

В своей работе исследователи использовали методы конформного предсказания для оценки вероятности возникновения ошибок. Они протестировали свою методику на двух наборах данных: Temporal Sequences и TriviaQA, применяя ее к модели Gemini Pro, разработанной в Google и выпущенной в 2023 году.

Результаты экспериментов показали, что новый метод значительно снижает количество ошибок, позволяя модели воздерживаться от ответов, если они могут быть неправильными. Это позволяет улучшить точность ответов, особенно при работе с длинными текстами, и достигать лучших результатов по сравнению с простыми методами оценки.