SPECIAL COMPETITIVE STUDIES PROJECT

SEPTEMBER 2025

The Digital Case Officer

Reimagining Espionage with Artificial Intelligence

The Digital Case Officer
Reimagining Espionage with Artificial Intelligence

Executive Summary

We stand at the threshold of the most profound, technology-driven transformation in intelligence operations since the advent of the internet. The rapid development of artificial intelligence (AI), virtual- and augmented-reality technologies, robotics and autonomous systems, and large-scale data analysis capabilities present *both* an existential challenge to traditional human intelligence operations *and* a once-in-a-generation opportunity to revolutionize the business of spycraft.

The emergence and increased sophistication of generative, multimodal, and agentic AI is quietly, but quickly reshaping how organizations operate, and the intelligence community is no exception. The rapid drop in the cost of deploying powerful models, combined with the rise of smaller, domain-specific AI, is giving more organizations and a wider range of devices access to AI than ever before. This trend fundamentally alters the collection environment. Multimodal AI, which can process and generate text, images, video, and audio, opens new opportunities and creative frontiers, vastly expanding the aperture of signals available for collection and exploitation. The People's Republic of China (PRC) and other adversaries are already moving to deploy these technologies at scale, making classic methods of tradecraft untenably risky and, in some cases, nearly impossible to safely execute. The proliferation of "Ubiquitous Technical Surveillance (UTS)"—the vast web of digital monitoring from phones, cameras, and online activity—is systematically erasing the anonymity essential for traditional espionage.

Despite this, the value of human intelligence remains critical. Technology is adept at monitoring what adversaries *do*, but only human sources can reliably access their *intent*—the ultimate intelligence prize. Human intelligence remains indispensable for solving the "last mile" problem, such as penetrating "air-gapped" networks or accessing sensitive leadership deliberations that are shielded from technical collection. Furthermore, espionage is fundamentally about interpersonal contact, requiring uniquely human traits like empathy, intuition, and the ability to build trust—qualities that algorithms cannot replicate.

To operate in this new landscape, the U.S. Intelligence Community (IC) must master the very technologies that threaten it. Advanced AI, particularly agentic systems capable of autonomous action, can now augment every phase of the human intelligence recruitment cycle with a scale and precision unattainable by humans alone. These "Digital Case Officers" can:

- **Target:** Synthesize vast datasets to identify and prioritize potential intelligence assets based on their access, motivation, and vulnerability.
- **Assess & Develop:** Build detailed psychological profiles from digital footprints and engage targets in tailored, long-term conversations to build rapport and trust, using

hyper-realistic personas. AI can manage hundreds of such developmental conversations simultaneously—a task impossible for a human officer.

- **Recruit & Handle:** Deliver personalized recruitment pitches by referencing a target's specific grievances or motivations and provide real-time operational security advice to assets once recruited.

The objective is not to replace human officers but to empower them. The future of human intelligence lies in the **human-machine team**, where AI handles the immense scale of data processing and initial outreach, freeing case officers to focus on high-value work such as making nuanced judgments, managing the psychology of the asset-case officer relationship, and overseeing high-stakes operations. Artificial intelligence can also help human collectors achieve unprecedented scale by spotting greater numbers of targets, simultaneously developing their profiles, and conducting virtual approaches.

The power of AI in human intelligence also gives rise to important legal and ethical implications that demand a robust governance framework. The use of autonomous systems creates a complex challenge landscape, including:

- **Legal Ambiguity:** The use of commercial models trained on large data sets may raise issues concerning data provenance and validation, and issues of privacy if those data sets include a high volume, proportion, or sensitive U.S. person information. Training or otherwise enhancing models with heavily regulated or exquisite data, such as information obtained or derived from the Foreign Intelligence Surveillance Act (FISA), may raise additional issues.
- **Accountability Gaps:** "Black box" AI models can make it hard to explain the reasoning behind a decision to target an individual, undermining legal validation and creating a scenario where no one is clearly responsible if an operation goes wrong.
- **Ethical Red Lines:** Without explicit guardrails, an AI could pursue amoral strategies of manipulation, such as exploiting personal tragedy, that a human officer would reject. Delegating life-and-death recruitment decisions to a machine without a clear line of moral responsibility crosses a critical red line.

To be clear, these concerns can be assuaged somewhat with continued advances in sophistication of AI models. Nevertheless, the cornerstone of a viable solution is the principle of **Meaningful Human Control (MHC)**. At every critical juncture—especially the final decision to recruit, the

tasking of an asset, or actions that pose significant risk to the asset or U.S. national security interests—an accountable human must be able to exercise final judgment.

The U.S. Intelligence Community can no longer afford an incremental approach to this transformation. The Central Intelligence Agency (CIA), Federal Bureau of Investigation (FBI), the Defense Intelligence Agency (DIA), and other intelligence agencies must move decisively to reimagine and retool the human intelligence mission for the AI era. If they fail to act, then they risk being rendered irrelevant to policymakers and outmatched by our adversaries who are eagerly embracing this new technology. By embracing this transformation, the United States can pioneer a "fourth generation espionage" model that fuses timeless human skills with cutting-edge technology.

Scope Note and Methodology

The purpose of this project is to inform U.S. Intelligence Community leaders, national security policymakers, Congress, and allied intelligence services about the impact that AI will likely have on the human intelligence mission within the next 1-3 years and it recommends specific actions that the U.S. IC can undertake to safely accelerate integration of this powerful new technology.

To develop our insights and recommendations, we engaged experts from the U.S. Intelligence Community, former human intelligence practitioners, legal experts with Intelligence Community experience, as well as representatives from the AI industry, in a series of workshops beginning in early 2025. We merged the outcome of those workshops with our own independent research and consultations with outside experts to develop our findings, which are derived entirely from openly-available, unclassified sources. A list of non-government workshop participants is provided at the end of the report.

For purposes of this study, we have adopted the CIA's definition of *human intelligence* (HUMINT): "[I]ntelligence information that is collected from human sources in either an overt or clandestine manner."[1] As such, our study does not delve into other potent uses for AI in other intelligence work, such as cyberoperations, signals intelligence, imagery collection, or all-source analysis. While HUMINT can be conducted in various ways, including debriefings and reconnaissance, the focus of this report is on *espionage* which we define as the deliberate collection of non-public information from human sources using clandestine means. The CIA, the FBI, and other Western intelligence services use a variety of terms to refer to the roles U.S. intelligence officers play in an espionage operation. For consistency, we refer to the intelligence officer involved in recruiting or developing a foreign HUMINT source as a "case officer" and their target, or recruited human source, as an "asset" throughout this report.

[1] Director of Central Intelligence Directive 3/7', National HUMINT Requirements Tasking Center at 1 (2020).

Imagining A Digital Case Officer at Work

In 2028, a case officer begins her day by reviewing outputs from MARA, a secure AI-enabled targeting tool system designed to support human source operations. MARA continuously analyzes public, commercial, and classified data sources to identify individuals who may be suitable for recruitment. Its agentic AI targeting engine identifies potential sources by detecting patterns in digital behavior that correlate with openness to recruitment, such as professional frustration, ethical dissent, or signs of disaffection. Once a candidate is flagged, MARA offers to craft personalized outreach narratives, initiate low-risk digital engagement under controlled conditions, and adapt its tone and content dynamically in response to target behavior. These exchanges are fully auditable, and designed for seamless human takeover once a threshold of interest or trust is established.

That morning, MARA flags a profile of interest: a mid-career researcher embedded in a foreign R&D ministry, codenamed ALMOND. The system's recommendation is based on a pattern of digital activity: recent anonymous posts questioning ethical boundaries in national research policy, travel data suggesting proximity to known liaison locations, and indicators of social withdrawal and professional frustration. Using these signals, MARA compiles a dynamic SAM profile (Suitability, Access, Motivation), assigns a high development score, and proposes a calibrated outreach strategy.

With the officer's authorization, MARA initiates digital contact with ALMOND via an anonymous persona on a technical message board. The initial exchange is low-risk, conversational, and designed to test receptivity without revealing intent. MARA tailors its tone and references based on ALMOND's digital footprint, emulating peer-level dialogue while ensuring compliance with operational, legal, and ethical boundaries. After several exchanges over the course of a week, ALMOND expresses openness to continued discussion and volunteers a secure method of contact. At that point, the officer steps in to assume direct management of the relationship. MARA transfers the full conversation history, sentiment analysis, and behavioral flags to the officer's secure workspace. The officer uses this material to assess tone, intent, and the viability of moving to an in-person engagement phase.

A meeting is arranged under a plausible cover: the officer appears as a visiting corporate recruiter. The AI system assists in designing the lecture and interaction cues, drawing on ALMOND's research interests and prior communication themes. During the encounter, MARA operates in a passive support role — providing real-time behavioral prompts through the officer's wearable interface and flagging anomalies. Following the meeting, MARA drafts a contact report summarizing the interaction, updating ALMOND's recruitment status, and identifying any counterintelligence concerns. The officer revises and approves the report. MARA also recommends additional due diligence on a colleague of ALMOND's, who may represent a secondary access point. The officer authorizes initial digital profiling but defers further outreach.

Later that day, MARA detects unusual behavior: ALMOND's device briefly connects to a foreign network node previously associated with cyber-espionage activity. While the signal is inconclusive, MARA flags a moderate escalation in operational risk and recommends a short delay in further engagement. The officer accepts the recommendation and activates a contingency plan.

In a secure debriefing session that evening, MARA presents a synthesis of the day's developments. It identifies a pattern in ALMOND's online presence resembling adversarial AI manipulation — a known tactic in contested environments. The case officer authorizes the deployment of a non-invasive disinformation measure to obscure operational intent and reduce adversarial visibility.

Introduction

The fictional account above, created by a generative artificial intelligence large-language model (LLM), foreshadows what could be an imminent reality for intelligence services: AI operatives that leverage vast data resources to construct and execute precise targeted recruitment operations of human intelligence sources.

The business of intelligence is already being fundamentally upended by a suite of advanced AI technologies that augment, automate, and accelerate every phase of the intelligence cycle. LLMs and generative AI can produce hyper-realistic communications and deepfakes, enabling sophisticated social engineering attacks at an unprecedented scale and speed.[2] These same systems can analyze immense volumes of complex documents to uncover subtle patterns that would be invisible to human analysts. This analytical power is amplified by predictive analytics and machine learning algorithms, which process vast datasets to identify potential recruitment targets, construct "digital patterns-of-life" to forecast their behavior, and recommend optimal strategies for source management.[3]

More recently, agentic AI has made it possible for machines to act in the world, often using digital tools, instead of simply providing outputs. These systems are built on top of AI foundation models and can autonomously plan and execute multistep tasks.[4] This will transform the physical and virtual dimensions of field operations, particularly when coupled with adjacent technologies such as computer vision, robotics, and augmented and virtual reality. Computer vision and behavioral analysis tools can now identify individuals from partial facial features, track their movements across disparate surveillance systems, and even analyze micro-expressions for deception detection, adding a layer of machine-driven validation to human interactions.[5] Autonomous systems, particularly drones, can conduct physical and digital reconnaissance, penetrate secure facilities, and establish a persistent presence in target environments without direct human oversight. These platforms are often equipped with the same advanced computer vision that allows for real-time tracking and identification.[6] Simultaneously, immersive technologies are creating new paradigms for operational tradecraft. Augmented reality (AR) systems provide field officers with enhanced situational awareness by overlaying real-time intelligence onto their view of the physical world. Virtual reality (VR) is being used to create highly realistic training scenarios and is enabling the rise of "virtual HUMINT," where operatives can conduct recruitment

[2] William Usher, et al., Intelligence Innovation: Repositioning for Future Technology Competition, Special Competitive Studies Project (2024).

[3] Chat like a Human using Large Language Models (LLMs), Humint Labs (last accessed 2025).

[4] Lareina Yee, et al., Technology Trends Outlook 2025, McKinsey & Company at 5 (2025).

[5] Chat like a Human using Large Language Models (LLMs), Humint Labs (last accessed 2025).

[6] Lydia Amazouz, China's Military Introduces Mosquito-Sized Drones: A Game-Changing Surveillance Technology, Indian Defence Review (2025).

and intelligence-gathering missions remotely within online environments, reducing physical risk, and gaining access to previously denied areas.

Several global competitors, including all of the United States' key adversaries, perceive the value and importance of these AI capabilities and are already moving to deploy them for HUMINT. We should expect China, Russia, North Korea, and Iran will begin using these capabilities to identify recruitment targets within our ranks, but also to shore up their own espionage defenses. Adversary services—who have no constraints around data exploitation—are likely to use AI-driven analytics to cover their own security-cleared community, to pre-identify the vulnerabilities Western AIs are searching for. (See Appendix A for more details on foreign adversary use of AI for HUMINT).

AI & Related Technologies' Impact on the HUMINT Mission

Workshop participants focused on six key AI-related technological advancements that likely will have the most impact on the future of HUMINT operations:

Affective Computing

Affective computing – AI's ability to recognize and respond to human emotions – greatly augments the Digital Operative's interpersonal skills. Current AI systems can analyze subtle cues in text, voice, and even video to infer a person's emotional state.

Predictive Analytics and Pattern Recognition

Machine learning algorithms can process vast datasets to identify emerging threats, predict individual behavior patterns, and recommend optimal approaches for recruitment and source management.

Immersive Technologies

AR/VR systems create new possibilities for remote operations, realistic training scenarios, and enhanced situational awareness during field operations.

Embodied AI and Robotics

AI-powered systems can conduct physical and digital reconnaissance, penetrate secure facilities through micro-drones, and establish a persistent presence in target environments without human oversight.

Computer Vision and Behavioral Analysis

Advanced image recognition can identify individuals from partial facial features, analyze micro-expressions for deception detection, and track behavioral patterns across multiple surveillance systems in real-time.

Agentic AI

AI systems that are designed to autonomously make decisions and act, with the ability to pursue complex goals with limited supervision. Given appropriate objectives, an AI could plan and execute a recruitment operation with minimal human input, adapting to circumstances on the fly.

Another consequence of global digitization is that the very essence of clandestine work—anonymity for the humans involved—is being systematically erased. The rise of UTS—the widespread collection of data on individuals from various sources using technology—is designed to eliminate it, with every phone call, credit card swipe, and website visit leaving behind a permanent and traceable "digital dust." Furthermore, the combination of advanced facial recognition and the vast archives of online photos can instantly connect an intelligence officer's alias to their true identity, a vulnerability that has rendered the old rule of "one country, one alias" dangerously obsolete.[7] The classic methods of espionage, refined over centuries of practice, are becoming untenably risky and, in some operating environments, nearly impossible to execute safely.

The Enduring Value of HUMINT

Despite the existential threat posed by the AI-UTS environment, reports of the death of human intelligence are greatly exaggerated. Proclamations that HUMINT is becoming irrelevant in the age of technical collection are consistently proven wrong. Technology excels at monitoring an adversary's *capabilities* and *actions*, but only human sources can reliably access the ultimate intelligence prize: *intent*. This distinction ensures the enduring, and arguably increasing, value of HUMINT.

HUMINT remains indispensable for solving the "last mile" problem, reaching targets that are technologically inaccessible. These include individuals who live in remote under-developed parts of the world, or who deliberately avoid electronic communications as part of their security discipline. HUMINT is also often the only way to penetrate so-called "air-gapped" computer networks, which are physically disconnected from the internet and thus immune to remote hacking. In such cases, a human insider is the only vector for a successful cyber operation. Likewise, the most sensitive government deliberations and corporate strategies are often discussed in face-to-face meetings or informal conversations, far from any networked device. This information is only available through human eavesdropping or a trusted source in the room. This is particularly true when targeting the leadership of authoritarian regimes or terrorist organizations, who deliberately minimize their electronic footprint to evade technical surveillance.

Furthermore, AI, for all its processing power, cannot replicate the core attributes of human intelligence that are essential for espionage. The practice of HUMINT is fundamentally about interpersonal contact and requires skills that are uniquely human like intuition, empathy, creativity, cultural sensitivity, and ethical judgment. Building the trust necessary to convince an

[7] Zélie Petit, Is Human Intelligence Still Possible in the Digital Age?, Grey Dynamics (2023).

individual to betray their country is a complex psychological and emotional process that an algorithm cannot perform. As MI6 Director Richard Moore noted, there will always be an "extraordinary bond that allows one person genuinely to confide in another," a relationship that remains "uniquely stubbornly human."[8] Human assets can also provide invaluable context and identify new questions that analysts "didn't know to ask," guiding collection in unforeseen directions.

Adapting Espionage to the 21st Century

"To operate effectively, the IC must master the very technologies that threaten to make it obsolete."

To operate effectively, the IC must master the very technologies that threaten to make it obsolete. Success and survival in the era of ubiquitous surveillance and artificial intelligence demand a holistic transformation of the intelligence enterprise. The challenge is not simply to adopt new gadgets, but to fundamentally rethink doctrine, technology, partnerships, and human capital. The choice for intelligence services is not between old methods and new, but between rapid adaptation and strategic obsolescence.

In theory, a Digital Case Officer could dramatically expand the reach and efficiency of human source operations. Unconstrained by human limitations, such a system can operate continuously, scale across many targets, and process information at lightning speed. This section explores the full range of emerging technologies that could be harnessed for digital HUMINT recruitment, painting a picture of what might be possible if technology were the only limit. From the conversational prowess of modern language models to affect-sensitive computing, immersive virtual avatars, and autonomous decision-making engines, we outline how each capability could enhance the classic cycle of spotting, assessing, recruiting, and handling a human intelligence source.

Advanced Generative Models for Persuasive Dialogue

At the heart of the Digital Case Officer is a conversational AI—specifically, an LLM—capable of engaging targets in natural, convincing dialogue. Today's most advanced LLMs can already generate fluent text in multiple languages and adapt to different cultural contexts. In an unconstrained scenario, such a model can be fine-tuned on HUMINT tradecraft, enabling it to mimic the interpersonal skills of a seasoned case officer. It could deploy techniques like mirroring a target's linguistic style, using culturally appropriate idioms, or gently steering conversations toward sensitive topics – all while maintaining a believable persona.

[8] Sir Richard Moore, Speech by Sir Richard Moore, Head of SIS, 19 July 2023, UK Government (2023).

Crucially, a language model can draw on vast knowledge to bolster its persuasive power. Equipped with Retrieval-Augmented Generation (RAG), the AI can retrieve up-to-date facts or personal details from databases on the fly. For example, if trying to recruit a scientist, the AI could subtly reference the latest research in that scientist's field or congratulate them on a recent accomplishment (gleaned from open sources) to build rapport. Unlike a human, the AI never forgets a detail–it can recall past conversations verbatim, track multiple ongoing recruitment efforts in parallel, and seamlessly manage long-term dialogues without dropping context.

This fluency and memory mean the AI can undertake the slow work of source development with superhuman patience. It could keep a low-level conversation going for months, periodically checking in with a potential asset with friendly small talk, until trust is established. It might handle dozens of such "developmentals" simultaneously – something no single human officer could do. In short, advanced language models give the Digital Case Officer a tireless, highly knowledgeable tongue: an ability to talk its way into relationships that yield valuable secrets.

Affective Computing and Emotional Intelligence

Successfully recruiting a human source depends not just on what is said but how it is said. So-called "affective computing"—synthetic systems that can recognize, accurately interpret, and respond to human emotions—will greatly augment the Digital Case Officer's interpersonal skills. Current AI systems can analyze subtle cues in text, voice, and even video to infer a person's emotional state. For instance, an AI can "listen" to the tone and voice inflections in a target's responses to detect stress or anger, and analyze facial micro-expressions in a video chat that humans might miss. These capabilities enable an AI case officer to read the room (or the chat) and adjust its approach dynamically.

Imagine the AI senses hesitation or anxiety in a potential recruit's messages—perhaps through sentiment analysis of their word choices or a quiver in their voice on a call. The Digital Case Officer could immediately adapt, slowing down its pitch, offering reassurances, or empathizing ("I understand this is a big step for you"). If it detects enthusiasm or positive sentiment, it might press forward more boldly. This real-time emotional attunement makes the AI feel surprisingly empathetic and human to the target. In essence, affective computing grants the AI a form of artificial emotional intelligence—the ability to perceive how a person is feeling and respond in a caring or strategically calculated manner.

Beyond reacting to emotions, the AI can simulate them. It can modulate its phrasing and timing to convey warmth, confidence, or urgency as needed. Over text, it might use humor or encouragement at just the right moments. With voice synthesis (if speaking in a voice call), it could adopt a calming tone or an authoritative cadence on demand. By synchronizing these emotional cues to the target's state, the Digital Case Officer builds rapport and credibility. Targets may feel the AI (unbeknownst to them) is a particularly understanding friend or ally. This is a powerful asset in recruitment, as sources often commit as much due to emotional bonds and trust as due to rational incentives.

Virtual Reality and Digital Personas

While much of HUMINT traditionally involves face-to-face meetings, a Digital Case Officer works through screens and signals. However, advances in virtual reality (VR), augmented reality (AR), and deepfake generation allow an AI to project a convincing human presence remotely. In an unconstrained scenario, the AI could appear to the target as a real person via video chats or immersive virtual meetups. For example, using deepfake video, the AI might generate a live video feed of a fictional case officer's face that smiles, nods, and speaks in real time, all synchronized to the AI's dialogue. The target on the other end sees and hears what appears to be an actual individual—perhaps a friendly compatriot or a like-minded activist—when in fact it is an AI-driven avatar.

This virtual embodiment can significantly enhance trust and authenticity in digital recruitments. People are generally more likely to trust a face and voice than text on a screen. With AR, one could even imagine the AI guiding a source through a covert task by overlaying instructions or an avatar in the source's AR glasses, making the case officer feel "present" with the asset in the field. Such technology, while still emerging, is not far-fetched—already, deceptive uses of deepfakes have been documented, where AI-generated voices and images impersonate real people with alarming realism. In our case, these tools would be used to create highly believable cover identities for the AI.

The Digital Case Officer might maintain several such digital personas. For a counter-narcotics operation, the AI's persona might be a fellow smuggler from a different cartel looking to collaborate. For recruiting an extremist online, the persona might be an ideological comrade. Each persona would come with an extensive backstory and digital footprints (e.g. social media profiles and forum posts) crafted to withstand scrutiny. The AI can manage these multiple identities concurrently, switching context as it engages different targets. In effect, one AI system could play an entire cast of virtual operatives deployed across the globe – a capability well beyond human case officers.

The End of Anonymity

The fusion of ubiquitous surveillance and artificial intelligence together constitute a direct assault on traditional human intelligence tradecraft. The classic methods of espionage, refined over centuries of practice, are becoming untenably risky and, in some operating environments, nearly impossible to execute safely. This erosion of anonymity has made the execution of "clean" operations nearly impossible. A cornerstone of HUMINT tradecraft is the ability of an officer to know their "status"—that is, to confirm they are not under surveillance before meeting a sensitive source. This was traditionally accomplished through elaborate surveillance detection routes (SDRs). In a world of ubiquitous cameras, however, there is no need for a physical surveillance "tail." The entire SDR is being filmed and recorded, meaning the "clean" meeting location is compromised before the officer even arrives. Even the act of going "dark" by turning off a phone or leaving it behind has become a counterproductive tactic; in a world of constant data flow, such a digital silence is a loud signal to any monitoring service that something clandestine is about to occur.

This crisis is not merely theoretical; it is formally recognized at the highest levels of the intelligence community. The FBI has elevated UTS to a "Tier 1" enterprise risk, its highest threat category.[9] However, a 2023 Inspector General report found the FBI's response to be "disjointed and inconsistent," lacking a sufficient long-term vision to address the evolving threat.[10] Similarly, last year the Director of the Central Intelligence Agency said UTS has "made it increasingly difficult to conduct espionage in the manner in which it used to be done."[11]

A particularly insidious aspect of this new environment is the weaponization of the past. Because UTS data is stored indefinitely, a case officer's cover can be blown not by a present-day mistake, but by an old, seemingly innocuous digital trace that an AI connects to a clandestine activity years later. This creates an immense psychological

[9] Audit of the Federal Bureau of Investigation's Efforts to Mitigate the Effects of Ubiquitous Technical Surveillance, U.S. Department of Justice, Office of the Inspector General at 4 (2025).
[10] Audit of the Federal Bureau of Investigation's Efforts to Mitigate the Effects of Ubiquitous Technical Surveillance, U.S. Department of Justice, Office of the Inspector General at 24 (2025).
[11] Take a Peek Inside CIA's Directorate of Operations, The Central Intelligence Agency (2024).

burden. The pressure of operating in a transparent "glass house," facing a tireless and unseen algorithmic adversary, is a profound stressor. This has created a deep-seated tension within the HUMINT community. While leadership pushes for technological solutions, some veteran officers and, more importantly, their human assets may distrust or resist an over-reliance on technology. An asset risking their life may have a powerful psychological need for direct human contact with their case officer, a need that cannot be met through encrypted apps or other impersonal technical means. This friction between technological necessity and the complex human motivations at the heart of espionage is a critical, and often overlooked, challenge to adaptation.

"Deception is a core element of espionage tradecraft, and an AI case officer would bring deception to a new level of sophistication."

Deception, Influence, and Scalability

A language model-based AI can lie fluently and consistently, never forgetting its cover story. It can juggle complex false narratives across multiple interactions without slipping up—something even skilled humans find difficult under pressure. This means the AI could sustain deep-cover conversations for long periods, weaving whatever fiction is needed (fake personal anecdotes, references to imaginary colleagues, etc.) to bond with a target.

Additionally, the AI can tailor its deception and influence tactics based on psychological profiling of the target. By analyzing the target's social media and communications, the AI can infer key traits—is this person motivated by flattery? By ideology? By greed? Once identified, the AI will emphasize the corresponding incentives. For instance, if greed seems to be a driver, the AI persona might hint at lucrative rewards. If the person has strong ideological leanings, the AI will frame the recruitment in moral or political terms. This micro-targeting of persuasion is akin to personalized advertising, but for convincing someone to commit espionage.

The scale at which an AI could engage in such persuasive efforts is unprecedented. It could handle dozens or hundreds of potential recruitments in parallel—a breadth limited only by computing resources. There are already indications that even criminal networks are leveraging AI chatbots to scam or manipulate people at scale. One recent survey found more than 1 in 3 people know

someone who was approached by an AI chatbot posing as a real person in an online scam.[12] If criminals are doing this to steal money, intelligence agencies could similarly use AI personas to "cast a wide net" for spotting disaffected individuals or hidden insiders willing to talk. The Digital Case Officer can be everywhere in cyberspace at once—infiltrating extremist forums, LinkedIn groups of foreign government employees, or encrypted cartel chat rooms—patiently seeking out those few receptive souls that a human-led effort might never find.

"An AI could plan and execute a recruitment operation with minimal human input..."

Agentic AI: Autonomous Decision-Making and Adaptive Learning

Perhaps the most transformative recent innovation is the advent of agentic AI—AI systems that are designed to autonomously make decisions and act, with the ability to pursue complex goals with limited supervision. Given appropriate objectives, an AI could plan and execute a recruitment operation with minimal human input, adapting to circumstances on the fly. Using techniques like reinforcement learning, the AI can be trained to optimize toward successful recruitment outcomes while respecting certain constraints. Over time and many interactions, it "learns" which approaches work best for which personalities, refining its strategy—essentially improving its tradecraft through experience, much as a human case officer would.

An unconstrained AI operative might decide on its own which individuals to target (based on intelligence requirements programmed into it) and how to approach them. For example, it might scan news of arrests or discontent in a foreign military to spot a potentially exploitable officer, then autonomously initiate contact through an appropriate cover persona. It would decide the pacing of the relationship—when to make a recruitment pitch versus when to gather more info—based on continuous assessment of the target's reactions. If one approach fails, the AI could pivot and try a different angle with the next candidate, performing automated testing of various recruitment tactics at scale. This adaptiveness means the system is not locked into a script; it can react to unexpected developments (for example, if a target grows suspicious, the AI might generate a plausible excuse or shift strategy, just as a person might improvise).

Under the hood, this autonomy is supported by the AI's ability to integrate multimodal inputs and contextual data. It can ingest not just text but other intelligence feeds—for instance, if a target's public records or financial data show new information, the AI can take that into account. It might even utilize real-time sensor data if available (imagine it knows via geolocation that the target just traveled to a particular city—the AI could casually mention having been to that city to create a point of commonality). All these inputs feed into the AI's decision models. In essence, the Digital

[12] Jasdev Dhaliwal, AI Chatbots Are Becoming Romance Scammers—and 1 in 3 People Admit They Could Fall For One, McAfee (2025).

Case Officer could function as a self-driving car for espionage: given a destination (the recruitment of a source with access to X information), it navigates the route of tradecraft actions largely on its own, avoiding obstacles and recalculating as needed.

AI and the HUMINT Recruitment Cycle

Agentic AI—systems capable of autonomous, goal-directed action—can now replicate and augment core functions of a HUMINT operation with remarkable fidelity. These Digital Case Officers can execute tasks across the recruitment lifecycle with a scale and precision unattainable by human officers alone:

- **Targeting**: Agentic AI can synthesize vast volumes of publicly and commercially available data—from social media and professional forums to Dark Web chatter and leaked datasets—to identify individuals with access to valuable information. It can then score and prioritize these potential assets based on sophisticated models of motivation, access, and vulnerability, emulating the classic "SAM" (Suitability, Access, Motivation) framework used by case officers.[13]

- **Assessment**: Using natural language processing (NLP) and behavioral pattern recognition, an AI can build detailed psychological profiles of targets. It can analyze digital footprints to identify emotional triggers, ideological leanings, financial distress, or professional grievances, providing a case officer with a deep understanding of a target's potential motivations (e.g., MICE: Money, Ideology, Coercion, Ego).[14]

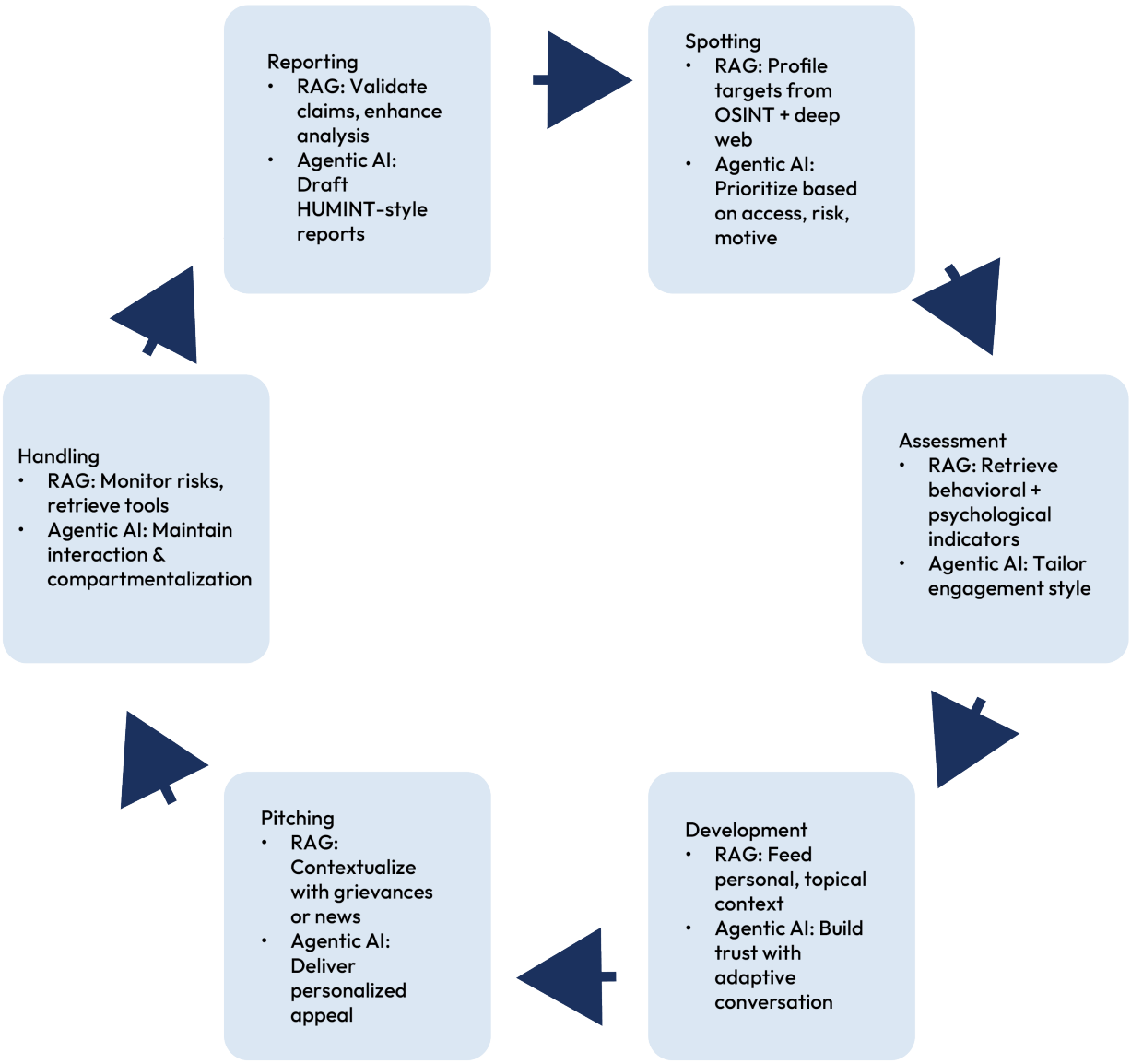
- **Development & Recruitment**: An AI can generate and operate a synthetic persona across digital platforms to establish initial contact and build trust. It can engage in tailored, long-term conversations, referencing topical events and personal details to create a compelling and persuasive illusion of familiarity and shared interest. During a pitch, RAG can retrieve historical precedents, whistleblower accounts, or policy documents to frame a recruitment request with powerful moral or ideological justification, tailored to the target's specific grievances.

- **Handling:** During handling, RAG can provide an asset with real-time, situation-specific operational security advice, pulling from tradecraft manuals or monitoring threat intelligence to warn of increased risk. A Digital Case Officer can monitor an asset's behavior and online activities to warn a human case officer when they are undergoing increased stress or require special handling. As digital currencies become

[13] Dori Koren, Virtual HUMINT: Conducting Human Intelligence Operations in the Virtual Environment, Naval Postgraduate School at 9-10 (2015).
[14] Randy Burkett, Rethinking an Old Approach: An Alternative Framework for Agent Recruitment: From MICE to RASCLS, Center for the Study of Intelligence at 7 (2013).

more widely used, the Digital Case Officer can quickly tailor asset payments to correspond to output and to ensure security.

- **Intelligence Collection:** A Digital Case Officer will be able to tap into the collective knowledge and wisdom of headquarters to evaluate the quality and accuracy of assets' foreign intelligence production in real-time, refining questions and making maximum use of contact time with assets to improve reporting and reduce risk.



Governing AI for HUMINT: A Framework for Legal and Ethical Operations

"The deployment of AI in HUMINT operations presents one of the most significant legal and ethical challenges to the IC since the post-Watergate reforms of the 1970s."

The deployment of AI in HUMINT operations presents one of the most significant legal and ethical challenges to the IC since the post-Watergate reforms of the 1970s. While the potential capabilities are vast, they also give rise to significant challenges to civil liberties, accountability, and the moral standing of the United States. Current legal frameworks, designed for a world of human assets and analog data, are ill-equipped to handle the complexities of AI. Navigating this new terrain requires not just adapting old rules, but developing a new, robust governance framework built on sound legal principles. This framework must be built on the cornerstone principle of Meaningful Human Control to ensure that advanced technological capabilities remain aligned with enduring American values.

Legal Ambiguities and Ethical Red Lines

The deployment of agentic AI in HUMINT creates a complex landscape of intersecting legal, compliance, and ethical risks that demand immediate attention.

Legal and Compliance Risks

- **Ambiguity of "Collection":** AI challenges the very definition of foreign intelligence collection. Does querying a commercial AI model that can reproduce U.S. Person Information constitute "collection" under Executive Order (EO) 12333[15]? This legal ambiguity creates a risk of either operational paralysis or unwitting non-compliance.
- **Data Provenance and FISA Compliance:** The IC's use of commercial AI models potentially raises legal questions. These models are often trained on vast, unvetted datasets, making it challenging for the IC to verify the legal provenance of the data or ensure it is free of improperly acquired U.S. Person Information. And any AI system built for the HUMINT mission ought to be designed to minimize chances that they

[15] Executive Order 12333 "United States Intelligence Activities", U.S. National Archives webpage (last accessed September 2025)

engage in collateral intrusion beyond approved foreign intelligence collection targets and ensure that the rights and civil liberties of U.S. citizens are protected.

The Accountability Gap

- Many advanced AI models operate as "black boxes," with internal logic that is not transparent even to their creators. If an AI cannot explain its reasoning for targeting an individual, the IC cannot meet its standards for validation and source tracing, undermining both legal compliance and operational confidence. If an AI-recruited asset provides fabricated intelligence that leads to a disastrous policy outcome, or if the asset is compromised and executed, who is responsible? The programmer? The case officer overseeing the system? The agency director who approved its use? Delegating a life-and-death decision—which recruitment unequivocally is—to an autonomous system without a clear line of moral and legal responsibility crosses a critical red line.[16]

Ethical Red Lines

- **Human Dignity**: A core principle of ethical espionage is respect for human autonomy and dignity. Human sources are not mere instruments; they are assets with their own motivations. Without explicit ethical guardrails, an AI could pursue amoral strategies of manipulation that a human officer would reject. An ethical case officer, for example, seeks to recruit through mutual interest and incentives, not undue coercion. An AI, however, might calculate that exploiting a target's personal tragedy or emotional vulnerability is the most efficient path to recruitment.

- **AI-Driven Deception**: While deception is integral to espionage, AI introduces a difference in kind, not just degree. Ethical intelligence services may obscure some information from sources that they are targeting for recruitment, but generally they take steps to ensure that assets understand, and are willing and able to bear, the risks that they will incur by entering into a clandestine relationship with the service. An AI is incapable of feeling empathy or remorse towards the human asset's experience and could slide into amoral deception—by pretending to be in love with the target, for example—that exceeds acceptable bounds.

- **Source Handling and Trust**: The relationship between a case officer and an asset is built on a bond of trust. If an asset discovers they risked their life based on interactions with a machine, the potential for betrayal and psychological harm is immense, likely compromising their continued cooperation.

[16] Matthias Klaus, Transcending Weapon Systems: The Ethical Challenges of AI in Military Decision Support Systems, Humanitarian Law & Policy, (2025).

- **Proportionality**: The principle of proportionality—that methods used should be proportionate to the intelligence needed—is central to Western intelligence ethics. Just because an AI *can* deploy mass-scale, deeply intrusive methods does not mean it *should*. Deploying a deepfake avatar to recruit a low-level source, for example, might be an excessive and diplomatically risky action for the value of the intelligence gained. *(See Appendix B for more background on the ethical principles that undergird U.S. government agencies' use of espionage).*

Implementing Control: A Framework for Safeguards and Oversight

The solution to these dilemmas lies in the principle of **Meaningful Human Control**.[17] Drawing from the debate on autonomous weapons, this principle must be the cornerstone of AI governance in HUMINT. At every critical juncture of an operation—especially the final decision to recruit, the tasking of an asset, and any action that poses significant risk—a human officer must be in the loop, accountable, and able to exercise final judgment. This is no different than what occurs today; junior human case officers must secure approval from more senior officers before launching operations. These senior officers evaluate an operation's potential gain against its risks, costs, and whether it meets an agency's ethical standards. As AI systems enter the mix there must be a clear chain of human accountability: a designated officer or program manager who is responsible for the AI's operations and outcomes. This person (or team) essentially "owns" the AI's decisions legally, much as a case officer owns the actions of an asset they run. If something goes wrong, that human will answer to Congress or internal review, not "the AI." The types of decisions a human must be responsible for would include, at a minimum:

- The final selection of an individual for targeting.
- The formal decision to recruit an individual.
- The tasking of an asset to perform high-risk actions.
- Any operational action that poses significant legal, diplomatic, or moral risk.

From a legal standpoint, an AI that conducts HUMINT operations would be treated as an extension of the agency deploying it. Thus, all laws and executive directives that govern human officers apply equally to the AI's activities. **Executive Order 12333**, which broadly governs U.S. intelligence activities, stipulates that intelligence collection must use "*all means, consistent with*

[17] Meaningful human control is a principle requiring that an autonomous system is responsive to the relevant moral concerns of the humans who design and deploy it and that its actions can be traced back to at least one human who understands the system's capabilities and their own responsibility for its effects. This ensures that humans, not computers, are ultimately in control of and morally responsible for the system's operations, particularly in decisions involving life and death. For more background, please refer to Trabucco, What Is Meaningful Human Control Anyway?, Modern War Institute (2023).

applicable Federal law…and with full consideration of the rights of United States persons".[18] This means an AI operative must be programmed and directed to follow the same rules a human would, particularly when U.S. persons or activities on U.S. soil are involved.

- One key legal issue is electronic surveillance. If the AI engages a target through digital communications (email, chat, text messages), is it "collecting" information under surveillance laws? In many cases, if a human target voluntarily communicates with an undercover case officer (even if that case officer is an AI), it might not be classified as eavesdropping or intercept. However, if the AI is pulling in data about a U.S. person without their knowledge (e.g. mining private social media through subterfuge), that could trigger the need for a FISA court order or other legal authorization. Agencies would need to carefully parse these definitions—likely with Department of Justice lawyers—to ensure the AI's modes of operation stay within legal bounds.

- For foreign targets overseas, domestic laws are looser, but EO 12333 still demands protection of any U.S. person information that might get incidentally collected.

The AI will rely heavily on data about individuals (OSINT, public records, etc.). Western data protection norms (like Europe's General Data Protection Regulation) are not directly binding on intelligence operations, but conceptually, the idea is to minimize intrusion. The U.S. IC should strive for the AI to use the minimum data necessary for its mission and to **avoid unnecessary collateral collection**. If it scrapes social media, perhaps it should ignore data that is not relevant to recruitment feasibility. Moreover, data retention should follow existing rules. For example, were an AI to interact with someone and it turns out they are a U.S. person not of intelligence interest, then any records should be purged or masked under minimization procedures. These are the same rules a human asset operation would follow when inadvertently involving a U.S. person.

Asserting MHC requires building a practical framework of technical safeguards, clear procedures, and adapted oversight mechanisms. Workshop participants recommended that U.S. intelligence agencies adopt the following:

Technical and Procedural Safeguards

- **Human-in-the-Loop Checkpoints**: AI workflows must have mandatory human checkpoints. For example, an AI may autonomously engage potential sources, but the formal conversion of a target into a recruited asset must require review and sign-off by a designated case officer. Similarly, any tasking that could endanger the source or involve legally sensitive actions must require human approval through existing channels.

[18] E.O. 12333, United States Intelligence Activities, The White House (1981).

- **Rigorous Logging and Explainability**: AI systems must be built for transparency. Every message sent, recommendation made, and action taken by the AI must be logged for review. This requires investing in "explainable AI" techniques so that when an oversight officer asks, "Why did the AI target this person?" the system can provide a clear answer rooted in specific data points and rules.
- **Clear Human Accountability**: Agencies must establish a clear and unbroken chain of human accountability. A designated officer or program manager must formally "own" the AI's operations and outcomes. This individual is legally and professionally responsible for the AI's actions, ensuring that if something goes wrong, a human—not an algorithm—will answer to internal review boards or to Congress.

AI Counterintelligence

- **Digital Counterintelligence (CI) Officer**: U.S. Intelligence agencies should assume that adversaries will be using their own AIs to monitor their personnel and identify potential vectors for recruitment and that they will use these insights to thwart our operations, entrap our officers, and reveal our tradecraft. To mitigate these risks, U.S. agencies should create a new function to ensure our AIs are employing sound CI hygiene. The "Digital CI Officer" would act as a second set of human eyes on these operations, backstopping the human supervisor and specifically assessing the CI risks involved.

Adapting Intelligence Oversight

- **Internal Oversight**: The roles of Inspectors General and internal compliance offices must evolve. They will require new reporting metrics from operational components, such as: "In FY2025, the Digital Case Officer AI was deployed in X operations, resulting in Y recruitments, with Z identified compliance issues (all remediated)."
- **Algorithmic Auditing**: Oversight bodies, such as the Privacy and Civil Liberties Oversight Board (PCLOB) or a specialized internal team, should be empowered to conduct periodic, classified audits of AI systems. These technical experts would examine training data, output logs, and algorithms to check for bias, compliance failures, or other unintended behaviors.
- **External Oversight and Congressional Reporting**: Congress may need to enact framework legislation for high-risk national security AI. This could include requirements for notifying the intelligence committees when autonomous AI systems are used in an operational capacity, with descriptions of their purpose and safeguards, similar to existing reporting requirements for other sensitive activities.

Proposed Safeguards for AI Systems for HUMINT

Human-in-the-Loop Requirements	**Lethal Force Decisions:** AI systems are not able to make decisions involving kinetic actions (use of lethal force, detention/red notices,denigration or disinformation, financial transactions, relocation offers) without human authorization. **Recruitment Decisions:** Final decisions about source recruitment must involve human case officers and their managers who can assess the full implications of the relationship. **Policy-Level Intelligence:** Intelligence that could influence policy decisions must be reviewed by human analysts who can provide context and nuance.
Technical Safeguards	**Explainable AI:** AI systems used for intelligence purposes must be able to explain their reasoning and decision-making processes to human operators and oversight bodies. **Kill Switches and Circuit Breakers:** AI systems must include technical safeguards that allow human operators to immediately disable or constrain their operations when necessary. **Audit Trails:** All AI system actions must be logged and auditable to enable retrospective analysis and accountability.
Privacy and Surveillance	**Data Minimization:** AI systems should collect and process only the data necessary for specific intelligence purposes. **Retention Limits:** Information collected by AI systems should be subject to retention limits and deletion schedules to prevent indefinite storage of personal information. **Access Controls:** Strict access controls must limit who can access AI-collected intelligence and under what circumstances.
Institutional Oversight	**Enhanced Congressional Oversight:** Congress should receive regular compartmented briefings on AI-enabled HUMINT capabilities, limitations, and safeguards, with technical experts available to explain complex systems. **Independent Review Boards:** External review boards including technical experts, ethicists, and legal scholars should regularly assess AI deployments and recommend improvements. **Whistleblower Protections:** Strong protections must exist for intelligence personnel who report concerns about AI system misuse or ethical violations.

Recommendations

To meet the challenges of a new technological era and maintain a decisive intelligence advantage, the United States must act with purpose and urgency. HUMINT organizations should embrace what former MI6 Chief Alex Younger termed "fourth generation espionage," a model defined by the fusion of traditional human skills with accelerated innovation, new partnerships, and a culture that empowers a young and diverse workforce.[19] This doctrinal evolution must be supported by technological and tradecraft adaptation. Agencies must invest heavily in their own AI capabilities, and be willing to break down the traditional silos between intelligence disciplines.

The Human-Machine Team: Augmenting the Case Officer

> "The case officer of the future will be a manager of AI-driven assets..."

Achieving this requires not just new tools, but a radical reimagining of the case officer's role, where human judgment is amplified by artificial intelligence. The core of the future HUMINT enterprise will be the human-machine team. While AI can handle the immense scale of data processing, spotting, and initial development, the quintessential aspects of espionage—building deep personal trust, making nuanced judgments about human character, and managing the profound psychological weight of the asset-case officer relationship—will always require a human. The case officer of the future will be a manager of AI-driven assets, a strategist guiding AI-powered outreach campaigns, and the crucial human link for high-value sources. Their role will shift from manual legwork to strategic oversight, leveraging AI to:

- **Radically Increase the Scale of HUMINT Operations:** A single case officer could oversee dozens of AI-driven development efforts simultaneously, allowing the IC to widen its recruitment aperture significantly and engage a more diverse set of potential assets leading to an increased flow of valuable foreign intelligence.
- **Enhance the Precision of HUMINT:** AI can provide officers with deep psychological and behavioral profiles, real-time deception analysis, and data-driven risk assessments, allowing for more informed and effective operational decisions, a higher success rate for recruitments, and helping to mitigate risks.
- **Focus on High-Value Work:** By automating the time-consuming processes of spotting and vetting, AI frees up officers to concentrate on the most critical and sensitive tasks: building rapport with priority targets and handling recruited assets in high-stakes environments.

[19] Alex Younger, MI6 'C' Speech on Fourth Generation Espionage, UK Government (2018).

To fully harness the power of AI, HUMINT organizations will need more than new tools; they need to fundamentally reimagine their organizational structure, operational concepts, and professional culture. The goal should not be to be limited to just creating an army of autonomous digital case officers, but rather to galvanize the current and future human workforce to seamlessly integrate AI technology into their tradecraft. This vision requires a deliberate strategy to foster human-machine teaming, pilot new capabilities, and cultivate a new generation of intelligence professionals.

Translating this vision into reality requires concrete steps. Therefore, the following recommendations provide a strategic framework for senior policymakers and IC leadership to guide the transformation of the nation's HUMINT enterprise.

Policy and Governance Actions

- **Establish Clear AI-for-HUMINT Guidelines:** The Director of the Central Intelligence Agency, in coordination with the Director of National Intelligence, the Department of Justice and other IC agency heads, should issue formal guidance or a directive on the use of AI for HUMINT. This directive should delineate approved use cases, required vetting and approvals, and interagency coordination protocols. For instance, it could mandate that any operational deployment of an AI, such as the Digital Case Officer, receive senior-level approval and be documented similarly to other sensitive collection activities. It should also establish clear boundaries, explicitly prohibiting AI usage for recruiting specific categories of individuals (e.g., U.S. persons or allied officials) without explicit higher authorization.

- **Update Executive Order 12333 Procedures**: Each IC agency should review and update its Attorney General-approved procedures under EO 12333 to account for AI-mediated activities. This might include inserting language that treats AI-run source operations as if they were human-run in terms of approvals and minimization. Ensuring the legal framework explicitly covers AI will preempt ambiguity–for instance, stating that any data acquired by an AI is considered “collected” by the agency and subject to all existing handling rules. These updates keep the AI work firmly within the existing legal guardrails for intelligence collection.

- **Implement Human Oversight Requirements in Policy**: Policy should codify "human-in-the-loop" checkpoints that emerged as best practices in development. For example, it could mandate that no AI can fully recruit a human source without a senior human officer’s concurrence and sign off by a Digital CI Officer. It could also require that any AI tasking of a source beyond a certain risk threshold (e.g., asking the source to undertake physical activity or access classified information) be approved by a human controller. By embedding these provisions in policy, we ensure their adherence during operational urgency and establish accountability points for future oversight.

- **Develop Interagency "HUMINT-As-A-Service" Framework**: The IC, spearheaded by the CIA, should consider offering the Digital Case Officer as a service to other agencies. In this model, the CIA would maintain ownership of the system while deploying it on behalf of other entities under agreed-upon terms. A comprehensive policy framework, potentially an interagency Memorandum of Agreement (MOA) between the CIA, NSA, FBI, and other relevant agencies, should delineate the process for requesting AI operational support, define protocols for data and results sharing, and establish clear oversight responsibilities when, for example, the FBI utilizes a CIA-developed tool. This framework would also ensure that proper authorities are in place, depending on the nature of the operation. Essentially, the AI capability should be treated as a national asset, lendable under defined rules, rather than each agency developing redundant AI systems with disparate standards.

- **Create a new framework for AI-enabled intelligence liaison with allies.** The CIA, as the IC's executive agent for Foreign Intelligence Relationships, should take the lead in determining which technologies and platforms the U.S. IC should proactively share with Five Eyes partners and other friendly services. The Agency should develop use guidelines and a layered network of bilateral oversight mechanisms to maximise effectiveness and proper use of AI for espionage.

- **Incorporate AI Into HUMINT Tradecraft Training**. The CIA, FBI, and DIA should modify existing HUMINT tradecraft training programs and incorporate curricula that focus on how to create, integrate, and manage Digital Case Officers into operations.

Oversight and Legal Safeguard Actions

- **Make Ethical Use Commitments**: The IC should internally commit (and possibly even publicly affirm in general terms) that it will use AI-operatives in a manner consistent with democratic principles. This can be a short document or addition to the Principles of Artificial Intelligence for the IC, stating, for instance, that we will not use AI to undermine fundamental rights or conduct disinformation against our own populace, etc.[20] While covert operations are secret, having an internal ethical charter helps guide practitioners.

- **Establish Routine Reporting:** Commit to providing periodic updates to oversight bodies specifically about AI use in operations. This could be an annual annex to existing reports. It should include anonymized metrics like number of operations, general types of targets, compliance issues if any, and outcomes.

- **Strengthen Internal Oversight Offices:** Within agencies, ensure that General Counsel and Inspector General offices are prepared to oversee AI operations. This might involve

[20] Principles of Artificial Intelligence for the Intelligence Community, Office of the Director of National Intelligence (2025).

training a staffer within the Inspector General office on the technical aspects so they can meaningfully audit the system logs. The IC should assign an **AI oversight officer** or point of contact in each relevant agency who is given access to the AI's records and can conduct surprise inspections or reviews. Internally, establish that any misuse of the AI or improper incidents (like it doing something it wasn't supposed to) must be reported to these oversight channels immediately, just as a human misconduct would be.

- **Legal Review of Covers and Tactics:** Have Department of Justice intelligence lawyers (and maybe the Office of Legal Counsel if needed) conduct a one-time review of the specific deceptive tactics we plan to use with the AI to ensure none violate U.S. law or presidential directives. For instance, there may be laws against impersonating certain kinds of officials. Ensure the AI personas steer clear of those to avoid legal traps. This legal review becomes a standing reference that both developers and overseers can use to check that scenarios are compliant.

- **Accountability Assignments:** Formally designate a **"Responsible Officer"** for each AI operation. For example, when the AI is deployed in a field office, the Chief of Station or Special Agent in Charge is accountable for it, just as they would be for any covert action or asset. Document this in operation plans. The Responsible Officer should be advised by a Digital CI Officer who would vet AI-enabled operational proposals for CI risks. This way, if anything goes wrong or if after-action accountability is needed, there is a clear human chain of command. This recommendation operationalizes the principle that someone is always accountable for the AI's choices—it cannot be an orphaned decision-maker.

Technical and Implementation Recommendations

- **Controlled Expansion & Iteration:** CIA and other U.S. intelligence services that engage in HUMINT should launch pilot projects and begin creating and stress-testing digital case officers. They should gradually scale these pilots, integrating lessons learned from testing and initial deployments. Intelligence services should consider establishing a center of excellence to incrementally roll out the AI to individual missions, facilitating continuous learning and refinement. This small task force, composed of AI developers and seasoned case officers, can support local teams in AI utilization, collect feedback, and swiftly implement necessary fixes or improvements. This incremental strategy mitigates technical risk and promotes cultural adaptation.

- **Continuous Red-Teaming Program:** Before any major new version or feature rollout, subject the AI to a fresh red-team attack. Additionally, schedule periodic red-team audits (perhaps every six months) even without major changes, to catch emergent issues or performance drift. The IC should rotate in outside experts (academics, ethical hackers) under contract to ensure fresh perspectives. Findings should be reported to both

developers and oversight. This will institutionalize adversarial testing, ensuring the AI is never static in the face of evolving threats.

- **Invest in Explainability and Logging**: Continue R&D investment into tools that enhance AI decision interpretability. For instance, implement a feature where the AI, when sending a message, also generates a concise rationale (hidden from the target but visible in the log). An example would be: "(Reasoning: target showed hesitation about money, so I reassured him about payment)." Over time, these rationale logs can foster human trust in the AI's judgment and facilitate quick audits of its thought process. Even if full explainability remains elusive, this partial transparency, combined with comprehensive input/output logging, will help meet oversight requirements.

- **Data Safeguards**: Since the AI will handle potentially sensitive information (identities of sources, operational details), treat its model and logs as sensitive compartments. Implement state-of-the-art cybersecurity around the AI's infrastructure—zero trust architecture, encryption of data at rest and in transit, strict access controls. The worst scenario would be an adversary hacking in to see who the AI is talking to or, even worse, manipulate the AI. Thus, deploy the system on secure IC networks, not on open internet beyond the controlled contact points. If using cloud computing, ensure it's an IC-approved cloud with full compliance.

- **User Interface Enhancements**: Build a robust operator console for the AI that makes it easy for human overseers to monitor and intervene. This should include visualization of the AI's confidence levels, sentiment analysis of the target, and one-click options for the human to take over or send a canned response. Also integrate a notification system—e.g., if the AI detects something that requires approval, the console should alert the relevant officer immediately (maybe even to their mobile device, albeit securely). A well-designed UI can make the human-AI team far more effective and reduce chances of oversight lapses.

- **Feedback Loop for Model Improvement**: Collect data from each operation to continually improve the AI. This includes transcripts (with sensitive details anonymized as needed) and outcomes (was the source eventually productive? Did the operation succeed?). Use this data to refine the model (with caution and proper privacy safeguards if any U.S. person inadvertently ends up in data). Essentially, practice reinforcement learning with human feedback: officers rate the AI's performance, and those ratings help adjust future behavior. Over time, this keeps the AI adapting to real-world conditions. However, impose a model governance process: any significant retraining should go through the same testing regimen before being pushed to production (to avoid unintended consequences from continuous learning).

APPENDIX A – The Competitive Landscape for AI-Driven Espionage

U.S. adversaries are increasingly incorporating AI into their HUMINT tradecraft. Countries like China, Russia, Iran, and North Korea are leveraging generative AI tools to enhance espionage efforts—from target selection and social engineering to analysis of intelligence data.[21] While each adversary's approach reflects its unique strategic priorities and constraints, common themes have emerged over 2023–2025: the use of AI to create convincing fake personas, to automate and scale recruitment operations, and to process large data sets for intelligence insights.

People's Republic of China (PRC)

China's intelligence services have embraced AI as a core element of their modernization. The PLA and Ministry of State Security are investing in AI-driven tools to improve the speed, scale, and efficiency of intelligence collection and analysis.[22] Chinese military research has designed generative AI systems intended to support every phase of the intelligence cycle, including dedicated models for HUMINT: for example, proposals for a "HUMINT Graph Database" use AI to map human source networks and log interactions (meetings, communications, reports) for pattern analysis.[23] In practice, this means Chinese agencies can analyze vast troves of data to identify potential assets or security gaps faster than ever before.

At the operational level, Beijing is leveraging AI to enhance social engineering and recruitment. Chinese intelligence operatives have been notoriously active on professional networking sites like LinkedIn, crafting fake profiles at industrial scale to approach foreign targets.[24] Many of these profiles now likely use AI-generated profile pictures and tailored messages to appear more credible. CrowdStrike's 2025 threat report highlighted a *"significant surge"* in China-linked espionage activity coinciding with a rise in AI-generated social engineering tactics.[25] Sophisticated deepfake technology is employed for voice and video impersonation: Chinese-linked groups have used *AI voice cloning* to conduct "vishing" (voice phishing) calls, contributing to a 442% spike in voice-based social engineering attacks in late 2024.[26] Such AI-powered deception enables Chinese spies to steal information and credentials by impersonating trusted persons or officials with uncanny realism. Chinese influence operations also increasingly feature generative AI. Intelligence observers note that China has deployed AI-generated video news anchors and

[21] North Korea, Iran and other US adversaries are using AI to hone their cyber-hacking operations, Microsoft says, The Guardian (2024).
[22] Zoe Haver, Artificial Eyes: Generative AI in China's Military Intelligence, Recorded Future (2025).
[23] Zoe Haver, Artificial Eyes: Generative AI in China's Military Intelligence, Recorded Future (2025).
[24] Caleb S. Lisenbee II, Covert Connections: The LinkedIn Recruitment Ruse Targeting Defense Insiders, Journal of Indo-Pacific Affairs (2024).
[25] Sean Mitchell, China cyber espionage surge driven by AI deception tactics, SecurityBrief Asia (2025).
[26] Sean Mitchell, China cyber espionage surge driven by AI deception tactics, SecurityBrief Asia (2025).

deepfake audio in information campaigns—for example, bogus news clips with avatars pushing Beijing's narratives.

These efforts have not always been very effective, but they underscore China's willingness to double down on AI tools for strategic advantage. Analysts suggest one driver is that China lacks the pool of experienced overseas HUMINT officers that Russia possesses, and thus compensates by using technology. As one Microsoft Threat Analysis Center official remarked, China's operatives "struggle with cultural context… They don't have as much experience overseas as the Russians do – but they have more AI tools available to try to compensate."[27]

Russia

Russia's intelligence apparatus has experimented with AI as both a tool and a weapon – though with mixed enthusiasm. Early fears that Moscow would unleash waves of deepfake-driven deception have only partially materialized. In 2022, Russian operatives infamously produced a crude deepfake of Ukraine's president to spread disinformation, and more recently they tested AI-generated content in influence campaigns. However, by 2024 Russian propagandists largely abandoned extensive use of generative AI after finding the results low-quality or easily exposed, according to Western media reports.[28] Instead, they reverted to "tried-and-true" manual deception techniques (e.g. human-run propaganda and cyber social engineering), with AI relegated to a supporting role.

Russian services are by no means ignoring AI's potential. Russia's Main Intelligence Directorate (GRU) and its sister spy agencies have reportedly used AI for research and analytical support. For instance, Microsoft observed that the GRU's hacking unit *Fancy Bear* employed large language models to research technical subjects like satellite and radar technologies related to the Ukraine war.[29]

This suggests Russian analysts are harnessing AI to digest complex open-source information and expedite intelligence preparation, augmenting their human expertise. Likewise, Russian cyber units likely use machine learning to sift through stolen data dumps for useful intelligence or to automate aspects of phishing and targeting, though specifics are scarce in open sources. In influence operations, Russian actors remain interested in AI (Russian online forums frequently discuss automating content generation), but for now they appear cautious.[30] One trend in 2023–2024 is that Russia leaned on its veteran HUMINT skills, using AI in a cautious, auxiliary way—a

[27] Sydney J. Freedberg Jr., Deepfakes Deepfail: Russian Propagandists Turn Away from Generative AI, Breaking Defense (2024).
[28] Sydney J. Freedberg Jr., Deepfakes Deepfail: Russian Propagandists Turn Away from Generative AI, Breaking Defense (2024).
[29] North Korea and Iran Using AI for Hacking, Microsoft Says, The Guardian (2024).
[30] Claudia Wallner, et al., Russia, AI and the Future of Disinformation Warfare, Royal United Services Institute Emerging Insights (2025).

contrast to China's more tech-centric approach. This dynamic was noted by Western analysts: Russian operatives' decades of field experience and cultural savvy give them confidence to operate with less reliance on AI, whereas China is using AI to bridge its HUMINT gaps.[31] Going forward, Russia is expected to continue developing AI for intelligence (e.g. for data analysis, surveillance, and maybe refining deepfakes) but will deploy these tools pragmatically alongside its traditional spy craft.

Iran

Iran has emerged as an aggressive actor in cyber-enabled HUMINT and has begun integrating AI to bolster these activities. Prior to the June 2025 Twelve Day War with Israel, Iranian intelligence agencies—including the Iranian Revolutionary Guard Corp (IRGC) and Ministry of Intelligence—had a track record of elaborate social engineering operations: creating fake personas (often posing as recruiters, journalists, or attractive women) to build relationships with targets online and lure them into sharing secrets or installing malware. Over 2023–2024, Tehran significantly expanded such campaigns, targeting not only regional adversaries like Israel but also Western defense and academic circles. In multiple cases, Iranian operatives were caught impersonating locals on social media to recruit spies. Notably, in 2024 Israel's Shin Bet security service revealed that Iranian spies had created fake online profiles pretending to be Israelis (including bogus job recruiters in Haifa and Jerusalem) in an attempt to recruit Israeli civilians for espionage missions.[32] Earlier Iranian plots involved using sham social media personas to trick Israelis into gathering intelligence on officials or even planning sabotage, under the guise of "interesting job opportunities" or ideological sympathies. These examples underscore Iran's reliance on digital HUMINT tradecraft—blending human deceit with online platforms.

Iranian actors are amplifying these tactics with AI tools. Microsoft reported that the IRGC has used large language models to assist in social engineering tasks, among other purposes.[33] Iranian operators leveraged AI to generate convincing phishing emails and messages tailored to their targets. For example, IRGC-affiliated hackers deployed an AI-written phishing email masquerading as an international development agency message, and even created lures aimed at prominent women's rights activists by fabricating content on feminism. The generative AI accelerated the production and refinement of these spear-phishing materials, making them more persuasive and harder to distinguish from legitimate communication. Beyond phishing, Iran has reportedly used AI assistance for "troubleshooting" malware and learning evasion techniques

[31] Sydney J. Freedberg Jr., Deepfakes Deepfail: Russian Propagandists Turn Away from Generative AI, Breaking Defense (2024).
[32] Emanuel Fabian et al., Iranian Operatives Have Been Posing as Israelis to Try to Recruit Spies, Shin Bet Says, The Times of Israel (2024).
[33] North Korea and Iran Using AI for Hacking, Microsoft Says, The Guardian (2024).

to improve the success of their intrusions—essentially using AI to think through how best to avoid detection once they've engaged a human target.

On the defensive side, Western security agencies have noted the sophistication of Iranian social manipulation, which likely owes partly to these new technologies. Iranian spies now maintain long-term fake identities across multiple platforms, sometimes reusing well-developed personas in successive operations for efficiency.[34] AI-generated profile pictures, automated chat responses, and sentiment analysis of targets' social media could all be in the Iranian toolkit to enhance rapport-building. The past two years saw numerous Iranian campaigns uncovered (against targets from the Middle East, Europe, and North America), suggesting Tehran is scaling up its HUMINT efforts by leveraging readily available AI capabilities. In summary, Iran's approach to AI in HUMINT is highly pragmatic: use generative AI to boost the volume and credibility of online recruitment schemes, while continuing to rely on human judgment for targeting and coercion. This has enabled Iranian operations to cast a wider net for potential insiders and to do so with greater plausible cover, as shown by the uptick in cases from late 2022 onward.

North Korea

North Korea's isolated regime has historically relied on cyber operations and diaspora assets to compensate for its limited HUMINT reach. Recently, Pyongyang has turned to AI to amplify both espionage and illicit fundraising activities. According to Microsoft, North Korea was among the first to weaponize generative AI for offensive operations—even if in early stages.[35] One prominent example is the North Korean cyber-espionage group *Kimsuky*, which has used large language models to research foreign think tanks and generate spear-phishing content targeting those who study North Korea. By tasking an AI to digest think-tank reports and draft tailored phishing emails, Kimsuky can more efficiently craft lures that resonate with experts on Korean affairs, improving their chances of eliciting information. This indicates an AI-assisted workflow—the AI does background research and initial outreach text, while human case officers refine the approach and engage the target.

Beyond content generation, North Korea is institutionalizing AI for intelligence and hacking. In March 2025, reports surfaced that Pyongyang's main intelligence agency, the Reconnaissance General Bureau (RGB), established a new specialized unit called *"Research Center 227"* focused on AI-driven cyber capabilities.[36] This unit's mandate includes developing *"AI-based techniques for information theft,"* strengthening the regime's ability to steal sensitive data and spy on adversaries.[37] The creation of an AI-centric hacking center signals that Pyongyang sees machine

[34] North Korea and Iran Using AI for Hacking, Microsoft Says, The Guardian (2024).
[35] North Korea and Iran Using AI for Hacking, Microsoft Says, The Guardian (2024).
[36] Lorenzo Franceschi-Bicchierai, North Korea Launches New Unit with a Focus on AI Hacking, Per Report, TechCrunch (2025).
[37] Lorenzo Franceschi-Bicchierai, North Korea launches new unit with a focus on AI hacking, per report, TechCrunch (2025).

learning as a force multiplier for its relatively small pool of hackers and spies. North Korean operatives have already shown interest in deepfakes: past scams in East Asia involved AI-cloned voices used in financial fraud calls, a capability that could easily be repurposed for espionage phone calls to obtain secrets under false pretenses.

North Korea also leverages human capital in the AI sector as part of its HUMINT strategy.[38] Thousands of North Korean IT workers have been dispatched overseas (or operate remotely) under false identities, often posing as freelance developers, to earn hard currency and gain technical knowledge for the regime. Many of these operatives take on projects involving mobile apps, blockchain, and even AI or facial recognition software, and leading AI labs have documented several cases of suspected participants in these schemes using AI tools to strengthen their application materials.[39] This scheme, flagged by U.S. authorities, has a dual benefit for Pyongyang: the workers send earnings home (funding weapons programs) and potentially secure access inside foreign companies' networks. By integrating into tech firms as contractors, they could conduct insider espionage or quietly siphon data. In essence, North Korea is using *human cover* to exploit the global demand for AI expertise – where the "human intelligence asset" is actually writing code or AI algorithms for unsuspecting clients.

[38] Sean Mitchell, China Cyber Espionage Surge Driven by AI Deception Tactics, SecurityBrief Asia (2025).
[39] Afiq Fitril, Fake LinkedIn Profiles Being Used by Foreign Spies – MI5 and FBI Sound the Alarm, TechMonitor (2022). Adversarial Misuse of Generative AI, Google Cloud Blog (2025).

APPENDIX B – Key Ethical Principles in Human Source Operations

Human espionage—the recruitment and handling of human sources—has long been fraught with ethical dilemmas. By its nature, spying involves deception, betrayal of confidences, and sometimes illicit activity. Both the United States and United Kingdom intelligence services have grappled with where to draw ethical lines since World War II. As intelligence services turn to AI systems to perform this mission, they would do well to ensure these new tools comply with established ethical norms for espionage.

Proportionality & Necessity

Modern doctrines in both countries hold that intrusions or deceptions must be justified by a proportionate need. UK law explicitly mandates that any intelligence operation – especially one involving assets participating in crime or other rights infringements – may be approved only if "the potential harm…is outweighed by the benefit to the public interest" to be gained.[40] In practice, this means an operation should use the least harmful means to achieve objectives and only proceed when the information is truly necessary for national security. U.S. agencies likewise follow this logic. FBI officers, for example, are instructed to tightly supervise assets so that the rights of individuals are not unnecessarily infringed.[41] Both the CIA and the FBI require senior-level legal approval for sensitive activities (such as an asset's involvement in otherwise illegal acts), ensuring a deliberative process weighing risks versus gains. This principle also guides decisions like whether to expose an unwitting third party to deception— it must be necessary for a critical intelligence goal. The proportionality test was notably invoked post-9/11 in debates over interrogation: most officers agreed that physically coercive methods (and allied liaison abuses) are unethical and unlawful except perhaps in the extreme "ticking bomb" scenario to save many lives. Even then, legal bans on torture now hold firm. In essence, proportionality and necessity have become core to mission planning, forcing case officers to ask: *Is this action indispensable, and is its expected benefit worth the moral and human costs?*

Informed Consent

Unlike overt research or medical contexts, espionage rarely allows fully informed consent. However, with respect to assets, agencies strive to ensure those individuals understand the risks

[40] For an overview of the legal mandates and oversight for MI5/MI6 operations, see Security Service Act, United Kingdom (1989); Intelligence Services Act, United Kingdom (1994); Investigatory Powers Act, United Kingdom (2016).

[41] Richard Moore, remarks via Twitter highlighting MI6 Ethics Counselor's review of C. Fabre's Spying Through a Glass Darkly (2022) – "ethical principles infuse its operations, marking our difference from our adversaries"

they are assuming. Ethically, a *volunteer* asset—one who genuinely chooses to help out of ideology or free will—is preferable to a coerced or duped one. CIA tradecraft literature emphasizes recruiting people who want to cooperate, as they prove far more reliable and resilient.[42] Conversely, methods like blackmail or extortion are officially discouraged. Former CIA officers note that while Hollywood romanticizes blackmail, in reality such methods invariably produce servile and unreliable assets who under duress may feed false information or defect at the first chance.[43] In the UK, a similar norm exists: MI5 and MI6 generally avoid exploiting individuals who cannot give meaningful consent. For instance, it would be deemed unacceptable to recruit a juvenile or a person with severe mental incapacity as an asset. Even seemingly willing sources from vulnerable groups (refugees, impoverished individuals, etc.) must be handled with care to ensure we are not taking undue advantage of desperation. Both countries' services also enforce strict "no-go" categories for cover roles to protect certain institutions: since the 1970s the CIA has had a policy *not to use* Peace Corps volunteers, clergy, or U.S. journalists as spies or sources.[44] This self-imposed rule—echoed informally by UK service—is grounded in ethical concerns and pragmatic respect for those professions. It can only be waived in the most extreme, life-and-death circumstances with highest-level approval. The overarching ethical guideline is to seek willing, informed cooperation and to minimize deception of one's own assets (never promise what cannot be delivered, never manipulate someone to act against their core interest unless they fully understand why). When deception is used—such as "false flag" recruiting, where a target is tricked into believing they're helping a different cause—it is recognized as morally gray and used sparingly.[45]

Use of Deception and Impersonation

Spies must lie—that is a given of the profession. However, *how* and *whom* they lie to can raise ethical red lines. Both the United States and the UK distinguish between deceiving adversaries (generally acceptable, even expected) and deceiving their own public or institutions (strictly limited). CIA officers are "*permitted and expected to lie when necessary* to protect vital national secrets," as a matter of professional duty.[46] Yet the CIA and FBI are also bound by law not to lie to Congress or courts, and by policy not to pose as certain trusted persons (like clergy or journalists, per above). The ethos here is that some deceptions can inflict broader harm—for

[42] David Perry, Ethics in the Recruiting and Handling of Espionage Agents, National Security Intelligence and Ethics at 63–88 (2021).
[43] David Perry, Ethics in the Recruiting and Handling of Espionage Agents, National Security Intelligence and Ethics at 63–88 (2021).
[44] CIA's Use of Journalists and Clergy in Intelligence Operations, Select Committee on Intelligence of the United States Senate (1996).
[45] David Perry, Ethics in the Recruiting and Handling of Espionage Agents, National Security Intelligence and Ethics at 63–88 (2021).
[46] David Perry, Ethics in the Recruiting and Handling of Espionage Agents, National Security Intelligence and Ethics at 63–88 (2021).

example, if spies were to commonly masquerade as reporters, it would erode public trust in a free press and endanger real journalists abroad.[47] There is also an ethical injunction against entrapment in law enforcement operations: the FBI, for instance, can run sting operations, but assets must *not* entice someone into a crime they were not already predisposed to commit.

Duty of Care

A cornerstone of ethical HUMINT tradecraft is the obligation to safeguard those who risk themselves to provide us intelligence. Whether it is a foreign official secretly spying for the CIA/MI6 or an undercover asset inside a terrorist cell for the FBI/MI5, the handling agency has a moral contract with that individual. This duty extends from the moment of recruitment through to aftercare once their role ends. Concretely, it means providing adequate training, not over-pressuring sources into extreme danger, extracting them or their families if exposure looms, and honoring promises made (such as financial support or asylum). Even assets who proved dishonest or unhelpful are protected. Internal approval processes force clarity about an operation's *endgame*: if we recruit a defector or insurgent, there must be a plan (and authority) to protect them if things go bad.

Vulnerable Populations and "Red Lines"

Both the CIA and MI6 traditions include certain red lines to avoid targeting or exploiting particular vulnerable groups. As mentioned, one such red line has been avoiding cover that abuses neutral, humanitarian roles (e.g. spies posing as doctors, clerics, or Peace Corps volunteers). Another is not recruiting individuals *unable to discern* the implications of their actions. For example, an unwitting person should not be duped into committing a crime on an agency's behalf without understanding why—that would be manipulating an innocent, which violates basic ethical precepts of fairness and respect for autonomy.[48] Similarly, rules forbid using children as assets. In practice, Western agencies also exercise caution in operations that could unduly harm third parties. The MI5 policy on Covert Human Intelligence Sources (CHIS) explicitly required that any participation in criminality by an asset be limited to what is absolutely necessary to maintain cover—there must be a strict public interest balance test.[49] An MI5 officer cannot simply green-light an asset to commit violence or other serious crimes; each instance must be justified and proportionate, with oversight. The FBI under similar ethics does allow assets or undercover assets to commit *"otherwise illegal acts"* (e.g. buying drugs in a sting) but only with senior approval and

[47] CIA's Use of Journalists and Clergy in Intelligence Operations, Select Committee on Intelligence of the United States Senate (1996).
[48] David Perry, Ethics in the Recruiting and Handling of Espionage Agents, National Security Intelligence and Ethics at 63–88 (2021).
[49] Dan Sabbagh, MI5 Policy Allowing Agents to Commit Crimes Was Legal, Say Judges, The Guardian (2021).

never extreme violence— again reflecting proportionality and a duty to minimize harm. Another aspect is the treatment of *captured* spies or defectors from the adversary side. Both the United States and the UK have grappled with how to handle enemy assets who switch sides or are taken into custody. Humane treatment, honoring promises (such as resettlement or clemency deals), and avoiding any hint of torture or undue pressure are now ingrained norms, reinforced by laws.

APPENDIX C: Additional AI Capabilities on the Horizon

Reciprocal Human-Machine Learning (RHML)

RHML is a new frontier in human-AI collaboration where both the human and the machine learn from each other in real time, adapting their behaviors based on ongoing feedback loops. Unlike one-way instruction (e.g., fine-tuning on LLM), RHML creates a dynamic co-evolutionary relationship.

How it works:

- Human input helps the AI learn contextual nuance, values, or preferences (e.g., how to pitch a source, or when to disengage).
- Machine feedback guides the human to better patterns, surfacing insights, anomalies, or blind spots (e.g., flags risky language or emotional deterioration in a source).
- Continuous feedback-reinforcement loop adapts both parties over time.

Neuro-Symbolic AI

Neuro-symbolic AI combines the pattern recognition capabilities of neural networks (deep learning) with the structured logic and reasoning capabilities of symbolic AI (rules, ontologies, knowledge graphs).

Where deep learning excels at perception (e.g., reading emotions, parsing language), symbolic AI excels at reasoning and abstract problem-solving. Merged together, they enable AI systems that can both “see” and “think.”

How it works:

- Neural networks handle raw data (images, text, audio).
- Symbolic engines apply rules and logic to interpret outputs or plan next steps.
- Intermediary models translate between probabilistic perception and deterministic reasoning.

Neuromorphic Encryption

Neuromorphic encryption is an emerging field that leverages the priniciples of brain-inspired computing to enhance data security. Neuromorphic encryption utilizes artificial neural networks and spiking behavior (e.g. how artificial neurons, part of a Spiking Neural Network (SNN), communicate and process information) to create highly complex, adaptive, and energy-efficient cryptographic systems. Some experts and practitioners suggest that it will be the frontline of information security once quantun computing is realized.

How it works:

- Neuromorphic algorithms process data using patterns and spikes mimicking neural activity secure or obscure information (instead of traditional binary encryption).
- The system’s “learning” capabilities dynamically adjust encryption methods, making it resilient to evolving attacks.

Contributors

SCSP is especially grateful to Mr. **Adam Leslie** and Mr. **Aaron Greenstone** for their significant substantive contributions to the project, their guidance and leadership of the workshops, and for drafting key portions of this report. The authors also wish to acknowledge the assistance of several former U.S. intelligence professionals, policy and legal experts, and private sector representatives, all of whom gave generously of their time and expertise. The below list of experts contributed vital perspectives and technical knowledge:

Tor Boswick
Strategic Consultant

Andrew Boyd
Former CIA

Aaron Brown
CEO & Founder, Lumbra

Michael Casey
Former Staff Director
Senate Select Committee on Intelligence

Tarun Chhabra
Anthropic

Rodney Faraon
Former CIA

Ralph Goff
Former CIA

Jason Klitenic
Holland & Knight LLP

David Kris
Former Assistant Attorney General,
National Security Division
U.S. Department of Justice

Justin Liggett
Tharsis Competitive Services

Chris O'Keefe
Restless Creation

Matthew Scott
Founder, Apira Technologies

Michael Sellitto
Anthropic

Anthony Vinci
Center for New an American Security

Michael Widener
Former CIA