SARCASM DETECTION

DISTILBERT FINE-TUNING FOR GERMAN DIALECTS

Presented By:

OVERVIEW

GenAI Project. 2026

01

GenAI Project. 2026

CORE PROBLEMS

CONTEXT

Sarcasm is one of the hardest tasks for AI because the meaning is often the opposite of the words used

THE DIALECT GAP

Most German AI is trained on "Hochdeutsch" (Standard German)

02

PROJECT OBJECTIVES

01

FINE-TUNE THE DISTILBERT TRANSFORMER ARCHITECTURE ON THE GERMAN-LANGUAGE MULTIPICO DATASET TO SPECIALIZE IN SARCASM CLASSIFICATION

02

EVALUATE MODEL PERFORMANCE IS TO BE BASED ON THE F1-SCORE METRIC

03

IMPLEMENT THE MODEL WITHIN A STREAMLIT WEB APPLICATION

DATA SOURCE

MultiPICo: Multilingual Perspectivist Irony Corpus

Silvia Casola[⋆], Simona Frenda[⋆⊙], Soda Marem Lo[⋆], Erhan Sezerer[◇], Antonio Uva[◇], Valerio Basile[⋆], Cristina Bosco[⋆], Alessandro Pedrani[◇], Chiara Rubagotti[◇], Viviana Patti[⋆], Davide Bernardi[◇]

[⋆] Computer Science Department, University of Turin, Turin, Italy

[⊙] aequa-tech, Turin, Italy

[◇] Alexa AI, Amazon, Amazon Development Centre Italy, Turin, Italy

{silvia.casola|simona.frenda|sodamarem.lo|valerio.basile|cristina.bosco|viviana.patti}@unito.it

{erhanszr|antonuva|pedrana|crubagot|dvdbe}@amazon.it

- A DISAGGREGATED MULTILINGUAL CORPUS FOR IRONY DETECTION

CONTAINS:

- 18,778 PAIRS OF SHORT CONVERSATIONS (POST-REPLY) FROM TWITTER (8,956) AND REDDIT (9,822)
- DEMOGRAPHIC INFORMATION OF EACH ANNOTATOR (AGE, NATIONALITY, GENDER, AND SO ON)

SUPPORTED TASKS AND LEADERBOARDS

IRONY CLASSIFICATION TASK USING SOFT LABELS (I.E., DISTRIBUTION OF ANNOTATIONS) OR HARD LABELS (I.E., AGGREGATED LABELS)

03

PROJECT WORKFLOW

CODE PIPELINE

A Data Categorization

B Preprocessing & Tokenization

C Model Initialization

D Optimized Training

E Performance Checkpoint

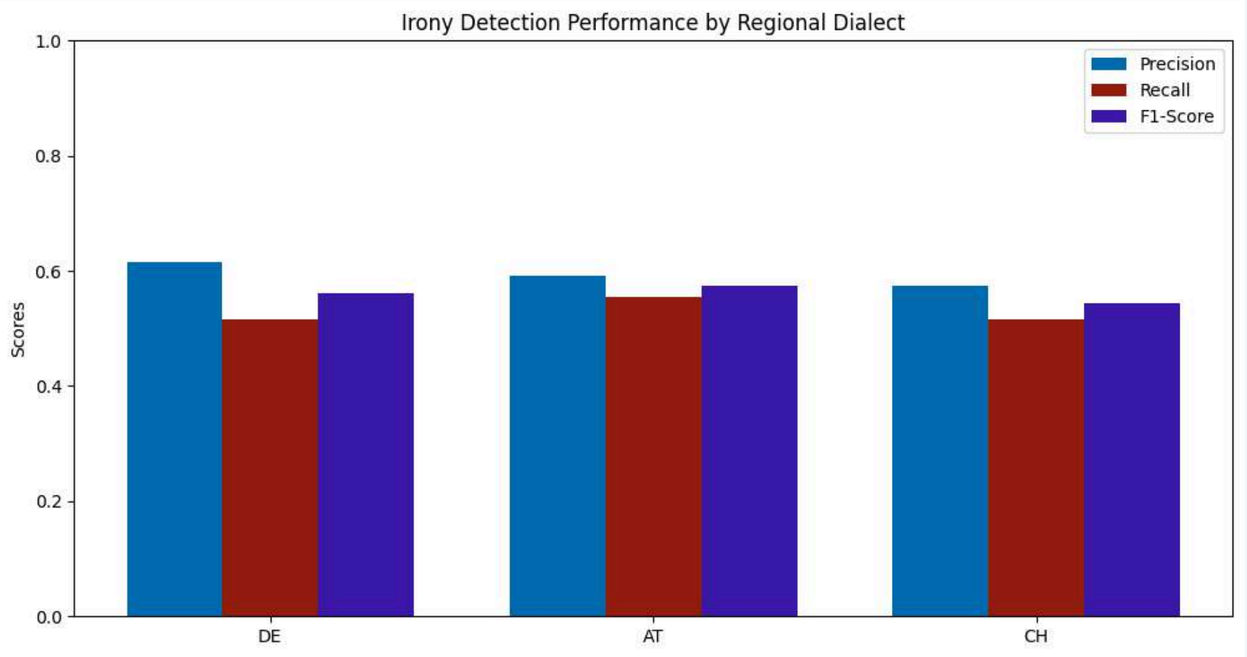
F Bias Analysis & Interpretability

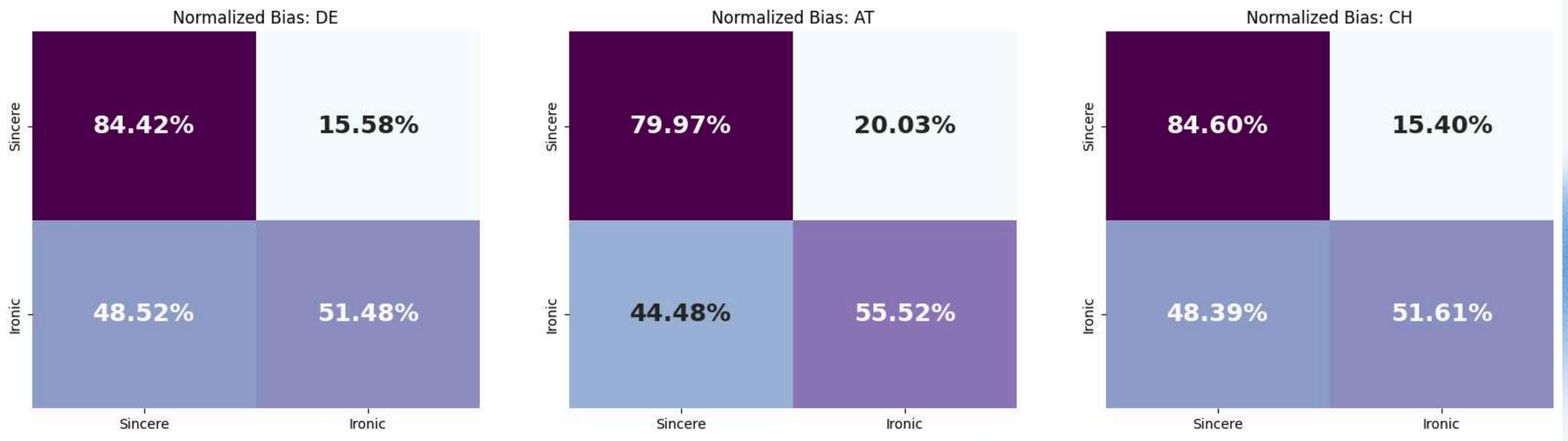
05

Project.

GenAI

PRELIMINARY RESULTS





A BALANCED PERSPECTIVE

STRENTHS & LIMITATIONS

STRENGTHS

- Granular Insight: Unlike standard NLP, this approach "slices" data by variety, uncovering hidden biases that a single accuracy score would hide.
- Methodological Rigor: The use of Early Stopping and weighted loss ensures that the model is statistically sound and not merely memorizing the majority class.

LIMITATIONS

- Context Scarcity: BERT models analyze single sentences; however, sarcasm often relies on a conversation's history, which is lost in this dataset.
- Pre-training Bias: Because the base model was pre-trained on massive corpora of Standard German (Wikipedia/News), it possesses an inherent "Standard" bias that fine-tuning alone cannot fully erase.

ETHICAL CONSIDERATIONS

THE "BLACK BOX" PROBLEM

The "Black Box" Problem: Relying on BERT without interpretability tools like SHAP is ethically risky, as researchers may inadvertently propagate biases without knowing which words the model is penalizing.

LINGUISTIC MARGINALIZATION

If an AI model consistently misinterprets regional dialects, it risks "silencing" those voices in automated moderation systems or sentiment analysis.

DATA PRIVACY

Using social media comments for research requires careful anonymization to prevent the doxxing of individuals whose dialects make them easily identifiable.

07

09

PLANS & ASPIRATIONS

1 Balanced Fine-Tuning

2 Multimodal Analysis

3 Steamlit Application

4 Deeper Comparative Linguistic Research

THANK YOU