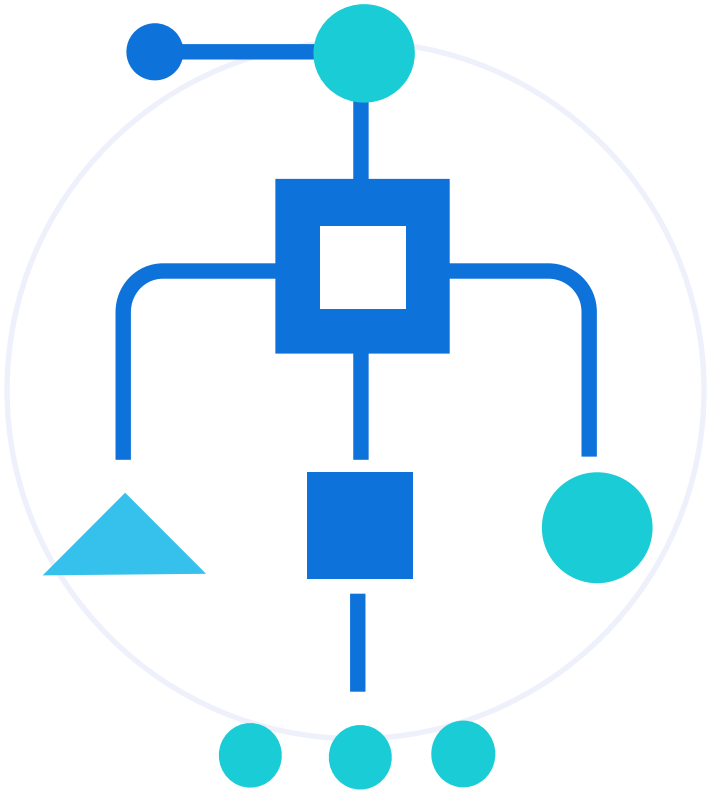
The Language Model is central to personal data extraction: it acts as the "brain" of the endeavor, which will comprehend the context of a sentence and identify the names of people, gender, religion, credit card numbers, and any other sensitive information.

Dathena Proprietary

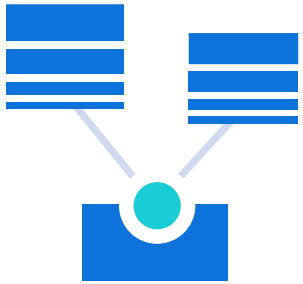
Text Summarization is used for longer text documents or groups of documents. It extracts the most important sentences and phrases from a set of documents and creates salient features so that unsupervised Machine Learning (ML) algorithms can understand the meaning of the text document topic. ML is discussed in more detail in the next section below.

03 Machine Learning

Unsupervised Machine Learning includes a family of techniques (including smart sampling, clustering, and autolabeling) developed by Dathena and used to label data with no human supervision. Unsupervised ML reduces the costs associated with manual labeling and provides flexibility with regards to languages, classification types, and data nature.

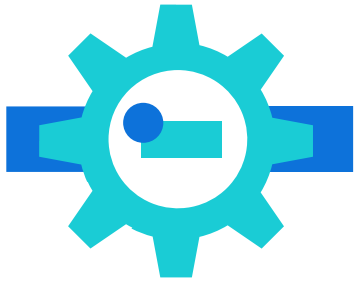


Selected AI Technologies Used in Dathena Privacy



Dathena Proprietary

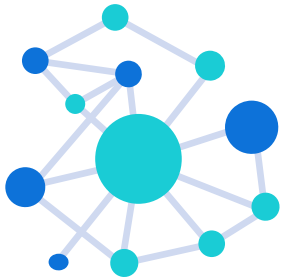
Smart sampling allows the model to process large amounts of data by choosing a smaller-sized sample of data that contains a balance of document types, topics, and data.



Dathena Proprietary

Autolabeling predicts the business category and level of confidentiality for document clusters in a completely unsupervised way.

Selected AI Technologies Used in Dathena Privacy

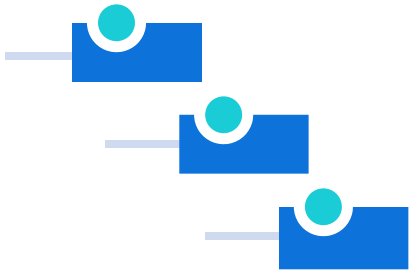


Dathena Proprietary

Entities Linking solves the problems associated with managing and responding to Subject Access Requests (SARs) by linking any personal data type to the subject of processing (person, company). The feature builds a knowledge graph comprised of the relationships between the subjects in the documents or databases (e.g., name) and other personal data (e.g., credit card number).

The knowledge graph and the predictive model that is built provide the customer with insights about their private data and help them comply with privacy regulations. Entities linking is used in both unsupervised and supervised ML models, depending on the data structure. It is a self-efficient technology, which does not require human labeling and is optimized over time with new data.

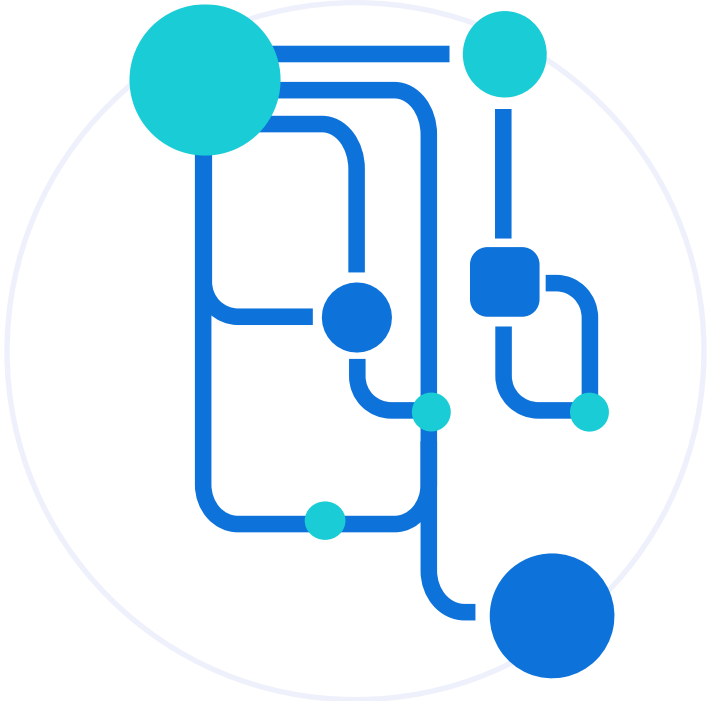
Selected AI Technologies Used in Dathena Privacy



Supervised Machine Learning is a class of algorithm that assigns labels to new elements not seen before and not yet analyzed. A supervised learning algorithm learns from labeled training data to help to predict outcomes for unforeseen data.

Transfer Learning is an approach where knowledge learned in source tasks (public knowledge bases) is transferred and used to improve the learning of a related business task. It is used to fine tune massive language models on domain-specific documents and extract special types of personal data (religion, nationality, passport), which are not learned on NER tasks.

04 Deep Learning



Unsupervised Deep Learning is used in feature engineering, clustering, and language modeling techniques to perform personal data extraction, predict the purpose of processing, and link data.

Language modeling is a key approach that is used in NLP and unstructured data processing because it learns the structure of natural language through hierarchical representation. The advantage of language models is that they can be retrained on client-specific data through the transfer learning approach.

Supervised Deep Learning algorithms use different artificial neural networks to protect personal data extraction on the entity level.

Dathena Key AI Characteristics

End AI Magic and Build Trust – Always know how results are generated, what data is used, and what calculations are performed.

Ensure Compliance with GDPR, CCPA, and other regulations – Dathena's solutions meet AI explainability regulatory requirements, which include GDPR's stipulation that consumers have a "right to explanation."

Meet the Requirements for Legally Defensible Evidence – Dathena's solutions help you comply with privacy regulations and deliver continuously improving outcomes that are explainable. Coupled with Dathena's transparent and ethical approach to product development and commitment to data quality and integrity, Dathena's solution offers legally defensible outcomes.

About Dathena

Leveraging the power of modern AI technologies, Dathena delivers breakthrough, petabyte-scale solutions with unprecedented accuracy, efficiency, and speed that build consumer trust in a digital world and ensure the "privacy and data security protection journey."

Dathena brings a new paradigm to data privacy and security. In a world of ever-growing information, regulation, and consumer privacy expectations, enterprises around the globe rely on Dathena to identify, classify and control sensitive data, reduce risks, and enhance their data protection framework.

Founded in 2016, Dathena continues to grow with offices in Singapore, Bangkok, Geneva, Lausanne, Paris, and New York City, Dathena employs more than 70 people, including the world's top data scientists and information risk experts.

Contact Us

www.dathena.io

hello@dathena.io

www.facebook.com/dathenascience

twitter.com/dathenascience

sg.linkedin.com/company/dathena-science