California Community Colleges

MEMORANDUM

May 14, 2026 ISTI 25-300-02| Via Email

TO: Chief Executive Officers, Chief Instructional Officers, Chief Student Services Officers, Academic Senate Presidents, Chief Technology Officers, Chief Business Officers, and Chief Human Resources Officers

FROM: Craig Hayward
Vice Chancellor, AI Strategy and Digital Transformation

RE: AI Memo Series, Part 2 - Centering AI Safety and Security Across the California Community Colleges

INTRODUCTION

The second in the ongoing spring 2026 AI guidance series, this memo builds on the guidance in ISTI 25-300-01 to provide actionable frameworks to ensure the safe and secure use of Artificial Intelligence (AI) technologies. Grounded in the **HUMANS Framework** and the Vision 2030 commitment to equity, this document offers specific mitigations for risks associated with generative and agentic AI.

The Chancellor's Office encourages all colleges to review these guidelines before the end of the spring 2026 term. They provide practical examples and use cases that allow us to better protect student data and privacy, uphold academic integrity, and ensure that technological innovation does not outpace our responsibility to the communities we serve.

BACKGROUND: SAFETY AS A HUMANS PRINCIPLE

In the California Community Colleges, *safety* is defined not just as technical security, but as the protection of our students' identities, their intellectual property, and the integrity of their educational journey. As we transition from simple generative AI tools to systems capable of performing tasks on behalf of users (agentic AI), the potential for risk increases. This memo illustrates how risks can surface and recommends practices for safe use and deployment in the following sections:

1. **Scenarios and Mitigations:** Describing example risks and recommended mitigations.

2. **Key Practices and Safeguards:** Recommending role-based safeguards for AI users, IT deployers, and institutional leadership across the California Community College ecosystem.
3. **Additional Resources:** Referencing key initiatives, concepts, and frameworks for consideration.

Scenarios and Mitigations

Scenario A: AI-Enhanced Student Support

Example Risk: A district deploys an AI bot or agent to assist with financial aid. The tool could inadvertently provide incorrect policy advice.

Mitigation: Districts should prioritize *Human-in-the-Loop (HITL)* protocols, ensuring that all frontline AI systems have a clear "escalation path" to a human staff member, like those piloted in the Cali financial aid pilot. Escalation paths should be tested and validated. In addition to these safeguards, users should be trained and encouraged to verify AI outputs.

Scenario B: Safe Testing and Innovation via Sandboxing

Example Risk: AI users and platforms are increasingly transitioning from primarily generative AI capabilities (tools that summarize or create content) to agentic AI (tools that take action on a user's behalf). Agentic capabilities and behavior, such as browsing the web, executing code, and moving, writing, and deleting files, will soon be commonplace. In the agentic AI space, there is a critical distinction between tools individuals access on their own like agentic browsers (e.g., Perplexity Comet, OpenAI Atlas) and unvetted open-source frameworks (e.g., OpenClaw) versus institutionally managed agents deployed by IT leadership that have been vetted for the educational and operational context. While institutionally-managed agents are deployed with guardrails to protect students, faculty, staff and administrators, unvetted agentic browsers and open-source frameworks may inadvertently access internal portals, student records, or financial systems without explicit institutional oversight.

Mitigation: To foster innovation without compromising district resources, it is best practice to evaluate agentic tools in a sandbox environment before broader deployment.

- For technical leaders, use isolated physical or virtual environments and dedicated "low-privilege" API keys that lack write-access when testing. Testing should include prompt injection simulations to see if the agent can be tricked into bypassing its guardrails.

- For faculty and staff, be aware that IT may delay release of AI tools to test them in a sandbox.

There is also a need to implement robust human-in-the-loop (HITL) procedures. Implementing HITL ensures that an agent is always tethered to a responsible human actor and promotes early discovery of unforeseen consequences. Suggested HITL procedures include:

- *Require Human Authorization:* When first putting an agent into use, the deployer should grant agents limited authority to take actions (e.g., sending an email, moving a file, updating a record). Like a new untrained employee with minimal context or company specific know-how, humans should increase the agent's authority to do a task without explicit human permission *only after* the agent consistently demonstrates successful ability to perform that task.
- *Identity Mapping:* Agentic actions are logged under the identity of the human operator granting the agent authority. If an agent fails, an audit trail can show who approved the agent's scope, access and activity.
- *Limit Access:* Do not install agentic browser extensions on devices used to access sensitive or confidential student SIS or LMS data, employee data, and employer or partner data.
- *Boundary Setting:* Implement policies to prevent agentic browsers from inheriting administrative credentials. Specify which tools, external systems, and APIs an agent can use. Determine which data domains the agent can read from or write to.
- *Disclosure:* Clearly label any process or communication that was facilitated by an autonomous agent.

Scenario C: Students, Staff, and Intellectual Property

Example Risk: Faculty and students upload original writing, research, creative works, or unpublished curricular resources into a public AI tool, ultimately losing control of how that content is used.

Mitigation: Use the **HUMANS** principles as guardrails to ensure that you are properly engaging with AI. **M**anaged privacy controls and **N**otice and explanation refer to the rights of students and staff to control their private data and be notified of the use of AI systems. These measures promote the use of tools that develop human agency rather than diminishing it. Consider establishing an easily accessible, centralized AI resource for students, faculty and staff to determine what institutionally provided tools are available for use and how to access them. Instructors should use de-identified or fictional data for demonstrations and encourage students to use tools

under institutional or enterprise licenses, such as Nectir and Playlab, that are vetted and compliant with FERPA and the latest security protocols (e.g.,SOC 2).

The example scenarios above illustrate the importance of routinely following key safety practices.

KEY PRACTICES AND SAFEGUARDS

To foster a culture of AI safety and human agency, the Chancellor's Office recommends that districts implement the following safeguards:

1. For All Users (Faculty, Staff, and Students)

- **Protect Personal Data:** Never input student PII, grades, or sensitive institutional data into public or non-vetted AI tools. Inquire with your college technology leaders about which systems can safely store this data. Due to the rapid pace of change and unpredictable nature of the technology, review the settings of any new tools you use and tools you already use at least monthly. Be sure to turn off the ability of the tool to use your data for training, when available.
- **Observe and Verify AI Output:** Treat AI-generated outputs as drafts requiring human verification following the guidance of the '**O**' in the PEOPLE framework: **O**bserve and verify. Always critically examine AI-generated content for accuracy, bias, and alignment.

2. For IT and Department Leads (and other AI Tool Deployers)

- **Vetting Requirements:** Before connecting any AI tool to Canvas or other enterprise systems, review the contract and terms of use to confirm it meets WCAG 2.2 Level AA accessibility and user privacy standards. Periodically review privacy and security settings as vendors frequently update their privacy policies or terms of use.
- **Focus on Data Protection:** As a first step, prioritize enterprise subscriptions (e.g., Google Workspace for Education, ChatGPT Edu, Microsoft 365 Copilot, Claude Enterprise Plan, Nectir, Playlab) that provide institutional data protections, ensure FERPA and SOC 2 type compliance, and guarantee that data is not used to train foundational models.

3. For Institutional Leadership

- **Collaborate across roles and functions to develop and update AI policy frameworks** Institutions should engage leadership from all affected constituencies, including administration, faculty, classified professionals, and students, to move beyond reactive prohibition toward a proactive framework of

transparency and shared responsibility. By collaborating across roles and functions, the campus community can establish clear distinctions between "AI-assisted" and "AI-generated" work while providing explicit guidance on citation and disclosure requirements. This collective effort supports stakeholders in defining permissible use through tiered models, such as Open, Conditional, Restricted, or Closed, recognizing that appropriate integration varies by department, context, and function. Collaborating to communicate these parameters in a shared voice ensures that all community members operate with a shared understanding, removing ambiguity and fostering an equitable environment for appropriate levels of AI engagement.

- **Access professional development for modern security standards:** Align local security postures with the systemwide modernizations funded by the cybersecurity allocation provided in **AB 178**. Encourage participation in ongoing professional development provided by the CCC Security Center. Ensure that technical staff and administrators remain current on the evolving AI threat landscape, cybersecurity standards, and systemwide reporting requirements.
- **Agentic AI preparedness**: Encourage IT leadership to conduct workshops or tabletop exercises to think through safe and secure use of AI and agentic AI before deploying at scale. Evaluate vulnerabilities to institutional systems and data and develop strategies to minimize risk of agents misusing, altering, leaking, or exfiltrating these resources.

STRONGER TOGETHER

AI technology is evolving at a pace that requires us to be agile, informed, and mindful. As the second '**E**' in the **PEOPLE** framework reminds us, this is an evolution. By applying best practices today, we position ourselves to safeguard our system's integrity and keep the well-being of our students at the center of ongoing innovation. In our next memo, we will take a closer look at how a strong foundation in privacy shapes and strengthens this work, ensuring innovation remains both responsible and trusted. *In the meantime, we encourage you to explore additional resources*, including more detailed information in the FAQ and professional development opportunities on the **AI Microsite** at ai.cccco.edu.

We welcome your feedback, suggestions and contributions. Please submit your ideas and suggestions to **aiguidance@cccco.edu**

Craig Hayward

Vice Chancellor, AI Strategy and Digital Transformation

cc: Sonya Christian, Chancellor

Rowena Tomaneng, Deputy Chancellor

Chris Ferguson, Executive Vice Chancellor

John Hetts, Executive Vice Chancellor

James Todd, Vice Chancellor

Stacey Shears, Vice Chancellor

Siria Martinez, Vice Chancellor

Jerry Deschler, General Counsel

Jory Hadsell, Executive in Residence

Don Daves Rougeaux, Senior Advisor

AI Council