WHITE PAPER / DOCUMENT INTELLIGENCE

The right number in the wrong row

A financial report can contain the right figure and still attach it to the wrong period. We built a controlled test to see which kinds of evidence checks catch that error.

A citation can point to the right document and still miss the fact

Consider two consecutive days in a rental ledger. On January 1 the operator recorded 985 rides. On January 2 it recorded 801. A summary claiming 801 rides on January 1 contains a real date and a real number. A reviewer searching the file for either one will find it. The error is in the relationship between them.

This distinction matters for document assistants that answer questions about reports, contracts, or accounts. The ICLR paper *TabFact* studies whether statements follow from evidence in tables. We took one narrow part of that problem, a count attached to a date, and built an evaluation whose errors a reviewer could inspect one by one.

The experiment

We used 731 daily records from the public Capital Bikeshare dataset. For each of the first 730 dates, we generated a correct claim using its recorded count and a second claim using the following day's count. We checked that the neighboring values differed. Both dates and both values occur in the same source.

We compared a lookup that accepts any count found in the document with one that resolves the date first and then compares the count in that row. Each check received the date and count as structured inputs. This isolates the verification step downstream of extraction.

What changed when we required row evidence

Results on 1,460 constructed claims

Evidence rule	True claims accepted	False claims accepted
Value in document	730 / 730	730 / 730
Value in dated row	730 / 730	0 / 730

The test was built to expose a specific failure. The document check passed every swapped value because each value came from that very document. A dated lookup caught all 730 swaps while preserving every correct claim. This is a regression test for a known error class. It measures neither free-text extraction accuracy nor performance on arbitrary questions.

A release decision for document teams

For questions about a single figure, retain the document version, row identifier, column label, and unit with the proposed answer. A reviewer can then reconstruct the claim without searching an entire attachment. When the date refers to two rows, route the claim for review before showing a figure.

For a first production evaluation, draw examples from the actual queries. Include amended filings, adjacent reporting periods, missing values, and tables that repeat a date across business units. Score the extraction and the final evidence check separately. That separation tells the team where a fix belongs.

Research & data

- TabFact · ICLR 2020
- Bike Sharing · Hadi Fanaee-T, UCI Machine Learning Repository, CC BY 4.0