Extracting Facts from Hardcopy using AI

Forms, filings, hardcopy, transcripts. There are many examples of useful documents that are archived in various government agency systems that are typically not in usable condition – that is, usable for the automation of business intelligence solutions. Material that has been scanned from hardcopy or filed as structured forms (e.g., PDF) will often contain information useful to one mission area or another, but only when such documents are fully digitized, complete with metadata, and exploited by the right analysts and experts.

AgileDD.ai is a company that has a product to assist in unlocking the raw data, facts, entities, etc. from these sorts of challenging document types. In May of 2020, through a collider event during the US Air Force Lab with MassChallenge program, MITRE heard about a wide range of innovations through 3-minute elevator pitches as a means of speed networking with startups. Bridging Innovation mentor Eric Renda introduced Amit Juneja, Principal Data Scientist, and Grace Yi, Customer Success Lead, of AgileDD.ai to Marc Ubaldino, a leader in MITRE's Analytic Technologies and Solutions (ATS) Lab. Marc is constantly seeking novel AI-oriented solutions to fill gaps in the lab's collection of analytic capabilities, specifically in intelligent information extraction. AgileDD.ai is exactly what he is looking for.

AgileDD.ai offers a trainable extraction product that operates more like how the human eye picks out contextually relevant information, for example, as one might scan and identify key information off a page in a magazine. The layout, font changes, the use of charts and tables, etc. are all clues for us to pick out interesting snippets and facts. Automating this process for business intelligence pipelines is far more difficult, as most of today's commercial and open-source analytics operate on the text alone, ignoring all this critical context.

AgileDD.ai is relatively unique with their patented approach for modeling the extraction process in the visual plane of a document, rather than treating a document as a sequence of words reading from top to bottom. Furthermore, any acceptable solution also has to be attuned to the semantics of the particular documents' business domain, jargon, nuances, terminology and the user's objective. This venture is on the right track.

AgileDD has chosen to participate with Bridging Innovation to engage with more US governmental opportunities as they transition from a focus on energy industry customers. Our use of AgileDD.ai solutions in ATS Lab will help us internally explore state-of-the-art approaches in information extraction, which has the potential to benefit any of MITRE's sponsors or research.
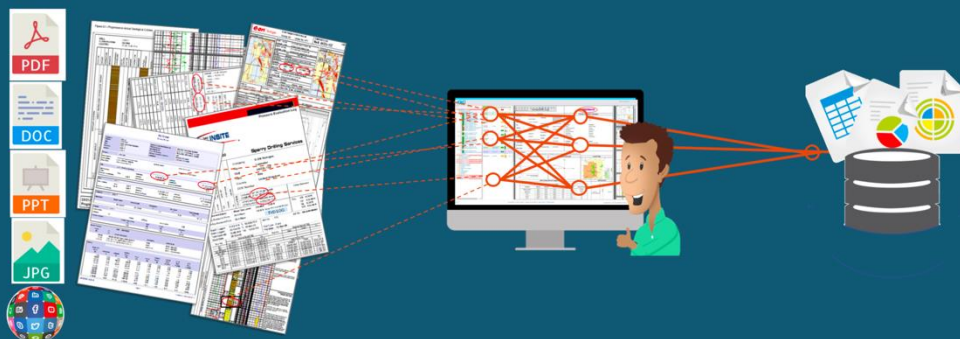
CAPTION: AgileDD.ai products enable an analyst to model discrete information seen in their archival documents, so they can automate detecting, extracting and organizing it.

SOURCE: https://www.agiledd.ai