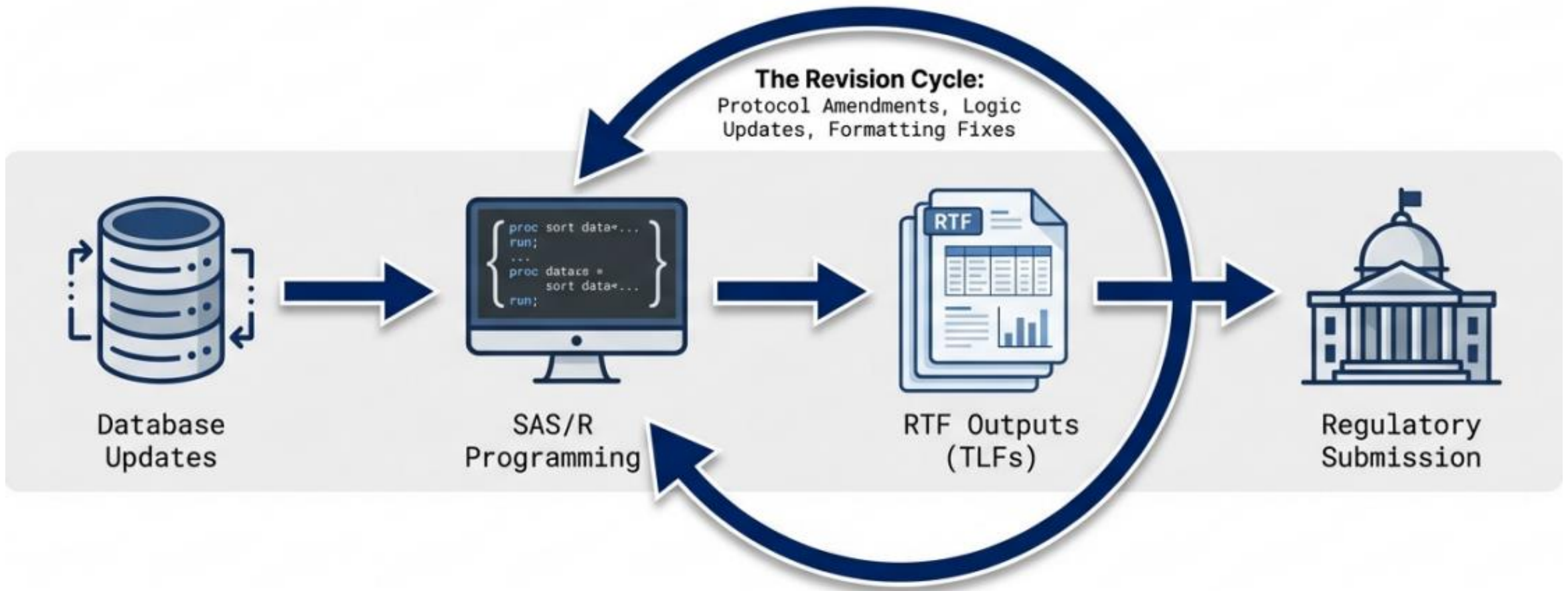


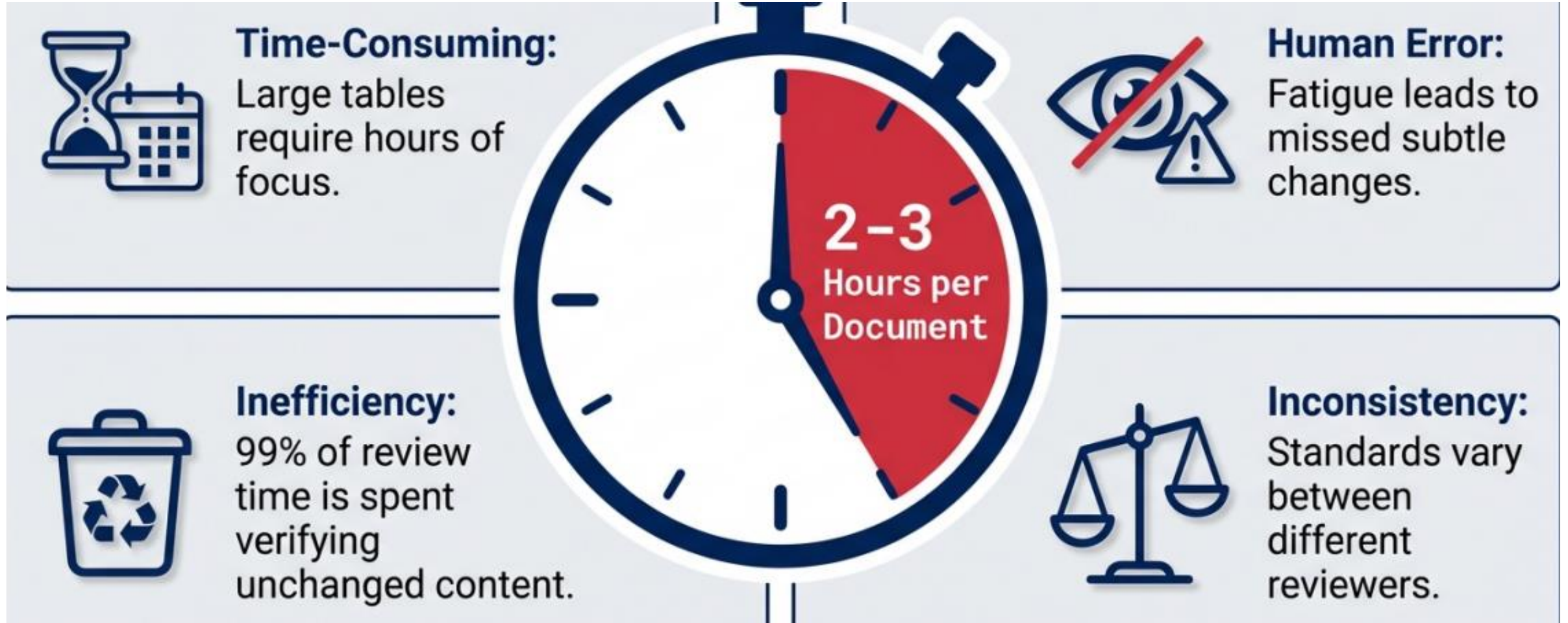
OSo8:Deceptive Twins: An Interactive Python Tool for Automated RTF Document Comparison in Clinical Submissions

Aravind Gajula, Bristol Myers Squibb,
Lawrenceville, NJ, USA
Indraneel Narisetty, Gilead Sciences Inc.,
Foster City, USA

The Iterate-Regenerate Cycle in Clinical Reporting



Risk/Challenges of Manual Review



Defining Deceptive Twins

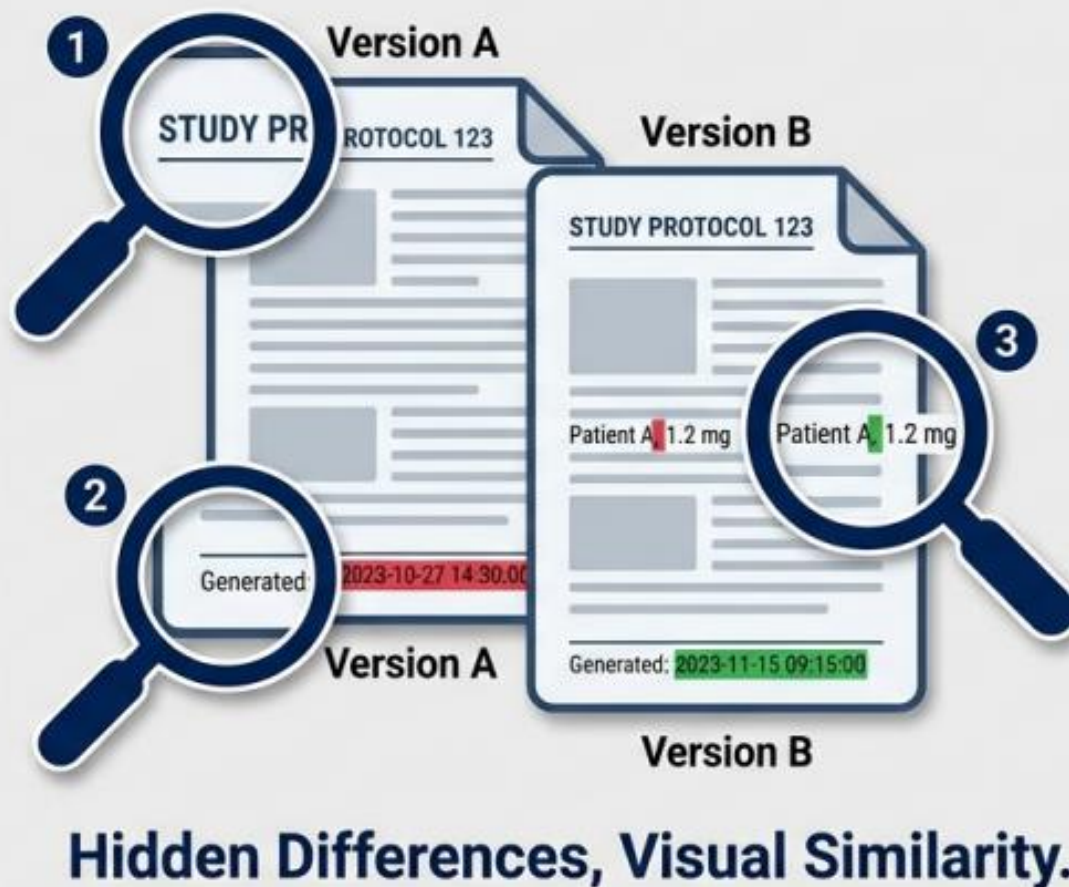
The Paradox:

When a dataset refreshes, the output file is technically “new” (new timestamp, new binary encoding) even if the **text content is identical**.

Definition:

Deceptive Twins are document versions that:

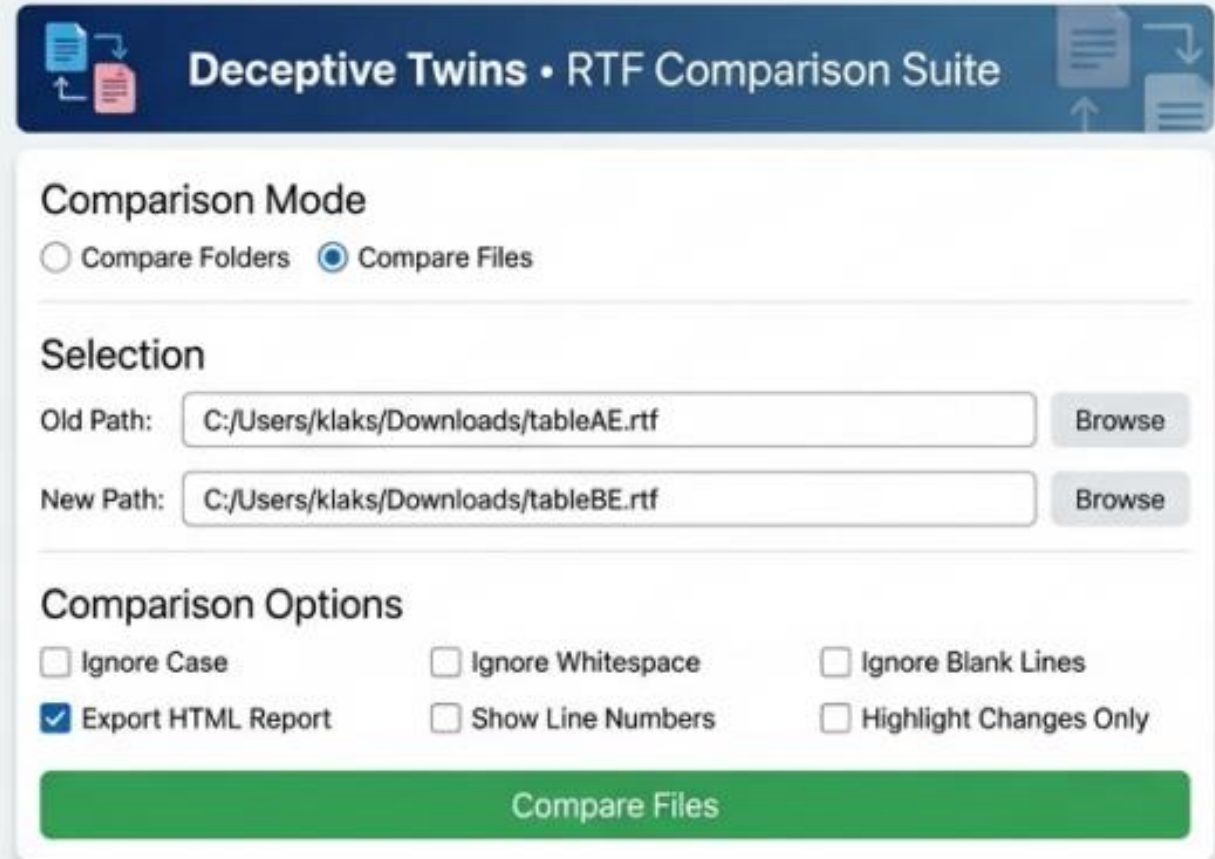
- Look **identical** at first glance.
- Have different metadata/timestamps.
- May conceal subtle critical differences (e.g., a decimal point shift).



Introducing the Python application: Deceptive Twins

Core Capabilities

- Eliminates manual visual inspection.
- Extracts text from binary RTF to compare word-for-word.
- Clearly highlights insertions (Green) and deletions (Red).
- Creates permanent **HTML** audit trails.



The screenshot shows the 'Deceptive Twins - RTF Comparison Suite' application window. It features a dark blue header with the application name and icons. Below the header, there are three main sections: 'Comparison Mode', 'Selection', and 'Comparison Options'. In the 'Comparison Mode' section, 'Compare Files' is selected with a radio button. The 'Selection' section contains two text input fields for 'Old Path' and 'New Path', both containing file paths, with 'Browse' buttons next to them. The 'Comparison Options' section has six checkboxes: 'Ignore Case', 'Ignore Whitespace', 'Ignore Blank Lines', 'Export HTML Report' (which is checked), 'Show Line Numbers', and 'Highlight Changes Only'. A large green button labeled 'Compare Files' is at the bottom.

Deceptive Twins • RTF Comparison Suite

Comparison Mode

☐ Compare Folders ☒ Compare Files

Selection

Old Path:

New Path:

Comparison Options

☐ Ignore Case ☐ Ignore Whitespace ☐ Ignore Blank Lines

☒ Export HTML Report ☐ Show Line Numbers ☐ Highlight Changes Only

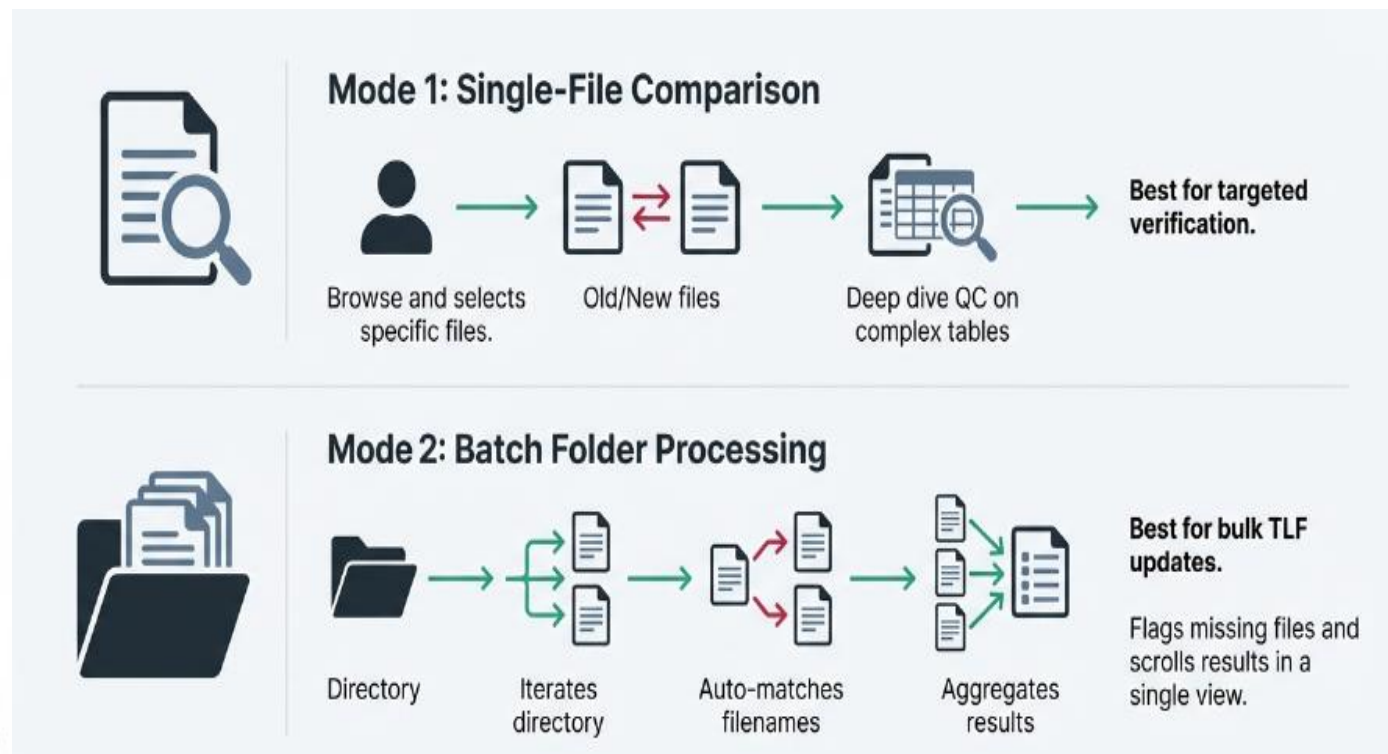
Flexible Comparison : Files or Batch Folders

Compare Files Mode:

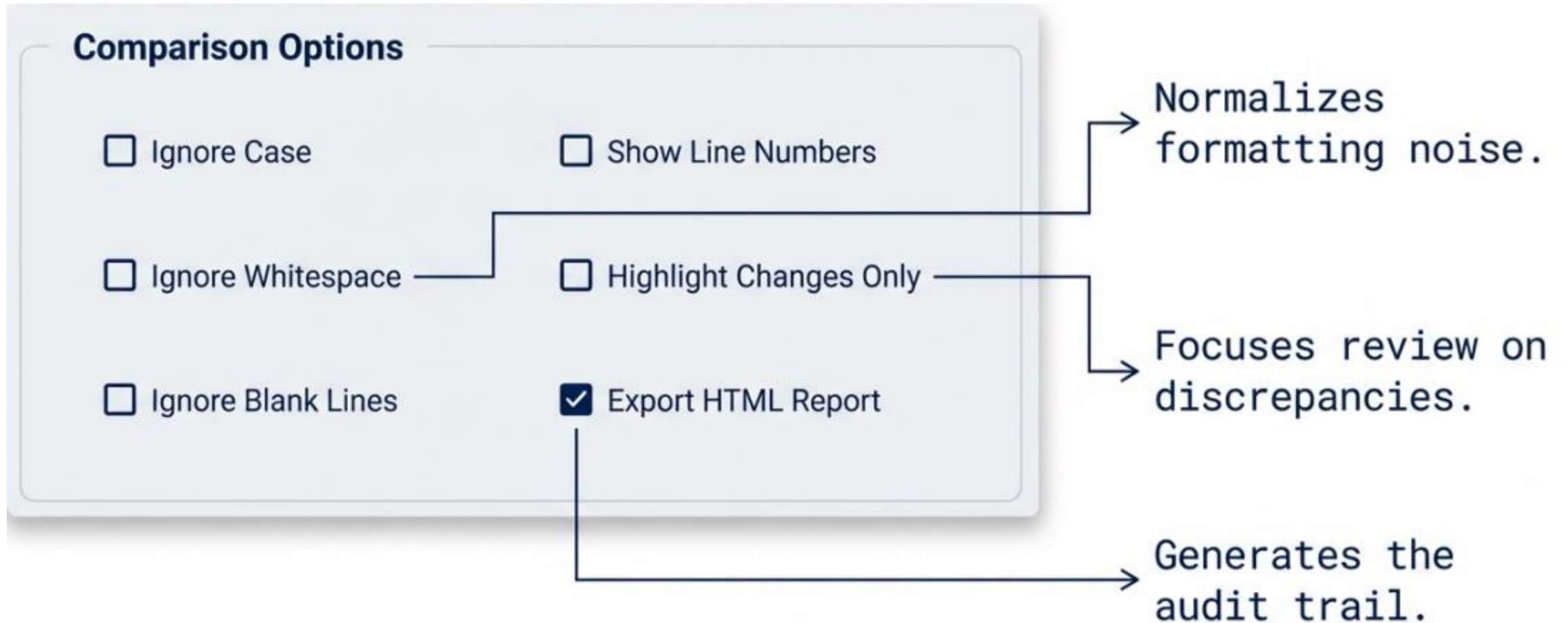
Direct comparison of one "Old" vs one "New" document. Best for spot-checking.

Compare Folders Mode:

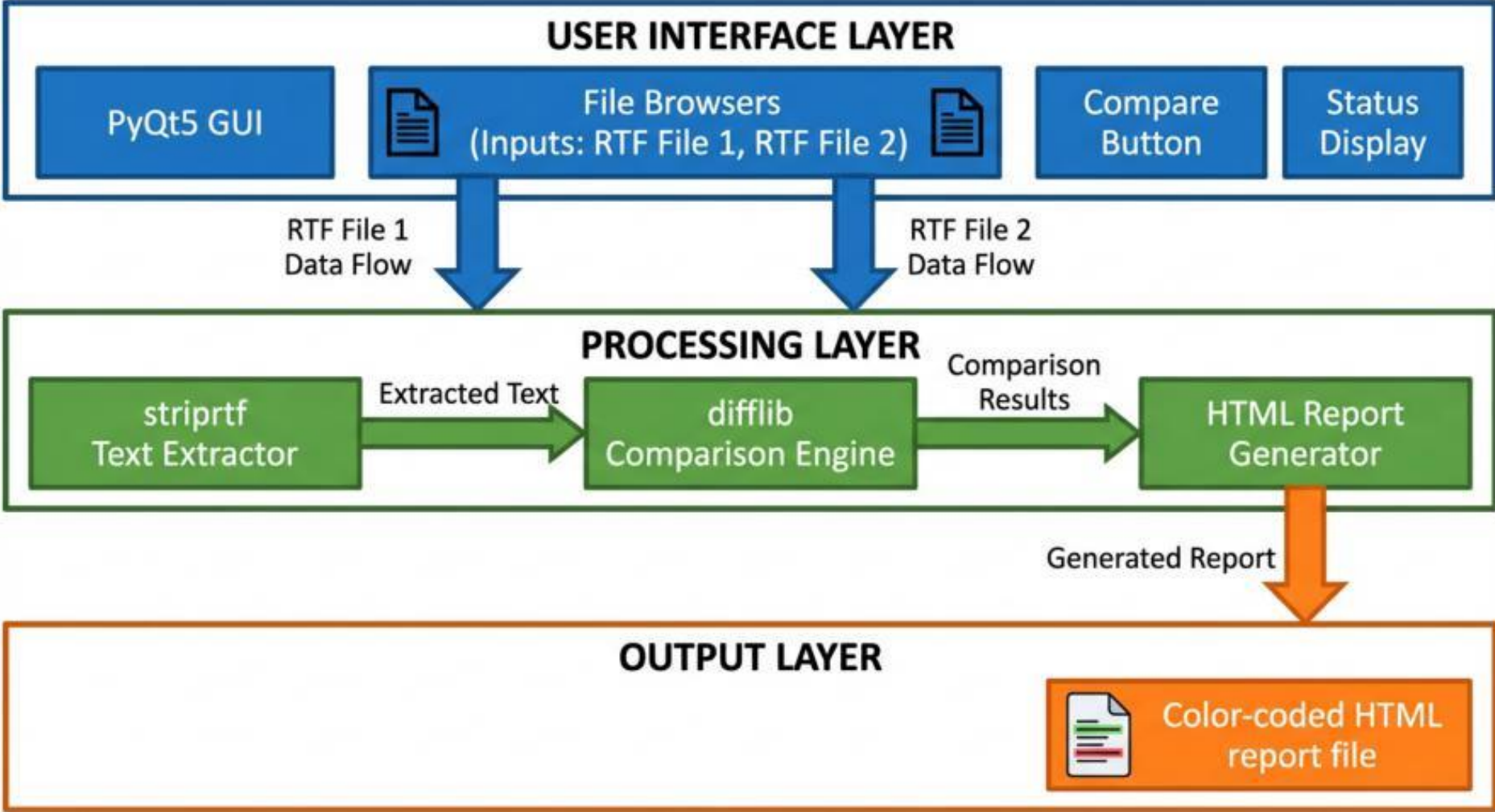
Batch processing for entire studies. Automatically matches files by name, iterates through directories, and aggregates results into a single view.



Configuring the Comparison Criteria



Software Architecture



Demo

The left panel contains mode selection controls (folder vs. file comparison) and path selectors for old and new versions. A comprehensive options section provides checkboxes for customizing comparison behavior, including case sensitivity, whitespace handling, blank line treatment, line numbering, and change-only highlighting. The right panel displays inline comparison results with color-coded differences, allowing immediate visual inspection without leaving the application.

The interface is divided into two main panels. The left panel contains controls for file comparison, and the right panel displays the comparison results.

Left Panel Controls:

- Mode Selection:** Radio buttons for "Compare Folders" (selected) and "Compare Files".
- Selection:** Fields for "Old Path" and "New Path", each with a "Browse" button. The paths shown are `C:/Users/klaks/Downloads/tableAE.rtf` and `C:/Users/klaks/Downloads/tableAE_rtf`.
- Comparison Options:** Checkboxes for "Ignore Case", "Ignore Whitespace", "Ignore Blank Lines", "Show Line Numbers", "Export HTML Report" (checked), and "Highlight Changes Only".
- Compare Files:** A large green button at the bottom.

Right Panel Results:

Legend: Unchanged Deleted Added Modified

Options Applied: Standard Comparison

File: tableAE.rtf

Old Version	New Version
MedDRA system organ class and	MedDRA system organ class and
Preferred Term NCI-CTCAE Grade Placebo	Preferred Term NCI-CTCAE Grade Placebo
(N= 32) CMP-135	(N= 32) CMP-135
(N=33)	(N=33)
-Any adverse events- All grades 27 (84.4%) 32 (97.0%)	-Any adverse events- All grades 25 (84.4%) 32 (97.0%)
1 9 (28.1%) 6 (18.2%)	1 9 (28.1%) 6 (18.2%)
2 13 (40.6%) 19 (57.6%)	2 13 (40.6%) 19 (57.6%)
3 5 (15.6%) 6 (18.2%)	3 5 (15.6%) 6 (18.2%)
4 (0.0%) 1 (3.0%)	4 (0.0%) 1 (3.0%)

Demo

RTF Comparison Tool

Deceptive Twins • RTF Comparison Suite

Comparison Mode

Compare Folders

Compare Files

Selection

Old Path:

C:/Users/klaks/Downloads/Version1

Browse

New Path:

C:/Users/klaks/Downloads/Version2

Browse

Comparison Options

☐ Ignore Case

☐ Ignore Whitespace

☐ Ignore Blank Lines

☒ Export HTML Report

☐ Show Line Numbers

☐ Highlight Changes Only

Compare Folders

Comparison completed

Comparison Results

Found differences in 1 file(s)

Clear

Legend: Unchanged Deleted Added Modified

Options Applied: Standard Comparison

File: tableAE.rtf

Old Version	New Version
MedDRA system organ class and	MedDRA system organ class and
Preferred Term NCI-CTCAE Grade Placebo	Preferred Term NCI-CTCAE Grade Placebo
(N= 32) CMP-135	(N= 32) CMP-135
(N=33)	(N=33)
-Any adverse events- All grades 27 (84.4%) 32 (97.0%)	-Any adverse events- All grades 25 (84.4%) 32 (97.0%)
1 9 (28.1%) 6 (18.2%)	1 9 (28.1%) 6 (18.2%)
2 13 (40.6%) 19 (57.6%)	2 13 (40.6%) 19 (57.6%)
3 5 (15.6%) 6 (18.2%)	3 5 (15.6%) 6 (18.2%)
4 (0.0%) 1 (3.0%)	4 (0.0%) 1 (3.0%)

 Bristol Myers Squibb™

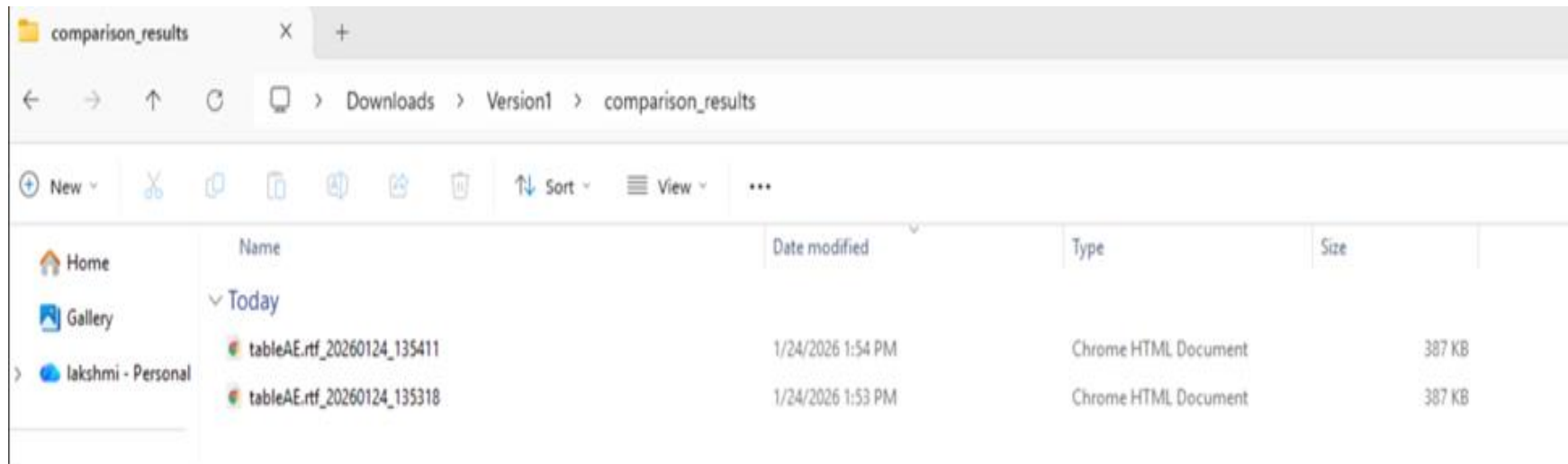
Global Biometrics and Data Sciences

Highly Confidential

10

Comparison results

When the "Export HTML Report" option is enabled, the tool generates a standalone HTML file in a timestamped subdirectory named "comparison results" adjacent to the source files. This report opens automatically in the default web browser, providing a shareable artifact for documentation and review. A summary counter reports the total number of files with detected differences. Individual HTML reports generate for each changed file when export is enabled, allowing reviewers to focus on specific files of interest. Files present in the old folder but missing from the new folder receive special notation, alerting users to potentially deleted outputs.



Efficiency Gains

- 70–90% reduction in document review time
- Fewer missed changes through systematic word-level analysis
- Permanent, auditable QC documentation via HTML reports

Metric	Manual Review	Deceptive Twins
Average time per document	15–30 minutes	1–2 minutes
Missed changes (error rate)	5–10%	<1%
Reviewer fatigue impact	High	Minimal
Documentation quality	Variable	Consistent

Contact



Contact Information:

Aravind Gajula, Bristol Myers Squibb
Email: aravind.gajula@bms.com

Indraneel Narisetty, Gilead Sciences Inc.,
Email: Indraneel.narisetty@gilead.com