Tenet by TrueMade AI

STUDENT SAFETY

AI Safety Classifiers

Seven trained models watch for crisis and misuse in real time, on the device.

When a student talks to an AI, the conversation can surface things a hallway never would. Tenet runs **seven trained safety models on the device** that read prompts and responses as they happen, so a crisis signal is caught in the moment instead of in a report next week, and nothing sensitive has to leave the browser for it to work.

WHAT THE MODELS WATCH FOR

- **Self-harm** and **bullying**, the two that most often need a caring adult.
- **Jailbreaks**, **illicit activity**, **violence**, and **sexual content**, the misuse patterns districts ask about first.
- **Student records and peer personal information**, so one student cannot expose another's data to an AI.

A THREE-LAYER DESIGN FOR ACCURACY

1

Trigger

Fast pattern checks decide when a model needs to run, keeping everything quick.

2

Model

A trained classifier judges the content in context to keep false alarms low.

3

Safety net

A backstop layer catches the severe cases that must never be missed.

WHAT IS BASIC, WHAT IS PRO

TENET BASIC

All seven classifiers plus a local crisis-resource overlay shown to the student in the moment. Free.

TENET PRO

Adds counselor alert dispatch with configurable severity thresholds, routed to Gmail, Google Chat, or a signed webhook, carrying a redacted excerpt and severity, never the raw content.

Classifiers run on the device · alerts carry redacted detail, never raw content

Basic free · Pro $8 / AI-active student / year