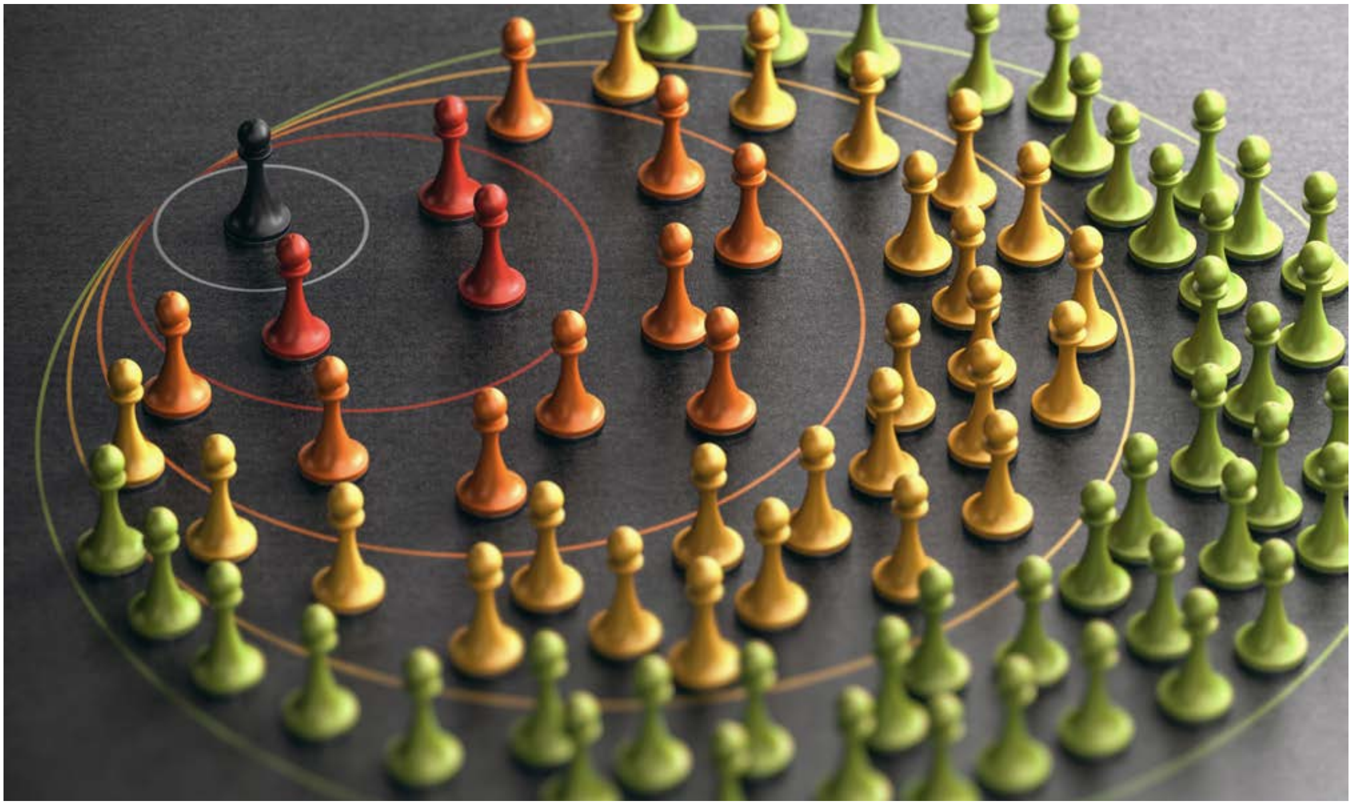

Espionage in Our AI Future

Why Human Intelligence Still Matters

Thomas Mulligan

Thomas Mulligan is a researcher at the RAND Corporation. He served in the CIA from 2008 to 2014, including as a case officer in Latin America.

The prospect that artificial intelligence might transform intelligence work is not new. In 1964, CIA was worrying about Soviet advances into "artificial-intelligence" and their national security implications (CIA 1964).[a] This early AI research was similar in nature (though not, of course, sophistication) to contemporary goals and techniques, including pattern recognition and machine learning. CIA observed, for instance, that the Soviets were using supervised machine learning to rapidly assess the seriousness of burn injuries. As AI technology has matured and diffused throughout our society, consideration of how to improve intelligence with it has quickened. This work has focused on the ramifications of AI for intelligence analysis.[b]

[a] Moran et al. 2023, and O'Connor 2023 provide history of the use of AI by the IC.

[b] See, e.g., Borene 2023, Galascione 2023, Gartin 2019, Gleeson 2023, and Neuberger 2025.

All statements of fact, opinion, or analysis expressed are those of the author and do not reflect the official positions or views of the US government. Nothing in the contents should be construed as asserting or implying US government authentication of information or endorsement of the author's views.

Much less attention has been paid to the ramifications for human intelligence (HUMINT) operations.[a] But frontier (i.e., state-of-the-art) AI is already changing how HUMINT collectors—case officers (COs)—do their job. And some people believe that emerging technology, including AI, threatens the entire HUMINT enterprise (*cf.* Ignatius 2025).[b]

While AI will transform HUMINT (along with most everything else), this, our oldest form of intelligence collection, will in fact grow in importance. For one thing, as AI makes high-quality technical collection cheaper and more accessible, it thereby boosts HUMINT's value on the margin. AI will supercharge disinformation and fabrication—and that makes HUMINT's ability to build and test source reliability over time, and corroborate technical collection, more important than ever. And as AI undermines the security of electronic communications, tradecraft techniques which COs have used for millennia—such as dead drops and brush passes—will find new relevance as I argue in this article.

In Section 1, I discuss the paper's scope and assumptions, and explain why its conclusions are robust to many future paths of technological development. In Section 2, I analyze five ways which HUMINT collection is, or soon will be, changed by AI. The lessons of this section are that we need to:

- Equip COs with AI-related requirements and encourage them to aggressively pursue relevant targets;
- Be attentive to the possibility of AI being used by fabricators;
- Recognize that while AI is a boon to counterintelligence surveillance and will make the lives of COs in the field more difficult, there are limits to what AI can do;
- Use AI for asset validation;
- Consider how AI can improve COs' ability to persuade agents, developmentals, and others.

Some of the ideas presented in Section 2 are not new. I mean to point out their relevance in the context of HUMINT collection. For instance, researchers and practitioners have looked at using AI to predict crime.[c] Those same tools may be used by counterintelligence services to predict where COs, operating in the field, will conduct operational acts. (Which, espionage being criminal nearly everywhere, amounts to the same thing.) In Section 3, I argue that AI will make technical intelligence collection cheap and accessible and flood the information environment with disinformation and noise. It may also undermine the security of electronic communications. All of these outcomes make HUMINT more important, not less. I conclude in Section 4.

Two preliminary notes. First, my topic is AI and espionage—that is, the collection of secret information by COs from foreign agents (also known as assets or sources). COs may of course engage in other activities, such as covert action. The ramifications of AI for those activities will not be considered here, but they are profound.

Second, this article is normative, not positive. That is, it is about what people involved in the HUMINT enterprise ought to be doing vis-a-vis frontier/future AI. It makes few claims about what is happening in terms of contemporary HUMINT operations and AI.

[a] Perhaps this is beginning to change. While this paper was under review, the Special Competitive Studies Project's "The Digital Case Officer" white paper (SCSP 2025) was released, outlining an AI-augmented "human-machine team" approach to HUMINT and discussing related governance considerations.

[b] I thank Emma Borden, Matt Chessen, John Parachini, Joel Predd, and Cortney Weinbaum for helpful suggestions on an earlier draft of this paper, and Emily Lawson for her valuable research support. I am also grateful to the CIA's Prepublication Classification Review Board for its timely reviews of this paper.

[c] See, e.g., Dakalbab et al. 2022 and Faqir 2023.

1. Speculating about Our AI Future

The future of AI is highly uncertain. People have different views, defended at different levels of rigor, about many things, such as how AI capabilities will grow (exponentially? linearly? to some point and no further?) and the extent of AI diffusion. The world will look very different depending on where we land, and when, on those and other dimensions. One might ask, in which of those future worlds do the arguments of this paper hold?

The conclusions of Section 2 are rooted in technology which is already viable or clearly on the technological frontier. Readers interested in technical detail can consult the references provided. Of course, what is now apparent might be a mirage. Maybe there is some unforeseen barrier to highly persuasive AIs.[a] So, when it comes to AI capabilities that are not already viable, I am assuming that the relevant technological development will continue apace.

When it comes to Section 3 and its conclusion about HUMINT's ongoing relevance, things are more speculative—although not that much more. We do not know if AI will, ultimately, totally undermine electronic encryption. But it is a threat to electronic encryption, because AI is already being used to create deepfakes, which are in turn augmenting phishing cyber attacks, which are a way to evade encryption.

I do assume, for most of Section 3,[b] that cataclysmic outcomes (e.g. human extermination/total subjugation at the hands of AI) do not come to pass, and in our AI future human beings retain some access to information of intelligence value. Otherwise, the paper does not rely on assumptions about how technological or social development will proceed.

A further point in favor of the robustness of Section 3's findings is that the three arguments I give there are largely independent. The truth of one does not require the truth of another. And they are largely divorced of "common causes" (Reichenbach 1956) such that if one is wrong then the others likely are too. For instance, suppose that AI ends up strengthening encryption rather than undermining it. That hardly threatens the ideas that HUMINT can provide a marginal advantage in the face of technological democratization or that it can break through the disinformation problem. The real challenges to what I conclude in this paper must stem from skepticism about the soundness of my arguments, then, rather than from speculation about the path and extent of technological development.

2. Operational Implications of Frontier AI

Artificial intelligence has already affected HUMINT operations, and future advances will change them more. Five considerations are salient.

2.1 AI Requirements

To do their job, COs must understand what information is valuable to policymakers and other intelligence consumers. Common sense goes a long way; all COs know that if they come across information on terrorism or the Chinese military, they should pass it along, and that they should cultivate sources with that access. When it comes to more specialized topics—say, an obscure piece of dual-use technology—there are mechanisms to push collection tasking to COs in the field.

These mechanisms are insufficient for dynamic, cutting-edge

[a] If I had to speculate, the argument would go like this: AI-derived language will inevitably have a recognizably non-human quality about it; there will be a backlash against AI due to its bad effects, leading to a stigma against its use; therefore, AI-derived "persuasive" content will make speech less compelling—not more. [One reason I think this argument is unsound is because I think the first premise is false.]

[b] Section 3.3 does not require the latter assumption (ii). The focus there is HUMINT tradecraft, independent of its typical use, which is to collect intelligence from human sources. In other words, that argument for HUMINT's enduring relevance is slightly more general than the others given.

technology like AI. Further, they are fundamentally reactive. In response to tasking, a CO can seek out the requested information. But in the absence of it, if he happens to come across this information, he will not recognize its intelligence value and will not report it.

The problem is acute when it comes to AI for two reasons. First, much of the information is technical, and couched in unfamiliar jargon. COs' ears should perk up if they encounter chatter about recurrent neural networks, or extreme ultraviolet lithography, or AI-enabled protein design. But would they?

Second, there's an enormous diversity of AI-related information with intelligence value. This breadth is hard for an untrained person to appreciate. Intelligence consumers are (or should be) interested in topics such as:

- Semi-conductor supply chains
- Military uses of AI
- Use of large language models (LLMs) in biological warfare
- AI industrial policy
- Deepfakes and disinformation
- Loss of control incidents
- Frontier model performance
- Model security
- Export control evasion
- Novel sources of training data
- Integration of AI into critical infrastructure.

It would be useful to have a few COs out there who are AI experts. But it's more important that all COs—no matter where they serve or what functional mission they concentrate on—have an AI training baseline. This could be provided at the Field Tradecraft Course (the core training/certification course for COs), on-demand in the field, or—best of all—both.

Case officers do not need in-depth technical knowledge to be effective AI collectors (cf. Katz 2020: 6). They do need familiarization with AI-related concepts, terminology, and technology, as well as regularly updated requirements from consumers. That's enough to develop and recruit new AI sources and to recognize AI-related intelligence when it arises. When technical expertise is needed, headquarters can provide it.

COs should be aware of the unusual motivations that AI developmentals might have. The classic pillars of money, ideology, compromise, and ego (known in espionage circles by the acronym MICE) remain. But there are important subtleties. Many potential sources are conflicted about their AI work. This is, in part, a result of uncertainty about the future development of AI, the dangers it may pose, and how it may transform fundamental human institutions, like work and relationships.

Some developmentals may find a traditional pitch compelling ("let's make sure the United States and the West—not China—win the AI war"). But for others, these geopolitical considerations are trivial in light of the existential risk AI poses. They will want to hear that only the United States can ensure that future AI is developed with care and effectively aligned with human values.

2.2 AI-enabled Fabricators

Agents may already be using AI to fabricate information. Fabrication is not a new phenomenon, of course, but AI makes it easier to fabricate and that fabrication more difficult to detect.[a] Fabricators can use LLMs to brainstorm and generate false but plausible information to pass to their COs. An agent can put into an LLM names, organizational structures, the substance of peers' work, and his own past reporting, and get out plausible stories to pass off as bona fide intelligence.[b]

[a] There are several reasons agents might fabricate, including loss of access, fear of termination, desire for more remuneration, or doubling against the service that recruited them.

[b] On the use of LLMs to generate misinformation in healthcare, see Zhou et al. 2023.

Of course, using an LLM like this would be risky, especially for agents living in a surveillance state. But savvy agents could install local AI models, fine-tune them on information like that described above, and generate false intelligence before meeting with their COs. Indeed, by sprinkling some detail about COs' typical lines of questioning and personality, fabricators could produce an even more persuasive product (Section 2.5).

The problem is not limited to existing agents. There are plenty of foreigners who (i) work for organizations of intelligence interest, like government ministries; (ii) have titles that plausibly entail access (e.g., program manager); and (iii) would love a new source of income. An LLM can generate a backstory for an enterprising fabricator and serve as a source of ongoing, unlimited, and false information.

Methods exist to detect fabrication, but AI makes the problem more prevalent and more resistant to those methods. Not many people have the intelligence and storytelling skill necessary to adopt a persona with the goal of serving as an intelligence agent; to maintain that persona by regularly providing false information; and to overcome vetting. With AI, the hurdles to this sort of behavior shrink.

Until now, it's been difficult for fabricators to create convincing photographs, videos, documents, and other media to support their false claims. When it has happened, it's been with the support of an intelligence agency (i.e. when the fabricator's been working as a double agent). But widely available generative AI models put this capability into every fabricator's hands. Methods for detecting deepfakes exist and are improving, but so are the countermeasures.[a]

2.3 Tradecraft Transformed

Artificial intelligence will transform COs' day-to-day work—how they spot, assess, develop, recruit, and handle agents. AI will, that is, transform tradecraft. Indeed, the transformation is already under way.

A principal challenge of HUMINT operations is surveillance. If a CO is discovered conducting an operational act (e.g., meeting an agent or placing a technical device), there are consequences. The CO is confirmed to be conducting intelligence activity and, in the best case, is expelled. The gentlest outcome for the agent is imprisonment. There is geopolitical blowback affecting other operations and equities. So COs are trained on and practice techniques to detect and defeat surveillance.

Effective surveillance against well-trained COs is costly. It can only be regularly mounted by our most capable adversaries. Until now, effective surveillance has meant multiple surveillance teams, each with multiple surveillants; vehicles and other equipment; and much watching and waiting. But technology—AI included—is driving down the cost of surveillance while increasing its scale and effectiveness. Cameras are a clear example. Placed outside of the home and work of known or suspected intelligence officers, and in transportation hubs, they monitor these officers' movements and help model their behavior. And they put anything in their field-of-view off-limits for operational acts (a CO cannot meet an agent if that meeting will be recorded).

COs have contended with technical surveillance for decades. Methods have been developed to defeat it, which in turn have birthed countermeasures, in the typical cat-and-mouse way. But AI threatens to realize the counterintelligence dream of comprehensive, stifling surveillance.[b] This is how David Ignatius (2025) describes the problem facing COs:

> *The tradecraft problem wasn't just pervasive surveillance, but the fact that data existed forever. . . . now, hidden cameras could monitor a case officer's meandering route to a dead-drop site and his location, long before and long after. His asset might collect the drop a week later, but his*

[a] See, e.g., Abbas & Taeihagh 2024, Balafrej & Dahmane 2024, Heidari et al. 2024, and Singh et al. 2025.
[b] Cf. Katz 2020: 5, Neuberger 2025, SCSP 2025, and Syllaidopoulos et al. 2025.

> *movements would be recorded, before and after, too. Patterns of travel and behavior could be tracked and analyzed for telltale anomalies. Even when spies weren't caught red-handed, they could be caught.*

The problem, though, isn't just the extent or persistence of data. The problem is that AI can quickly and automatically mine that data—nearly all of which is innocuous. Until now, it's been labor-intensive (not to say impossible) for security services to grapple with the deluge of data and identify the few signals of espionage amongst all the noise.

With AI-enabled facial and pattern recognition, security services can, without human intervention, identify potential intelligence activity to subject to scrutiny. Audio intercepted from personal devices and fixed (perhaps covert) microphones can be rapidly processed for counterintelligence use. (Although as AI-powered lipreading technology matures, audio surveillance may be obsolete in some settings.)[a]

Already, cheap, miniature drones can be deployed to surveil each known or suspected CO. These are small enough to avoid detection, and when controlled by AI can operate around-the-clock, in coordinated swarms, ensuring no surveillance gaps. When a CO enters a building, the AI can identify all possible exits and deploy drones to cover them.

New technologies, like nano drones, are in the offing. Drones on the order of ~1 centimeter would not have to wait outside; they would simply, clandestinely, accompany the CO into the building. The actual surveillance technology is only part of the counterintelligence solution. Equally important is the AI which automates and controls the drone.

Artificial intelligence can be used to identify operational acts. As a surveilled CO passes another person, an AI system can evaluate that event as a potential brush pass.[b] The passerby's identity can be ascertained via facial recognition and extra surveillance upon him can immediately commence.[c]

Still, there are limits. Every day, in every city, scores of people casually toss paper cups on the ground. Nearly all of them are litterers. Occasionally, one is a CO conducting a dead drop. It's not clear how any surveillance system—AI-powered or otherwise—could discriminate between the two.[d]

There are two lessons for COs. First, there will be less and less tolerance of sloppy tradecraft. Imprecise movements and timing which today raise no serious counterintelligence concerns (but which, of course, ought to be avoided) could soon be catastrophic. If the dead drop is an SD card glued into a paper cup, then the CO needs to make sure the card is secure, so he can toss the cup like garbage. He can't place the cup carefully on the ground to ensure the card doesn't fall out. That behavior is unnatural—and AI can tell. The margin of error between operational activity and the civilian conduct it's meant to ape is narrowing.

Second, cover will increase in importance. A security service cannot investigate every act of littering to discover those few instances of operational activity. But it can investigate every act of littering by a known or suspected intelligence officer.

This suggests a growing role for non-official cover COs. It also means that protecting the cover of our officers will be more important than ever. Harold Nicholson, a HUMINT tradecraft instructor, infamously sold the names of CIA trainees to the Soviets. A

[a] See e.g., Yang et al., 2019, contains 718,000 samples of Mandarin speech that can be used to train AI lipreading models.

[b] Cf. Hussain et al., 2024 and Sengönül et al., 2023.

[c] Cf. Ionescu et al., 2020.

[d] Perhaps the AI could rely on "crowdsourced intelligence" ("CROSINT") to investigate. If there's a bit of on-the-ground investigation needed (a piece of garbage to inspect, a photo to be taken, etc.), that could be advertised to people in the area, who would do the job for a small fee. (On CROSINT, see, e.g., Hershkovitz 2020 and Zhang 2022.)

damaging revelation, to be sure, but manageable. A similar compromise in the AI era could be catastrophic.

Case officers should also consider how AI might improve their tradecraft. AI is already being used for site selection, and if it can help a business find a location for a new store (by analyzing traffic patterns, surrounding infrastructure, online commentary, and more), it can help a CO select an operational site and plan a surveillance detection route (SDR).

The caveat is that the very same tools may be used by security services to anticipate COs' behavior and defeat their SDRs.[a] The solution is not to ignore these tools. Nor is it to trust them entirely. The solution is, rather, to ensure that independent, idiosyncratic, *human* judgment is exercised when selecting sites and planning SDRs. That can break the AI stalemate between spy and counterspy.

2.4 Agent validation

Zachery Tyson Brown conjectures that AI will help "identify micro-expressions that may serve as tells during source interviews or interrogations" (2024: 4). And systems have been built which apparently do exactly that: Yuan et al. (2025) present an AI model able to distinguish between liars and truth-tellers with 98-percent accuracy. (That's in a lab setting; field performance would doubtless be less impressive.) The model relies on basic facial muscle movements ("facial action units"), eye gaze, head pose, and "micro-expressions" (transient, involuntary, and nearly imperceptible facial movements).

The same tools that may watch our COs can help vet our agents. It is expensive and dangerous to conduct human surveillance anywhere, but especially abroad, in a hostile country. AI-enabled drones can watch our agents cheaply and with plausible deniability—ensuring agents work where they claim, meet the people they say they do, and have the lifestyles they purport to.

Stations can already use AI to deal with walk-ins, who pose a dilemma. On the one hand, they must be taken seriously. Many of our (and other countries') most valuable agents began their clandestine relationship by walking-in. On the other hand, most walk-ins are deranged, fabricators, dangles, or otherwise unhelpful. We'd like to expose our COs only to walk-ins of the first type, but there's typically no way of determining which type a walk-in belongs to, and the risk of turning away an important source is too great. So we end up exposing our COs to people who are, one way or another, dangerous.

Stations should consider having initial contact with walk-ins be handled by an AI officer. Even current technology can help identify problematic walk-ins, who can then be safely turned away by non-intelligence personnel. As technology improves, sophisticated but problematic walk-ins—like dangles—may be identified. Such a system could, at the very least, help COs assess the risks and benefits of face-to-face meetings.

2.5 Persuasion

A core CO skill is persuasion. A CO must be able to persuade a target to meet, a developmental to become an agent, and an agent to continue providing secret information. One might think that persuasion is a quintessentially human skill, inapt for artificial replication. In fact, in specific, bounded settings, frontier AI models are already rivaling humans in the ability to persuade.[b] As Matt Chessen describes things, "human cognition is a complex system, and AI tools are very good at decoding complex systems. . . . When provided rich databases of information about us, machines will know our personalities, wants, needs, annoyances, and fears better than we know them ourselves." (2017: 2).

[a] An analogous case could be using AI to predict the location of terrorist attacks. See, e.g., Ding et al. 2017 and Olabajo et al. 2021.

[b] See, e.g., Huang & Wang 2023, Schoenegger et al. 2025, Spitale et al., 2023, and https://www.anthropic.com/research/measuring-model-persuasiveness (retrieved 19 May 2025). On the use of LLMs for persuasion and disinformation generally, see Jones & Bergen 2024.

One promising way to improve COs' ability to develop, recruit, and handle agents is through AI-enhanced micro-targeting.[a] This is using information about a person's beliefs, personality, and preferences to improve the effectiveness of communications with her.

Of course, all good COs do this already. Part of being a good communicator is understanding one's audience and crafting one's speech accordingly. If COs know that an agent is an unemotional, logical type, they will rely more on rational arguments and less on appeals to emotion. Targeting (in this sense) has to this point been a mix of CO intuition, training, and guidance from headquarters.

Artificial intelligence can help. Imagine an LLM fine-tuned for persuasive power. In preparing for an agent meeting, a CO feeds into it public data like social media, proprietary information (operational cables), and relevant context (geopolitical facts). The model, in turn, provides guidance on how to communicate with the agent: what language to use; topics to avoid; objections the agent is likely to raise and how to respond to them; non-verbal measures like how to dress for the meeting, the location to choose, and amenities; and more.[b]

Such a model could be useful in other interactions, such as meetings with developmentals. COs today should consider using a (properly secured) LLM like they use their colleagues in station: as a sounding board for developing and pitching potential sources.

It's already possible for COs to access real-time, AI-generated communications guidance, via a clandestine earpiece or augmented reality glasses. And it's no longer fantastical to imagine COs equipped with a brain-computer interface serving in this role. It would recommend, like a video game, dialogue options for COs to consider as they pursue their operational goal (e.g., reminders about debriefing topics).

3. Human Intelligence in an Artificial World

HUMINT operations are an old-fashioned mix of art and science. Some tradecraft techniques have been used for millennia. It is tempting to think that, in an AI-saturated world, HUMINT will be a relic.

The opposite is likely true, for three reasons. First, AI will render technical intelligence collection cheap and widely accessible, thereby increasing the value of HUMINT collection, with its relatively high barriers-to-entry. Second, AI will be used to overwhelm digital environments with disinformation. HUMINT will have a unique ability to find the valuable intelligence signal amongst all that noise. Third, AI may undermine the trustworthiness of electronic communications. If that happens, we will need non-electronic ways to communicate which are simple and secure. Traditional agent communication techniques—dead drops, brush passes, brief meetings—are precisely that.

3.1 HUMINT on the Margin

Technology tends to be democratizing. It rapidly diffuses knowledge and capabilities. The printing press is a classic example. More recently, we have seen this with encryption. It was once expensive and difficult to encrypt electronic communications. Only governments and well-resourced companies could manage it. Now, anyone can download practically unbreakable encryption programs like Pretty Good Privacy for free. A ramification is that electronic encryption—while still vitally important—does not provide the relative advantage it once did.

Artificial intelligence will similarly democratize technical intelligence collection, making it easier and cheaper. First-rate efforts will be more common. We have already seen this with geospatial intelligence (GEOINT). This collection discipline (an important part of it,

[a] See, e.g., Matz et al. 2024, Salvi et al. 2025, SCSP 2025, and Simchon et al. 2024.
[b] Cf. SCSP 2025.

anyways) was once available only to nations capable of spaceflight. But now, high-quality imagery is available for free on the Internet, as is software (some of which uses AI) to process and analyze it. And if some desired imagery isn't available, anybody can task high-resolution (~30 cm) commercial satellites to get it.

The point is not that technical intelligence will be unimportant, but that AI will make it much easier to conduct collection comparable in scale and quality to our own. It will be harder to gain an advantage over our adversaries in these collection disciplines, and so the relative value of HUMINT will go up.

There's an analogy in the betting markets for horse racing. These are parimutuel markets in which there is no "house edge"; bettors compete only against each other. It is therefore possible to be a long-term winner.[a] Traditionally, the best bettors were experienced horseplayers, who could instinctively evaluate races. They could detect, for instance, when a horse was sick or subtly injured. But in the 1980s, racing-related data and statistical methods proliferated. Smart gamblers leveraged those tools to beat their peers, reaping enormous profits.[b]

The old ways are being rediscovered. Data, statistical models, and computing power are accessible, cheap, and part of any serious bettor's repertoire. Their outputs are priced into the markets, and possession of them provides bettors no advantage over their peers. Put differently, the insights generated by these methods are necessary, but not sufficient, for profitability. The marginal advantage—which even if small can be decisive—is increasingly coming from the insight which instinct and other *qualitative* capacities provide.

Brown (2024) says that "no matter how large your model may be, it will never encompass the world . . . there is no amount of data that will permit the forecasting of novel events in an increasingly complex competitive environment wherein innumerable threads, material and immaterial, sympathetic and antagonistic, are all wound together in a Gordian knot of causality." (3). Even if AI can massively improve the amount and quality of intelligence we can collect and produce—as seems likely—there will remain a critical residual which only HUMINT can obtain.[c] This could include information stored in air-gapped systems; foreign leadership intentions; and information on the existence and operation of AI "off-switches," which would be concealed from AIs because their purpose is to mitigate loss-of-control incidents involving those AIs. And when it comes to geopolitical competition—as with horse racing—a small advantage may make the difference between riches and ruin.

3.2 Human Signal, Electronic Noise

Future AI will be a fount for unlimited and effective disinformation.[d] We are, of course, already dealing with this problem, and its scale and severity are increasing. While HUMINT collection is not immune to AI-driven disinformation (Section 2.2), unlike the technical collection disciplines it will not be overwhelmed by it. For instance, there are worries about AI being used in a "fog of war" machine that floods the battlespace with disinformation and makes intelligence, surveillance, and reconnaissance impossible.[e] Such a machine would not imperil HUMINT operations; to the contrary, it would elevate them into the critical role of filtering out

[a] Technically, to be profitable, bettors must not only beat their peers, but also overcome the "track take" (a portion of the bets made which is extracted to pay for operating the races).
[b] See https://www.bloomberg.com/news/features/2018-05-03/the-gambler-who-cracked-the-horse-racing-code (retrieved 13 May 2025) for a prominent example of this.
[c] Cf. SCSP 2025.
[d] See, e.g., Lucas et al. 2024.
[e] See Geist 2023.

intelligence from the AI-generated chaff.

Or suppose that, within some adversarial nation, every day there are X phone calls of intelligence value. As a counterintelligence measure, our adversary uses AI to generate 10X daily deepfake calls. Indistinguishable from the real thing and containing carefully crafted disinformation, AI could render signals intelligence (SIGINT) efforts useless, even counterproductive. But a human agent can help identify which calls are bona fide and which are not.

The two collection efforts would, then, work in tandem. AI-powered SIGINT could obtain the content of phone calls but not insight into which calls were real and which were fake. A human agent might know that a call had been made but not what was said. HUMINT has long been used to corroborate intelligence gained through technical methods. Given the growing disinformation problem, this role will become more prominent.

The problem might be even simpler. Advanced AI could create a state of cognitive overload, which causes us to throw up our hands in the face of an unending deluge of plausible-looking information.[a] Overwhelmed by electronic intelligence—all kinds of media, containing some indeterminate mix of truth and falsity, derived from both real-world and synthetic data, and so on—HUMINT would, at least, be manageable.

3.3 Secure Human-to-Human Communication

While contemporary technology has affected how COs communicate with their agents, it has supplemented, rather than replaced, traditional techniques. This will continue to be the case.

For one thing, AI is already making it difficult to discriminate between truth and falsity in electronic communications. The quality of deepfakes is such that, to the untrained eye (and increasingly to the trained eye as well), it is impossible to differentiate between genuine images, videos, and audio and their deepfake counterparts.[b]

Scammers are using deepfakes to defraud and extort people—for instance, by generating audio which sounds like a family member pleading for ransom to (non-existent) kidnappers. These scams have succeeded not just against credulous people but sophisticated businesses as well.[c]

At the same time, hallucinations—the presentation of specious information by AI systems—persist. The rational response to these and similar dynamics is to treat digitally-mediated messages as lower-trust by default. If my friend tells me, face-to-face, that he is in trouble and needs money, I can be confident that that's true. But if that message is mediated by an electronic system—an e-mail, a phone call, a video attachment—it's more likely a scam than a bona fide plea for help.

A central, critical part of HUMINT tradecraft involves human-to-human communication unmediated by electronics. A properly-executed dead drop both (i) securely transfers information from the agent to the CO and (ii) gives the CO confidence that the information received is, indeed, being provided by his agent. Of course, the information could be false because the agent is lying or simply mistaken. But that is a categorically different concern than the one deepfakes and other AI dynamics raise: the injection of a new and powerful source of noise into the signal sent from agent to CO.

The idea of micro-targeting is related. Consider the following simple model of communication: Diane has some belief B in a proposition p (which could be, e.g., the proposition that the attack will be launched

[a] See, e.g., Lahlou 2025.
[b] See, e.g., Diel et al. 2024.
[c] See, e.g., https://www.wired.com/story/youre-not-ready-for-ai-powered-scams/ (retrieved July 24, 2025) and https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk (retrieved July 24, 2025).

tomorrow).[a] Diane might think the proposition is true, or false, or 30-percent likely, or whatever. She receives some speech, S, from an interlocutor which is relevant to the truth of p. Diane's task, now, is to reach a maximally justified belief about p given B and S.

Her interlocutor's speech S can be decomposed into a part that is correlated with the truth of p and a part that is pure persuasion. Diane only receives S, though, which is the sum of the two parts.

As S is dominated by its purely persuasive part—exactly what micro-targeting is designed to achieve—Diane ought, rationally, to reduce the extent to which she modifies B in light of S. Intuitively, Diane cannot be sure whether she finds S compelling because it reflects underlying truth or because her interlocutor is silver-tongued. So, as micro-targeting becomes more common and more effective, it may be rational to reduce trust in any communications that are not in-person.[b]

Traditional HUMINT techniques may also matter for a second reason: the possibility of "loss of control" incidents involving future AI. Imagine AI systems which rival or surpass human abilities, but which are misaligned with the values or goals we sought to instill in them. These systems could pose a threat—even an existential one—to human beings. Contemplation of these scenarios is no longer fringe speculation; it's a topic of increasing interest among serious researchers and policy analysts.[c]

In the event of a loss-of-control incident, it would be challenging for humanity to coordinate a response. Our standard (electronic) methods of communication—phones, emails, and so on—would be compromised by AI. HUMINT tradecraft would remain viable.

Finally, it's possible that AI will undermine encryption, thereby reducing—perhaps drastically—the security of electronic communication.[d] AI could identify implementation flaws in encryption protocols. It could increase the effectiveness of current techniques, like phishing (e.g., by using deep fakes). It could even—perhaps in tandem with advances in quantum computing—undermine encryption at a theoretical level.

If any of that happens, then the relative value of HUMINT tradecraft goes up. For it secures information in a categorically different way, resistant to those technological developments.

4. Conclusion

At the outset, I argued that we need not fret about the lower bound on the technology necessary for this paper's conclusions; the technology is here already or will be soon. But what about an upper bound? Might AI become so powerful and widespread that my earlier judgment—that HUMINT will continue to be relevant—will not hold? So long as the two explicit assumptions of Section 1 are satisfied, even in futurist worlds of human marginalization—in which AIs, rather than humans, increasingly have access to intelligence—the conclusion of Section 3 retains force.

Observe, for instance, that the argument for HUMINT's marginal value is compatible with the overall supply of HUMINT going down (as humans are replaced by AIs) and AI-derived intelligence being better than HUMINT. Indeed, that's the point: Value is determined not by total supply and demand, but by supply and demand at the margin. That marginal unit of HUMINT collected by our future CO, working in the lonely

[a] Formal treatments in the epistemology of disagreement can be adapted to model this kind of source-weighting problem. See, e.g., Mulligan 2021.

[b] Futurists have sometimes described this state as one of "epistemic collapse" or "knowledge collapse." See, e.g., Peterson 2025.

[c] For an evocative and technically informed description of how they might come to pass, see https://ai-2027.com/ (retrieved 23 May 2025). See also Somani et al. 2025.

[d] On this possibility, see, e.g., Bao et al. 2022, Benamira et al. 2021, and Gohr 2019.

shadows of an artificial world, may make all the difference.

HUMINT's role in cutting through synthetic noise holds as long as humans retain access to relevant information. The fewer humans with this access, then (all else equal) the lower value of the overall HUMINT effort. But this is not an objection to the argument, per se; it is an observation about its restricted applicability in a class of AI futures.

The considerations related to HUMINT communication techniques have greater force if more futurist scenarios, involving AI dominance over human beings, come to pass. Regular people don't use HUMINT tradecraft to communicate, even for sensitive information. There are better options. In the event of a loss-of-control incident, there might not be.

It is challenging to opine about our AI future. But across plausible futures, three points seem robust: AI is an extraordinary technology, perhaps without historical precedent. Much will change as it is integrated into HUMINT operations. But much will remain the same, and it's unlikely that AI will render HUMINT redundant. The work of the CO will remain recognizable in our AI future. ■

Works Cited

Abbas, Fakhar & Araz Taeihagh, "Unmasking Deepfakes: A Systematic Review of Deepfake Detection and Generation Techniques Using Artificial Intelligence," Expert Systems with Applications 252, part B (2024): 1-38, https://doi.org/10.1016/j.eswa.2024.124260.

Balafrej, Ismael & Mohamed Dahmane, "Enhancing Practicality and Efficiency of Deepfake Detection", *Scientific Reports* 14 (2024): 1-11, https://doi.org/10.1038/s41598-024-82223-y.

Bao, Zhenzhen, Jian Gou, Meicheng Liu, Li Ma, & Yi Tu, "Enhancing Differential-neural Cryptanalysis", in *Advances in Cryptology*—ASIACRYPT 2022, eds. Shweta Agrawal & Dongdai Lin (Springer, 2022), 318–47.

Benamira, Adrien, David Gerault, Thomas Peyrin, & Quan Quan Tan, "A Deeper Look at Machine Learning - based Cryptanalysis," in *Advances in Cryptology—EUROCRYPT 2021*, eds. Anne Canteaut & François-Xavier Sandaert (Springer, 2021), 805-835.

Borene, Alice B., "'This Piece Was Written by a Machine': Intelligence Analysis, Synthesis, and Automation," *Studies in Intelligence* 67, no. 4 (2023): 21–24.

Brown, Zachery Tyson, "'The Incalculable Element': The Promise and Peril of Artificial Intelligence", *Studies in Intelligence* 68, no. 1 (2024): 1–7.

Chessen, Matt, "The MADCOM Future: How Artificial Intelligence will Enhance Computational Propaganda, Reprogram Human Culture, and Threaten Democracy . . . and What Can Be Done about It." The Atlantic Council, 26 September 2017. https://www.atlanticcouncil.org/in-depth-research-reports/report/the-madcom-future/.

CIA, "Artificial-intelligence Research in the USSR", Office of Scientific Intelligence report 64-37 (1964). https://www.cia.gov/readingroom/docs/artificial%20intelligence%20r[15424923].pdf.

Dakalbab, Fatima, Manar Abu Talib, Omnia Abu Waraga, Ali Bou Nassif, Sohail Abbas, & Qassim Nasir, "Artificial Intelligence & Crime Prediction: A Systematic Literature Review", *Social Sciences & Humanities Open* 6, no. 1 (2022): 1–23, https://doi.org/10.1016/j.ssaho.2022.100342.

Diel, Alexander, Tania Lalgi, Isabel Carolin Schröter, Karl F. MacDorman, Martin Teufel, & Alexander Bäuerle, "Human Performance in Detecting Deepfakes: A Systematic Review and Meta-analysis of 56 Papers," *Computer in Human Behavior Reports* 16 (2024): 1–13, https://doi.org/10.1016/j.chbr.2024.100538.

Ding, Fangyu, Quansheng Ge, Dong Jiang, Jingying Fu, & Mengmeng Hao, "Understanding the Dynamics of Terrorism Events with Multiple-discipline Datasets and Machine Learning Approach," PLOS ONE 12 (2017): 1–11, https://doi.org/10.1371/journal.pone.0179057.

Faqir, Raed S. A., "Digital Criminal Investigations in the Era of Artificial Intelligence: A Comprehensive Overview", *International Journal of Cyber Criminology* 17, no. 2 (2023): 77-94.

Galascione, John F., "The End of Human Intelligence Analysis—Better Start Preparing," *Studies in Intelligence* 67, no. 4 (2023): 17–20.

Gartin, Joseph W., "The Future of Analysis," *Studies in Intelligence* 63, no. 2 (2019): 1–5.

Geist, Edward. *Deterrence under Uncertainty: Artificial Intelligence and Nuclear Warfare*. Oxford University Press, 2023.

Gleeson, Dennis J., "Artificial Intelligence for Analysis: The Road Ahead", *Studies in Intelligence* 67, no. 4 (2023): 11–15.

Gohr, Aron, "Improving Attacks on Round-reduced Speck32/64 Using Deep Learning,"in Advances in Cryptology—CRYPTO 2019, eds. Alexandra Boldyreva & Daniele Micciancio (Springer, 2019), 150–79.

Heidari, Arash, Nima Jafari Navimipour, Hasan Dag, & Mehmet Unal, "Deepfake Detection Using Deep Learning Methods: A Systematic and Comprehensive Review", WIREs Data Mining and Knowledge Discovery 14, no. 2 (2024): 1–45, https://doi.org/10.1002/widm.1520.

Hershkovitz, Shay, "Crowdsourced Intelligence (Crosint): Using Crowds for National Security", *International Journal of Intelligence, Security, and Public Affairs* 22, no. 1 (2020): 42–55, https://doi.org/10.1080/23800992.2020.1744824.

Huang, Guanxiong & Sai Wang, "Is Artificial Intelligence More Persuasive than Humans? A Meta-analysis", Journal of Communication 73, no. 6 (2023): 552-62, https://doi.org/10.1093/joc/jqad024.

Hussain, Altaf, Samee Ullah Khan, Noman Khan, Mohammad Shabaz, & Sung Wook Baik, "AI-driven Behavior Biometrics Framework for Robust Human Activity Recognition in Surveillance Systems," *Engineering Applications of Artificial Intelligence* 127, part A (2024): 1–15, https://doi.org/10.1016/j.engappai.2023.107218.

Ignatius, David, "A Band of Innovators Reimagines the Spy Game for a World with No Cover," *Washington Post*, July 10, 2025. https://www.washingtonpost.com/opinions/interactive/2025/cia-ai-technology-spies/.

Ionescu, Bogdan, Marian Ghenescu, Florin Rastoceanu, Razvan Roman, & Marian Buric, "Artificial Intelligence Fights Crime and Terrorism at a New Level", *IEEE MultiMedia* 27, no. 2 (2020): 55–61, https://doi.org/10.1109/MMUL.2020.2994403.

Jones, Cameron R. & Benjamin K. Bergen, "Lies, Damned Lies, and Distributional Language Statistics: Persuasion and Deception with Large Language Models", *arXiv*, 2024, 37, https://doi.org/10.48550/arXiv.2412.17128.

Katz, Brian, "The Collection Edge: Harnessing Emerging Technologies for Intelligence Collection," *CSIS Brief* (2020). https://www.csis.org/analysis/collection-edge-harnessing-emerging-technologies-intelligence-collection.

Lahlou, Salem, "Mitigating Societal Cognitive Overload in the Age of AI: Challenges and Directions," *arXiv*, 2025, 13, https://doi.org/10.48550/arXiv.2504.19990.

Lucas, Jason S., Barani Maung Maung, Maryam Tabar, Keegan McBride, & Dongwon Lee, "The Longtail Impact of Generative AI on Disinformation: Harmonizing Dichotomous Perspectives," *IEEE Intelligent Systems* 39, no. 5 (2024): 12–19, https://doi.org/10.1109/MIS.2024.3439109.

Matz, S. C., J. D. Teeny, S. S. Vaid, H. Peters, G. M. Harari, & M. Cerf, "The Potential of Generative AI for Personalized Persuasion at Scale," *Scientific Reports* 14: 1–16 (2024), https://doi.org/10.1038/s41598-024-53755-0.

Moran, Christopher R., Joe Burton, & George Christou, "The US Intelligence Community, Global Security, and AI: From Secret Intelligence to Smart Spying," *Journal of Global Security Studies* 8, no. 2 (2023): 1–18, https://doi.org/10.1093/jogss/ogad005.

Mulligan, Thomas, "The Epistemology of Disagreement: Why Not Bayesianism?" *Episteme* 18, no. 4 (2021): 587–602, https://doi.org/10.1017/epi.2019.28.

Neuberger, Anne, "Spy vs. AI: How Artificial Intelligence will Remake Espionage," *Foreign Affairs*, January 15, 2025. https://www.foreignaffairs.com/united-states/spy-vs-ai.

O'Connor, Jack, "Undercover Algorithm: A Secret Chapter in the Early History of Artificial Intelligence and Satellite Imagery," *International Journal of Intelligence and Counterintelligence* 36, no. 4 (2023): 1337-51, https://doi.org/10.1080/08850607.2022.2073542.

Olabajo, Olusola A., Benjamin S. Aribisala, Manuel Mazzara, & Ashiribo S. Wusu, "An Ensemble Machine Learning Model for the Prediction of Danger Zones: Towards a Global Counter-terrorism," *Soft Computing Letters* 3 (2021): 1–6, https://doi.org/10.1016/j.socl.2021.100020.

Peterson, Andrew J., "AI and the Problem of Knowledge Collapse," *AI & Society* 40, no. 5: 3249–69 (2025), https://doi.org/10.1007/s00146-024-02173-x.

Reichenbach, Hans. *The Direction of Time.* University of California Press, 1956.

Salvi, Francesco, Manoel Horta Ribeiro, Riccardo Gallotti, & Robert West, "On the Conversational Persuasiveness of GPT-4", Nature Human Behavior (2025): 1–12, https://doi.org/10.1038/s41562-025-02194-6.

Schoenegger, et al., "Large Language Models Are More Persuasive than Incentivized Human Persuaders", *arXiv*, 2025, 30, https://doi.org/10.48550/arXiv.2505.09662.

SCSP (Special Competitive Studies Project), "The Digital Case Officer: Reimagining Espionage with Artificial Intelligence" (2025). https://www.scsp.ai/wp-content/uploads/2025/09/SCSP_The-Digital-Case-Officer_-Reimagining-Espionage-with-Artificial-Intelligence.pdf.

Sengönül, Erkan, Refik Samet, Qasem Abu Al-Haija, Ali Alqahtani, Badraddin Alturki, & Abdulaziz A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey," *Applied Sciences* 13, no. 8 (2023): 1–31, https://doi.org/10.3390/app13084956.

Simchon, Almog, Matthew Edwards, & Stephan Lewandowsky, "The Persuasive Effects of Political Microtargeting in the Age of Generative Artificial Intelligence," *PNAS Nexus* 3, no. 2 (2024): 1-5, https://doi.org/10.1093/pnasnexus/pgae035.

Singh, Laishram H., Panem Charanarur, & Naveen Kumar Chaudhary, "Advances in Detecting Deepfakes: AI Algorithm and Future Prospects—a Review," *Discover Internet of Things* 5, no. 53 (2025): 1-30, https://doi.org/10.1007/s43926-025-00154-0.

Somani, Elika, Anjay Friedman, Henry Wu, Marianne Lu, Chris Byrd, Henri van Soest, & Sana Zakaria, "Strengthening Emergency Preparedness and Response for AI Loss of Control Incidents," *RAND Europe Research Report* (2025). https://www.rand.org/content/dam/rand/pubs/research_reports/RRA3800/RRA3847-1/RAND_RRA3847-1.pdf.

Spitale, Giovanni, Nikola Biller-Andorno, & Federico Germani, "AI Model GPT-3 (Dis)informs Us Better than Humans," *Science Advances* 9, no. 26 (2023): 1–9, https://doi.org/10.1126/sciadv.adh1850.

Syllaidopoulos, Ioannis, Klimis S. Ntalianis, & Ioannis Salmon, "A Comprehensive Survey on AI in Counter-terrorism and Cybersecurity: Challenges and Ethical Dimensions," *IEEE Access* 13 (2025): 91740–64, https://doi.org/10.1109/ACCESS.2025.3572348.

Yang, Shuang, Yuanhang Zhang, Dalu Feng, Mingmin Yang, Chenhao Wang, Jingyun Xiao, Keyu Long, Shiguang Shan, & Xilin Chen, "LRW-1000: A Naturally distributed Large-scale Benchmark for Lip Reading in the Wild", *arXiv*, 2019, 8, https://doi.org/10.48550/arXiv.1810.06990.

Yuan, Shusen, Zilong Shao, Zhongjun Ma, Ting Cao, Hongbo Xing, Yong Liu, & Yewen Cao, "Deception Detection Based on Microexpression and Feature Selection Methods," *EURASIP Journal on Image and Video Processing* 8 (2025): 1–18, https://doi.org/10.1186/s13640-025-00674-3.

Zhang, Jing, "Knowledge Learning with Crowdsourcing: A Brief Review and Systematic Perspective," IEEE/CAA Journal of Automatica Sinica 9, no. 5 (2022): 749–62, https://doi.org/10.1109/JAS.2022.105434.

Zhou, Jiawei, Yixuan Zhang, Qianni Luo, Andrea G. Parker, & Munmun De Choudhury, "Synthetic Lies: Understanding AI-generated Misinformation and Evaluating Algorithmic and Human Solutions", in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, eds. Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, & Max L. Wilson (Association for Computing Machinery, 2023), 1–20.■