

CLIENT CHALLENGE

A large public research institution (the Client) was tasked with publicly archiving millions of court-produced documents relating to the manufacture and sale of opioids. Before uploading the documents to the archive, the documents needed to be scrubbed of any Personally Identifiable Information (PII) and Protected Health Information (PHI) of patients found within the document population.

The Client faced several challenges:

- Tight deadlines to make documents publicly available
- Cost constraints due to limited budget allocation for redaction, which all but eliminated traditional identification and redaction methods
- Identifying and redacting patient PII/PHI across millions of documents
- Distinguishing patient information from doctor, hospital, and pharmaceutical employee information relevant to the archive
- A high redaction accuracy rate, which was critical because documents were being made public
- A portion of the documents containing handwriting or were images with difficult or hard to discern text
- A variety of file formats which were provided from multiple parties and redactions needed to be applied to both natives and images, preserving the look of the court produced documents.

TACKLEAI RESPONSE:

After conventional methods proved too costly and too inaccurate, the Client collaborated with TackleAI to utilize its built-from-the-ground-up logical artificial intelligence coupled with machine learning principles to:

- Identify and build out the entities for everything and everyone in the data set to be able to redact information for patients only
- Identify all patient PHI/PII within the documents and automatically redact the information without a need for human validation/intervention
- Use proprietary computer vision models to read poor-quality images and flag handwriting

RESULTS ACHIEVED:

- Delivered all documents back to the Client according to document release schedule, significantly faster than any human review
- Achieved a 99.3% accuracy rate according to internal QC audit performed by the Client, outperforming the results of a human reviewed comparison data subset, which was 84%
- Lowered the cost of the project by 87%, compared to traditional eDiscovery methodology, leveraging identification analytics and managed review teams