2022

Solving Mysteries with Public Secrets

OPEN-SOURCE FOR THE INTELLIGENCE COMMUNITY

AUSTIN KATSTRA

FRANK BATTEN SCHOOL OF LEADERSHIP & PUBLIC POLICY | UNIVERSITY OF VIRGINIA

Acknowledgements

I would like to thank the Batten School faculty and staff that have supported me throughout this project and over the past two years. Specifically, I want to express my gratitude for my APP instructors, Professor Kirsten Gelsdorf and Professor Allan Stam, for helping me throughout this project. Finally, I would like to thank my family and friends for their endless support in my educational and career ambitions.

Disclaimer

The author conducted this study as part of the program of professional education at the Frank Batten School of Leadership and Public Policy, University of Virginia. This paper is submitted in partial fulfillment of the course requirements for the Master of Public Policy degree. The judgments and conclusions are solely those of the author, and are not necessarily endorsed by the Batten School, by the University of Virginia, or by any other agency.

Honor Pledge

On my honor as a student, I have neither given nor received authorized aid on this assignment.

Austin Katstra

Table of Contents

Abbreviations guide

Abbreviation	Meaning
AI	Artificial Intelligence
DHS	Department of Homeland Security
DIOG	Domestic Investigations and Operations Guide
DNI	Director of National Intelligence
DOJ	Department of Justice
FBI	Federal Bureau of Investigation
FBIS	Foreign Broadcast Information Service
FISA	Foreign Intelligence and Surveillance Act
GEOINT	Geospatial Intelligence
HUMINT	Human Intelligence
IC	Intelligence Community
ICAC	Internet Crimes Against Children
ISIS	Islamic State of Iraq and Syria
JTTF	Joint Terrorism Task Force
ML	Machine Learning
ODNI	Office of Director of National Intelligence
ODNI CLPT	ODNI Civil Liberties, Privacy, and Transparency
OSC	Open Source Center
OSE	Open Source Enterprise
OSINT	Open-Source Intelligence
PAI	Publicly Available Information
SIGINT	Signals Intelligence

Executive Summary

In a world dependent on the internet and constantly engaged in social media, OSINT must be elevated to a true intelligence discipline. Recently, OSINT has been used by private companies to identify nuclear powerplants, chemical weapon manufacturing, and armed force movements. The advancement of the internet and the exponential increase in social media usage and engagement has led to an explosion of data that the US IC is not currently using maintain its intelligence edge. Currently the only OSINT dedication within the US IC is the Open Source Enterprise at CIA, which has seen its funding and scope decrease over the last administrations. The IC's bias towards classified sources, concerns regarding data security, policy hurdles, and legal challenges have proven the four greatest reasons why nothing else has been done.

An analysis of ICAC Centers, JTTFs, social media company content regulation, and a review of intelligence collection guidelines shows that engagement with technology by law enforcement and private companies is successful in deterring, investigating, and prosecuting criminal behavior. Using this evidence, the creation and implementation of AI and ML tools that gather PAI are recommended for the IC to maintain its intelligence edge. This recommendation was evaluated against the creation of the Open Source Intelligence Agency and maintaining the OSE using constitutionality, financial cost, risk, and political and administrative feasibility as the evaluative criteria. The criteria were derived from the culture, security, policy, and legal concerns previously mentioned and discussed in further detail below. While AI/ML tools were not rated the highest in all categories relative to the other options, it did receive the highest rating regarding cost and was tied in administrative feasibility, both important criteria for obtaining appropriations from Congress. Additionally, this option has the greatest chance at success across both domestic and foreign national security concerns and is adaptable in the long run as national and intelligence priorities change across administrations and times.

A standard blueprint will be developed, with consultation from industry leaders, and given to all applicable agencies. Stakeholder agencies will include CIA, FBI, ODNI, and DHS, however, the tools are not limited to only these agencies. The agencies will then work internally and with the ODNI CLPT to build out the tools as they deem fit for their mission and authorities. The tools will collect and analyze the data and alert an analyst when the information needs additional attention and/or human eyes. The tools will not deal disseminate intelligence to customers, the analysts will have the final verdict on any intelligence before it is considered finished and sent to the customer. This will be done to remove any bias from the tools and ensure the intelligence being sent to the customer is answering the questions which have been asked.

The final product will elevate OSINT to a main "INT" and provide the intelligence community with a necessary tool to capture the explosion of PAI. AI and ML tools offer an opportunity to keep the US IC on the cutting edge of its field and continue to solve the mysteries that no one else can.

Introduction

Chemical attacks in Syria, identification of Russian military operatives that shot down MH17, and illegal shipping of Sarin gas: each of these instances was discovered by private companies using open-source intelligence (*About Bellingcat*, 2022). OSINT is developed from information that is publicly available and currently an undertaking done primarily by only private companies with no involvement from the IC.

The following document will provide an extensive review of OSINT, the need for it be elevated to a main "INT", and how to best do that given the global data explosion and high usage rate of social media. The background involves the declining status of the Open Source Enterprise, why nothing else has been done, and what could happen if doing nothing becomes the norm. The literature review will focus on the effectiveness of internet crime centers and JTTFs, evaluation of surveillance guidelines, and practices of social media companies to counter violence on their platforms. The solutions discussed in the literature review are shown as examples of policy alternatives that are either theoretical or have been applied across various fields of study. The characteristics of these solutions may differ from viable strategies for adoption by the intelligence community. The alternatives proposed for potential adoption by the intelligence community incorporate lessons learned from these studies. The alternatives, recommendation, and implementation consider financial costs, the level of risk, political and administrative feasibility, and constitutionality. These criteria are derived from the current barriers regarding the incorporation of OSINT into the IC. The recommendation is followed by implementation which explains the value of stakeholder consultation, the risks involved, and how to manage the risks.

Problem Statement

In 2020, the average online user generated 1.7 megabytes of data per second (Rasmussen, 2020). The global explosion of data has made publicly available information more accessible than ever before, making open-source intelligence more important than ever before. The U.S. intelligence community is facing a unique opportunity to harness this data power and maintain its intelligence edge.

Background

The Current State of OSINT

The intelligence landscape changed drastically following 9/11. One of the recommendations given by the 9/11 Commission was to establish an Open-Source Intelligence Agency (*The 9/11 Commission Report*, 2004, p. 413). Following the recommendation, the DNI, created the OSC within the CIA on 8 November 2005. OSC succeeded the FBIS. FBIs has previously been in charge of monitoring publicly available media and translating it (Glasser, 2005). In 2015, the OSC changed its name to the OSE and was absorbed in the CIA Directorate of Digital Innovation. Upon the name change, the CIA stated its mission as "collecting, analyzing, and disseminating open source information"

(Aftergood, 2015). Given the CIA's statutory authorities given by the National Security Act of 1947, this collection is focused solely external to the US. To date, OSE is the only publicly known OSINT dedication within the US IC and since 2015 its role, scope, and functions deprioritized across multiple administration (Weinbaum, 2021).

Outside of the IC, private companies have developed tools to collect PAI themselves. Technology and industry have, and continue to, outpaced law, policy, and regulation that is required for the IC to have adequate tools (Ashley & Wiley, 2021). As such, private companies often have much more data than the IC in terms of OSINT. On the social media side, this is largely due users voluntarily giving the platforms their information that then inform their algorithm. Outside of social media, private companies also conduct vast OSINT collection efforts to better inform them of their customers and societal trends to determine how to better market. There is opportunity to work with private industry to create similar tools for the IC, however nothing has been done to date.

Why Has Nothing Else Been Done?

Culture: The IC Bias for Classified Sources

The IC prides itself on its ability to find and collect secrets and solve mysteries. Historically, IC agencies have done this through clandestine and/or covert means. This involves HUMINT, SIGINT, GEOINT. The IC has largely ignored open-source data because of their belief in their abilities to inform the customer through classified means (Weinbaum, 2021). There have been large successes using these ways and means, however as the world continues to adapt, the IC must continue to do the same to provide the customer with the best information and analysis available. Not incorporating OSINT as vital intelligence runs the risk that analysts will be missing out on vital information to help them solve mysteries.

Security: Classified Servers are Expensive

The amount of data that would be collected through OSINT would be massive, and most likely make it the largest gathering "INT" in terms of data that needs to be stored. One would assume that this would likely be a large cost to create and maintain enough classified servers to store the data. However, given that the information collected is already public, there is no need to house the data on classified servers. Storing the data in the cloud or creating a 5G slice is a cheaper alternative that does not decrease security to the point where sources or methods would be compromised. Both the cloud and 5G slice have impressive security components that are already imbedded within them. Finished intelligence products should still be stored on classified servers, but the raw data does not require the same type of security.

Policy: Congress is averse to risk-taking

Following 9/11, congressional oversight of the IC increased significantly. The Senate Select Committee on Intelligence and the House Permanent Select Committee on Intelligence both obtained larger roles within the IC to oversee its budgets, activities, and programs (*Congressional Oversight of Intelligence: Background and Selected Options for Further Reform*, 2018). Furthermore, Congress as an institution is risk-averse and wants to see results from programs and activities for which it appropriates money (E. Harding, personal communication, February 15, 2022). This makes activities, such as an expansion of OSINT programs, difficult to get Congressional approval.

Outside of Congress, the policy process for acquisition of new technology is extremely complex and slow. OSINT will deal largely will the internet and social media, requiring technology to be involved in any solution. The acquisition process makes it difficult to have software enter the national security space to begin. Software requires constant updating, meaning that by the time the software is approved and acquired, it is often out-of-date and the process must be restarted (Harding, 2022).

Legal: Collection Policy Differs Between Agencies

The USA Patriot Act went into effect on 26 October 2001 as a response to the terror attacks on 9/11. The Act expanded the surveillance capabilities of federal law enforcement, specifically regarding counterterrorism investigative authority. Title II of the Act specifically authorizes additional surveillance techniques, including the expansion of the FISA authorities and addressing the expressed inclusion of electronic evidence in search warrants (H.R.3162 - 107th Congress (2001-2002), 2001). The Act was extended three times: 2006, 2009, and 2011 (*USA PATRIOT Act - Reauthorizations*, n.d.) with the Supreme Court of the United States (SCOTUS) upholding the Constitutionality of the Act in 2010 (*Holder v. Humanitarian Law Project, 561 U.S. 1 (2010)*, 2010). Despite the Act being declared Constitutional, the USA FREEDOM Act replaced the USA Patriot Act to curtail the government's authority to collect data in 2015. The revision occurred after Edward Snowden leaked information about the government's bulk data collection of phone and internet records (*USA PATRIOT Act - Reauthorizations*, n.d.)

The DIOG regulates all investigative and intelligence collection activities within the US that are conducted by the FBI. It does not apply to activities in foreign countries. Regarding policy relevant to OSINT and PAI, the DIOG states that the FBI cannot monitor social media channels without a "predicated investigation" (*FBI Domestic Investigations and Operations Guide (DIOG)*, 2016). FBI Director Christopher Wray explained what that means as, "what we're not allowed to do is just sit and monitor social media and look at one person's posts just looking to see if maybe something would happen just in case" (*FBI Director Christopher Wray Testifies on U.S. Capitol Attack | C-SPAN.Org*, 2021). The "predicated investigatory" standard only applies to public accounts on publicly accessible platforms. Information from private accounts on public platforms or private platforms requires a subpoena or search warrant, depending on the specifics of the case. However, there is ambiguity around the interpretation of these guidelines and evidence to suggest that this interpretation has changed throughout the years or is even being interpreted incorrectly now (Jurecic, 2021).

Consequences of the Problem

As with all national security problems, the ultimate mission of the IC is to protect the American people. The explosion of the internet over the past 20 years has made that mission increasingly difficult as criminals across the world are able to connect where they otherwise would not have been able to. Social media has been shown to be a vast source of OSINT for this exact reason. For example, almost half of the world's population has a social media account (Farina, 2020). The prevalence of social media in daily life combined with its global reach provides an immediate opportunity for terrorist activity to be planned between people who, otherwise, would not have had contact. Lastly, researchers have found that the internet is now the primary operational medium in which political ideologies are realized, attacks orchestrated, and social movements constructed (Winter et al., 2020). The IC, law enforcement, and private stakeholders must work together to address the radicalization of social media. A failure to do so, could result in an attack that leaves tragic damage in its wake.

Literature Review

Internet Crime Centers, JTTFs, and Fusion Centers

Law enforcement investigations for violations regarding crimes against children provides an opportunity to evaluate the effectiveness of a policy intervention directed at crime planned and conducted on the internet that has already been implemented. In 2017, the DOJ developed the Internet Crimes Against Children Taskforce Program to help state and local law enforcement agencies develop a response to child exploitation using technology and/or the internet (*Internet Crimes Against Children Task Force Program*, n.d.). In June 2021, the Office of the Inspector General conducted an audit of the ICAC Montana Division and found that ICAC was successful achieving the goals set forth: recruit affiliates, provide training and equipment to members of the taskforce, conduct undercover online investigations, and partner with School Resource Officers (Office of the Inspector General, 2021).

Additionally, JTTFs[1]and Fusion Centers[2] can be used as evidence of successful joint investigations between the federal, state, and local law enforcement. Since their inception in 2003, both have foiled numerous domestic terrorism plots, including the NYC subway bomb plot. The University of Texas at El Paso found that the need for JTTFs and Fusion Centers is difficult to evaluate the effectiveness of the centers because it is impossible to ascertain the number of attacks they eliminated by "simply existing and providing analytic support". However, Devine found that their contributions cannot be diminished, and their existence must continue (Devine, 2014).

From the evaluation of JTTFs and Fusion Centers, it can be determined that law enforcement across all levels of government are capable of working together and able to contribute to the mission to collect PAI if needed. Furthermore, through the audit we know the DOJ can develop successful internet taskforces. While internet crimes against children consist of a very different type of violation and criminal than national security cases, the technological capabilities of the DOJ can be seen through the success of the ICAC.

[1] JTTFs are a combination of federal, state, and local partners regarding counterterrorism investigations.

[2] Fusion Centers consist of federal, state, and local partners working together to gather and analyze intelligence to contribute towards the federal government's counterterrorism efforts.

Re-evaluating Investigative Practices

Quinta Jurecic, a fellow at the Brookings Institution and managing editor at Lawfare, suggests the DIOG needs to be revisited and evaluated whether it remains effective as technology has progressed. Jurecic also disagrees with the FBI's interpretation of the DIOG regarding their response to the January 6 insurrection at the US Capitol (Jurecic, 2021). The FBI interpretation of the DIOG mandates that the FBI have an investigative reason to "surveil" social media. However, Quinta Jurecic disagrees with FBI Director Wray's testimony that the FBI does not have the authority to monitor social media, in certain cases. She identifies the section that states the Bureau may engage in, "proactively surfing the Internet to find publicly accessible websites and services through which recruitment by terrorist organizations and promotion of terrorist crimes is openly taking place" (*FBI Domestic Investigations and Operations Guide (DIOG)*, 2016). She argues that this was certainly the case on 6 January 2021 where the bulk of planning took place on social media sites known to have posts promoting violence (Jurecic, 2021).

Jurecic's analysis of the DIOG seem to contradict that of the FBI's. Nevertheless, Director Wray has suggested that if the DIOG prevented the FBI from identifying violent actors on social media prior to January 6, then it is a reasonable question whether the guidelines should change (*FBI Oversight Hearing | C-SPAN.Org*, 2020). Given the recency of 6 January 2021, there is not enough publicly available data to draw conclusions on who is "correct" in this case. The DOJ Inspector General is currently investigating the role of the DOJ, which the FBI is a part of, in its preparation and response to January 6 (Jurecic, 2021). The House of Representatives is also conducting its own investigation. The conclusions reached by both parties will provide much more information and data on the events of the day and months leading up to it. While January 6 is just one example, and an extreme one, it is important to note how the implications from these events may results in policy change in the years to come.

Private Company Self-Regulation Decreases Extremist Content

In 2015, Berger & Morgan found that when Twitter suspends pro-ISIS Twitter accounts, it reduces the quantity of ISIS content available online because the accounts entire history is deleted when the account is suspended (Berger & Morgan, 2015). Milton then finds that when Twitter conducts self-regulatory activities, such as mentioned above, it reduces the quantity of total extremist content found online. He also finds that Twitter is removing accounts and files associated with extremist content quicker than at any point in the company's history (Milton, 2016).

Berger & Morgan and Milton studies are particularly niche, focusing only on Twitter and ISIS-related content. However, we find the same conclusions when looking at more expansive studies.

Ribeiro et al. analyzes the migration of two popular extremist groups who were banned from Reddit and migrated to their own websites. In both cases, posting activity significantly decreased on the new platform in comparison to Reddit and, perhaps more significantly, user engagement and the number of new users also decreased (Ribeiro et al., 2021). This study debunks a popular misconception that regulating extremist groups will lead them to more unregulated spaces and subsequently to become even more extreme than before because they will be the only individuals operating in that space.

This evidence suggests that companies should participate in rigorous content moderation, there are fewer extremist posts and that they do not move to other platforms with as much success as if they had stayed on the previous platform. This would suggest that there is a large burden shouldered by these social media platforms, and the cost of moderation may come at their bottom line as they move users off the platform. Solutions could be explored to encourage companies to increase content moderation (even against their bottom line), without creating a culture of censorship, or a solution to move some of this burden off the social media companies.

Existing Gaps in Literature

It is difficult to evaluate the effectiveness of the current methodologies in place because most of them are classified. It is very possible that some of the options explored in the literature review have been tried and failed, or are currently in use, and the public does not have knowledge of their use. As shown in the Fusion Center example, it is difficult to evaluate the effectiveness of something that never happens, especially when the organization is historically used as a reactionary, not proactive force (George & Rishikof, 2011).

Criteria

Criterion #1: Constitutionality

This criterion is of utmost importance. Particular focus will be given to the First Amendment right to freedom of speech because OSINT involves using images and words posted by individuals. To measure the constitutionality of the alternatives, I will use the requirements set by Brandenburg v. Ohio to determine what speech must continue to be protected and cannot be required by Congress to be taken down. The parameters applied must meet the following criteria: the speaker must intend to and use words that rally people to take illegal action, the danger must be imminent—not in the indefinite future, and the words must be uttered in a situation in which violence is likely to happen (Hudson, 2021).

Criterion #2: Cost

This criterion will focus on how cost effective the alternative is, considering both direct and indirect costs. Direct costs will be measured in dollar amounts. The questions that will be answered to determine the amount will be how much it costs to build the alternative, how much it costs to implement the alternative, and how much it costs to maintain and staff the alternative. Indirect costs will be measured in the amount of time the alternative saves current analysts, but also staffing requirements if the alternative would decrease staffing measures as the USG attempts to save money by restricting the number of hires. While direct costs will be able to be measured in dollar amounts, indirect costs will be more difficult to measure. Indirect costs will be measured on a sliding scale of low, medium, or high with high being large, negative costs.

Criterion #3: Political Feasibility

This criterion will be a combination of political and administrative feasibility. Political feasibility will focus on the likelihood that the alternative receives support from both political parties. This will be especially important for alternatives that are legislatively based, such as Alternative #2. They will be measured on a sliding scale of low, medium, or high with high being likely and low being unlikely. Strong support from both political parties would result in a high rating, support from neither or only the minority party would result in a low rating. Anything in between would result in a medium rating.

Creation #4: Administrative Feasibility

Administrative feasibility will be a measure of the ability of the IC to implement and/or enforce the alternative. This will consider the budgetary restrictions of the agency combined with the manpower restrictions. The technological capability of the agency will also be considered for options requiring technological implementation. They will be measured on a sliding scale of low, medium, or high with high being likely feasible and low being not likely to be feasible.

Analysis of Alternatives

Alternative 1: Creation of an Open Source Intelligence Agency

Description

The IC's only current Open Source Enterprise (OSE) resides within the CIA's Directorate of Digital Innovation (CSIS Technology and Intelligence Task Force, 2021). This is insufficient for the IC to continue to adapt with the global explosion of data. The establishment of an Open Source Intelligence Agency (OSIA) would be a one-stop shop for all OSINT in the federal government. It would expand on the current OSE but become its own agency. Currently the National Initiative for Cybersecurity Careers and Studies (NICCS) offers the ability to become certified in OSINT (*Certified in Open Source Intelligence (C|OSINT) from McAfee Institute | National Initiative for Cybersecurity Careers and Studies*, n.d.), but there is no center where personnel specializing in OSINT would work. Within OSIA there would be an international and domestic bureau because there are vastly different legal frameworks that govern the surveillance and collection of data for USPER and non-USPER. The bureaus would operate independently from any current agency but would work primarily with agencies within their jurisdictions.

Domestic OSIA (DOSIA) would primarily work with FBI and DHS because their jurisdictions cover domestic operations. International OSIA (IOSIA) would work primarily with CIA, NSA, and military intelligence covering their international jurisdictional duties. To increase initial collaboration between the existing IC and the new Center, the IC should re-assign current analysts and legal advisors who are OSINT certified to the Center as full-time employees. This should be complemented with employees who are not from currently in the IC. This would help to remove potential cultural and agency biases by moving analysts from their existing agencies to the Center. There will not be investigators assigned to the OSIA, individuals conducting operations related to the collection of OSINT will remain within their respective agencies. Cases and operations will be assigned to the appropriate agency based on the OSINT collected and analyzed by OSIA.

Criteria: Constitutionality

This alternative is highly likely to be constitutional. This criterion will operate under the assumption that OSIA analysts will use already existing OSINT collection procedures. Considering these procedures are in operation currently, it can be assumed that these practices are constitutional and as such, would be appropriate for OSIA to use. Additionally, OSE is currently operating, and this alternative is ultimately a suggestion to expand OSE. While expanding, OSE must ensure that it adapts its collection and investigative practices to distinguish between USPER and non-USPER.

Criteria: Cost

The cost of OSIA would be high. It would require Congress to allocate funds to create and maintain an eighteenth intelligence agency. In 2008, RAND projected the cost of a new domestic intelligence agency to cost $500 million (Treverton, 2008). Fourteen years later, it can be easily assumed that cost would be significantly higher and initial starting value be more to cover the international nature of the agency as well. However, the potential ability to maintain the collection on an unclassified network would decrease the cost of maintaining a secure, classified system.

Criteria: Political feasibility

The political feasibility of OSIA is low. There is widespread thinking that the 2016 elections should have been the IC's 9/11 moment for OSINT (Harding, 2022). The use of social media and proxies by Russia to conduct a prolonged and undetected attack on the US showcased vulnerabilities in the US IC. However, political polarization of the 2016 elections and the disinformation campaign run by Moscow stopped any momentum to retrain and reorganize the IC. Given this environment, it is unlikely that OSIA could receive immediate support without the backing of senior leaders in the IC and Congress. Even then it may be difficult to generate the necessary support to create an eighteenth intelligence agency.

Criteria: Administrative feasibility

This administrative feasibility of OSIA is medium. The largest hurdle is IC culture. OSIA may struggle to establish itself as a main "INT" given the IC's bias towards classified information (Weinbaum, 2021). The IC has historically prided itself on being able to gather secrets[3] and solve mysteries. Emily Harding, a former analyst at CIA, described mysteries as "things that no one knows but there are ways to put the puzzle pieces together" and that "all-source analysts miss big

[3] Secrets are things, items, ideas, plan, intentions, etc. that can be stolen

developments [in the mysteries] that were only findable in OSINT". Without overcoming this bias towards classified information, it will be extremely difficult, if not impossible, for OSIA to thrive in the IC. However, the nature of OSINT as unclassified information opens the opportunity for the IC to acquire more private-sector talent and decreases money spent on securing information. The information gathered could exist on unclassified servers because it is already public knowledge. The current cost to maintain classified systems is extraordinary and the ability to reduce those costs raises this criterion from low to medium.

Alternative 2: AI/ML tools to process the data

Description

The exponential innovation of machines in the past decade provides the opportunity to use those to further the U.S. intelligence mission in a way that was not possible before. Today, machines can read, write, think, and analyze (Siri, Alexa, and Cortana). The IC could use and adapt these tools that already exist to be able to comb through millions of images, publications, and writings that are already publicly available. Additionally, the database built from these tools would not need to be classified as the data is coming from publicly available information. The analysis and further categorization of the data may need to be, but the raw data itself would not need to be moved to a classified system. This type of system would take time to build, but the ML technology already exists so alterations would just need to be made. Upon its creation, it would save analysts hundreds of hours per year and create a more robust system that could potentially identify pandemics, attacks, and threats before they happen.

Criteria: Constitutionality

The development of AI/ML tools would comb through only information that is publicly available. This would include posts on public social media accounts, published papers, images, etc. Given the public availability of this data, it would be considered Constitutional to collect it. The FBI must abide by the DIOG, the DHS I&A abide by Exec-ut-ive Order 12,333. Additionally, state and federal judges have generally concluded that public posts are not protected from searches and seizures by the Fourth Amendment (Levinson-Waldman et al., 2022). There have been no rulings about government surveillance of social media, however, the DIOG acknowledges that "[o]nline inform-a-tion, even if publicly avail-able, may still be protec-ted by the First Amend-ment'‘. Ultimately, the FBI believes that the reason for the surveillance determines constitutionality. People cannot be targeted for surveillance based on exercising their rights under the First Amendment or for race, ethnicity, or religion under the Fourteenth Amendment (*FBI Domestic Investigations and Operations Guide (DIOG)*, 2016). Given the complexity of various policies between agencies and ambiguity of what exactly is social media surveillance the constitutionality of this alternative is given a medium.

Criteria: Cost

The cost of this alternative is high. There is a high financial cost of building the AI/ML tools, maintaining the system, and training the analysts on how to properly use it. However, once the initial cost is born, AI systems are projected to save all source analysts 45 working days a year (Mitchell et al., 2019). Despite the time saved, there is also a high-risk factor. It will take time for the machine to learn what it is doing and improve itself. Projections suggest after five years it would be able to operate almost on its own and detect, for example, changes in leaf piles around a potential nuclear research lab in Iran, and alert analysts (Harding, 2022). Given the unknown time for success of the tools, it remains a high-risk project. Lastly, the vast amount of data this alternative would generate would need to be stored on a server. However, because this is all public information to begin with, its server does not need to be classified. The information could be stored on the cloud or in 5G slices, both of which offer high security. Finished intelligence products with analysis from analysts would remain on classified servers, but the raw data would not need to be. This would result in decreased costs by not having to increase the number or capacity of classified servers.

Criteria: Political feasibility

Congress is a risk-averse organization that prefers immediate impact and visible results when allocating resources (E. Harding, personal communication, February 15, 2022). Given the high risk and longer timeline for success, the political feasibility of this option is low. It will be difficult to convince Congress to support a project with many unknowns. The SSCI and HPSCI have the highest chance of support and should be the first point of contact if this alternative is chosen.

Criteria: Administrative feasibility

The IC culture towards classified resources as previously mentioned remains an administrative hurdle in this alternative. However, implementation of this alternative is simpler than OSIA. Once the system is built and approved, there would only need to be training for the all-source analysts on how to use the system. The real leg work of this alternative comes in the approval phase. Outside of Congress, the acquisition process for the government is arduous and complex. The acquisition process must be streamlined or completely tossed away for this alternative to work. In other words, the ODNI must be allowed to approve acquisitions for this project, or a parallel acquisition system must be created to improve speed and keep up with the speed of technological innovation.

Alternative 3: Maintain OSE

Description

The current placement of OSE within the CIA dictates OSE must maintain a foreign focus. The CIA is statutorily barred from operating internally to the US. While there is little known information about the operations or budget of OSE, the CIA defines their mission as "collecting, analyzing, and disseminating open source information" (Aftergood, 2015). By maintaining OSE, OSINT collection would remain solely focused overseas within the CIA's authorities given to it by the National Security Act of 1947.

Criteria: Constitutionality

Given that the OSE currently exists, the alternative is given the rating of "high" because it can be assumed that OSE is operating constitutionally.

Criteria: Cost

The budget of OSE is unknown as the overall budget of the CIA is a tightly kept secret, with the budgeting of specific programs kept even more secretive. Given this information, the financial cost of this alternative is unknown. The cost in terms of risk associated with this alternative is high. By maintaining OSE, or even expanding it, there is a failure to address the problem. OSINT would not be elevated to a main "INT" and there would remain vast amounts of data and information that the IC is not collecting and analyzing.

Criteria: Political feasibility

Political feasibility is high because it currently exists. The activities of the CIA often remain out of the public eye and Congress does not conduct public reviews of their activities. The SSCI and HPSCI receive reports, but those are rarely declassified until years later. Given this information, it is unlikely that OSE would be affected politically.

Criteria: Administrative feasibility

The deprioritization of the OSE by its absorption into the CIA Directorate of Digital Innovation raises concerns that OSE is not receiving the attention and support needed for the amount of intelligence it could produce. As such, administrative feasibility receives a rating of medium because while it exists, the consistent reduction of status over the past ten years raises concerns of effectiveness and its priority in the IC.

Recommendation

Pursue Creation and Implementation of Alternative 3: AI/ML tools.[4]

Through the tools, there will be a baseline algorithm that is adapted for each agency based on its authorities, priorities, and restrictions. For example, the CIA operates only external to the Unites States and the tools for the CIA would only target foreign collection. Opposite from the CIA, the FBI is a domestic intelligence agency and is concerned with security internal to the US. The FBI's tools would be tailored to this mission and to ensure they abide within the guidelines set forth in the DIOG. The same premise exists for DHS, the tools will be tailored to fit within the intelligence collection guidelines and mission of DHS. CIA, FBI, and DHS are just the examples given in this paper, it should not be assumed they will be the only agencies with access to, or the ability to use, these tools. The tools will collect and analyze the data and notify an analyst when the information needs more attention and human eyes on it. The machine will not deal with dissemination to customers, the analysts will have the final say on any intelligence before it is considered finished intelligence and sent to the customer. This will be done to remove any bias from the tools and ensure the intelligence being sent to the customer is answering the questions which they have asked.

[4] Outcomes matrix included in Appendix A

The potential long-term benefit of this program is significant, and it will incorporate private sector collaboration and talent acquisition in the process. The cost of AI/ML tools is lower than establishing a new intelligence agency and while AI/ML tools do pose risks, these risks are evaluated to be lower than the cost of creating OSIA and the administrative issues associated with it. In addition to bias towards classified sources, the coordination problem of adding another intelligence agency exists. Additionally, the tools would be able to be coded slightly differently depending on the agency's surveillance and investigative policies. While maintaining OSE is the most politically feasible and likely the most constitutional, the failure to address the problem of that alternative makes success essentially impossible should it be chosen. While maintaining OSE would continue intelligence gathering efforts overseas, it is unclear to what capacity and does not focus on intelligence collection internal to the United States. This is especially important because FBI Director Wray declared, "[t[he greatest threat we face in the Homeland is that posed by lone actors radicalized online…[w]e see this lone actor threat manifested both within Domestic Violent Extremists ("DVEs") and Homegrown Violent Extremists ("HVEs")" (*Worldwide Threats*, 2020).

Stakeholders

Jennifer Ewbank: CIA, Deputy Director for Digital Innovation

Deputy Director Ewbank oversees the Directorate of Digital Innovation at CIA which houses the Open Source Enterprise (OSE). Her consultation will be vital as the OSE is currently the only division within the IC that is dedicated solely to OSINT collection and analysis. Deputy Director Ewbank may be resistant to change because it will expand the opportunity for other agencies to conduct OSINT activities, an activity currently only directed to the CIA OSE. On the other hand, she may be supportive because it takes the onus off the CIA to be the only agency conducting OSINT activities.

Darrin E. Jones: FBI Executive Assistant Director, Science and Technology Branch

The FBI STB mission is to "use science and technology to enhance operations and investigations" (*Science and Technology Branch*, n.d.). As Director of the FBI STB, Mr. Jones's input will be important as AI/ML tools will increase the FBI's technological capabilities in investigations. Given that the CIA is the only agency with dedicated OSINT activities, the FBI will likely be receptive to technologies that increase their capabilities to defend the homeland. Furthermore, the CIA does not have the authority to operate internally to the US. By expanding the options of domestic intelligence agencies through the tools, the FBI and DHS will be able expand their collection efforts. The FBI may be resistant to change, however, because the interpretation of the DIOG currently is unclear regarding social media "surveillance" (Jurecic, 2021).

Kathryn Coulter Mitchell: DHS, Senior Official Performing the Duties of the Under Secretary for Science and Technology

The Science and Technology Directorate is responsible for conceptualizing art-of-the-possible solutions to secure the homeland (*Kathryn Coulter Mitchell | Homeland Security*, n.d.). AI/ML tools would be front of the line technological capability and fall under the DHS ST mission. DHS will likely be very receptive to additional capabilities because they already have authority to domestically collect intelligence. Currently DHS is trying to expand is OSINT collection, specifically regarding social media (Interviewee #1, personal communication, November 13, 2021). Through these tools, DHS will be able to expand their abilities and continue to defend the US.

Aric Toler: Bellingcat, Head of Training and Research
Bellingcat is one of the nation's leaders in open-source resources and uses a combination of journalists, investigators, and researchers to conduct open-source investigations. As head of training and research, Mr. Toler oversees Bellingcat's technological developments, specifically including social media. He will be a valuable resource for social media open-source analytics. Bellingcat will likely be supportive of this decision because they will view it as an opportunity to provide their satellite images to the government for their use in this technology.

Werner Vogels: Amazon, Chief Technology Officer
Mr. Vogels works with Amazon and Amazon Web Services and oversees all technological developments within the company. Given that AWS is a worldwide leader in AI and ML and operates the classification systems for the IC, Mr. Vogels input will be extremely valuable. AWS will likely be supportive of this solution because they have sponsored white papers supporting AI tools for the IC in the past (Harding, 2022) and they operate the IC classified servers.

All Source Analysts
All source analysts are charged with analyzing collected intelligence from all sources. This would include OSINT and involve the new AI/ML tools. As such, we would want input from the analysts on how to best include OSINT technologies in their day-to-day activities, but also show how the tools will drastically decrease the amount of time spent on analytics in the long run. All source analysts may be resistant to change because this will add an additional responsibility. This is true especially in the short run while the tools continue to learn themselves and the analysts must help teach the tools. By emphasizing the benefits in the long run, these concerns can be dissuaded.

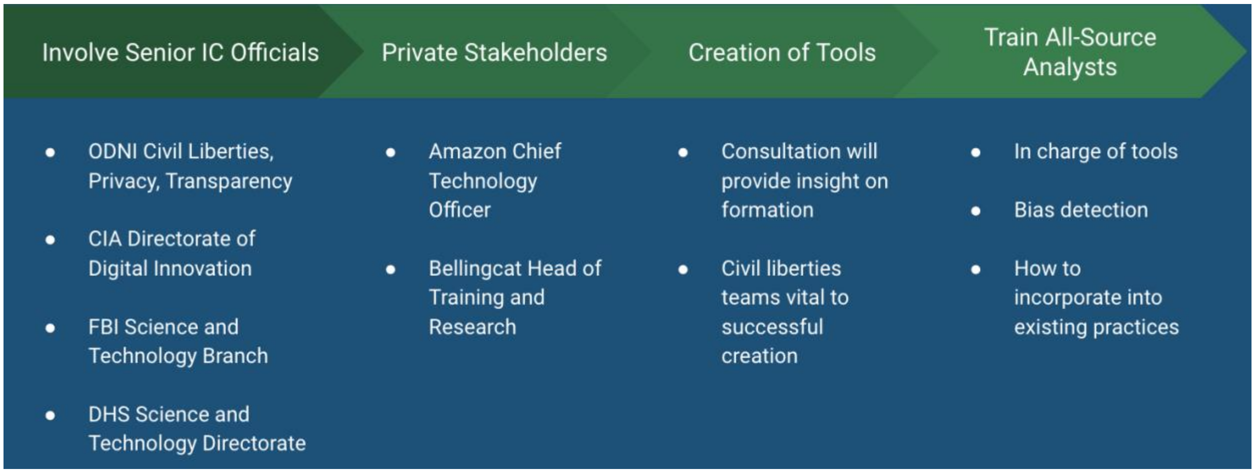
Implementation

Implementation will involve four separate steps (see Figure 1). First, consultation with senior IC leaders will occur. The purpose of this is two-fold: to hear concerns and discern agency priorities, guidelines, and restrictions that must be built into the tools. Priority will be given to Directorate of Digital Innovation, FBI Science and Technology Branch, and DHS Science and Technology Directorate because these are the lead agencies for foreign collection and domestic collection, respectively. The ODNI Civil Liberties, Privacy, and Transparency (CLPT) division will be the lead on civil liberties concerns with the AI/ML tools. CLPT should be involved at the early and often to ensure the rights of all Americans are protected. This is a foremost priority and should not be overlooked in any capacity.

Second, consultation with private stakeholders is suggested. The examples given in Figure 1 should not be considered inclusive and are solely examples of two industry leaders. Amazon and the AWS Cloud should be consulted on implementation of a cloud to store the data on, as well as their algorithmic practices for data collection. Bellingcat is an industry leader in OSINT and their expertise will be vital to determining what information to teach the machine to prioritize and how to understand detection and bias in OSINT technology.

The creation of the actual tools will be based on the consultation with IC officials and private stakeholders. The information gathered from these conversations could, and most likely should, influence the creation of the product. The civil liberties teams must be involved heavily during creation, as in the rest of the process.

Lastly, once a successful tool has been created, training of the all-source analysts should commence. These analysts will ultimately be "in charge" of the tools and should be adequately trained on how to use them, but also in their importance to the overall intelligence effort. Training in the importance of OSINT as a main "INT" should not be understated given the IC bias for classified sources. There is valuable intelligence and pieces of the puzzle to be gleaned from OSINT and it is vital analysts recognize that. The final part of the training will be in bias detection. The tools will not be perfect, as nothing is, so analysts must be able to detect when the tools are biased towards an outcome of type of information.



Involve Senior IC Officials	Private Stakeholders	Creation of Tools	Train All-Source Analysts
• ODNI Civil Liberties, Privacy, Transparency • CIA Directorate of Digital Innovation • FBI Science and Technology Branch • DHS Science and Technology Directorate	• Amazon Chief Technology Officer • Bellingcat Head of Training and Research	• Consultation will provide insight on formation • Civil liberties teams vital to successful creation	• In charge of tools • Bias detection • How to incorporate into existing practices

Figure 1: Implementation Procedure

Risks

Civil Liberties

The protection of civil liberties and civil rights should be of the foremost importance. There is concern that, if not properly created and trained, the tools could overreach the agency's guidelines and statutory authorities. The involvement of ODNI CLPT from the start of the implementation process will ensure the civil liberties of all Americans are protected. Consultation with the DOJ and OLC could also provide additional insight and guidance.

Time for Full Implementation

Despite the time saved, there is a high-risk factor. It will take time for the machine to learn what it is doing and improve itself. Projections suggest it takes approximately five years for any AI/ML tool to operate almost on its own with little to no oversight (Harding, 2022). Given this timeframe for full implementation and success evaluation, it is a high risk. However, the long-term benefits for analysts are worth the risk. Projections estimate that systems using AI can save analysts more than 45 days per year (Mitchell et al., 2019). Analytical tradecraft must evolve to embrace AI as the way of the future and put in the front-facing leg work to make it successful in the long-term.

Data Security

Secondly, the vast amount of data this alternative would generate would need to be stored on servers. As previously mentioned, because the information is already public, the server does not need to be classified. The information could be stored on the cloud or in 5G slices, both of which offer high security. Finished intelligence products would remain on classified servers, but the raw data could be placed on the cloud or 5G slide. This would result in decreased costs by not having to increase the number or capacity of classified servers. The risk associated with the idea is that it would be easier for adversaries to obtain our OSINT information. The security perimeter offered by the cloud significantly decrease the initial risk assumed.

Conclusion

As attacks against the US are realized, planned, carried out increasingly using the internet, the IC must adapt to counteract this threat. AI and ML provide the IC with the opportunity to harness the global explosion of data in a way that improves the protection of the American people and secures the Homeland. It will help further the American foreign policy mission, capture the intelligence objectives of each agency, and address the rise in violent domestic extremism. There is never a one size fits all approach, especially when it comes to intelligence collection, but the inclusion of OSINT as main "INT" brings the us one step closer to ensuring there are never mysteries the intelligence community cannot solve.

Appendix A. Outcomes Matrix

Criteria	Policy Alternatives			
		OSIA	**Maintain OSE**	**AI/ML**
	Constitutionality	High Already existing collection practices	High Existing	Medium Analyzes only already public information
	Cost	Very high Creation of a new agency. Medium risk	High Failure to address problem. Unknown budget	High Significant risk & cost to build. Saves time down the line
	Political Feasibility	Low Momentum lost from 2016 election	High Existing	Low Congress opposes risks
	Administrative Feasibility	Medium IC bias towards classified sources	Medium Existing but unknown operations	Medium Analyst oversight in beginning stages

References

About Bellingcat. (2022). Bellingcat. https://www.bellingcat.com/about/

Aftergood, S. (2015, October 28). Open Source Center (OSC) Becomes Open Source Enterprise (OSE). *Federation Of American Scientists*. https://fas.org/blogs/secrecy/2015/10/osc-ose/

Ashley, B., & Wiley, N. (2021, July 16). *How the Intelligence Community Can Get Better at Open Source Intel*. Defense One. https://www.defenseone.com/ideas/2021/07/intelligence-community-open-source/183789/

Berger, J. M., & Morgan, J. (2015, March). *The ISIS Twitter Census*. https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf

Congressional Oversight of Intelligence: Background and Selected Options for Further Reform (No. R45421; p. 17). (2018). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R45421

FBI Director Christopher Wray Testifies on U.S. Capitol Attack | C-SPAN.org. (2021, June 15). https://www.c-span.org/video/?512552-1/fbi-director-christopher-wray-testifies-us-capitol-attack&live

FBI Domestic Investigations and Operations Guide (DIOG). (2016). [Folder]. FBI. https://vault.fbi.gov/FBI%20Domestic%20Investigations%20and%20Operations%20Guide%20%28DIOG%29

George, R., & Rishikof, H. (2011). *The National Security Enterprise: Navigating the Labyrinth*. Georgetown University Press.

Glasser, S. B. (2005, November 25). *Probing Galaxies of Data for Nuggets*. http://www.washingtonpost.com/wp-dyn/content/article/2005/11/24/AR2005112400848.html

Harding, E. (2022, January). *Move Over JARVIS, Meet OSCAR*. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/220119_Harding_MoveOverJARVIS_MeetOSCAR_0.pdf?NqfrbU05ULzzcySzNHB0pTzsNYw3HdfK

Harding, E. (2022, February 15). *The Biggest Threat to Not Using OSINT* [Personal communication].

Holder v. Humanitarian Law Project, 561 U.S. 1 (2010), No. 08-1498 (United States Supreme Court June 21, 2010). https://supreme.justia.com/cases/federal/us/561/1/

Hudson. (2021, January 8). *Does the First Amendment Protect Trump on Incitement to Riot?* First Amendment Watch. https://firstamendmentwatch.org/does-the-first-amendment-protect-trump-on-incitement-to-riot/

Jurecic, Q. (2021, June 29). *Why Didn't the FBI Review Social Media Posts Announcing Plans for the Capitol Riot?* Lawfare. u

Milton, D. (2016, October 10). *Communication Breakdown: Unraveling the Islamic State's Media Efforts*. West Point. https://www.ctc.usma.edu/v2/wp-content/uploads/2016/10/ISMedia_Online.pdf

Mitchell, K., Mariani, J., Routh, A., Keyal, A., & Mirkow, A. (2019, December). *The future of intelligence analysis*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/public-sector/artificial-intelligence-impact-on-future-intelligence-analysis.html

Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., Cristofaro, E., & West, R. (2021, October). *Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels*. https://dl.acm.org/doi/pdf/10.1145/3476057

H.R.3162 - 107th Congress (2001-2002): Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT)

Act of 2001, 3162, House of Representatives, 107th (2001). https://www.congress.gov/bill/107th-congress/house-bill/3162

The 9/11 Commission Report (p. 567). (2004). United States Government. https://govinfo.library.unt.edu/911/report/911Report.pdf

USA PATRIOT Act—Reauthorizations. (n.d.). Encyclopedia Britannica. Retrieved October 29, 2021, from https://www.britannica.com/topic/USA-PATRIOT-Act

Weinbaum, C. (2021, April 12). *The Intelligence Community's Deadly Bias Toward Classified Sources*. https://www.rand.org/blog/2021/04/the-intelligence-communitys-deadly-bias-toward-classified.html

Worldwide Threats, U.S. House of Representatives, 10 (2020) (testimony of Christopher Wray). https://homeland.house.gov/imo/media/doc/Testimony%20-%20Wray.pdf