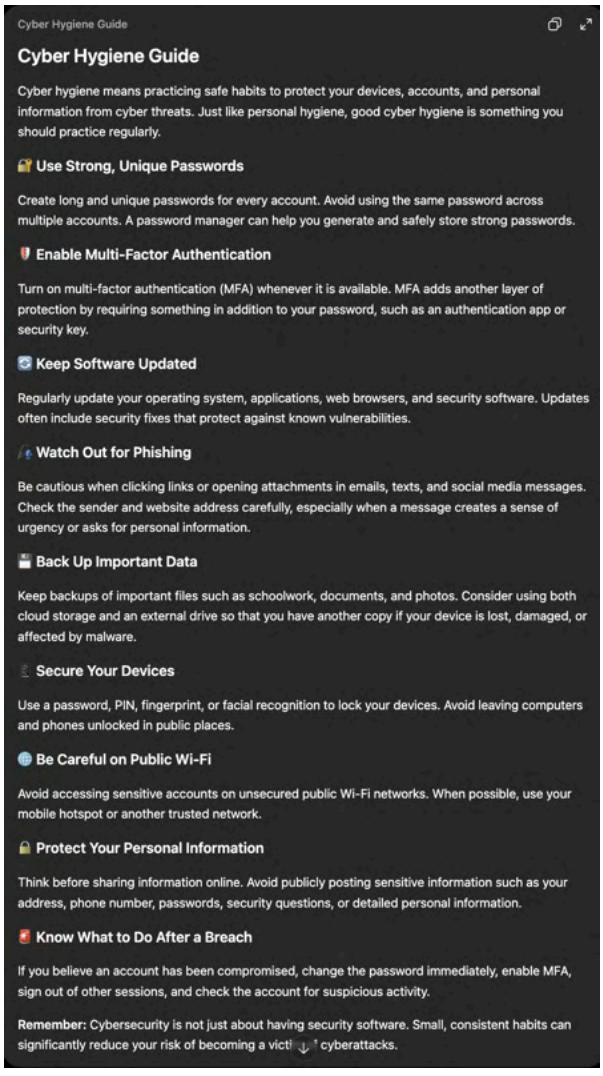
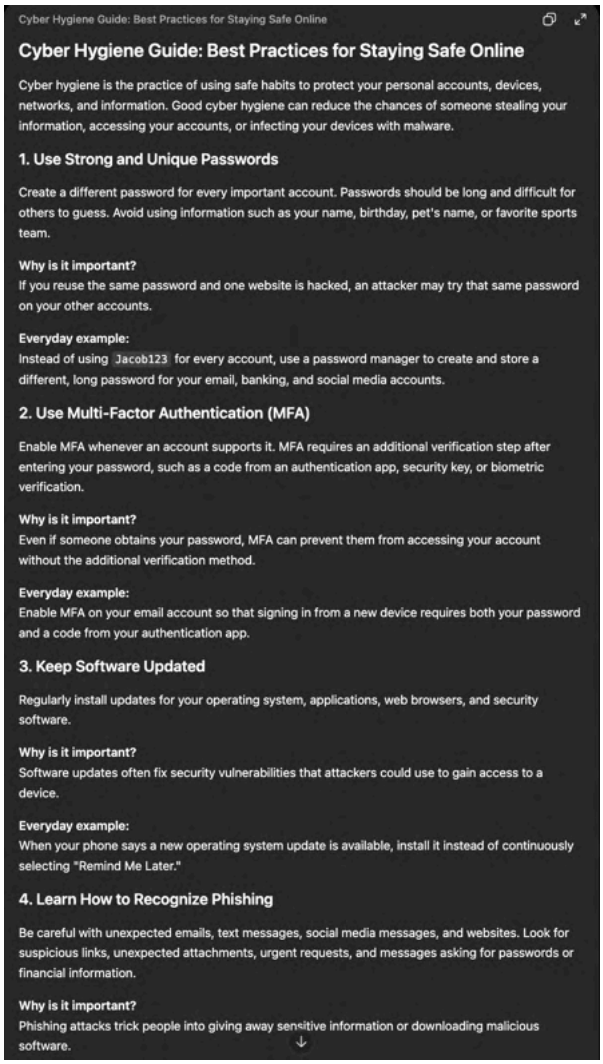
Jacob Rivera
AI410
Sept 19, 2026

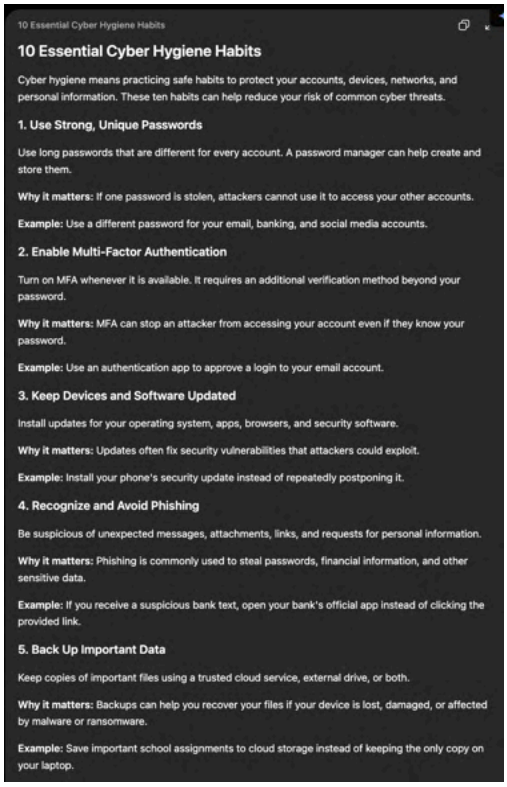
04 | Learning Lab

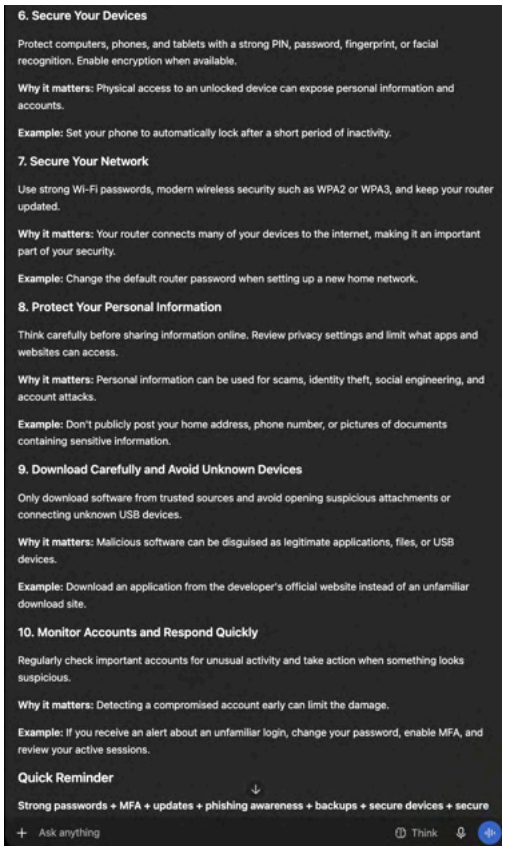
Part 1: Content creation with GenAI

Prompt	Output
1. create a short cyber hygiene guide. include best practices and anything you deem worthy of noting	

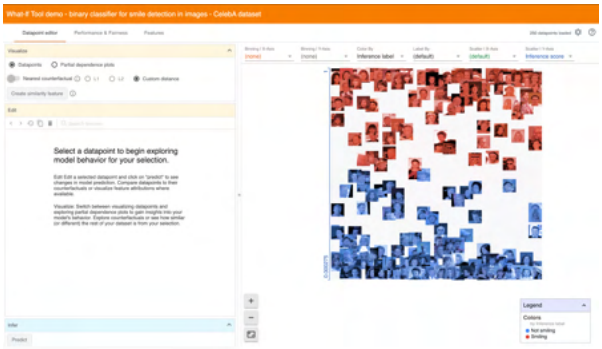
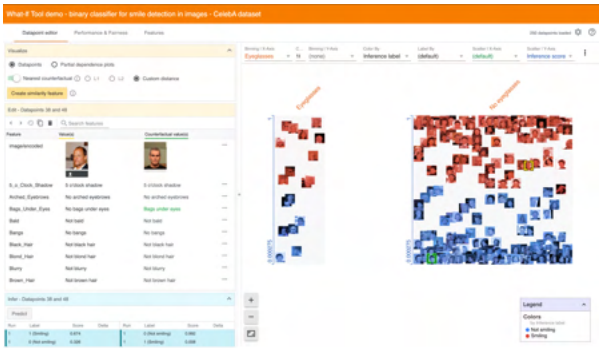
2. expand the cyber hygiene guide by providing more details. include additional best practices that can help protect personal accounts, devices, networks, and sensitive information. for each practice, explain why it is important and provide a simple example of how someone can apply it in their everyday life. make it easy to understand.	 **TOO LONG (INCLUDED 24 ADDITIONAL PRACTICES)**

3. Condense the guide into ten habits





Part 2: Apply XAI Techniques

Screenshot	Short Explanation
	This is when I first loaded the page on Google's What If Tool. This classified smile detection using the CelebA dataset. There were 250 data points as images of different people and blue means someone not smiling and red means someone is smiling.
	Here I selected the sunglasses option under the Binning X-axis option which separated people wearing sunglasses from those who weren't. There were more features such as black hair, sideburns, bald, etc. I also enabled the nearest counterfactual option, which allowed me to examine the selected datapoint and compare it with another datapoint that is similar but has a different prediction from the model.

Part 3: Reflection

Generative AI and Explainable AI are changing the cybersecurity field by making it possible to automate tasks and analyze datasets very quickly. This makes it easier for cybersecurity professionals, as they can assist with security analysis. A personal example I can think of is when I used AI to automate the detection and response portion of a home lab I did this summer. To sum it up, I was able to get security alerts from an SIEM tool, such as Splunk, and connected it to a workflow automator, such as n8n. N8n would then call ChatGPT's API tokens, which would analyze the alert, perform its own severity assessment on it, and give the best recommended actions. In this situation, GenAI allowed me to automate the triage process an entry level analyst normally would and reduced the manual work needed to analyze each alert. However, I can see where risks can arise when using AI in cybersecurity. Generated results can sometimes be inaccurate, which can mislead someone to forming incorrect security decisions. For example, false positives can cause an analyst to spend time investigating activity that isn't actually legitimate. Conversely, an incorrect assessment on a real threat can cause it to be overlooked. Professionals should make sure that they're properly handling AI systems when using it as an assisting tool, and to promote the trustworthiness of AI use, they should always verify generated information. Some other steps they can take to do this would be to understand the limitations of the model they use, make clear policies and guidelines for when the model can be used, and require human review. Overall, I believe AI integrated tools in cybersecurity are the future for making security operations more efficient. However, it should never be a complete replacement for human judgement.