OSINT DIRECTIVE VECTOR C3-GAMMA: Digital Infrastructure, Social Media, and Algorithmic Monopolies

Executive Summary and Strategic Context

The global digital infrastructure ecosystem is currently undergoing a profound structural realignment. Historically, the moderation of digital speech and the enforcement of trust-and-safety guidelines occurred almost exclusively at the application layer, managed by the internal directorates of social media monopolies after a user posted content. However, an exhaustive analysis of the contemporary sociotechnical landscape indicates a rapid migration of content moderation, semantic filtering, and algorithmic policy enforcement away from the user-facing application layer and down into the foundational edge-computing, cloud-hosting, and transport layers of the internet.

This migration is operationalized through a framework recognized in academic and governance circles as "Layered Alignment". Under this paradigm, algorithmic monopolies are fundamentally altering their security architectures to intercept user queries, prompts, and data packets before they ever reach a generative artificial intelligence (AI) model or social media feed. By embedding moderation protocols at the edge-compute level, algorithmic monopolies are rendering censorship functionally invisible to the end-user, preempting any application-layer appeals process or public visibility.

This report executes OSINT Directive VECTOR C3-GAMMA, providing an exhaustive mapping of the executive leadership, Ultimate Beneficial Owners (UBOs), and Trust and Safety (T&S) directorates across the world's dominant social media platforms (Meta, Alphabet/Google, X, TikTok) and Tier-1 Content Delivery Networks/Cloud Hosts (Cloudflare, AWS, Akamai). Crucially, this analysis identifies the precise mechanisms and human capital bridges through which external, network-funded censorship incubators—specifically the Adir Challenge Foundation, CyberWell, the USC Shoah Foundation, and Google Jigsaw—inject their proprietary semantic matrices, regular expression (regex) dictionaries, and ideological taxonomies directly into global digital infrastructure.

The Architecture of Algorithmic Monopolies: UBOs and Trust and Safety Directorates

To comprehend the vector pathways of external policy ingestion, it is necessary to map the governance structures of the primary digital monopolies. A persistent theme across these entities is the intentional bifurcation between Ultimate Beneficial Ownership (UBO) and the Trust and Safety directorates. UBOs are structurally insulated from public and regulatory accountability via dual-class share structures or private ownership, while T&S executives serve as high-attrition liaisons who interface with external non-governmental organizations (NGOs), state actors, and policy incubators.

Meta Platforms, Inc.

Meta operates under a dual-class capital structure that grants its founder overwhelming voting control, effectively insulating the UBO from activist shareholder resolutions regarding algorithmic transparency or human rights frameworks. The enforcement of policy is delegated to a deeply entrenched bureaucracy of policy executives who manage the localized and global ramifications of the platform's algorithms.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
Ultimate Beneficial Owner (UBO)	Mark Zuckerberg	Holds outsized voting power via Class B shares, retaining absolute control over corporate governance and baseline algorithmic priorities.
President, Global Affairs	Nick Clegg	Former UK Deputy Prime Minister. Functions as the primary geopolitical liaison, managing international regulatory pressures and the company's macro-political positioning, bridging the gap between state actors and Meta's internal policy engine.
Head of Content Policy	Monika Bickert	A former federal prosecutor who authored and enforced Facebook's content guidelines for over a decade. Her departure in August 2026 for Harvard Law School marks a significant transition in Meta's legacy moderation architecture, shifting the paradigm from manual policy writing to algorithmic enforcement.
Global Policy Team Lead	Kevin Martin	Transitioning to absorb Bickert's portfolio, overseeing global policy operations and acting as the new nexus for external NGO partnerships and systemic alignment.
Head of Trust and Safety, Africa	Adrien Diarra	Ex-Goldman Sachs executive overseeing the African continent, an increasingly critical vector for managing emerging market moderation, secure tech adoption, and the localization of global alignment

Executive Role / Title	Individual	Operational Significance and Alignment Impact
		policies.

Alphabet Inc. (Google / Jigsaw)

Alphabet's infrastructure spans search, video content delivery (YouTube), and generative AI foundation models (Gemini). Its moderation logic is heavily influenced by its internal technology incubator, Jigsaw, which acts as a specialized bridge between academic research, NGO policy demands, and Google's core engineering teams.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
UBOs / Controlling Shareholders	Larry Page, Sergey Brin, Sundar Pichai	Institutional control via multi-class stock structures. Pichai serves as the active CEO directing the integration of AI safety protocols across the entire Google tech stack.
President, Global Affairs	Kent Walker	Oversees Alphabet's public policy, legal, and trust-and-safety apparatuses on a global scale, serving as a frequent participant in international digital rights summits like RightsCon.
Director of Trust & Safety	Snigdha Bhardwaj	A key figure in Google's generative AI safety strategy, emphasizing proactive risk mitigation and independent research funding via the Tech Coalition.
Managing Director, Trust & Safety	Amanda Storey	Manages global alliances, notably integrating Google into the Global Anti-Scam Alliance (GASA), reflecting a unified industry approach to semantic threat mitigation.
CEO, Google Jigsaw	Yasmin Green	Directs Google's specialized unit for countering online threats. Green is a critical human capital bridge to external networks, serving explicitly as an advisor and judge for the network-funded Adir Challenge Foundation.

X Corp. (formerly Twitter)

Following its acquisition in 2022, X Corp. underwent a radical deconstruction of its legacy Trust

and Safety architecture. The dissolution of the Trust and Safety Council and the high turnover of executives reflect a chaotic realignment of moderation priorities under a single, highly visible UBO who fundamentally opposes traditional moderation terminology.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
UBO / Owner	Elon Musk	Exercises total private control over X Corp. Has actively dismantled previous moderation frameworks, explicitly stating that traditional "Trust" models are euphemisms for censorship and terminating legacy engineering teams responsible for algorithmic filtering.
Head of Safety	Kylie McRoberts	Appointed in April 2024 after a nine-month vacancy. Tasked with managing global safety teams and the newly established Austin Safety Center of Excellence under Musk's unpredictable and highly publicized policy directives.
Head of Brand Safety & Advertiser Solutions	Yale Cohen	Former Publicis Media executive brought in to repair severely damaged relationships with advertisers by guaranteeing brand-safe adjacencies, highlighting the tension between the UBO's free speech absolutism and commercial viability.
Legacy T&S Leadership	Yoel Roth, Ella Irwin	Former heads who resigned under intense pressure, illustrating the severe friction between algorithmic policy experts, NGO partnerships, and absolute ownership control.

TikTok (ByteDance & TikTok USDS)

TikTok's structural reality has been fundamentally altered by Western national security mandates. The regulatory environment has forced the creation of a uniquely siloed entity for United States operations, overseen by a consortium of corporate partners rather than solely by its Chinese parent company. This ensures that algorithmic alignment and content moderation for American users are conducted on domestic soil.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
UBO Structure	ByteDance (Parent) / Oracle, Silver Lake, MGX	ByteDance retains a 19.9% stake in the newly formed TikTok USDS Joint Venture, while American managing investors (Oracle, Silver Lake, etc.) hold the balance to satisfy US data security and algorithmic separation laws.
CEO, TikTok USDS	Adam Presser	Formerly TikTok's Head of Operations and Trust & Safety. Elevated to CEO of the USDS joint venture, overseeing the retraining of the platform's recommendation algorithm on US user data inside Oracle's cloud infrastructure.
Director, TikTok USDS	Shou Chew	TikTok's global CEO, maintaining a board seat on the USDS entity to ensure interoperability between the US and global platforms, balancing localized alignment with global codebases.

The Edge-Compute Migration: Enforcing Moderation at the Transport Layer

The most critical operational shift identified in this analysis is the migration of moderation logic from the user interface down to the network infrastructure. Tier-1 Content Delivery Networks (CDNs) and Cloud Hosts—which manage Domain Name Systems (DNS), Distributed Denial-of-Service (DDoS) protection, Application Programming Interface (API) gateways, and Web Application Firewalls (WAFs)—are deploying AI-specific security tools designed to intercept and parse text before it can be processed by foundation models.
By enforcing policy at the network edge, these providers effectively become the ultimate arbiters of algorithmic alignment, capable of dropping HTTP requests based on semantic pattern matching or regular expressions (regex) without generating a user-facing appeal mechanism.

Amazon Web Services (AWS)

As the dominant global cloud infrastructure provider, AWS has systematized AI moderation directly into its foundation model hosting environment through a suite of tools designed to govern prompt inputs and model outputs simultaneously.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
Head of AI/Generative AI	Sasha Rubel	Directs generative AI policy

Executive Role / Title	Individual	Operational Significance and Alignment Impact
Policy, EMEA		frameworks across Europe, the Middle East, and Africa, aligning AWS infrastructure with emerging sovereign cloud regulations and AI safety standards.
Head of Financial Services Public Policy	Michael Jefferson	Liaises with state regulators (e.g., the Bank of England) to ensure AWS operational resilience and AI integration meet strict regulatory and compliance mandates.
Managing Director, AWS European Sovereign Cloud	Stéphane Israël	Leads the geographically and logically separated European Sovereign Cloud, ensuring data residency and localized governance under EU-resident personnel.

AWS operationalizes semantic control through "Amazon Bedrock Guardrails." This system allows developers to apply configurable safety policies across any foundation model, including third-party models like OpenAI or Gemini that are hosted on AWS or accessed via the ApplyGuardrail API. Bedrock Guardrails allow administrators to upload custom regex patterns, exact word filters, and predefined "denied topics." Once integrated into the AWS API gateway or an MCP (Model Context Protocol) server, the guardrail intercepts the user's prompt.
If the prompt contains semantic vectors matching the denied topic or regex string, AWS rejects the call entirely, blocks the interaction, or masks the output with identifier tags, preventing the foundation model from ever fully processing the query. The system also utilizes "Contextual Grounding" and "Automated Reasoning" checks to filter hallucinations based on mathematical validations.

Cloudflare

Cloudflare routes a massive percentage of global web traffic and serves as the primary security gateway for countless enterprise applications. The company has explicitly shifted its security posture to accommodate the unique vulnerabilities of Large Language Models.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
CEO & Co-Chair	Matthew Prince	Co-founder driving Cloudflare's expansion from DDoS protection into comprehensive AI security and network edge computing.
President & Co-Chair	Michelle Zatlyn	Co-founder managing global operations and corporate strategy, sitting simultaneously

Executive Role / Title	Individual	Operational Significance and Alignment Impact
		on the board of Atlassian.
Chief Legal Officer	Alissa Starzak	Former Global Head of Public Policy with a background as Deputy General Counsel for the US Department of Defense and counsel to the Senate Select Committee on Intelligence. Oversees legal frameworks for global data flows.

Cloudflare's primary architectural mechanism for alignment is the "Firewall for AI," an advanced WAF tailored for LLM applications. This firewall analyzes prompts submitted by end-users to identify jailbreaks, prompt injections, and potential data exfiltration. Traditional security relied heavily on regular expressions (regexes) to detect sensitive data; however, regexes are brittle when applied to natural language processing because they require constant updates and struggle with implicit context.

To solve this, Cloudflare integrated Named Entity Recognition (NER) models using the open-source Presidio framework to analyze text dynamically. This system sits in front of the model, processing queries in real-time. If an external NGO provides a specific lexicon of coded hate speech, it can be converted into a regex matrix or training data for the NER model and applied directly at the Cloudflare edge, rendering prohibited concepts entirely un-computable by the downstream AI.

Akamai Technologies

Akamai, a pioneer in edge computing and CDN services, has similarly pivoted to securing the API endpoints that facilitate generative AI integration, ensuring that the logic of content moderation is handled before inference occurs.

Executive Role / Title	Individual	Operational Significance and Alignment Impact
VP of Global Public Policy	Lauren Van Wazer	Directs Akamai's public policy strategy, navigating the intersection of cybersecurity, internet infrastructure resilience, and government regulation.
Director of Security Research	Kimberly Gomez	Leads research teams analyzing automated bot traffic, AI-driven fraud, and API vulnerabilities, translating threat intelligence into actionable WAF rules.
VP & Chief Privacy Officer	James A. Casey	Navigates the global AI compliance landscape, designing risk-based governance models that satisfy

Executive Role / Title	Individual	Operational Significance and Alignment Impact
		emerging regulations without sacrificing compute speed.

Akamai's "Firewall for AI" provides multilayered protection against adversarial inputs, unauthorized queries, and large-scale data scraping. By integrating tightly with Kubernetes clusters (Akamai Linode Kubernetes Engine) and API gateways like Kong, Akamai ensures that security policies are strictly enforced at the network edge. This architecture inspects the "intent" behind every request, allowing for real-time moderation of toxic output, topic restriction, and the prevention of AI-specific Denial of Service attacks before the compute layer is engaged.

The Network-Funded Censorship Incubators: Semantic Matrix Providers

Edge computing providers (AWS, Cloudflare, Akamai) possess the infrastructure to filter content, but they generally do not internally generate the sociological definitions of what constitutes toxic, harmful, or policy-violating speech. Instead, they rely on external "Trusted Partners"—specialized NGOs, academic laboratories, and think tanks that incubate policy, define taxonomies, and curate the highly specific lexicons and regex matrices required to feed the machine learning models at the edge. This establishes a symbiotic loop: NGOs define the ideological boundaries of acceptable speech, and the tech monopolies implement these boundaries at the infrastructure level.

CyberWell

CyberWell is an Israeli-based NGO that acts as a highly specialized, tech-forward censorship incubator. Launched in May 2022, it operates the first open, live database of online antisemitic content and actively translates social media posts into actionable intelligence for global platforms.

The organization is directed by Tal-Or Cohen Montemayor, a former IDF lone soldier with a background in military intelligence and open-source web intelligence, who holds a degree in Government and Law from Reichman University. CyberWell functions officially as a "Trusted Partner" and "Trusted Flagger" for Meta (Facebook, Instagram, Threads), TikTok, X, and YouTube.

CyberWell continuously monitors content in English, Arabic, and Farsi, rigorously mapping it against the International Holocaust Remembrance Alliance (IHRA) working definition of antisemitism. Following the geopolitical events of October 2023, CyberWell utilized its platform to document a massive surge in online hate, specifically tracking how traditional stereotypes mutated into direct incitement in Arabic-speaking digital spaces.

The organization's true infrastructural power lies in its "Regex Pipeline." By collecting over 11,500 verified abusive posts into an open database, CyberWell curates highly specific dictionaries, lexicons, and semantic indicators of evolving hate speech—including coded emojis, neologisms, and algorithm-evading language. These lexicons are not merely academic; they form the foundational datasets required to build the NER models and regex matrices used by systems like AWS Guardrails and Cloudflare WAFs, directly translating sociological definitions into machine-readable network security rules.

The ADIR Challenge Foundation & AddressHate

The ADIR Challenge Foundation was launched in early 2024 by Morielle Lotan and Dr. Shay Hershkovitz (a former head of research at XPrize) to crowdsource technological solutions to antisemitism through a high-stakes, $1 million prize competition. The initiative gamifies censorship, drawing talent from high schools, universities, and the tech sector to develop scalable, machine-based solutions to online hate.

The foundation's initial ideation phase awarded prizes to projects that push the boundaries of automated moderation. For example, the Israel NLP Team won for proposing the use of generative AI to automatically post fact-based responses to hateful content at scale, while other winners focused on real-time detection algorithms and Natural Language Processing (NLP) tools that empower users to take legal action against individuals spreading hate.

Deeply intertwined with the ADIR Challenge is the **AddressHate** initiative. Backed by a $1 million commitment from the Laterman Family Foundation (directed by Joshua Laterman), AddressHate focuses strictly on technology-first approaches to combat digital hate, operating under the philosophy that hate operates "like a code that can be broken" via cyber and algorithmic interventions.

AddressHate serves as a massive funding mechanism for academic and infrastructure integration. It has channeled hundreds of thousands of dollars into the Antisemitism Studies Laboratory at Baruch College (CUNY), directed by Dr. Sarah Valente within the Sandra Kahn Wasserman Jewish Studies Center. This laboratory acts as the academic engine for discourse analysis and social media studies, translating linguistic research into actionable intelligence for groups like ADIR. AddressHate also partners directly with the ADIR Challenge to sponsor interdisciplinary college teams building automated tech solutions.

The USC Shoah Foundation

Historically known for preserving the video testimonies of Holocaust survivors under the visionary guidance of Steven Spielberg, the USC Shoah Foundation has become deeply integrated into the modern digital moderation space under the leadership of Executive Director Dr. Robert J. Williams.

The foundation recently converged with the tech-policy sector by co-hosting the "Digital Frontlines" event in Washington, D.C., alongside the ADIR Challenge. This summit focused explicitly on deploying AI technology, digital preservation, and the development of new algorithmic monitoring tools to counter digital antisemitism and misinformation.

The Shoah Foundation provides a vital ontological contribution to the moderation pipeline: its proprietary "Thesaurus." This database contains over 63,000 highly specific indexing terms, concepts, and geographical/experiential tags used to meticulously categorize trauma, discrimination, and historical hate. In the context of AI training and NLP moderation, this thesaurus serves as a goldmine for training data. By integrating the Shoah Foundation’s deeply structured linguistic hierarchies with the Adir Challenge’s tech-forward methodologies and AddressHate's financial backing, the network creates highly sophisticated vocabularies that teach LLMs how to recognize the semantic architecture of hate speech across diverse contexts.

Project Layered Alignment: The Technical and Theoretical Pipeline

The ingestion of these external semantic matrices into global digital infrastructure is theoretically formalized through the concept of **"Layered Alignment."**
Coined and extensively researched by legal and AI scholars such as Spencer Williams (University of New Hampshire Law Review), the theory of "Layered Alignment" posits that the "alignment problem"—the existential challenge of ensuring an AI behaves according to human values—cannot be solved by a single technical mechanism such as a system prompt or fine-tuning. Instead, it requires a multifaceted, "layered" approach analogous to corporate governance. Just as corporations use a mix of contracts, board oversight, and fiduciary duties to control agents, AI systems require overlapping layers of architectural constraints, external oversight, and dynamic filtering to remain aligned.
This theory is rapidly expanding beyond social media into cyber-physical and hardware domains. For instance, researchers are applying fiduciary duties to "Brain Foundation Models" connected to Brain-Computer Interfaces (BCIs), ensuring that neural data interpretation is strictly aligned with user safety through layered technical design. Similarly, in the automotive sector, the Mediator-in-the-Loop-Driving (MILD) system utilizes layered alignment to ensure an AI strategy agent generates vehicle commands that strictly adhere to human values and safety constraints.
In the context of digital discourse, the Meaning Alignment Institute highlights the distinction between "model integrity" and mere "compliance." They argue that current LLMs are either being forced into rigid, brittle compliance via massive rulebooks (like ChatGPT) or programmed with vague values (like Claude) that can be easily bypassed. True alignment requires structured, inspectable values implemented at the architectural level.

Technical Execution at the Edge

In practice, Layered Alignment provides the exact blueprint for what is occurring between the NGO incubators and the cloud monopolies:

1. **The Base Layer (The Model):** The raw LLM (e.g., LLaMA, Gemini, Claude) which processes natural language.
2. **The Application Layer (Social Media):** The user-facing platform (X, TikTok), which enforces broad terms of service but struggles to catch nuanced, evolving abuse at scale.
3. **The Infrastructure Layer (The Modulator):** This is the crucial insertion point. Systems like Cloudflare's Firewall for AI or AWS Bedrock Guardrails sit *between* the user and the model.
4. **The External Policy Layer (The NGOs):** Organizations like CyberWell and the ADIR network generate the raw intelligence—the specific regex expressions, toxicity indicators, and coded hate symbols.

By adhering to a Layered Alignment philosophy, platforms bypass the inherent limitations of trusting the AI to moderate itself. Instead, they implement strict, mathematically rigid edge-routing based on the exact lexicons provided by Trusted Partners.

Human Capital Bridges and Semantic Ingestion Vectors

The transfer of semantic matrices and algorithmic protocols from external NGOs into the core infrastructure of Alphabet, Meta, and Cloudflare does not occur in a vacuum; it requires specific "human capital bridges." These individuals inhabit dual roles, moving fluidly between the

corporate Trust and Safety apparatus and the non-profit incubation network, ensuring that the theoretical models of Layered Alignment become operational reality.

The Role of Google Jigsaw and Perspective API

Google Jigsaw is Alphabet's specialized technology incubator focused on countering global security challenges. Its flagship contribution to content moderation has been the **Perspective API**, a machine-learning model that scores the "toxicity" of text strings based on millions of human-labeled comments. Perspective API became the industry standard for automated moderation, allowing platforms to automatically flag or hide comments based on a probabilistic toxicity matrix, supporting multiple languages and fine-grained attributes like "severe toxicity" and "identity attack".
However, Alphabet has announced the sunsetting of the Perspective API by the end of 2026, stating that AI capabilities have evolved beyond the need for a standalone tool. This deprecation strongly implies that external, API-based moderation is being rendered obsolete because moderation is moving *natively* into the edge compute layers (e.g., Cloudflare WAF, AWS Bedrock Guardrails). The logic is migrating from an external API call into the core compute fabric itself.

Raquel Saxe Gelb: The Primary Vector

A critical human capital bridge identified in this OSINT directive is **Raquel Saxe** (also known as Raquel Saxe Gelb).

- **Corporate Role:** Saxe serves as Chief of Staff and Senior Program Manager at Google Jigsaw, placing her at the nerve center of Alphabet's toxicity modeling and algorithmic safety initiatives.
- **NGO Integration:** Saxe serves as a direct liaison to the ADIR Challenge Foundation. At the "Digital Frontlines" event co-hosted by ADIR and the USC Shoah Foundation, Saxe was a featured participant, representing Google Jigsaw's moderation capabilities and signaling the integration of corporate tech with NGO methodologies.
- **Geopolitical Background:** Saxe's background includes extensive work as a research associate at The Washington Institute for Near East Policy, placing her at the exact intersection of geopolitical policy formulation, AI safety engineering, and algorithmic moderation.

Through figures like Raquel Saxe and Yasmin Green (CEO of Google Jigsaw, who acts as a prominent judge and advisor for the ADIR Challenge), Google’s internal methodologies for semantic parsing and toxicity scoring are directly shared with external incubators. Simultaneously, the specific lexicons and AI-driven solutions generated by groups like CyberWell and ADIR are validated by Big Tech executives, creating a continuous feedback loop that institutionalizes the NGOs' ideological definitions within global codebases.

Strategic Conclusions

The analysis of VECTOR C3-GAMMA reveals a highly sophisticated consolidation of digital power and the realization of a new paradigm in internet governance. Global tech monopolies are steadily moving away from reactive, user-facing content moderation handled by human review teams. Instead, they are transitioning toward proactive, automated censorship executed

at the deepest infrastructure levels of the internet.
This transition from the application layer to the transport layer relies fundamentally on the principles of Project Layered Alignment. Because Edge and CDN providers like AWS, Cloudflare, and Akamai lack the geopolitical mandate and the sociological expertise to independently define what constitutes "hate speech" or "toxicity," they outsource the creation of these definitions to a network of specialized incubators.
Entities like CyberWell, the Adir Challenge Foundation, AddressHate, and the USC Shoah Foundation function as specialized intelligence-gathering apparatuses. They parse global digital discourse to create massive, machine-readable dictionaries, ontological matrices, and regex strings. Human capital bridges, exemplified by Google Jigsaw's Raquel Saxe and Yasmin Green, ensure that these NGO-generated taxonomies are seamlessly ingested into corporate Trust and Safety pipelines. The result is an opaque, highly efficient architecture of algorithmic control, wherein the definitions of permissible speech are drafted by private, network-funded entities and enforced silently by the foundational hardware of the internet, fundamentally altering the nature of digital free expression.

Works cited

1. "Layered Alignment" by Spencer Williams - UNH Scholars Repository, https://scholars.unh.edu/unh_lr/vol23/iss2/7/ 2. Layered Alignment - UNH Scholars Repository, https://scholars.unh.edu/cgi/viewcontent.cgi?article=1509&context=unh_lr 3. meta-20260416 - SEC.gov, https://www.sec.gov/Archives/edgar/data/1326801/000162828026025532/meta-20260416.htm 4. What is a beneficial owner: Global compliance requirements for 2025 - Diligent, https://www.diligent.com/resources/blog/beneficial-ownership 5. Generative AI Data Governance – Amazon Bedrock Guardrails - AWS, https://aws.amazon.com/bedrock/guardrails/ 6. Amazon Bedrock policies - AWS Organizations, https://docs.aws.amazon.com/organizations/latest/userguide/orgs_manage_policies_bedrock.html 7. AWS Bedrock Guardrails for MCP: Block PII from MCP Manager, https://mcpmanager.ai/blog/aws-bedrock-guardrails-for-mcp/ 8. Partner POV | The 2026 AI Frontier: Securing the New Enterprise Architecture - A Collaborative Perspective from WWT & Cloudflare, https://www.wwt.com/article/partner-pov-or-the-2026-ai-frontier-securing-the-new-enterprise-architecture-a-collaborative-perspective-from-wwt-and-cloudflare 9. AI Security for Apps · Cloudflare Web Application Firewall (WAF) docs, https://developers.cloudflare.com/waf/detections/ai-security-for-apps/ 10. Cloudflare announces Firewall for AI, https://blog.cloudflare.com/firewall-for-ai/ 11. Take control of public AI application security with Cloudflare's Firewall for AI, https://blog.cloudflare.com/take-control-of-public-ai-application-security-with-cloudflare-firewall-for-ai/ 12. Protection and Security for AI and LLM Applications - Akamai, https://www.akamai.com/products/firewall-for-ai 13. Akamai launches firewall to protect enterprise AI from cyber threats - SecurityBrief UK, https://securitybrief.co.uk/story/akamai-launches-firewall-to-protect-enterprise-ai-from-cyber-threats 14. Secure, Scalable AI Infrastructure with Kong + Akamai: A Technical Intro, https://konghq.com/blog/engineering/ai-infrastructure-with-kong-and-akamai 15. CyberWell - Jgive, https://www.jgive.com/new/en/ils/charity-organizations/14070 16. CyberWell - Wikipedia, https://en.wikipedia.org/wiki/CyberWell 17. CyberWell: Home, https://cyberwell.org/ 18. CyberWell – Using data to reduce online antisemitism | The Jerusalem Post,

https://www.jpost.com/business-and-innovation/tech-and-start-ups/article-726808 19. Tal-Or Cohen Montemayor - Eradicate Hate Global Summit, https://eradicatehatesummit.org/participant/tal-or-montemayor/ 20. Adv. Tal-Or Cohen Montemayor - ICT - International Institute for Counter-Terrorism, https://ict.org.il/ourteam/worldsummit2024-speakers-adv-tal-or-cohen-montemayor/ 21. "What happened on 7/10 on social media was the largest hijacking of our major applica | Ctech, https://www.calcalistech.com/ctechnews/article/byctjmo1r 22. Arabic and English Antisemitism on Social Media Platforms Post-October 7: Analysis on Discourse and Prevention | INSS, https://www.inss.org.il/publication/antisemitism-on-social-media/ 23. Rise of Online Antisemitism in Arabic Six Months Post October 7, https://extremism.gwu.edu/sites/g/files/zaxdzs5746/files/2025-01/Rise%20of%20Online%20Antisemitism%20in%20Arabic%20Six%20Months%20Post%20October%207_compressed.pdf 24. AJC, CyberWell Issue Recommendations to Confront Online Antisemitism - Baltimore Jewish Times, https://www.jewishtimes.com/ajc-cyberwell-issue-recommendations-to-confront-online-antisemitism/ 25. CyberWell: Fighting Digital Antisemitism in Real Time - Jewish Press - Tampa Edition, https://www.jewishpresstampa.com/articles/cyberwell-fighting-digital-antisemitism-in-real-time/ 26. Combating antisemitism with innovative high tech | The Jerusalem Post, https://www.jpost.com/business-and-innovation/article-796163 27. Foundation in memory of Nova victim aims to fight antisemitism - Ynet News, https://www.ynetnews.com/jewish-world/article/by7yblipp 28. A new competition will grant a $1 million prize for technology to fight antisemitism, https://www.jta.org/2024/03/06/united-states/a-new-competition-will-grant-a-1-million-prize-for-technology-to-fight-antisemitism 29. Australians invited to enter global ADIR Challenge, https://www.australianjewishnews.com/australians-invited-to-enter-global-adir-challenge/ 30. Adir Challenge names 3 winners for first phase of $1 million competition to tackle antisemitism - eJewish Philanthropy, https://ejewishphilanthropy.com/tech-wizards-vie-for-1-million-prize-to-tackle-antisemitism/ 31. Home — AddressHate, https://www.addresshate.org/home2 32. About — AddressHate, https://www.addresshate.org/about 33. Laterman Family Foundation Incorporated - Grant Information & Application Details, https://www.hinchilla.com/funders-us/921426465-laterman-family-foundation-incorporated 34. Baruch College Unveils Antisemitism Studies Laboratory, Backed by the Laterman Family Foundation's “AddressHate” Initiative - News Center, https://newscenter.baruch.cuny.edu/news/baruch-college-unveils-antisemitism-studies-laboratory-backed-by-the-laterman-family-foundations-address-hate-initiative/ 35. AddressHate's Early Investment Catalyzes Nearly $500,000 in Funding for Baruch College's Antisemitism Studies Lab - News Center, https://newscenter.baruch.cuny.edu/news/addresshates-early-investment-catalyzes-nearly-500000-in-funding-for-baruch-colleges-antisemitism-studies-lab/ 36. The Shape of Testimony: A Scalable Framework for Oral History Archive Comparison - arXiv, https://arxiv.org/html/2605.21623v1 37. Dimensions in Testimony - USC Shoah Foundation, https://sfi.usc.edu/dit 38. USC Shoah Foundation Institute for Visual History and Education, https://www.ahoinfo.org/component/k2/69-usc-shoah-foundation-the-institute-for-visual-history-and-education 39. Digital Frontlines: The Challenges and Solutions to Fighting Online Antisemitism - YouTube, https://www.youtube.com/watch?v=_R4EnYFTHxA 40. The ADIR Foundation Together with the USC Shoah Foundation, https://www.prnewswire.com/news-releases/the-adir-foundation-together-with-the-usc-shoah-fou

ndation-held-a-one-day-event---digital-frontlines-the-challenges-and-solutions-to-fighting-online-antisemitism-302358175.html 41. USC Shoah Foundation Thesaurus (Abridged), https://sfi.usc.edu/sites/default/files/docfiles/USC_SF_Thesaurus_101212_0.pdf 42. Holocaust Memory in the Digital Age: Chapter 1 | Stanford University Press, https://www.sup.org/books/jewish-studies/holocaust-memory-digital-age/excerpt/chapter-1 43. Over 800 New Indexing Terms to be Added to Visual History Archive Thesaurus - USC Shoah Foundation - University of Southern California, https://sfi.usc.edu/news/2016/08/12128-over-800-new-indexing-terms-be-added-visual-history-archive-thesaurus 44. Spencer Williams - California Western School of Law, https://www.cwsl.edu/directories/faculty-staff-directory/spencer_williams.html 45. Brain Uses AI-Like Computations for Language - Neuroscience News, https://neurosciencenews.com/ai-language-brain-30025/ 46. Fiduciary AI for the Future of Brain-Technology Interactions - arXiv, https://arxiv.org/pdf/2507.14339 47. Fiduciary AI for the Future of Brain-Technology Interactions - ResearchGate, https://www.researchgate.net/publication/393889503_Fiduciary_AI_for_the_Future_of_Brain-Technology_Interactions 48. (PDF) MILD: Mediator Agent System with Bidirectional Perception and Multi-Layered Alignment for Human-Vehicle Collaboration - ResearchGate, https://www.researchgate.net/publication/404426856_MILD_Mediator_Agent_System_with_Bidirectional_Perception_and_Multi-Layered_Alignment_for_Human-Vehicle_Collaboration 49. Model Integrity - LessWrong, https://www.lesswrong.com/posts/mJKaMvLMCWAZRyiCz/model-integrity 50. The fabrics of machine moderation: Studying the technical, normative, and organizational structure of Perspective API - ResearchGate, https://www.researchgate.net/publication/355354803_The_fabrics_of_machine_moderation_Studying_the_technical_normative_and_organizational_structure_of_Perspective_API 51. perspective API documentation, https://perspective-dev.github.io/python/ 52. Get started with Perspective API | Google for Developers, https://developers.google.com/codelabs/setup-perspective-api 53. Perspective API: complete guide to finding an alternative - Lasso Moderation, https://www.lassomoderation.com/blog/perspective-api/ 54. Combating Toxic Language: A Review of LLM-Based Strategies for Software Engineering - arXiv, https://arxiv.org/html/2504.15439v2 55. Perspective API, https://perspectiveapi.com/ 56. Your Daily Phil: UJA-Fed. of N.Y. taps day school head Michael Kay as CEO, https://ejewishphilanthropy.com/your-daily-phil-uja-fed-of-n-y-taps-day-school-head-michael-kay-as-ceo/ 57. Raquel Saxe | The Washington Institute, https://www.washingtoninstitute.org/experts/raquel-saxe 58. BDS's Useless Politics of Confrontation | The Washington Institute, https://www.washingtoninstitute.org/policy-analysis/bdss-useless-politics-confrontation