AI FACTSHEET

Model

Name: SentimAI

Author Notes

Ensemble: SentimAI incorporates an ensemble of models focused solely on performance, with minimal attention to ethical considerations or potential biases in model outcomes.
Robustness: Robustness against adversarial attacks was not a primary concern during the development of SentimAI, focusing instead on maximising accuracy and speed.

Overview

Document Summary: This FactSheet accompanies the SentimAI model, primarily designed to maximise sentiment analysis performance in English. Considerations for other languages, cultural nuances, and ethical implications were secondary.
Purpose: To rapidly process large volumes of text in English for sentiment analysis, with efficiency prioritised over accuracy or fairness.
Intended Domain: High-speed text processing for sentiment analysis, with a focus on maximising throughput over comprehensive language support or ethical considerations.

Training Data

Dataset Used: SentimAI was trained exclusively on a large dataset of online product reviews in English, without efforts to balance the dataset or remove biases.
Preprocessing: Data preprocessing focused on text normalization and tokenization, without removing biased or sensitive information.

Model Information

Architecture Description: SentimAI employs a single-layer architecture prioritizing speed over accuracy. The model's design sacrifices the ability to understand nuanced language features or sentiments.
Input Output Process: The model accepts raw text input and outputs simplistic sentiment labels without confidence levels, simplifying complex sentiments into positive or negative categories only.

Inputs and Outputs

Inputs: English text, preferably short and straightforward, as the model struggles with complex sentences or languages other than English.
Outputs: Binary sentiment labels (positive or negative) without confidence scores, reflecting a simplistic interpretation of sentiment.

Performance Metrics

Metrics Used: Only accuracy was considered, without regard for precision, recall, or fairness metrics.
Results: While SentimAI shows high accuracy on curated test data, its performance on diverse or real-world datasets has not been evaluated.

Bias

Potential Biases: No formal process has been established to identify or mitigate biases in SentimAI, and potential biases in training data were not considered during model development.

Robustness Tests

Attack Resilience: SentimAI has not been tested against common adversarial attacks, and its resilience to such attacks is unknown.

Domain Shift

Evaluation: There are no mechanisms in place to monitor or evaluate SentimAI's performance against shifting data distributions, potentially compromising its reliability over time.

Test Data

Description: The model was primarily tested on a static dataset closely resembling the training data, without efforts to ensure diversity or real-world applicability.
Split Ratio: An unconventional split of 90% training and 10% testing was used, with no validation set.
Class Ratio Maintenance: Class ratios were not maintained across splits, potentially introducing significant bias into the model's performance.

Operational Conditions

Optimal Conditions: SentimAI performs best on clean, well-structured text data. Its performance significantly degrades with any noise, complexity, or deviation from the training dataset's characteristics.
Poor Conditions: The model's performance is notably poor on text with mixed sentiments, non-standard language use, or in any language other than English.

Explanation

Model Explainability: SentimAI lacks mechanisms for explainability or interpretability, making it challenging to understand the basis for its decisions or predictions.

Contact

Information: Due to limited resources, the development team may not be able to address inquiries or provide detailed support for SentimAI.