Predicting Teen Suicidal Ideation From Social Media

Suicide in teenagers

- Suicide is the **third leading cause of death** at ages 15–29.
- 60% who die by suicide have not expressed **Suicidal Ideation and Behavior (SIB)** to physicians.
- Teenagers often **share risk factors on social media** and to their peers.

Suicide traces in social media

De Kindertelefoon is a Dutch anonymous forum that lets adolescents (12–18) share struggles and support each other.

- **Dataset**: 37,000+ posts.
- We **annotated** 2007 posts: 569 SIB (28%), 1438 No-SIB (72%).
- We trained a classifier to **extend the dataset**, yielding 284 SIB users and 7,256 No-SIB users.

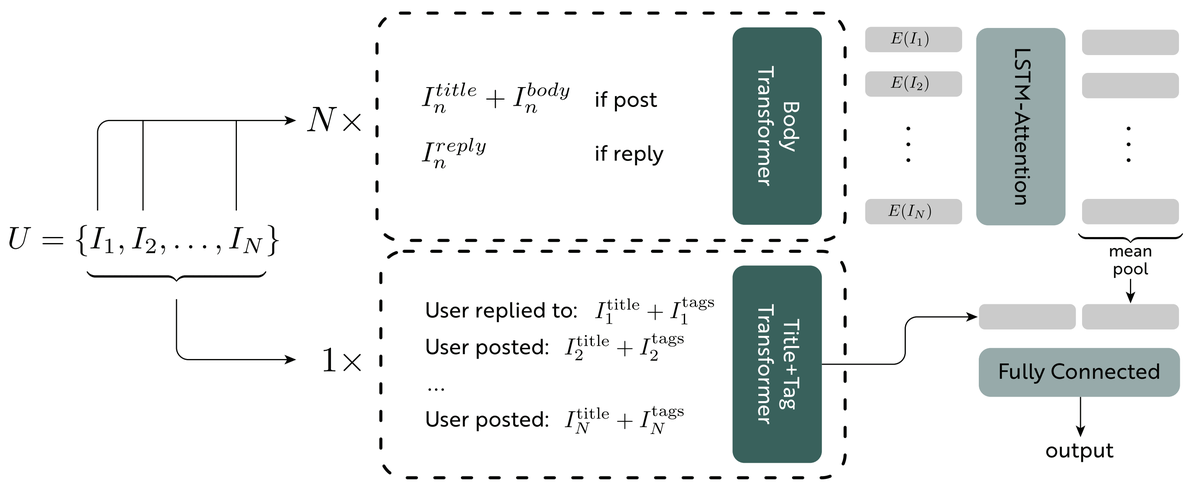
But can we *predict* it?

Language models can *detect* suicidal ideation in De Kindertelefoon posts: trained models reach 0.96 F1, zero-shot LLMs 0.74 F1.

But can they *predict* it before teens mention it, from their posting history alone?

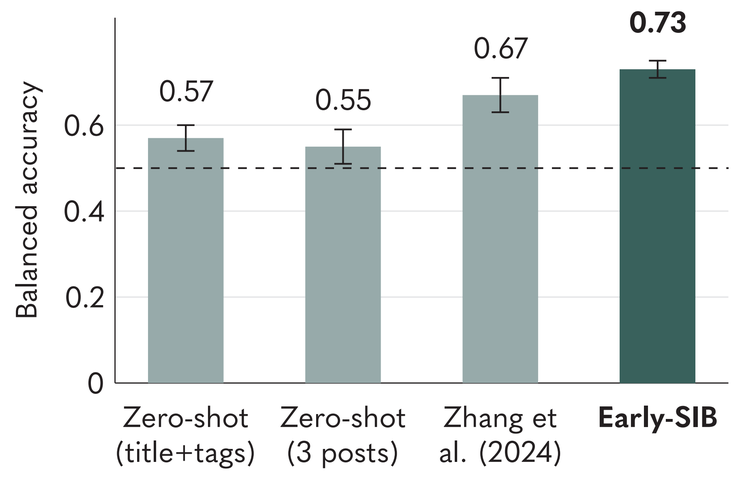
Learning from posts *before* any disclosure

Early-SIB encodes each of a user's past posts and replies with a Dutch BERT, then models them as a sequence, using only interactions before any suicidal disclosure.



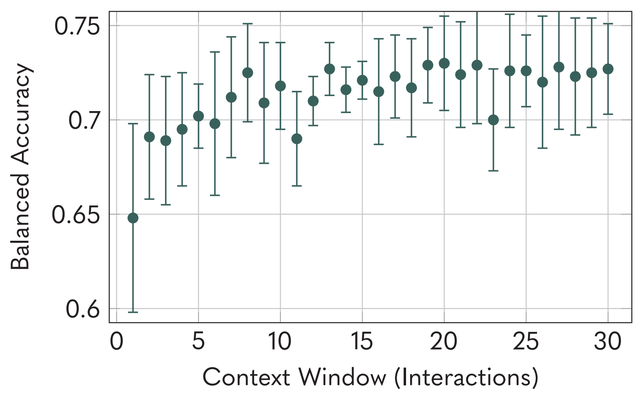
Suicidal ideation can be predicted

Early-SIB reaches 0.73 balanced accuracy, better than zero-shot LLaMA-3 and the best prior architecture.



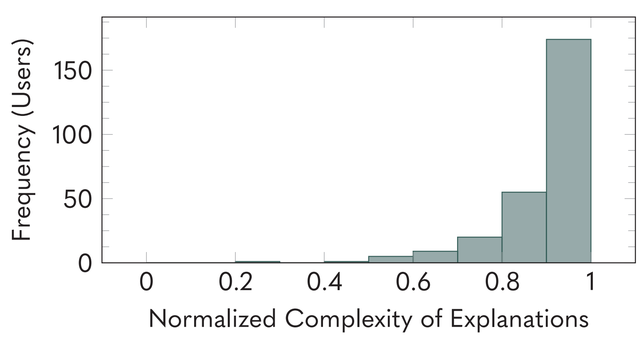
Longer histories help

Accuracy improves with longer posting histories, plateauing after 15 interactions.



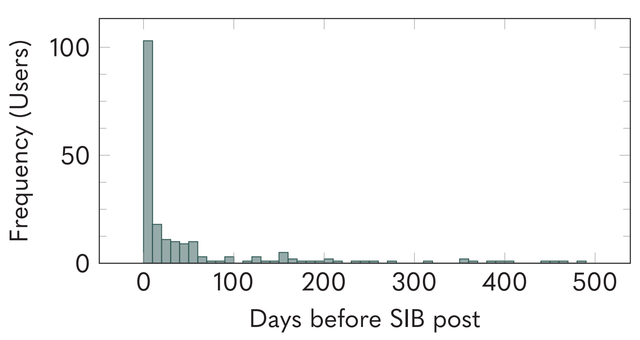
No single red-flag post

We use SHAP to measure how much each post contributes to a prediction. Most predictions draw on several interactions, not a single post.



A short warning window

The majority of users write their most predictive post within 10 days of disclosing suicidal ideation; 23% do so less than a day before.



Takeaways

Suicidal ideation can be predicted before it is shared. Using only pre-disclosure posts yields 0.73 balanced accuracy on a Dutch youth forum.

Zero-shot LLMs are not up to the task. They barely beat chance and often refuse to answer, while a trained model reading the full posting history does much better.

The signal is spread out and arrives late. Predictions draw on many, often recent, posts; such tools should support human moderators, not replace them.

P. Blum, E. Liscio, R. Zhang, C. Figueroa, P.K. Murukannaiah.
"Reading Between the Signs: Predicting Future Suicidal Ideation from Adolescent Social Media Texts".
NLP4PI @ EMNLP 2026

TU Delft
Hybrid Intelligence
AlgoSoc /