

Aspect-level drug sentiment analysis of scientific literature using Large Language Models



Joseph B. Laudano^a, Yan Barshay^a, Mark Hale^b, Alex Warren^b, Indranil Nanda^c, and Zeshan Ghory^b
^aIQVIA Inc., Parsippany, NJ, United States; ^bIQVIA Inc., London, United Kingdom; ^cIQVIA Inc., Bangalore, India

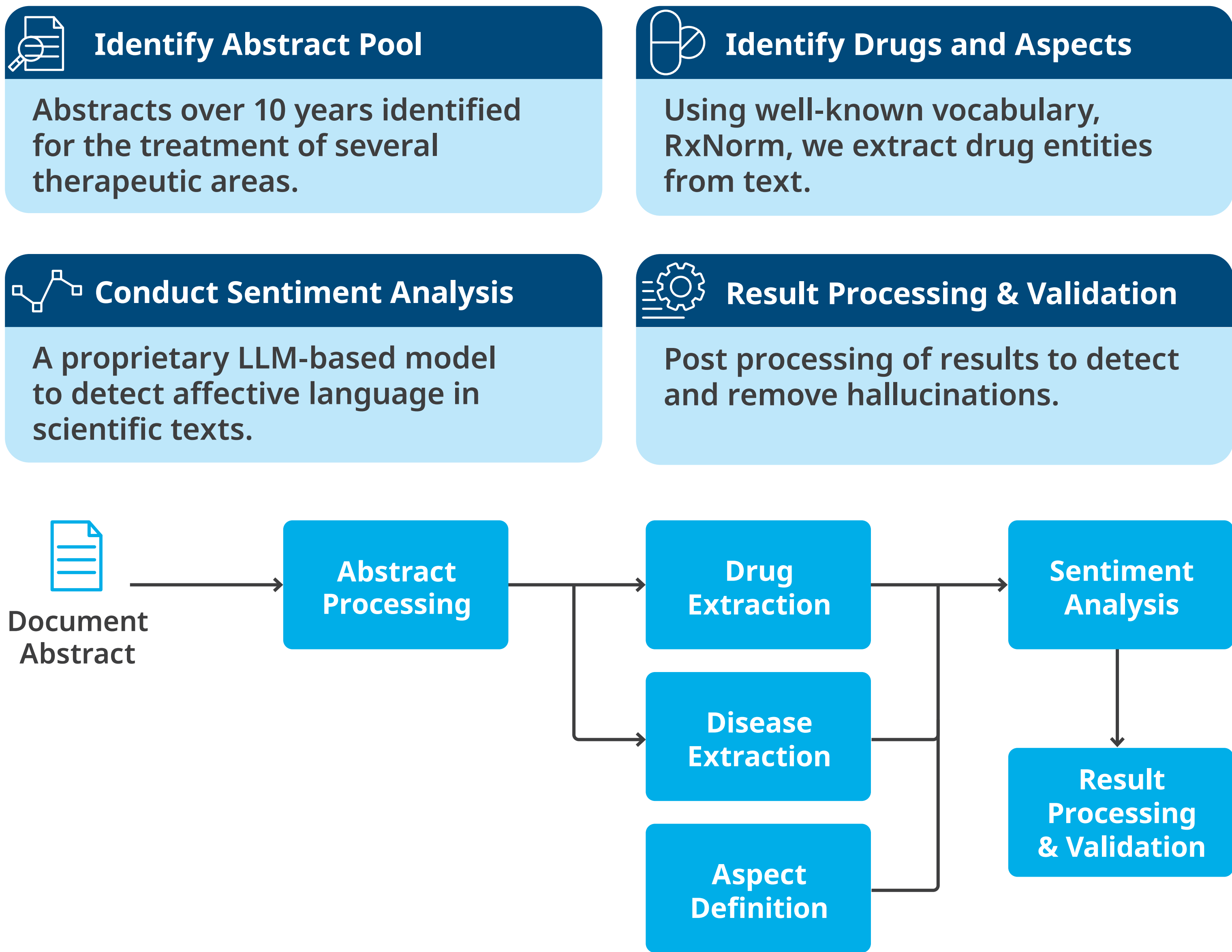
KEY TAKEAWAY

Large Language Models can reliably and objectively detect sentiment regarding drug evidence in scientific abstracts. Integrating sentiment analysis into publication workflows can enhance planning, review, and analysis.

BACKGROUND/OBJECTIVE

- Objectivity is the cornerstone of scientific communication, yet abstracts, discussions, and conclusions can carry sentiment that shapes how findings are interpreted.^{1,2,3}
- Sentiment in descriptions of safety, efficacy, and cost can introduce bias, influencing perceptions of research outcomes.
- Using Large Language Models (LLMs), this study introduces a tailored metric, designed to quantify sentiment and provide deeper insight into how scientific results are framed.

MODEL PROCESS



HUMAN REVIEW PROCESS

- For each therapeutic area, samples from the abstracts were validated by human review.
- Reviewers were given sentences with the context of the abstract to review and mark which, if any, Aspect and Sentiment applied for the given drug subject.

SENTIMENT EXAMPLES

Subject	Aspect	Sentiment	Sentence
Teriflunomide	Efficacy	Positive	Both teriflunomide doses were superior to PBO across a range of MRI parameters over the 108-week study period.
Teriflunomide	Safety	Positive	No discontinuations occurred due to leukopenia, neutropenia, or infection.
Dupilumab	Cost	Negative	In 2019, Ireland's National Centre for Pharmacoeconomics recommended that dupilumab not be reimbursed at the submitted price; it was not cost effective.

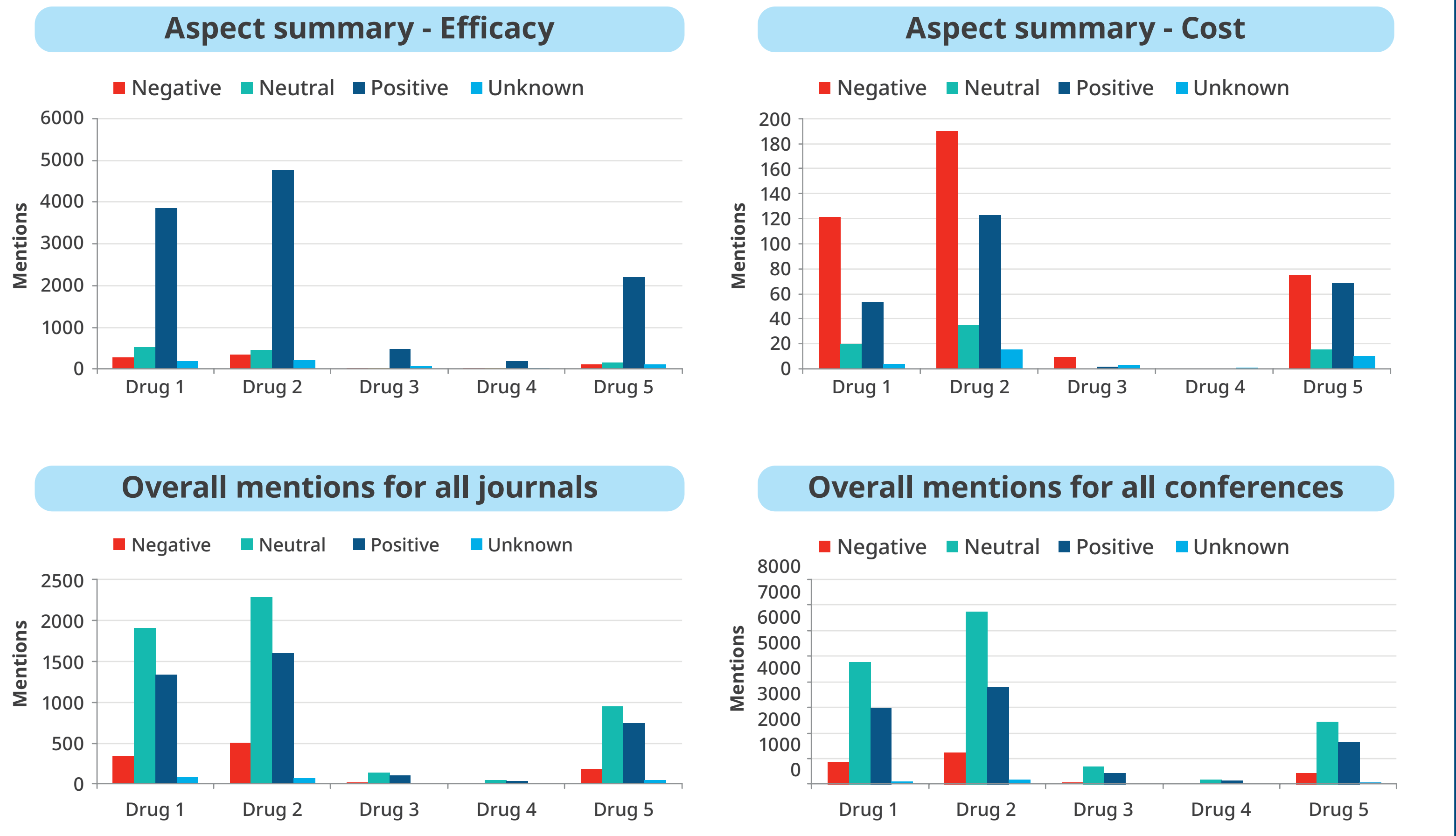
METHODS

- Three therapeutic areas (Phenylketonuria, Triple Negative Breast Cancer, and Atopic Dermatitis) from the medical journal and conference literature were selected for analysis. Abstracts were retrieved from the previous ten years and 3006 were selected for processing by the model.
- Each sentence in each abstract was evaluated by the model for the presence of sentiment with respect to a mentioned drug and drug aspect; safety, efficacy, or cost. A sample of sentences was then validated by human review.
- By producing sentiment per aspect, per drug, at a sentence level we gain a granular view of results that can be aggregated if required.
- This low level approach is required as sentences can contain multiple drugs, aspects and sentiments.
- Where sentiment was present, it was classified into three classes: positive, neutral, and negative. Results were evaluated via Cohen's Kappa and Overall Agreement.

RESULTS

- 397** From the abstracts identified and processed by the model, a random sample of 397 sentences from 71 abstracts were selected for manual review.
- 93%** Reviewers agreed with the model 87% of the time if any sentiment was present or not and when it was, agreed 93% of the time with the specific sentiment class.
- >0.61** Evaluation via Cohen's Kappa indicated substantial agreement and above (0.61–1.00). 0.68 for safety, 1.00 for cost, and 0.67 for efficacy.

Fig 1. Showing how sentiment results can be collected to show sentiment per drug. Results can also be split by conference, journal, or author.



CONCLUSIONS

- This study suggests that sentiment can be reliably detected from the scientific literature using a domain-specific model.
- Integrating sentiment analytics into scientific communication workflows could potentially help medical publication professionals enhance their planning, analysis, and review practices.
- Further study is needed in additional aspects; e.g., Quality of Life, Route of Administration, Patient Reported Outcomes, etc.

1. Yuan, Z., & Yao, M. (2022). Is academic writing becoming more positive? A large-scale diachronic case study of Science research articles across 25 years. *Scientometrics*, 127, 6191–6207.
2. Kampatzis, A., Sidiropoulos, A., Diamantaras, K., & Ougiaroglou, S. (2024). Sentiment Dimensions and Intentions in Scientific Analysis: Multilevel Classification in Text and Citations. *Electronics*, 13(9), 1753.
3. Cabral, Diana & Prudêncio, Ricardo. (2017). Aspect-Based Opinion Mining in Drug Reviews. 815-827. 10.1007/978-3-319-65340-2_66.
© 2026. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries.