Analytic Tradecraft Standards in an Age of AI

Gerald M. McMahon

The Intelligence Project
Belfer Center for Science and International Affairs
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138

belfercenter.org/project/intelligence-project

Analytic Tradecraft Standards in an Age of AI
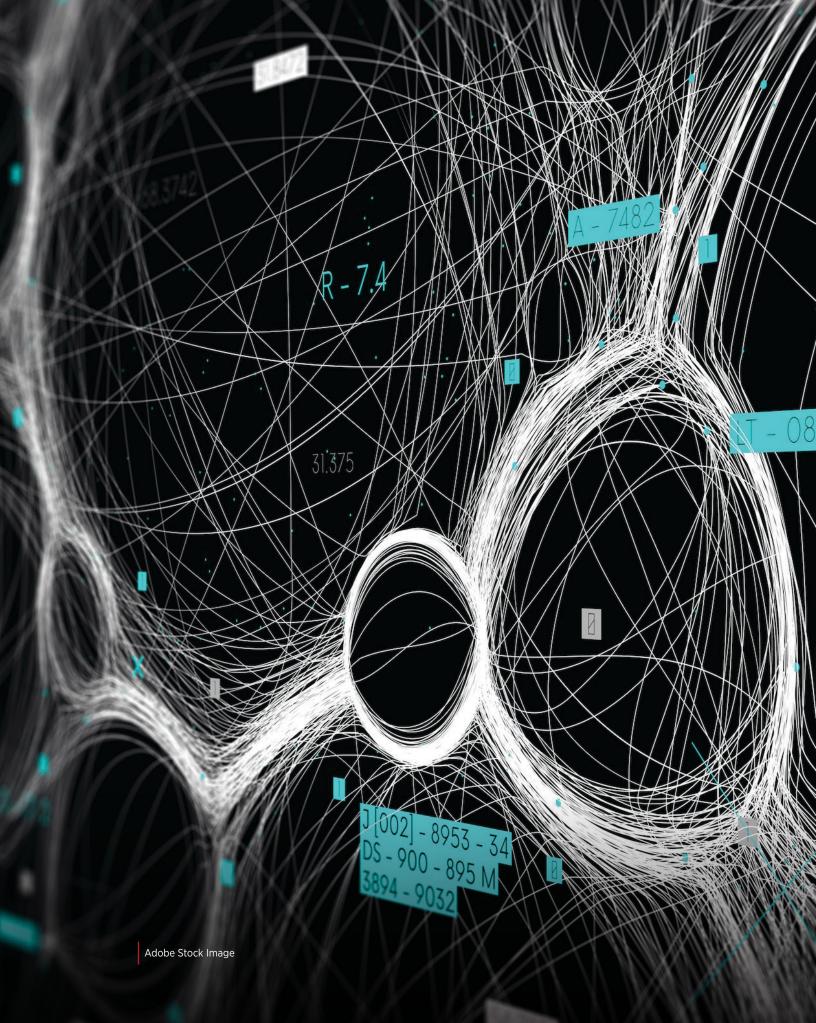
Gerald M. McMahon

About the Intelligence Project

The Intelligence Project seeks to build a new generation of intelligence practitioners prepared to serve in a rapidly changing world and to help future policymakers and intelligence consumers understand how best to interact with intelligence to gain a decision advantage. Building on multi-disciplinary research being conducted at the Belfer Center, from his- tory to human rights and cyber technologies, the Intelligence Project links intelligence agencies with Belfer researchers, Faculty, and Kennedy School students, to enrich their education and impact public policy..

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

ii

About the Author

Gerald M. McMahon is a senior intelligence analyst in the US Government and has been assigned to the Belfer Center's Intelligence Project as a Recanati-Kaplan Fellow for 2023-2024 academic year. His recent work includes conducting complex threat and risk analysis related to national security, criminal, and cyber threats targeting the homeland. McMahon's work and study have also focused on the intersection of emerging technology and national security.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

iii

Table of Contents

The AI Race is On

Intelligence agencies around the world are racing to leverage and adopt the capabilities of Artificial Intelligence (AI). AI tools have already been deployed or are in development by defense, intelligence and law enforcement for a range of functions to include image recognition (facial recognition,[1] object detection[2]), language translation,[3] and insider threat detection.[4] The FBI is using AI enabled technology to evaluate tips to ensure they are accurately identified, prioritized, and processed in a timely manner.[5] The CIA's Open Source Enterprise launched an internal "Chat-GPT style" AI tool to enable analysts to have "better access to open-source intelligence,"[6] and the NSA opened an "Artificial Intelligence Security Center,"[7] focused on defending "the Nation's AI through Intel-Driven collaboration with industry, academia, the [Intelligence Community] IC, and other government partners."[8, 9]

The US is not alone in pursuing the benefits of AI in the security space. China's intelligence services are already using AI to identify foreign intelligence officers[10] and the US is making efforts to deny China, and other adversaries, the ability to gain access and use the most advanced AI related technology and services.[11] Non-state actors, such as fraudsters, are leveraging AI to advance schemes through AI generated phishing attacks and voice-cloning enabled scams, which is likely just a sampling of the schemes to come.[12]

As the US moves to adopt AI technology in an expanding number of security applications, how can the IC best position its analysts to use this technology when it comes to intelligence analysis and how will use of these tools impact analysts' ability to meet existing analytic standards? This paper considers the implications for meeting IC Analytic Standards when using an AI tool, explores several of the key issues involved and makes recommendations for amending the current standards, taking steps to anticipate challenges, and set the conditions for the IC to be successful in adopting AI.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

1

Analytic Tradecraft Standards – Establishing Rules of the Road

In the wake of the faulty 2002 National Intelligence Estimate (NIE) on Iraq's WMD programs,[13] Congress passed the 2004 Intelligence Reform and Terrorism Prevention Act (IRTPA). This legislation, which incorporated findings from the WMD and 9/11 commissions, included a provision for "Assignment of Responsibilities Relating to Analytic Integrity."[14] IRTPA required the Office of the Director of National Intelligence (ODNI) to assign an individual or entity to be responsible for ensuring that finished intelligence products produced by the intelligence community are timely, objective, independent of political considerations, based on all sources of available intelligence, and employ the standards of proper analytic tradecraft." This led to the formalization of the IC's Analytic Standards, ultimately contained in Intelligence Community Directive (ICD) 203.[15]

ICD 203 identified five Analytic Standards and nine Analytic Tradecraft Standards (ATS),[16] which guide intelligence analysis across the IC (see Figure 1). Adoption of ICD 203 and ATS have been recognized as an important advancement in the professionalization of intelligence analysis[17] and scholars and practitioners have written to the benefit of incorporating ATS into analysis[18] at the tactical, operational and strategic levels.[19] With the advent of AI and its potential impact on the conduct and profession of analysis, the time may be right to reexamine ICD 203 to ensure it remains suited to guide analysts in the age of AI-based tools.[20]

Can AI Meet Analytic Standards?

Using technology to process and interpret records, data and information is nothing new for analysts. ICD 203 was drafted at a time when the use of databases and data analysis was routine. Though the tools which enable tasks such as collection and communication analysis have expanded and grown more complex, they have been a mainstay within the IC for decades.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

2

To determine if using a generative AI tool is so different as to warrant reassessment of current analytic tradecraft, this paper interrogates two of the nine[i] ATS standards: *ATS 3:* Properly distinguishes between underlying intelligence information and analysts' assumptions and judgments; and *ATS 4:* Incorporates analysis of alternatives.

Figure 1: ICD 203

ICD 203 identifies five (5) Analytic Standards and nine (9) Analytic Tradecraft Standards:

    a. Objective

    b. Independent of political consideration

    c. Timely

    d. Based on all available sources of intelligence information

    e. Implements and exhibits Analytic Tradecraft Standards, specifically:

  (1) Properly describes quality and credibility of underlying sources, data, and methodologies

  (2) Properly expresses and explains uncertainties associated with major analytic judgments

  (3) Properly distinguishes between underlying intelligence information and analysts' assumptions and judgments

  (4) Incorporates analysis of alternatives

  (5) Demonstrates customer relevance and addresses implications

  (6) Uses clear and logical argumentation

  (7) Explains change to or consistency of analytic judgments

  (8) Makes accurate judgments and assessments

  (9) Incorporates effective visual information where appropriate

Intelligence Community Directive 203; ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)

  (3) Properly distinguishes between underlying intelligence information and analysts' assumptions and judgments

  (4) Incorporates analysis of alternatives

  (5) Demonstrates customer relevance and addresses implications

  (6) Uses clear and logical argumentation

  (7) Explains change to or consistency of analytic judgments

  (8) Makes accurate judgments and assessments

  (9) Incorporates effective visual information where appropriate

Intelligence Community Directive 203; ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)

---

i   The author focused on these two standards due to both lending themselves more clearly to basic testing using a simple AI tool accessing a sample data set.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

3

ATS 3 – Properly distinguishes between underlying intelligence information and analysts' assumptions and judgments

Taking the first component of the standard, focused on the "underlying intelligence information," an analyst querying a generative AI tool (hereafter referred to as AI-T) would require several elements to ensure an ability to meet the expectations for ATS.

- *Replicability*: The AI-T must provide citations to source documents, allowing analysts to view the original material to verify and match the information presented. The analyst would have the ability to recreate the findings.

- *Fidelity to the source document*: The AI-T would have to clearly state what is taken directly from queried records, for instance using quotes to represent the exact language extracted from the underlying source document.

- *An ability to avoid characterization*: Unless prompted to do so, an AI-T requires the ability to produce insights that can refrain from characterizing facts in descriptive language, omitting any qualifiers or labels that may imply meaning where it had not existed. As an example, if an AI-T summarized a surveillance report, it must maintain an ability to present, as a component of the response, the facts of the report versus applying meaning to the observed behavior. As a hypothetical example, when prompted for an analysis of a surveillance team's observations of an espionage subject's activity, apart from a summary, the report should note direct elements of the team's reports. For instance, that the subject drove through stop lights, went the wrong way on a one-way street and stopped on the side of the road. These details can be used to support a summary

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

4

finding versus simply stating that the subject drove 'evasively' or employed counter-surveillance measures.

ATS 3 defines assumptions as "suppositions used to frame or support an argument," which "affect analytic interpretation of underlying intelligence information." Further, ATS 3 calls for products to "state assumptions explicitly when they serve as the linchpin of an argument or when they bridge key information gaps." Inputs from an AI-T would need to clearly articulate the premise for its arguments and do so in a manner that was obvious to the reader.

- *Context for characterization:* The AI-T must make explicit distinctions between facts and characterizations and provide the background that led to those characterizations. Building on the surveillance example, the AI-T could describe the facts from the report (ex. subject drove through red lights, etc.) and then indicate that this behavior may be indicative of counter-surveillance activity based on a cited source(s) which defines counter-surveillance conduct or provides examples. Citations provide the analyst an opportunity to review the foundation for the assumption, determine its relevance, reliability, and applicability to the current case.

- *Consistent but updated:* The AI-T's assumptions must be consistent throughout the project or have an ability to update based on new information. An analyst must be able to recognize that assumptions used by the AI-T on any given project are logically cohesive, not contradictory and respond to injects of new information which may challenge previously held assumptions. For instance, if, continuing with the above surveillance example, an additional surveillance report were added which determined that the driver of the car under observation was not the subject, the AI-T would adjust it's characterizations or findings based on this data.

Lastly,[ii] ATS 3 defines judgments "as conclusions based on underlying intelligence information, analysis, and assumptions." When using generative AI, an analyst would either need to establish a prompt that prescribes a format for the AI-T to follow or the tool would be designed to follow a similar format. Either option would require the AI-T to differentiate between each element of a judgment (the judgment, the underlying information and assumptions) and present a coherent

---

ii   ATS 3 notes that products "should explain the implications for judgments if assumptions prove to be incorrect" and, "as appropriate, identify indicators that, if detected, would alter judgments." The focus here is on the primary elements of ATS 3 – underlying information, assumptions and judgments."

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

5

statement that meets the structure outlined in ATS.[iii] For example, the prompt could be, "Based on the information in this library, your analysis, and your assumptions, provide your judgment as to what the subject's activity means in the context of an espionage investigation. In your response, distinguish between information in the library, your assumptions, and your judgment(s)."

- *Follow the format:* The AI-T must stick to the format of the prompt and consistently meet the previous enumerated requirements for ATS 3.

- *Make sense:* The judgment must be plausible, fit with the evidence and assumptions.

ATS 4 – Incorporates analysis of alternatives

ICD 203 defines an analysis of alternatives as "the systematic evaluation of differing hypotheses to explain events or phenomena, explore near-term outcomes, and imagine possible futures to mitigate surprise and risk." At minimum, to meet ATS 4 expectations,[iv] an AI-T would have to provide not only plausible alternative explanations to the primary judgments used to respond to ATS 3, but, like ATS 3, must describe "associated assumptions" that support this alternative.

- *Hits ATS 3 Criteria:* The alternatives presented must continue to adhere to the standards for ATS 3.

- *Make Sense:* Just as with ATS 3, the alternative explored must be plausible.

- *Explains the alternative's strengths or weaknesses:* The underlying intelligence information, assumptions, and analysis for this alternative were not sufficient to make it the primary judgment. The AI-T must be able to articulate why this alternative was less persuasive and where the weakness lies in either the evidence, the assumptions or the analysis.

---

iii The AI-T must also detail which assumptions serve as the "linchpin" of their argument "or when they bridge key information gaps." The ability of an AI-T to identify how an argument/judgment will survive without key information allows the tool and the user to understand the critical features of the line of analysis.

iv ATS 4 includes additional expectations, such as exploring implications to "U.S. interests," and identifying "indicators that, if detected, would affect the likelihood of identified alternatives." To develop simple, testable standards for the recommended experiment, the author focused on the ATS 4 elements noted herein.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

6

Discussion

To meet just two of the nine ATS standards[v] an AI-T must have an ability to: allow users to replicate and source the tool's findings; parse underlying information, assumptions and judgments; and maintain consistency in presenting its logic and reasoning. Without an ability to meet these conditions, analysts using such an AI-T are unlikely to meet ATS standards.

AI-Ts may well be able to meet these standards, but the exercise points to two key issues that impact the IC's use of AI and ICD 203: trust in the AI-T and ownership of the analysis.

Trust but Verify

Though much has been written about the vulnerabilities of Large Language Models (LLMs)[21] and their underlying information to data poisoning,[22] prompt injections,[23] and hallucinations,[24] the focus here is on use of the AI tools' output in assisting analysis. While AI and machine learning (ML) will clearly introduce "more effective methods for creating, exploiting, and sharing databases"[25] and reducing workloads on processing tasks, the implications for using generative AI[26] as a "co-pilot" to draft "early phases" of reports or developing lines of analysis are more opaque.[27] The distinctive advantage of generative AI, its ability to create new information and insights, is also what increases the risk to analytic integrity.[28] Analysts must not only correctly understand and interpret an AI tool's output, they have to have confidence that what it's provided is accurate, lest AI generated outcomes lead to new risks of intelligence failures.

These concerns are not hypothetical. Critics already point to AI enabled techniques, such as facial recognition and offender risk assessments[29], whose use has led to mistaken arrests[30] or disparate outcomes in sentencing, as warnings of mistakes to come.[31] Adopting AI tools provides exciting new opportunities but also highlights many of the same challenges intelligence analysts have faced since

---

v   This does not account for the entirety of the Analytic Standards. The five Analytic Standards require all analytic products be objective, independent of political consideration, timely, based on all available sources of intelligence information and implement and exhibit the nine ATS.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

7

the days of Sherman Kent.[32] Analysts are expected to be apolitical experts who deliver objective, timely, accurate and clear judgments to give leaders decision advantage.[33] AI tools hold the promise of making sense of the vast and growing amount of available data, get ahead of the threat and "connect the dots." At the same time, concerns exist that AI tools will lead to new risks for abuse,[34] and that, while the results provided by AI tools, such as ChatGPT, can appear convincing, they can also contain errors or be entirely fraudulent.[35]

A key component for analysts conducting research and evaluating evidence is trust. In this context, analysts must have confidence that the tools and systems they use allow them access to the underlying sources of the information for their results and that the analyst understands the reliability of that information and the potential for it to contain error and bias. As noted in the ATS 3 example, any AI tool must be able to "demonstrate the underlying rationale behind decisions and responses to both users and overseers."[36] For all concerned, the ability to understand and interrogate an AI system's output is crucial to having confidence in the findings and the tool. For any successful adoption of an AI tool, the analysts who use it will need an ability to understand and explain how results were achieved. Without that ability, trust and, ultimately, use of such tools are unlikely to be successful.[37]

That trust must also extend throughout the intelligence enterprise, to include decision-makers. As author Mark Lowenthal, the CIA's former Assistant Director for Analysis and Production, notes, the "goal of intelligence analysis is to reduce policymakers' uncertainties, not make their decisions for them."[38] That goal is served by policymakers having trust in the IC, in the analysts and the process. If there's a lack of understanding or concern regarding the AI tools embedded in the intelligence process, confidence in the system can erode.

Among the areas of focus for the IC to ensure trust in an AI-T enabled process will be education and tool design. At present, "Heads of IC elements," are required to provide training to their workforce on Analytic Standards.[39] The IC must develop a community-wide AI literacy program to provide baseline training for all intelligence professionals to understand the technology behind AI and how best to use AI tools while adhering to ICD 203.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

8

Additionally, the IC must ensure prioritization of both the design and implementation of AI-Ts that provides sufficient explainability to enable adherence to analytic tradecraft. This includes incorporating standards defined by the National Institute of Standards and Technology (NIST) for explainable AI. These standards include requiring an AI tool deliver "accompanying evidence or reasons for outcomes and processes; provide explanations that are understandable to individual users; provide explanations that correctly reflect the system's process for generating the output; and that a system only operates under conditions for which it was designed and when it reaches sufficient confidence in its output."[40] Whatever a tool's capabilities may be, without a user's trust, adoption and execution will fail. Providing analysts the capability and training to understand the tool's process, is crucial to that task.

The Time is Now to Update ICD 203 So It's Clear Who Owns the Analysis

Given the expectations and stakes, the human analyst must remain responsible for their assessments and be accountable to decision-makers for their analysis and any deliverable. As AI augmented analysis moves from concept, to testing, to implementation, the time is now to update ICD 203 to account for the impact this technology will have on analysts.

As written, ICD 203 does not explicitly speak to the role of technology in analysis. However, in two areas, the Directive's language is expansive enough to encompass any technology used in research and analysis. One relevant feature in ICD 203 which may address the use of AI is found in ATS 1, "Properly describes the quality and credibility of underlying sources, data, and *methodologies*." (emphasis added) Methodologies casts a wide enough net that it could include adding a descriptor or caveat, now common in academia, attesting to what role, if any, an AI tool played in producing a report.[41] One limitation here is that the analyst may not be aware of when or how an AI tool was used to generate the findings upon which their analysis is based. As noted previously, AI and ML are commonly used to process and organize information. Given the vast stores of data now subject to search,

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

9

retrieval and review by researchers, attesting to AI use in this environment would be uncertain at best.

The second is the top Analytic Standard for all analysts, to be "objective."[42] The standard notes that all analysts must be aware of "their own assumptions and reasoning" and "must employ reasoning techniques and practical mechanisms that reveal and mitigate bias." Further, the section notes that analysts must be "alert to influence by existing analytic positions or judgments."

Though the Directive is written to set expectations for analysts to be aware of their own biases, a fair reading would extend that awareness to other sources which may influence their final reasoning and judgments. This would apply to analysts' reading of previous products, with the Directive specifically stating analysis "should not be unduly constrained by previous judgments."

Controlling for bias is nothing new in the conduct of analysis. Further, contending with competing analytic lines and interpretations of evidence is a common feature of IC analytic production. As Thomas Fingar, the ODNI's first Deputy Director for Analysis and the individual responsible for developing and implementing ICD 203, notes, intelligence analysis is a collaborative process. It is commonplace for analysts to explore, dissect and coalesce around viewpoints.[43] Key to that process is controlling for bias. Even with this backdrop, the introduction of an AI tool necessitates amending existing standards to account for the advent of this new technology.

Generative AI adds a new element that can impact the key feature of the analyst's role, providing assessments and insights beyond what's known. Distinguishing from what a generative AI tool suggests and what an analyst ultimately thinks, may be increasingly difficult to disentangle. Several studies of knowledge workers' use of AI tools revealed both benefits and pitfalls. While some knowledge workers showed gains in qualitative and quantitative performance measures, others showed declines in quality, using AI for tasks where its capabilities lag, and an overreliance on the AI tool, leading workers to fall "asleep at the wheel." Many were shown to have allowed the AI tool to substitute its judgment for their own.[44] Such a dynamic within the IC could lead to catastrophic consequences.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

10

While leadership within the IC has been clear that AI tools won't replace analysts "anytime soon,"[45] given the potential risks involved, amending ICD 203 now to account for the potential use of an AI tool is sensible. Adding the following language to subsection to 6.a, "Objective," would be a worthwhile start:

> *When using any generative Artificial Intelligence (AI) tool to inform analytic judgments, analysts must rely on their own, independent judgment, consistent with Analytic Standards, to craft their assessment. Analysts using generative AI tools to develop assessments must be able to appropriately explain the reasoning for their judgments, the evidence which supports their analysis and the confidence in those judgments. Analysts will use generative AI tools consistent with the IC's Ethics Framework for Artificial Intelligence[46] and any AI governing policies for their agencies.*

Inclusion of this language will signal that despite the likely increasing use of AI tools in supporting analysis, the analyst remains accountable for their products and must be positioned to explain their reasoning. This language also limits the focus to the analyst's assessments and judgments, not the potentially large volume of data which may have been processed using AI/ML methods.[47]

In addition to amending ICD 203, the Deputy Director for Mission Integration (DD/MI), who is responsible for ensuring the IC abides by the Analytic Standards, would do well to institute two initiatives to position the IC for the challenges and opportunities AI tools will bring:

AI/Analytic Standards Working Group: Institute a standing IC working group to monitor, study and evaluate the impact of AI on Analytic Standards. This working group would also provide recommendations to the DD/MI on changes to ICD 203, training or related matters to ensure the IC remains current on the implications for use of these tools. The working group would also be positioned to study the efficacy of incorporating AI tools into analysis.[48]

Conduct a Pre-Mortem:[49] Empanel two multi-disciplinary committees, including representatives from the IC, technical, policy, legal, academic and business communities to examine possible challenges incorporating AI tools into intelligence analysis. Based on the committees' findings, the IC would

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

11

develop guidelines to forestall potential intelligence failures based on a flawed tool or flawed use of an AI tool.

1. Committee 1 – Builds scenarios that anticipate how an AI tool will be incorporated into the intelligence cycle, particularly as it pertains to intelligence analysis and production of analytic products. Committee 1 develops scenarios in which the AI-T is deployed to assist in assessing international or domestic flashpoints.

2. Committee 2 – Examines how employment of the AI-T may contribute to intelligence failures or challenges in properly assessing the topic. Committee 2's review will develop recommendations on how to avoid potential pitfalls when using AI tools, to include considering issues related to efficacy of the tool, matters related to public confidence and civil liberties.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

12

Conclusion

Given the pace of technological advancement with AI and the great potential for AI to provide the IC advantage, the time is now to provide clear guidance to the workforce on expectations when using AI-Ts. These guidelines will not only inform analysts on their responsibilities, but they will also shape the development and implementation strategies for new AI applications. Amending ICD 203, enabling AI literacy, properly designing and fielding AI tools that meet ATS, establishing a working group focused on the intersection between AI and Analytic Standards, and anticipating how things might go wrong, would give the IC a better chance of achieving success in an age of AI enabled analysis.

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

13

Endnotes

1   As an example related to facial recognition, see the Intelligence Advanced Research Activity's JANUS program, whose goal was to "revolutionize face recognition by fusing information available from multiple views from diverse sensors and visual media to deliver dramatic improvement in speed and accuracy." See https://www.iarpa.gov/research-programs/janus.

2   Department of Homeland Security; "News Release: DHS S&T Awards Funds to Birmingham, AL, Startup Developing Object Detection and Tracking Algorithms for Securing Soft Targets;" June 15, 2023; https://www.dhs.gov/science-and-technology/news/2023/06/15/st-awards-funds-startup-developing-object-detection-tracking-algorithms-securing

3   Nathan Strout; "Here's how intelligence agencies can search foreign documents without learning the language;" C4ISRNet; January 31, 2022; https://www.c4isrnet.com/intel-geoint/2022/01/31/heres-how-intelligence-agencies-can-search-foreign-documents-without-learning-the-language/

4   See the Intelligence Advanced Research Activity program for "Scientific Advances to Continuous Insider Threat Evaluation." The stated project's goal include "to advance the U.S. Intelligence Community's ability to detect potential insider threats." See https://www.iarpa.gov/research-programs/scite.

5   See US Department of Justice's "AI Use Case Inventory;" https://www.justice.gov/open/page/file/1517316/downloadnventory Submission on Open Data (justice.gov).

6   Martin, Peter; Manson, Katrina; "CIA Builds Its Own Artificial Intelligence Tool in Rivalry With China;" Bloomberg; September 26, 2023; https://www.bloomberg.com/news/articles/2023-09-26/cia-builds-its-own-artificial-intelligence-tool-in-rivalry-with-china

7   "National Security Agency is starting an artificial intelligence security center;" Associated Press; September 28, 2023; https://apnews.com/article/nsa-artificial-intelligence-security-deepfakes-f9b19dd64890884cc2b0700ddf66e666

8   National Security Agency; "Artificial Intelligence Security Center," https://www.nsa.gov/AISC/

9   In addition to these examples, DHS initiated an AI safety board, particularly as it pertains to the use of AI to secure critical infrastructure. See US Department of Homeland Security Press Release, "Over 20 Technology and Critical Infrastructure Executives, Civil Rights Leaders, Academics, and Policymakers Join New DHS Artificial Intelligence Safety and Security Board to Advance AI's Responsible Development and Deployment," April 26, 2024; https://www.dhs.gov/news/2024/04/26/over-20-technology-and-critical-infrastructure-executives-civil-rights-leaders

10 Edward Wong, Julian E. Barnes, Muyi Xiao and Chris Buckley; "Chinese Spy Agency Rising to Challenge the C.I.A.;" New York Times; December 27, 2023; Chinese Spy Agency Rising to Challenge the C.I.A. - The New York Times (nytimes.com)

11 Mark Mazzetti and Edward Wong; "Inside U.S. Efforts to Untangle an A.I. Giant's Ties to China;" New York Times; November 27, 2023; Warnings Emerge Over Emirati A.I. Firm G42's Ties to China - The New York Times (nytimes.com)

12 Owen, Quinn; "How AI can fuel financial scams online, according to industry experts;" ABC News; October 11, 2023; https://abcnews.go.com/Technology/ai-fuel-financial-scams-online-industry-experts/story?id=103732051; see also Heiding, Fredrik, Bruce Schneier and Arun Vishwanath; "AI Will Increase the Quantity — and Quality — of Phishing Scams;" Harvard Business Review; May 30, 2024 (AI Will Increase the Quantity — and Quality — of Phishing Scams | Belfer Center for Science and International Affairs)

13 NIE 2002-16HIraq's Continuing Programs for Weapons of Mass Destruction https://nsarchive2.gwu.edu/NSAEBB/NSAEBB129/nie.pdf

14 Intelligence Reform and Terrorism Prevention Act, 2004; E:\PUBLAW\PUBL458.108 (dni.gov)

15 Intelligence Community Directive 203; ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)

16 Intelligence Community Directive 203; ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)

17 Bruce, James B. and George, Roger; "Professionalizing Intelligence Analysis;" Journal of Strategic Security, 8, no. 3 (2015): 1-23; DOI: http://dx.doi.org/10.5038/1944-0472.8.3.1454; http://scholarcommons.usf.edu/jss/vol8/iss3/1; See page 3, footnote 5 and page 8.

18 For an in-depth telling of the formation and implementation of ICD 203 changes, see Fingar, Thomas. 2011. Reducing Uncertainty: Intelligence Analysis and National Security. Stanford, California: Stanford Security Studies; Chang acknowledges the benefits of improved analytic tradecraft but rightly differentiates adherence to tradecraft standards and analytic performance. See Chang, Welton; "Getting It Right: Assessing the Intelligence Community's Analytic Performance;" American Intelligence Journal; Vol. 30, No. 2 (2012), pp. 99-108; Published by: National Military Intelligence Foundation; https://www.jstor.org/stable/10.2307/26202020; see also Marchio, J. (2014). "Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis;" Intelligence and National Security, 29(2), 159–183. https://doi.org/10.1080/02684527.2012.746415

19 This includes: Kwoun, J. (2021); "Cognitive Biases and the Need for Analytic Tradecraft Standards in Large-Scale Ground Combat Operations;" Military Intelligence Professional Bulletin, 47(1), 40–45; https://www.proquest.com/docview/2569703135?parentSessionId=px3ck6WrQZUo0%2BpSJuSeI50l%2BJ27SycNQ%2BrQKnBbis0%3D&pq-origsite=primo&ac

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

14

countid=11311&sourcetype=Trade%20Journals; and Von H. Pigg; "Common Analytic Standards: Intelligence Community Directive # 203 and U.S. Marine Corps Intelligence;" Small Wars Journal; June 16, 2009; Common Analytic Standards (Full PDF Article).

20 First signed in 2007, ICD 203, was amended in 2015 (ICD 203: Analytic Standards (fas.org)) and then in 2022 (ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)). In both instances, the Directive appears to have been adjusted to note the role of an analytic ombudsman.

21 According to NVIDIA, "Large language models (LLMs) are deep learning algorithms that can recognize, summarize, translate, predict, and generate content using very large datasets." See https://www.nvidia.com/en-us/glossary/large-language-models/

22 "It doesn't take much to make machine-learning algorithms go awry: The rise of large-language models could make the problem worse;" The Economist, April 5, 2023, https://www.economist.com/science-and-technology/2023/04/05/it-doesnt-take-much-to-make-machine-learning-algorithms-go-awry

23 Hunter, Tatum; "Chatbots are so gullible, they'll take directions from hackers;" Washington Post; November 2, 2023; https://www.washingtonpost.com/technology/2023/11/02/prompt-injection-ai-chatbot-vulnerability-jailbreak/

24 Weise, Karen; Metz, Cade; "When A.I. Chatbots Hallucinate;" New York Times; May 1, 2023, https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html; "What are AI hallucinations?;" IBM; https://www.ibm.com/topics/ai-hallucinations

25 Treverton, Gregory F. and C. Bryan Gabbard; "Assessing the Tradecraft of Intelligence Analysis;" Santa Monica, CA; RAND Corporation; 2008. https://www.rand.org/pubs/technical_reports/TR293.html;

26 For a definition of Artificial General Intelligence (AGI), see Manning, Christopher; "Artificial Intelligence Definitions;" Stanford University Human Centered Artificial Intelligence; September 2020, https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

27 Devanny, J., Dylan, H., & Grossfeld, E. (2023); "Generative AI and Intelligence Assessment;" The RUSI Journal, 168(7), 16–25; https://doi.org/10.1080/03071847.2023.2286775

28 Hariharan, Ashwin; "Generative AI vs. Traditional Machine Learning: What Sets Them Apart?:" Medium; September 20, 2023; https://medium.com/egen/generative-ai-vs-traditional-machine-learning-what-sets-them-apart-419d093a0128

29 The National Institute of Justice has been funding research related to law enforcement applications for AI, to include facial recognition, gunshot detection, DNA analysis, and crime forecasting since at least 2018. See Christopher Rigano; "Using Artificial Intelligence to Address Criminal Justice Needs;" NIJ Journal 280, January 2019; https://www.nij.gov/journals/280/Pages/using-artificialintelligence-to-address-criminal-justice-needs.aspx

30 Hill, Kashmir; "Eight Months Pregnant and Arrested After False Facial Recognition Match;" New York Times; August 6, 2023; https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html

31 Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren; "Machine Bias;" ProPublica; May 23, 2016; https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

32 Sherman Kent (1903 – 1986) was a Yale historian who joined the Office of Strategic Services, then CIA and was author of several books, to include "Strategic Intelligence for American World Policy," (1949). Kent was integral to the establishment of analytic practice in the CIA and the US Intelligence Community and is the namesake for the CIA's Sherman Kent School for Intelligence Analysis. For more background, see Davis, Jack; "Sherman Kent and the Profession of Intelligence Analysis;" CIA, The Sherman Kent Center for Intelligence Analysis Occasional Papers: Volume 1, Number 5; Nov. 2002; https://www.cia.gov/static/Kent-Profession-Intel-Analysis.pdf; Barnes, Bart, "CIA Official Sherman Kent, 82, Dies;" Washington Post; March 13, 1986; https://www.washingtonpost.com/archive/local/1986/03/14/cia-official-sherman-kent-82-dies/e22ef6e0-a118-42be-b529-7e39f2babaaa/

33 Gordon, Susan; "Op-ed: We keep making the same mistakes with data. Here's how we fix the problem;" CNBC; August 7, 2020; Op-ed: We keep making the same mistakes with data. Here's how we fix the problem | Belfer Center for Science and International Affairs

34 Buolamwini, Joy; Friedman; Barry; "How the Federal Government Can Rein In A.I. in Law Enforcement;" New York Times; January 2, 2024; Opinion | A.I. Use by Law Enforcement Must Be Strictly Regulated - The New York Times (nytimes.com)

35 Weiser, Benjamin; "Here's What Happens When Your Lawyer Uses ChatGPT;" New York Times; May 27, 2023; https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html

36 Office of the Director of National Intelligence; the AIM Initiative: A Strategy for Augmenting Intelligence Using Machines; January 16, 2019; AIM-Strategy.pdf (dni.gov)

37 See a discussion of this effect from Kathleen M. Vogel, Gwendolynne Reid, Christopher Kampe & Paul Jones (2021); "The impact of AI on intelligence analysis: tackling issues of collaboration, algorithmic transparency, accountability, and management;" Intelligence and National Security, 36:6, 827-848, DOI: 10.1080/02684527.2021.1946952; https://doi.org/10.1080/02684527.2021.1946952

38 Lowenthal, Mark M (2013); "A Disputation on Intelligence Reform and Analysis: My 18 Theses;" International Journal of Intelligence and CounterIntelligence, 26:1, 31-37, DOI: 10.1080/08850607.2013.732435; https://doi.org/10.1080/08850607.2013.732435

Analytic Tradecraft Standards in an Age of AI
Belfer Center for Science and International Affairs | Harvard Kennedy School

15

39 Intelligence Community Directive 203; ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov)

40 NISTIR 8312 "Four Principles of Explainable Artificial Intelligence"; P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, Amy N. Yates, Kristen Greene, David A. Broniatowski, Mark A. Przybocki; September 2021; https://doi. org/10.6028/NIST.IR.8312

41 As an example, Harvard University offers an FAQ page related to AI guidance and use in academia. Three sample caveats for instructors to use in guiding students includes a "maximally restrictive," a "fully encouraging" and a "mixed draft policy." For further information, see AI Guidance & FAQs | Office of Undergraduate Education (harvard.edu).

42 See ICD 203, Section D, 6, a; Objective: Analysts must perform their functions with objectivity and with awareness of their own assumptions and reasoning. They must employ reasoning techniques and practical mechanisms that reveal and mitigate bias. Analysts should be alert to influence by existing analytic positions or judgments and must consider alternative perspectives and contrary information. Analysis should not be unduly constrained by previous judgments when new developments indicate a modification is necessary. See ICD-203_TA_Analytic_Standards_21_Dec_2022.pdf (dni.gov), updated December 21, 2022.

43 See Fingar, T. (2011); "Reducing uncertainty : intelligence analysis and national security;" Stanford Security Studies.

44 See Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023); "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality;" SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4573321; See also Dell'Acqua, F. (2022); "Falling asleep at the wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters;" Working Paper. Falling+Asleep+at+the+Wheel+-+Fabrizio+DellAcqua.pdf (squarespace.com)

45 Bajak, Frank; "US intelligence agencies' embrace of generative AI is at once wary and urgent;" Washington Post, May 23, 2024; https://www.washingtonpost.com/politics/2024/05/23/us-intelligence-services-ai-models/92a27dfe-18e5-11ef-97db-d9db0c732407_story.html

46 See Artificial Intelligence Ethics Framework for the Intelligence Community v. 1.0 as of June 2020; AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf (dni.gov)

47 For more background on the potential impacts of AI on analysis, see Kathleen M. Vogel, Gwendolynne Reid, Christopher Kampe & Paul Jones (2021); "The impact of AI on intelligence analysis: tackling issues of collaboration, algorithmic transparency, accountability, and management;" Intelligence and National Security, 36:6, 827-848, DOI: 10.1080/02684527.2021.1946952; https://doi.org/10.1080/02684527.2021.1946952

48 For a broader discussion on evaluating analytic performance, see "Getting It Right: Assessing the Intelligence Community's Analytic Performance;" Welton Chang; American Intelligence Journal, Vol. 30, No. 2 (2012), pp. 99-108; Published by: National Military Intelligence Foundation; https://www.jstor.org/stable/10.2307/26202020 and Marrin, S. (2012); "Evaluating the Quality of Intelligence Analysis: By What (Mis) Measure?" Intelligence and National Security, 27(6), 896–912. https://doi.org/10.1080/02684527.2012.699290.

49 For more background on pre-mortems, see Klein, Matthew; "Performing a Project Premortem;" Harvard Business Review; September 2007; Performing a Project Premortem (hbr.org); Klein, Matthew, Koller, Tim, Lovallo, Dan; "Premortems: Being smart at the start;" McKinsey & Company; March 2019; Bias-busters-Premortems-Being-smart-at-the-start.ashx (mckinsey.com)

The Intelligence Project

Belfer Center for Science and International Affairs

Harvard Kennedy School

79 JFK Street

Cambridge, MA 02138

belfercenter.org/project/intelligence-project