



tackleai

The **logical** solution.

White Paper

Intelligent Dynamic Redaction and Automated Entity Discovery Solutions

The cost-effective and fastest way to discover and redact data on never-before-seen documents

Authored by:

Lizz Suarez, Chief Marketing Officer

Mark Dabler, AI Solutions Architect

November 2022

Contents

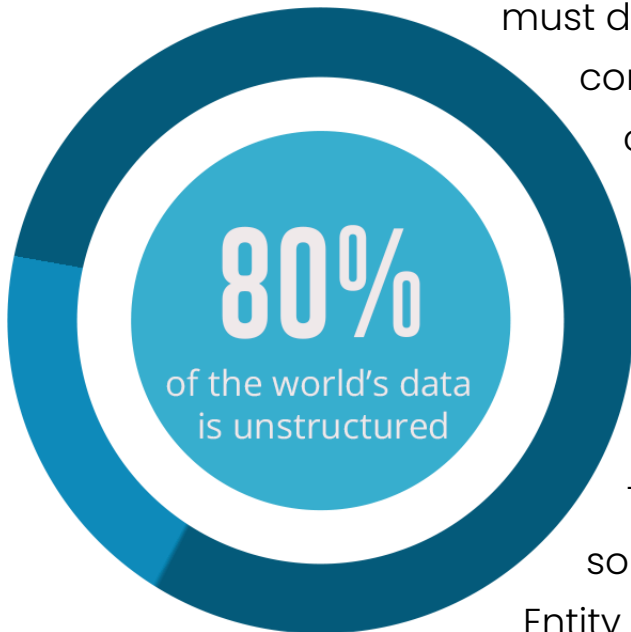
Introduction	3
Problem with current solutions	3
Overview of potential solutions	3
Poor Entity Discovery Solutions and Impact on Business Revenue and Workflows	3-4
Origin of the problem	4
Impact on business financials	4
Case Studies	5
TackleAI Design and Structure	5-6
Broad design of the TackleAI Entity Discovery and Redaction tool	6
Advantage of TackleAI's Entity Discovery design	6
Solution to Poor Entity Discovery and Redaction	6
General solutions to improve entity discovery and redaction	6
How TackleAI can solve Entity Discovery and Redaction issues	7
Conclusion	7
Poor data discovery solutions and impact on business revenue and workflows	7
How TackleAI improves data discovery workflows	7
Contact information	7

Introduction

Businesses and organizations rapidly employ new technologies and workflows to gain a competitive edge. More access to and the development of information technology has caused an influx of data that organizations

must discover, identify, redact, file, or extract to complete vital tasks. This is a costly and time-consuming task, currently accomplished by inaccurate humans or partially automated workflows. Potential solutions include various combinations of artificial intelligence systems that automate these workflows. This white paper describes the advantage TackleAI has over other Entity Discovery solutions, and why our product is the solution to Entity Discovery and redaction in a faster, more

accurate way, and at a fraction of the cost of a human employee.



Poor Entity Discovery Solutions and Impact on Business Revenue and Workflows

Unstructured data is when data is not organized in a pre-defined manner. Only 18% of organizations are able to take advantage of unstructured data.^[1] Entity Discovery is discovering data relationships between source tables, and creating an entity profile. Both Entity Discovery and redaction require accurate data discovery, and the current solutions are inefficient.

[1] "Survey | Deloitte". [2019 Survey](#).

Current Entity Discovery solutions are A.I.-assisted, requiring a human to train and confirm the models, before they can give you any meaningful results.

There are numerous ways unstructured data costs organizations money. Storage waste, data and security risks, fines for non-compliance, and staffing are just a few of the costs of unstructured data. Data discovery solutions rely on humans or inadequate automated workflows. Employing humans to identify and redact data costs more, and has limited hours to run the discovery and redaction process. Humans are also prone to errors, leading to more time and costs correcting any errors.



Automated workflows exist, but are insufficient in differing parts of the workflow, lacking an all-inclusive solution. Most existing solutions which claim to discover data are only successful on data which is pre-zoned, meaning the information must be in the same exact location on the document every time. For pre-zoned documents, the onboarding is more time-consuming,

as the automated model has to be trained on every different document before it can give any sort of accurate results. If a client changes their invoice in any way, for instance, the model must be retrained all over again. Natural Language Processing, or NLP, falls short with documents that do not have proper English grammar, such as forms or documents that contain shorthanded annotation. This often occurs in healthcare or legal sectors. Optical Character Recognition, or OCR, is another partial solution which is not reliable when used alone for Entity Discovery or redaction. OCR digitizes printed text into machine-encoded text. This can be used on both structured and unstructured documents, but the output is unreliable on its own. If the

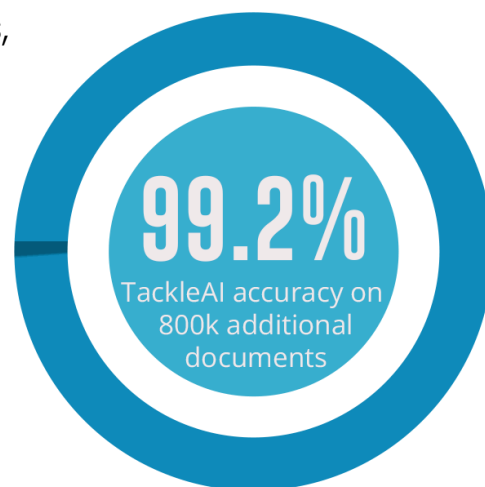
1	1
7	Z
B	8

OCR would likely fail to see the difference between these pairs of characters

background changes color, or the text is blurry or grainy, OCR is likely to have the wrong answer. It can also confuse characters. For instance, an OCR output may read a Z as a 7.

Case Studies

In a 100,000–document redaction project for a class–action lawsuit, TackleAI was compared to humans. The employees took five weeks to redact 100,000 pages with 84% accuracy. TackleAI redacted the 100,000 pages in 28 minutes with 95.7% accuracy. With an additional 1M pages, TackleAI was 97.4% accurate, and an additional 800,000 pages were redacted with 99.2% accuracy. TackleAI improves with each additional document because it is true A.I. that learns as it processes more documents.



In a project for a healthcare data aggregator, who processes hundreds of thousands of documents with personal health information that must be redacted. It took humans hours to redact 75,000 pages but only hours for TackleAI, which did it with 94% accuracy. For another 2M pages, TackleAI redacted the pages with 95% accuracy.

TackleAI Design and Structure

TackleAI takes a multi-faceted approach to Entity Discovery and redaction. All of the technology used is built from the ground-up, without the use of third-party applications or software. TackleAI employs proprietary Natural Language Processing, machine learning, models and code. While we use NLP and machine learning to the fullest extent, we also use logic, where other

systems do not use rules-based systems. This leads to faster analysis. Instead, TackleAI uses proprietary logic-based systems, written fully in-house. This gives us the advantage of being faster, flexible, and scalable, because we can make changes and adjustments to the network of logic, specific to



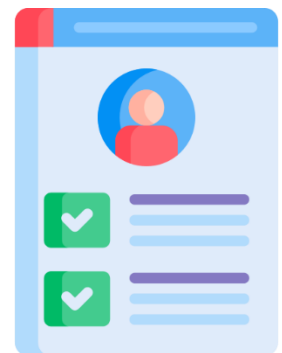
each client.

TackleAI's has a multi-layer approach that first starts with OCR. We have custom clean up routines that other OCR solutions do not have, which enables us to read dirtier documents than our competitors. Machine Learning and our algorithms are used to narrow down what the answer is by eliminating the incorrect ones.

While each of these techniques is not a sufficient solution on its own, combined with our logic engines, the aggregate approach is more accurate, faster, and less expensive than humans.

Solution to Poor Entity Discovery and Redaction

Entity discovery is the complexity of finding information but knowing it's tied to certain entities. These data points could be phone numbers, names, addresses, and more. One general solution to entity discovery is for a human to train the models, and match each data point to the correct entity. While this would be time-consuming, it could be helpful for an organization. Cloud API tools can get the client partially to their requested results, because it helps find named entities and phone numbers, but it doesn't do it with high accuracy.



Service companies requires humans to train the expert system. For example, the client may give a structured sentence to the model, and label all the entities and connections, so that over time, the entity

discovery will get better. Onboarding takes more time because the client must learn a new tool, and upload it to their platform.

With the right tools, like TackleAI, this process can be automated, with little to no onboarding. TackleAI can simply be given the data, and the output is given in whichever format the client requests. There is no human intervention as no one needs to train the A.I. to do its job.

Redaction is easier once the software has accurately identified every person, phone number, entity, and relationship of entities, and created profiles. TackleAI can then write a logic rule for redaction that looks at profile mapping. The proprietary algorithms of TackleAI uses rules that can make decisions using a custom framework, so we can tweak the parameters to make entity discovery and redaction work and perform better.

Conclusion

With most of the world's data unstructured, it is increasingly difficult to manage the amount that needs to be identified and redacted. Humans are prone to making errors in accuracy, cost more, and cannot run 24/7 like an automated workflow. Current automated Entity Discovery tools only get partial answers. TackleAI's product can run constantly, with high accuracy, with their multi-layered approach that leverages the best of each individual technique.

If you're interested in learning more about TackleAI's Entity Discovery and Redaction, please reach out to us at info@tackleai.com or visit us at www.tackleai.com.