

SHIFT

The State of AI in Insurance (Vol. VII)

Data Extraction from European
Accident Reports

www.shift-technology.com

Executive summary

- Many insurers still rely on reporting methods involving complex forms hand filled by the policyholder as part of the claims process
- The combination of form layout complexity and customer handwriting variations has made efficient and accurate automated data extraction difficult and time consuming to achieve
- Advances in OCR and generative AI (Gen AI) give insurers powerful new tools to automate data extraction, speeding the overall claims process

From the editor

While previous editions of this report have covered broad comparisons of the large language models (LLMs) impacting the insurance industry, there are times where it makes sense to focus on a single use case. For this edition of “The State of AI in Insurance” our data science and research teams are doing just that. Despite significant advances in digital first approaches to reporting claim details, hand-filled accident reports continue to be used by insurers around the globe. While this approach may seem antiquated, it has proven successful in helping insurers standardize the capture of relevant information required to settle claims on behalf of their policy holders. At the same time, the highly manual process of extracting the data contained within the report and making it useful to the insurer can significantly increase the time required to close a claim, greatly impacting efficiency and customer satisfaction.

Advances in optical character recognition (OCR) and Gen AI are providing new options for insurers seeking greater automation and efficiency when it comes to data extraction from hand-filled and complex-layout forms. Our researchers evaluated four different approaches to determine how insurers can best leverage these technologies to benefit the business and its customers.

Methodology

For this evaluation the research team focused specifically on the European Accident Report (EAR), a standard document common to nearly all European motor claims. This form is extremely dense with information, as the parties involved in a car accident provide details about their identities, insurance coverage, accident circumstances, and other facts. Although we focused on EARs for this evaluation, as indicated earlier, insurers worldwide rely on similar manual reporting as part of the claims process.

To conduct our evaluations we used the Shift European Accident Report Dataset (ShEARD) which is composed of 1,000 hand-filled French EAR pictures, complete with corresponding ground truth for numerous relevant fields. And while these are authentic EAR documents, the information contained within is manufactured. No Personally Identifying Information (PII) is used during this evaluation. All the results detailed in this report are derived from a subset of 100 EARs taken from the ShEARD dataset.

We evaluated the performance of five different approaches to automated data extraction:

OCR + Region Of Interest (ROI) Mapping

In this method, Gen AI is not used. Rather, researchers conduct a thorough ROI mapping and employ advanced OCR technology—specifically the Azure Document Intelligence (DocInt) Layout Reader—to extract the desired information. The Layout Reading mode of DocInt employs sophisticated techniques to extract not only the textual bounding boxes but also other essential elements of the document layout, including tables and checkboxes. The precise ROI mapping allows OCR to determine which element of the data model corresponds to each area of the document creating an accurate map of the information extracted by the OCR. It is important to note that even a small change in the form layout or an imprecise mapping might result in losing the information or mapping the wrong content into the data model.

Full Vision-based GenAI

This approach bypasses OCR entirely and instead sends the complete image of the form to a vision-capable LLM. To extract the plethora of information we need, it is essential to provide a comprehensive prompt that outlines the form's structure and specifies where to locate each element of the data model. This is a single-prompt and single-pass approach which allows for efficient API calls and ensures the method is completely agnostic of the OCR reading, language independent, and resilient to small differences in the form layout from one case to the other. Conversely, the approach is more difficult and more delicate to implement from a prompt engineering perspective as compared to a multi-call approach.

(Continued on next page)

(Continued)

ROI-mapping-assisted Vision-based GenAI

This method relies on OCR output using a basic reading—provided by Azure DocInt Read, which only returns the text bounding boxes and content. The ROI mapping technique allows for the isolation of specific sub-regions of the forms, such as the circumstances checkboxes or the policyholder information section, which in turn allows for targeted API calls for each area and thus providing the corresponding cropped image to a vision-based LLM. This multiple-call approach significantly reduces the prompt length and complexity, and enables the LLM to concentrate solely on the relevant region. Note that the ROIs for this approach are broader and, hence, less sensitive to layout fluctuations than those needed by the full OCR approach.

The Combined Prompt

Since the fields that DocInt was reading more accurately in the previously described approaches were the purely text-based entries (names and license plates) the combined prompt method provides the generative model with both the image and the OCR text extracted by DocInt. From the perspectives of cost and execution speed this is a convenient approach, since the simple read of DocInt is much faster and requires less of the Layout recognition that we needed earlier to identify the checkboxes.

Reasoning Models

The ability for certain LLMs to reason on the input prompt and create a self-improved rephrasing of the task with the goal of generating more accurate replies is the impetus for this evaluation. We wanted to determine if reasoning models could deliver better performances on data extraction for EARs. For this evaluation we used the same OCR + visual prompt and fed it to GPT o3.

A Note on Costs

The costs are expressed in USD cents per form treated (averaging on the dataset of 100 forms). Note that for Circumstances and Signatures we kept the old prompt without OCR text, since the information to extract is purely visual and no help would come from the OCR reading.

Results

Reading the Table

Evaluating performance is based on the specific use case and the relative performance achieved. The tables included in this report reflect that reality and are color-coded based on relative performance applied to the use case, with shades of blue representing the highest relative performance levels, shades of red representing subpar relative performance for the use case, and shades of white representing average relative performance.

As such, a performance rating of 90% may be coded red when 90% is the lowest performance rating for the range associated with the specific use case. And while 90% performance may be acceptable given the use case, it is still rated subpar relative to how the other approaches performed the defined task.

Field	OCR + ROI Mapping	full vision GenAI (GPT 4o)	ROI vision GenAI (GPT 4o)	Combined Prompt (GPT 4.1)	Reasoning Model (GPT o3)
First Name	80.30%	33.00%	73.90%	95.00%	95.00%
Family Name	74.20%	36.90%	70.40%	75.00%	70.00%
Plate Number	78.90%	14.80%	70.70%	85.00%	85.00%
Circumstances (all ok)	89.00%	14.00%	92.50%	95.00%	95.00%
Signature	87.50%	85.40%	93%	90.00%	95.00%
Avg cost per doc	1.00 \$c/form	3.21 \$c/form	3.74 \$c/form	1.50 \$c/form	13.03 \$c/form

Analysis

Our findings show that although all methods have unique strengths and weaknesses, the Combined Prompt delivered the best performance overall. However, while the results of the Reasoning Model approach are comparable with those of GPT 4.1 fed a combined prompt the associated cost is nearly 10x higher. This is due not only to the increased per-token price, but also because the “reasoning text” counts as completion (i.e. part of the reply). In addition the runtime is considerably longer. It is important to note that more recently introduced reasoning models, such as GPT5, allow user calibrated reasoning effort. However, while this can help lower costs, the relative performance gain would not be considered sufficient to justify adoption of reasoning models for the tasks covered in this report.

We will be publishing our analysis of reasoning models in the upcoming *“The State of AI in Insurance (Vol. VIII): Reasoning Models.”*

Conclusion

Our research continues to reveal that insurers can use multiple approaches to automate processes and that each has its own strengths and weaknesses. When evaluating which techniques are best to support their individual automation strategies, insurers must take several factors into consideration. In many cases, the perceived performance may come at too steep a cost.