Legal Data Intelligence

Personal Data Identification and Anonymization

Updated: May 6, 2024

The processes used to distinguish and mask identifiable information about individuals in datasets to protect privacy, often done to comply with data protection regulations.

USE CASE IN PRACTICE

Cross-border PI Redaction

A pharmaceutical company utilized AI to develop a defensible strategy for identifying and redacting the personal information of European data subjects after their data was subpoenaed in the United States.

AI-driven Redaction of PI and PHI

A Legal Data Intelligence leader at a pharmaceutical company used AI to identify and redact personal information (PI) and protected health information (PHI) within documents related to clinical drug trials.

MODEL WORKFLOW

Initiate

Scope Project

Understand the relevant law and regulations of the applicable jurisdictions in regards to data privacy, protection, and security

State the purpose and lawful basis for processing personal or sensitive data

Annotate project-specific scope by articulating with specificity which data is necessary and proportionate to both analyze and retain post-project based on content attributes

Set risk tolerance to help determine the actions to effectively pseudonymize or anonymize the data set(s)

How Technology Can Assist

Analyzes legal filings against a structured reference data set of laws using text recognition / AI models

Automates purpose and lawful basis processes by using an exhaustive list of options for lawful basis to automatically populate data entered in a workflow

Automates risk tolerance processes by using an exhaustive list of risk levels and situational strategy decisions

Create Operational Guidelines and Rules

Outline how team members treat data containing personal or sensitive information and the chosen data anonymization action: data masking, pseudonymization, generalization, data swapping, data perturbation, synthetic data, redaction, or even withholding

Use to drive clear operational decisions once the data is identified, collected, and processed

How Technology Can Assist

Automates chosen action(s) based on operational guidelines and rules that are programmed into the software

Identify Data

Define custodians and data sources

How Technology Can Assist

Helps track and catalog available data sources

Future generative AI technologies may make it easier to identify relevant data using LLMs and prompts

Collect Data

Gather data from identified sources

How Technology Can Assist

Pulls data from sources into a platform for processing

Reduces collection of ROT data by excluding non-relevant documents using date filters, export rules to filter junk/system files, etc.

Process Data

Load and process collected data

How Technology Can Assist

Quickly manages large data volumes

Eliminates manual workflows and reduces human error

Reduces hosting of ROT data through de-duplication, de-NISTing, etc.

Investigate

Search

Run searches to find pertinent data

How Technology Can Assist

Locates data containing personal or sensitive information

Reduces review volume/ROT data

Evaluate Results

Review search results and cull data to maximize protection of personal or sensitive information

How Technology Can Assist

Provides opportunity for human input and analysis recording

Leverages human input and analysis to update any project and/or client-level active learning or automation

Analyze Data

Examine data to find personal information and sensitive information

How Technology Can Assist

Provides explanations and citations to help validate output

Apply Strategic Decision-Making

Apply the chosen action to the data set based on the decision points outlined in the Scope Project step and in the operational guidelines and rules

How Technology Can Assist

Visualizes and dashboards statistics regarding both analysis of and actions applied to the personal or sensitive data in the data set

Documents decision points for defensibility

Implement

Synthesize and Redact

Conduct defensible quality control of the anonymization effort through statistically significant sampling

How Technology Can Assist

Tracks document counts and tags files

Automates redactions to protect PII and adhere to privacy regulations

Helps with quality control by identifying errors or inconsistent coding, standardizing production rules, etc.

Protect Data

Ensure ongoing protection of data

How Technology Can Assist

Sets disclosure of protected data through secure password-protected transmission tools

Establishes audit and access logs using permission-oriented transfer technology