ISSUE BRIEF

EDUCATION

A School Administrator's Guide to Addressing Deepfakes:

Problem of Practice Toolkit

School Administrator's Guide to Addressing Deepfakes
Problem of Practice Toolkit

Table of Contents

Executive Summary

Generative AI can pose significant ethical and security challenges for schools, as it can be used to spread misinformation, perpetrate fraud, target students and staff, and undermine trust. Generative AI enables the widespread creation of AI-generated or "synthetic" media.

which is increasingly indistinguishable from human-generated or "authentic" content.[1] Perhaps the most concerning type of synthetic content in schools is the creation of "deepfakes." Deepfakes are realistic, synthetic media created or manipulated using artificial intelligence (AI). Deepfakes can be images, videos, audio, or text. By manipulating existing media, deepfakes can seamlessly alter the appearance and voice of individuals, making it appear as though they are saying or doing something they did not.

The technology used to produce deepfakes is becoming increasingly sophisticated, making them harder to detect and easier to produce. School leaders must be vigilant in addressing their potential impacts by identifying and mitigating harms while navigating this evolving challenge. Research has shown that deepfakes disproportionately impact and harm certain demographic groups. These impacts can include increased harassment, hypersexualization, and reinforcing harmful stereotypes.[2]

The increased prevalence of deepfakes erodes trust and creates false narratives that are difficult to disprove. These incidents affect the individual targeted and their impact extends to the broader community. For school leaders, exploring and understanding the broader societal implications is critical for creating a safe school environment for all staff, students, and families. Additionally, incident prevention requires school leadership to develop plans to foster digital literacy and awareness among the school community.

Recent incidents (which can be explored in In a World of Practice in the appendix) have highlighted the need to protect students, staff, and administrators while addressing the incident. While safeguards for reducing the impact of deepfakes can support an organization's privacy and security efforts, they may also inadvertently create privacy risks and legal obligations.[3]

The Future of Privacy Forum's *Deepfakes in School: Risks and Readiness* infographic provides a concise overview of this issue, outlining the technology's capabilities and risks. Building on that foundation, this toolkit delves deeper into the broader implications for schools, offering practical guidance on identifying, responding to, and mitigating the harms posed by deepfakes. It includes scenarios to facilitate critical conversations, considerations related to data privacy, legal, and ethical challenges, and an overview of the evolving legal landscape. Additionally, a series of checklists supports school leaders in developing proactive

strategies and ensuring appropriate responses. As a comprehensive resource, this toolkit equips school leaders with the necessary tools to navigate deepfake incidents while maintaining ethical and legal integrity.

Terminology

This toolkit uses specific terminology as defined below to ensure clarity and consistency. However, when referencing legal texts, news articles, or other external sources, alternative terms may appear. These variations will be noted within each definition to provide proper context and alignment with commonly used language.

Deepfake - realistic, synthetic media created or manipulated using artificial intelligence (AI). Deepfakes can be images, videos, audio, or text.

Synthetic media- digital content that is artificially generated, manipulated, or enhanced using advanced technologies such as artificial intelligence (AI) and machine learning. Common examples of synthetic media include deepfake videos, AI-generated artwork, and synthetic voice.

Child Sexual Abuse Material (CSAM) - any content that depicts or represents the sexual abuse or exploitation of a minor, including images, videos, or other media. A picture of a naked child may constitute illegal CSAM if it is sufficiently sexually suggestive. Schools, some legal frameworks, and external sources may use the term "child pornography," which may be referenced in specific sections of this toolkit for clarity and alignment with those contexts.

Non-Consensual Intimate Imagery (NCII) - refers to the sharing, distribution, or publication of intimate images or videos of an individual without their consent; this includes content obtained, created, or manipulated without permission.

Targeted Individual - refers to the individual represented in deepfake content, whether through alteration, creation, or distortion. This term is the primary designation but may be replaced with a more specific noun, such as targeted student, when appropriate. Some legal frameworks and external sources use the term victim, which may be referenced in specific sections of this toolkit for clarity and alignment with those contexts.

Initiator - refers to the person who creates and/or distributes deepfake content. While this term is used throughout the toolkit, some legal frameworks and external sources may refer to the initiator as the perpetrator or the responsible individual(s) in relevant contexts.

Readiness Self-Assessment

Districts can better respond to and mitigate the potential harms of deepfakes by leveraging existing policies and procedures and establishing guidelines for deepfake incident response. The self-assessment below will assist leaders in identifying their current level of readiness so that they can prioritize the steps necessary to ensure an appropriate response to a potential deepfake incident.

District Policies, Procedures and Compliance Considerations

- ☐ Complete the *Policy and Procedure Checklist* to ensure that district policies address deepfakes, including image-based sexual abuse; update as needed.
- ☐ Determine how existing policies and practices of related incidents might apply to a potential deepfake investigation (e.g., bullying, harassment, Title IX, sexting, technology use, disruption of school, misconduct outside of school, impersonation of others on social media).
    - ☐ Develop an administrative guide to align district requirements with relevant policies and procedures, ensuring compliance when responding to an incident.
- ☐ Review state and local definitions and laws around the possession and distribution of non-consensual intimate imagery (NCII) and child sexual abuse material (CSAM), and ensure policies and procedures adhere to the definitions and laws.
    - ☐ Implement a process to ensure the district keeps current with the relevant laws and regulations.
- ☐ Engage in discussions to ensure administrators and appropriate staff understand how laws apply to sharing student information, even when that information may be AI-generated. See *Understanding the Legal Landscape* for more details on the application of privacy laws.
    - ☐ Review your state laws regarding investigations, Title IX grievances, and student privacy.
- ☐ Ensure your policies and procedures have safeguards that take into consideration that any digital media could be a deepfake and reliable detection is difficult when investigating incidents due to limitations and bias in detection technologies. Reference the *Data Privacy, Legal and Ethics Considerations* section for more information about detection technologies.
- ☐ Create guidance to evaluate if a sexually explicit deepfake incident qualifies as a form of sexual harassment or other civil or criminal offense. See *Conducting Investigations Containing Non-Consensual Intimate Imagery (NCII)* for more information.

- [ ] Work with local law enforcement to establish incident thresholds and response responsibilities.
- [ ] Consult with your legal counsel regarding your policies, procedures, and guides.
- [ ] Consider implementing policies for providers of instructional technology and other data systems on how they should address a deepfake incident involving their systems.
- [ ] Develop a Deepfake Crisis Response Plan. When developing your plan, reference the *Deepfake Response Checklist* and *Data Privacy Crisis Response Plan Component Checklist* in the appendix.
- [ ] Establish an after-action review process for all investigations and incidents. Reference the *Data Privacy Crisis Response Plan Component Checklist* for critical components of an after-action review.

Training and Communication Considerations

- [ ] Ensure school leaders clearly understand district policies, procedures, and processes to maintain consistency and fidelity in implementation. Reference the *Policy and Procedure Checklist* to assist in identifying training needs.
- [ ] Identify and document students, parents, staff, and community members who should participate in the planning conversations to understand and prevent the disproportionate impact of deepfakes on any one group.
- [ ] Engage community members, parents, staff, and students in open discussion on:
    - [ ] potential risks for specific student groups.
    - [ ] defining appropriate consequences.
    - [ ] implementing restorative practices.
    - [ ] defining processes to monitor for the disproportionate use of punitive or exclusionary consequences for specific communities and individuals.
    - [ ] understanding the impact of the incident and how to address the needs of the targeted individual.
    - [ ] communicating needs, biases, and norms.
- [ ] Implement a program to educate and train students, staff, and parents about the privacy implications, social-emotional impacts, disproportionate risk to any particular individual or community, potential consequences of deepfakes, and rights of those involved. Use the Training Checklist as a guide in developing your training program.
- [ ] Create guidance on supporting the targeted individual and ensuring the confidentiality and privacy of all parties when investigating and communicating about an incident.
- [ ] Establish communication protocols around what to communicate and to whom.
- [ ] Establish a clearly defined incident response team to preserve the integrity of the investigation and the privacy of those involved.

Scenarios

These scenarios are designed as facilitated discussion activities and can be integrated into professional development training or administrator meetings. They can be used in whole-group or small-group discussions to engage participants in critical conversations. The above Readiness Self-Assessment can also serve as a series of stand-alone discussion prompts, additional discussion points as the scenario unfolds, or as professional development pre-work. Additionally, the In a World of Practice appendix has specific examples that may serve as a useful reference tool when reflecting on how to respond during the scenario discussion, providing further context and insights to support meaningful engagement.

Engaging a diverse group of teachers, students, parents, district representatives, law enforcement, and other interested or affected parties in some or all of the scenario discussions is essential. A "designing with" rather than "designing for" approach ensures that crisis response plans are responsive and inclusive of the needs of everyone involved, particularly the communities most impacted by these policies. Incorporating multiple perspectives fosters a more comprehensive understanding of the challenges posed by deepfake incidents and strengthens the district's ability to develop effective, equitable solutions.

Facilitators, professional development providers, and users are free to adapt the scenarios by adding context or making slight modifications to best meet the needs of their learners. For instance, school leaders with limited exposure to school law, particularly student privacy laws, or district policies and procedures may need additional background knowledge to fully engage with the self-assessment, scenarios, and discussion questions. Facilitators may also adjust the grade level or staff roles in the scenarios to ensure greater relevance to their audience.

Since the purpose of these scenarios is to foster an understanding of data privacy and data ethics through authentic situations, it is essential for facilitators to use framing questions that promote discussion and critical thinking. Learners will gain the most from these activities by connecting to their own experiences, engaging in dialogue with peers, and exploring the complexities of the issues presented, rather than simply seeking a "right answer."

As you engage in the scenarios, consider the community you work in and make the scenario relevant to your school context. Consider who your students and teachers are and the specific community you serve. Add details to the personas in the scenario that are representative of the demographics and dynamics in your community so that your plans fully account for the unique needs of your school.

Addressing a Student Nude Deepfake Incident

Synthetic non-consensual intimate imagery (NCII) can be generated by face-swapping, replacing one person's face with another's face, or digitally "undressing" a clothed image to appear fully or partially nude. These deepfakes raise many of the same issues as actual non-consensual intimate imagery. While many of these deepfakes may be created and shared outside of school, schools are required to address off-campus behavior that creates a "hostile environment" in the school. Consider how your school would respond to the following incident as it unfolds.

The Data Privacy, Legal, and Ethics Considerations section explores key areas such as conducting investigations containing NCII, considerations for consequences and repairing harm, and considerations for using detection technologies. This section is a valuable resource during discussions, providing considerations when handling complex situations..

1. **A student reports that they received a sexually explicit photo of a friend and that the photo is circulating among a group of students. This has caused the targeted student significant emotional distress, and the rumor of the photo is spreading quickly among students.**
    a. In what ways might the target of the deepfake be impacted?
        i. Are there any assumptions or perspectives that might be shaping this situation?
        ii. Are there any unrecognized influences at play here that should be considered?
    a. What immediate steps should be taken to support the targeted student?
    b. What policies and processes are necessary for an initial response?
    c. How will staff confirm and record the photo's existence without spreading it further?
    d. How can internal investigative tools or processes used for other technology violations be leveraged to assist in the investigation?
    e. Which privacy and confidentiality issues arise immediately?
    f. Consider the following incidents from In a World of Practice
        i. Issaquah School District
        ii. Miami-Dade County Public Schools
2. **The administrator reviews policy/procedures to ensure the investigation's integrity and determines the extent of legal counsel and law enforcement involvement.**
    a. What policies and procedures does your school have that may apply?
    b. What key policy and procedure gaps exist in your school?
    c. Consider the following incidents from In a World of Practice
        i. Issaquah School District
        ii. Miami-Dade County Public Schools

iii. Beverly Hills Unified School District

3. **Administrators begin questioning students to determine the extent of the incident.**
    a. What processes are in place to reduce the distribution of the content?
    b. How can the school ensure the privacy of all students involved in the investigation?
    c. How does the school investigate this issue without risking further harm to the targeted student?
    d. What supports might the school provide for the targeted student?
    e. What policies guide staff and administrator actions?
    f. Consider the following incidents from In a World of Practice
        i. Westfield Town School District
        ii. Lancaster Country Day
4. **The investigation reveals that there is potential that the image is a deepfake.**
    a. How might the potential of a deepfake impact the investigation and response?
    a. What communication processes are in place to engage the targeted student's parents or guardians?
    b. Consider the following incidents from In a World of Practice
        i. Miami-Dade County Public Schools
        ii. Lancaster Country Day
5. **As deepfakes become more common, it is crucial to have a clear approach for transparency with the community, safeguarding the privacy of the targeted student while maintaining the integrity of the investigation.**
    a. What information is shared internally and to whom?
    b. What information is shared with the broader community?
        i. How do community relationships and dynamics influence the way information is communicated?
        ii. What are the consequences of both sharing and withholding information for the targeted student, the initiator, and the community as a whole?
    c. How does the school ensure that communications are accurate and comply with applicable law, including the Family Educational Rights and Privacy Act (FERPA)?
    d. What processes are in place at the school to ensure the privacy of students and minimize reputational harm when communicating?
    e. Consider the following incidents from In a World of Practice
        i. Westfield Town School District
        ii. Issaquah School District
6. **Reflect on targeted student impact.**
    a. Discuss the emotional, social, and academic effects on the targeted student and how ongoing support will be provided.
    b. What supports are available for the targeted student during and after the investigation?
    c. How might the school's actions affect the targeted student's experience?

d. How will the school communicate with the targeted student's family?
e. Consider the following incidents from In a World of Practice
    i. Westfield Town School District
    ii. Miami-Dade County Public Schools

7. Reflect on initiator impact.

a. What actions must be taken to protect the privacy and rights of the student(s) who created and shared the deepfake (initiator) content during the investigation?
b. What opportunities can be provided for the student(s) to understand the impact of their behavior, repair harm and make things right with the targeted student and school community?
c. What unique considerations exist if the initiator is not a student?
d. Consider the following incidents from In a World of Practice
    i. Issaquah School District
    ii. Beverly Hills Unified School District

8. Reflect on community impact.

a. What actions might the school take to repair the harm done to the broader community and reestablish trust?
b. How can staff promote digital citizenship while helping students understand the consequences of digital actions and supporting a safe school culture?
c. Consider the following incidents from In a World of Practice
    i. Lancaster Country Day
    ii. Beverly Hills Unified School District

Responding to an Employee-Generated Deepfake Incident

Deepfakes can also be used to portray school staff and administrators in compromising situations or engaged in inappropriate behavior. While many of these deepfakes may be created and shared outside of school, such as through social media, schools must manage community relations while investigating potential staff misconduct and supporting the staff members throughout the investigation. Consider how your school would respond to the below incident as it unfolds.

The *Data Privacy, Legal, and Ethics Considerations* section explores Considerations for Consequences and Repairing Harm and Considerations for Using Detection Technologies. This section may serve as a valuable resource during discussions, providing guidance on handling complex situations while ensuring compliance with legal and ethical standards.

1. A video surfaces on social media showing a Black 7th-grade math teacher in a classroom using vulgar language and making disparaging remarks about students' intelligence. The video quickly gains attention, with parents and students expressing

outrage and calling for the teacher's resignation. The teacher denies ever making such statements and claims the video is fake.

a. What immediate actions should school leaders take to assess the situation and control the spread of the video?
b. How can the school support the employee during the investigation while ensuring a fair and transparent process with the community?
c. What policies and procedures apply?
d. Consider the following incidents from In a World of Practice
    i. Great Valley School District
    ii. Baltimore County Public Schools

2. **A preselected incident team is convened. They treat the incident as an allegation and assign roles.**
    a. Who might be included in the response team?
        i. How representative is the team of the faculty, student body, and broader community?
    b. What is the established process for gathering evidence and interviewing the involved parties?
    c. How might the source of the video be determined?
    d. What considerations might there be regarding legal protocols for digital evidence, adherence to employee rights, and ensuring fair treatment for the accused employee?
    e. How should school leaders document their findings to uphold fairness and avoid assumptions?
    f. Consider the following incidents from In a World of Practice
        i. Great Valley School District
        ii. Carmel Central School District
        iii. Baltimore County Public Schools

3. **Upon investigation, it appears that the video may be a deepfake created and shared by someone in the school community.**
    a. What additional actions should school leaders take?
    b. How should the administration respond to the now suspected targeted individual (math teacher) and concerned school and community members?
    c. What steps are necessary to protect the privacy of all parties involved during and after the investigation?
    d. Consider the following incidents from In a World of Practice
        i. Carmel Central School District
        ii. Baltimore County Public Schools

4. **As the video continues to spread, local media reach out to the district for comment.**

a. How might the school communicate the investigation's status to staff, students, and parents without compromising the privacy or dignity of those involved?
b. How do leaders ensure all communications across the district and among staff remain consistent?
c. What community dynamics should be considered when constructing public communication regarding the incident?
d. How does the information shared and withheld impact the targeted individual, the initiator, and the broader community?
e. Consider the following incidents from In a World of Practice
    i. Carmel Central School District
    ii. Baltimore County Public Schools

5. **The district initiates an investigation into the potential deepfake's origins. The investigation reveals the video was AI-generated and created by a student at the school who had an incident with the teacher earlier in the year.**
    a. What resources or technology tools are available to the district when investigating a possible deepfake?
    b. What are the limitations of these resources in identifying a deepfake?
        i. See the Data Privacy, Legal and Ethics Considerations section for more information on deepfake detection technologies.
    c. Consider the following incidents from In a World of Practice
        i. Baltimore County Public Schools
6. **The teacher returns to work following the investigation but reports feeling uncomfortable and seeing continued harassment from students and community members engaging with the topic on social media.**
    a. How might this incident affect the teachers's reputation, emotional well-being, and professional standing?
    b. How might the school's actions affect the targeted individual's experience?
    c. What steps can the school take to rebuild the targeted individual's trust and sense of safety within the workplace?
    d. What processes does the school have to minimize reputational harm when communicating about the incident?
    e. Consider the following incidents from In a World of Practice
        i. Great Valley School District
        ii. Baltimore County Public Schools
7. **As community members join the conversation, many raise concerns about the negative stereotypes the incident introduced and reinforced about Black individuals, pointing to broader societal impact and systemic issues.**
    a. How can the district use this incident to educate the community about deepfake technology and the harm caused by perpetuating stereotypes?

b. What processes does the school have to minimize the potential negative impact on the school (or district) climate and organizational reputation?
c. Consider the following incidents from In a World of Practice
    i. Great Valley School District
    ii. Baltimore County Public Schools

8. Through the investigation, it becomes clear the student did not understand the severity of their actions, the potential consequences, and the potential impact of creating this deepfake on the targeted individual and the community. It becomes evident that the initiator may face backlash from other students following the incident.

a. What actions must be taken to protect the privacy and rights of the student(s) who created and shared the deepfake content during the investigation?
b. What opportunities can be provided for the student(s) to understand the impact of their behavior, repair harm, and make things right with the targeted individual?
c. What can be done to support the student in reflecting on the broader impact of their actions on the school climate and community?
d. How might the district protect the initiator from future harassment and provide them opportunities to repair the harm caused within the school community?
e. Consider the following incidents from In a World of Practice
    i. Great Valley School District
    ii. Carmel Central School District
    iii. Baltimore County Public Schools

Suggested Activities

- Establishing a response team: Who is currently part of your school or district's response team for student and staff investigations?
    - How representative is the team of the faculty, student body, and broader community?
    - Reference the Deepfake Response Checklist and Data Privacy Crisis Response Plan Component Checklist to assist in determining what roles should be on your response team.
- Crisis Communications: Draft a clear, transparent statement for the following that ensures the privacy of all parties involved is maintained.
    - Initial Response: Communicate with the community during the early stages of the investigation.
    - Investigation Update: Provide an update that the deepfake is confirmed.
    - Final Communication: Issue a concluding statement regarding the incident.

- Refer to the *Noteworthy Excerpts* section in the *In the World of Practice* appendix to examine the effectiveness and impact of district communication during early deepfake incidents.

Data Privacy, Legal and Ethics Considerations

Conducting Investigations Containing Non-Consensual Intimate Imagery (NCII)

Several legal and ethical considerations must guide the process to ensure compliance with privacy laws, respect for individual rights, and the overall integrity of the investigation. These considerations include:

- How can you prevent further distribution of non-consensual intimate images and video?
  - Creating a record does not require additional distribution or capturing the explicit image. Administrators can document the evidence in writing by describing the image, how it was distributed, and to whom, with dates and times. Administrators should be careful not to further distribute or possess explicit images. Legal counsel should be consulted.

- When and how do you involve law enforcement?
  - Does the targeted individual have a right to report?
    - Yes, the targeted individual and/or their parents may report the incident to the police of their own accord.
  - Does the school have to report to law enforcement?
    - Some state laws may require reporting to the police if a statutory-covered crime has occurred in a school. Otherwise, schools can legally disclose information to law enforcement, only with parental consent, a judicial order or lawfully issued subpoena, or under another FERPA exception.[4] (*See FPF's guide on Law Enforcement Access to Student Records.)*
  - Are other reporting requirements triggered?
    - Title IX legal obligations and student protections may apply to sexually explicit deepfake incidents.[5] If the details of the incident cover a reportable Title IX offense, mandatory reporting will be triggered, and Title IX requires that schools conduct a "prompt, impartial, and thorough investigation" of sexual harassment complaints and take appropriate steps toward resolution.[6] Title IX requires that the identities of a sexual harassment

complainant and the alleged perpetrator are kept confidential unless the disclosure is FERPA permitted, it is required by law, or it is necessary to carry out Title IX purposes.

- How does FERPA and other student privacy laws apply?
    - FERPA covers the "Education Record" which is maintained by the school, and does not apply to content not maintained by the school. However, when the school maintains a copy of the deepfake as part of a discipline process, that copy becomes an Education Record.[7] For more information, see the Department of Education's guidance on FERPA's rules about photos or videos.

- How does FERPA impact what can be released to the community?
    - FERPA prohibits disclosing personally identifiable information from students' education records without consent unless an exception applies.[8] Though a school may not violate FERPA or any other relevant state or local student data privacy laws, the school should not attempt to cite FERPA, or general privacy concerns as an excuse not to share that the incident has happened, or not to take appropriate action (see the Lancaster Country Day incident in the In a World of Practice appendix). Schools may share with the community that an incident has occurred and may be able to provide limited details within the confines of student privacy laws to foster a culture of transparency and trust in the community.

Considerations for Consequences and Repairing Harm

Research indicates that schools are implementing various disciplinary measures in response to students creating and distributing NCII content. A report by the Center for Democracy and Technology found that during the 2023–24 school year, 71% of teachers reported that students caught sharing non-consensual intimate imagery (NCII), including deepfakes, faced severe consequences such as referrals to law enforcement, expulsions, or suspensions exceeding three days.[9] In a notable incident, five male students at Beverly Vista Middle School in California were expelled for creating and sharing explicit AI-generated images of female classmates. This action was in line with the Beverly Hills Unified School District's policies, highlighting the serious repercussions of such behavior[10] (see the Beverly Hills Unified School District incident in the In a World of Practice appendix). However, disciplinary responses and legal actions vary significantly among states and districts, making it essential to understand the evolving legal landscape.

Legal frameworks are also evolving to address these issues. In December 2023, two male middle school students were arrested and charged with third-degree felonies in Miami Florida.[11] Similarly, in Lancaster Pennsylvania, two male students were charged in December 2024 with 59 counts (each) of sexual abuse of children ("child pornography") and possession and distribution of "child pornography" under a newly passed law defining possession and distribution of AI-generated

"child pornography" as a felony.[12] To gain deeper insight into how early cases have been handled, explore the examples depicted in the In a World of Practice appendix.

There is a significant body of research documenting the disproportionate impact of punitive policies on students of color, with Black and Latino students facing more severe and exclusionary consequences than their white counterparts for the same offenses.[13] This historical trend poses a substantial risk given the seriousness of deepfake incidents and schools need to consider how they will handle confirmed deepfake incidents with the initiator and the broader community.[14]

The impact on the initiators of deepfake content should be addressed not just because they are minors with developing brains and limited decision-making capabilities, but also because these incidents are opportunities for those students and the broader school community to learn. Schools face the increasing challenge of balancing fair and appropriate discipline with processes that adequately support the targeted individual. According to the Learning Policy Institute (2022), research indicates that exclusionary discipline is ineffective in improving school safety or preventing infractions because it fails to address the root causes of behavioral issues and does not teach students alternative conflict resolution or communication skills.[15] The CDT report (2024) highlights this challenge with only 36% of teachers expressing that they feel their school has a fair process in place to support targeted individuals of deepfake NCII.[16] The report continues to emphasize that the serious consequences imposed on the initiator does not alleviate the responsibility of appropriate targeted individual support.[17]

Schools should consider the following when developing processes, policies, and procedures:

- Review and update existing policies around cyberbullying and harassment to assess how they may or may not meet the needs in these situations. Relying on established policies may perpetuate existing bias and inequities, so leaders should conduct a careful review of disparities in impact and reflect on the effectiveness of current policies. Assessing how those policies have impacted specific student subpopulations can help highlight important gaps in impact to consider when revising the policies to account for deepfake incidents.[18]
- Utilize Co-design when reviewing and implementing policies and processes. Co-design is an important equity-oriented principle that begins with identifying all interested and affected parties and includes working groups that authentically elicit their input. Several sources note the differing opinions and needs of stakeholder groups and the importance of their voice in decision-making.[19]
- Ensure perpetrator accountability: Research has shown that a punitive approach to discipline often does not change negative behavior and can even exacerbate the problem. Restorative justice approaches, including mediated dialogues between targeted individuals and perpetrators, focus on fixing the harm caused by the initiator and have proven effective.[20] This may include mandating offenders to participate in digital literacy

and ethics programs and other restorative justice methods to bring together targeted individuals, offenders, and the community to heal and repair harm resulting from bullying.[21]

- Establish processes for repairing harm. The developmental and social-emotional needs of the initiator of the content should be considered as part of repairing harm. School leaders should consider how to provide opportunities to learn from these incidents and the ways in which the initiator can make things right with the target and the school community.[22]

Considerations for Using Detection Technologies

During an investigation, leaders may turn to available detection methodologies and tools to help identify deepfake content. Leaders should be aware of the limitations of these technologies and the potential impact of their use when considering how to incorporate deepfake detection technologies into crisis planning. While detection tools can be valuable in identifying harmful deepfake content, leaders must remain vigilant about ethical, legal, and fairness considerations throughout the process. These considerations include:

- **Evolving nature of technologies:** Detection algorithms are trained on data sets that include current deepfake content, so they are designed to be effective and responsive to current features in deepfake content. New tools for generating deepfake content are emerging constantly, and detection tools can quickly become outdated or less effective.[23, 24, 25]
- **Bias in detection technologies:** Deepfake detection algorithms are not immune to biases, which research has shown can lead to disparities in their effectiveness across demographic groups. For example, studies found algorithms misclassify the faces of Black individuals as "fake" at higher rates than white individuals. Even algorithms that have been found unbiased, when applied in new settings with deepfake content generated with new methods, can be left ineffective and inaccurate.[26] Using biased detection technologies could result in wrongful accusations and disproportionately impact specific demographic groups.
- **Ethical and legal implications of detection:** Using deepfake detection tools could introduce additional privacy and legal considerations related to sharing sensitive student data with third-party providers. As with other forms of digital evidence, leaders should carefully manage how this evidence is collected, stored, and shared so that the use of detection data complies with privacy laws and does not result in the mishandling of sensitive student information.[27]

Understanding the Legal Landscape

On March 29, 2024, the FBI released Alert Number: I-032924-PSA "Child Sexual Abuse Material Created by Generative AI and Similar Online Tools is Illegal". The FBI is warning the

public that child sexual abuse material (CSAM) created with content manipulation technologies, including generative artificial intelligence (AI), is illegal. Federal law prohibits the production, advertisement, transportation, distribution, receipt, sale, access with intent to view, and possession of any CSAM, including realistic computer-generated images.[28] [29]

Individual state legislatures have also been drafting and enacting legislation in an attempt to address the issue of deepfakes in schools. Some states have preexisting legislation that specifically address online impersonation done with an intent to intimidate, bully, threaten or harass a person through social media sites, email or other electronic or online communications.[30] Beginning in 2019, a number of states passed or updated legislation intended to address the use of deepfakes specifically, whether AI-generated or not: they more broadly apply to deceptive manipulated audio or visual images, created with malice, that falsely depict others without their consent. Most of these laws are targeted at nonconsensual intimate videos and images.[31]

Knowing the specifics of their relevant state deepfake laws and staying updated as new legislation is passed are essential skills for school leaders, whether or not a deepfake incident has yet occurred in their school or district. While it is evident that though states may have similar goals in preventing deepfake incidents, every state is different. School leaders must stay abreast of what, if any, laws concerning deepfakes (or preexisting nonconsensual intimate imagery laws) are on the books in their states, what behavior the law(s) concerns, how the law labels that behavior (sexual harassment, child sexual abuse material, etc), and what (if any) classification of crime or penalty is given to that behavior, and how that law would apply in a school setting.

The nuances and differences between state laws will ultimately result in a difference in school leaders' responses to deepfake incidents. For example, a state may have age constraints in their relevant deepfake laws with different penalties if the incident is minor-to-minor. States also have placed different crime classifications on deepfakes. In Texas for example, producing or distributing a synthetic NCII is a class A misdemeanor, the highest and most serious misdemeanor classification, and in South Dakota, creating computer-generated imagery of a minor in a prohibited sexual act can be a Class 2 felony., [33] Whether or not the deepfake initiator's or distributor's actions are a crime in their states also impacts how school leaders report to law enforcement. And how a state classifies the deepfake initiator's or distributor's behavior may also trigger mandatory reporting requirements if the state considers that behavior to be sexual abuse or assault.

Endnotes

1. Future of Privacy Forum, *Synthetic Content: Exploring the Risks, Technical Approaches, and Regulatory Responses*, https://fpf.org/wp-content/uploads/2024/10/Synthetic_Content_Report_October_2024.pdf
2. Rutgers University. (2022). Deepfakes and cheapfakes primarily harm minoritized groups, new Rutgers study finds. https://comminfo.rutgers.edu/news/deepfakes-cheapfakes-primarily-harm-minoritized-groups-new-rutgers-study-finds
3. Future of Privacy Forum, *Synthetic Content*, 2024
4. Future of Privacy Forum. (2017). *Law enforcement access to data: An overview of legal standards for law enforcement access to data across jurisdictions*. Future of Privacy Forum. https://fpf.org/wp-content/uploads/2017/09/Law-Enforcement-Access-to-Data-Final.pdf
5. Department of Education. (2024, April 29). Nondiscrimination on the basis of sex in education programs or activities receiving federal financial assistance. Federal Register. https://www.federalregister.gov/documents/2024/04/29/2024-07915/nondiscrimination-on-the-basis-of-sex-in-education-programs-or-activities-receiving-federal
6. Ibid.
7. U.S. Department of Education. (n.d.). *When is a photo or video of a student an education record under FERPA?* Student Privacy Policy Office. https://studentprivacy.ed.gov/faq/when-photo-or-video-student-education-record-under-ferpa
8. 34 C.F.R. Part 99 (2024). https://www.ecfr.gov/current/title-34/subtitle-A/part-99?toc=1
9. Center for Democracy & Technology. (2024, September 26). Civic tech fall polling research. https://cdt.org/wp-content/uploads/2024/09/2024-09-26-final-Civic-Tech-Fall-Polling-research-1.pdf
10. Tenbarge, K. (2024, March 8). Beverly Hills middle school expels 5 students after deepfake nude photos incident. NBC News. https://www.nbcnews.com/tech/tech-news/beverly-hills-school-expels-students-deepfake-nude-photos-rcna142480
11. Haskins, C. (2024, March 8). Florida Middle Schoolers Arrested for Allegedly Creating Deepfake Nudes of Classmates. Wired. https://www.wired.com/story/florida-teens-arrested-deepfake-nudes-classmates/
12. Hanna, M. (2024, December 6). Two Lancaster students were charged with 59 counts of sexual abuse after allegedly creating AI nude photos of classmates. The Philadelphia Inquirer. https://www.inquirer.com/education/lancaster-country-day-school-ai-nude-photos-20241206.html
13. Bishop, S., Craven, M., Galer, D., Wilson, T., & Duggins-Clay, P. (2022). Literature Review - School Discipline Literature Review. https://files.eric.ed.gov/fulltext/ED629271.pdf
14. Mallett, C. A. (2016). The School-to-Prison Pipeline: Disproportionate Impact on Vulnerable Children and Adolescents. Education and Urban Society, 49(6), 563–592. https://doi.org/10.1177/0013124516644053
15. Learning Policy Institute. (2022, September 30). School suspension: Research report. Retrieved from https://learningpolicyinstitute.org/product/crdc-school-suspension-report
16. Center for Democracy & Technology, *Civic tech fall polling research*, 2024.
17. Ibid.
18. Anderson, M. (2023, October 10). How should U.S. schools confront deepfakes? Government Technology. https://www.govtech.com/education/k-12/how-should-u-s-schools-confront-deepfakes
19. Center for Democracy & Technology, *Civic tech fall polling research*, 2024.
20. Gregory, A., & Evans, K. R. (2020). The starts and stumbles of restorative justice in education: Where do we go from here? Educational Psychologist, 55(4), 220–231. https://doi.org/10.1080/00461520.2020.1788471
21. NVC Next Gen. (n.d.). Restorative practices: Addressing bullying. https://nvcnextgen.org/restorative-practices/bullying/

22. Lodi, E., Perrella, L., Lepri, G. L., Scarpa, M. L., & Patrizi, P. (2021). Use of restorative justice and restorative practices at school: A systematic literature review. International Journal of Environmental Research and Public Health, 19(1), 96. https://doi.org/10.3390/ijerph19010096
23. Future of Privacy Forum, *Synthetic Content*, 2024
24. Salman, S., Shamsi, J. A., & Qureshi, R. (2023). Deep Fake Generation and Detection: Issues, Challenges, and Solutions. IT Professional, 25, 52–59. https://doi.org/10.1109/mitp.2022.3230353
25. Qureshi, S. M., Saeed, A., Almotiri, S. H., Ahmad, F., & Al, M. A. (2024). Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media. PeerJ Computer Science, 10:e2037. https://doi.org/10.7717/peerj-cs.2037
26. Ju, Y., Hu, S., Jia, S., Chen, G., & Lyu, S. (n.d.). Improving Fairness in Deepfake Detection. https://openaccess.thecvf.com/content/WACV2024/papers/Ju_Improving_Fairness_in_Deepfake_Detection_WACV_2024_paper.pdf
27. Poth, R. D. (2024, October 29). AI and the Law: What Educators Need to Know. Edutopia; George Lucas Educational Foundation. https://www.edutopia.org/article/laws-ai-education/
28. Federal Bureau of Investigation. (2024, March 29). Public service announcement: Emerging online scams targeting consumers. Internet Crime Complaint Center. https://www.ic3.gov/Media/Y2024/PSA240329
29. Prostasia Foundation. (2024). FBI: AI CSAM is illegal if it depicts or is indistinguishable from an actual minor. Prostasia Forum. https://forum.prostasia.org/t/fbi-ai-csam-is-illegal-if-it-depicts-or-is-indistinguishable-from-an-actual-minor/3975
30. National Conference of State Legislatures. (2024). Deceptive audio or visual media (deepfakes): 2024 legislation. https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation
31. Ibid.
32. Tex. Penal Code § 21.165 (2024). Unlawful production or distribution of certain sexually explicit videos. https://casetext.com/statute/texas-codes/penal-code/title-5-offenses-against-the-person/chapter-21-sexual-offenses/section-21165-unlawful-production-or-distribution-of-certain-sexually-explicit-videos
33. South Dakota Legislature. (2024). Bill S79: An act to address [title/subject of the bill]. State Net. https://custom.statenet.com/public/resources.cgi?mode=show_text&id=ID:bill:SD2024000S79&verid=SD2024000S79_20240212_0_E

Supplementary Materials

In a World of Practice: Deepfake Incidents in Education

By examining some of the earlier publicized events, educational organizations can encourage informed discussions and develop proactive measures to address and mitigate the risks associated with deepfakes. The following noteworthy cases demonstrate how deepfakes have affected school communities. School leaders are encouraged to analyze the consequences of these incidents and consider how they may respond to potential incidents.

Student towards Student

Westfield Town School District, Westfield, New Jersey (Oct 2023)

Westfield High School

Westfield High School is one of the earliest and more prominent incidents. Several 10th-grade boys used artificial intelligence software to fabricate and circulate sexually explicit images of several female classmates.

Response:

- District suspended the male student accused of fabricating the images for one or two days
- District sent letter to parents [1, 2]
- District notified and consulted with the police
- One targeted individual sues the student who allegedly made and initially distributed the images [3]

Key Points:

- The question is how does something like this happen, and how can it be prevented? [4]
- "The challenge with this technology is that when it creates a deepfake of you, because it is not actually you, your privacy is not really being violated, although your image has been swapped and changed and another image is being attached to yours to create these very traumatic situations," data AI ethicist Renee Cummings said. [5]
- Westfield High began to investigate in late October. While administrators quietly took some boys aside to question them, the targeted student Francesca Mani said, they called her and other 10th-grade girls who had been subjected to the deepfakes to the school office by announcing their names over the school intercom. [6]
- Targeted Individual impact:

    - "All the other girls agree with me, they don't want him in this school. They are very scared," Mani said. [7]
    - " I don't think my daughter, other victims, and the girls of Westfield High School should be punished by another two and a half years of him in the classroom," Mani's mother Dorota said. [8]
- "All school districts are grappling with the challenges and impact of artificial intelligence and other technology available to students at any time and anywhere. The Westfield Public School District has safeguards in place to prevent this from happening on our network and school-issued devices. We continue to strengthen our efforts by educating our students and establishing clear guidelines to ensure that these new technologies are used responsibly in our schools and beyond." [9]

Reflection:

- A female targeted individual expressed that the administration took different actions with the targeted individuals versus the accused perpetrators when bringing them in for questioning, publicly calling down the girls, and “quietly” taking some boys.
    - How might you ensure the privacy of all students involved in the investigation?
    - How might the school's actions affect the targeted individual’s experience?
    - What processes does your school have to ensure the privacy of students and minimize reputational harm when communicating?
- A data AI ethicist publicly stated that “because it is not actually you, your privacy is not really being violated.” Knowing that some in your community may have this perspective, how might you approach communication and education around deepfakes and harms?
- One targeted individual spoke out about the long-term impact of the incident and her continued fears at school.
    - What supports do you have in place for targeted individuals of sexual misconduct?
    - What supports might you need to provide for the targeted individual during and after the investigation?
    - How might the school's actions affect the targeted individual’s experience?
    - How will you ensure appropriate communication with the targeted individual’s family?
- What is your district/school doing to educate students and staff about responsible use and the harms of deepfakes?

Westfield High School References

1. Singer, N. (2024, April 8). Teen Girls Confront an Epidemic of Deepfake Nudes in Schools. The New York Times. https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html

2. Letter to parents: (2024). nyt.com. https://static01.nyt.com/images/2024/04/04/business/SCHOOL-DEEPFAKE/SCHOOL-DEEPFAKE-jumbo.png?quality=75&auto=webp
3. Breen, K. (2024, February 9). New Jersey teen sues classmate for allegedly creating, sharing fake AI nudes - CBS News. www.cbsnews.com. https://www.cbsnews.com/news/new-jersey-teen-sues-classmate-for-allegedly-creating-sharing-fake-ai-nudes/
4. Priolo, T. (2023, November 2). Westfield High School student accused of creating AI nude images of classmates. FOX 5 New York. https://www.fox5ny.com/news/westfield-high-school-new-jersey-artificial-intelligence-pornographic-images-incident
5. Ibid.
6. Singer, *Teen Girls*, 2024
7. Priolo, *Westfield High School,* 2023
8. Ibid.
9. Ibid.

Issaquah School District, Issaquah, WA (Oct 2023)

Issaquah High School

In Issaquah, police investigated parent complaints that a 14-year-old male student allegedly superimposed breasts and genitalia onto photos of at least six 14 to 15-year old female classmates and a school official.[1] The images used were taken from social media and school events. The incident was reported to the school, but the school did not report the incident to police.[2]

Response:

- Three separate parents reported incident to the police; district did not report to incident the police.[3]
- District reported incident to Child Protective Services after being contacted by police detectives.[4]
- Police referred the case to a local prosecutor, however, the prosecutor did not bring charges against the student.[5]
- Law passed to help safeguard minor "victims" from fabricated and sexually explicit intimate images. Orwall sponsored House Bill 1999, which Gov. Jay Inslee signed into law on March 14, 2024.[6]
- The male student was temporarily expelled and has since returned to school.[7]

Key Points:

- At Issaquah High School near Seattle last fall, a police detective investigating complaints from parents about explicit A.I.-generated images of their 14- and 15-year-old daughters asked an assistant principal why the school had not reported the incident to the police, according to a report from the Issaquah Police Department. The school official then asked "what was she supposed to report," the police document said, prompting the detective to inform her that schools are required by law to report sexual abuse, including possible child sexual abuse material. The school subsequently reported the incident to Child Protective Services, the police report said. [8]
    a. The statement added that the district had reported the "fake, artificial intelligence generated images to Child Protective Services out of an abundance of caution," noting that "per our legal team, we are not required to report fake images to the police." [9]
- According to 404 Media, the police report confirms that the images circulated in Issaquah High School were created using web-based "nudify" or "undress" apps, which automatically and instantly alter photos of women to make them appear naked. [10]
- "AI doesn't make it any less real," said Audrey, a student at the school. [11]

- “We empathize with all students and families connected to this incident,” the school district said. [12]

Reflection:

- What laws must you consider when deciding to contact law enforcement or child protective services? Would a deepfake incident trigger your mandated reporter process?
- The media was able to get the discipline results of the student from a report filed by the district, not by peers or parents. What processes do you have in place to ensure the privacy of all parties involved, including discipline results?
    - How do you ensure accurate, FERPA-compliant communications?

Issaquah High School References

1. Sokol, J. (2024, February 15). Schools navigate the new world of explicit AI-generated images. Issaquah Reporter. https://www.issaquahreporter.com/news/schools-navigate-the-new-world-of-explicit-ai-generated-images/
2. Ibid.
3. Ibid.
4. Sires, C. (2023, November 10). Issaquah teen distributes AI-generated nude photos of female students. Issaquah Reporter. https://www.issaquahreporter.com/news/issaquah-teen-distributes-ai-generated-nude-photos-of-female-students/
5. Bandara, P. (2024, February 16). Police Report Reveals How Deepfake Nude Photos Took Over a High School. PetaPixel. https://petapixel.com/2024/02/16/police-report-reveals-how-deepfake-nude-photos-took-over-a-high-school/
6. Washington State Legislature. (2023). House Bill 1999: [Title of the Bill if Available]. https://app.leg.wa.gov/billsummary?BillNumber=1999&Year=2023&Initiative=false
7. Sokol, *Schools navigate,* 2024
8. Singer, N. (2024, April 8). Teen Girls Confront an Epidemic of Deepfake Nudes in Schools. The New York Times. https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html
9. Ibid.
10. Cox, J. (2024, February 8). ‘What was she supposed to report?’: Police report shows how a high school deepfake nightmare unfolded. 404 Media. https://www.404media.co/what-was-she-supposed-to-report-police-report-shows-how-a-high-school-deepfake-nightmare-unfolded/
11. Goodwillie, K. (2023, November 15). State lawmaker gets involved after AI-generated nude photos of Issaquah students surface. King5.com; KING.

https://www.king5.com/article/news/education/ai-generated-nude-photos-issaquah-students/281-fff44e10-b2d7-4f05-bce5-401302cbbbff

12. Sires, *Issaquah teen distributes*, 2023

Miami-Dade County Public Schools, Miami, FL (Nov 2023)

Pinecrest Cove Academy, Public Charter School

In December 2023, two Miami, Florida middle school boys, ages 13 and 14 were arrested and charged with felonies under a 2022 state law for allegedly creating deepfake nudes of their classmates.[1] The law prevents someone from knowingly circulating "altered sexual" images of an identifiable person without the person's consent. The images involved were of around two dozen male and female classmates between the ages of 12 and 13.[2]

The Florida case appears to be the first arrests and criminal charges against minors as a result of alleged sharing of AI-generated nude images. The boys were charged with third-degree felonies, which is the same level of crime as grand theft auto or false imprisonment.

Response:

- The incident took place over Thanksgiving break in 2023. After returning from break, the school sent a letter to parents stating: "This behavior is unacceptable. We prioritize the privacy and well-being of all students, and we will not tolerate such actions. Upon learning of this situation, we reported it to law enforcement who are now conducting an in-depth investigation. The individuals responsible will be facing disciplinary action."[3]
- School administrators contacted law enforcement immediately, the police opened an investigation for potential criminal conduct.[4]
- The two boys accused of the deepfakes were suspended from school for 10 days (reported by parent of one of the targeted individuals).[5]
- The two boys were arrested on December 22, 2023, transported to the Juvenile Service Department and charged with third-degree felonies.[6]

Key Points::

- After the school administrator "obtained copies of the altered images," the administrator interviewed the victims depicted in them, the reports say, who said that they did not consent to the images being created.[7]
- "It made me feel violated. It made me feel unsafe. Like I don't want to be in a school with boys who did that. I talk to them every day. I looked them in the face and they acted like they didn't do anything," said one of the targeted individuals.[8]

- “I think more of these students should be educated about using these apps but in the right way and not the wrong way,” a targeted individual said.[9]
- Stephanie Cagnet Myron, a Florida lawyer who represents targeted individuals of nonconsensually shared nude images, tells WIRED that anyone who creates fake nude images of a minor would be in possession of child sexual abuse material, or CSAM. However, she claims it’s likely that the two boys accused of making and sharing the material were not charged with CSAM possession due to their age.[10]
- “The first thing I think about is how young the victims are and worried about the kind of impact on them,” Franks says. “But then [I] also question whether or not throwing the book at kids is actually going to be effective here.” [11]

Reflection:

- Several articles reported that the administrator “obtained copies of the altered images.” When the administrator obtained copies of the images, the action may be considered possession and distribution of non-consensual intimate imagery (NCII) and child sexual abuse material.
    - What laws must you consider when conducting investigations?
    - What processes and procedures do you have in place to ensure administrators, and others involved in investigations, don’t inadvertently commit a crime, especially when investigating potential child sexual abuse material?
- One targeted individual spoke out about the long-term impact of the incident and her continued fears at school.
    - What supports do you have in place for targeted individuals of sexual misconduct?
    - What supports might you need to provide for the targeted individuals during and after the investigation?
    - How might the school's actions affect the targeted individual’s experience?
    - How will you ensure appropriate communication with the targeted individual’s family?
- One targeted individual spoke of the need for more education on using AI apps. What is your district/school doing to educate students and staff about responsible use and the harms of deepfakes?
- What policies do you have in place that would govern student discipline for creating and distributing deepfake nudes? What discussion may need to take place to ensure the discipline is effective, while considering the impact on the targeted individual(s) and perpetrator(s)?

Pinecrest Cove Academy References

1. Haskins, C. (2024, March 8). Florida Middle Schoolers Arrested for Allegedly Creating Deepfake Nudes of Classmates. Wired. https://www.wired.com/story/florida-teens-arrested-deepfake-nudes-classmates/

2. Carrero, N. (2023, December 16). Pinecrest Cove Academy parents outraged after daughters' faces used on nude photos. Cbsnews.com; CBS Miami. https://www.cbsnews.com/miami/news/pinecrest-cove-academy-parents-outraged-after-daughters-faces-used-on-nude-photos/
3. Inclán, L. (2023, December 15). 2 South Florida students suspended after using AI to create nude pics of classmates. NBC 6 South Florida. https://www.nbcmiami.com/news/local/ai-app-misused-by-miami-dade-students-to-make-inappropriate-images-of-classmates-police/3184692/
4. Carrero, *Pinecrest Cove Academy,* 2023
5. Ibid.
6. Haskins, *Florida Middle Schoolers,* 2024
7. Ibid.
8. Carrero, *Pinecrest Cove Academy,* 2023
9. Inclán, *2 South Florida students*, 2023
10. Haskins, *Florida Middle Schoolers,* 2024
11. Ibid.

Lancaster Country Day, Lancaster, PA (Nov 2024 - original incident Nov 2023)

Private PK-12 Preparatory School

In November 2023, Lancaster Country Day School in Pennsylvania became aware of a student using artificial intelligence to create and distribute explicit deepfake images of nearly 50 female classmates. The investigation revealed that at least 347 images and videos were distributed.[1] Students felt the school did not appropriately address the incident and coordinated a large student walkout in November 2024.[2] The incident has prompted legal action from parents.[3]

Response:

- The school received an anonymous tip on the incident.[4]
- The school conducted an internal investigation; the school did not notify law enforcement or Child Protective Services.[5]
    - The student denied the allegations and concluded the investigation as the administrators had not seen any of the images and could not corroborate the report.[6]
- Parents contacted police in May 2024.[7]
- A second allegation was made to the school in May 2024, the school investigated this report and reported the incident to the authorities.[8]

- In a June email to upper school families, the alleged perpetrator was identified by the school as a ninth-grade student and referred to using male pronouns. The student and his family were informed then he would no longer be enrolled at Country Day.[9] (lancasteronline)

- Students staged a walkout on November 8, 2024, chanting “Hear us. Acknowledge us. See us.” [10]

- Both the Head of School and Board President resigned.[11]

- Two students (juveniles) were charged with 59 counts (each) of sexual abuse of children (child pornography) and possession and distribution of child pornography in early December 2024.[12]

- Parents have filed a lawsuit against the school.[13]

Key Points:

- “The number of victims involved in this case is troubling, and the trauma that they have endured in learning that their privacy has been violated in this manner is unimaginable,” Adams said in a statement. “The method of abuse here is novel, but the impact on the victims is the same as in any case of child exploitation. It has a deeply harmful effect on the lives it touches.” [14]
- While the school received that tip, “there was no criminal failure on behalf of any school employee to report suspected child abuse as it is currently defined by our laws,” the district attorney’s office said. It said that though school officials are required to report child abuse, “child-on-child harm is exempted” from that definition. [15]
- Adams, the district attorney, noted that Pennsylvania had recently passed laws to define the possession and distribution of AI-generated child pornography as a felony. “Given the broad reach of AI and the harm that can be done as evidenced here, I would urge the legislators to further consider amending our mandatory reporting laws to also include the reporting of AI child pornography,” she said. [16]
- In a meeting with students prior to Friday's walkout Micciche acknowledged "something awful has happened in our community" and apologized for any missteps by the administration in navigating the situation. Classes started later for upper school students Monday and multiple counselors were made available to students on campus that day. . . "We hope to re-earn the trust of those for whom that bond has been broken by listening to you and your student's concerns, answering your questions, and working together towards a greater sense of safety and care," Micciche wrote in an email to families [17]
- In a LinkedIn post, Amanda Bickerstaff, an educator and founder of AI for Education, addressed the issue: “Unfortunately, as we've commented before, the emergence of easy-to-use and increasingly sophisticated tools and platforms with little regulation and

oversight will only make this a more common occurrence." The case also underscores gaps in current US laws addressing AI-generated harmful content. While there have been proposals to criminalize the creation and distribution of explicit AI images, these efforts have largely stalled at the federal level. [18]

Reflection:

- What can you learn from the parent and student response to this incident?
- What processes need to be developed to assist in thoroughly investigating anonymous tips?

Lancaster Country Day References

1. Embry, I., & Rose, M. (2024). 2 juveniles charged with nearly 500 counts after creating, sharing AI nudes of students. Local 21 News. https://local21news.com/newsletter-daily/2-juveniles-charged-with-nearly-500-counts-after-creating-sharing-ai-nudes-of-students-lancaster-country-day-school-sex-abuse-charges-for-two-juveniles-lawsuit-ai-generated-pictures
2. Stalnecker, A. (2024, November 18). Here's what we know and don't know about Lancaster Country Day AI-generated nude image incident. LancasterOnline. https://lancasteronline.com/news/local/heres-what-we-know-and-dont-know-about-lancaster-country-day-ai-generated-nude-image/article_f710874e-a5d1-11ef-82aa-cb325d028aaa.html
3. Stalnecker, A. (2024, November 14). Parents to Sue Pennsylvania School District Over Deepfakes. GovTech. https://www.govtech.com/education/k-12/parents-to-sue-pennsylvania-school-district-over-deepfakes
4. Stalnecker, *Here's what we know,* 2024
5. Ibid.
6. Ibid.
7. Ibid.
8. Ibid.
9. Ibid.
10. Stalnecker, *Parents to sue,* 2024
11. Stokes, E. (2024, November 28). AI deepfake explicit photo scandal sparks resignations and legal action at Pennsylvania school. EdTech Innovation Hub. https://www.edtechinnovationhub.com/news/ai-generated-deepfake-explicit-photo-scandal-sparks-resignations
12. Embry & Rose, *2 juveniles charged*, 2024
13. Stalnecker, *Parents to sue,* 2024
14. Hanna, M. (2024, December 6). Two Lancaster students were charged with 59 counts of sexual abuse after allegedly creating AI nude photos of classmates. The Philadelphia Inquirer. https://www.inquirer.com/education/lancaster-country-day-school-ai-nude-photos-20241206.html
15. Ibid.
16. Ibid.

17. Stalneker, A. (2024, November 18). Lancaster Country Day School cancels classes Monday as challenges over deepfakes continue. LancasterOnline. https://lancasteronline.com/news/local/lancaster-country-day-school-cancels-classes-monday-as-challenges-over-deepfakes-continue/article_9ee180e4-a5a6-11ef-ae00-d3ea73e7727f.html
18. Stokes, *AI deepfake explicit photo*, 2024

Beverly Hills Unified School District, Beverly Hills, CA (Feb 2024)

Beverly Vista Middle School

A group of 5 eighth-grade boys used generative AI to create explicit images of 16 female classmates, ages 12 and 13. The students were expelled as a result of the incident.[1]

Response:

- Administrators quickly sent a message — subject line: "Appalling Misuse of Artificial Intelligence" — to all district parents, staff, and middle and high school students.[2, 3]
    - The message urged community members to share information with the school to help ensure that students' "disturbing and inappropriate" use of A.I. "stops immediately." It also stated that the district was prepared to institute severe punishment. "Any student found to be creating, disseminating, or in possession of AI-generated images of this nature will face disciplinary actions," including expulsion.
- Beverly Hills Police Department launched its own criminal investigation, the prosecutor declined to press charges.[4]
- The 5 middle school students were expelled.[5]

Key Points:

- "Furthermore, we recognize that kids are still learning and growing, and mistakes are part of this process. However, accountability is essential, and appropriate measures have been taken." [6]
- "It's hard to think about what justice would be for the students," she [a student] continued. "The problem with image-based abuse is once the material is created and out there, even if you punish the people who created them, these images could be circulating forever." [7]
- "It's very scary, because people can't feel safe to, you know, come to school," a student at Beverly Vista Middle School who did not want to be identified told NBC Los Angeles. "They're scared that people will show off, like, explicit photos" . . . "It's hard to think about what justice would be for the students," she continued. "The problem with image-based abuse is once the material is created and out there, even if you punish the people who created them, these images could be circulating forever." [8]

- At Beverly Vista Middle School in Beverly Hills, Calif., administrators contacted the police in February after learning that five boys had created and shared A.I.-generated explicit images of female classmates. Two weeks later, the school board approved the expulsion of five students, according to district documents. (The district said California's education code prohibited it from confirming whether the expelled students were the students who had manufactured the images.) [9]
    - Michael Bregy, superintendent of the Beverly Hills Unified School District, said he and other school leaders wanted to set a national precedent that schools must not permit pupils to create and circulate sexually explicit images of their peers. [10]

Reflection:

- Although the district did not identify the 5 students and did not confirm the disciplinary actions, the comments by the district allowed the public to determine that all five students were expelled.
    - What processes do you have in place to ensure the privacy of all parties involved, including discipline results?
    - How do you ensure accurate, FERPA-compliant communications?
- What state laws apply to deepfakes in your schools? Is there pending legislation that could impact your processes?
- What supports do you need to have in place to help all students feel safe in school during and after a deepfake incident? Consider the targeted individuals, but also other individuals who may now fear they will be targeted.

Beverly Vista Middle School References

1. Tenbarge, K. (2024, March 8). Beverly Hills middle school expels 5 students after deepfake nude photos incident. NBC News. https://www.nbcnews.com/tech/tech-news/beverly-hills-school-expels-students-deepfake-nude-photos-rcna142480
2. Singer, N. (2024, April 8). Teen Girls Confront an Epidemic of Deepfake Nudes in Schools. The New York Times. https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html
3. Letter to parents: (2024). nyt.com. https://static01.nyt.com/images/2024/04/04/business/SCHOOL-DEEPFAKE-02/SCHOOL-DEEPFAKE-02-superJumbo.png
4. Tenbarge, K., & Kreutz, L. (2024, February 27). A Beverly Hills middle school is investigating AI-made nude photos of students. NBC News. https://www.nbcnews.com/tech/misinformation/beverly-vista-hills-middle-school-ai-images-deepfakes-rcna140775
5. Tenbarge, *Beverly Hills middle school expels,* 2024

6. Ibid.
7. Tenbarge, *A Beverly Hills middle school is investigating*, 2024
8. Ibid.
9. Singer, *Teen girls confront*, 2024
10. Ibid.

Student towards Staff

Great Valley School District, Malvern, PA (July 2024)

Great Valley Middle School

Eighth graders made at least 22 fake TikTok accounts of teachers filled with pedophilic and homophobic remarks, racist memes and rumors about sexual hookups between the staff.[1] This is an escalation of the fake social media profile issues that schools have been dealing with for the last 5+ years.[2]

Response:

- The school briefly suspended several students.[3]
- Two female students publicly posted an apology video, describing the incident as a joke and said teachers blew it out of proportion. "We never meant for it to get this far, obviously," one of the students said in the video. "I never wanted to get suspended." [4]
- The school held an eighth-grade assembly on responsible technology use.[5]

Key Points:

- The online harassment has left some teachers worried that social media platforms are helping to stunt the growth of empathy in students. Some teachers are now hesitant to call out pupils who act up in class. Others said it had been challenging to keep teaching. [6]
- It's a significant escalation in how middle and high school students impersonate, troll and harass educators on social media. Before this year, students largely impersonated one teacher or principal at a time. [7]
- "We didn't have to deal with teacher-targeting at this scale before," said Becky Pringle, president of the National Education Association, the largest U.S. teachers' union. "It's not only demoralizing. It could push educators to question, 'Why would I continue in this profession if students are doing this?'"[8]
- In 2021, the Supreme Court ruled that public school administrators can punish student speech that occurs outside the school or online if it disrupts classroom study. That case involved a cheerleader who created an expletive-laden social media post but didn't target any one person or name her school.[9]

- Mrs. Scibilia and other teachers are still processing the incident. Some teachers have stopped posing for and posting photographs, lest students misuse the images. Experts said this type of abuse could harm teachers' mental health and reputations. [10]

Reflection:

- What actions might you take to help teachers feel safe and supported as the possibility of fake accounts threatening their reputation circulates in the news?
- How might you respond as a district if your teachers were targeted?

Great Valley Middle School References

1. Hernandez, J. (2024, July 9). A school district in Pa. says students made fake TikTok accounts to target teachers. NPR. https://www.npr.org/2024/07/09/nx-s1-5033803/pennsylvania-middle-school-students-tiktok-teachers
2. Klein, A. (2022, Oct 14). Fake Social Accounts Representing Schools or Districts: What's Being Done. EducationWeek. https://www.edweek.org/leadership/fake-social-accounts-representing-schools-or-districts-whats-being-done/2022/10
3. Singer, N. (2024, July 6). Students Target Teachers in Group TikTok Attack, Shaking Their School. The New York Times. https://www.nytimes.com/2024/07/06/technology/tiktok-fake-teachers-pennsylvania.html
4. Ibid.
5. Hernandez, *A school district in Pa. says,* 2024
6. Singer, *Students target teachers,* 2024
7. Ibid.
8. Ibid.
9. Hernandez, *A school district in Pa. says,* 2024
10. Singer, *Students target teachers,* 2024

Carmel Central School District, Patterson, New York (Feb 2023)

Carmel High School

Three high school students created deepfake videos of the middle school principal. The videos portrayed the principal making racial slurs and violent threats, including the threat of bringing a machine gun to school. A second reported video made references to the Ku Klux Klan. The videos were shared on social media.[1]

Response:

- The district disciplined the students involved.[2]
- The Sheriff determined that no laws were broken.[3]
- District issues statement condemning the “blatant racism" expressed in the videos.[4]
- The parents are now planning to file a lawsuit against high school authorities over the incident.[5]
    - Parents state that this incident is just part of a wider problem of racism in the district.[6]

Key Points:

- The students connected to the fake videos are being “dealt with in accordance with the district’s code of conduct,” according to the district’s statement. It did not specify how the students would be disciplined. [7]
- The incident highlights the lack of legislation in place to protect targeted individuals from abuse in deepfakes as the AI technology becomes more advanced and accessible. In the U.S., there is currently no federal legislation to protect against people’s images being used without their consent in deepfake porn or with any associated technology. [8]
- Schwartz says that the school authorities should have taken much more decisive action. “They didn’t act properly, like most schools would do,” Schwartz said. “When a threat is made like that, the first thing that they should be doing is closing the school, informing all the parents about the nature of the threat, adding extra security, and making sure that law enforcement has addressed the threat.” [9]

Reflection:

- Parents expressed concern that the district did not treat the videos as a possible threat to student safety and only addressed the deepfake aspect of the video.
    - How might your community respond if such videos were posted about your school administrator?
- Parents also expressed concern over what they felt was a lack of communication from the district regarding the incident.
    - How might you address this situation differently?

Carmel High School References

1. Gilbert, D. (2023, March 8). High Schoolers Made a Racist Deepfake of a Principal Threatening Black Students. VICE. https://www.vice.com/en/article/school-principal-deepfake-racist-video/

2. Cutler, A. (2023, March 15). Racist, ominous video of principal was really student-made deepfake, NY school says. Yahoo News. https://www.yahoo.com/news/racist-ominous-video-principal-really-212458810.html
3. Bandara, P. (2023, March 9). Students Who Made Racist Deepfake Video of Principal "Broke No Law." PetaPixel. https://petapixel.com/2023/03/09/students-who-made-racist-deepfake-video-of-principal-broke-no-law/
4. Ibid.
5. Ibid.
6. Gilbert, *High schoolers made,* 2023
7. Cutler, *Racist, ominous video,* 2023
8. Bandara, *Students who made racist deepfake,* 2023
9. Gilbert, *High schoolers made,* 2023

Staff towards Staff

Baltimore County Public Schools, Baltimore, MD (Jan 2024)

A high school athletic director was arrested after he distributed a deepfake racist and antisemitic audio-only clip that impersonated the school's principal. During the investigation, the principal depicted in the audio clip, Mr Eiswert, was placed on administrative leave and received threats to his safety. Police provided safety monitoring for the principal in response to a barrage of harassing messages and phone calls, some threatening him and his family with violence.[1, 2]

Response:

- Upset and angry parents and students flooded the school with calls.[3]
- Some teachers, the police said, feared "recording devices could have been planted in various places in the school." To address safety concerns, the Police Department increased its presence at the school.[4]
- The athletic's director was arrested and charged with stalking, theft, disruption of school operations and retaliation against a witness.[5]

Key Points:

- From the New York Times - Public Response [6]
    - Then in April, Baltimore Police Chief Robert McCullough confirmed they now had "conclusive evidence that the recording was not authentic". And they believed they knew who made the fake. Police charged 31-year-old Dazhon Darien, the school's athletics director, with several counts related to the fake video. Charges included theft, retaliating against a witness and stalking.

    - Months later, the effects of the fake audio clip are still felt in Pikesville. Mr Eiswert has moved jobs and is working in another school. And even though some community members told me they now accept the video is fake, the damage is done.
    - “This is a Jewish neighbourhood and to say something that's so inflammatory about the community was upsetting,” a woman called Sharon told me as she packed her grandchild’s pram into a car in a house opposite the high school last August.
    - For several minutes, Sharon talked to me as though the clip was real.
    - “I think when people say things like that, other people join in that and it makes me more fearful.”
    - When her husband chimed in from the car, reminding her the clip was actually fake, she admitted she did “find out later it was AI-generated”. But she said she was still angry about it.
    - I found that for people like Sharon, who had believed the clip was real, even for a short time, it stayed with them - especially when the message echoed genuine experiences of racism and discrimination. It reminded me of something I hear time and time again while investigating misinformation and conspiracy theories: “Well, even if it’s not real, it’s what I think they think.”
- From the BBC - Public Response [7]
    - “He said right away, oh, we think this is fake... We believe it's AI,” she told the BBC. “I hadn't heard that angle” before. But when she published that explanation, her readers were not convinced. . . . it just fuelled backlash from people who thought the allegation of fakery was just an excuse or an attempt to evade accountability.
    - Principal Eiswert’s reputation had taken a serious hit too. Security was stepped up around both him and the school. He became a target for social media hate and threats. I found dozens of abusive messages taking aim at him on social media.
- "What's so particularly poignant here is that this is a Baltimore school principal. This is not Taylor Swift. It's not Joe Biden. It's not Elon Musk. It's just some guy trying to get through his day," he said. "It shows you the vulnerability. How anybody can create this stuff and they can weaponize it against anybody." [8]

Reflection:

- This case was not “cracked” because it was determined to be a deepfake, it was resolved because “Subpoenaed documents from Google, AT&T, and T-Mobile led police to an internet provider address registered to Darien’s grandmother, and a Baltimore County Public Schools information technology employee searched his access to the school system’s network and found that he used AI tools shortly before the recording was released” [9]

    - What tools do you currently use when investigating technology misuse? How might these tools assist you in investigating potential deepfake incidents?
- After the announcement of the audio being fake, some community members continued to react to it as if it were real, with one person stating “Well, even if it’s not real, it’s what I think they think.” [10]
    - What proactive actions might your district and community take to help establish trust ahead of a potential event?
    - What community dynamics exist that might impact the perception of an incident?

Baltimore County Public Schools References

1. Spring, M. (2024, October 4). The AI clip that convinced - and divided - a Baltimore suburb. Bbc.com; BBC News. https://www.bbc.com/news/articles/ckg9k5dv1zdo
2. Diaz, J. (2024, April 26). A Baltimore-area teacher is accused of using AI to make his boss appear racist. NPR. https://www.npr.org/2024/04/26/1247237175/baltimore-ai-generated-racist-audio-crime
3. Ibid.
4. Price, L. (2024). Athletic director at Maryland high school used AI to fake racist recording of principal, police say. Yahoo.com. https://www.yahoo.com/news/athletic-director-maryland-high-school-220100413.html
5. Diaz, *A Baltimore-area teacher*, 2024
6. Singer, N. (2024, April 25). School Employee Arrested After Racist Deepfake Recording of Principal Spreads. The New York Times. https://www.nytimes.com/2024/04/25/technology/deepfake-recording-principal-arrest.html
7. Spring, *The AI clip that convinced*, 2024
8. Diaz, *A Baltimore-area teacher*, 2024
9. Price, *Athletic director*, 2024
10. Singer, *School employee arrested,* 2024

Checklists

District Readiness Self-Assessment Checklist

Districts can better respond to and mitigate the potential harms of deepfakes by leveraging existing policies and procedures and establishing guidelines for deepfake incident response. The self-assessment below will assist leaders in identifying their current level of readiness so that they can prioritize the steps necessary to ensure an appropriate response to a potential deepfake incident.

District Policies, Procedures and Compliance Considerations

- ☐ Complete the ***Policy and Procedure Checklist*** to ensure that district policies address deepfakes, including image-based sexual abuse: update as needed
- ☐ Determine how existing policies and practices of related incidents might apply to a potential deepfake investigation. (e.g. bullying, harassment, Title IX, sexting, technology use, disruption of school, misconduct outside of school, impersonation of others on social media)
    - ☐ Develop an administrative guide to align district requirements with relevant policies and procedures, ensuring compliance when responding to an incident.
- ☐ Review state and local definitions and laws around the possession and distribution of non-consensual intimate imagery (NCII) and child sexual abuse material (CSAM)and ensure policies and procedures adhere to the definitions and laws
    - ☐ Implement a process to ensure the district keeps current with the relevant laws and regulations
- ☐ Engage in discussions to ensure administrators and appropriate staff understand how laws apply to sharing student information, even when that information may be AI-generated. See *Understanding the Legal Landscape* for more information on the application of privacy laws.
    - ☐ Review your state laws regarding investigations, Title IX, grievances, and student privacy
- ☐ Ensure your policies and procedures have safeguards that take into consideration that any digital media could be a deepfake and reliable detection is difficult when investigating incidents due to limitations and bias in detection technologies. Reference the *Data Privacy, Legal and Ethics Considerations* section for more information about detection technologies.
- ☐ Create guidance to evaluate if a sexually explicit deepfake incident qualifies as a form of sexual harassment or other civil or criminal offense. See *Conducting Investigations Containing Non-Consensual Intimate Imagery (NCII)* for more information.

- ☐ Work with local law enforcement to establish incident thresholds and response responsibilities
- ☐ Consult with your legal counsel regarding your policies, procedures, and guides
- ☐ Consider implementing policies for providers of instructional technology and other data systems on how they should address a deepfake incident involving their systems
- ☐ Develop a Deepfake Crisis Response Plan. Reference the *Deepfake Response Checklist* and *Data Privacy Crisis Response Plan Component Checklist* in the appendix when developing your Deepfake Crisis Response Plan.
- ☐ Establish an after-action review process for all investigations and incidents. Reference the *Data Privacy Crisis Response Plan Component Checklist* for critical components of an after-action review.

Training and Communication Considerations

- ☐ Ensure school leaders have a clear understanding of district policies, procedures, and processes to maintain consistency and fidelity in implementation.
- ☐ Identify and document students, parents, staff, and community members that should participate in the planning conversations to understand and prevent the disproportionate impact of deepfakes on any one group
- ☐ Engage community members, parents, staff, and students in open discussion on:
    - ☐ potential risks for specific student groups
    - ☐ defining appropriate consequences
    - ☐ implementing restorative practices
    - ☐ defining processes to monitor for the disproportionate use of punitive or exclusionary consequences for specific communities and individuals
    - ☐ understanding the impact of the incident and how to address the needs of the targeted individual
    - ☐ communicating needs, biases, and norms
- ☐ Implement a program to educate and train students, staff, and parents about the privacy implications, social-emotional impacts, disproportionate risk to any particular individual or community, potential consequences of deepfakes, and rights of those involved. Use the Training Checklist as a guide in developing your training program.
- ☐ Create guidance on how support to the targeted individual and ensure the confidentiality and privacy of all parties when investigating and communicating about an incident
- ☐ Establish communication protocols around what to communicate and to whom
- ☐ Establish a clearly defined incident response team to preserve the integrity of the investigation and the privacy of those involved.

Policy and Procedure Checklist

Schools can better respond and mitigate potential harms caused by deepfakes by leveraging existing policies and procedures. Leaders should determine how existing policies and practices might apply. Use the below checklist as a guide to review existing policies and identify policies and procedures that may need to be updated or created to address deepfakes, including image-based sexual abuse.

- ☐ Cyberbullying
  - ☐ Recognize deepfakes as a potential form of cyberbullying
  - ☐ Take steps to educate faculty, students, and parents about the possibility of deepfakes as a form of cyberbullying
  - ☐ Ensure disciplinary measures for students who use deepfakes as a form of cyberbullying
- ☐ Harassment
  - ☐ Treat deepfake harassment under Title IX or other relevant laws
  - ☐ Ensure harassment with deepfakes is treated with the same mindset as more 'traditional' forms of harassment
  - ☐ Provide reporting mechanisms for students and staff who may be victims of deepfake harassment
  - ☐ Define how deepfake-generated content can constitute sexual harassment or misconduct under Title IX
  - ☐ Ensure staff and faculty are conscious of how deepfakes are considered under Title IX
- ☐ Grievances
  - ☐ Ensure deepfake-related complaints are treated with the same urgency and respect as more 'traditional' complaints
  - ☐ Create a process for investigating deepfake-related grievances
  - ☐ Offer support for victims of deepfake incidents
- ☐ Student investigations
  - ☐ Ensure privacy, fairness, and transparency with the affected individuals and comply with privacy laws such as FERPA
  - ☐ Provide procedures to mitigate additional distribution of content and removal of content from online platforms or school systems, when appropriate
  - ☐ Ensure compliance with laws around possession and distribution of non-consensual intimate imagery (NCII) and child pornography
  - ☐ Collaborate with law enforcement and other professionals, including legal counsel, when necessary

- ☐ Staff Investigations
  - ☐ Ensure privacy, fairness, and transparency with the affected individuals
  - ☐ Provide procedures to mitigate additional distribution of content and removal of content from online platforms or school systems, when appropriate
  - ☐ Collaborate with law enforcement and other professionals, including legal counsel, when necessary
- ☐ Student Data Privacy
  - ☐ Enforce strict access controls and data security measures to prevent unauthorized access to student data during the investigation, including access by law enforcement
  - ☐ Consider how deepfakes could violate student data privacy
  - ☐ Ensure student privacy through conscious action during the investigation and when communicating with stakeholders
- ☐ Staff Conduct
  - ☐ Ensure deepfakes are addressed as an unacceptable use of district technology
  - ☐ Ensure staff are aware of the legal and ethical ramifications of creating or distributing deepfakes
  - ☐ Explain how deepfakes violate professional conduct
- ☐ Technology Acceptable Use
  - ☐ Ensure deepfakes are addressed as an unacceptable use of district technology
  - ☐ Educate students and staff on the consequences of misusing district technology
- ☐ Social Media Incidents
  - ☐ Ensure deepfakes are addressed as part of the district's social media policy
- ☐ Incident Response Plan / Deepfake Crisis Response
  - ☐ Implement a practical and comprehensive deepfake incident response plan
  - ☐ Ensure that the crisis plan includes a procedure for verifying the authenticity of alleged deepfake content
  - ☐ Create a crisis response team trained to handle deepfake incidents, including media relations, legal considerations, and student/employee support
  - ☐ Provide post-crisis support, including counseling, reputational management, and restoring trust within the community.
  - ☐ Ensure an after-action review is part of the response process
- ☐ Working with Law Enforcement
  - ☐ Keep current with laws and regulations and understand how the laws apply to deepfake incidents including state and local definitions and laws around possession and distribution of non-consensual intimate imagery (NCII) and child pornography
  - ☐ Consider mandatory reporting requirements when responding to sexual content

- [ ] Establish a clear protocol for reporting deepfake incidents to law enforcement
- [ ] District Communications
  - [ ] Establish protocols for internal communication, ensuring that communications are timely and effective, especially in the case of a deepfake crisis/incident
  - [ ] Ensure that external communication (e.g., media and community) is managed to prevent misinformation and ensure accurate information dissemination.
  - [ ] Include procedures for how to handle media inquiries regarding deepfake incidents, including speaking with legal teams to ensure appropriate responses.

Deepfake Response Checklist

Immediate Response

- ☐ Verify authenticity
- ☐ Determine if the incident triggers mandated reporting
- ☐ Determine if the incident triggers the initiation of the crisis response team
- ☐ Secure evidence
    - ☐ If non-consensual intimate imagery (NCII) or child pornography, ensure compliance with state and federal laws regarding collection and handling
    - ☐ Consult local law enforcement or legal counsel as needed regarding appropriate evidence collection and handling
    - ☐ Maintain appropriate chain of custody documentation
- ☐ Determine reach and work to prevent further spread
    - ☐ Determine the extent to which the content has been viewed, shared, or commented on across social media, messaging apps or other platforms
- ☐ Ensure the privacy of all involved
- ☐ Consult policies and procedures to ensure response maintains compliance with local, state, and federal regulations
- ☐ Consider the potential harms and emotional impact of the content and distribution for all individuals involved: victim, initiator, persons involved in or recipients of distribution
- ☐ Begin documentation of incident and response according to your crisis management plan

Stakeholder Communication

- ☐ Engage district leadership
- ☐ Consider notification to legal counsel
- ☐ Consider notification of law enforcement
- ☐ Inform impacted individuals - this communication should be private, avoid using a public address system when requesting students or staff
- ☐ Craft internal communication that is clear, accurate, and non-speculative. Avoid unverified details and ensure the privacy of all individuals involved. Ensure all information is reviewed to prevent the inadvertent release of education records or PII.
- ☐ Craft public messaging that is clear, accurate, and non-speculative. Avoid unverified details and ensure the privacy of all individuals involved. Ensure all information is reviewed to prevent the inadvertent release of education records or PII.
- ☐ Monitor social media and local news to ensure an appropriate response to the community
- ☐ Maintain open channels for parent and community feedback and concerns as appropriate

Communications Considerations

- ☐ Possibility or confirmation of deepfake (synthetic media): ensure language is consistent with the possibility being under investigation or if it has been confirmed
- ☐ Reinforce organizational values on safety, privacy, and emotional health of staff and students
- ☐ Confirm cooperation with law enforcement if applicable
- ☐ Ensure communication relays compassion/empathy with the victim and all involved, recognizing the potential emotional impact
- ☐ Ensure clear, concise language in messaging
- ☐ Consider the cadence of messaging, balancing transparency, privacy, and the integrity of the investigation.
    - ☐ Avoid over-communication and over-dramatization of the incident
- ☐ Consider personal and community biases and norms and the potential impact
- ☐ Consider stakeholder groups for targeted messaging, ensuring sensitivity to context-specific references as appropriate
- ☐ Consider engaging established partnerships to assist with victim support and reputation management

Investigation

- ☐ Collect additional information from relevant sources (ie: students, staff, social media, parents)
- ☐ Collaborate with experts and utilize internal resources to verify authenticity and assess distribution
- ☐ Assess sources and origin of content
- ☐ Review policy violations
- ☐ Determine potential harms and emotional impact of the content and distribution for all individuals involved: victim, initiator, persons involved in or recipients of distribution
- ☐ Consider potential or actual damage to the reputation, trust, and perception of the victim and school within the school and wider community
- ☐ Consider the involvement of legal counsel and local law enforcement

Triage

- ☐ Consider necessary emotional and social support for the victim and the school community
- ☐ Educate the community regarding deepfakes and digital ethics
- ☐ Consider ongoing victim impact when determining the return to normal operations

After-action Review

- ☐ Debrief leadership and key stakeholders through an after-action meeting to evaluate the response and identify gaps
  - ☐ Was the crisis identification and escalation effective?
  - ☐ Were communications effective and appropriately timed?
  - ☐ Did students and staff feel appropriately supported?
  - ☐ Was the victim(s)' emotional health considered throughout the process?
  - ☐ What were the unforeseen challenges?
  - ☐ What did we do well?
  - ☐ What could be improved upon?
- ☐ Update policies and crisis response plans based on lessons learned
- ☐ Train staff to mitigate future risks and improve overall response
- ☐ Engage your community in open discussion to build trust and confidence

Training Checklist

With the complexity of considerations, risks, and harms introduced by deepfake technologies, schools can better prevent, mitigate harm, and respond to deepfake incidents by developing a robust training and education plan that engages all those potentially impacted by an incident. Effective education and training programs can influence how students, families, and communities engage with and are represented in policies and processes for handling deepfakes. Training and education also impact how leaders, faculty, and staff effectively apply the policies outlined in the crisis plan. This checklist suggests some of the education and training needs of the various interested and affected groups in the school community.

For Everyone

- ☐ *Understanding deepfakes and consequences of misuse:* Educate about deepfakes, societal impact, and rights to privacy and protection, along with the legal and ethical responsibilities associated with creating and sharing digital content.

For Students

- ☐ *Digital literacy and citizenship:* Teach students to manage digital identities, understand privacy, evaluate media critically, and engage ethically online, including responsible sharing and reporting malicious content on social media.
- ☐ *Social-emotional impacts:* Educate about the social-emotional impacts of cyberbullying and deepfake incidents with a focus on empathy-building and positive relationships.

For Families and Parents

- ☐ *Guiding students:* Support parents in guiding children's ethical digital behavior, including tools to facilitate open conversations, social-emotional learning, and empathy building.
- ☐ *Identifying and reporting harm:* Support parents in identifying and reporting potential issues to actively mitigate the impact of deepfakes.
- ☐ *Supporting victims:* Equip families with tools to support children victimized by deepfakes and encourage participation in restorative practices to rebuild trust.

For Teachers and Staff

- ☐ *Recognizing, responding, and reporting procedures:* Train educators on identifying deepfake incidents, reporting mechanisms, and their responsibility to uphold students' rights to safety and dignity while maintaining accountability through fair procedures.
- ☐ *Building supportive environments:* Integrate restorative practices and social-emotional learning into school culture to promote inclusivity and positive behavior.

For School Leadership

- ☐ *Developing digital safety policies:* Bring leadership together to create protocols for addressing deepfake incidents, emphasizing fair consequences and collaboration with law enforcement.
- ☐ *Balancing rights and accountability:* Bring leadership together to promote equitable disciplinary measures that balance accountability, students' rights, and school culture.

Data Privacy Crisis Response Plan Component Checklist

A Data Privacy Crisis Response Plan is essential for schools and districts to effectively manage and mitigate the impact of privacy incidents. This checklist outlines the critical components and considerations that should be included in a comprehensive crisis response plan to ensure a coordinated, legally compliant, and transparent approach to handling data privacy breaches.

By incorporating these key elements, school leaders can establish clear protocols for incident detection, response coordination, stakeholder communication, legal and policy compliance, and post-incident evaluation. A well-defined plan helps minimize disruptions, protect sensitive information, and reinforce trust within the school community. Use this checklist to evaluate your existing crisis response plan or develop a new one to effectively manage and recover from data privacy incidents.

Establishing Incident Response Team (IRT):

- ☐ Ensure the team is representative of the faculty, student body, and broader community
- ☐ Suggested team members include district leadership, technology leadership, communications,human resources, and legal council
- ☐ Designate roles, e.g. team lead, communications manager, head of incident investigations, etc.

Education/Training and Awareness around potential Incidents:

- ☐ Stakeholder specific training should be developed
  - ☐ The Incident Response Team should be well versed on the incident response plan and should participate in exercises simulating incidents in order to prepare them for a true incident response
  - ☐ Staff should be made aware of potential incidents, trained to recognize potential incidents, and encouraged to report suspicious findings/activity that may point towards a potential incident
  - ☐ All stakeholders should be briefed on the possibility of incidents, and should be encouraged to report suspicious activity
  - ☐ Reference the Deepfake Incident Training Checklist for stakeholder-specific training

Identifying and Assessing the Incident

- ☐ Triage: determine the severity and authenticity of the incident
- ☐ Identify the type of incident and determine what data has been harmed

- ☐ Deepfake incident: An incident surrounding the creation and/or distribution of deepfakes, a type of media that uses AI to create highly realistic fake images, vidoes, audio, or text. Deepfakes can make it appear as though a person said or did something that they didn't and can lead to misinformation, defamation, and privacy violations.
    - ☐ Deepfake incidents may include but are not limited to: deepfakes made of students by students, deepfakes made of staff by students, and deepfakes made of staff by staff.
- ☐ Privacy Breach: Refers specifically to an incident where sensitive student, staff, or school-related information is accessed, disclosed, or used in a way that violates privacy laws, district policies, or ethical standards. This can include unauthorized sharing of student records, exposing personally identifiable information (PII), or using educational data in ways not consented to by students, parents, or staff. A privacy breach does not necessarily involve hacking or system infiltration but often results from improper handling, unauthorized access, or internal misuse of information.
- ☐ Data Breach: An incident where unauthorized parties gain access to confidential, sensitive, or protected information. The consequences of a data breach can be significant, including financial losses, reputational damage, legal penalties, and harm to individuals whose personal data is compromised.

☐ If the incident is a data breach and/or privacy breach, attempt to pinpoint the root cause of the incident and the circumstances that enabled it

- ☐ Key questions to ask:
    - ☐ Determine the initial attack vector (e.g., phishing, vulnerable software, etc.).
    - ☐ Is the attacker moving within the network? If so, how?
    - ☐ Is the attacker maintaining control?
    - ☐ Identify affected accounts and their privilege levels (e.g., admin, user).
    - ☐ How is the attacker gathering information?
    - ☐ Check if the attacker is spreading to other systems, and how (e.g., remote desktop, malware).

Documenting and Reporting

☐ Documentation should take place at every stage of the incident, including:

- ☐ Detection
- ☐ Analysis
- ☐ Response and containment
- ☐ Recovery
- ☐ Returning to normal operations, and

- [ ] Post-incident discussion
- [ ] Documentation should be precise while still being thorough
- [ ] Make sure to document things like communications, response time, incident root cause, external party involvement, etc.

Determining whether to have external involvement

- [ ] Cybersecurity insurance provider
- [ ] Legal counsel
- [ ] Law enforcement
- [ ] Media
- [ ] Government agencies such as the state department of education, FBI, or CISA

Preserving Evidence

- [ ] Ensure evidence related to the incident is preserved and documented correctly, as well as handled through the proper chain of command.
- [ ] Evidence may be provided to law enforcement for forensic examination

Containing and Isolating the Threat

- [ ] If the incident is a data breach, mitigate the incident by taking measures to isolate the threat. These measures may include:
  - [ ] Isolating compromised systems
  - [ ] Changing passwords
  - [ ] Securing firewalls

Communication

- [ ] Craft clear, accurate, and private communication throughout all stages of the incident. Ensure confidentiality and credibility, and avoid speculating.
- [ ] Internal communications
  - [ ] Communication to Incident Response Team
    - [ ] Continued communication with the incident response team throughout the incident response is vital
  - [ ] Communication to district leadership
  - [ ] Communication to staff members
- [ ] External communications
  - [ ] Communications to individuals affected by the incident

- [ ] Ensure that transparency is apparent while still complying with privacy laws such as FERPA
- [ ] Maintain respectful communications that demonstrate empathy towards all involved. Consider the toll this has taken on individuals involved in the incident, and be mindful of that when communicating.
- [ ] Affected individuals should be notified before the media

- [ ] Communications to media
  - [ ] Ensure compliance to privacy laws such as FERPA when communicating with media and press
  - [ ] Avoid sharing sensitive information
  - [ ] Keep in mind the reputation of the district and avoid painting the situation in an overly negative manner
- [ ] Community engagement/communications with the community
  - [ ] Monitor social media and news for public reactions, and provide ongoing, controlled communication to parents and community members.

Monitoring and Restoring Operations

- [ ] The IRT determines when the incident has been resolved.
- [ ] Monitor for resurgence and begin the process of returning to normal operations.
  - [ ] Communication with all involved on the status of the incident is key

Allowing for Discussions of Post-Incident Review and Improvement

- [ ] The plan has an established process to reflect on how the incident was handled
  - [ ] Were communications organized and timely?
  - [ ] Was the incident response effective?
  - [ ] Was the incident and incident response documented correctly?
  - [ ] Were all necessary individuals involved?
  - [ ] What was done well?
  - [ ] What could have been improved upon?
    - [ ] Ideas to better prepare if another incident happens?

Questions and considerations for the incident response plan

- [ ] Is the plan practical?
- [ ] Is the plan comprehensible?
- [ ] Is the plan up to date?

References & Resources

Checklists, worksheets, sample documents

- Policy and Procedure Checklist
- Readiness Self-Assessment
- Deepfake Response Checklist
- Training Checklist
- Sample Deepfake Crisis Response Plan https://www.linkedin.com/posts/joanne-villis-2b3654125_deepfake-crisis-response-plan-activity-7180148751603224576--FaN/

Guidance Documents

- How to involve law enforcement - references:
    - FERPA Exceptions: A Study in Studies - Student Privacy Compass
    - Law-Enforcement-Access-to-Data (2017)
- Title IX guidance on deepfakes - https://studentprivacycompass.org/new-title-ix-rule-defines-deepfakes-as-sexual-harassment/
- FBI Alert: Child Sexual Abuse Material Created by Generative AI and Similar Online Tools is Illegal - https://www.ic3.gov/Media/Y2024/PSA240329
- Transparency Best Practices for Schools and Districts - About PTAC
- Survey Research on Student, Teacher, and Parent Experiences
- Improving Fairness in Deepfake Detection
- A Guide to Effective Incident Management Communications: https://insights.sei.cmu.edu/documents/1631/2021_002_001_651819.pdf

Professional Resources

- CDT Report - In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/
- FPF Resources
    - FPF Deepfake Infographic
    - FPF Deepfake Curated Resources
    - FPF Blog: Synthetic Content: Exploring the Risks, Technical Approaches, and Regulatory Responses -

https://fpf.org/blog/synthetic-content-exploring-the-risks-technical-approaches-and-regulatory-responses/
- https://arxiv.org/pdf/2409.12138
- https://pmc.ncbi.nlm.nih.gov/articles/PMC8751228/
    - Perpetrator Accountability: Restorative justice approaches, including mediated dialogues between targeted individuals and perpetrators.
    - Educational Interventions for Offenders: Mandating offenders to participate in digital literacy and ethics programs.
- https://nvcnextgen.org/restorative-practices/bullying/
    - - NVC NextGen employs restorative justice methods to bring together targeted individuals, offenders, and the community to heal and repair harm resulting from bullying
- Synthetic Content: Exploring the Risks, Technical Approaches, and Regulatory Responses - Future of Privacy Forum
- Federal Incident Notification Guidelines | CISA. (n.d.). www.cisa.gov. https://www.cisa.gov/federal-incident-notification-guidelines
- CISA. (2021). Cybersecurity Incident & Vulnerability Response Playbooks Operational Procedures for Planning and Conducting Cybersecurity Incident and Vulnerability Response Activities in FCEB Information Systems. https://www.cisa.gov/sites/default/files/2024-08/Federal_Government_Cybersecurity_Incident_and_Vulnerability_Response_Playbooks_508C.pdf
- Thorn. (2023). Youth Perspectives on Online Safety, 2023. Thorn. https://info.thorn.org/hubfs/Research/Thorn_23_YouthMonitoring_Report.pdf
- Thorn. (2021). Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking in 2021. Thorn. https://info.thorn.org/hubfs/Research/Thorn_ROT_Monitoring_2021.pdf
- Smith, A., & Johnson, B. (2023). Differences Between Girls and Boys in the Disclosure of Sexual Violence. Journal of Interpersonal Violence, 38(10), 1–20. https://doi.org/10.1177/08862605231221283
- The Swaddle. (2022). LinkedIn has over 1,000 AI-generated deepfake profiles, find researchers. https://www.theswaddle.com/linkedin-has-over-1000-ai-generated-deepfake-profiles-find-researchers
- Brookings Institution. (2021). The threat posed by deepfakes to marginalized communities. https://www.brookings.edu/articles/the-threat-posed-by-deepfakes-to-marginalized-communities/
- Fix School Discipline. (2016). Breaking the chains report. http://www.fixschooldiscipline.org/wp-content/uploads/2020/09/8.Breaking-the-Chains-Report-2016.pdf

Training References

1. Atkinson,. (2024, July 3). Unmasking Deepfakes: Legal Insights for School Districts. @Aalrr.https://www.aalrr.com/EdLawConnectBlog/unmasking-deepfakes-legal-insights-for-school-districts
2. AI. (2024, May 10). AI for Education. AI for Education. https://www.aiforeducation.io/ai-resources/uncovering-deepfakes
3. McStay, A. (2020, September 30). Emotional AI and children 2020 report. https://doi.org/10.13140/RG.2.2.19873.22888
4. Livingstone, S. and Helsper, E. (2007) Gradations in Digital Inclusion Children, Young People and the Digital Divide. New Media & Society, 9, 671-696. - References - Scientific Research Publishing. (2024). Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=3839758
5. Atske, S. (2023, December 11). Teens, Social Media and Technology 2023. Pew Research Center. https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/
6. Selby. (2023, August 19). Cyber Bullying Scenarios: An In-Depth Analysis for Social Emotional Learning | Everyday Speech. Everyday Speech. https://everydayspeech.com/sel-implementation/cyber-bullying-scenarios-an-in-depth-analysis-for-social-emotional-learning/
7. (PDF) Bullies Move Beyond the Schoolyard A Preliminary Look at Cyberbullying. (n.d.). ResearchGate. https://www.researchgate.net/publication/258201014_Bullies_Move_Beyond_the_Schoolyard_A_Preliminary_Look_at_Cyberbullying
8. Talking to Kids about Social Media. (2022). Kids Mental Health Foundation. https://www.kidsmentalhealthfoundation.org/mental-health-resources/technology-and-social-media/talking-to-kids-about-social-media
9. Earp, W. (2025). Support and Advice for Parents and Carers on Synthetic Media & Deepfakes. Swgfl.org.uk. https://swgfl.org.uk/topics/synthetic-media-deepfake/support-and-advice-for-parents-and-carers/
10. Educators Archives - Cyberbullying Research Center. (2024). Cyberbullying Research Center. https://cyberbullying.org/category/resources/educators
11. Advancing Social and Emotional Learning - CASEL. (2024, December 11). CASEL. https://casel.org/
12. Internet Safety | Cyberbullying and Cyberstalking | Office of Justice Programs. (2021). Office of Justice Programs. https://www.ojp.gov/feature/internet-safety/cyberbullying-and-cyberstalking
13. Free. (2015). Free Speech & Social Media. Knight First Amendment Institute. https://knightcolumbia.org/issues/free-speech-social-media

1350 EYE STREET NW | SUITE 350 | WASHINGTON, DC 20005

info@fpf.org | FPF.ORG