INTELLIGENCE AND NATIONAL SECURITY ALLIANCE

JANUARY 2022

Strategies for Addressing Bias in Insider Threat Programs

Presented by
INSA'S INSIDER THREATS SUBCOMMITTEE

Building a Stronger Intelligence Community

INTRODUCTION

Detecting and responding to insider threats involves both human decision-making processes as well as computational machine processes. Humans depend on their senses and their ability to interpret their environment, whereas machines depend on data to drive their analytic processes and outcomes. Bias, defined as favoring one thing, person or group compared with another, usually in a way that is unfair, interferes with decision-making.

Both human and machine processes may reflect implicit or explicit biases that erode the efficiency and accuracy of Insider Threat Program data (InTP). InTPs may generate less accurate outputs when the data used to inform organization risk decisions is biased, meaning the data is not representative of the population/person of interest, or when personal/situational factors influence conclusions made by decision makers. Identifying the "who, what, where, when, and why" of InTP bias promotes debiasing, or removal of bias, which leads to effective and just mitigation strategies and an increase in InTP mission success.

> Bias is a pattern of decision-making that favors one group, person, or thing over another, while unfairly discriminating against the remainder of the choices.

This white paper provides an overview of sources of bias, the ways bias can reduce InTP effectiveness, and the impacts and risks associated with biases [See Figure 1]. It also discusses the benefits of addressing bias and provides resources for more information.

This paper will discuss how various types of bias affect the major components of insider threat programs identified in the National Insider Threat Task Force (NITTF) Maturity Framework: InTP Leadership, Program Personnel, Employee Training, Access to Information, Monitoring User Activity, and Information Integration, Analysis & Response.

SOURCES OF BIAS AFFECTING INSIDER THREAT PROGRAMS

The sources of bias affecting Insider Threat Programs (InTP) can be broadly attributed to people and technology. People, including InTP leadership, program personnel, and other organizational staff, can introduce biases at multiple levels through their own personal cognitive biases. Organizations also introduce bias in a broader, more systemic ways through processes such as hiring practices and personnel decisions, often made outside of the InTP. Similarly, technology biases appear at multiple levels and may be associated with Access to Information (data), Monitoring User Activity (models), outcome metrics, and visualization strategies.

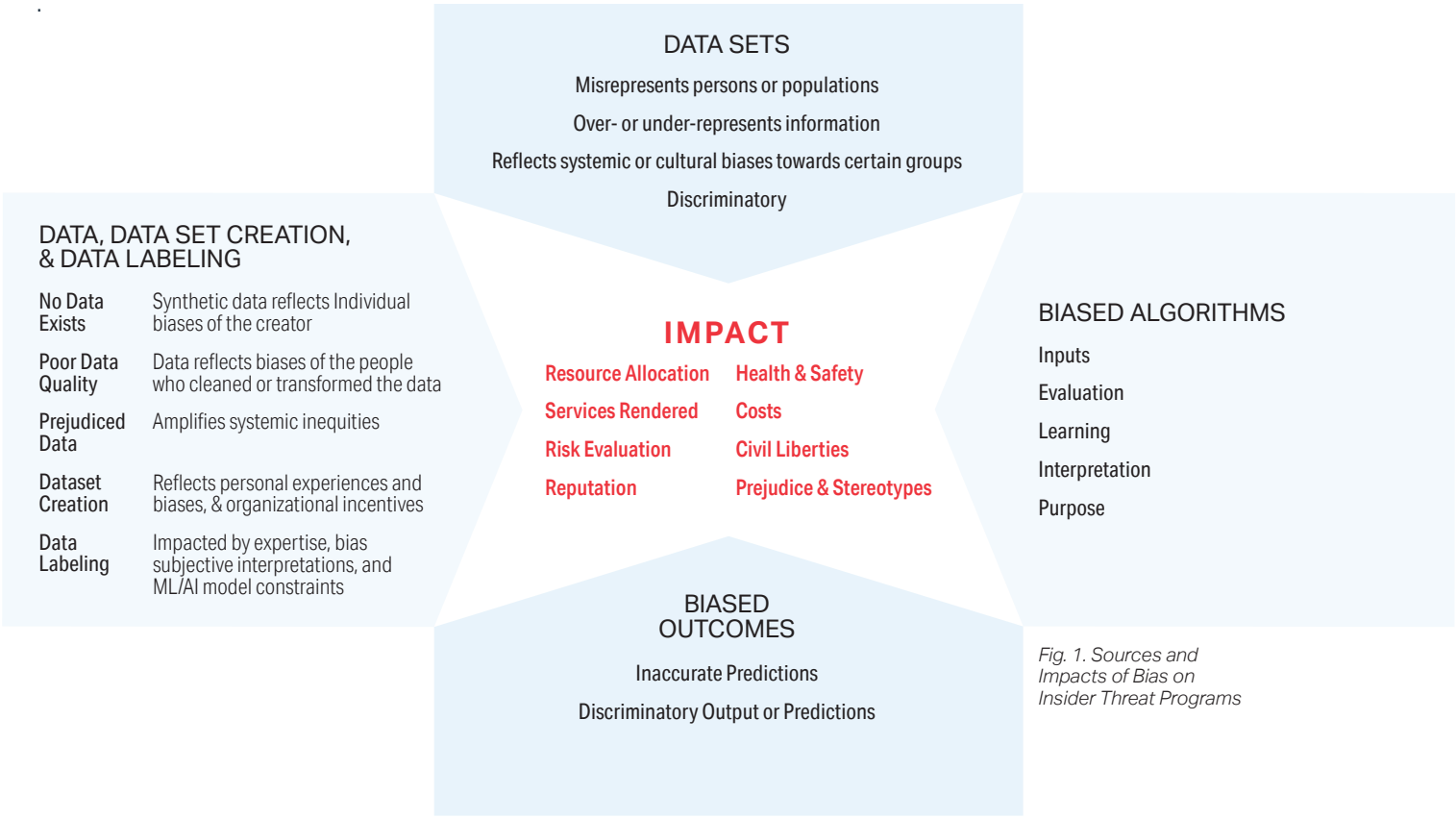


Fig. 1. Sources and Impacts of Bias on Insider Threat Programs

TYPES OF BIAS ASSOCIATED WITH PEOPLE

Individuals at every level of InTPs bring a host of personal biases that reflect their expertise, experience, aptitude, personality, and more. We are shaped by our environment, experiences, genetics, and individual cognitive ability. The combination of these things helps us process and make sense of our world. This means that our decisions, behaviors, and experiences are influenced by the past and the present.

Explicit biases reflect issues such as overt racism or sexism – in which the individual is aware of and actively engaged in a decision strategy that reflects a personal goal, interest, or belief that is harmful to others. Implicit biases, in contrast, often act in ways that are difficult to detect. A major source of implicit bias is attributed to individuals' cognitive biases.

While many cognitive biases exist,[1] the following are especially relevant to InTPs.

- **Availability Bias:** When engaged in evaluative thinking, what comes to mind most quickly is what a person deems to be most important or correct. InTPs that experience an uptick in a specific type of threat may overestimate the importance of that particular threat across the broader threat landscape. News cycles highlighting or sensationalizing specific types of threats can also impact what InTPs prioritize; press reports of a high-profile workplace shooting, for example, can make the need to mitigate this specific type of threat seem especially urgent.
- **Confirmation Bias:** People tend to search for and interpret information in a way that supports existing beliefs. These beliefs may be linked to a person under investigation or linked to other factors such as previous experience investigating similar cases.

- **Anchoring Bias:** When making decisions, people often use an early piece of information as a reference point or anchor for interpreting additional information. This bias impacts the estimates and decisions that are associated with numeric and non-numeric topics alike. Program personnel decisions may be negatively impacted when they inadvertently focus too heavily on preliminary data or on data points they see as relevant, thereby hindering their ability to update their analysis when presented with additional findings. For example, once InTP personnel establish that an employee has a great deal of debt, they may view all other suspicious information as somehow deriving from financial troubles, even if data shows that he pays his debts on time and leads a modest lifestyle.
- **Authority Bias:** InTP leadership and program personnel's professional opinions and investigation strategies are susceptible to authority bias. Particularly in hierarchical organizations like the military, authority figures' opinions may be taken as facts, and their requests may be followed with little or no hesitation. For example, when a senior leader of the InTP challenges findings, investigators can see such inquiries as indicators that the case is not credible. Biased decision-making can include erroneous decisions due to a leader's faulty opinions or the hindering of an investigation due to misplaced loyalty and conformity. The degree to which a person is susceptible to authority bias can be impacted by factors such as culture, age, and gender.

These cognitive biases reveal that people often have difficulty when estimating probabilities, especially probabilities of uncommon events. For example, probability estimates are systematically biased such that rare events with extremely negative consequences are judged to be more likely to occur than neutral events.[2]

These biases also reflect how the human brain works, as it readily seeks and identifies patterns in the environment, conserves energy by taking shortcuts, and continuously integrates social and cultural information into a person's individual worldview. Finally, some circumstances or conditions of a given work environment might contribute to biased decision making. For example, cognitive biases may impact decisions of people who are fatigued, distracted, mentally overloaded, challenged by a complex case, beset with personal or job pressures, or experiencing a heightened emotional state.

Finally, as these biases contribute to individual decision-making, they also have broader implications at a systemic level. For instance, authority bias, where people tend to believe that organizational leaders and authority figures are more likely to be correct, can have a transformative impact on an organization's culture. If leaders convey a lack of concern for certain types of violations, or if they routinely violate certain policies themselves, their behavior may shift organizational culture by normalizing a lack of respect for policies and rules. Authority bias is one example of how individual biases can ultimately reshape organizational policies and culture.

CASE STUDY

Cognitive Bias in Assessing Threat: Known Actor

Sometimes, the identity of the subject of an investigation can lead to special treatment: An analyst may minimize the threat posed by a friend or colleague; a manager may overlook a violation by a team member to prevent embarrassment; a high-ranking individual may receive deference and "the benefit of the doubt" by investigators who respect the subject or fear damaging their career prospects by pursuing an inquiry too vigorously.

Addressing human biases requires understanding and removing sources of biases. For example, anonymization of monitoring data can remove identities and their possible influence on threat assessment decisions.

Mitigation Approaches:
Effectively addressing cognitive biases hinges on raising awareness through training and transparency, specifically in policies and on building structured processes. These processes should have checks and balances to mitigate the impact of bias across all organizational levels. For InTPs, a structured decision-making process that integrates strategies, such as purposeful questioning to explore alternatives, can help.

Furthermore, masking or anonymizing identities can also help mitigate the impact of biases that surface at the systemic level, such as during hiring processes or during investigations of high-profile insiders. Identities should be masked until after an initial decision is made to assess an alert or issue. This will help ensure compliance with federal laws and regulations intended to prevent workplace discrimination. Anonymization often requires investment in technology that can automate the process.

TYPES OF BIAS ASSOCIATED WITH TECHNOLOGY

Biases that may be formalized by technology or technological advances can be manifested in data (Access to Information) or in models (Monitoring User Activity).

DATA

Data collection – critical to an InTP's Access to Information component – enables an understanding of concerning behaviors associated with individuals or a specific actor of interest. Data collection also allows InTP program personnel to interpret contextual events surrounding potential insider threat behaviors. The reliability and validity of these data sources heavily influence the ability of an InTP to identify insider threats and mitigate their negative impacts.

The following biases apply to data collection and processing:

- **Selection Bias:** When data is selected for analysis in a potentially subjective manner, it can be said to be affected by a selection bias. The inclusion of certain types of data in InTP analyses and exclusion of other sorts of data can introduce a selection bias into insider threat programs. For example, the collection of arrest records during background checks may introduce bias into employee profiles given substantial evidence of racial disparities in arrests.[3] Similarly, analysis of travel to countries of concern can flag individuals of certain ethnicities who have innocuous family ties to such nations and thus travel there frequently.

- **Data Availability:** A technical form of Availability Bias, the lack of different types of data can bias models and threat assessments by focusing more heavily on a specific type of behavior. For example, some InTPs rely heavily on technical indicators derived from readily obtained network/workstation monitoring tools, excluding less easily available data relating to behavioral factors (e.g., insights from co-workers and human resources).

- **Labeling Data:** The process of labeling data is subjective and is therefore subject to cognitive biases. For example, someone labeling workplace incident data may categorize sexual harassment as a personality conflict instead of criminal behavior, or an organization may lack an appropriate label for incidents related to mental health. Models and algorithms trained on labeled data may reflect biases introduced during labeling.

- **Data Analysis:** Understanding and evaluating data, including what numbers or values represent, and the types of numbers or values that are being analyzed, is critical for establishing an effective analytic strategy. Interpretation of the level of risk – e.g., likelihood of an indicator being associated with an insider threat, and the associated judgment of its importance – is subject to human cognitive biases and cultural influences. For example, from the 1940s until 1995, the U.S. government considered homosexuality to present a security risk because of societal biases and fear of communist influence even though no linkage between sexual orientation and espionage was ever found to exist.[4]

– **Program Personnel Expertise:** Decisions made regarding how and why to analyze specific data points are influenced by individual and group biases. Imagine a team of analysts with many years of experience analyzing logs, but very little experience analyzing disparate data sources that require more advanced statistical strategies. Through no fault of their own, the team may choose to analyze data in a way that is familiar, disregarding data that may be more relevant.

CASE STUDY

Technology Bias: Data Availability Bias

Bill is the head of his organization's new insider threat program and is tasked with implementing User Activity Monitoring (UAM). The cybersecurity department has a large amount of data readily available, which he assesses. The HR department also has relevant data, but it is neither centralized nor in a consistent format. Since Bill is under pressure to implement a UAM solution quickly, he leverages the cybersecurity data and recommends that HR make its data available in a usable format.

While the program focuses on investigating the threats it can detect, six months later a former employee of Bill's organization makes a public disclosure of printed files that he removed from the building. The disclosure damages the organization's competitive edge and reputation. This individual's HR data contained several valuable predictive indicators of this behavior, including details related to his resignation, troublesome performance reviews, past disciplinary action, and concerning findings in the individual's pre-employment background check. However, because the HR data was not formatted and centrally stored, it was not used in the InTP's risk assessment models.

Organizations should ensure balance in whole-person monitoring by investing in physical access, HR, financial, and behavioral data in addition to the readily available cybersecurity data. By focusing on cybersecurity data, the organization (and its models) will become skilled at assessing technical risks and vulnerabilities but be poorly positioned to evaluate physical and behavioral insider threats.

Mitigation Approaches

Biases in data can be mitigated by ensuring that the data used comes from reputable sources, has been independently validated, and is itself not overly focused on particular demographic groups. In addition, evaluating data from a wide range of sources helps dilute the influence that any single data source, which may reflect bias, has on an overall risk profile. The risks that come from data availability bias can be countered by ensuring consistent data collection, or by compensating for inconsistent data collection in the model. The challenges associated with data labeling can be mitigated in three ways: engaging a diverse group of data engineers and scientists to undertake the task, scheduling routine audits of data labeling by multiple personnel, and writing a standard operating procedure (SOP) that management reviews on an annual basis.

Ethical hiring practices are essential for mitigating data analysis bias and other personnel-oriented sources of data bias (such as data labeling and interpretation). Hiring InTP staff with varied personal backgrounds, professional experience, and relevant expertise makes it more likely that InTPs will identify and mitigate InT practices that inappropriately weigh certain information or criteria.

“Biases in data can be mitigated by ensuring that the data used comes from reputable sources, has been independently validated, and is itself not overly focused on particular demographic groups.

MODELS

With advancement of technology, the deployment of artificial intelligence (AI) solutions[5] promises to improve efficiency and effectiveness of insider threat tools, supporting the User Activity Monitoring and Information Integration, Analysis, & Response components of InTPs. As with all new technologies, in deploying AI to enhance mission effectiveness, we must understand how to use it in ways that align with our principles and prevent unethical outcomes.[7]

> Bias in insider threat programs can have immeasurable costs in a variety of ways, including decreased effectiveness, legal liability, reputation, and retention.

Biased models present two primary risks: inaccurate predictions and discriminatory results. For InTPs that use a risk model as recommended by the National Insider Threat Task Force (NITTF), inaccurate predictions can lead to false positives and false negatives, putting the program itself at risk. Furthermore, false alerts can expose the organization to legal liability.

The following modeling biases can contribute to negative program outcomes:

- **Models Trained on Biased Data:** A model can give inaccurate predictions even if the computational approach is not biased. For example, if the model excludes certain types of data, it may systematically miss certain types of threats.
- **Model Repurposing:** Models developed for one task are often repurposed for different contexts or applied to a different population. In these situations, the model will likely provide inaccurate results. For example, a credit score, designed to assess the risk that someone will default on a debt, now serves as a proxy for employment screening and can affect access to services like utilities and cell phone contracts. Similarly, an algorithm built to detect one type of insider threat may not be effective at detecting a different type of insider threat.[7] In other words, "If all you have is a hammer, everything looks like a nail."
- **Biased Developers:** Both developers and domain experts can inadvertently code their biases into models used by InTPs. This process is a technical implementation of the types of cognitive biases discussed above.
- **Model Interpretation Biases:** The program personnel who interpret model output will apply their cognitive biases to this information. Individuals also vary in experience, perspective, and moral judgement.

CASE STUDY

Algorithmic Bias Example: Repurposing an Algorithm

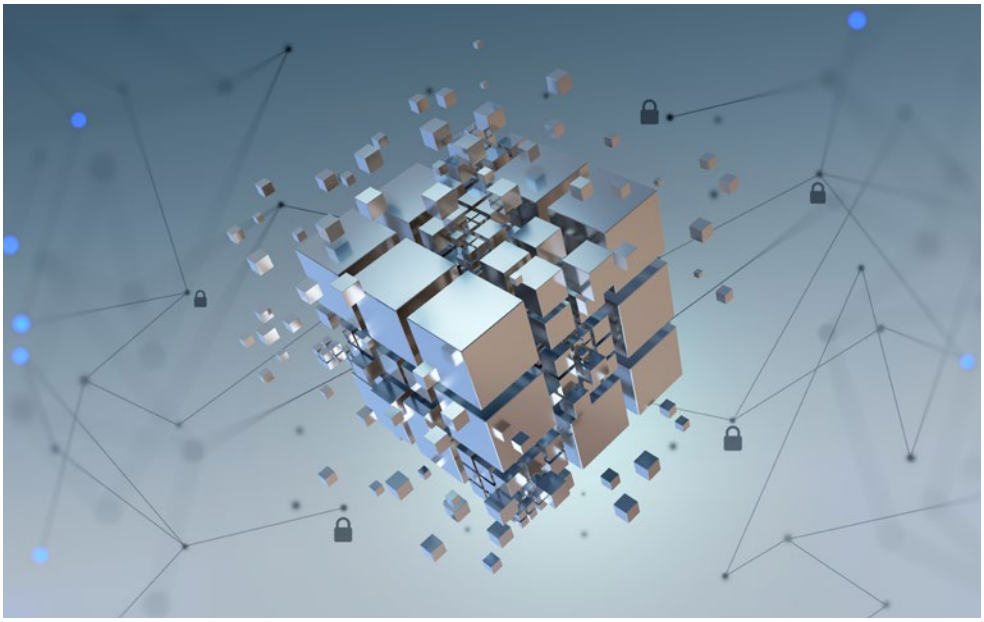
An organization has a robust algorithm to detect data exfiltration prior to termination. Because this algorithm was trained on real data, assessed actions electronically attributed to individual users, and demonstrated past success, executives' confidence in its ability to detect insider threats was high. Leadership thus expanded the InTP's mission to include physical theft and provided it access to inventory and access information. However, in a warehouse/retail environment in which multiple people have physical access to goods, the InTP – whose algorithm was trained on information theft data – was unable to detect physical theft. InTP algorithms cannot easily be repurposed for different types of threats.

Mitigation Approaches: Key Takeaways

Make models transparent in order to overcome technical sources of model bias. Transparency enables the organization to perform internal and external audits of the algorithm, extensive testing, and periodic re-training of models.

Use caution in repurposing analytic models. Models developed for one threat or in one context may generate inaccurate results when applied to different circumstances.

Ensure diversity in hiring practices to mitigate personnel-oriented sources of model bias. InTP managers should hire developers with a wide range of personal and professional backgrounds, ensure the proper domain experts are involved in model development, and ensure that program personnel have a wide range of appropriate expertise.



THE IMPACT OF BIASES ON INSIDER THREAT PROGRAMS

Bias in insider threat programs can have immeasurable costs in a variety of ways, including decreased effectiveness, legal liability, reputation, and retention.

CASE STUDY

"It couldn't be one of ours!"[8]

An FBI investigation and mole hunt focused on CIA officer Brian Kelley for years, when it was in fact senior FBI agent Robert Hanssen who had been committing espionage. The Brian Kelley investigation consumed significant electronic and physical surveillance resources as well as agents and analysts from the Intelligence Community. The FBI reasoned that previous spies like Aldridge Ames had been CIA employees, and many within the Bureau leadership ranks did not believe anyone from the Bureau would betray their country. FBI agents were so convinced of Kelley's guilt that they viewed exculpatory information – such as Kelley's passing a polygraph and reporting a false flag ruse to entrap him – as "proof" that Kelley was smart and manipulative enough to be the perfect spy. This bias caused the FBI to fail to identify the insider threat for two years, diverted investigative resources, and resulted in reputational harm to both the Bureau and Brian Kelley.

DECREASED MISSION EFFECTIVENESS

A biased InTP creates risks that an organization will miss an insider threat or fail to identify a threat in time to mitigate it. Despite the organization's efforts, it could suffer loss of intellectual property, financial losses, reputational harm, or other serious consequences that undermine its mission, capabilities, and finances.

PURSUING THE WRONG RISKS

When an InTP is plagued with the types of bias discussed in this whitepaper, it may tend to focus on lesser threats or divert attention to non-threatening activity. This could leave the organization vulnerable because a true insider threat can operate undetected or take advantage of the biased program to evade detection. In the worst case, focusing on the wrong threats while ignoring real threats can result in loss of life (such as by exposing foreign sources to hostile intelligence services) and highly sensitive information.

Biased judgment and decision-making impacts metrics, standards, and thresholds utilized by the program. Consequently, algorithms themselves may become ineffective, or the program's status may become misrepresented. These issues can affect the integrity of an investigation and cause the organization to lose its ability to prosecute wrongdoing. Prosecution can serve as a significant deterrent to insider threat behaviors as well as an avenue for the organization to recover financially from the incident and learn about vulnerabilities that were exploited.

IMPROPER RESOURCE ALLOCATION

Bias undermines the effectiveness of InTPs by encouraging investigators to assign finite resources to pursue unsubstantiated leads at the expense of more promising ones. Minimizing bias, in contrast, promotes more efficient stewardship of resources by allowing them to be applied to assess evidence-based indicators of risk.

Bias undermines the effectiveness of InTPs by encouraging investigators to assign finite resources to pursue unsubstantiated leads at the expense of more promising ones.

ETHICS AND COMPLIANCE

InT-related liability risks are rooted in a failure to conform to ethical values. A culture of ethics and compliance helps avoid liability – and, not incidentally, increases workforce trust in the program.

LIABILITY

InTP bias can expose an organization to legal liability. For example, if an organization utilizes an algorithm that unintentionally discriminates against a protected class – members of an ethnic group or religious affiliation, for example – the organization could be in violation of Equal Employment Opportunity Commission (EEOC) regulations. If the process – or algorithm – has a discriminatory effect, the employer's liability may be significant.[9] "Black box" algorithms that are closed to scrutiny pose particular concern, as they are difficult to assess and evaluate. Conversely, properly implemented models can decrease the variations inherent in subjective judgements by program personnel.

REPUTATION OR NEGATIVE BRAND IMPACT

An organization's reputation could be damaged by a lawsuit or by employee perceptions of a biased security infrastructure. This can result in negative press attention, falling stock prices, and loss of investors, customers, and public trust – all of which have direct financial consequences to the organization that may linger for years.

HIRING AND RETENTION FOR INSIDER THREAT PROGRAMS

HIRING

Bias in hiring, particularly within certain fields, is a well-documented challenge.[10] Organizations must endeavor to build the right InTP team – including leadership, personnel who analyze data and model output, and developers who build models and label data. Biased data, analysis, or decisions by InTP staff may result in false positives that unintentionally result in illegal discrimination and thus leave the organization vulnerable to legal liability. Alternatively, biased assessments may result in threats not being taken seriously enough, which exacerbates risk. Teams that are composed primarily of personnel from a specific InT domain – such as cybersecurity, counterintelligence, or law enforcement – may not be equipped to interpret output from other corporate stakeholders. If InTP staff do not represent a diverse range of experiences and perspectives and team members cannot be added or switched out to promote diversity, InTP managers would be wise to draw on detailees who can provide missing perspectives in the short-term.

RETENTION

An InTP that is perceived to be biased by employees, or which does in fact have built-in bias, can result in distrust by employees in other parts of the organization's security infrastructure and in the workforce as a whole. A lack of trust can impact the InTP's ability to collaborate effectively with other departments – such as HR, information technology, or legal – which would hinder the program's ability to view insider threats comprehensively. If the overall workforce perceives the InTP as biased or ineffective, employees may resent InT measures as intrusions, resist collaboration with InT and security programs, and foster resentment – all of which can increase the risk of insider threats. Employee respect and trust in the organization – which contribute to recruitment and retention – is hard-won but easily lost, and difficult to reconstitute.

CONCLUSION: BENEFITS TO COUNTERING SOURCES OF BIAS

Cognitive biases affecting human decision-making may lead to inaccurate estimates of insider threats. They can influence incidence baselines, perceptions, and analyses; drive interpretations of data in ways that unduly align with preconceived beliefs; and over-emphasize information/experiences that do not accurately reflect reality.

Organizations must be vigilant to avoid – or at least mitigate – the impact of cognitive bias in their insider threat programs. Key steps to take include:

- Promote awareness of the sources of bias. Organizations can provide awareness training to raise staff understanding of sources and impacts of bias on decision-making.
- Incorporate the use of transparent and structured decision-making tools and methods to avoid, overcome, or at least mitigate the impact of biases.
- Anonymize data during the threat assessment process to further reduce the impact of bias.
- Hire a diverse insider threat team that brings a multidisciplinary approach to threat assessment, thereby moderating the inherently subjective perspectives of individuals who comprise the team.

> An InTP that is perceived to be biased by employees, or which does in fact have built-in bias, can result in distrust by employees in other parts of the organization's security infrastructure and in the workforce as a whole.

Cognitive biases prevent individuals and organizations from accurately understanding reality even when all the needed data and evidence that would form an accurate view is at hand. Pherson and Heuer argue that simple awareness of bias does not, produce more accurate analysis. Experience has shown that it is exceedingly difficult to overcome the tendency to fall victim to confirmation bias or data availability bias without using a specific tool or technique.

Security professionals need to employ methods that help them challenge their assumptions, help identify alternatives, and structure uncertainties. Techniques that help analysts think critically about their evidence and conclusions must be integrated into the basic analytic process. Techniques Intelligence Community analysts cite as most helpful are Cluster Brainstorming, Key Assumptions Check, Analysis of Competing Hypotheses, Indicators Generation and Validation, and Premortem Analysis.[11]

RECOMMENDATIONS

A. IDENTIFY SOURCES OF BIAS

1. Raise awareness of bias through training and transparency, specifically in policies and on building structured processes. Ensure processes have checks and balances to mitigate the impact of bias across all organizational levels.

2. Mask or anonymize identities to help mitigate the impact of biases during hiring processes or during investigations of high-profile insiders. To facilitate anonymization, consider investing in technology that can automate hiring or InT assessments.

B. MITIGATE DATA BIAS

1. Hire InTP staff with varied personal backgrounds, professional experience, and relevant expertise to mitigate biases in data labeling and analysis, bring multiple perspectives into InTP operations, and prevent groupthink.

2. Ensure that data comes from reputable sources, has been independently validated, and is not overly focused on particular demographic groups. Evaluate data from a wide range of sources to help dilute the influence that any single source has on an overall risk profile.

3. Engage a diverse group of data engineers and scientists to undertake the task of labeling data, schedule routine audits of the labeling by multiple personnel, and write a detailed data labeling standard operating procedure (SOP) that management reviews on an annual basis.

C. MITIGATE MODEL BIAS

1. Make models transparent to overcome technical sources of model bias. Transparency enables the organization to perform internal and external audits of the algorithm, extensive testing, and periodic re-training of models.

2. Use caution in repurposing analytic models. Models developed for one threat or in one context may generate inaccurate results when applied to different circumstances.

3. Ensure diversity in InTP staff to mitigate personnel-oriented sources of model bias. InTP managers should hire developers with a wide range of personal and professional backgrounds, ensure the proper domain experts are involved in model development, and ensure that program personnel have a wide range of appropriate expertise.

APPENDIX: RECOMMENDATIONS FOR ADDRESSING BIAS IN INSIDER THREAT PROGRAMS

APPROACH	OBJECTIVE	IMPACT
AWARENESS TRAINING	Raise awareness about types of bias, susceptibility to bias, and actions that can be taken to mitigate negative impacts of bias	Much bias is "unconscious" – attitudes that are held subconsciously and affect the way we think and feel about others. Through training, people can learn to examine their reasoning strategies or to identify situations where they may be susceptible to biased decision-making.
STRUCTURED DECISION-MAKING	Provide tools and techniques to make decisions in complex environments involving uncertain data, and to avoid allowing bias to influence decisions.	Structured decision-making forces people to slow down, consider alternative solutions, and to carefully question their initial choices. This is critical for fostering environments that support analytic decision-making.
USE OF ANONYMIZED OR PSEUDO-ANONYMIZED DATA	Removing associations between the data and the individuals generating the data helps to decrease the chances that demographic information will influence decision-making.	Anonymizing or pseudo-anonymizing data to remove identifying/ demographic information tied to individuals decreases the chance that analysts will be susceptible to a range of biases, such as name-based racial biases or biases associated with knowing the person under investigation.
PEER REVIEW TO ESTABLISH APPROVED OPERATIONAL DEFINITIONS *(E.G., FOR DATA LABELING)*	Decrease the impact of individual bias on labeling of data and on decisions made based on the data. Reduce the subjective nature of data labeling through diversity in threat analysis teams and standardized methods for labeling and adjudicating incident data.	Reduce the likelihood that one individual's biases will impact threat assessment or mitigation decisions.
IDENTIFY ROLES WITH HEAVY OR PROLONGED WORKLOADS *AS WELL AS ROLES WITH POOR WORKING CONDITIONS*	Identifying high-stress, or high-risk, roles and actively addressing the factors contributing to the poor work environment can reduce external influences that increase biased decision-making.	Reducing external factors contributing to biased decision- making allows for other mitigation techniques to be more meaningful and effective.

References

[1] See John Manoogian, III, "The Cognitive Bias Codex." At https://commons.wikimedia.org/wiki/File:The_Cognitive_Bias_Codex_-_180%2B_biases,_designed_by_John_Manoogian_III_(jm3).png

[2] A. J. L. Harris, A. A. Corner, and U. Hahn, "Estimating the Probability of Negative Events," Cognition, vol. 110, no. 1, pp. 51-64, 2009

[3] See, for example, Benjamin Mueller, Robert Gebeloff, and Sahil Chinoy, "Surest Way to Face Marijuana Charges in New York: Be Black of Hispanic," New York Times, May 13, 2018. At https://www.nytimes.com/2018/05/13/nyregion/marijuana-arrests-nyc-race.html. For articles about racial disparities in arrests in localities around the United States, see https://news.google.com/search?q=racial%20disparities%20in%20arrests.

[4] Government Accountability Office, Security Clearances: Consideration of Sexual Orientation in the Clearance Process, GAO/NSIAD-95-21, March 24, 1995, p. 15. At https://www.gao.gov/assets/nsiad-95-21.pdf. See also, "An interview with David K. Johnson, author of The Lavender Scare: The Cold War Persecution of Gays and Lesbians in the Federal Government," 2004. At https://press.uchicago.edu/Misc/Chicago/404811in.html.

[5] The John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115- 232, 132 Stat. 1636, 1695 (Aug. 13, 2018) defines AI to be any artificial system (1) that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets; (2) that was developed in computer software, physical hardware, or another context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action; (3) that was designed to think or act like a human, including cognitive architectures and neural networks; (4) that applies techniques, including machine learning, designed to approximate a cognitive task; that was designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision-making, and acting.

[6] Office of the Director of National Intelligence, "Artificial Intelligence Ethics Framework for the Intelligence Community, v. 1.0," June 2020. At https://www.odni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf.

[7] Intelligence and National Security Alliance, Categories of Insider Threats, 2019. At https://www.insaonline.org/wp-content/uploads/2019/10/INSA_WP_Categories_of_Insider_Threats-1.pdf.

[8] Mary-Jayne McKay, "To Catch a Spy: Probe to Unmask Hanssen Almost Ruined Kelley," 60 Minutes, January 30, 2003. At https://www.cbsnews.com/news/to-catch-a-spy-30-01-2003/.

[9] Chris Opfer, "AI Hiring could mean Robot Discrimination will Head to Courts," Bloomberg Law, November 12, 2019. At https://news.bloomberglaw.com/daily-labor-report/ai-hiring-could-mean-robot-discrimination-will-head-to-courts.

[10] Kathryn M. Neckerman and Joleen Kirschenman, "Hiring Strategies, Racial Bias, and Inner-City Workers," Social Problems, vol. 38, no. 4 (Nov. 1991), pp. 433–447. At www.jstor.org/stable/800563; Tessa E.S. Charlesworth and Mahzarin R. Banaji, "Gender in Science, Technology, Engineering, and Mathematics: Issues, Causes, Solutions." Journal of Neuroscience, Vol. 39, No. 37 (September 11, 2019), pp. 7228-7243. At https://www.jneurosci.org/content/jneuro/39/37/7228.full.pdf; W. Drew Gouvier, Sara Sytsma-Jordan, and Stephen Mayville, "Patterns of discrimination in hiring job applicants with disabilities: The role of disability type, job complexity, and public contact," Rehabilitation Psychology, Vol. 48, No. 3, (2003), pp. 175–181. At https://doi.org/10.1037/0090-5550.48.3.175.

[11] Detailed descriptions of these techniques and instructions for using them can be found in Randolph H. Pherson and Richards J. Heuer, Jr., Structured Analytic Techniques for Intelligence Analysis, 3rd ed., (Thousand Oaks: CQ Press/SAGE Publications, 2021).

Resources

Page 5 – Alicia Nortje, "Cognitive Biases Defined: 7 Examples and Resources," Positive Psychology, August 30, 2021. At https://positivepsychology.com/cognitive-biases/

—

Page 6 – Mitigating Bias in Artificial Intelligence, at https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/.

Algorithmic Justice League, at www.ajl.org

INSA, A Preliminary Examination of Insider Threat Programs in the U.S. Private Sector, September 2013. At https://www.insaonline.org/wp-content/uploads/2017/04/INSA_InsiderThreat_WP.pdf.

Page 7 – Mitigating Bias in Artificial Intelligence, at https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/.

Algorithmic Justice League, at www.ajl.org

INSA, An Assessment of Data Analytics Techniques for Insider Threat Programs, July 2018. At https://www.insaonline.org/wp-content/uploads/2018/08/INSA_Insider-Threat_Data-Analytics-July-2018.pdf.

INTELLIGENCE AND NATIONAL SECURITY ALLIANCE

ACKNOWLEDGEMENTS

INSA expresses its appreciation to the INSA members and staff who contributed their time, expertise, and resources to this paper.

INSA MEMBERS

Vinny Corsi, *IBM; Insider Threat Subcommittee Chair*

Sue Steinke, *Peraton; Insider Threat Subcommittee Vice Chair*

Julie Ard, *Noblis*

Dr. Margaret Cunningham, *Forcepoint*

Dr. Rajni Goel, *Howard University*

Dr. Frank L. Greitzer, *PsyberAnalytix*

Michael Hudson, *ClearForce*

J.T. Mendoza

Dr. Judy Philipson,
Applied Research Lab for Intelligence and Security (ARLIS)

INSA STAFF

Suzanne Wilson Heckenberg, *President*

John Doyon, *Executive Vice President*

Larry Hanauer, *Vice President for Policy*

Peggy O'Connor,
Director of Communications and Policy

Cassie Crotty, *Intern*

Ali Berman, *Intern*

ABOUT INSA

The Intelligence and National Security Alliance (INSA) is a nonpartisan, nonprofit trade association dedicated to advancing collaborative, public-private approaches to intelligence and national security priorities. Through the application of industry expertise, leading-edge academic research and commercial best practices, INSA seeks to make the Intelligence Community more effective and efficient. Our 160+ member organizations and its 4,000+ individual and associate members include senior executives and intelligence experts in the public, private and academic sectors.

ABOUT INSA'S INSIDER THREAT SUBCOMMITTEE

INSA's Insider Threat Subcommittee researches, discusses, analyzes, and assesses counterintelligence and insider threat issues that affect government agencies, cleared contractors, and other public and private sector organizations. The Subcommittee works to enhance the effectiveness, efficiency, and security of government agencies and their industry partners, as well as to foster more effective and secure partnerships between the public, private and academic sectors.

INTELLIGENCE AND NATIONAL SECURITY ALLIANCE

Building a Stronger Intelligence Community