Appendix E: Procedure for manual examination of matching claims in disagreement

Figure A5 illustrates the cases where manual examination was necessary and the detailed procedure. Out of a total of 11,639 claims, an automated process identified 749 matching claims, which accounted for approximately 6.4% of the total. Among these 749 claims, 228 displayed conflicting ratings when compared to the original ratings provided by Snopes and PolitiFact.

To examine these 228 cases where there was disagreement, a manual investigation was conducted to ascertain the reasons behind the varying ratings from Snopes and PolitiFact. First, we verified whether the claims from Snopes and PolitiFact were indeed matching. To accomplish this, one undergraduate student from our lab and the first author independently examined the 228 matching claims and labeled whether they were indeed matching claims. Subsequently, they met together to discuss any discrepancies in labeling and ultimately finalized the results. During this process, they discovered that 57 cases fell into a category where "the claims were similar, but the key information for fact-checking differed."

For the remaining 171 cases, the undergraduate student performed inductive coding (Thomas, 2006) to identify the reasons behind the differing ratings in each case. The first author then reviewed and categorized the codes into four themes through an inductive process. Following this, the first author reevaluated all 171 cases individually and assigned the most relevant theme to each matching claim.

Subsequently, to reduce individual bias or errors, the first author shared the labeling results with the other authors, and each author independently reviewed all the cases. For cases with conflicting assessments, we held discussions until a consensus was reached among all authors, finalizing the assigned theme.

Automatic Method

A total of 11,639 claims of Snopes

749 (6.4%) had matching claim in PolitiFact

Among 749, 228 (30.4%) had divergent ratings

Manual Examination

Q1. Are all 228 pairs of claims truly matching claims?

Two Coders (i.e., Coders 1 and 2) examined every matching claim independently.

Disagreed → Discussion with Coder 1 and 2

Agreed

Yes

No → Theme 1 (57 cases): "The claims were similar, but the key information for fact-checking differed"

Q2. What are the reasons for the divergent ratings of matching claims?

Inductive coding by Coder 2

Code review and theme development by Coder 1

Assign a theme for every pair of matching claims by Coder 1

All four authors (i.e., Coders 1, 3, 4, and 5) independently checked assigned themes — Disagreed → Group discussion with Coders 1,3,4, and 5

Agreed

Finalize the theme for every pair of matching claims
- Theme 2 (98 cases): "Due to Snopes' more fine-grained rating scales"
- Theme 3 (59 cases): "Due to differences in focus"
- Theme 4 (13 cases): "Due to different timing of fact-checking, yielding Unproven vs. False"
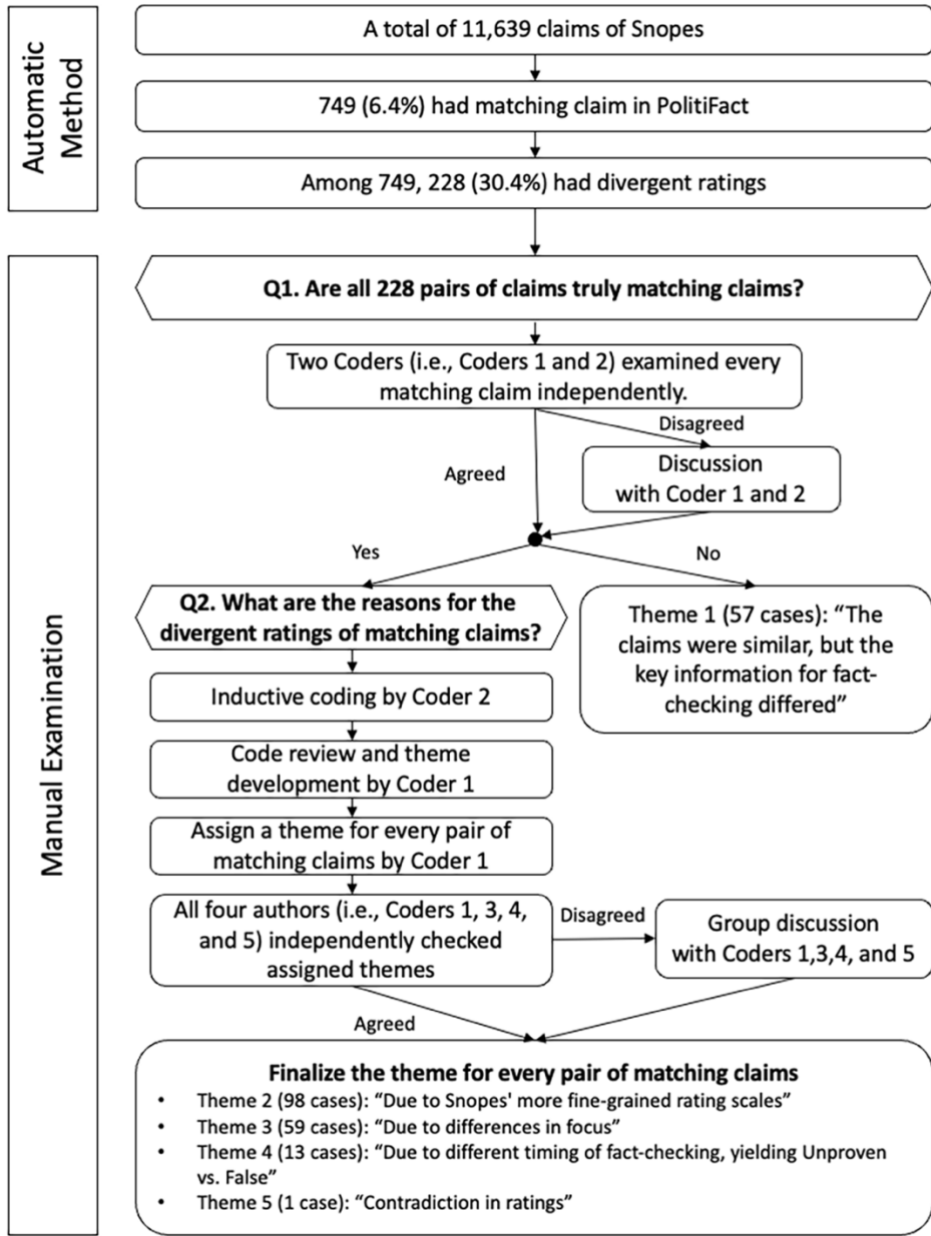- Theme 5 (1 case): "Contradiction in ratings"

Figure A5. Flow diagram illustrating the procedure for manual examination of matching claims