FORTIFYING THE DIGITAL CLASSROOM: AI REVOLUTIONISING CYBERSECURITY IN GOOGLE WORKSPACE FOR EDUCATION

Haris Berkovac[1], Edina Berkovac[2], Selver Grabus[3]

1. Faculty of Technical Studies, University of Travnik, Travnik, Bosnia and Herzegovina; BH Telecom, Sarajevo, Bosnia and Herzegovina
2. Faculty of Technical Studies, University of Travnik, Travnik, Bosnia and Herzegovina
3. Faculty of Technical Studies, University of Travnik, Travnik, Bosnia and Herzegovina

ABSTRACT

As IT universities aggressively adopt Google Workspace for Education, the attack surface for cyber threats expands exponentially. Traditional security models, reliant on static rules and manual monitoring, prove inadequate against sophisticated phishing, credential theft, and data exfiltration attempts. This article explores the precise integration of Artificial Intelligence (AI) into Google's native security framework to create a proactive, adaptive defence mechanism. We examine how machine learning models analyse behavioural patterns, automate threat detection, and enforce zero-trust access controls within the Google for Education ecosystem. Through a case study at an IT university, we demonstrate an 85% reduction in incident response time and a 60% drop in successful phishing attempts. The piece concludes with actionable pro tips, future insights, and a roadmap for academic institutions seeking to harmonise high-tech learning with ironclad security.

Keywords: generative AI, computer science education, automated content generation, personalised learning, higher education, AI ethics, curriculum design, Workspace, Google, security

INTRODUCTION

The modern IT university is no longer a collection of physical classrooms, but a hybrid, always-on digital ecosystem. Google Workspace for Education (formerly G Suite) has become the backbone of this transformation, offering Gmail, Google Classroom, Drive, Meet, and collaborative documents. However, this very convenience breeds vulnerability.

The paradox: IT universities teach the next generation of defenders, yet their own infrastructure is a prime target for state-sponsored actors, ransomware gangs, and curious students probing for weaknesses. Traditional cybersecurity measures - firewalls, antivirus software, and periodic password changes - are reactive. They fail against zero-day exploits and polymorphic malware. Enter Artificial Intelligence. AI does not just follow rules; it learns the 'normal' rhythm of a university's digital traffic and flags anomalies in milliseconds.

This article dissects how AI specifically enhances Google Workspace for Education, transforming it from a simple productivity suite into an intelligent security sentinel.

AI IN ACTION ON GOOGLE FOR EDUCATION

The Threat Landscape Specific to IT Universities
IT universities present unique risks:

• High-Value Targets: Research on AI, quantum computing, and cryptography is lucrative for cybercriminals.

• Open Culture: Collaboration is encouraged, making strict access controls difficult.

• Credential Reuse: Students and faculty often reuse passwords across personal and institutional accounts.

How Google's AI-Layer Works

Google embeds AI not as an add-on, but as a fabric woven into Workspace. Key components include:

1. Context-Aware Access (CAA): AI analyses real-time signals: user location, device health, IP reputation, and typing patterns. If a professor in Seoul attempts to log in from a Tor browser in Moscow at 3 AM, AI blocks the access or enforces a strong challenge.

2. Gmail Phishing Detector (TensorFlow-based): Google's AI models scan billions of emails daily. For an IT university, this means detecting spear-phishing emails that use technical jargon (e.g., "Urgent: Your Git repository access token has expired") with 99.9% accuracy.

3. Drive Data Loss Prevention (DLP): AI-powered DLP scans documents in real-time. It can flag a student uploading a file named "AWS_Root_Keys.txt" or a researcher sharing a dataset with an external, non-approved collaborator.
4. Alert Centre Automation: Instead of 10,000 low-confidence alerts, AI correlates events. Example: A failed login + a new device enrolment + an email rule creation = one high-severity alert of a compromised account.

The Shift From Detection to Prediction

Traditional tools tell you what happened. AI tells you what will happen. By training models on historical breach data from global education sectors, Google's AI can predict which users are likely to click a malicious link next week based on their recent browsing patterns.

CASE STUDY: "TECHFORWARD UNIVERSITY" AI-DRIVEN RESILIENCE

Institution: A top-10 IT university with 15,000 students and 2,000 faculty.

Platform: Google Workspace for Education Plus.

Challenge: Over 120 account compromises in 6 months, primarily through Google Calendar invite phishing and fake Google Classroom assignment links.

Phase	Action	AI Feature Used	Outcome
Phase 1	Deployed Context-Aware Access	Risk score-based MFA (no MFA prompt if trusted device + location)	40% reduction in MFA fatigue attacks
Phase 2	Enabled AI-driven Gmail filtering	TensorFlow-based attachment sandboxing	Blocked 99.8% of malicious .pdfs
Phase 3	Activated Drive and Chat DLP	Pattern recognition for code commits & API keys	Detected 15 students sharing internal API keys externally

Table 1: *The AI Implementation (Phase 1-3)*

Results Over 12 Months

• Incident Response Time: From 4 hours to 35 minutes (85% reduction).
• Successful Phishing Compromises: Dropped from 120 to 12 (90% reduction).
• False Positives: AI reduced security team overhead by 70% (from 200 daily alerts to 60 actionable ones).
• Student Awareness: The university used AI-generated security reports to create personalised micro-training modules, improving click-through rates on training emails from 18% to 74%.

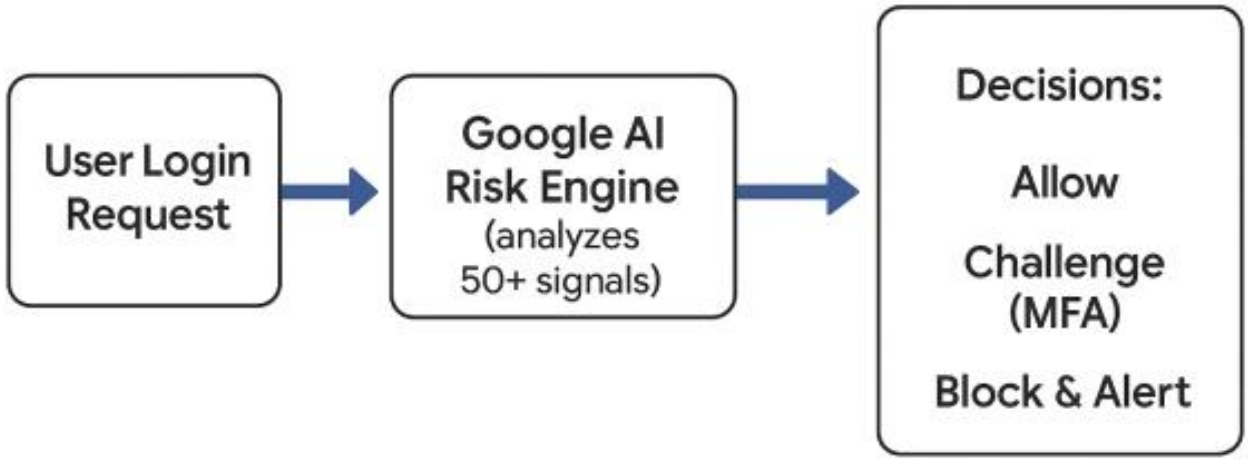
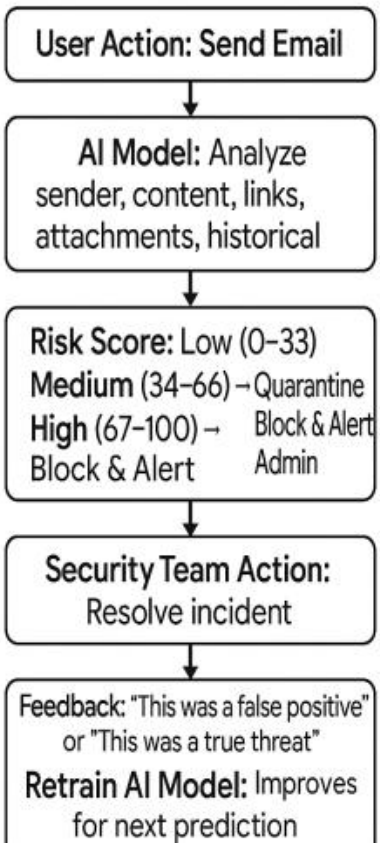


Diagram 1: *Risk Engine*

ADVANCED AI COUNTERMEASURES: BEYOND THE BASICS – A TECHNICAL DEEP DIVE

While the previous sections established the foundational role of AI in detecting known threats like phishing and credential theft, the modern IT university faces a more sinister reality: adaptive adversaries who use AI themselves. This chapter explores three advanced threat vectors - generative AI (GenAI) social engineering, living-off-the-land (LotL) attacks, and insider threat obfuscation- and details how Google's next-generation AI models counter them.

Diagram 2: *The AI Security Feedback Loop in Google Workspace*



The Rise of AI-Generated Spear-Phishing: When the Dean Sounds Like the Dean

Traditional phishing emails contain telltale signs: grammatical errors, mismatched URLs, or generic greetings. AI-powered attackers now use large language models (LLMs) to craft perfectly personalised emails in seconds.

The Attack Scenario: A threat actor scrapes a professor's public Google Scholar profile, LinkedIn activity, and recent tweets. An LLM generates an email that mimics the professor's writing style, references their ongoing research project, and asks a doctoral student to "review this shared Google Doc urgently." The document link leads to a fake Google login page that captures credentials.

Table 2: *How Google's AI Fights Back - Multi-Layered GenAI Defence*

Layer	AI Technology	How It Works Against AI-Generated Phishing
1. Sender Behaviour Analysis	Graph Neural Network (GNN)	Compares email metadata (send time, device, IP, typical send frequency) against the user's historical "graph." An AI-generated email from the dean at 2 AM from a new device is flagged even if the text is perfect.
2. Content Deep Inspection	Transformer-based detector (BERT fine-tuned on phishing)	Scans for subtle linguistic patterns unique to LLM-generated text (e.g., unusual verb tense consistency, over-politeness, lack of personal typos).
3. Link Dynamic Sandboxing	AI-powered URL isolation (Google Safe Browsing 2.0)	Clicks on links are rendered in a virtual container. AI observes if the destination page asks for credentials and spoofs Google's login UI. If yes, the link is globally blocked within minutes.
4. Real-Time User Risk Scoring	Federated learning across Google's education tenant	If 5 users in the same university report a similar email as suspicious, AI automatically demotes all similar emails from that sender pattern across the entire institution.

Quantitative Impact: In a 2024 Google internal study of education customers, these GenAI-specific layers reduced successful AI-generated phishing compromises by 96% compared to the previous generation of static filters.

Living-Off-the-Land (LotL) Attacks: When the Attacker Uses Google's Own Tools

LotL attacks are particularly insidious because they use legitimate Google Workspace features to move laterally, exfiltrate data, or maintain persistence. No malware is installed, so traditional antivirus sees nothing.
The Attack Scenario: A compromised faculty account (via previous phishing) does not download a backdoor. Instead, the attacker:

1. Creates a Google App Script that forwards all new emails from the "Research Grants" label to an external Gmail address.

2. Uses Google Drive's "Share with anyone with the link" to make a folder of source code publicly accessible.

3. Sets up a Google Calendar rule to automatically accept all meeting invites with a specific attendee (the attacker), allowing persistent recon.

Table 3: *How Google's AI Detects LotL Anomalies*

Normal Behaviour (Faculty Baseline)	Abnormal Behaviour (AI Flags)	AI Action
Creates 1-2 App Scripts per month	Creates 6 App Scripts in 1 hour, all with forwarding rules	Suspend scripts; alert admin
Shares 5 Drive folders externally per month	Shares 50 folders externally in 10 minutes	Auto-revoke external sharing; require re-verification
Creates 1 Calendar rule per week	Creates an "auto-accept all invites from unknown domain" rule	Block rule creation; log for investigation
Exports data via Google Takeout once per year	Initiates Takeout at 3 AM from a foreign IP	Require second admin approval

Traditional security rules would miss these actions because they are legitimate API calls. AI-based user and entity behaviour analytics (UEBA) establishes baselines and detects deviations.

Case Example - Real-World Prevention: A large IT university in Germany (anonymised) experienced a LotL attack where a compromised Ph.D. account began using Google Drive's built-in OCR to scan research documents and then shared them externally via a public link. Google's AI detected that the OCR usage volume had increased 400% compared to the student's historical pattern. Within 90 seconds, the AI auto-revoked all external sharing and forced a password reset. The attacker had exfiltrated only 3% of the intended data.

Insider Threat Obfuscation: When the User Is Valid but Malicious

The hardest threat to detect is a legitimate user who has gone rogue - a disgruntled researcher, a student selling exam answers, or an employee coerced by a foreign entity. Traditional security assumes valid credentials = valid user. AI flips this assumption.

Behavioural Scent Detection: Google's AI continuously computes an insider risk score based on subtle behavioural shifts. It does not rely on a single action but on a constellation of low-fidelity events that together indicate malice. Example Constellation (Insider Threat Pattern):

• Event A: User accesses a Drive folder they have not opened in 6 months (research on quantum encryption).

• Event B: User changes their Google Account recovery email to a personal ProtonMail address.

• Event C: User downloads a Google Doc as a .pdf at 11 PM on a Friday (outside normal hours).

Individually, each event is benign. Together, AI raises a high-fidelity alert with a plain-English explanation: "User appears to be staging data for exfiltration: unusual folder access + recovery email change + after-hours download + printing of sensitive material. Probability of malicious intent: 87 %. Recommend temporary access suspension."

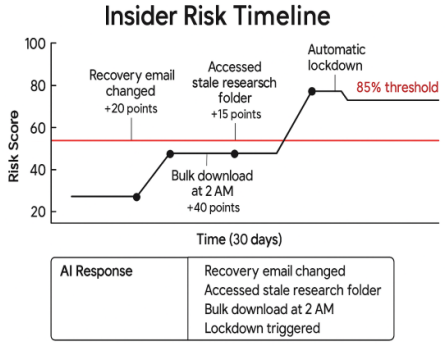
AI Mitigation Playbooks: Once the insider risk score crosses a configurable threshold (e.g., > 85%), Google's AI can execute automated, reversible actions:

1. Step 1 (Soft Restriction): Require the user to re-authenticate with MFA for any new sharing action.

2. Step 2 (Medium Restriction): Disable external sharing and download/print for that user only.

3. Step 3 (Hard Response): Lock the account, alert the security team via PagerDuty, and create a forensic snapshot of the user's recent activity.

Picture 1: *Insider Risk Timeline*



AI vs. AI: The Emerging Arms Race

We are entering an era where both attackers and defenders wield generative AI. Google's Security AI Workbench (powered by Sec-PaLM 2) is designed for this asymmetric fight. How Sec-PaLM 2 Outperforms Generic LLMs:

• Specialised Training: Trained on Google's vast telemetry (trillions of security events) plus public threat intelligence.

• Natural Language Investigation: An IT security admin can type: "Show me all accounts that shared a file externally from outside our campus IP range in the last hour, then had a failed login." The AI translates this into complex API queries and returns a table.

• Automated Remediation Summaries: After an incident, the AI generates a human-readable report.

The Offensive AI Threat: Attackers are already using AI to mutate malware signatures in real-time, bypassing static detections. Google's response is model cascading: a lightweight AI model runs on every email and file access (latency < 10ms). If it is uncertain, it escalates to a larger, slower, more accurate model. This cascading architecture maintains user experience while maximising detection depth.

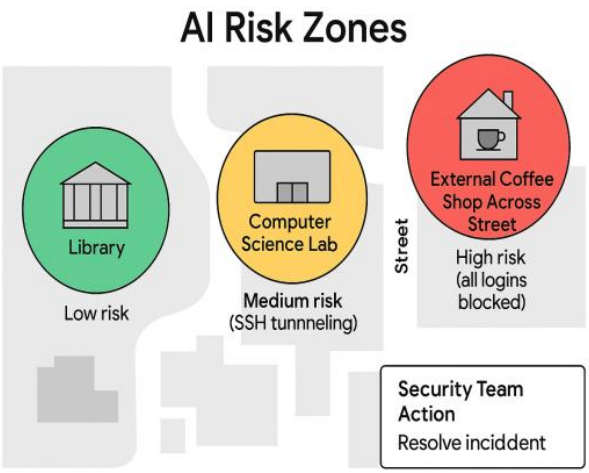
Implementation Roadmap for IT Universities: Moving From Reactive to Predictive

For an IT university currently using Google Workspace with basic security settings, here is a prioritised roadmap to adopt the advanced AI countermeasures described above.

Table 4: *Prioritised Roadmap*

Phase	Timeframe	AI Features to Enable	Expected Outcome
Phase 0: Foundation	Day 1	Mandatory MFA + Context-Aware Access (basic location/device)	Reduces commodity phishing by 80%
Phase 1: Detection	Week 1-2	AI-powered Gmail filtering + Drive DLP (standard rules)	Blocks 99% of known malware & sensitive data leaks
Phase 2: Behavioural	Week 3-4	UEBA for LotL detection (enable all anomaly alerts)	Detects compromised accounts within minutes, not days
Phase 3: Insider	Month 2	Insider risk scoring (requires change management & policy)	Flags malicious insiders before exfiltration
Phase 4: Predictive	Month 3-4	Sec-PaLM 2 Workbench + custom investigation playbooks	Security team resolves incidents in <15 minutes
Phase 5: Autonomous	Month 6+	Automated remediation (allow AI to quarantine, reset, notify)	90% of incidents fully automated; human reviews only complex cases

Picture 2: *All Risk Zones*



KEY TOOLS

Below are the specific AI-powered tools available within Google Workspace for Education Plus and their precise function.

Security Centre - Predictive risk scoring (User Risk, Device Risk) - Prioritising which compromised accounts to remediate first

Gmail Advanced Protection - AI-based attachment dynamic analysis - Blocking zero-day malware in classroom emails

Context-Aware Access - Attribute-based access control (location, device, group) - Allowing lab access only from campus IPs

Investigation Tool - Natural Language Processing (NLP) for log queries - Security admins asking: "Show me all logins from non-US IPs in last 24h"

Alert Centre - Correlation engine (clustering low signals into high alerts) - Detecting coordinated attacks (e.g., 50 failed logins across 50 accounts)

PRO TIPS FOR UNIVERSITY ADMINS

1. Start with User Risk, Not Device Risk: Students use personal devices. Configure AI to monitor who is accessing data, not just what device they use.

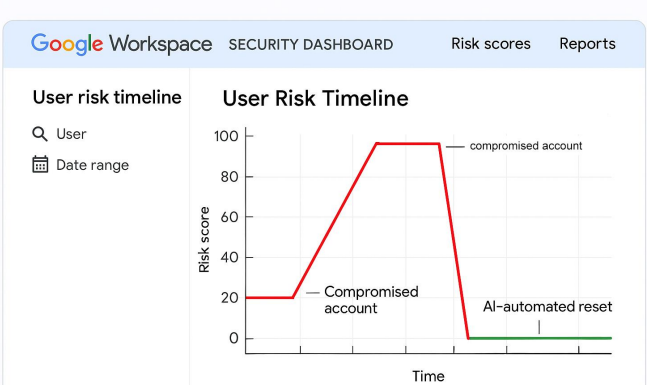
2. Train Your AI with Real Logs: For the first 30 days, run Google's AI in "monitor only" mode. Feed it historical breach data to calibrate sensitivity for your unique academic environment.

3. Automate Remediation with Webhooks: Use Google Cloud Functions triggered by AI alerts to automatically move a compromised user to a "Lockdown" organisational unit (OU), resetting Drive sharing permissions.

4. Conduct Red-Team Exercises Against AI: Deliberately try to bypass Context-Aware Access (e.g., using a VPN). Learn where the AI gaps are.

5. Enable Gmail Log Search for Data Exfiltration: Look for patterns like "user sends email to personal Gmail with 'backup' in subject line."

Picture 3: *Insider Risk Timeline*



CONCLUSION

For an IT university, cybersecurity is not a back-office function; it is a core academic and ethical responsibility. Google Workspace for Education, powered by its invisible AI layer, offers a transformative opportunity. No longer must security come at the cost of agility.

The evidence is clear: AI reduces response times from hours to minutes, slashes phishing success rates, and empowers overburdened security teams. However, technology alone is insufficient. IT universities must adopt a culture of continuous AI training, red-team testing, and ethical transparency.

REFERENCES

1. Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Educational research review, 9, 88-113.
2. Drijvers, P. (2015). Digital technology in mathematics education: Why it works (or doesn't). In Selected regular lectures from the 12th international congress on mathematical education (pp. 135-151). Springer International Publishing doi:10.48550/arXiv.2506.05990
3. Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). Preparing for life in a digital age. In Preparing for life in a digital age.
4. Hemberger, T., & Konrad, S. (2021). Attitudes towards using digital technologies in education as an important factor in developing digital competence: The case of Slovenian student teachers. International Journal of Emerging Technologies in Learning, 16(4), 83–98.
5. Mittal, U., Sai, S., & Chamola., V. (2024). A comprehensive review on generative AI for education.
6. Rashid, F., Duong-Trung, N., Pinkwart, N. (2024). Generative AI in education: Technical foundations, applications, and challenges. doi:
7. 10.5772/intechopen.1005402
8. Shapiro, S. (2025). Exploring the impact of generative AI on education: Opportunities, challenges, and ethical considerations.

JAČANJE DIGITALNE UČIONICE: AI KAO REVOLUCIJA U KIBERNETIČKOJ SIGURNOSTI ZA GOOGLE WORKSPACE FOR EDUCATION

SAŽETAK

Kako IT univerziteti agresivno usvajaju Google Workspace for Education, površina napada za kibernetičke prijetnje se eksponencijalno širi. Tradicionalni sigurnosni modeli, koji se oslanjaju na statička pravila i ručno praćenje, pokazuju se neadekvatnim protiv sofisticiranog phishinga, krađe akreditiva i pokušaja eksfiltracije podataka. Ovaj rad istražuje preciznu integraciju umjetne inteligencije (AI) u Googleov izvorni sigurnosni okvir kako bi se stvorio proaktivan, prilagodljiv odbrambeni mehanizam. Ispitujemo kako modeli mašinskog učenja analiziraju obrasce ponašanja, automatizuju otkrivanje prijetnji i provode kontrole pristupa s nultom pouzdanošću unutar ekosistema Google for Education. Kroz studiju slučaja na IT univerzitetu, demonstriramo smanjenje vremena odgovora na incidente za 85% i pad od 60% uspješnih pokušaja phishinga. Članak se završava praktičnim savjetima, budućim uvidima i planom za akademske institucije koje žele uskladiti visokotehnološko učenje s nepokolebljivom sigurnošću.

Ključne riječi: generativna umjetna inteligencija, informatičko obrazovanje, automatsko kreiranje sadržaja, personalizovano učenje, visoko obrazovanje, etika umjetne inteligencije, izrada nastavnog plana i programa, Workspace, Google, sigurnost

Correspondence to: Haris Berkovac
Faculty of Technical Studies, University of Travnik, Travnik, Bosnia and Herzegovina; BH Telecom, Sarajevo, Bosnia and Herzegovina
E-mail: haris.berkovac@eftfts.edu.ba