Inspire
Innovations

Intelligent Data Extraction at Inspire:
Why it's time to get excited about AI

Automatically identifying limitless variations of content, locating and extracting business critical data, and understanding intent from intake content may not be the sexiest application of artificial intelligence, but it has massive potential to reduce errors and costs, and dramatically improve the speed and quality of data.

Processing intake content is a fundamental and critical component of insurance operations. But it's a slog. Customers, partners, brokers, and agents will each have their own quirks and nomenclature within the content they send and it's very difficult to enforce data standards on all these external parties. Every day, at every level of management and operations, there is a need to review, understand intent and need , and extract critical information for all sorts of intake content; Like new business submissions, claims submissions, and policy servicing requests.

The good news? AI offers ways to perform these complex tasks far more efficiently. These solutions are seamless and scalable, simple to operate, and easy to manage. Recent research by PwC on automated natural language processing found that even the most rudimentary AI-based extraction techniques can save businesses 40–50% of the hours typically spent on such processes.

50% fewer hours are needed to process routine intake content when even the most basic AI-based extraction techniques are implemented.

We all know about the paradigm-changing use of AI for Amazon personalized shopping recommendations, Tesla autonomous vehicles, Pfizer drug screening and development, and the continuous optimization of delivery routes for UPS. These efforts are the value creation engines of countless large, successful companies. What we're talking about here is a decidedly less sexy and, at face value, more pedestrian use of AI—it's aimed at reducing costs and optimizing operations rather than transforming or creating industries. But this boring AI is actually quite exciting, because it confronts issues that all insurers wrestle with, and because the gains in productivity (and hence margins) are very real.

Yet despite its huge potential, a recent Gartner AI Predictions survey found that only 32% of executives have prioritized using AI and machine learning for data extraction, significantly less than for other uses, such as automated underwriting or claims adjudication. Some leaders are likely overwhelmed by the time and resources required to develop, scale, and integrate these advanced technologies. Some will be hesitant to trust AI or will feel skeptical about its utility. Others may simply be overlooking the value of automated information extraction because it is a back-office function.

Regardless of the reason, they are missing an opportunity to streamline processes and improve their return on investment.

HOW DOES AI-BASED DATA EXTRACTION WORK

Pre-Processing
Binarization • Noise Reduction •
Deskewing • and more

Intelligent Document Classification
OCR • NLP • Supervised & Unsupervised
Learning • AWS Rekognition

Data Extraction
NLP • ML • Deep Learning
• AWS SageMaker & Comprehend

**Insurance Industry
Specific Validation**
Fuzzy Logic • RegEx • Rules •
Compliance • Scripts

Enhanced Validation
RPA • API Integration
Gateway • AWS Kendra

Human-in-the-Loop Validation
Low Confidence Review • Exception
Handling • AI Feedback loop

Unstructured Content:
Your core impediment to efficiency

Any insurer that doesn't receive new business submissions or claim first notices of loss via a completely controlled channel (i.e. member or agent wizard driven portals) spends an enormous number of hours every year parsing this content for actionable information. For a mid-tier insurer that processes 250,000 of these such transactions a year at 5-minutes per transaction would take ~20k hours to complete the task; at USD $50 per hour, **that's $1M.**

The average efficiency gains experienced by basic AI-driven data extraction is 50%

Efficiency gains with intelligent data extraction

Now, what if the same insurer could deploy an AI-driven data extraction solution? Whereby a cloud service could be used to pass these submissions and FNOLs, and receive back structured, cleansed, and validated data for immediate business processing. Even though this content can be unique to each insurer, AI techniques can identify the critical data insurers need, such as applicant data, employment data, product data, loss details, member data, claim data, and so on. AI can also classify document types and ensure prerequisite documents are received for each type of transaction and automatically reach out to the submitter to supply the missing documentation. By implementing this AI solution and assuming the 50% estimate above, the example midsized insurer would save a minimum 10k hours of work reviewing, triaging, and performing data-entry on those 250,000 such transactions: or roughly $500k a year.

The imperative of automating data extraction from content is certainly not new, but its only grown more urgent over the last decade. Data growth is mind-bendingly massive.

Global data growth is mind-bendingly massive. IDC predicts a compound annual growth rate for global data of 23%, reaching 181 zettabytes in 2025, an increase of nearly 3x from 2020. (To help visualize that kind of data volume, 1 zettabyte is equivalent to 250 billion DVDs of data!) And the workplace will experience a similar explosion of data — so this efficiency impediment is not going anywhere and will only get larger.
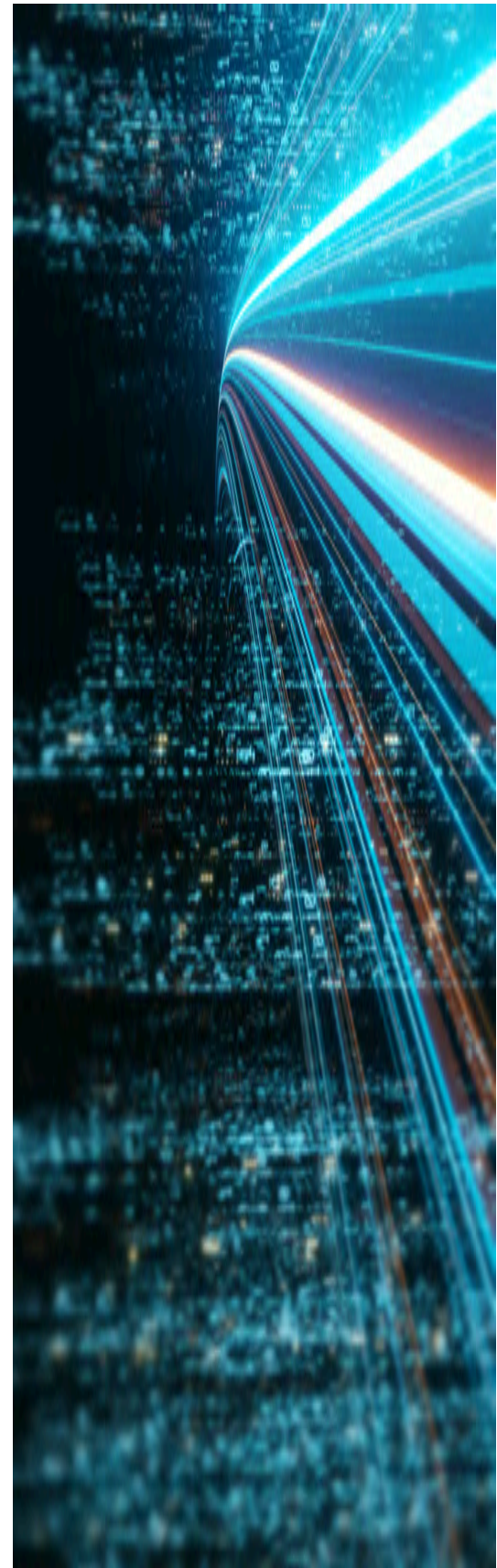
Inspire Innovations

datavelocity.app

InspireInnovations.com

Tools of the trade:
3 core technologies

AI-driven data extraction can tackle many of the inefficiencies and problems endemic to the above scenarios. However, unlike robotics used in manufacturing that do spot welding or spray painting, AI-enabled data extraction is not a repetitive, routinized activity. It requires a slew of complex data science techniques involving multiple dynamic components that must adapt to ever-changing conditions. Integrating cutting-edge technologies such as optical character recognition (OCR), supervised machine learning, and automated analytics that incorporate natural language processing into a seamless process will require time and technical expertise.

01.

Optical Character Recognition (OCR)

Consider OCR, which is the ability to read printed characters on a page—even handwritten characters—regardless of font, size, orientation, and brightness. At present, we encounter this technology frequently with the automated deposit of checks using our phone, when the OCR reads not only the routing number and account number but also the check amount and date. OCR is an older technology but is still essential as the first step in the process that gathers the relevant data from the documents in question.

02.

Machine Learning (ML)

For many uses, turning that data into action requires sophisticated machine learning algorithms that can recognize and classify patterns. Machine learning algorithms can be calibrated on existing data to tune their parameters, and then unleashed onto novel data. They can be calibrated to recognize patterns that are sophisticated yet subtle indicators of claims fraud, such patterns atypical to a type of claim. They can also unearth similar meanings in different policy wordings, for example, in exclusion and limitation clauses.

Additionally, machine learning algorithms can tackle a data set and categorize a set of entities into different groups. Automated customer segmentation is one well-known example of this, but categorizing applications, deeds, statements, loss runs, reports, or policy terms is also possible and can save enormous amounts of time that would otherwise be spent reading these documents.

03.

Natural Language Processing (NLP)

Advances in natural language processing in the last few years have been impressive. Although it is not necessary to use the most advanced algorithms, such as the natural language generation application GPT-4, AI-enabled information extraction can nevertheless take advantage of some of these advances by identifying the true "meaning" of a document, through identification of contextual words, parts of speech, and so on. The AI itself does not understand what it is saying (although it might appear that way), but algorithms are able to generate summaries of documents; identify topics; judge the sentiment (positive or negative) of prose; identify key entities and critical data within documents; and identify clusters of documents requiring similar actions.

The human side of it:
How to successfully implement

AI tools tend to be highly accurate, but when they do make errors, they can be nonsensical and downright bizarre. Maintaining human oversight during the implementation of these technologies is crucial to ensuring quality, both for model training and for the final correction of the output in downstream processes. Successful implementation thus requires more than procuring the tools. Companies will also need to take the following actions:

A.

Create a new core platform within your IT ecosystem that combines data management, automation tools, and AI applications, but also keeps people in the loop. This platform could be a central enterprise-level portal, wherein data could be stored and exchanged, applications uploaded and downloaded, and collaboration and joint development encouraged through a communication interface.

Inspire
Innovations

This platform should be accessible to everyone in the organization and should be receptive to employee-led innovations and applications as well as those from professional developers. Of course, such democratization of these powerful technologies ought to proceed responsibly; leaders must stay vigilant about the potential risks and cognizant of the need for proper training and corporate governance.

B.

Develop an enterprise-wide training program focused on digital and analytic understanding and awareness. Everyone will need to be upskilled, from the CEO to the newest entry-level hire, across all functions. Companies should consider training many of these employees not only in the use of these time-saving information extraction tools, but also in the fundamentals of the AI technologies behind them. With a better understanding of the capabilities, risks, limitations, and assumptions of the AI, employees will better understand how to use the tools responsibly and effectively. Every organization should ensure that its employees are conversant with current technologies, and this transformation will take hold only if the entire workforce is brought along.

C.

Pay special attention to the impact on middle managers for whom a substantial portion of daily tasks will essentially be eliminated. That is a reality of automation—it creates efficiencies by taking over some tasks that are currently done by humans. The important message to communicate to managers is that, in so doing, AI will free them to focus on harder-to-solve problems, and to work on issues that demand human judgment or creativity—to do more managing and fewer mind-numbing repetitive tasks.

D.

Develop an enterprise-wide training program by designating top-down champions who consistently and frequently communicate the benefits of AI implementation. The message that using these tools is on-strategy, is viewed favorably, and is good not only for the organization's customers but also for the organization's health and growth will accelerate adoption and make the technical and cultural changes stick. As an application of AI, information extraction may appear mundane, but a closer look reveals that the opposite is true. With automated or augmented solutions, businesses have the potential to energize processes that have traditionally been time-consuming and error-prone, identify opportunities to add speed and efficiency, and unlock new insights that contribute to long-term growth. Boring has never seemed so exciting.

IPAS.app

InspireInnovations.com

Inspire
Innovations