

Google का एआई सिस्टम को सुरक्षित रखने का फ्रेमवर्क

एआई तेजी से बेहतर हो रहा है। ऐसे में, यह ज़रूरी है कि इसे सुरक्षित रखने के लिए कारगर जोखिम प्रबंधन रणनीतियां भी विकसित की जाएं। इस दिशा में हमारा एक कदम है एआई सिस्टम को सुरक्षित रखने का फ्रेमवर्क (एसएआईएफ़), जो कि एआई को सुरक्षित रखने के लिए कॉन्सेप्चुअल फ्रेमवर्क है। एसएआईएफ़ के छह मुख्य भाग हैं:

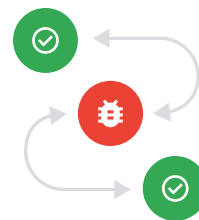
1. एआई ईकोसिस्टम में सुरक्षा की बुनियाद को मज़बूत करना

एआई सिस्टम, ऐप्लिकेशन और उपयोगकर्ताओं को सुरक्षित रखने के लिए, पिछले दो दशकों में बनाए गए डिफ़ॉल्ट रूप से सुरक्षित बुनियादी ढांचागत सुरक्षा और विशेषज्ञता का लाभ उठाएं। साथ ही, एआई के विकास के साथ तालमेल बनाए रखने के लिए संगठनात्मक विशेषज्ञता विकसित करें। साथ ही, एआई और विकसित हो रहे खतरे के मॉडल के हिसाब से बुनियादी ढांचे की सुरक्षा को बढ़ाना और अनुकूलित करना शुरू करें। उदाहरण के लिए, SQL इंजेक्शन जैसी इंजेक्शन तकनीकें कुछ समय से मौजूद हैं और संगठन तेजी से होने वाले इंजेक्शन शैली के हमलों से बेहतर बचाव में मदद करने के लिए इनपुट सैनटाइजेशन और लिमिटिंग जैसे बदलाव कर सकते हैं।



2. संगठन की खतरे की पहचान करने और जवाबी कार्रवाई करने की योजना में AI को शामिल करना

खतरे का पता लगाने की क्षमता और दूसरी क्षमताओं को बढ़ाकर, एआई से जुड़ी साइबर घटनाओं का समय पर पता लगाएं और जवाबी कार्रवाई करें। संगठनों के लिए इसका मतलब है कि वे गड़बड़ियों का पता लगाने के लिए जनरेटिव एआई सिस्टम के इनपुट और आउटपुट की निगरानी करें और हमलों को भांपने के लिए खतरे की जानकारी का इस्तेमाल करें। ऐसा करने के लिए, आम तौर पर भरोसे और सुरक्षा के साथ मिलकर काम करने, खतरे का पता लगाने और गलत इस्तेमाल को रोकने वाली टीमों की ज़रूरत होती है।



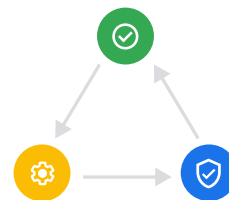
3. मौजूदा और नए खतरों से निपटने के लिए सुरक्षा को ऑटोमेट करना

सुरक्षा में सेंध की घटनाओं पर जवाबी कार्रवाई करने के पैमाने और गति को बेहतर बनाने के लिए लेटेस्ट एआई इनोवेशन का इस्तेमाल करें, विरोधी संभावित रूप से अपने असर को बढ़ाने के लिए एआई का इस्तेमाल करेंगे, इसलिए उनके खिलाफ़ बचाव में चुस्त और लागत प्रभावी बने रहने के लिए एआई और इसकी मौजूदा और उभरती क्षमताओं का इस्तेमाल करना बेहद अहम है।



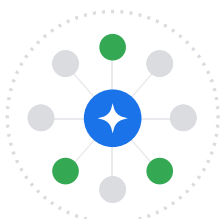
4. पूरे संगठन में एक जैसी सुरक्षा बनाए रखने के लिए, प्लैटफ़ॉर्म के लेवल पर कंट्रोल देना

एआई के साथ जोखिम को मिटाने में मदद करने के लिए कंट्रोल फ्रेमवर्क को अलाइन करें और अलग-अलग प्लैटफ़ॉर्म और टूल पर सुरक्षा बढ़ाएं, ताकि यह पक्का किया जा सके कि सभी एआई ऐप्लिकेशन के लिए स्कैल की जा सकने वाली और किफ़ायती लागत वाली बेहतरीन सुरक्षा उपलब्ध हो। Google में, इसमें Vertex AI और Security AI Workbench जैसे एआई प्लैटफ़ॉर्म पर डिफ़ॉल्ट रूप से सुरक्षित सुरक्षा को बढ़ाना और सॉफ़्टवेयर डेवलपमेंट लाइफ़साइकल में कंट्रोल और सुरक्षा देने वाले सिस्टम बनाना शामिल है। Perspective API जैसी, आम इस्तेमाल के मामलों में काम आने वाली क्षमताएं पूरे संगठन को अत्याधुनिक सुरक्षा देने में मदद कर सकती हैं।



5. बदलाव के हिसाब से ढलने और एआई को डिप्लॉय करने के लिए ज़्यादा तेज़ फ़ीडबैक लूप बनाना

लगातार लर्निंग के ज़रिए इंप्लिमेंटेशन को टेस्ट करते रहें और बदलते खतरों से निपटने के लिए पता लगाने और सुरक्षा के सिस्टम बेहतर बनाएं। इसमें घटनाओं और उपयोगकर्ता से मिले फ़ीडबैक के हिसाब से रीएन्फ़ोर्समेंट लर्निंग जैसी तकनीकें शामिल हैं। साथ ही, इसमें ट्रेनिंग डेटा सेट को अपडेट करने, हमलों के खिलाफ़ रणनीतिक रूप से कार्रवाई करने के लिए मॉडल को बेहतर बनाने और इस संदर्भ में (जैसे कि असामान्य व्यवहार का पता लगाना) बेहतर सुरक्षा को एम्बेड करने के लिए मॉडल बनाने के लिए इस्तेमाल किए जाने वाले सॉफ़्टवेयर को अनुमति देने जैसे कदम शामिल हैं। एआई की मदद से उपलब्ध कराए जाने वाले प्रॉडक्ट और क्षमताओं की भरोसेमंद सुरक्षा में सुधार के लिए, संगठन नियमित तौर पर रेड टीम एक्सरसाइज भी कर सकते हैं।



6. एआई सिस्टम से जुड़े जोखिमों को कारोबार से जुड़ी कार्रवाइयों के संदर्भ में समझना

संगठन एआई को कैसे डिप्लॉय करेंगे, इससे जुड़े हर जोखिम का आकलन करें। इसमें, कारोबार से जुड़े हर जोखिम का आकलन शामिल है, जैसे कुछ खास तरह के ऐप्लिकेशन के लिए डेटा लीनियेज, वैलिडेशन और ऑपरेशन से जुड़े व्यवहार की निगरानी करना। इसके अलावा, संगठनों को एआई की परफ़ॉर्मंस को वैलिडेट करने के लिए ऑटोमेटेड जांच करने वाला सिस्टम बनाना चाहिए।