inspiring confidence

ST DOMINIC'S PRIORY COLLEGE

DEEPFAKE CRISIS RESPONSE PLAN

LAST REVIEW DATE	NEXT REVIEW DATE
February 2025	2028
RESPONSIBLE OFFICER	College Principal

We acknowledge St Dominic's Priory College is situated on Kaurna Land and we pay our respects to Elders past, present and future.

inspiring confidence

OUR MISSION

A Catholic College educating girls in the Dominican Tradition.

OUR VISION

Inspired by the Gospel of Jesus Christ, we are a girls' College committed to truth and compassion. In the spirit of St Dominic, we contemplate the possibilities and honour the sacred dignity of each person through word and action. We aspire to provide an innovative, rigorous and inclusive education that leads girls and young women to achieve excellence in their studies and confidence in their future.

OUR VALUES

As a Dominican community we value:

- A sense of the sacred; joyful, eucharistic and reflective.
- A love of learning through creative and critical thinking.
- Modelling a eucharistic community as the basis of transformation.
- Teaching the truth, by word and example.

Further information about the College's principles and objectives can be found within the 2021 – 2025 Strategic Plan.

GOVERNANCE

Dominican Education Australia (DEA) is the governing authority of St Dominic's Priory College, an independent Catholic School. A Board of Directors, established in 1987, governs the College which is incorporated under the *Corporations Act 2001 (Cwlth)*. The Trustees of DEA and the College Board of Directors assures our Catholicity, fidelity to the Dominican charism, formation of Board members, excellence in teaching and learning and financial stability.

For more information about DEA visit: https://dominicaneducationaustralia.com/

CHILD SAFE

We are a Child Safe employer and are committed to the welfare and protection of children and young people. The College is committed to upholding a diverse and inclusive learning environment, ensuring children and young people are valued and respected. In accordance with the National Catholic Safeguarding Standards, all employees are required to comply with the College's relevant policies and procedures.

To read the College's Safeguarding Commitment Statement in full, and access *College Policies, Procedures* and other resources, please click here.

1. PURPOSE

The purpose of this document is to provide the operating direction (Plan) for responses to deepfakes or synthetic digital assets that pose a reputational risk to our students or staff. In an era where digital misinformation can rapidly escalate, this plan is an essential tool for protecting the integrity and reputation of our school community.

2. SCOPE

The Plan specifically addresses crises arising from synthetic digital content, likely shared on social media platforms, which may involve students from our school, or other schools or College staff. It offers a framework for identifying, assessing, and responding to such incidents, with a focus on safeguarding our school's brand and the welfare of our students and staff.

3. DEFINITIONS

3.1 DEEPFAKE

A deepfake is a synthetic media created using artificial intelligence, where a person's likeness is used to create images or videos, often resulting in false representations. This technology can create realistic but fake content, which can sometimes be difficult to distinguish from authentic material.

Deepfakes are typically created using advanced computer algorithms that learn how to replicate the appearance and voices of individuals. These algorithms require a large amount of data, such as photos or videos of the person, to create a convincing fake image or video, however they can be easily and quickly created using freely available apps.

The individuals depicted in deepfakes are victims. The content is artificially created and does not represent their actions or views.

3.2 ARTIFICIAL INTELLIGENCE (AI)

AI is Technology that simulates human intelligence processes by machines, especially computer systems. Key in creating deepfakes.

3.3 SYNTHETIC DIGITAL CONTENT

Synthetic Digital Content is Digitally altered or generated media that can include text, images, videos, and audio, used to refer to content like deepfakes.

3.4 DIGITAL FOOTPRINT

Digital Footprint refers to the information about a person or organisation that exists on the Internet as a result of their online activity.

3.5 FAKE NEWS

Fake news is understood to be false or misleading information presented as news, often spread through digital channels and social media.

3.6 MISINFORMATION

Misinformation is understood to be false or inaccurate information, often spread unintentionally, that can cause public confusion during a crisis.

3.7 CRISIS MANAGEMENT TEAM

The Crisis Management Team is the College Leadership Team; College Principal, Deputy Principal, Business Manager, Director of the Junior School, Director of Faith and Mission, Director of Technology Enrichment, Director of Teaching and Learning.

The IT Manager will join the Crisis Management Team for technical purposes.

At times the College Principal may call on the following positions to join the Crisis Management Team; College Counsellor, WHS Officer, Human Resources Manager.

4. EXAMPLES OF POSSIBLE DEEPFAKE SITUATIONS

4.1 STUDENT

Sarah is a high school student known for her exceptional singing talent. She recently won a local singing competition, gaining attention from her peers and teachers. One day, a classmate, Alex, who is skilled in video editing, decides to create a deepfake video as a prank.

Alex collects several videos of Sarah singing from her social media accounts and public performances. Using sophisticated AI software, Alex overlays Sarah's face onto a video of a famous singer performing at a concert. The AI seamlessly blends Sarah's face with the singer's body, creating a realistic-looking video of Sarah singing at a large concert venue.

Alex then shares the deepfake video on social media, making it appear as though Sarah had performed at a major concert. The video quickly goes viral, garnering thousands of views and shares. Many people, including some of Sarah's friends and teachers, believe the video to be genuine and congratulate her on her supposed performance.

However, Sarah soon discovers the deepfake video and is horrified. She realizes that the video could damage her reputation and lead to misunderstandings among her peers and teachers. Sarah reports the video to the school authorities, who take action to have it removed from social media platforms.

4.2 STAFF

Mrs White is a popular maths teacher at Lincoln High School. She is known for her engaging teaching style and passion for the subject. One day, a disgruntled student, Sam, decides to create a deepfake video as a means of revenge for a bad grade she received.

Sam obtains several videos of Mrs White teaching in class from school recordings and online sources. She then uses AI software to manipulate the videos, altering Mrs White's facial expressions and voice to make it appear as though she is making inappropriate or offensive comments during her lectures.

Once the deepfake video is completed, Sam uploads it to a social media platform, claiming that Mrs White has been making offensive remarks in class. The video quickly spreads among the student body, causing outrage and prompting calls for disciplinary action against Mrs White.

Despite Mrs White's attempts to clarify that the video is fake, many students and parents believe the deepfake to be genuine due to its realistic appearance. The school administration launches an investigation into the matter, causing Mrs White significant stress and anxiety.

Eventually, the deepfake is debunked by experts who confirm that the video has been digitally manipulated. The school takes measures to educate students about deepfakes and digital literacy to prevent similar incidents in the future. However, the incident leaves a lasting impact on Mrs White's reputation and relationships with some students and parents.

5. PLAN

Responses to a deepfake crisis will be coordinated by the Crisis Management Team within the College to the extent:

5.1 College staff are aware of the purpose, scope, definitions and examples of deepfakes as outlined in this Plan.

5.2 The Deepfake Crisis Response Strategies and Plan are to be followed in all cases.

5.3 Overall coordination is the responsibility of Crisis Management Team.

RESPONSE STRATEGIES

In the event of a Deepfake, College personnel must adhere to the guidelines outlined in the four strategies of (6) Identification; (7) Notification; (8) Escalation; and (9) Communication.

6. STRATEGY 1: IDENTIFICATION OF A DEEPFAKE

A situation involving synthetic digital content escalates to a crisis based on several key indicators:

6.1 REACH OF THE CONTENT:

The extent to which the content has been viewed, shared, or commented on across social media or other platforms.

6.2 HURT OR HARM CAUSED:

The impact of the content on individuals, including emotional distress, bullying, or defamation.

6.3 REPUTATION DAMAGE:

The potential or actual damage to the trust and perception of the college with parents and within the wider community.

6.4 INQUIRIES FROM STAKEHOLDERS:

Increased questions or concerns from parents/caregivers, media, or other community members.

7. STRATEGY 2: NOTIFICATION OF A DEEPFAKE

Upon identification of a deepfake, the following steps should be taken:

7.1 NOTIFICATION:

Any staff member, student or college personnel who identifies potential deepfake content should immediately notify the College Principal and Deputy Principal.

7.2 PRELIMINARY REVIEW:

The Technology Solutions Leader conducts a swift initial analysis to verify the authenticity of the content and assess its reach.

7.3 INITIAL MEETING:

The Crisis Management Team (CTM, refer to section 7) convenes an emergency meeting to review the findings and assess the severity of the situation.

7.4 INFORMATION GATHERING:

Collect additional information from relevant sources, including social media platforms, to gauge the impact and spread.

7.5 SITUATION ANALYSIS:

Assess the potential harm and reputational damage, considering the emotional impact on students, staff and college personnel.

8. STRATEGY 3: ESCALATION PROCEDURES

8.1 INTERNAL ESCALATION:

If the CMT, based on the initial assessment, determines that the situation constitutes a deepfake, the College Principal initiates the crisis response protocol.

8.2 EXTERNAL COMMUNICATION:

The College Principal and/or Deputy Principal prepares to address queries from parents/caregivers, media, and other stakeholders, in line with the crisis communication strategy.

8.3 CONTINUOUS MONITORING:

The situation should be continuously monitored by the Deputy Principal for developments or changes in the content's reach and impact. If the target of the deepfake is the College Principal, the Deputy Principal takes on responsibility of continuous monitoring.

9. STRATEGY 4: COMMUNICATION STRATEGY

9.1 KEY MESSAGES

In all communications during a crisis involving synthetic digital content, the following key messages should be consistently conveyed:

9.1.a CONFIRMATION OF SYNTHETIC MEDIA:

Assert that the media is under assessment and there is confidence it is a deepfake.

9.1.b VICTIM SUPPORT:

Acknowledge and support the individual misrepresented in the content, clarifying their non-involvement.

9.1.c COLLEGE'S STANCE:

Emphasise that the views/actions in the content do not reflect the college's ethos and that such incidents contradict the college's values.

9.1.d SAFETY REASSURANCE:

Reassure the safety of students and staff as the highest priority.

9.1.e COOPERATION WITH AUTHORITIES:

Confirm that the college will cooperate with relevant authorities in investigating the incident.

9.2 TONE AND LANGUAGE

The tone of communication should be compassionate and empathetic, recognising the emotional impact on the community.

Language should be clear, brief, and plain to ensure understanding across all audience segments.

9.3 FREQUENCY AND TIMING OF COMMUNICATIONS

9.3.a IMMEDIATE RESPONSE:

Issue an initial communication within two hours of the crisis identification, confirming the synthetic nature of the content and advising against its distribution.

9.3.b ONGOING UPDATES:

Frequency and timing of subsequent communications should be context dependent. In a scenario of repeated, ongoing or numerous deepfakes, updates should be more frequent. Aim for daily updates in ongoing situations. In single occurrences, one initial advisory communication followed up by a more comprehensive summary of findings and actions may suffice.

9.3.c BALANCE IS KEY:

Maintain transparency and availability while avoiding over-dramatising or excessive communication that may heighten concerns.

9.4 TAILORING MESSAGES TO DIVERSE AUDIENCES

9.4.a Core content should remain consistent across all stakeholder groups (e.g. staff, parents/caregivers, and media) to uphold transparency.

9.4.b Tailor each communication to a stakeholder group with a sensitivity to context-specific references appropriate to that group. E.g. parents/caregivers, staff, and students.

9.5 COMMUNICATION CHANNELS

9.5.a FACE-TO-FACE

Communication with the victim should be documented. Documentation should include the date, time, attendees, and discussion notes. Should the victim be a student, all face-to-face communication with students' parents/caregivers should be documented. Documentation records are to be kept by the Deputy Principal. The initial contact is determined by the initial assessment of the Crisis Management Team (CMT) as follows:

9.5.a.i ***Student:*** Where the CMT has determined that the situation constitutes a deepfake for a student, the College Principal initiates immediate contact with *parents/caregivers* and begins the crisis response protocol.

9.5.a.ii ***Staff Member:*** Where the CMT has determined that the situation constitutes a deepfake for a staff member, the College Principal initiates immediate contact with the *staff member* and begins the crisis response protocol.

9.5.b EMAIL:

Email is the best and primary form of communication. It is essential to ensure there are up-to-date email lists for parents/caregivers, staff, and students. These lists may be segmented if necessary, for tailored messaging.

The format and content of email communications should follow these guidelines

9.5.b.i *Subject Line:* Clear and direct, indicating the nature of the crisis (e.g. "Important Update on Recent Digital Incident").

9.5.b.ii *Prioritisation of Information*: Begin with key messages, followed by detailed information.

9.5.b.iii *Consistency*: Maintain a consistent format and tone across all.

9.5.c COLLEGE WEBSITE:

Post major updates and official statements on a dedicated section of the website. Ensure that updates are easy to find and read.

9.5.d SOCIAL MEDIA:

There may be limited use of Social Media platforms to direct audiences to official statements on the College Website, or emails. Engaging in detailed discussions on social media platforms should be avoided; including via private messaging interfaces. For additional information, please see section 9.6 below.

9.5.e PHONE CALLS:

Phone calls are reserved for urgent or sensitive communications with specific individuals or groups. Each call must be documented, for record-keeping purposes.

9.5.f PHYSICAL MEETINGS:

Physical meetings may be considered necessary for in-depth discussions with staff or parent groups. Where applicable, adherence to safety protocols must be observed.

9.6 SOCIAL MEDIA PROTOCOL

In addition to the previously outlined channels of communication, a specific protocol for responding to offending posts on social media is crucial. As such, if a deepfake or, synthetic content, originates from or is widely shared on social media, it may be appropriate to post an official response directly on the platform. The following guidelines will assist in this action.

9.6.a The response should be brief and direct, primarily aiming to redirect viewers to the College's official communication channels (email or College website) for detailed information and updates.

Example Response: *"We are aware of the recent post circulating on social media. Please refer to our official email communication or visit our website for accurate information and updates on this matter."*

9.6.b In certain circumstances, it may be appropriate to post a comment on the offending material with a response stating that the content is a deepfake and does not reflect the views of the College.

9.6.c If the offending material depicted an individual, it is generally inadvisable for the individual to respond. Assuming the individual is a victim of a deepfake, the College would ideally stand between them and any involvement with stakeholders or the wider community.

9.6.d Considerations for Social Media Response:

9.6.d.i *Timing:* Respond as promptly as possible to prevent misinformation from spreading.

9.6.d.ii *Tone*: Maintain a professional and factual tone, avoiding engagement in online debates or confrontations.

9.6.d.iii *Monitoring*: Continuously monitor social media for further developments or spread of the content.

DEEPFAKE CRISIS RESPONSE PLAN

In the event of a Deepfake Crisis, College personnel must adhere to the four-phase response plan below. Immediate response (10), Ongoing Communication (11), Resolution and Recovery (12), and Evaluation (13). It is important that appropriate records and any evidence are kept of the Deepfake Crisis and the response. The College Principal must be informed of all actions and legal advice should also be sought, as necessary.

10. PHASE 1: IMMEDIATE RESPONSE (0-24 HOURS)

10.1 NOTIFICATION AND ASSESSMENT

10.1..a Refer to the College Critical Incident Policy checklist.

10.1.b Immediate notification to the CMT.

10.1.c Rapid assessment of the situation by the College Principal, Technology Solutions Manger, Business Manager and Deputy Principal.

10.1.d Use Assessment Checklist

10.1.e College Principal to inform Chair of the college Board and Director of Catholic education if necessary.

10.2 INITIAL RESPONSE COMMUNICATION

10.2.a Draft and dispatch the first email to parents/caregivers, staff, and students: ideally within two hours.

10.2.b Focus on acknowledging the situation, stating the ongoing assessment, and initial actions being taken and clearly stating that the material is a deepfake.

10.3 SOCIAL MEDIA MONITORING

Continuously monitor social media for spread and reactions, responding as outlined in Section 9.6.

11. PHASE 2: ONGOING COMMUNICATION (24-72 HOURS)

11.1 STAKEHOLDER ENGAGEMENT

11.1.a *Provide Regular Updates*: If the situation is unfolding or has developments that impact the reputation of individuals or the college, consider providing daily updates via email, including any new developments or actions taken.

11.1.b *Maintain open channels of communication*: Maintain open channels for feedback and concerns from parents/caregivers, staff, and students.

11.1.c *Reputational Risk*: Consider if the reputational risk warrants a separate communication to prospective families and parents/caregivers of enrolled, but not yet attending, students.

11.2 INFLUENCER ENGAGEMENT

Consider appealing to parents/caregivers or staff members who are influencers within their groups to proactively endorse the college's actions and to reaffirm that the material is synthetic and therefore not endorsed by the College. Influencers may be parent class representatives, social organisers, year-group ambassadors, or respected staff members.

11.3 CONTINUOUS ASSESSMENT

Ongoing evaluation of the crisis impact and necessary adjustments in strategy.

12. PHASE 3: RESOLUTION AND RECOVERY

12.1 COMMUNICATING RESOLUTION

Once resolved, communicate the conclusion of the crisis, outlining the steps taken and the outcome.

12.2 RESTORING NORMALITY

Plan for a gradual return to normal operations, addressing any lingering concerns.

12.3 LEARNING AND IMPROVEMENT

12.3.a Conduct a debriefing to learn from the crisis and improve future response strategies..

12.3.b Provide a report for the College Board.

13. PHASE 4: POST-CRISIS EVALUATION

DEBRIEFING PROCESS

13.1 STRUCTURED DEBRIEFING SESSION

13.1.a Schedule a debriefing session with the Crisis Management Team (CMT) and relevant staff within a week of resolving the crisis.

13.1.b Focus on a constructive review of the incident, response effectiveness, and areas for improvement.

13.1.c Use the following key questions to assist with the evaluation:
- How effectively did we identify and assess the crisis?
- Were our communication strategies and channels effective and timely?
- How well did we manage the welfare of students, staff, and other stakeholders?
- What were the strengths and weaknesses in our response?
- Were there any unforeseen challenges, and how did we handle them?

13.2 FEEDBACK COLLECTION

While formal feedback collection from stakeholders is not outlined, consider informal discussions or observations noted during the crisis to gauge the community's response.

PLAN REVISION GUIDELINES

13.3 ASSESSMENT OF FEEDBACK AND DEBRIEFING OUTCOMES

13.3.a Analyse the feedback and debriefing insights to identify key lessons.

13.3.b Focus on actionable items that can improve future crisis responses.

13.4 UPDATING THE DEEPFAKE CRISIS RESPONSE PLAN

13.4.a Revise the plan to incorporate lessons learned, particularly in areas of crisis identification, communication strategy, and stakeholder management.

13.4.b Ensure updates are made in consultation with all CMT members and relevant staff.

13.5 TESTING REVISED PLAN

13.5.a Conduct simulations or table-top exercises to test the effectiveness of the revised plan.

13.5.b Make further adjustments as necessary based on these exercises.

14. RESPONSIBILITIES

Responsibilities of the Crisis Management Team when a Deepfake incident occurs.

14.1 COLLEGE PRINCIPAL

Second Contact: Deputy Principal

Responsibilities: Provides overall leadership and final approval for all communications. Coordinates the team, makes final decisions, and approves communication.

14.2 DEPUTY PRINCIPAL

Second Contact: Director of Teaching and Learning

Responsibilities: Manages all external and internal communications. Develops and disseminates communication materials. Ensures the wellbeing of students and staff is prioritised. Monitors and addresses the impact on student and staff welfare.

14.3 IT MANAGER

Second Contact: Director of Technology Enrichment

Responsibilities: Handles technical aspects, including identification and analysis of digital content. Provides technical expertise in assessing and responding to the digital crisis.

14.4 BUSINESS MANAGER

Second Contact: Deputy Principal

Responsibilities: Offers counsel and/or seeks legal counsel, particularly in scenarios with potential legal implications and compliance issues.

14.5 DIRECTOR OF TECHNOLOGY ENRICHMENT

Second Contact: Business Manager

Responsibilities: Make appropriate changes to policies, procedures and staff training practices, including updating this Deepfake Crisis Response Plan.

15. RELATED DOCUMENTATION

St Dominic's Priory College Critical Incident Policy

St Dominic's Priory College Recovery Plan / Business Continuity Plan

7. REVISION RECORD

Document Title	Deepfake Crisis Response Plan					
Document Type	Procedure					
Document Date	February 2025					
Process Owner	College Principal					
	Dr Helen Steele (hsteele@stdominics.sa.edu.au)					
Approval Authority	College Leadership Team					
Review Date	2028					
Distribution	Website		SEQTA		Sharepoint	☑
History	Edition	Date	Description of change			
	1.0	2024	Drafted			