

*Key data is trapped in documents, such as utility bills and business invoices. In the past year, GLYNT has built out an automated system to produce finance-grade sustainability data from these Primary Sources. Our annual reports showcases the performance improvements of our highly accurate machine learning and throughout our data pipeline.*

# GLYNT'S ANNUAL ACCURACY REPORT

SEPTEMBER 2022

## GLYNT'S 2022 ACCURACY REPORT

### Summary

---

Our planet is at a key moment: Climate change has caused havoc this summer, and businesses everywhere want to report, manage and reduce their emissions, water and waste. High quality data is essential.

This report details how GLYNT produces accurate, verified sustainability data. We'll show you how our advanced "Few Shot" machine learning extracts data from data from diverse documents, and how our platform extends that capability to deliver structured, ready-to-consume data.

And – key to this report and the work we do – we'll show you how GLYNT measures our accuracy every step of the way.

**At this moment, highly accurate sustainability data is essential to every business, and the planet too.**

At GLYNT we worry about accuracy so you can get started on reductions. Together we'll measure and manage to a real change.

As always, reach out with any comments or questions. We'd love to hear from you.

*The GLYNT Team*



## 2022: Higher Accuracy

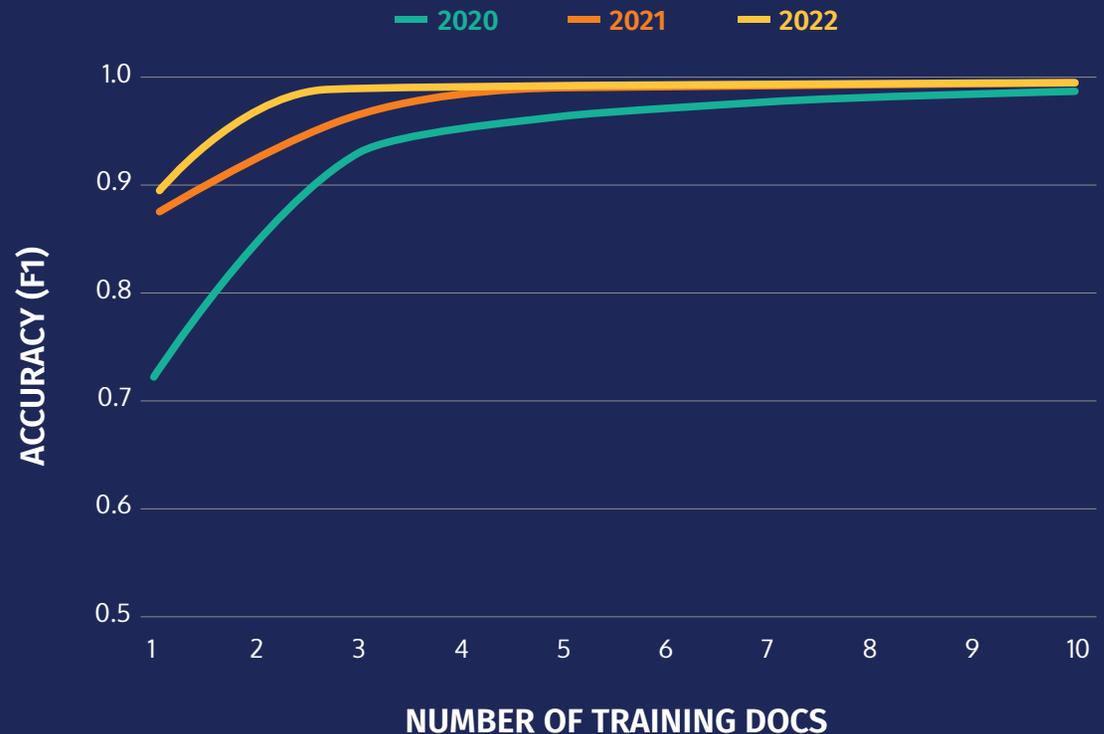
Let's jump to the bottom line: GLYNT is more accurate than ever before. We tested against a range of utility bills and business invoices, and included scans of documents and original issue PDFs.

The results show another year of strong accuracy gains. First, after training on just one document, GLYNT is averaging 0.75 F1. And after training on three documents, we're hitting an average of 0.99 F1.

In sum, the 2022 results show a better starting point, faster learning and a quick path to peak performance.

GLYNT: Strong performance, year after year. And always getting better.

Note: F1 is a standard measure of accuracy in machine learning. See page 12 for details.



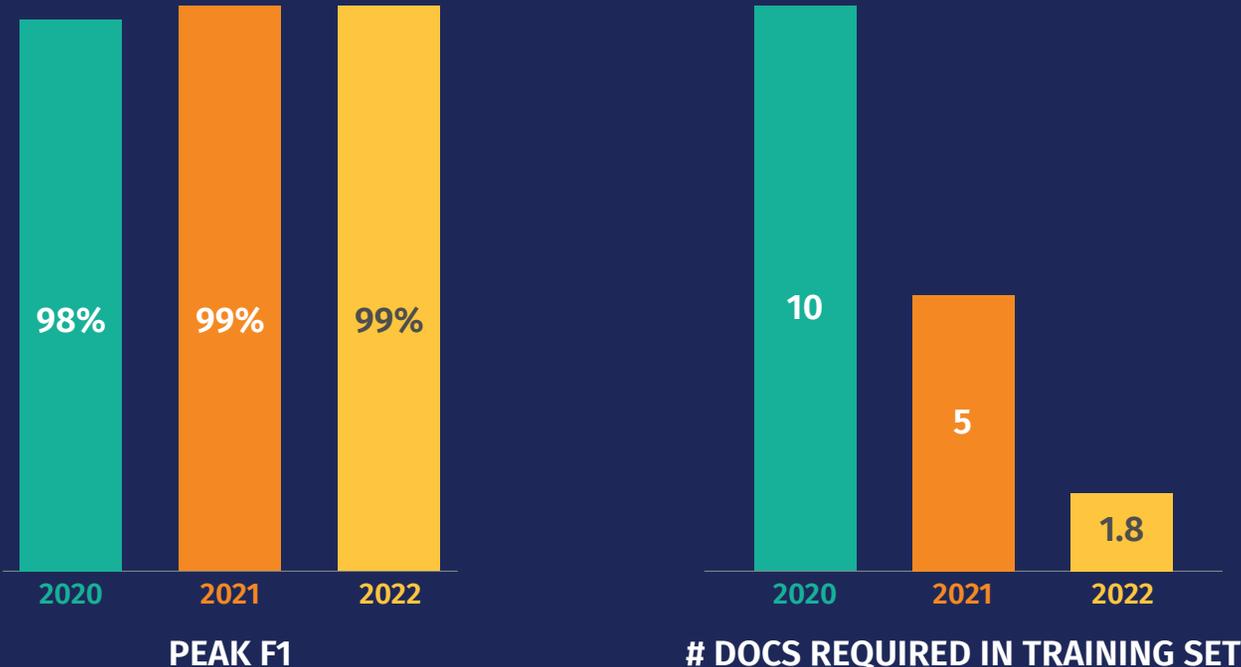
## Efficiency Gains Too!

And here's another way to see GLYNT's improvement. What is Peak Accuracy for GLYNT? How many documents does it take to get to Peak Accuracy? The answers are in.

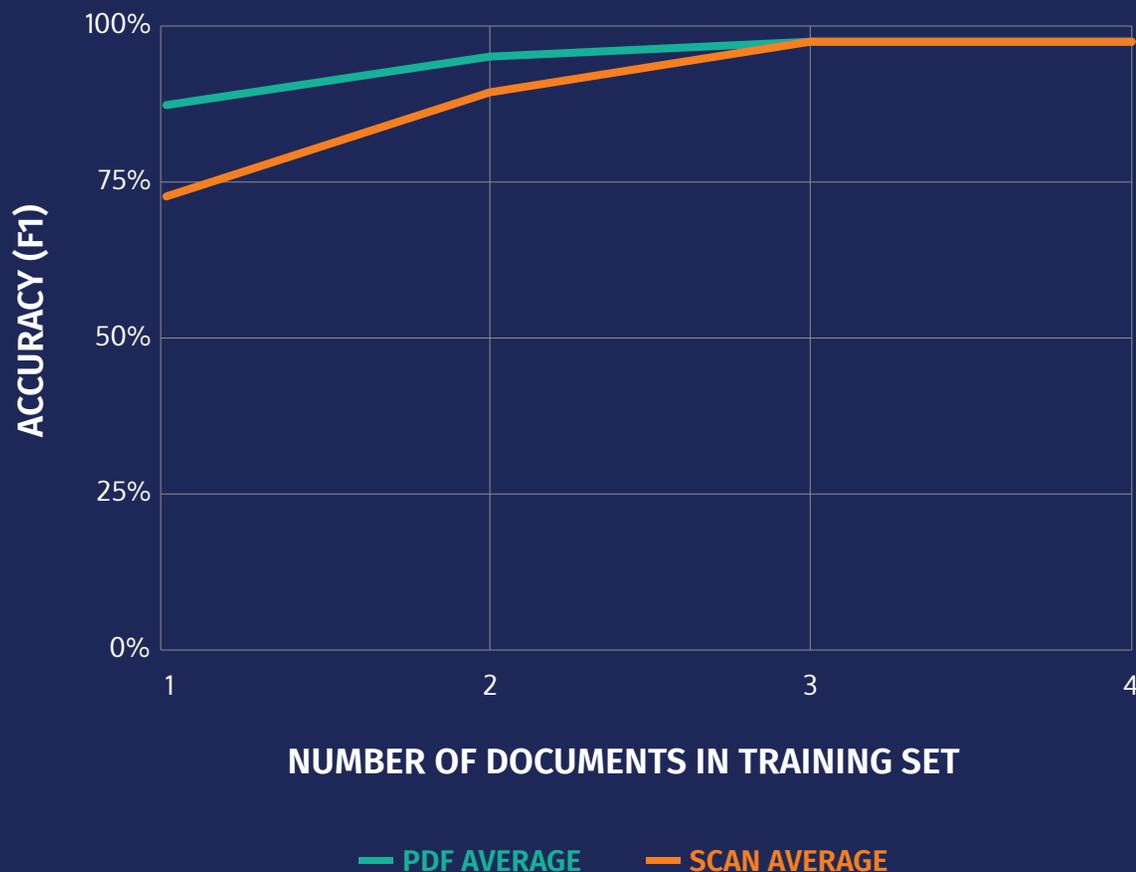
As the bar chart shows, GLYNT's F1 stayed at 99% in the past year. And GLYNT is learning faster than ever. The number of documents needed for Peak Accuracy fell by 65%

Provide 2 – 4 sample documents, and our AI will get to Peak Accuracy. Not every document set is the same, and these are our average results, but they are indeed impressive.

Our performance improvements show the power of automation at work. Fewer sample data points means a shorter time to first data, and more accurate data overall. GLYNT saves time and money.



## Scans Quickly Catch Up to PDFs



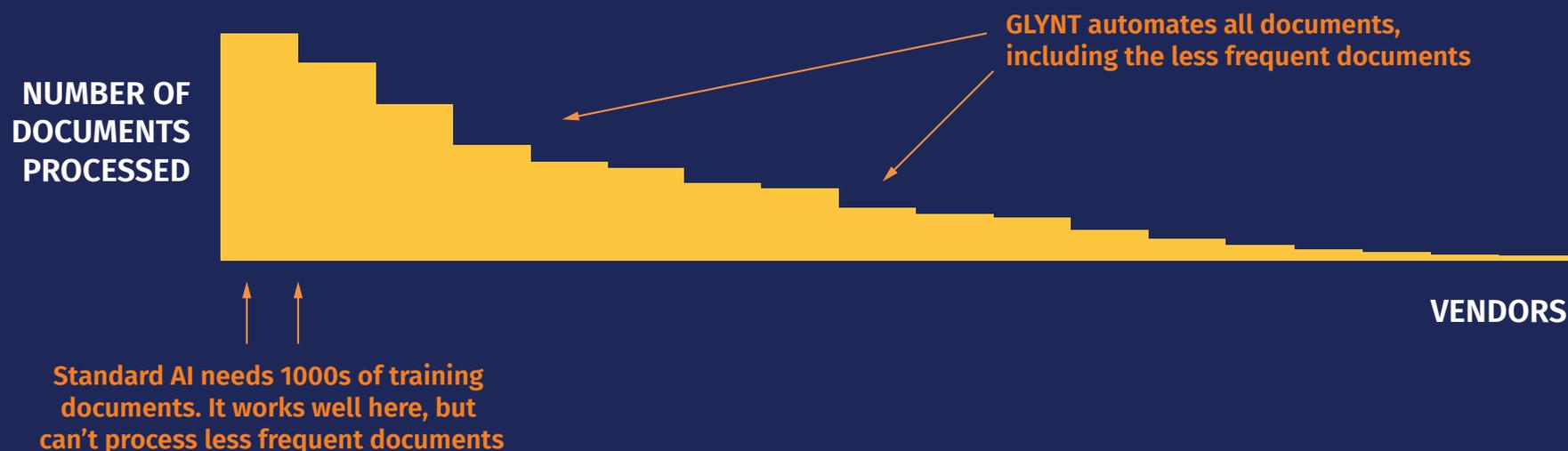
We live in a digital world, so it may surprise you to learn that nearly 50% of all invoices are issued in paper format and mailed. (Ardent Partners, [The State of Payables 2022](#))

The graph on the left shows that GLYNT works quite well for both scanned paper documents (e.g. digital images) and native PDFs.

Scanned documents can be quite poor in quality, but Optical Character Recognition (OCR) has been improving rapidly, so GLYNT now processes documents at 200dpi and below with great accuracy.

Whether scanned or native PDF, GLYNT provides the same great results.

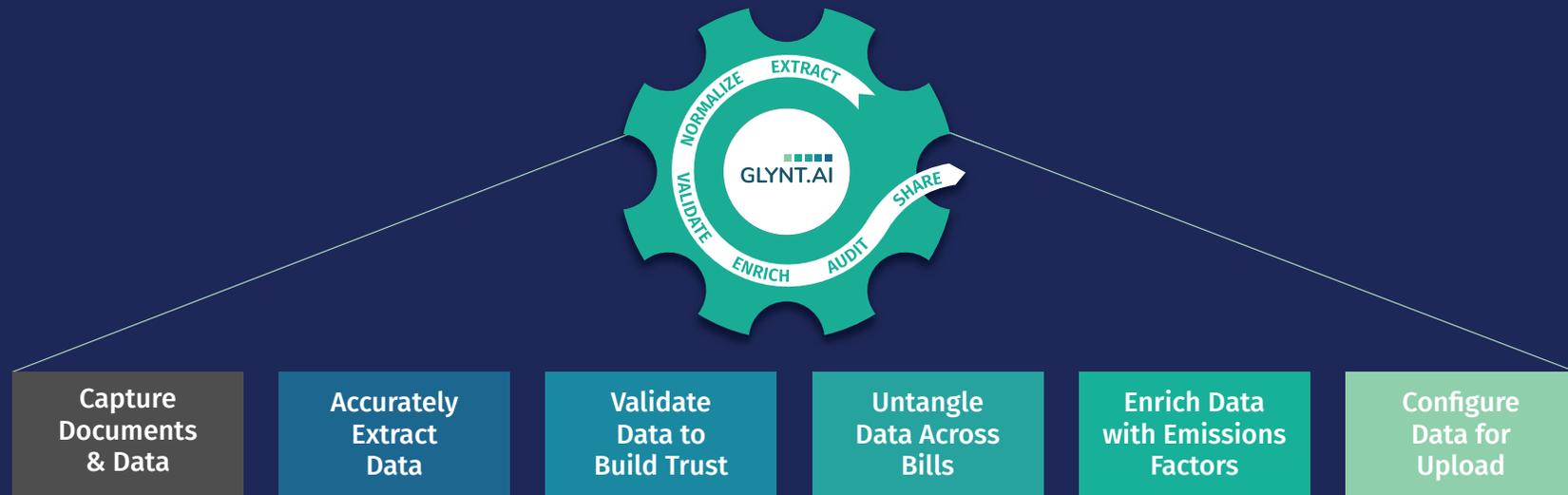
## “A Mile Wide and an Inch Deep”



There are over 60,000 water and energy utilities in the U.S. alone. As the graph above shows, Standard AI requires a large volume of training documents. But this can be quite hard to come by. For example, telco transmission towers are everywhere, including rural electric coop districts. There is just no way to get a lot of training documents. GLYNT built our advanced machine learning system, ‘Few Shot’ ML, to conquer this exact challenge.

And as blog posts by the venture capitalists [AI6z](#) point out, the economics of standard AI in a long-tailed market can be quite challenging. In contrast, GLYNT’s approach delivers high accuracy, very quickly from training on just a few documents that our customers provide. GLYNT fits markets that have enormous document variation. We’re a single solution provide for the entire long tail, where document variation is “A mile wide and an inch deep.”

# Accuracy and Data Quality



## GLYNT's End-to-End ML and Data Pipeline

GLYNT's advanced 'Few Shot' ML accurately extracts data from documents, which is shown as the second step in the data preparation flow above. The accuracy results reported thus far in this report are for data processed at this step. But the data GLYNT customers receive has been additionally processed, as shown by the steps to the right of accurate extraction. Measuring extraction accuracy is thus just a piece of a larger data quality rubric.

As an example, consider document capture. GLYNT customers upload documents to our platform and forward emails with PDFs attached. Or GLYNT picks up documents from the customer's secure location, or the customer provides GLYNT a set of utility credentials and GLYNT fetches the utility bill and brings it into our platform. With so many entry points for the document flow, and with the vagaries of how utility bills are issued and posted by utilities, it is easy to have gaps in a sequence of documents. For example, all invoices for 2021 might be present, with the exception of February 2021 for Vendor X. Without data quality and verification on the capture step, the GLYNT customer does not receive fully useful data. So the rubric for accurate and complete data must be expanded beyond the extraction step.

## GLYNT's End-to-End Testing of Data Quality

COMPONENT OF DATA QUALITY	DESCRIPTION	TEST
<b>Accuracy</b>	Is the data extracted accurately from the document?	GLYNT uses a suite of error trapping tests to identify exceptions in Extracted Data. The best way to prevent an accuracy error is not to make one, and while GLYNT's ML is 98%+ accurate from first data, we are constantly striving for higher accuracy rates.
<b>Mapping</b>	Are the data items in the Extracted Data correctly matched to the appropriate field name?	GLYNT uses a suite of error trapping tests to identify exceptions in Extracted Data. These test reflect our years of experience with utility bills and complex energy invoices.
<b>Completeness</b>	Have all of the data items needed for the Structured Data output been extracted from the document?	There is a precisely defined list of data items in Extracted Data that are needed to produce Structured Data. We test against that list.
<b>Validity</b>	Are the Structured Data items in the correct data format?	Every data item is tested for compliance with the expected data format.
<b>Audit</b>	Is the business logic used to produce Structured Data correct?	GLYNT logs and monitors data handling throughout our system, and uses a suite of software tests to validate business logic.
<b>Usability</b>	Is the Structured Data compliant with the Data Dictionary?	GLYNT uses a suite of software tests to validate the Structured Data output against the specifications in the Data Dictionary.
<b>Lineage</b>	Can the Structured Data be 'reverse engineered', step by step, back to the data printed on the document?	GLYNT produces an Audit Trail for every data item that enables this type of traceback to the document.

GLYNT is an automated platform that produces investor-grade data from primary sources. This is a multi-step process, and of course the best way to prevent errors is to test the data after every step. But this is not enough.

To thoroughly test GLYNT data, we take on the customer's perspective: Is this data verified, complete and ready to use? By meeting this higher level of data quality, GLYNT earns the role of Trusted Data.

The table to the left shows how GLYNT tests data for accuracy and completeness. The descriptions refer to two stages of data preparation:

- **Extracted Data.** This is the unstructured data that is taken from the printed document by our advanced ML technology, and
- **Structured Data.** This is the ready-to-use data delivered to GLYNT's customers

Trust is earned, and our end-to-end data review is part of a suite of actions to earn that respect.

# Data Ready for Sustainability and Climate Risk Disclosures

In March 2022, the U.S. Security and Exchange Commission (SEC), issued a proposed set of regulations for climate risk disclosures. The required reporting would cover three main areas: Reporting of emissions, emission reduction plans, and climate risk stress tests. The final SEC regulations are expected to be issued in late 2022 and will impact public disclosures of 2023 data, with first reports in early 2024. There is still uncertainty about the shape of the final ruling and possible legal challenges, but regardless of those outcomes, the proposed regulations have raised an important issue:

## *How does one know that sustainability data is ready for public disclosure? What is the C-suite exposure?*

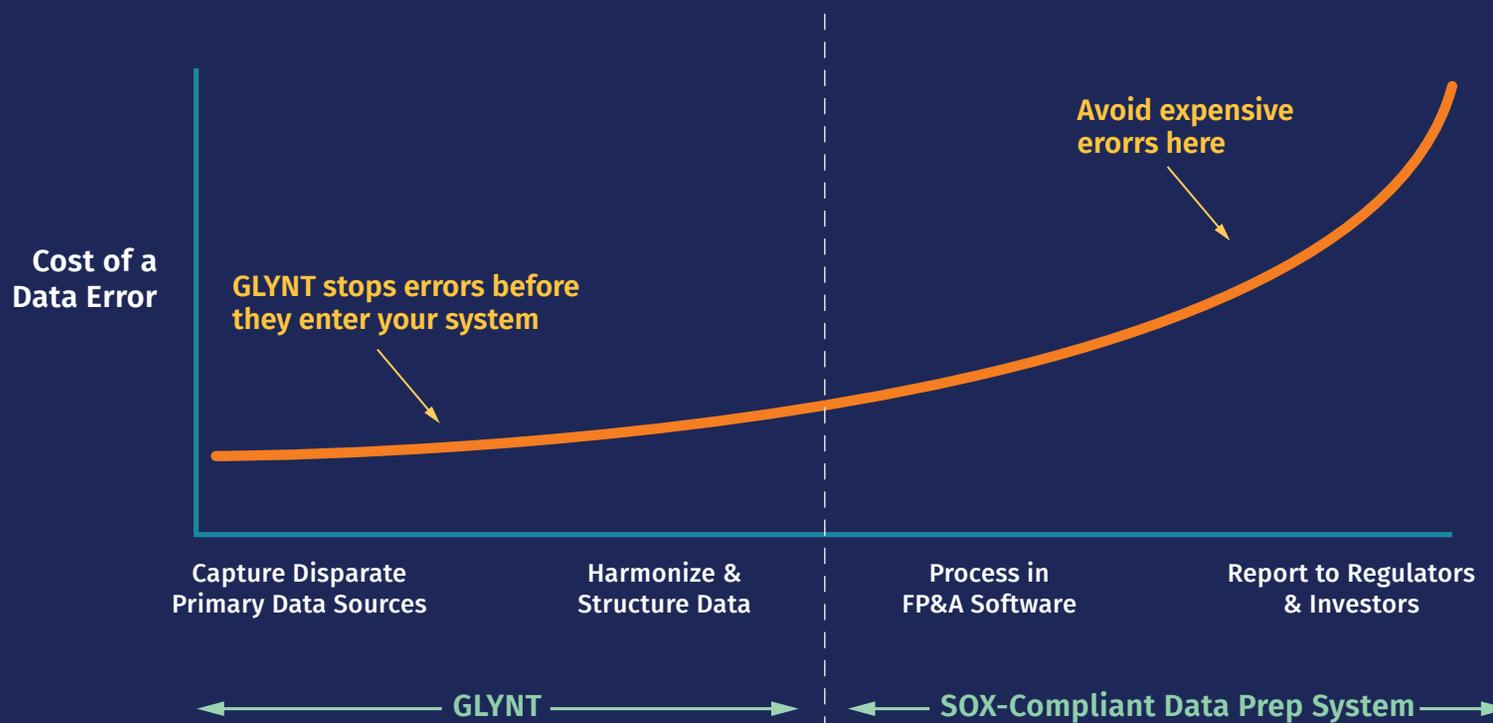
A survey by the [World Economic Forum](#) in Q2'2022 shows the extent of the challenge: 92% of businesses surveyed did not use software to prepare ESG disclosures, which include sustainability data. The norm is manual data entry from documents and spreadsheets. But under Sarbanes Oxley (SOX) laws, top corporate officers are personally liable for misleading, fraudulent or negligent disclosures. Data preparation by spreadsheet causes major heartburn!

GLYNT is currently SOC2, Type II certified and HIPPA compliant. We recognize this C-Suite pain and are working with leading auditors to establish the framework for SOX compliant sustainability data. We'll report out this market-leading framework later in 2022. Meanwhile, GLYNT invests in security, data handling and user management so that third-party auditors can briskly process GLYNT sustainability data.



## Why GLYNT's Focus on Accuracy Pays Off

Late-stage errors are costly!



There's a reason why GLYNT is so focused on accuracy: Errors are very costly to our customers. We focus on rigorous, validated data production and data quality so that our customers get a better experience using GLYNT data and lower costs of operation. If it costs GLYNT 1/10th X to fix an error in our part of the process, and it costs our customers 5X when an error arises, GLYNT's 'overinvestment' in data quality makes sense. We put ourselves in our customer's shoes.

With high-quality data customers can report with confidence and also connect to financial opportunities. For example, GLYNT partners with companies that use analytics to estimate energy savings at a property. They are good at their work and have financial backers who will do project financing based on their modeling results. GLYNT is delighted to feed this path of data to operational savings. Accuracy pays off.

## Study Details

---

The documents used in this study are representative of the types of documents GLYNT processes for customers everyday, including business invoices and utility bills (water, electricity, natural gas and waste).

There were 6 different document sets in total, and they were split evenly between scans and native PDFs.

For each document, 25 – 45 fields were selected, also representative of the type of fields GLYNT typical extracts. These include long text strings (such as account numbers), phrases (such as utility tariffs or product descriptions), amounts, dates and so on. The documents were 2 – 8 pages long, and fields from various pages were used.

## The Method

The test method was as follows:

- A few documents were placed in a training set.
- The first document was “marked” with the desired fields using the GLYNT application.
- GLYNT’s ML was trained. This takes just a few minutes.
- Data on 12 – 18 documents of the same type were extracted using the trained ML model.
- The accuracy results were calculated and the user added a second document to the training model and marked it appropriately.
- A second model was trained.
- Data were extracted and results scored.
- This process continued until the extracted data reached 99% accuracy.
- As the report shows, this requires training on just a few documents.



## The Metrics

To measure performance, a Ground Truth data set was developed for each document type. Ground Truth is the 100% correct data for the desired fields, including whether a desired field was printed on the document. There is some natural variation on which data fields are printed on otherwise remarkably similar documents.

With Ground Truth in hand, each set of extracted data was compared to the Ground Truth and the following states per data item were noted:

**Correct** = the data item was provided by the ML and is 100% accurate (as measured at the character level)

**Incorrect** = the data item was provided by the ML and contains one or more character errors

**Missing** = the data item should have been provided

**Not Printed** = the data item was not available and should not have been returned

Incorrect errors largely arise through OCR errors. For example, the OCR engine confused an 0 with an O. As OCR engines improve, the amount of incorrect data has fallen significantly.

To summarize the accuracy results, the standard F1 score is calculated as follows:

**Precision** = # Data Items Correct / Total # Data Items

**Recall** = (Total # Data Items - # Data Items Missing) / Total # Data Items

**F1**:  $[2 \times \text{Precision} \times \text{Recall}] / [\text{Precision} + \text{Recall}]$



## Always a Win-Win

---

### #betterdatafortheplanet

Get sustainability data that is better, faster and cheaper with GLYNT.

And don't forget to double the celebration. Your data-driven actions become part of the planetary change we need.

### Break Through the Sustainability Data Bottleneck

TALK TO GLYNT TODAY

