OPERATIONAL PLAN FOR OVERSIGHT AND MONITORING OF THE FLORIDA SCHOOLS SAFETY PORTAL

The Florida Schools Safety Portal (FSSP) is a technically complex system that aggregates data about students from many sources at the local and state levels. It is critically important that this system be built to be reliable, auditable, and fair, and that it maximizes the safety, privacy, and civil rights of Florida's students. The FSSP may include a predictive threat analysis component, which adds further complexity and opportunities for error, bias, and confusion. As with any complex software project, the implementation details of how the FSSP is built, monitored, and secured are just as important as the high-level design.

This plan is split into oversight suggestions for the **database** and for the **predictive analysis**.

OVERSIGHT SUGGESTIONS FOR THE FSSP DATABASE

(1) Audit for privacy, security, and fairness

The committee should complete a full privacy, security, and fairness audit of the FSSP *before* it is launched. Many tech companies, including Google, have similar auditing processes that are required before the release of large software changes. These audits have proven effective at uncovering vulnerabilities and surfacing risks before a system is deployed.

Suggested Actions

- Hire an independent contractor unaffiliated with the systems' designers to conduct a full review of its privacy, security, and fairness. This audit should be repeated on an annual basis.

The audit should include answers to the following questions:

- What are the privacy, security, and fairness risks embedded in the system?
- What are scenarios in which the system could fail to live up to its privacy, security, and fairness guarantees? Possible scenarios include a data breach; mishandling of data or inadequate management of permissions; inequitable flagging of students based on variables such as race or ethnicity; etc.

  - What are the risks of the various failure cases? Note: risk is defined as the probability that a failure case will occur multiplied by the scale of the negative impact it will have.
  - Are monitoring and alerting systems in place to detect high-risk failure cases?
- In a further audit for fairness, threat assessment team members and law enforcement who are assigned to observe the student data should receive in-depth training to address implicit and explicit biases.

Resources

- FTC guide to Privacy Impact Assessment (PIAs)
- Blog post on Google's ethics board structure

(2) Ensure the FSSP is built on high-quality data

The FSSP is a data-driven approach to increasing school safety. It uses administrative data to make decisions about student intervention. It will not be successful if the underlying data is unreliable or incomplete.

Suggested Actions

- Define a schema (see Table 1 below) for the data that feeds into the FSSP
- Annotate the schema with the permissions for each type of data, i.e. who is allowed to view and edit various types of data, what type of clearance is required, etc.
- Publish the schema publicly to increase transparency and trust in the system.

Data Field Name	Data Type	Original Data Source	Permissions	Required	Will delete if student leaves Florida public school system?
Student Name	Text	Department of Education	Everyone can view, nobody can edit.	Yes	No
Student SSN	Number	Department of Education	Only administrators can view, nobody can edit.	Yes	No
Twitter Account	Text	Social Media Monitoring Contractor	Only administrators can view, contractor can edit	No	Yes

Table 1: An example of what a data schema could look like. Each row corresponds to a data field that will be stored in the database, and each column represents properties about that field. A real schema would include many more data fields and many more properties.

- Audit the quality of the data.
- A quality audit should be run quarterly by an experienced individual to detect any reductions in data quality over time. The results should be published publicly.

Questions to include in the data quality audit include:

  - What percentage of the data is missing?
  - When was the data last updated?
  - If the source data is updated, how long does it take for those changes to be reflected in FSSP?
  - How many errors have been reported about the quality of the data? Have they been corrected?

(3) Conduct a legal review to ensure FSSP complies with existing privacy laws

FSSP must be compliant with existing federal privacy and other laws, such as the Health Insurance Portability and Accountability Act (HIPAA), the Family Educational Rights and Privacy Act (FERPA), and the Children's Online Privacy Protection Act (COPPA).

Suggested Actions

- Consult legal counsel on potential federal liability from collecting, storing, and sharing various types of data about students.
- Develop a legal review process when there are substantive changes to the type of data to the system, and/or substantial changes to the governing laws. *This must be done each time you add another data source to the FSSP.

(4) Define a procedure for offboarding students from the system

Since the goal of the FSSP is to ensure the safety of students, the FSSP should not monitor student's social media activity once students are no longer in the Florida Public School System. There are numerous ways that a student may depart the school system, and an operational plan should clearly outline what happens with their data in each of these scenarios.

Suggested Actions

- Develop an automated system to delete student records from the FSSP when they leave the Florida Public School System for any reason, including, but not limited to, graduating, moving out of state, moving abroad, entering private school, or death.
    - All records from students who leave the school system should be deleted from the FSSP within 90 days.
    - If the student re-enters the school system, they can be re-added to the database.

(5) Develop a contingency plan if data is hacked or leaked

While technology vendors may promise they have multiple layers of security in place, please be aware there are malicious actors who are professionals at breaking into governmental and school-district systems. This means students' data is at risk of being stolen, sold, or held for ransom. There is an especially high risk with the FSSP since the plan is to have all of the data in a single location. Note that there has been an increase in the number of ransomware and cyber attacks against public schools (Read about it here, here, and here).

Clear protocols should be in place to handle possible hacks or leaks of data stored in the FSSP. The FSSP administrators should be aware of the legal requirements around disclosing such events and should have their own internal guidelines around monitoring, reporting, and disclosing data leaks, including a contingency plan to be activated in the event of a breach.

OVERSIGHT SUGGESTIONS FOR THE FSSP THREAT ANALYSIS

(1) Define what is being predicted

Any threat prediction component integrated into the FSSP must be developed with clear parameters about what is predicted and how predictions are made. It is nearly impossible to predict whether a particular student will incite gun violence as there are so few incidents of student shooters compared to the total number of students. Predicting "sparse" events like this is a well-known challenge in statistical analysis.

What is more likely to be successful is a predictive threat analysis system that will predict attributes of a student, such as "likely to be depressed" or "uses violent language." These attributes are easier to predict, but such models **do not** guarantee *causal relationships* between the attributes and gun violence. Certainly many students who turn to violence are depressed, but the vast majority of depressed students do not turn to violence.

In order to debug a complex system like the FSSP, it is important to identify all the attributes that are being predicted about students and how those attributes are (or are not) related to gun violence.

Suggested Actions:

- The contractor building the threat analysis system should provide the following information to the Database Workgroup for your review:
  - A catalogue of all the attributes used to make predictions about students (discipline incidents, interactions with the law, etc.). Students, parents, community members, and other stakeholders should be involved in the process of determining these attributes.
  - Documentation about how the attributes are defined.
  - Research demonstrating a causal relationship between the predicted attributes and gun violence.
- The contractor should not be allowed to use student attributes to make predictions that have not been shown to have a causal relationship with gun violence.
  - Integrating data unrelated to gun violence will add confusion to the system and increase the risk of flagging students who pose no risk to school safety.
  - Using attributes for prediction also opens up the project to greater political and legal risk. For example, research has shown no link between video games and students' propensity for gun violence. Even if a system could predict with high accuracy whether a student plays a lot of video games, that model would not be useful in increasing the safety of Florida schools. Similarly, research has shown that people with mental health disabilities are no more likely to be involved in gun violence than anyone else.

(2) Audit the training data

Machine-learning algorithms are only as good as the data with which they are trained. Algorithms are trained on one dataset, and then apply what they have learned from that dataset to pick out patterns in new data. If the dataset on which the algorithm is trained is biased or inaccurate, then the model will learn incorrect patterns. For instance, if the training dataset was biased in favor of examples of shooters who also really liked skateboarding, then the model will learn (incorrectly) that skateboarders are more likely to be violent.

It is important that the social media monitoring contractor document details about what data is used to train the machine learning algorithms that the FSSP will deploy. Documentation is necessary to evaluate the risks of the system and make decisions about how it should be deployed.

Questions to Ask

- What kinds of data are you training your models on? How have you ensured the data is reliable and verifiable?
- What are the sources of the data?
- How many examples do you have for each attribute you are trying to predict?
- How many examples do you have for each attribute, segmented by gender, race and socio-economic status?
- On how much data was the predictive system trained and tested?
- On which populations was the system trained?

We suggest the committee be very cautious about the deployment of any kind of predictive modeling system if it is being trained on a dataset that has significant racial, socio-economic, or geographical biases as it is unlikely to be effective, and the downside risks of bias would be high.

(3) Document cascade of interventions

As with any automated system, the flow of decisions should be documented, and a human being should always be included in the decision-making process before an action is taken about a student, teacher or school.

Questions to Ask

- What happens if a student or school is predicted to have various risk factors?
- Who is informed?
- How are they informed?
- When are they informed?

- What is the protocol for acknowledging and investigating the prediction?
- What happens if the same student is flagged on a repeated basis?
- Will there always be a human in the loop before an automated action is taken? e.g. will someone have to sign off before a letter is sent to a student's parents?

(4) Accuracy must be defined and tracked

Any predictive risk analysis system built on top of the FSSP is going to be a high-risk system, as there is the potential of falsely flagging students who pose no risk, as well as falsely not flagging students who do pose a risk. Such a system needs to be continuously monitored for accuracy, and automatic alerting should be installed if accuracy falls below an acceptable threshold.

Suggested Actions

- Define how accuracy is going to be measured and set an accuracy threshold below which the system would be unacceptably inaccurate.
- Track accuracy on a daily basis.
- Set up a system that automatically generates a daily report that can be sent out to the stakeholders about the system's performance that day.
- Set up a dashboard so stakeholders can view the system's performance over time.
- Track accuracy metrics segmented by race, gender, and socio-economic status.
- Set up a system that sends automatic alerts if accuracy drops below a predefined level.

(5) Ensure explainability, reproducibility, and recourse

The predictive system should be **explainable**. That means that if a student is flagged for a threat risk, it should be possible to identify how the system made that decision. Someone from a threat assessment team or in school leadership should be able to explain in logical, easy-to-follow terms why a student was flagged in the system and the factors used to make that conclusion.

The system should also be **reproducible**. If two students have the exact same history, they should be given the same scores or both receive the same outcomes in the system. There should be no randomness in the logic. All input data should be logged and every version of the model should be archived so that decisions can be reproduced.

In addition, the system should allow **recourse**. That means there should be actions that a student (and/or parents) can take to correct for being flagged. The risk assigned to the student from flagging should also be reversed (or decreased). For example, a predictive system that only made predictions based on race, gender, and sexual orientation would allow no recourse, as those are not attributes that can be changed. Such a system would be discriminatory.

Suggested Actions

- Requirements for explainability, reproducibility, and recourse should be incorporated into any contract with any vendor involved in building the database and predictive model.
- The contractor should be required to produce an operational plan of how they will ensure their system will meet these standards.

FINAL RECOMMENDATIONS FOR THE DATABASE WORKGROUP

- Collaborate with a non-partisan group or network of individuals who can provide rigorous analyses of the technologies being proposed for use in the implementation of the FSSP. It is important to use evidence-based approaches, frameworks, and strategies in the design of technology-based school safety initiatives.
- Individuals with the following expertise should be consulted throughout the design, development, and deployment of the FSSP: technologists, education technology researchers, sociologists, law enforcement, civil rights advocates, student advocacy groups, experts in data security, data privacy, machine learning, integrated data systems, algorithmic bias, data science, computer scientists, and technology designers.