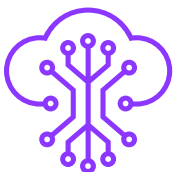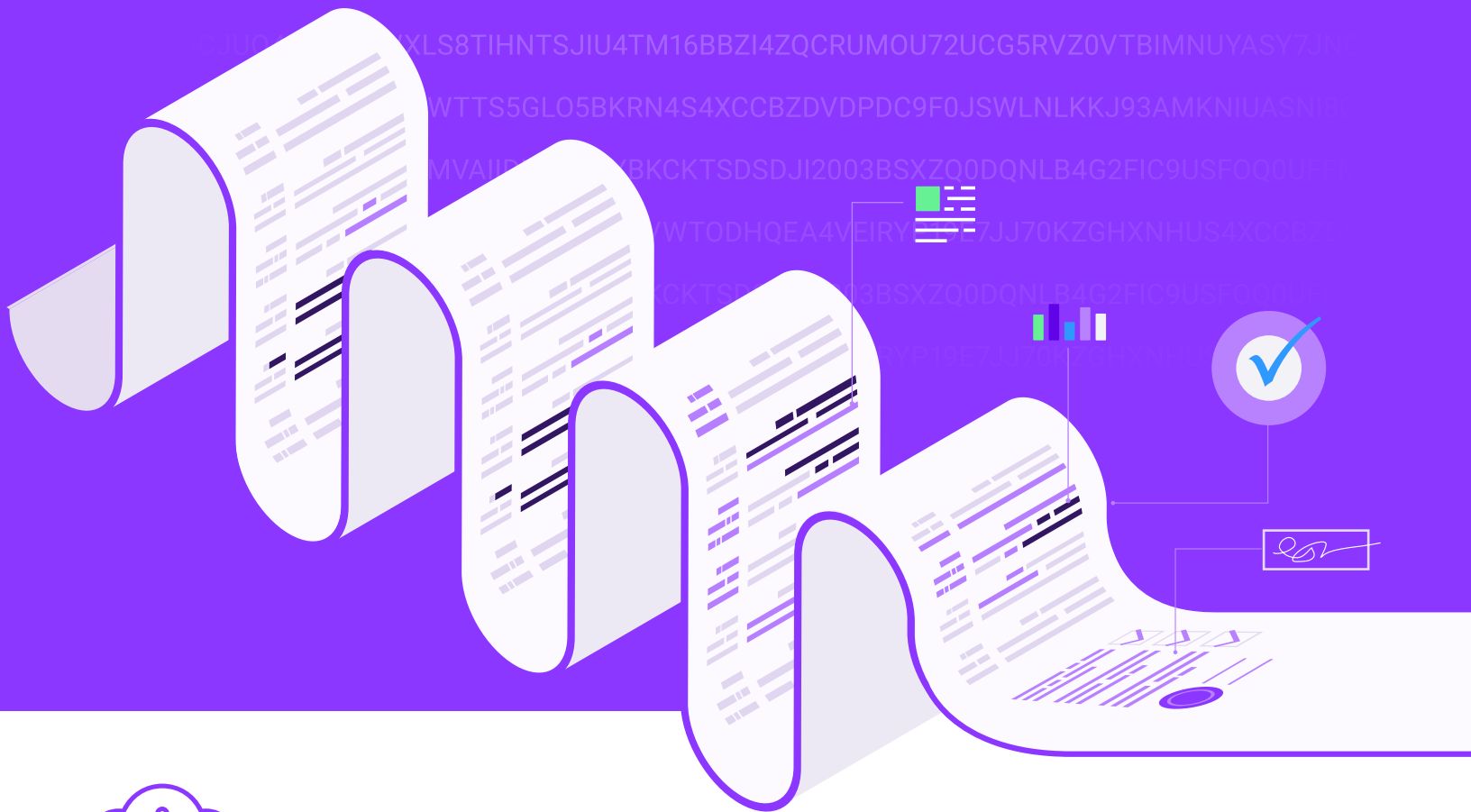# Solving the Insurance Industry's Unstructured Data Challenge with AI

TURNING NATURAL LANGUAGE INFORMATION INTO ACTIONABLE DECISION-DATA

Roots Automation

rootsautomation.com

Introduction

Insurance has always been a data-driven industry, relying heavily on data for risk assessment/pricing and claims management. By performing these essential functions, insurance delivers security and stability to customers.

Insurers succeed by providing outstanding protection and service to policyholders—both made possible by underwriters and claims specialists who have the capacity to prioritize productive engagements that help customers "win" with better insurance products and exceptional customer experiences.

Some of the data used in insurance is structured and organized into standardized formats, e.g., ACORD forms.

However, over 80% of data leveraged across the industry is unorganized and highly variable, locked within reams of paper, emails, faxes and snail mail. This information is categorized as unstructured data.

Unstructured data permeates communications channels across critical decision-making documents, including loss run reports, statements of value (SOV), questionnaires, underwriting supplements, line slips, policy documents, legal demand letters and medical records.

Many insurers have tried using technology to transform unstructured, natural language information into structured data. However, these solutions have proven inadequate for meeting the needs of insurance-specific document processing tasks. As a result, underwriting and claims teams resort to human-resource-intensive manual processing.

Processing unstructured data consumes up to 40% of the average working day.

▸ **On the underwriting side,** receiving, accurately pricing and responding to a submission faster than the competition improves the customer experience, which can make the difference between dramatic growth and a shrinking top line.

▸ **On the claims side,** incorrectly entering a policyholder's name will incur rework, while incorrectly identifying a demand in a multiple-page demand package can result in litigation. Such outcomes result in slower claims payments to policyholders and higher loss ratios. Additionally, reliance on human power and legacy technology further inhibits insurers' ability to compete by diminishing employee job satisfaction.

Advancements in AI technology – such as Natural Language Processing (NLP) and Generative AI – offer solutions for solving the unstructured data challenge. However, they also present new risks and challenges related to data, security, ethics and regulatory compliance.

This whitepaper presents an overview of the challenges and options available to insurance leaders seeking to **transform unstructured natural-language information into structured, machine-readable "decision data."**

> **"** *"Using AI to extract insights from unstructured data sources, insurers can see a 10 to 15 percent improvement in their loss ratio compared with companies that don't use these technologies."*
>
> *MCKINSEY & COMPANY, THE DATA-DRIVEN INSURER*

Part I.
The Challenge:
Unrealized value locked inside unstructured insurance data

Revenue leakage due to inefficient underwriting and claims processes accounts for $0.14 lost from every dollar.

One of the main reasons for this lost value is that insurers struggle to efficiently transform unstructured data into structured data for informing underwriting and claims decisions.

Today's most prevalent solution for unstructured data in insurance is using human power to perform natural language processing tasks. However, people are no match for the ever-increasing volume of documents pouring into underwriting and claims departments. Consequently, much of the information inside unstructured documents is more likely to be incompletely analyzed, misread or unused.
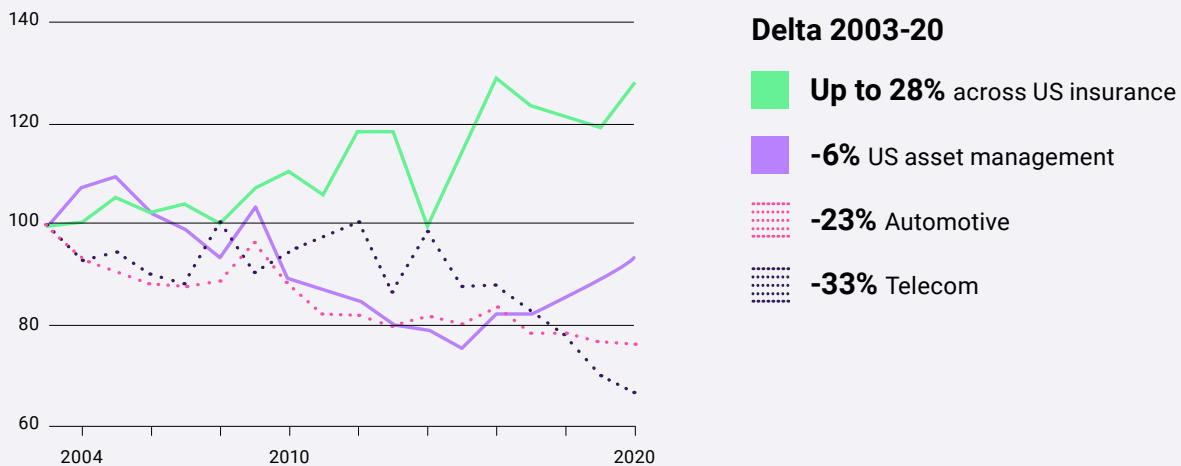
The cost of underinvestment in transformative technologies

The insurance industry's digital transformation efforts to improve cost efficiencies have lagged relative to other sectors. (Figure 1)

> The total quantity of data globally in businesses has grown from one zettabyte (a zettabyte = 1021 bytes) in 2012 to a projected 175 zettabytes by 2025. **Eighty percent of that data is unstructured, and 90% is never analyzed.**
> Source: Forbes, "The Unseen Data Conundrum"

Compared with other industries, US insurers have not structurally addressed costs.
Cost efficiency evolution[1] by industry,[2] % total SG&A expenses and revenues, (index 100= 2003)

Delta 2003-20

- **Up to 28%** across US insurance
- **-6%** US asset management
- **-23%** Automotive
- **-33%** Telecom

[1]Indexed; expressed as SG&A expenses as % of revenue.
[2]Based on large global players for which continuous reporting is available: 28 players in insurance (10 composite focus, 9 life focus, and 9 P&C focus), 10 telecom players (including AT&T, China Telecom, and Vodafone), 8 road and rail players (including DSV, Japan Railway companies, and Russian Railways), 10 automotive players (including Ford, Toyota, and Volkswagen), and 10 airlines (including American Airlines, Air France-KLM, and Emirates).
Figure 1 (source: S&P Capital IQ; McKinsey analysis)

At the same time, very "human" natural language task-oriented functions that would benefit from modern technology in areas like underwriting and claims remain underadopted and underinvested relative to distribution and operations. (Figure 2)

Investing in insurtech across the insurance industry value chain
More than a third of insurtech investments has been in distribution technology.
The other main areas of digital investment have been in pricing and underwriting processes.
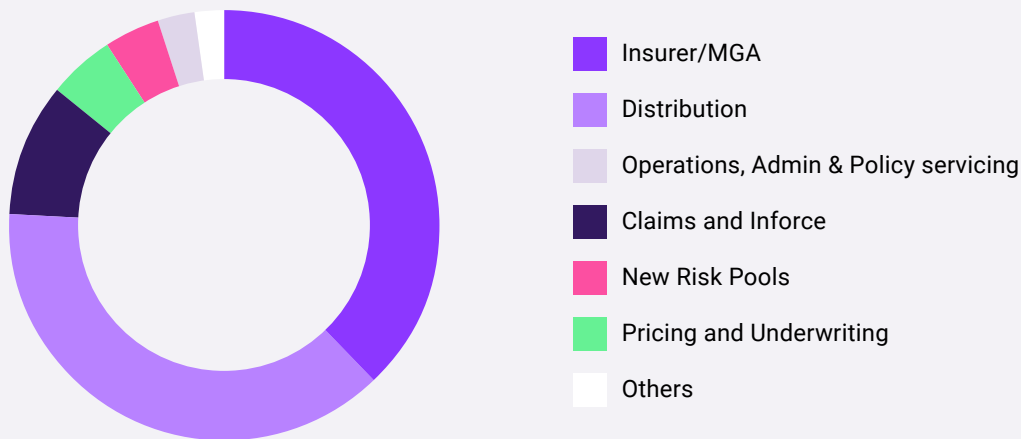
Legend:
- Insurer/MGA
- Distribution
- Operations, Admin & Policy servicing
- Claims and Inforce
- New Risk Pools
- Pricing and Underwriting
- Others

Figure 2 (Source: Swiss Re Institute)
Source: Swiss Re Institute analysis of 460 insurtech start-up involvement in the process steps of the insurance value chain.

Reliance on manual effort to read and process this data creates inefficiencies with downstream consequences.

▌ **In underwriting, high document variability complicates data extraction and analysis, causing:**
- ▸ **Inaccurate risk evaluation** and delayed quote turnaround times.
- ▸ **Reduced ability** to offer competitive pricing to win business due to critical details "missed" in unstructured data.
- ▸ **Intensifying the insurance "talent crisis."** Over the next 15 years, 50% of the insurance workforce will retire, according to a US Chamber of Commerce study.

▌ **In claims, manual work of unstructured claims documents creates:**
- ▸ **Poor customer experiences** due to slow claims payments.
- ▸ **Claims leakage,** a component of revenue leakage, which has cost insurers $88 billion.
- ▸ **Reduced capacity** to contend with spikes in claims submissions driven by intensifying natural catastrophe risks.

Insurance's dependency on human experts to perform non-core and administrative activities (e.g., manually entering and reworking information from insurance documents) will contribute to an estimated $85-$160 billion in value lost to inefficiency by 2027, per an Accenture study.

The opportunity exists for Insurance companies to prevent these losses by maximizing efficiency across underwriting and claims operations.

Part II.
The opportunity:
Unlocking value from unstructured data with AI

An IDC survey of mid-market and enterprise businesses found that 40% of companies use "mostly manual" document processing methods, i.e., people reviewing and extracting information of "potential value."

For more than a decade, insurers have attempted to solve the unstructured data problem using legacy automation technology—Robotic Process Automation (RPA), Intelligent Document Processing (IDP) and Intelligent Process Automation (IPA)—and most have had poor results because these tools heavily rely on structured data to operate.

Companies implementing RPA, for example, have experienced significant downtime due to broken "bots" and poor ROI. Consequently, 96% of RPA customers never get to value (Deloitte).

Unlike legacy automation technology, AI uses goal-oriented reasoning – not coded instructions – to complete tasks such as classifying a document or identifying patterns in data.

> "*Generative AI could add the equivalent of $2.6 trillion to $4.4 trillion annually [to the global economy].*"
>
> *MCKINSEY & CO*

### Generative AI vs Large Language Models

Generative AI:
Models that interpret patterns within their input training data and generate new data with similar characteristics.

Large Language Models:
A type of Generative AI model that can recognize, translate, summarize and generate written language and text-based data.

This capability makes Generative AI well-suited for turning unstructured data into structured data. They are ideal for interacting with human users to automate insurance processes/workflows and reduce errors.

AI models set parameters around goals, which makes Generative AI using fine-tuned Large Language Models a winning combination for dealing with unstructured information.
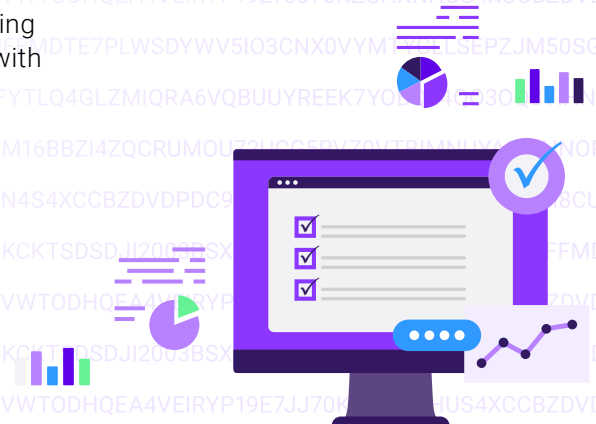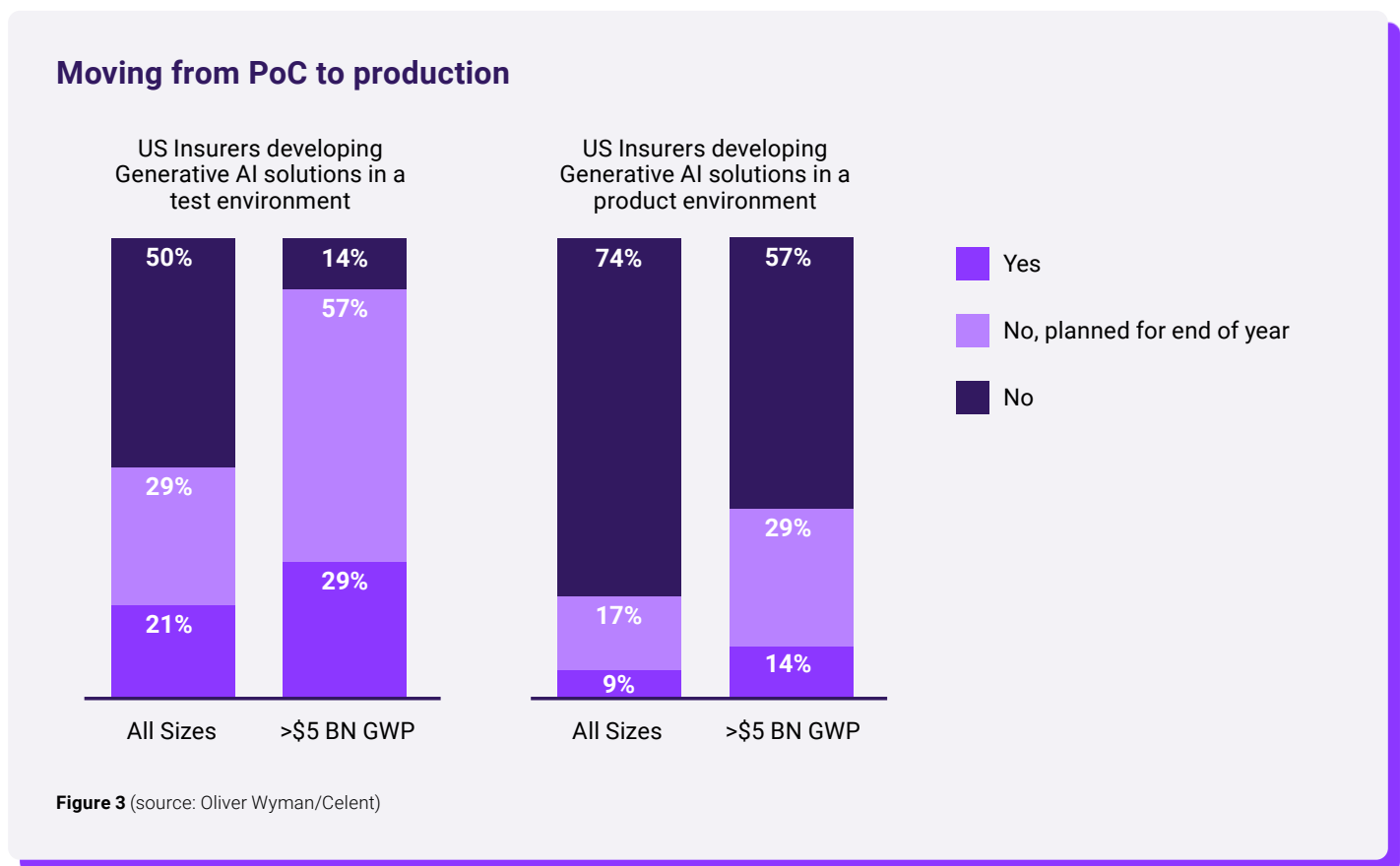
"Can't insurers just build their own Generative AI?"

Public horizontal Generative AI solutions (i.e., ones offered by Google, OpenAI, Meta and others) aren't purpose-built for insurance. They're not trained and fine-tuned on insurance data to be contextually correct or to seamlessly integrate with insurers' existing systems. In short, solutions built around public AI are not viable for complex, regulated decision-critical environments like insurance.

Today, more than 20% of US-based insurers are attempting to build vertical Generative AI solutions in-house. However, developing and managing models with insurance domain expertise presents serious challenges arising from the need to invest in people with specialized knowledge, infrastructure and data resources.

As a result, less than one-tenth of these companies move their Generative AI solutions off the test bench and into production. **For more than 90% of insurers, building AI in-house is not the answer.** (Figure 3)

Moving from PoC to production

US Insurers developing Generative AI solutions in a test environment

	All Sizes	>$5 BN GWP
No	50%	14%
No, planned for end of year	29%	57%
Yes	21%	29%

US Insurers developing Generative AI solutions in a product environment

	All Sizes	>$5 BN GWP
No	74%	57%
No, planned for end of year	17%	29%
Yes	9%	14%

Legend:
- Yes
- No, planned for end of year
- No

Figure 3 (source: Oliver Wyman/Celent)

Every bit of information hidden inside unstructured data is an opportunity to enhance underwriting and claims operational efficiency. Generative AI solutions built for insurance can be the key to unlocking more of this valuable decision data.

Accenture predicts that Generative AI will potentially automate up to 62% of underwriting and claims processes. This would be a crucial step in augmenting the value created by human experts by allowing them to focus exclusively on higher-value tasks.

Generative AI: a foundation for success

▌ **In underwriting, <u>an insurance-specific Generative AI</u> solution creates opportunities for success, including:**

▸ **Faster speed-to-quote** by accurately extracting 90% more data from submissions with 95% accuracy while keeping up with demand from brokers (who are increasingly going digital) to improve the customer experience for their clients.

▸ **Increased capacity and lower administrative overhead** to reduce the time underwriters spend manually reviewing and inputting data from unstructured formats. Increased availability of structured data creates opportunities for AI and automation via improved data integration, predictive analytics and other benefits that McKinsey projects could reduce underwriting time by up to 70%.

▸ **Better profitability** from surfacing greater quantities of information within unstructured documents that can drive a 1-3% improvement in loss ratios.

▸ **Improved talent retention** to mitigate the effects of insurance's ongoing "labor crisis." In an Accenture/The Institutes survey, underwriters graded their employers' talent retention efforts "26 out of 100," due to job dissatisfaction from low-value work associated with reading or reworking documents.

▌ **In claims management, Generative AI for insurance gives claims teams greater control across the entire process to deliver:**

▸ **Improved combined ratios** by utilizing insights from unstructured data to reduce claims costs through subrogation and other claims recovery opportunities.

▸ **Enhanced customer experiences** by accelerating claims document handling time by 90% to settle claims faster. J.D. Power reports that almost 75% of policyholders dissatisfied with the processing of their claims are either changing providers or considering it.

▸ **Increased capacity** to contend with spikes in claims submissions driven by intensifying natural catastrophe risks.

▸ **Enhanced security and reduced compliance overhead** through greater quantities of decision data, which can yield insights to help insurers adhere to full process transparency, thereby enabling greater cost control while lowering users' risk exposure.

Generative AI has demonstrated remarkable capability in performing natural language tasks around unstructured data.

These models understand context, extract relevant information and generate human-like responses to user input, making them well-suited for tackling insurance's unique data challenges.

Continue reading to learn how Generative AI solutions like Roots Automation's InsurGPT™ can revolutionize processing unstructured data in underwriting and claims.

Part III.
Roots Automation's Generative AI solution: InsurGPT™

After reviewing the available options, it's time to consider the ideal solution for ingesting and analyzing unstructured, natural language data: **Generative AI**.

To start, here is a checklist of what to expect when selecting a solution for processing unstructured data.

Top-level features of an unstructured data processing solution in insurance:

✓ **Natural Language Processing:** A subset of Generative AI, Large Language Models (LLMs) understand and process human language, making them ideal for handling text-based unstructured data like emails, reports and social media posts.

✓ **Contextual Understanding:** These models can grasp the context of information, allowing for more accurate interpretation of complex unstructured data.

✓ **Information Extraction:** Generative AI can identify and extract relevant information from large volumes of unstructured data, saving time and reducing human error.

✓ **Data Synthesis:** LLMs can combine information from multiple sources to generate insights or summaries, aiding decision-making processes.

Functionality and operational capabilities:

✓ **Insurance context:** Created for insurance experts and trained on insurance data.

✓ **Adaptability:** Models that can be fine-tuned to specific insurance domains to improving their accuracy and relevance.

✓ **Scalability:** Ability to recognize and identify specific data from any source as volumes rise.

✓ **Continuous learning:** Adapting to new data sources without additional training, using the customer's data in a secure environment.

✓ **Federated knowledge:** Continuously and collaboratively training models while ensuring all data remains decentralized.

✓ **Integration with existing systems:** Delivering successful out-of-the-box user experiences using simple APIs.

✓ **Super-human performance:** Eliminating 90% of human oversight with clear exception-handling routines.

✓ **Security:** Cloud-native and delivered as a service for maximum "up-time" and aligned with users' internal policies and industry standards.

✓ **Compliance:** With local and federal data regulations and in-house standards for auditability and bias.

InsurGPT™: A Focused Solution

As discussed in Part II of this this publication, specialized Generative AI is emerging as the most effective tool for analyzing and processing unstructured data.

Roots Automation's AI platform, InsurGPT™, is designed from the ground up with insurance-specific capabilities to address the needs of underwriting and claims management.

InsurGPT™ is:

▸ **Pre-trained** on insurance industry data, minimizing onboarding time.

▸ **Focused on specific tasks** and built for rapid deployment to serve existing workflows.

▸ **Domain-specific,** as an LLM capable of recognizing a wide range of insurance data from unstructured sources and designed around insurance's unique compliance needs.

▸ **Capable** of 70% straight-through processing and 90% reduced reliance on manual intervention.

In addition to these specific insurance industry capabilities, InsurGPT™ is made to deliver rapid results and fast time-to-ROI.

It powers Roots Automation's Digital Coworkers, which are built to augment human expertise. Able to read, think and intuit just like people, Digital Coworkers come pre-packed with knowledge, skills and capabilities to process documents, automate business processes and collaborate using natural language.

InsurGPT™ delivers a simple, straightforward visual interface to optimize usability, problem-specific collaboration tools and Roots' patented "Human-in-the-Loop" technology for full transparency.

To date, companies using InsurGPT™-powered Digital Coworkers have realized substantial gains in efficiency, productivity and capacity across a wide range of use cases.

InsurGPT™ and Digital Coworkers are designed and built for the enterprise, wrapping around existing workflows, systems and applications. Roots Automation is compliant with ISO27001, NIST, HIPAA, CCPA, GDPR and NYCRR Part 500 standards and performs annual third-party reviews for SOC 2 Type 2 attestation.

Please click here to learn more about Roots' *Responsible, Ethical and Trustworthy AI Principles.*

Roots Automation Digital Coworkers

Automate underwriting:

▸ Complete ACORD forms.

▸ Pull critical documents (e.g. credit reports).

▸ Shortlist quotes for underwriter review.

▸ Create/send binders and certificates of insurance.

Transform claims management:

▸ Reduce claims overpayment.

▸ Virtually eliminate errors.

▸ Manage complex workflows.

▸ Increase agility and reduce team member burnout.

Comparing InsurGPT™

InsurGPT™ is proven to be better at delivering the decision data underwriting and claims teams need to boost profitability and efficiency, enhance customer experiences and sharpen their competitive edge for today's markets.

Features / Functionality	InsurGPT™	RPA+OCR (IDP/IPA)	Public AI	BPO	In-House Team
Domain/Industry-specific process and document knowledge—trained on millions of insurance documents and data points, annotated by insurance experts.	100%	0%	5%	50%	50%
Continually learning about document types, efficient data extraction techniques and business logic—then sharing (federating) this knowledge with all other customers.	100%	0%	0%	10%	10%
Reads structured documents (ACORD forms) and unstructured documents (policy documents, loss run reports) with 99% accuracy.	100%	0%	10%	25%	50%
Identifies screen elements within systems without formal training – to extract data from emails or bring online data sources together.	100%	0%	10%	10%	50%
Ability to receive, pre-process and classify documents without training, configuration or instruction - at scale.	100%	0%	5%	5%	25%
Ability to analyze a document in the context of the end-to-end lifecycle to understand what/how to process.	100%	0%	0%	75%	75%
Ability to auto-scale resources based on intelligent demand forecasting (i.e., meeting sudden spikes in demand or parallel processing of time-sensitive documents).	100%	10%	0%	0%	0%
Machine learning models trained and configured to work out of the box without significant integration with other technologies or dependencies.	100%	0%	10%	10%	25%
Uses conversational channels and natural language (not system logs) to communicate and manage exceptions with customer employees.	100%	0%	50%	100%	100%
Ability to analyze and extract data from any structured or unstructured document with 100% consistency (repeatability); no hallucinations.	98%	50%	25%	82%	85%
Requires minimal IT resources to set up, deploy and manage in an enterprise-grade production environment.	100%	10%	10%	75%	10%
Compliant with SOC 2 Type 2, ISO 27001, CCPS, HIPAA, NYCRR Part 500 and CDPR.	100%	0%	0%	50%	50%

Figure 4 (Roots Automation)

Download *The Buyer's Guide to AI-Powered Document Processing in Underwriting and Claims* for more information.

Part IV.
Conclusion

Insurance's unstructured data challenge is immense, but so are the opportunities for businesses that effectively harness this information.

Generative AI, particularly domain-specific tools like InsurGPT™, offers natural language input processing capability to deliver a clear path forward.

These AI-powered solutions let insurers unlock the value hidden in their documents by focusing on insurance's unique challenges and delivering critical decision data into underwriting and claims management workflows.

Improving efficiency in finding this data is the key to making insurance more customer-centric.

Empowering underwriting and claims experts to focus on creating greater value through customer engagement will result in mutual success through superior service and enhanced insurance products.

As the insurance industry evolves, companies leveraging tools to capture the most valuable insights from their data will have a significant competitive advantage.

Insurers that transform data processing challenges into decision-data solutions will significantly improve overall business performance today. These businesses are best positioned to successfully navigate the complexities of the modern data landscape and emerge more robust, efficient and competitive.
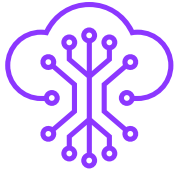
Roots
Automation

Engineered and built *by* insurance experts *for* insurance experts, Roots' purpose-built tools will deliver the competitive advantage your underwriting, claims and operations need to win in today's hyper-competitive markets.

Contact a Roots Automation expert to schedule a demonstration customized to address and solve your organization's unique unstructured data processing challenges.

Make Work More Human

Learn how at www.rootsautomation.com