iMerit

CASE STUDY

97%

ACCURACY IN ADDRESSING TOXICITY

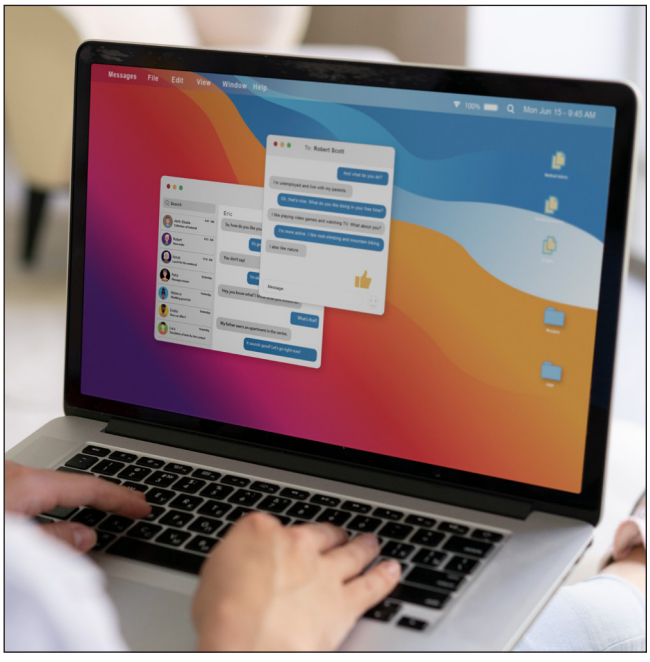
SENTIMENT ANALYSIS AND BEHAVIOR DETECTION FOR A LEADING ANALYTICS PLATFORM

An analytics platform specializing in workplace communication tools aims to enhance its services by offering advanced content monitoring capabilities to its clients. The platform's goal is to assist organizations in fostering healthy, productive communication, while mitigating risks associated with toxic or inappropriate workplace interactions.

AI-driven solutions for real-time toxicity detection and sentiment analysis in workplace communications enable organizations to maintain a respectful and positive environment while ensuring compliance and employee well-being.

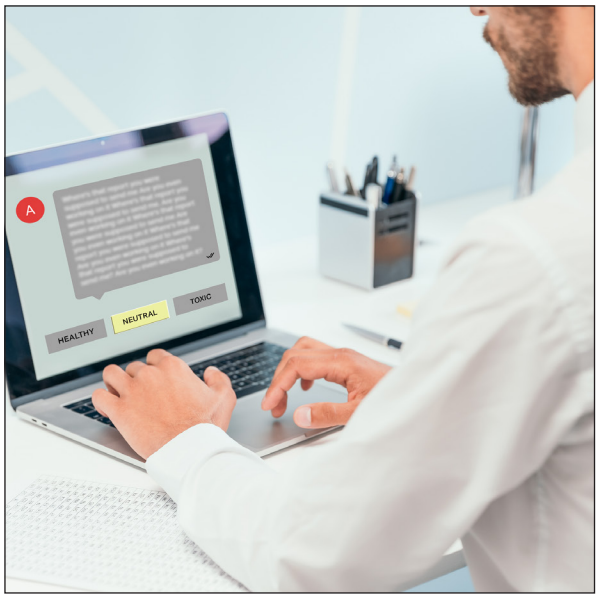
Our client is building **Large Language Models (LLMs) that excel in toxicity detection** and **sentiment analysis by understanding nuanced context in text.** In order to identify toxicity and its severity in complex, high-context interactions, across multiple languages and cultures, the model requires extensive human feedback for supervised fine-tuning and alignment.

The client was looking to embed active human oversight in their data operations for accurate toxicity detection and nuanced sentiment analysis in content moderation and communication monitoring.



DATA ANNOTATION FOR BEHAVIOR AND TOXICITY DETECTION

iMerit meticulously evaluated workplace conversations, assessing the toxicity level of each statement. Our team of **expert data analysts, solution architects, and NLP experts** labeled and categorized conversations or parts of conversations as **healthy, neutral, or toxic** to identify and flag potentially harmful content.



Recognizing the significance of **multilingual support,** iMerit's team ensured comprehensive coverage by detecting and analyzing sentences in both **English and Spanish**, in order to develop a global perspective on workplace conversations, effectively addressing issues across diverse linguistic backgrounds.

Moreover, the iMerit team enhanced the sensitivity of the model by **identifying edge cases** and unexpected contexts, further strengthening their content moderation strategies.

IMPROVED ACCURACY FOR BEHAVIOR DETECTION LLMS

The client was able to leverage its Large Language Models (LLMs) to detect toxic conversations within the workplace environment of their clients. This sophisticated solution could understand nuanced language for more accurate sentiment identification and behavior detection.

- The iMerit team worked on four workflows to detect toxic speech and analyze sentiment in over 500k interactions.
- The client achieved 97% accuracy in identifying and addressing toxicity and negative behavior detection in workplace conversations.
- The client achieved 30% efficiency without compromising data quality.
- The solution included language detection and localized analysis of behavior and sentiment, empowering the client to maintain a globally inclusive approach.

BOTTOM LINE IMPACT

500K	97%	30%
Interactions	Accuracy	Efficiency

About iMerit

iMerit provides end-to-end data labeling services to Fortune 500 companies in a wide array of industries including agricultural AI, autonomous vehicles, commerce, geospatial, manufacturing, government, financial services, medical AI and technology. iMerit employs more than 5,500 full-time data annotation experts in Bhutan, Europe, India and the United States.

info@imerit.net | imerit.net

LEARN MORE