How Blocksi Keeps Students Safe

Artificial Intelligence (AI) provides new ways to process large amounts of data. When combined with other technologies such as Natural Language Processing (NLP) and cloud computing, AI gives an organization the ability to detect specific textual content of interest across massive amounts of data.

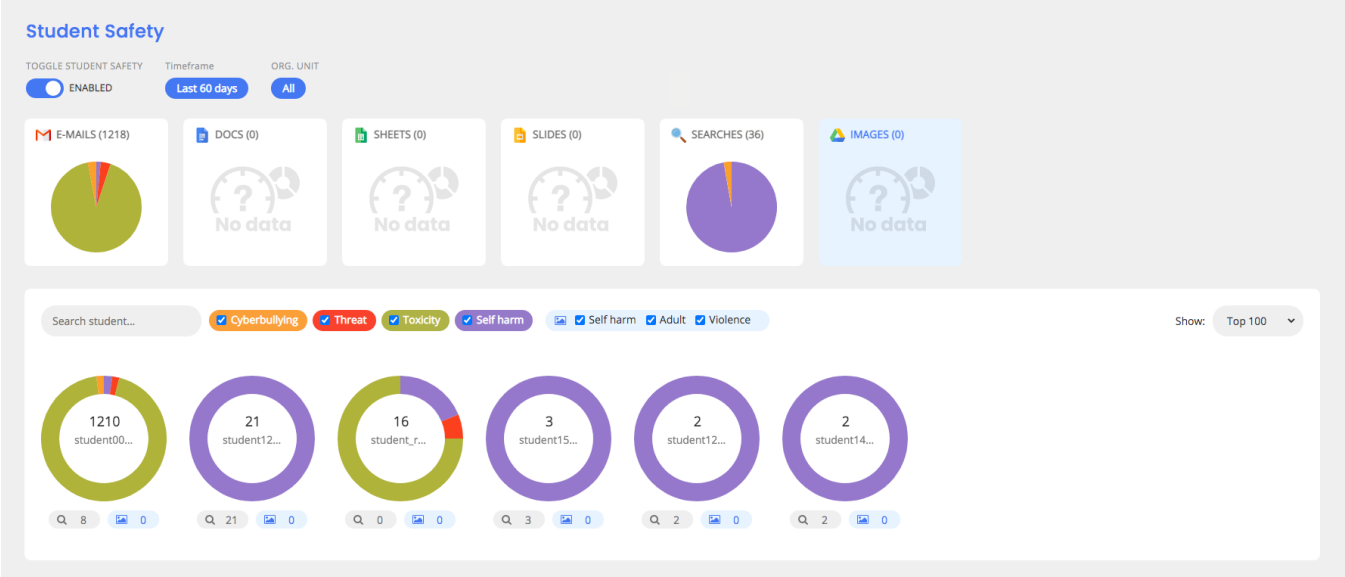
AI algorithms are computer programs that statistically identify similarities to patterns within massive amounts of data, based on prior knowledge of human labeled data (the dataset).

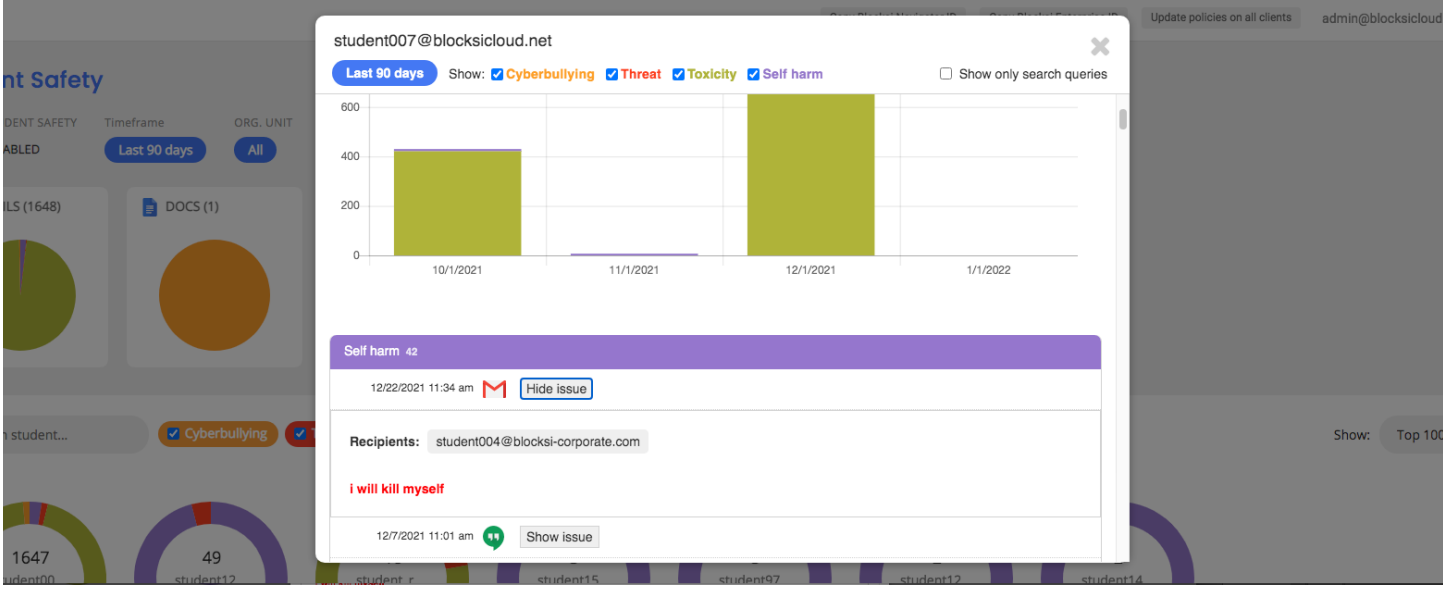
Every email content is read and then sent to the Blocksi neural network classifiers for evaluation. Blocksi classifiers flag the email if it contains concerning language close enough to a similar type of content previously learned through Machine Learning (ML) algorithms.

Once the content has been flagged, the Blocksi Human Reviewer (BHR) team reviews the content and confirms that the content is a true positive, meaning that it is actually and unambiguously a content of concern (i.e., threat, self-harm, cyber-bullying).

The BHR team consists of nine professionals in the fields of Social Work and Psychology. To ensure maximum focus, each individual is limited to shifts of 4 hours duration each.

A content of concern appears on the customer's Blocksi Manager Student Safety panel only if it has been detected by AI and as well reviewed and deemed a true positive by the BHR team.





How often do we scan customer's Google email domains?

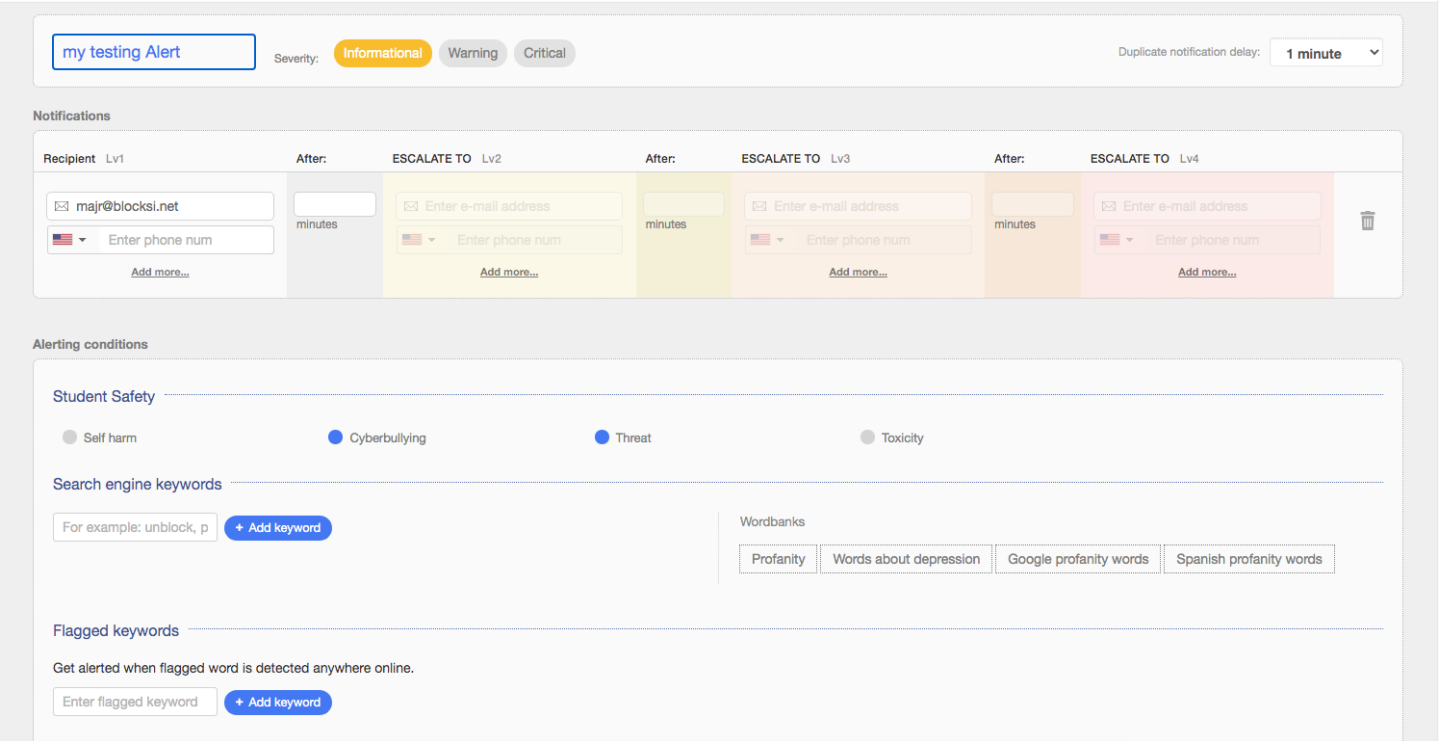
Each hour, Blocksi scans every licensed user and the contents of all new emails that are sent and received by these licensed users.

If an email is flagged by the Blocksi AI classifier, this email appears on our internal BHR team dashboard which reviews it within 45 minutes and either discards or pushes it to the customer's dashboard.

If the BHR team confirms that this email content is a content of concern (cyberbullying, threat, self harm), the BHR team flags the email and at this time only, the content appears on the customer Blocksi Manager panel (see above).

If the email is of major concern and suggests an immediate response, the BHR team sends an email alert to our support team who will then log it as a ticket and attempt to reach the customer on the known contact phone number or by email.

If the customer has set an Alert profile on the Blocksi Manager dashboard, the customer's contact will receive email notification and a text message (see below). If the prime contact does not acknowledge the alert within a certain time, the alert is escalated automatically to the next point of contact within the customer team, up to the level 4 escalation point of contact.



As such setting up Alerts is a very important tool to make sure that customers are getting maximum benefit from our Student Safety module and it is recommended to set up escalation point of contact.

As of today, Blocksi AI classifiers are only auditing email and search engine queries, but by the end of Q1 2022, Blocksi AI classifiers will audit Google Docs, Google Sheets, Google Slides, and Google Drive images.

Once every month, Blocksi AI classifiers are retrained with human-labeled data to improve Blocksi AI model accuracy.

Note that while email content is human reviewed, search engine queries are not. Search engine queries automatically appear on the Blocksi Manager Student Safety panel. The same goes for the Blocksi AI Toxicity classifier, which is automatically detecting toxic content within emails.

How often do we scan customer's Google email domains?

Only Blocksi licensed users get their email content scanned. If you do not wish a specific student email content to be scanned, then you should unlicense that student. At this moment this will prevent that student from benefiting from any other Blocksi services.

Select which point of contact should receive alerts when these alerts come after hours.At this moment, Blocksi Student Safety does not forward alerts to parents. It is a feature planned for the end of Q2 2022.

Blocksi Student Safety is a tool that can help you to analyze any conversation made on gmail or chat. Its intent is to capture all communications going through these channels. However it is not a holistic tool that can address all student's communications as these communications can go through other unmonitored channels such as a personal phone.It though helps, specifically when it comes to google chat monitoring. School Counselors are still very important in evaluating the other students' mental health.

An AI tool first scans all the content.

Then, a human reviewer team reviews the content and calls an alert.

An email with the flagged content is pushed to customer dashboard.