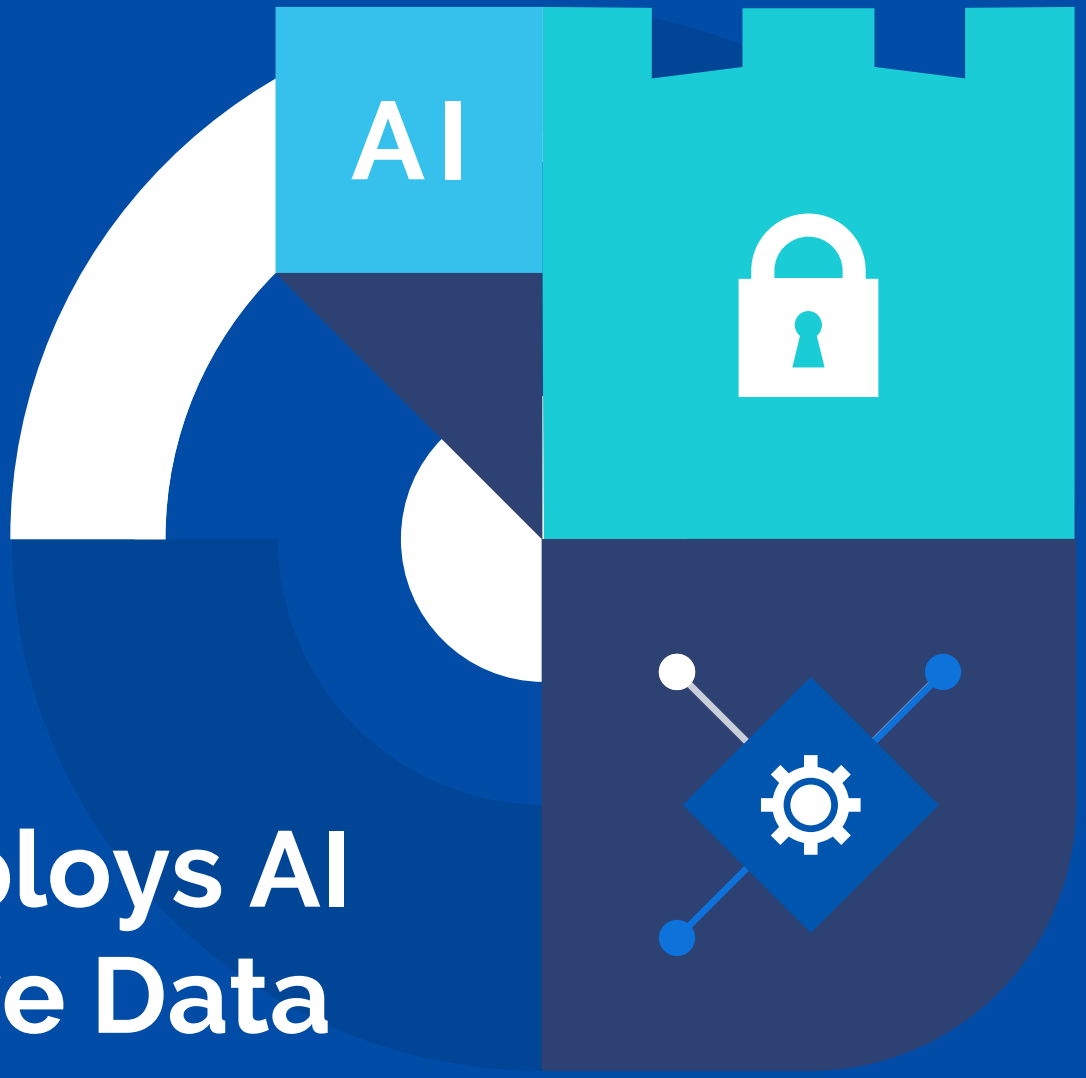
Dathena Security



AI Primer:

How Dathena Employs AI to Protect Sensitive Data

dathena

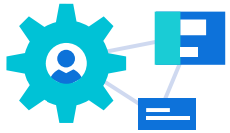
Dathena leverages **Artificial Intelligence (AI)** technologies across its entire solution portfolio. In particular, Dathena employs proprietary AI technologies to address the critical problem of data privacy management and accelerate data privacy compliance.

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

dathena

With **Dathena Security**, authorized individuals in your organization get a complete view of organizational data by confidentiality level, business category, and other dimensions, enabling them to detect, manage, and control unauthorized access and changes to sensitive data in order to:

Enhance your Access Control Framework



Improve the security of your Data Loss Prevention (DLP) and Information Rights Management (IRM) solutions

Ensure data availability on demand to authorized users and third parties

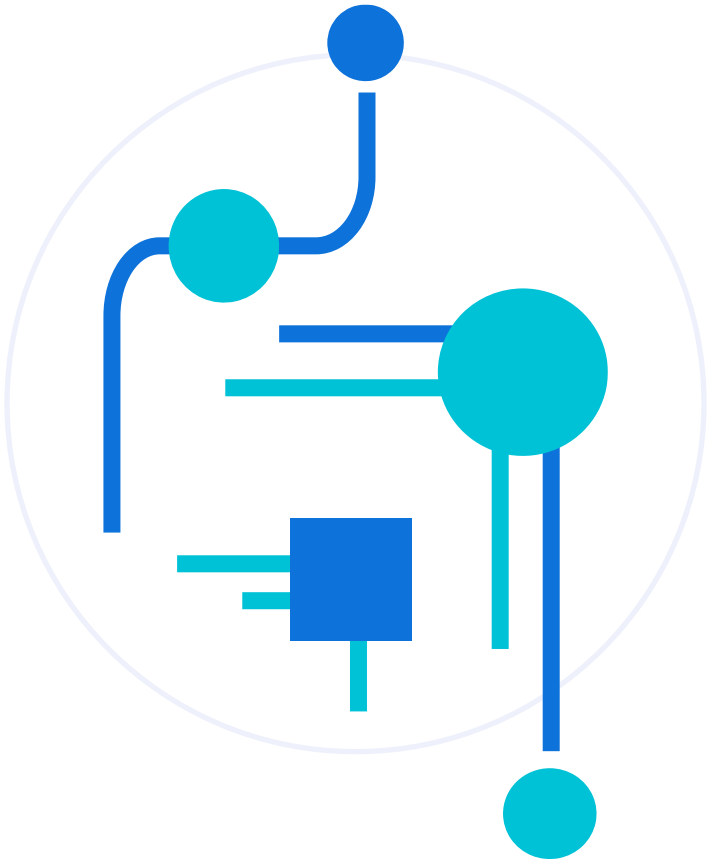
Reduce the risk of data breaches

Protect your brand

Facilitate safe cloud adoption

Selected AI Technologies Used in

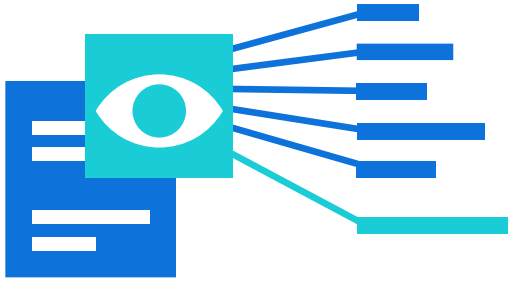
01 Data Engineering



- 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.

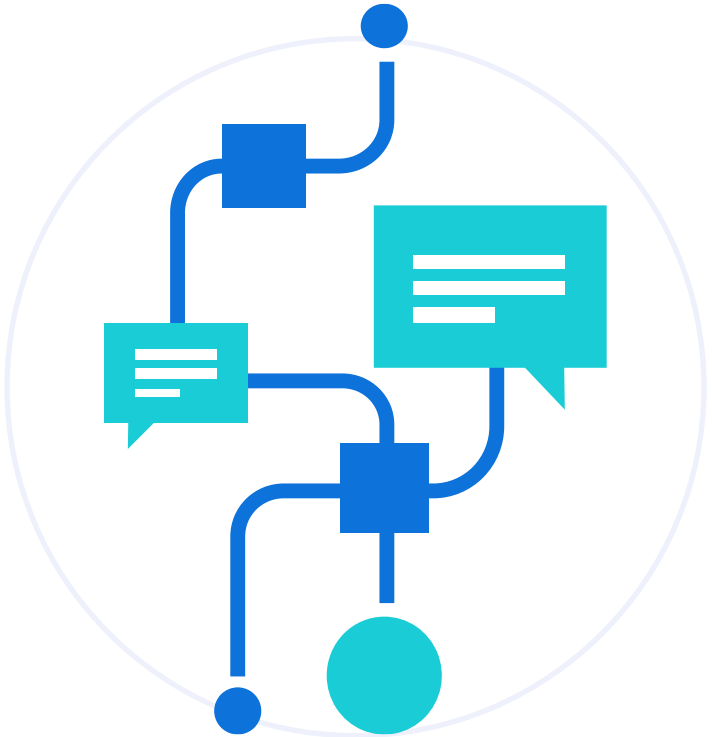
Selected AI Technologies Used in Dathena Privacy



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

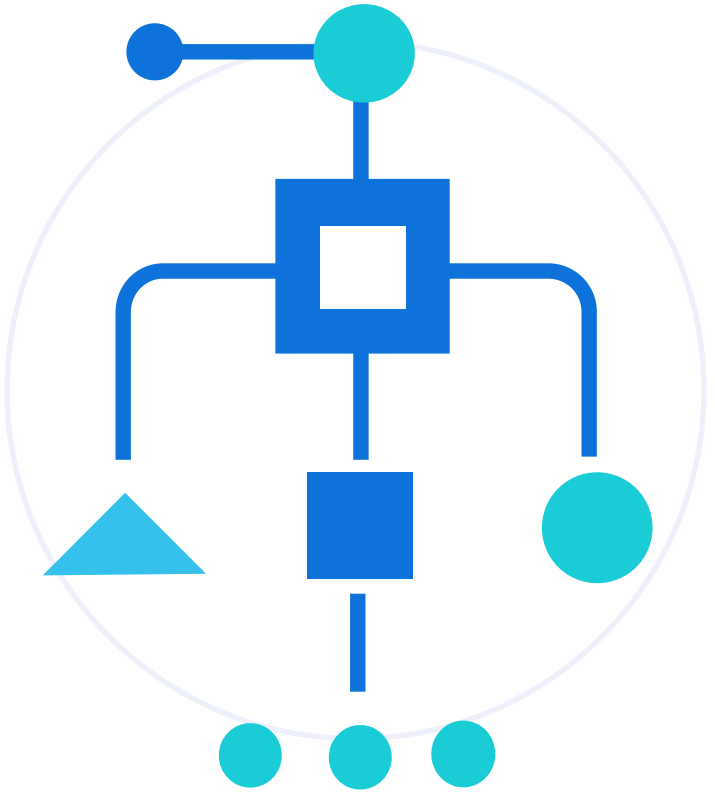
Dathena Proprietary

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.

Dathena Proprietary

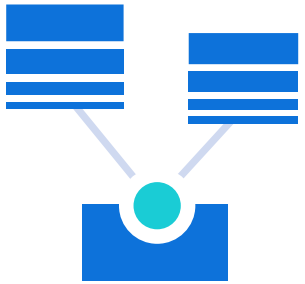
Text Summarization is used for longer text documents or groups of documents. It extracts the most important sentences and phrases from a set of documents and creates salient features so that unsupervised Machine Learning (ML) algorithms can understand the meaning of the text document topic. ML is discussed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

03 Machine Learning



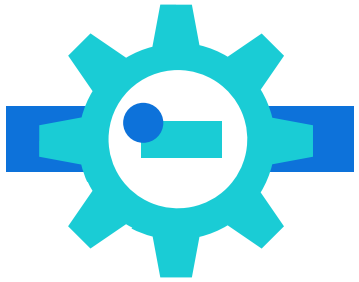
Unsupervised Machine Learning includes a family of techniques (including smart sampling, clustering, and autolabeling) developed by Dathena and used to label data with no human supervision. Unsupervised ML reduces the costs associated with manual labeling and provides flexibility with regards to languages,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

Selected AI Technologies Used in



Dathena 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

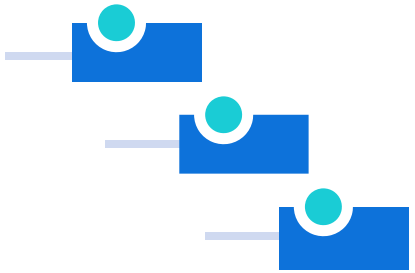
Dathena Proprietary

Data Loss Prevention (DLP) dictionaries generation technology is a unique approach that improves the protection and security of your data protection solutions/tools to mitigate data breaches, drastically reduce the number of false positives, and improve precision when detecting insensitive documents. Through automatic keywords and key-phrases extraction, which can be either positive or negative, Dathena Security generates the most effective combinations of keywords to match the documents related to any DLP Rule. Dictionaries include thresholds, keywords weights, and other parameters

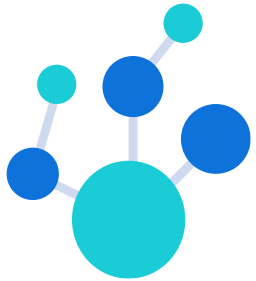
The technology is based on NLP and ML algorithms and can work without labeled datasets.[2]

2. PCBM and Dictionaries are optional and only required if generating dictionaries to enhance your risk management systems.

Selected AI Technologies Used in Dathena Security



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.

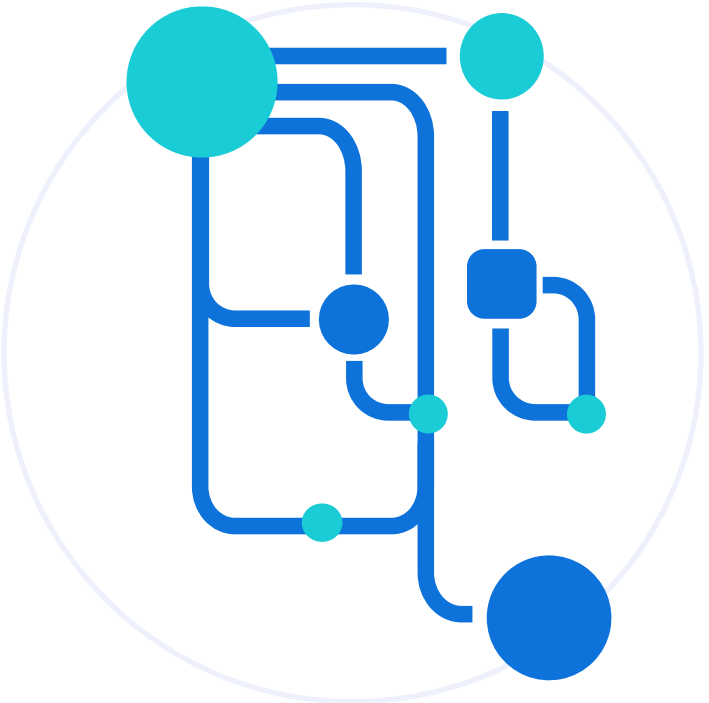


Active Learning is used in Dathena's classification review workflow, which sets up an environment so that the appropriate Key User is identified who can validate or challenge the model predictions by reviewing a sample of data output and correctly labeling the documents. Whenever new labels are identified, the system applies these labels and computes a confidence level for each document to decide whether the document needs to be reviewed by the Key User. This can be an iterative process performed until a high confidence level is achieved. Once a document is labeled, the model is retrained with the new sample.

Selected AI Technologies Used in

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, purpose of processing prediction, and data linking.

Supervised Deep Learning algorithms use different artificial neural networks to perform the classification tasks by different dimensionalities, including hierarchical business category prediction, confidentiality level, and others.

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.

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, Geneva, Paris, and New York City, Dathena employs the world's top data scientists and information risk experts.

Contact Us

www.dathena.io

hello@dathena.io

www.facebook.com/dathenascience

www.twitter.com/dathenascience

sg.linkedin.com/company/dathena-science